

**Final Report:**

**An Electronic Digital Speech-Corpus  
and Searchable Dictionary/  
Database of the Tamil Verb**

பேச்சுத் தமிழ் மின்தொகுப்பு மற்றும்  
தேடும் வசதி கொண்ட மின்னகராதி / தமிழ்  
வினைகளுக்கான மின்தரவு

**Chief Editor and P.I.:**

**Harold F. Schiffman  
Department of South Asia Studies  
University of Pennsylvania**

**Associate Editors:**

**R. Vasu, Prathima Christdas,  
and S. Palanisamy**

**Contributing Editors:**

**Thanam Baba, K. Karunakaran, Ganesh Krishnamurti,  
C. M. Pradeep, and S. Radhakrishnan**

November 15, 2007

## 0.1 Introduction

### 0.1.1 Background.

This dictionary has been in preparation for at least 25 years, and has been the result of a collaboration between myself, as chief editor and various associate editors since 1978. It has benefited from the contributions of many people and has been funded by several different agencies, foundations, academic institutions, and endowments. We are deeply grateful for the support provided by the American Institute of Indian Studies, the National Endowment for the Humanities, the Smithsonian Institution, the Department of Education (International Research and Studies Program), the South Asia Language Resource Center (SALRC), the Consortium for Language Teaching and Learning, the University of Washington Graduate School Research Fund, the Penn Research Foundation, the South Asia Studies Department and South Asia Center at the University of Pennsylvania, and private sources.

From its earliest conception, this dictionary has involved the use of computers in the processing and editing of the data. Early attempts involved the use of punch-cards and processing on a mainframe, but after funding from the NEH was obtained in 1984, data entry began on a microcomputer and most processing has been done with what used to be known as micro- or mini-computers as the power of these machines has increased and the ability to do “desk-top” publishing has improved. Computers have also been used to develop an electronic Tamil font using METAFONT. Together with  $\text{\TeX}$  and  $\text{\LaTeX}$ , data originally input on a PC in romanized form has been converted to a form that can be photocomposed and printed totally by electronic means. The original plan for this dictionary was to publish it in book form, but now it is possible to deliver it in electronic form (on DVD, and an html versions) and the ‘printable’ form, is included here in PDF format.<sup>1</sup>

Computers have essentially changed the conception of what is a ‘text’ since the shape of a work can be changed almost instantly by altering the output, sometimes radically. Since dictionaries tend to try to capture generalities about the lexicon of a language at a particular point in time, and can never in any sense be considered *finished*, we have thought from the beginning that this first version of *An English Dictionary of the Tamil Verb* would not be the final or definitive version. We are painfully aware of its many inadequacies, and were it not for the feasibility of rapidly revising and improving the data-base to issue newer versions, we would not place it before the public at this time. Many dictionary projects have consumed decades of their compilers’ lives, and many compilers of dictionaries did not live to see their final product. We did not wish to emulate James Murray, the chief editor of the *Oxford English Dictionary*, who, despite devoting decades of his life to its editing, died before it was anywhere near completion. We therefore welcome *constructive criticism* of this product, so that revisions of it may be made that will improve it. We expect that in particular, that portion of it that is devoted to *spoken* Tamil will arouse the most commentary; for some, any use at all of spoken Tamil is considered to be contributing to the corruption of the language; for others, consensus on what is ‘standard’ spoken Tamil (SST; for more on this, see below) has not been reached, so what may be acceptable to one speaker may be of questionable acceptability to others.

This dictionary was undertaken because of a number of needs that were not being met by existing or previously-extant English-Tamil dictionaries. The main goal of this dictionary is to get an English-knowing user to a Tamil *verb*, irrespective of whether he or she begins with an English verb or some other item, such as an adjective; this is because what may be a verb in Tamil may in fact not be a verb in English, and vice versa. The web and DVD versions of this dictionary are searchable, so that if a particular English verb the user wants a Tamil equivalent for is not one of the main entries, inputting the search item should take the user to the English synonym file, which will give the user the Tamil verb. For example, we do not have a main entry for ‘pounce’ but this item does appear as a synonym for ‘jump, leap’, and some other verbs, so searching for ‘pounce’ will get the user to a Tamil verb. Our original conception was therefore to specifically concentrate on supplying the kinds of information lacking in all previous attempts at capturing

---

<sup>1</sup>Originally it was hoped that a print version of this dictionary could be published, but with demands for considerable subsidies for printing and publication, it was decided that an electronic version, originally conceived to fit on a CD, with sound files for the spoken examples, would be more feasible. Because of the size of the audio files, however, we have been forced to opt for a DVD version. A pdf version of the “paper” format will be included on the DVD’s that can be printed out, for users who prefer a paper version. And it is also planned to have an HTML version, with sound files, hosted on a website at the University of Pennsylvania’s Van Pelt Library (South Asia section: <http://www.library.upenn.edu/collections/sasia/>), at the Chicago Digital Dictionaries of South Asia site (<http://dsal.uchicago.edu/dictionaries/list.html>), and perhaps elsewhere.

the equivalencies between English and Tamil. In particular we have focused on the following problematical areas:

1. **Verb classes:** English-Tamil dictionaries, both current and previously extant, do not provide the user with any information about the *morphological class* of the Tamil verb, nor do they give information as to whether a verb is *transitive* or *intransitive*. This kind of information is readily available in Tamil-English dictionaries (for example, in Fabricius 1972), but not in English-Tamil dictionaries.<sup>2</sup>
2. **Spoken Tamil:** No English-Tamil dictionaries give information about the *spoken* or *colloquial* pronunciation of Tamil, or whether a verb found in Literary Tamil is also used in spoken Tamil.<sup>3</sup> Information about spoken Tamil is harder to get than Tamil in any other form, since no electronic databases exist for ST, and because many speakers of Tamil do not consider ST to be worth devoting any attention to; for non-Tamils attempting to learn Tamil, however, ST is necessary for day-to-day functioning in a Tamil environment, and this dictionary is intended to meet their needs, not primarily the needs of Tamil speakers.
3. **Example Sentences:** Currently extant English-Tamil dictionaries give few if any *example* sentences illustrating the morphological and/or syntactic frames that verbs occur in.
4. **Modern Usage:** Most extant English-Tamil dictionaries are now seriously out of date, since their compilers have often simply replicated the data found in previous dictionaries, with the result that the English represented in them is that of the 18th or 19th centuries; the Tamil forms given are also lacking in modernity, but for other reasons.
5. **Syntactic Complexity of the Verb Phrase:** Because the Tamil verb is morphologically complex, and the verb phrase therefore syntactically *very* complex, we decided to focus only on the Tamil verb. Tamil nouns are, in contrast, morphologically fairly simple and the noun phrase is remarkably uncomplicated—Tamil nouns have no gender distinctions (except where there is biological gender), no agreement, and no marking of adjectives as to number or gender. The Tamilnadu government has spent much time and energy creating lexica and glossaries for various modern usages for Tamil, but from what we can gather, these have mainly generated new nominal terminology, not verbs. This is partly because Literary Tamil cannot borrow verbs easily, i.e. it cannot take a 'foreign' word and add Tamil morphological material to it, such as tense marking and person-number-gender marking, which all Tamil finite verbs must have. (Spoken Tamil has no problem with borrowings or other innovative word-formation devices, but ST is, as already mentioned, not deemed worthy of being used in such contexts.)

The reasons for the above deficiencies have to do with the goal of such previous dictionaries. Most have been conceived as being primarily for the use of speakers of *Tamil* who wished to know or use English effectively, rather than the opposite. Tamil speakers of course already know intuitively (though not explicitly) the verb class and transitivity status of Tamil verbs, and how to use them in sentences. They also know spoken Tamil usage by virtue of having learned it as a first language, and do not need to be taught anything about it. They also use English primarily in its written form, with the result that modern British or American colloquial usage receives short shrift. This dictionary attempts to correct these shortcomings by providing verb-class and transitivity status for all Tamil entries, by giving examples of Tamil usage and their English equivalents in sentence-frames, and by providing spoken pronunciation and examples whenever a Literary Tamil form has a spoken equivalent. Sometimes these are not at all historically cognate—when a Tamil verb like கூறு III tr. is rendered in spoken Tamil, it must be replaced by சொல்லு III tr. *sollu*, since no Tamil speaker would ever produce a spoken equivalent of கூறு, \**kuuru* in his or her colloquial speech. As for English equivalents and translations of example sentences, we have tried to stay close to modern colloquial North American usage, which means that certain rare or archaic verbs that would be found in an English

<sup>2</sup>An exception is Rajaram 1986, which may be the first English-Tamil dictionary to give verb classes and transitivity specifications. The Cre-A *Tamil-Tamil Dictionary* (Subramanian 1992) also gives information about the form of the tense-markers of verbs, but does not give 'classes.'

<sup>3</sup>Rajaram 1986 gives the pronunciation of *Literary Tamil* verbs, but gives no information about ST pronunciation.

dictionary, such as ‘malt’ or ‘smite’ will not appear. We also give ‘informal’ equivalents as synonyms in some cases, and these are marked as ‘inf.’—such forms as ‘be blown away’ or ‘diss’ that may not even appear in English dictionaries yet. This also means that the English translation of the example sentences may not be a ‘word-for-word’ or literal translation of the Tamil, but rather a colloquial equivalent that captures the overall meaning.

A word about English synonyms is in order. The field (see below) that contains English synonyms may occasionally show an English verb that may not be the exact synonym of the main entry, or it may seem far-fetched or even inaccurate as an equivalent of the Tamil verb given. We have relied on two on-line lexical sources for this—the Oxford Reference Core of on-line English dictionaries (<http://www.oxfordreference.com/views/GLOBAL.html?authstatuscode=200>), and WordNet (<http://wordnet.princeton.edu/>). Users of this dictionary who disagree with our choice of English equivalents may therefore direct their objections to the compilers of these two sources, rather than to us. We present them primarily as suggestions of possible synonyms of related verbs.

### 0.1.2 The Structure of an Entry.

Entries in this dictionary are given in the following way:

1. **English entry:** the ‘head word.’
2. **Tamil equivalent**, (in Tamil script), in the form of the verb *stem* (usually identical to the singular imperative.) If more than one Tamil equivalent is appropriate, subsequent equivalents are given after items (3–5) below.
3. **Verb Class and transitivity specification:** If no Spoken Tamil form is given or if verb class is different from that of ST. Additionally, any idiosyncratic syntactic information, such as whether a verb requires a dative subject, or occurs only with third–person subjects, or are ‘defective’<sup>4</sup> or idiomatic, may also be given here. But we do not wish to complicate this field with too much grammatical or syntactic information, but rather have the user rely on the syntactic and morphological information given in the example sentence. In some cases we then refer the user to the discussion of this particular issue in a grammar of Tamil (Schiffman 1999) for further information.
4. **Spoken Tamil pronunciation** in the spoken form of the verb stem. If no spoken pronunciation is given, it means the Literary Tamil form is not used in ST. Rarely, a ST form is given without an Literary Tamil entry, meaning there is no Literary equivalent of the (usually borrowed) ST form.
5. **English definition(s)**<sup>5</sup>
6. **Additional Tamil entries** followed again by items 3–5 (above), if applicable.
7. **Example sentences** or phrases, in Literary Tamil, Spoken Tamil, and an English translation. The Spoken Tamil phrase is also provided with an *audio version*, which can be accessed on the DVD and HTML versions of the dictionary by clicking on the symbol for a loudspeaker.<sup>6</sup>
8. **Tamil Synonym(s)** or near-synonyms, where appropriate.

### 0.1.3 Pronunciation

As noted above, spoken equivalents of Literary Tamil Verbs are given in a transliterated form just after the LT entry in Tamil script. No pronunciation of LT forms is given; it is assumed that if one can read Tamil, one can pronounce it. When we fail to provide a Literary Tamil form with a spoken equivalent, it means that

<sup>4</sup>That is, do not have all the ‘principal parts’ that other verbs have.

<sup>5</sup>Again: note that we have used two on-line lexical sources for this—the Oxford Reference Core of on-line English dictionaries (<http://www.oxfordreference.com/views/GLOBAL.html?authstatuscode=200>), and WordNet (<http://wordnet.princeton.edu/>)

<sup>6</sup>The recording of the ST example sentences was made possible by the grant from the South Asia Language Resource Center, which was outsourced to a company called Audioworld, in Chennai.

to our knowledge this verb is not widely used in spoken Tamil,<sup>7</sup> or at least not in standard educated spoken Tamil. (Some dialects may in fact use some items, and their use by certain speakers might be accepted by others, but if this is the case, we are not aware of this.) We welcome suggestions by Tamil speakers or others knowing English and Tamil well, if they know of equivalents we have missed. We also welcome any or all *constructive* suggestions about equivalencies, about meanings, synonyms, idioms, nuances, or any other aspect of this work. Since we can change entries easily, we expect to be able to periodically issue (a) revised version(s) of it that can benefit from the knowledge of more English and Tamil-knowing people than just ourselves. As such it may truly become a collaborative effort of those who know and care about the Tamil and English languages to make it, in the end, a more truly useful work.

### 0.1.3.1 Standard Spoken Tamil.

The question of whether there exists a variety of spoken Tamil that is ‘standard’ is a somewhat difficult issue.<sup>8</sup> Many linguistic scholars have broached the issue and have various conclusions to offer; the consensus seems to be that a standard spoken Tamil, if it does not already exist, is at least ‘emergent’ and can be described as that variety that one hears used in the Tamil ‘social’ film, and on the radio and in the production of ‘social’ dramas, both live and on radio and television; it is the variety that emerges when educated speakers of various local and social dialects meet in college and university hostels in Tamilnadu and must, perhaps for the first time in their lives, speak a variety of Tamil that is understandable (and acceptable) to other Tamils from vastly different parts of the Tamil-speaking world. It is also the variety spoken by the hero and heroine in films, as contrasted with other characters in a movie, who may use other forms, often for comic or rustic effect. Speakers of Tamil learn in this way which forms are stigmatized and which are more acceptable, since lampooning or ridiculing the speech of certain characters evokes laughter from the audience. Tamils who wish to avoid being thus ridiculed therefore try to emulate speech forms that are acceptable and comprehensible to the largest number of speakers; this means avoiding regionalisms, caste-specific forms, rustic or vulgar forms, or anything stereotypical of a particular place or community. In recent years this kind of inter-caste, inter-regional dialect has most typically resembled the higher-caste, educated speech of non-Brahman groups in Tamilnadu; according to some it is neither from the far north (i.e. Madras/Chennai) or from the far southern reaches of Tamilnadu (e.g. Kanniyakumari District), but rather from urban areas in the more ‘central’ districts of Tamilnadu, such as Thanjavur, Trichy, or Madurai. In cases of doubt as to whether a form is acceptable or not, speakers tend to lean more toward Literary Tamil, and may choose a form that is not actually found in any spoken regional or social dialect, but is known from Literary Tamil. Since Literary Tamil is the form that all educated speakers know, it can be a repository from which general forms can be chosen; this is another example of Labov’s maxim (Labov 1971:450) according to which non-standard languages in contact with a standard one will vary in the direction of the standard. Here it is not in a formal context, but in a context of avoiding stigmatization.<sup>9</sup>

For example, speakers may model their choice of the past neuter form of verbs on the Literary Tamil past அது *adu*, e.g. வந்த-அது *vand-adu*, rather than the form found very commonly in many non-Brahman dialects, i.e. *-cci* or *-ccu*, e.g. *vandu-cci* ‘it came’ (which is not found per se in Literary Tamil with this verb, but has spread from Class 3 verbs, or from the prototypical pasts in *-ற்று* of verbs like போ *poo* ‘go’ and ஆகு *aa(hu)* ‘become’, which have spoken pasts *pooccu* and *aaccu*, from Literary Tamil போயிற்று and ஆயிற்று, respectively.) Other speakers may choose the *ccu/i* forms unequivocally, so that no hard and fast rules can be given. Many of our entries are variable on this, with many examples of both types.

In fact though we conclude that while some consensus does exist as to what spoken Tamil entails, the situation must be described as being VARIABLE and fluid. Individual speakers may vary considerably, even in their own speech, depending upon whom they are talking to, or what the topic of conversation is. These

<sup>7</sup> Here again, there must be some caveats; some verbs are used only in certain tenses, or only in the negative, or only in certain idioms or collocations. One cannot give them as an equivalent of an English form without severely restricting or carefully defining their limited usage.

<sup>8</sup> I have dealt with the issue of whether there is a ‘standard’ form or not in my 1998 article (Schiffman 1998).

<sup>9</sup> Part of the problem with claiming that a standard Spoken Tamil does or does not exist stems from the fact that for some, a standard must be set by an ‘academy’ or an ‘authoritative’ agency of some sort, and cannot be thought to emerge from usage. But other languages with ‘standard’ (or thought-of as standard) forms, such as so-called RP English, arrived at these standards through customary usage, i.e. the usage of upper-class British boys educated at ‘public’ (i.e. private) schools such as Rugby, Eton, and Harrow, beginning in the 19th century. No academy was involved in setting the RP standard.

phenomena have been noted by many linguists working in the field of sociolinguistics, and are not limited to Tamil. Speakers may vary depending on social characteristics such as their place of birth, their community of origin, their level of education, their socio-economic status, their sex (male vs. female), their age, their occupation, and any other social markers one may isolate. Part of this variability is sociolinguistic, while part of it stems from the fact that Spoken Tamil is grammatically different from LT, and when grammaticalization occurs in a language, there is variability that stems from this kind of change, and not just from sociolinguistic variables. This is particularly true of variability in the aspectual system of ST (see Schiffman 1999: 3.10 for more on this.)

Given this kind of fluidity, we have made our own decisions about what form might be given that would be acceptable to most speakers, forms that would be neutral as to most social characteristics (except that it would not be typically Brahman, nor from the lowest non-Brahman usage.) This is based on our own observations of Tamil usage, and in particular from close study of the Tamil film and the Tamil radio play. We have also decided to err on the side of clarity, so that even though some phonologically reduced forms such as *maari* for *maadiri* (LT மாதிரி) ‘way, manner, type’ are common in rapid speech, we replace them with *maadiri*, since the latter is always acceptable, and learners will not be confused by the form. But if users are confused by a colloquial form that they do not recognize, they can always compare it with the LT form, which will spell out more explicitly the morphological information in question.

### 0.1.3.2 Morphology

This brings up another issue, namely, that ST can not be arrived at by simply substituting spoken forms for LT forms on a one-to-one basis. LT and ST differ *grammatically*, as well as phonologically, lexically, and syntactically. There are many examples of these differences in this dictionary, but let us enumerate a few of them here.

- **Plurality** LT marks plurality of nouns much more systematically than does ST. The typical plural marker in LT is கள் and may occur with both animate and inanimate nouns. ST, on the other hand, does not typically mark plurals on inanimate nouns, or if it does, may use the form *ellaam* (LT எல்லாம் ‘all’) instead of the marker derived from LT கள் which would typically be *nga(l)*, which is used with animate nouns, and also shows up in plural pronouns, e.g. நீங்கள் *niinga(l)*. Note that the final *l* is given in parentheses, because it also does not appear in ST unless a suffix follows.

Furthermore, plurality is also not marked in verbs that agree with neuter (or inanimate) nouns, so although LT can mark neuter nouns and the verbs that agree with them as plural, ST does not.<sup>10</sup>

- **Aspect** Aspectual marking on ST verbs is also more complex and varied than in LT verbs, which will be obvious in many of the example sentences.<sup>11</sup>
- **Passive** LT verbs may be marked for passive, using a verbalizer படு *padu* which otherwise means ‘suffer’ or ‘experience’, but ST avoids passive verbs, so any example of passive in spoken Tamil will sound artificial (cf. Schiffman 1999:3.15.7; in most cases we give examples where the ST verb is active, with an overt subject of some sort, even if the LT example is passive and does not have an overt subject.)
- **Negative forms** The negation of verbs in ST differs from the way negative verbs are formed in LT in a number of ways. One of the more interesting ones is that there is a new form that I refer to in my grammar as the ‘obstinate negative’ (Schiffman 1999:6.5.6) which uses the quotative verb என் (*e)n-* plus the future negative மட்ட- *maatt-* in a construction with the surface form ‘X says it won’t verb’, e.g. LT உப்பு வராமட்டேன் என்கிறது *uppu varamaatteengadu* ‘The salt says ‘I won’t come out’, i.e. ‘The salt refuses to come out (of the saltshaker).’ This quotative verb is also used in a number of other constructions, detailed in HS 1999 7.3. These constructions involve the expression of intent, the expression of facts, a number of onomatopoeic expressions, and some others. It is clear that even if in LT என் is a lexical verb, it does not seem to retain this function in ST, but instead has been

<sup>10</sup> For more information on this question see Schiffman 1999 2.2 ff.

<sup>11</sup> See Schiffman 1999, 3.10 for more on this.

grammaticalized as something else, i.e. it lacks the lexical notion ‘say.’ In fact, the form given above (உப்பு வரமட்டேன் என்கிறது) would probably not occur in LT at all, and some linguists even deny they ever use this form.

Other differences in negative forms between LT and ST morphology are given in Schiffman 1999 6.5 ff. Often, the difference is one where LT has complex morphology, such as in the negative conditional, e.g. வராவிட்டால் ‘If it doesn’t come’, but where ST uses the quotative verb, i.e. *varalle-ṇṇaa* ‘ibid.’ The form *varaṭṭaa* also occurs, of course (HS 1999: 6.10.2).

### 0.1.3.3 Transliteration.

The Roman transliteration chosen represents something between a traditional ‘phonemic’ transcription and a more phonetic one, i.e. it is closer to IPA-phonetic without getting into fine phonetic detail that is actually predictable from a general knowledge of Tamil. Unlike some Indian languages, Tamil does not have a single standard transliteration system. Authoritative sources such as the Madras University *English-Tamil Dictionary* (Chidambaranatha Chettiar 1965), the Madras University *Tamil Lexicon*, and Burrow and Emeneau’s *Dravidian Etymological Dictionary* use different transliterations, especially for some of the laterals and rhotics, where true confusion reigns. We therefore follow Burrow and Emeneau in using an ‘r’ with two underdots, i.e. ṛ for the ‘retroflex frictionless continuant’ that is the last sound in the word ‘Tamil’ (தமிழ்). I am of course fully aware that others use other symbols for this sound, but I have dealt with this issue in two separate studies of Tamil rhotics and liquids (Schiffman 1975, 1980) and have also discussed this in Schiffman 1999, and am convinced that the use of fricatives or various kinds of laterals for this sound is misconceived. There is in fact no lateral articulation with ṛ since the tongue is retracted but does not touch the palate the way laterals do; similarly, the tongue is not close enough to the palate to render a fricative sound. Moreover, the sound represented in Tamil as ṛ patterns like rhotics, i.e. like [r], for example, by not being able to be geminated, which is not a restriction on Tamil laterals. It should never be represented as anything but a rhotic sound. More recent acoustic and articulatory data obtained using different techniques, such as palatography, magnetic resonance imaging (MRI), and magnetometer (EMMA) studies (Narayan, Byrd and Kaun 1999), show that this sound should be placed among the rhotic group, based on the fact that like other rhotic sounds, ṛ involves central airflow. As they note “Laterals, by contrast, exhibit airflow along the sides of the tongue.” (Ibid. pg. 1993)

One problem that has not been addressed by modern technology is that while LaTeX can produce ṛ and this symbol will appear in the ‘print’ version of this dictionary, HTML does not as yet have an ‘r’ with two underdots, so we are forced to use an ‘r’ with one underdot, i.e. ṛ̣, which may be confusing to users who know Hindi or other languages that use this symbol for a flapped retroflex ‘r’, because this is not what this ‘retroflex frictionless continuant’ is phonetically. To make matters worse, popular transcriptions, such as those used in public signing, transliterations of personal names, etc. typically do not mark differences in vowel length, retroflexion,<sup>12</sup> and other distinctions. This is unfortunate, but scholars and others have not been able or willing to agree on a standard transliteration, so we have chosen one that we hope can be used by lay persons as well as scholars, and provide below a chart showing the correspondences between some of these systems, where their differences are significant.

- **Initial Stop Consonants.** The Tamil stop consonants ப, த, ட, ச, க் are usually represented in initial position as [p, t, ʈ, s,] and [k]. Where words borrowed from Indo-Aryan, English, Portuguese, or some other language preserve voiced stops in initial position, we use the Roman letters [b, d, ɖ, j] (but sometimes also [s]), and [g]. Actually retroflex consonants never occur in initial position in Tamil words, so they will never occur in Literary Tamil, and in ST only in borrowed words, e.g. from Telugu or Hindi. One exception to this rule is for onomatopoeic forms, which *do* often have voiced or retroflex consonants in initial position, as well as some other unusual and unpredictable occurrences of various sounds.<sup>13</sup>
- **Medial Voiceless Geminate Stops.** Medially, stops are voiceless in Literary Tamil only if geminate, i.e. doubled, and we follow this convention as well, writing two consecutive consonants [pp, tt, ʈʈ, cc,

<sup>12</sup>Or, retroflex consonants are represented by ordinary letters, but dental consonants are then marked with ‘th’, ‘nh’, or some other letter, which leads to a word like பத்து *pattu* ‘ten’ being spelled *paththu*, which we find objectionable.

<sup>13</sup>For more on this, see Schiffman 1999, 7.5

and kk] for what Literary Tamil writes as ப்ப், த்த், ட்ட், ச்ச், and க்க். Geminate [cc] is to be understood as being phonetically an affricated palatal voiceless stop, similar to the ‘ch’ of English ‘cheese.’ Some speakers also have this sound in initial position in words like சின்ன *cinna* ‘small’ while others prefer *sinna*. Geminate Literary Tamil ற்ற never occurs in ST as such, but is almost always realized as [tt], e.g. பற்றி ‘about’ is *patti* in ST. Occasionally, a Tamil word spelled with ற்ற is borrowed directly (as it were) into ST, but the alveolar sequence is rendered as [tr], as in the item வெற்றி ‘victory, win, triumph’ which comes out as *vetri*.<sup>14</sup> After short vowels, the phonetic value of geminate consonants is to be understood as longer in duration than a single voiceless consonant; after long vowels, geminate consonants are actually not as long as after a short vowel, but we represent them as double and long, anyway, in order to avoid the confusion that results when a single stop occurs intervocalically. That is, Literary Tamil போடு ‘put, place, serve’ is *poodu* in ST, but போட்டு ‘having put, placed, served’ is *poottu* in our transcription, rather than *\*pootu*.

- **Post-nasal Stops.** The stop consonants ப், த், ற், ட், ச், க் following nasals are always voiced in both Literary and Spoken Tamil. Except for -ற்-, which does not occur in such clusters in ST<sup>15</sup>, they are represented in this transcription by the roman letters [b, d, r, ɖ, j], and [g.] Thus தம்பி ‘younger brother’ is *tambi*, அங்கே ‘there’ is *ange*, கெஞ்ச ‘supplicate’ is *kenju*, etc. It should be noted that the transcription [d] for post-nasal clusters such as [vandɖ] வந்து ‘having come’ is phonetically actually [vanɖu], or even more explicitly [vanɖɨ] but we do not usually give this level of phonetic detail.
- **Medial Stop Consonants.** Single stop consonants (i.e. not double or geminated) in medial position, (i.e. between two vowels) in Tamil are typically lax and fricativized. Thus the stops consonants ப், த், ற், ட், ச், க் in medial position are actually phonetically [b] (or sometimes [v] or even [β]), [ð], [r], [pD] [s], and [h], respectively. That is, some of them are lax and voiced, some are flapped (e.g. the retroflex stop), but some (the palatal and velar) are only lax but not usually voiced (although the velar may in some speech be voiced as well, i.e. pronounced [ɣ], in some dialects.) In order to stick to a Roman transcription that does not require elaborate phonetic symbols that complicate our typography, we will use the voiced Roman letters [b] (or [v]), [d, r, ɖ] and the voiceless fricatives [s] and [h] for these lax and sometimes fricativized intervocalic Tamil stops. As noted above, Tamil, unlike some Indian languages, does not have a single standard transliteration system. Our choice was one that could be used by lay persons as well as scholars, hence the use of [h] for intervocalic -க்- rather than a morphophonemic //k// or a phonemic /g/. This does not follow a purely linguistic (e.g. phonemic) convention, or those used in most other dictionaries, but we have found through years of teaching Tamil that most dictionary representations of Tamil are not phonetic enough to permit non-Tamils to approximate usual spoken pronunciations.

In all of our phonetic representations we give preference to those that preserve morphological clarity, so that even though it may be common for many speakers to convert Literary Tamil intervocalic retroflex ட் to a flapped rhotic alveolar [r] in certain cases, we still represent these as *ɖ*. Since the completive marker (Literary Tamil விடு) may be realized in the speech of many people as non-retroflex, i.e. வந்துவிடுகிறேன் ‘I am definitely coming’ may come out in ST as *vandirreē*, we prefer the transcription *vandireen*, as this shows more clearly that there is a completive aspect marker விடு present, even if it is realized phonologically only as *ɖ*. Otherwise it is hard to explain to the learner why the past of it is always, in all dialects *vandireē* Literary Tamil வந்துவிட்டேன். This will help avoid confusion with perfect forms also marked with இரு *iru*, contrasting with completive -விடு.

- **Laterals and Rhotics.** The Tamil sonorants ல், ள், ற், and ழ் are represented in our transliteration as *l, ɭ, r,* and *ɻ*, respectively. Tamil ற் is not treated as a sonorant in its underlying form, but as an alveolar stop. Neither in Literary Tamil nor in ST can it occur in initial position; intervocalically it does occur in ST, where in most dialects it is indistinguishable from ற், so in this position it is simply transliterated by *r*. In ST, Literary Tamil clusters such as ள்ற are simplified in non-Brahman dialect to [ɻɻ] after short vowels (e.g. Literary Tamil என்று ‘having said’ becomes simply *-ɻɻu*); after long vowels in ST ள்ற becomes simply *ɻ*, i.e. தோன்று ‘seem, appear’ is *toonɻu* or (to be truly ‘phonetic’) [toɻɻɨ].

<sup>14</sup> This term is common in sports reports, both oral and written.

<sup>15</sup> An exception would be the rather artificial word நன்றி for ‘thanks’, pronounced in ST *nandri*



In many spoken dialects of Tamil the sonorant  $\text{ṛ}$  is merged with  $\text{ṅ}$  and never occurs in ordinary speech. But because this sound is sociolinguistically highly preferred, and foreigners who are able to pronounce it are praised for their efforts, we give  $[\text{r}]$  as the usual transliteration of  $\text{ṛ}$ , even though many speakers, even educated ones, may be heard to use  $\text{ṅ}$ . That is, we give *maṛe* for Literary Tamil மழை ‘rain’, even if many speakers actually say *maṅe* or even *male*.<sup>16</sup>

There are certain stem-final consonants in this group that tend to be not realized at all in ST, while others occur but with an epenthetic  $[\text{u}]$  vowel<sup>17</sup> added. Thus it is typical for the final  $\text{ṅ}$  in Literary Tamil pronouns அவள், அவர்கள், நீங்கள் (‘she, they, you’) to be deleted in final position in ST: *ava*, *avanga*, *niinga*. The same is true of  $\text{ṅ}$  that occurs as the final segment in certain person-number-gender endings, as in இருக்கிறீர்கள் ‘you are located’ which is realized phonetically as *irukkriinga* or *irukkiinga* if nothing follows it. If anything follows as a suffix, however,  $\text{ṅ}$  is not deleted in ST: நீங்களா? ‘you?’ is realized in ST as *niingaḷaa*, essentially phonetically the same as the spelling pronunciation of Literary Tamil. Final  $\text{ṅ}$  of other words may be treated in different ways by different ST dialects. Some dialects delete  $\text{ṅ}$  in நாள் ‘day’ to give *naa*, but others add an epenthetic  $[\text{u}]$ : *naaḷu*. We give preference to the dialects that do the latter, i.e. preserve morphological clarity.<sup>18</sup>

Monosyllabic words with short vowels ending in laterals (there are none that end in rhotics) such as பல் ‘tooth’, கள் ‘toddy’, சொல் ‘say’, etc. are realized in ST with doubled laterals and an added epenthetic  $[\text{u}]$  vowel: *kallu*, *kallu*, *sollu*, etc., and all of the  $[\text{u}]$ ’s are, of course, phonetically  $[\text{ɥ}]$ <sup>19</sup>, which we deal with in more detail below. Such doubled laterals, like doubled stops in general, are in fact twice as long as a single segment, if they follow a short vowel; when they follow a long vowel, they are not in fact twice as long, as Lisker (Lisker 1958) has shown.

- **Nasals.** Literary Tamil has graphemes for a labial nasal  $\text{ṃ}$ , a dental nasal  $\text{ṅ}$ , an alveolar nasal  $\text{ṇ}$ , a retroflex nasal  $\text{ṣ}$ , a palatal nasal  $\text{ṅ}$ , and a velar nasal  $\text{ṅ}$ . Spoken Tamil does not need this many phonemic or phonetic distinctions, requiring only  $[\text{m}, \text{n}, \text{ɳ}]$  and  $[\text{ŋ}]$ . We therefore transliterate  $\text{ṃ}$  as  $[\text{m}]$ , both  $\text{ṅ}$  and  $\text{ṇ}$  as  $[\text{n}]$ ,<sup>20</sup> as already noted.  $\text{ṣ}$  as  $[\text{ɳ}]$ , while  $\text{ṅ}$  and  $\text{ṅ}$  which usually occur only before a homorganic stop (i.e. the palatal and velar stop, respectively) can be transcribed with  $[\text{ɲ}]$  and  $[\text{ŋk}]$  with the assumption that English speakers, at least, will pronounce these with palatal and velar articulations automatically.  $\text{ṅ}$  occasionally occurs in prevocalic position in ST, in which case we do transliterate it as  $[\text{ŋ}]$ , e.g. ஞாயிறு ‘Sunday’ *ñāayiru*.

Monosyllabic words with short vowels ending in alveolar nasals (those that end in retroflex nasal) follow the pattern of the laterals mentioned above (there are no monosyllabic words ending in a labial nasal) such as என் ‘my’, ஒன் ‘your’ convert the nasal segment into nasalization of the vowel:  $[\text{ẽ}]$ ,  $[\text{õ}]$ ,

Monosyllabic words with short vowels ending in retroflex  $\text{ṣ}$ , such as பெண் ‘girl’, கண், ‘eye’ etc. follow the pattern of doubled laterals with an added epenthetic  $[\text{u}]$ : *poṇṇu*,<sup>21</sup> *kaṇṇu* in ST.

- **Glides.** The Tamil glides  $\text{ṽ}$  and  $\text{ṽ}$  are usually transliterated as  $[\text{v}]$  and  $[\text{y}]$ , respectively. In certain cases  $\text{ṽ}$  is closer phonetically to  $[\text{w}]$  or even to the bilabial  $[\text{β}]$ , similar to Spanish ‘v’ in ‘vaca’, but we ignore this degree of phonetic detail. Often  $\text{ṽ}$  and  $\text{ṽ}$  in Literary Tamil forms will not have any phonetic representation at all in ST, since intervocalic  $\text{ṽ}$  in particular is deleted in weak positions, resulting in forms like ST *koṇḍaa* from Literary Tamil கொண்டுவா. In such cases no representation of  $\text{ṽ}$  will appear in ST.<sup>22</sup> Similarly, the presence of  $\text{ṽ}$  in Literary Tamil often conditions palatalization of adjacent consonants in ST, with subsequent assimilation or deletion of the  $\text{ṽ}$  in ST. Thus, Literary Tamil ஆய்ந்து ‘five’ is *anju* in ST—the  $\text{ṽ}$  causes palatalization of the dental  $\text{ṅ}$ , after which  $\text{ṽ}$  disappears, leaving

<sup>16</sup>Most Tamils will deny that they merge these two sounds, because of the notion they grow up with that the sound  $\text{ṛ}$  is so quintessentially Tamil that it must be preserved against all change. (In fact, this sound is more consistently used by Malayalam speakers in Malayalam, but that does not earn much credence among Tamils.)

<sup>17</sup>For its pronunciation, see below.

<sup>18</sup>In this sense we take bits and pieces of different dialects as ‘standard’, since this is more transparent pedagogically, even though there may be no speaker who actually replicates each and every pronunciation we prefer.

<sup>19</sup>This is an unrounded high back vowel, such as found in some southern American English pronunciations of ‘book’, i.e.  $[\text{bʊk}]$ ; this vowel is also common in Japanese.

<sup>20</sup>With the exception that clusters  $\text{ṅṅ}$  usually become  $[\text{ŋŋ}]$

<sup>21</sup>This form also undergoes vowel rounding; for an explanation see below.

<sup>22</sup>See Schiffman 1993 for more details on the deletion of intervocalic  $[\text{v}]$ .

only ST [nj]. In final position also, many Literary Tamil  $\acute{y}$ 's are deleted, e.g. the adverbial ending -ஆய் is realized simply as [aa] in ST. Monosyllabic words that end in  $\acute{y}$  in LT may have an epenthetic [i] added, as in நாய் 'dog', ST *naayi*.

In contrast, Tamil words beginning with the mid-vowels [e] and its long counterpart [ee] have an automatic [ʸ] onset in Tamil (as in many South Dravidian languages). Thus என்ன 'what' is phonetically [ʸenna]. Since this is automatic and predictable, we do not supply this [ʸ] glide in our transcription, but transcribe என்ன always as *enna*. Similarly, the rounded vowels o], [oo], [u], and [uu] are preceded in initial position with an automatic [w]-glide in Tamil. ஊர் 'town, city' is phonetically [ʷuru], but again since these w-glides are predictable (in fact hardly even salient to a Tamil speaker), we do not provide them. In connected speech in non-sentence-initial position, these automatic glides are usually deleted: அவர் பேர் என்ன? 'What is his name?' is in ST *avar peer-enna* rather than *avar peer-ʸenna*.

- **Oral Vowels.** The Tamil vowel system consisting of five cardinal vowels அ, இ, உ, எ, ஓ and their long counterparts ஆ, ஈ, ஊ, ஏ, ஔ are represented in our transliteration as [a, i, u, e, o] and [aa, ii, uu, ee, oo], respectively. The diphthong ஐ is usually simplified to [e] in ST; thus the accusative அவரை 'him' is *avare* in ST. This is actually phonetically [e], i.e. *avare*, but we represent it as [e] for simplicity. In monosyllables ஐ is not monophthongized, but the [i] element is lengthened, or followed by a glide [y]. Thus the verb வை 'put, keep' becomes *vayyi* in isolation (e.g. as a verb stem or imperative), but in more complex morphological constructions, e.g. followed by tense-markers, will change to [e]: வைத்தேன் 'I put, kept' will be *vacceen* or *vecceen*. Here the [i] element triggers palatalization of the த் to produce [c].

A special note must be made of the phonetic qualities of the short உ vowel when it occurs after the first syllable of a word, and in particular in final position. Its pronunciation in initial syllables is [u], but after the first syllable its phonetic quality is unrounded and somewhat fronted, i.e. more like IPA [ʊ] or [ɯ]. This is similar in quality to the short 'oo' vowel in 'book' as pronounced in southern American English, to the Russian 'jery' and to the final [u] in Japanese. Since it is again totally predictable when a Tamil உ will be pronounced in this way, we do not represent it as different from phonetically rounded [u]. This pronunciation is not different from the spelling pronunciation of Literary Tamil [u], so anyone with a knowledge of the pronunciation of Literary Tamil will have no trouble predicting this. The only exception to this is that Tamil nicknames or shortenings of longer names that end in [u] will in fact be phonetically [ʊ], e.g. *baɽu* 'Balu'.

This situation is complicated by the fact that in ST, many short [i] vowels (phonetically [i] also merge with [ɯ]). For example, the vowel of the past tense-marker of Class 3 verbs spelled இன் as in வாங்கினேன் 'I bought, acquired, fetched' is often pronounced like [ɯ] in ST: [vaangʊneen]. Some linguists who have worked on ST have regularly substituted [ɯ] in these positions, but since this pronunciation is predictable, and differs from the spelling of Literary Tamil, we do not give either [u] or [ɯ] here, but transcribe it as [i], leaving it to the knowledge of the speaker to provide the correct phonetic realization.

The Literary Tamil diphthong ஔ [au] is rare even in Literary Tamil, and does not occur in our data except in loan words, e.g. English 'pound', which we would represent as *paunḍu* or *pavunḍu*.

- **Long Vowels.** In final position in ST, Literary Tamil long vowels are often shortened, so that what may be written with a long vowel in LT may always occur short in ST. Thus அங்கே 'there' is always *ange* in ST, unless followed by another vowel, as in அங்கேயே 'there (emph.)' *angeeye*. Here the non-final ஏ remains long, but the final one is shortened. Sometimes to keep morphological processes clear, however, we represent long vowels in final position as long (in transcription), even though they are phonetically short. In rapid speech, moreover, long vowels anywhere in a word will be shorter than when the word is in isolation, and short vowels may be completely deleted.
- **Nasal Vowels.** Spoken Tamil possesses a set of nasal vowels [ã], [õ], [ẽ], and [ũ], some of which also have long counterparts [ãã], [õõ], and [ẽẽ]. These nasal vowels are not found in Literary Tamil, but arise from the nasalization of vowels followed by ன் or ம் in final position. Thus, [ã] arises from the sequence -அன் in Literary Tamil, e.g. அவன் 'he' becomes [avã]; in some dialects அவன் becomes [avẽ] instead, which accounts for some instances of [ẽ]. [õ] arises from the Literary Tamil sequence -அம், so

that Literary Tamil மரம் ‘tree’ becomes [marō]; [ũ] arises from the Literary Tamil sequence -உம், e.g. நீங்களும் ‘you, too’ becomes [niingalũ] in ST.

Long nasal vowels [ãã], [õõ], and [ẽẽ] may have several sources in Literary Tamil. [ãã] may result from the nasalization of both Literary Tamil -ஆன் or -ஆம், i.e. வந்தான் ‘he came’ becomes [vandãã], but இருக்கலாம் ‘it may be, let it be’ also has final [ãã], i.e. [irukkalãã]. Because of the shortening of long vowels in final position, these long nasal vowels also are shortened finally; but to preserve morphological clarity we usually represent them as long in our transcription. The long vowel [õõ] arises from the nasalization of the sequence ஓம், found typically as the marker of second person plural, as in வந்தோம் ‘we came’ [vandõõ]. Because of the shortening rule, however, it may be realized phonetically as [vandõ], but we usually avoid this representation, again for morphological clarity. The long vowel [ẽẽ] usually arises from nasalization of the Literary Tamil sequence -ஏன், found most typically in the person-number-gender ending for first person singular, as in வந்தேன் ‘I came’, i.e. [vandẽẽ]. Again, by the shortening rule this usually becomes [vandẽ], but for clarity we avoid this representation. It does not become short in monosyllabic environments, so Literary Tamil ஏன் ‘why?’ remains long: [yẽẽ], contrasting with என் ‘my’, which is [yẽ].

Monosyllabic words with long vowels ending in Literary Tamil ன் usually do not nasalize, but instead an epenthetic [u] ([யு]) is added, e.g. மாண் ‘stag’, becomes [maanu]. Literary Tamil words ending in ண் also do not produce nasalized vowels in ST, but if position final, simply add [u], e.g. ஆண் ‘male’ becomes ST *aanu*.

- **Vowel Shifting.** A number of other differences between vowels in Literary Tamil and their realization in ST have to do with certain phonological changes in the Tamil vowel system since Tamil orthography was fixed.

- **Lowering.** Literary Tamil words with short high vowels இ [i] and உ [u] in an initial syllable followed by *one* consonant and the vowel அ [a] or ஐ [ai] lower these vowels to எ [e] and ஒ [o] respectively in ST. Thus Literary Tamil forms like இலை ‘leaf’ and குழந்தை ‘child’ become *ele* and *korande*, respectively, whereas forms like இல்லை ‘no, not’, where the vowel is followed by a double consonant, do not exhibit lowering. This change is totally regular, so that even some borrowed words, such as English ‘glass’, borrowed usually as கிளாஸ் may, in some dialects, where the word has been nativized, become *keḷaas* or even *kelas*.
- **Rounding.** Another process that is less regular, and may therefore still be in progress as a sound change, is the rounding of short and long front vowels இ, ஈ, எ and ஏ ([i, ii, e], and [ee]) to their corresponding back vowels உ, ஊ, ஓ, and ஔ ([u, uu, o], and [oo].) This occurs usually when the initial consonant is a labial ([m, v, p]) and the following consonant is retroflex. Some forms that have undergone this change are socially quite acceptable, but others are considered to be somewhat substandard or casual (or even “vulgar”) so many speakers avoid this kind of rounding, or deny that they do it even when it is observed in their speech. Thus the rounding of the vowel in Literary Tamil பெண் ‘woman’ to ST *poṇṇu* is quite normal, but the following are on a kind of sliding scale of acceptability: *porandadu* for Literary Tamil பிறந்தது ‘it was born’ (this form undergoes lowering first), [puḍi] for Literary Tamil பிடி ‘like’, *vuudu* for Literary Tamil வீடு ‘house’, *voṇṇum* for Literary Tamil வேண்டும் ‘want, need, must’, etc. Different speakers would rank these differently, but the general scale of acceptability is as given. We try to avoid what are considered the most egregious of these, but in an attempt to remain colloquial, some may be present in our examples. Sometimes the conditioning factor does not even include a retroflex consonant, as in the example of பிற, where the following consonant is alveolar; in extreme cases no second conditioning consonant is present, and an initial labial alone is sufficient to cause rounding, as in Literary Tamil மிதக்கும் ‘it will float’ becoming, in some dialects, *modakkum*. This is obviously an example of a sound change in progress, and is therefore sociolinguistically marked.
- **Other Changes.** There are a few other changes in vowel quality from Literary Tamil to ST that cannot be described under the previous rubrics. These are mostly idiosyncratic, but may have to do with what is often called vowel harmony, i.e., vowels changing in order to agree with another vowel in height or rounding. Thus, for example Literary Tamil கொடு ‘give’ has a high vowel in its

first syllable in ST, i.e. *kuḍu* rather than *\*koḍu*, which is the reading pronunciation of LT. There is no good explanation for this change, except that the height of the two vowels agree; but there are many other counterexamples. Since this dictionary is neither an etymological nor a historical dictionary, its concern is not to explain such changes, but merely catalogue them. Another vowel harmony example is the change of the dative case marker க்கு in LT to [kki] in ST when following a front vowel, e.g. பட்டிக்க்கு ‘to grandmother’ becomes *paattikkki* in ST.

## 0.2 Transitivity and Verb Classes.

As noted earlier, this dictionary provides the verb class and transitivity specification lacking in most English-Tamil dictionaries, without which non-Tamils cannot reliably produce grammatically correct forms of the verb.

### 0.2.1 Classification of the Tamil Verb.

Tamil verbs have been classified in a number of different ways, depending on the shape of the morphemes used to mark tense. The **Tamil Lexicon** uses a schema involving 13 separate classes; other scholars have proposed smaller numbers of classes but each involve subclasses and there are always exceptional forms that do not fit neatly into any kind of scheme. We use a modification of the 7-class scheme given in Arden (1942:148-9) and Fabricius (1972:vi-vii), which is known in the literature as “Graul’s classification”. We differ, however, in that instead of Graul’s seven classes marked by roman numerals, we use arabic numerals.

This classificatory scheme does not work perfectly for ST, however, so we have proposed additional subclassifications for certain of Graul’s classes, notably verbs in classes II (2) and VI (6), where stem-final front vowels trigger palatalization of the past tense markers ந்த and த்த to [nj] and [cc], respectively. Verbs of classes II (2) and VI (6) whose stems do not meet these conditions are not specially marked, but those that do are marked 2b and 6b, respectively. Thus the verb தெரி, 2 intr., with stem-final ி triggering palatalization of the past marker ந்த to [nj] in spoken is classified as 2b, and சமை, VI(6) tr., which has a palatalized past tense marker [cc] in ST, is classified as 6b.

In addition, in classes II, III, and V there are a number of irregular verbs that do not follow the regular rules in one or another tense form, so we have marked certain verbs as members of subclass 2c, or 3b, or 5c, to try to fit them into the class they most closely resemble. Most irregularities or complexities of the Tamil verb are to be found in the *past* tense; were it not for these past tense problems, most Tamil verbs could be classified simply as “strong” or “weak”, depending on whether they have doubled consonants in the present (கூற-) and future (஁஁) tense marker, or single consonants (஁ற) and (஁஁). This strong-weak scheme is used by some scholars and in some pedagogical materials for Tamil as a general description of the verb, but it is not adequate to predict the details of the past tense. It is useful if the discussion centers on the formation of infinitives or of neuter futures, and the strong-weak distinction also correlates in some ways with transitivity and intransitivity. But this correlation is not perfectly regular and can only serve as a mnemonic device when the exact classification cannot be remembered.

In ST, subclassifying classes 2, 3, 5, and 6 into palatalizing, non-palatalizing, and otherwise irregular helps to take care of most areas where ST forms are different from Literary Tamil, but it does not take care of all. In some cases, ST verbs are members of totally distinct classes from their Literary Tamil counterparts, and this is particularly true of members of Literary Tamil classes I(1) and V(5)—Literary Tamil class I is a very small class, i.e. has very few members, and given that some of these Literary Tamil verbs are never used in ST, while other members shift to another class, class I(1) as a spoken class is an almost empty set. There are, however, a few members that retain and duplicate the morphology of the Literary Tamil set, so it must be retained. The Literary Tamil verb செய், I tr. ‘do, make’ changes to class 2b in ST: its past is [senj-] rather than the Literary Tamil செய்த்- [seyd-] (the

verb பெய் ‘rain’ also shifts to 2b in ST). But verbs like அழு ‘weep’ and உழு ‘plow’ remain in class 1 in spoken, with pasts in [d] analogous to Literary Tamil single த் pasts: அழுதேன் = [aɾudeen] ‘I wept.’

Another Literary Tamil class that has lost members in ST is class 5—this is a class with final sonorants, and many irregular and unpredictable things happen to these stems in both Literary Tamil and in ST. In ST some of these are not used, or are used only with an aspectual auxiliary. For example கல் ‘learn’, with Literary Tamil past கற்ற- usually occurs in ST only with aspectual இரு or கெள், e.g. [katt-iru] or [kattu-kkoo], as in [tamiɻ engee kattirukkiinga?] ‘Where (in the world) did you learn Tamil?’ Other Literary Tamil class V verbs are not used at all, or only in certain idiomatic expressions. The verb காண் ‘see’ is not common in ST as a main verb, only in certain collocations or idioms such as [kanaa kaṇu] ‘have (i.e. see) a dream’, or in the archaic frozen negative form [kaṇoom] ‘(I) don’t see (a thing, etc.)’. Other class 5 verbs are shifted to class 3, e.g. தின் ‘eat’ which has the Literary Tamil past தின்றேந் ‘I ate’ is realized as [tinnineen] (with usual shortening of the cluster [nnin] to [nn]) and the Literary Tamil future தின்பேன் would be [tinnuveen]. But not all speakers do this, and some retention of this class, even at a minimal level, must be recognized. For example, the quotative verb என் [en], phonologically reduced as it is, is in standard ST more or less a predictably class 5 verb, with past in [ṇṇ-], present [ngr-] and future [mb-], although the future may also occur as [nnuv-], e.g. [appaḍi-ngraan] ‘that’s what he says’ or [appaḍi-mbaan] ‘that’s what he’ll say’. The pervasive use of this verb as a grammaticalized quotative marker and embedding marker in Tamil guarantees that it will appear very often in conversation and spoken texts, so despite the fact that என் may be almost the only verb that retains features of class 5 morphology, its functional load is high in the language.

## 0.2.1.1 Graul's Verb classes.

As the verb classifications are dependent on the form of the tense markers used with different stems, we give first a chart showing the shape of these tense markers in the different classes, followed by a chart showing illustrations of typical members of the various classes.

Tamil Verb Class Tense Markers			
Verb Class	Present	Past	Future
1	-ஓற்- [-r-]	-த்- [-d-]	-வ- [-v-]
2	-ஓற்- [-r-]	-ந்த்- [-nd-]	-வ- [-v-]
2b	-ஓற்- [-r-]	-ந்த்- [-nj-]	-வ- [-v-]
3	-ஓற்- [-r-]	-இன்- [-in-]	-வ- [-v-]
3b	-ஓற்- [-r-]	-ன்- [-n-]	-வ- [-v-]
3c	-ஓற்- [-r-]	-ன்ன- [-nn-]	-வ- [-v-]
4	-ஓற்- [-r-]	-ட்ட்- [-tt-]	-வ- [-v-]
5.	-ஓற்- [-gr-]	-ன்ற- [-nṇ-]	-ப்- [-b-]
5b	-ஓற்- [-kr-]	-ற்ற- [-tt-]	-ப்- [-pp-]
5c	-ஓற்- [-kkr-]	-ட்ட்- [-tt-]	-ப்- [-pp-]
6.	-க்கூற்- [-kkr-]	-த்த்- [-tt-]	-ப்ப்- [-pp-]
6b	-க்கூற்- [-kkr-]	-த்த்- [-cc-]	-ப்ப்- [-pp-]
7	-க்கூற்- [-kkr-]	-ந்த்- [-nd-]	-ப்ப்- [-pp-]

Tamil Verb Classes, with Examples of Typical Verbs				
Class	Verb Stem	Present	Past	Future
1	அழு 'weep' [aɻu]	அழுகிறேன் [aɻurēē]	அழுதேன் [aɻudēē]	அழுவேன் [aɻuvēē]
2	உட்கார் 'sit' [okkaaru]	உட்காருகிறேன் [okkaarrēē]	உட்கார்ந்தேன் [okkaandēē]	உட்காருவேன் [okkaaruvēē]
2b	உடை 'break' [oɖe]	உடைகிறது [oɖeyradu]	உடைந்தது [oɖenjadu]	உடையும் [oɖeyum]
2c	வா 'come' [vaa]	வருகிறேன் [varrēē]	வந்தது [vandadu]	வருவேன் [varuvēē]
3	வாங்கு 'buy' [vaangu]	வாங்குகிறேன் [vaangurēē]	வாங்கினேன் [vaanginēē]	வாங்குவேன் [vaanguvēē]
3b	போ 'go' [poo]	போகிறேன் [poorēē]	போனேன் [poonēē]	போவேன் [poovēē]
3c	சொல் 'say' [sollu]	சொல்லுகிறேன் [sol(lu)rēē]	சொன்னேன் [sonnēē]	சொல்லுவேன் [solluvēē]
4	போடு 'put' [pooɖu]	போடுகிறேன் [pooɖurēē]	போட்டேன் [pooɖtēē]	போடுவேன் [pooɖuvēē]
5.	உண் 'eat' [uɻɻu]	உண்ணுகிறேன் [uɻɻurēē]	உண்டேன் [uɻɻdēē]	உண்பேன் [uɻɻuvēē]
	என் 'quote' [-(e)n(u)]	என்னுகிறேன் [(e)ngrēē]	என்றேன் [(e)nɻēē]	என்பேன் [(e)nɻuvēē]
5b	கேள் 'ask' [keɻu]	கேட்கிறேன் [keɻkurēē]	கேட்டேன் [keɻtēē]	கேட்பேன் [keɻpēē]
5c	காண் 'see' [kaaɻu]	காண்கிறேன் [kaaɻurēē]	கண்டேன் [kaaɻdēē]	காணுவேன் [kaaɻuvēē]
6.	பார் 'see' [paaru]	பார்க்கிறேன் [paakkrēē]	பார்த்தேன் [paattēē]	பார்ப்பேன் [paappēē]
6b	சமை 'cook' [same]	சமைக்கிறேன் [samekkkrēē]	சமைத்தேன் [sameccēē]	சமைப்பேன் [sameppēē]
7	நட 'walk' [naɖa]	நடக்கிறேன் [naɖakkrēē]	நடந்தேன் [naɖandēē]	நடப்பேன் [naɖappēē]

## 0.2.2 Transitivity.

The notions 'transitive' and 'intransitive' as applied to verbs, is used to express the difference between actions that have an effect on someone or something else in the sentence, vs. actions that do not affect anyone or anything else, except for perhaps the subject of the action.

Most grammars of Tamil have discussed the transitivity status of Tamil verbs as being a case of either transitive or intransitive, i.e., as if this distinction were exactly parallel to that of English or some other western language. Actually any cursory examination of the Tamil verb will reveal that the semantic distinction so clearly marked in the morphology, i.e., the distinction between pairs like ஓடு [ooɖu] and ஓட்டு [ooɖtu], which is usually glossed as 'run' vs. 'cause to run' or 'run of one's own volition' vs. 'run something' is not as simple when all the verbs of the language have been taken into account.<sup>24</sup> Some researchers on Tamil, such as Paramasivam 1979, have rejected the dichotomy between transitivity and intransitivity as inadequate for Tamil, and have opted for a distinction known as 'affective' vs. 'effective', which is felt to more adequately capture the distinction. We have opted to stick with the transitivity/intransitivity distinction, however, because it is our experience that

<sup>24</sup>This is not just the case with Tamil, but with many other languages of the world, as has been shown very clearly by the research of Hopper and Thompson (1980), for example.

American students, at least, if they have any familiarity with this distinction, know it in this way, rather than as ‘affective/effective.’

In fact Hopper and Thompson show that verbs must be scaled for their *degree* of transitivity, since ‘blaming’ or ‘seeing’ is in some sense less transitive than ‘breaking’ or ‘killing’, actions which have a definite effect on an object, whereas to be blamed or seen does not affect the ‘target’ of the action in the same way. Thus to refer to உடை 2b, intr. as an intransitive kind of *breaking* since the process or person who caused the breaking is not known is also not as neat a distinction as one would like, even though the morphology of Tamil gives us to உடை’s, one ‘intransitive’, i.e. without known agent, as in கண்ணாடி உடைந்தது [kaṇṇaadi oḍenjadu] ‘the glass broke’, the other ‘transitive’, as in அவன் கண்ணாடியை உடைத்தான் [avan kaṇṇaadiye oḍeccaan] ‘He broke the glass.’ These ‘intransitives’ are also usually possible only with a third-person, often neuter, ‘subject,’ i.e. ‘glass.’ Yet to think of glass as the ‘subject’ of ‘intransitive’ breaking but as the object or target of transitive breaking (when the agent of the action is known), is illogical.<sup>25</sup>

### 0.2.2.1 Noun-verb Compounds

Another problem found in Tamil that we have struggled with in an attempt to give a clear and unambiguous picture of this problem is that of verbs such as குற்றம் சொல் *kuttam sollu* ‘blame’, (i.e. ‘say-blame’) that combine a nominal object with a verb stem in a kind of compound, such that the noun, though functioning as the object of the verb, is not marked with the normal accusative marker.<sup>26</sup> However, although no accusative is marked, no other noun in the sentence may be marked as the object of that verb either, so that the structure of English ‘he blamed his sister for the problem’, with ‘sister’ as the object of the verb ‘blame’, cannot be replicated in Tamil. In Tamil the object of the verb குற்றம் சொல் *kuttam sollu* must be marked with some other case, in this instance the ‘locative’ of animates, i.e. இடம் which in ST would be *kitte*.

- அவன் தன் தங்கையிடம் கஷ்டத்தை பற்றி குற்றம் சொன்னான்
- *avan tangecci kitte kaṣṭatte patti kuttam sonnaan.*
- ‘He sister-LOC problem-ACC. about blame-said.’
- ‘He blamed his sister for the problem.’

The case marking for the item glossed ‘problem’ may also be problematical, i.e., is not immediately predictable. In the above example the phrase ‘for the problem’ might also be marked with dative plus ஆக [aaha] ‘for the sake of, on behalf of’, i.e. கஷ்டத்துக்காக *kaṣṭattukkaaha*; the case-marking of ‘sister’ might also be realized with the postposition பார்த்து *paattu*, which is derived historically from the past participle of the verb பார் *paru* ‘see’, but must be glossed in modern Tamil something like ‘direct the attention toward’ when used with verbs like சிரி *siri* or குலை *kole* ‘bark’ to indicate the person being laughed or barked at.

We have considered a number of ways to deal with this issue, but because of the complexity of the syntax here, have opted to continue to refer to a noun-verb compound as a *transitive* verb, even though its object is incorporated into the verb. Given the various optional syntactic possibilities mentioned above, we ask the user to rely on the syntax of the example sentence given, rather than provide complex syntactic frames.

<sup>25</sup> But as anyone who has dealt with young children knows, an argument is often likely to ensue between the parent and the child over who the agent of the breaking was, with the parent claiming that the action was transitive and that there had to be an agent, while the child argues that the action had no cause and no agent—“it just *broke*.” Parents typically contend this is not the case, and that responsibility or blame has to be assigned; children, even when found with rocks in their hands, attempt to deny this contention.

<sup>26</sup> This kind of structure is common in many South Asian languages.



### 0.2.2.2 Dative Subjects

Another issue that enters into the discussion of this area is that of verbs that take ‘dative subjects.’ These are stative verbs (i.e., they describe psychological *states*, rather than actions) whose semantic subject is marked with the dative case, and the verb itself marked with a neuter person-number-gender marker. The object (or target of the action) is marked accusative if animate; otherwise it is unmarked, i.e. nominative. Such Tamil stative verbs as பிடி *pidi* ‘like’, தெரி *teri* ‘know’, வேண்டும் *veeṇum* ‘want, need’, போதும் *poodum* ‘suffice’ கிடை *keḍe* ‘be available’<sup>27</sup> take the subject in the dative and if the object (or target, i.e. that which is known, liked, wanted, etc.) is animate, it is marked accusative. For example, எனக்கு அவரைத் தெரியும் *enakku avare teriyum* ‘I know him’ has the subject in the dative and the object in the accusative, with no nominative case marking possible.

On a scale of transitivity, such verbs are obviously very low, and in normal usage either the dative-marked subject or the object, or both, may be missing, i.e. a well-formed sentence can consist simply of தெரியுமா *teriyumaa* ‘Do you know?’ or வேண்டாம் *veeṇḍaam* ‘(I) don’t want (it)’. In our data, these are always marked “3 pn + dat” for “third-person-marked verb with dative subject”. Since these verbs are either low on the transitivity scale (or in some cases definitely *intransitive*, as with போதும் *poodum* ‘suffice’ and கிடை *keḍe* ‘be available’), we may not give the usual “tr.” or “intr.” specification, but mark it as ‘defective’. Some Tamil verbs can be used dative-statively and also with first and second-person subjects, so when this happens, this is also marked.

Our solution to this problem is to issue caveats but not to attempt a wholesale reclassification or scaling of transitivity for the Tamil verbs. We continue to use the (probably archaic) bipolar scale of transitivity, with the two உடை’s above given the traditional ‘intransitive/transitive’ labels, often with information about restrictions on person and number of ‘subject.’ Were it not for the fact that Tamil usually marks the distinction between intransitive and transitive morphological differences in the *tense*-marking of the two types, and that there are tense markers for *all* tenses in Tamil (unlike English, where only the past is morphologically marked) it would not be obvious to most non-Tamils that distinctions must be kept separate. English, for example, has only a small set of verbs that are paired in this way, one being transitive and the other intransitive. Even these (sit/set, lie/lay, fall/fell, rise/raise) are not kept separate by many speakers. In Tamil either the stem itself is different (such as the (c)vc/(c)vcc-type exemplified by ஓடு/ஓட்டு *oḍu/oḍṭu*, ‘run’ vs. ‘drive’ or there is an alternation (c)vNC-/(c)vCC- (as with திரும்பு/திரும்பு *tirumpu/tiruppu* ‘return’, or the differences are marked in the tense markers, usually with weak types for intransitive and strong types for transitive. More work needs to be done on the ways that Tamil marks the distinction between transitive/effective and intransitive/affective verbs; since the database for this dictionary can be easily searched, we hope future researchers will use it to look at various lexical patterns that have yet to be analyzed or described for Tamil.

### 0.2.2.3 Verb-verb compounds

Another feature that is common in the verbal system of most Indic languages is the tendency to ‘compound’ verbs. By this is meant the use of two verbs adjoined in such a way that only the last one has tense and person-number-gender marking, while the previous one(s) occurs in a form known in Tamil as an ‘adverbial participle’ (which is commonly referred to by the abbreviation AVP.) Thus where English or other languages might conjoin two sentences such as ‘I went to the store’ and ‘I saw him’ to get ‘I went to the store and saw him’ Tamil (and other Indic languages) typically has a sentence like ‘Having gone to the store, I saw him’, i.e. *naan uurukku pooyi, avare paatteen*. To complicate matters, aspectual verbs are also adjoined in this way (cf. Schiffman 1999: 3.10) as well, with the aspectual verb marked for tense and PNG, but not the lexical verb, which occurs in the AVP form. Beyond this, we also find that verbs are compounded in this way to in effect create new lexical verbs; since Tamil does not borrow *verbs* easily from other languages, it creates new ones by combining existing verbs, e.g. the verb ‘teach’ can be rendered as சொல்லி கொடு *solli kuḍu* ‘say and give; having said, give.’ A study of the verbs in this dictionary will show that a large number of them have been

<sup>27</sup> See Schiffman 1999 for a more complete description of these ‘dative-stative’ verbs, 3.5.2.1

‘created’ this way, either with aspectual verbs, or with other lexical verbs, or both. Certain lexical verbs tend to recur often in these combinations, especially when the result is a transitive verb: ஆக்கு *aakku* ‘make s.t. become’; உட்டு *uttu* ‘feed, nourish’; செலுத்து *seluttu* ‘make s.t. go’, and so on. The last example here is instructive, because it itself is an example of an intransitive verb made transitive by the addition of -உத்து *-uttu*, which is a frequent way to create transitive verbs.<sup>28</sup>

But it is even more interesting because செல் *sel* alone does not occur in Spoken Tamil; but as a derived transitive, *seluttu* is acceptable in spoken when combined with other *verbs*, though not with nouns as the object. Again, this phenomenon needs to be studied; attention to it will reveal other interesting patterns, such as the fact that when விழு *viru* ‘fall’ is made transitive by adding -உத்து *-uttu*, what we get is a form with a long vowel, but with only த்து *-ttu* suffixed to it: வீழ்த்து *viruttu* ‘bring down, make s.t. fall, defeat.’. Another common verbalizer is படுத்து *paḍuttu* ‘cause to be made’ even though *paḍu* as a marker of passive is not used in ST.

### 0.2.3 Hyphenation.

A word is in order about the manner in which Tamil is hyphenated, since this differs from English style and practice. Before the advent of the printing press, Tamil was written on palm-leaf manuscripts and other flat surfaces without any spacing between what we now recognize as individual words. When space required beginning a new line, scribes began the new line with the next Tamil “character,” irrespective of whether it was the beginning of a new word. No punctuation (hyphens, dashes, etc.) was used. After Tamil began to be printed with moveable type, and Europeans inserted spaces between “words,” breaking still occurred wherever a “character” ended, i.e. in a word like பார்த்துக்கொண்டிருந்தார்கள் after பார் or after த் or after துக் or after க்கொ or after கொண்ட or after இரு or after இருந் or after இருந்தா or இருந்தார் etc. That is, one could break a word after a vowel or after a consonant that was followed by another consonant, but not *before* a vowel, and not before a consonant that was arose from doubling by *sandhi*. The lack of punctuation in this hyphenation process (i.e. hyphenless hyphenation) has been a problem in terms of computerization of the Tamil font, because the hyphenation routine of T<sub>E</sub>X, which formats lines automatically and is not governed by human input, requires that when words are broken, a symbol of some sort, such as a hyphen in a roman font, must be printed at the end of the line. We hesitated to introduce a hyphen into Tamil punctuation, even though Tamil now uses other kinds of roman-font punctuation, since as periods (‘full-stops’), exclamation points, question marks, and commas. Our solution has been to develop a hyphen that is one pixel wide and is therefore invisible when printed. Users of TamilT<sub>E</sub>X need not concern themselves about hyphenation since it will try to break words after a character, in the Tamil tradition. There is no need to insert soft-hyphens, as is sometimes done with plain T<sub>E</sub>X to get rid of lines that are too long because of incorrect hyphenation.

Hyphenation of romanized Tamil, i.e. the spoken Tamil transcription, has been a more difficult problem to deal with, since T<sub>E</sub>X thinks it is dealing with English words unknown to its spelling checker, rather than with another language. We are still working on this.

### 0.2.4 Special symbols.

In the entries a number of special symbols have been used to avoid spelling out repetitious terminology. The symbol  $\rightsquigarrow$  has been used to introduce an example sentence, one that illustrates the use of a verb being defined. The symbol  $\approx$  is used to indicate that what follows is a synonym of the Tamil being defined. In addition a number of abbreviations for grammatical categories are also used.

<sup>28</sup> For more on verbalizers, such as அடி *adi* ‘beat’, see Schiffman 1999 3.15.2.

Special Symbols and Abbreviations	
Symbol or Abbreviation	Use or Meaning
~	Introduces an example sentence
≈	Introduces a synonym
ADJ	Adjectival Participle
INF.	Infinitive
3pn + dat.	Verb is marked for third person; semantic subject is in dative case
s.o.	someone
o.s.	oneself
s.t.	something
esp.	especially
usu.	usually
(inf.)	informal
(metaph.)	metaphorical
(fig.)	figurative
AVP	Adverbial Participle
tr	transitive verb
irreg.	irregular form
def.	defective form
intr	intransitive verb
HS99 (number)	Cf. Schiffman 1999, section indicated

### 0.2.5 Other possible uses of this dictionary

Since this dictionary exists in a database that can be searched, and because there are few other searchable databases of *spoken* Tamil, we anticipate that many uses can be put to the database that will illuminate some features of Tamil that have hitherto been difficult to access. For just one example, there is the fact that as Tamil creates new lexical verbs primarily by compounding existing lexical verbs, or by marking a lexical verb with aspectual morphology, there is much to be learned about this process that this database can contribute to. As we worked through the material ourselves, we discovered that there are many and varied ways to make an intransitive verb transitive (or causative), many of these not discussed in extant grammars of Tamil. We have mentioned this earlier, but more examples can be seen in what follows.

Just one example of this is the way Tamil can add a suffix த்து *-ttu* to an intransitive verb, for which we have many examples in the language, such as படு *paḍu* 'experience, feel' + த்து *-ttu* படுத்தது *paḍuttu* 'cause to experience; make'. This verb then becomes a general 'causativizer' in Tamil, which, combined with other verbs in their AVP form, is found widely throughout the entries here. Another very common example of this is the verb நட *naḍa* 'run, walk' which can be made causative by adding த்து *-ttu* it *-ttu*, i.e. நடத்தது *naḍattu* 'run something, make s.t. go, operate'.

What was not obvious to us beforehand was the existence of many other verbs like this, such as the following:

- விழு *viru* 'fall' வீழ்த்து *viṛttu* 'bring down, defeat, fell'. Note that in this case, the short இ *i* is lengthened when suffixed.
- தாழ் *taar* 'be ruined, decline' தாழ்த்து *taarṭtu* 'ruin, destroy'. By the addition of aspect markers, such as கொள் *kol* 'self-benefactive' we can get தாழ்த்திக்கொள் *taarṭtikko!* 'make lower, discredit, downgrade, debase, devalue; cheapen, abase, humble, humiliate, disgrace, dishonor; behave unworthily; humble o.s.'
- ஆழ் *aar* 'be deep, profound' can be transitivized by adding த்து *-ttu* to get ஆழ்த்து *aar* 'put to shame; further, with various aspect markers, other forms can be derived, such as ஆழ்த்திக்கொள் *aarṭtikko!* 'involve o.s. deeply in; throw o.s. into, immerse o.s. in

- நிகழ் *nikar* means ‘resemble, be similar to;’ by the addition of த்து *ttu* we get நிகழ்த்து *nikar<sub>ttu</sub>* meaning ‘create, form, work (a miracle), deliver a speech’
- கவிழ் *kavir* means ‘turn upside down, invert (o.s.)(intr.)’ and by the addition of த்து *ttu* we can get கவிழ்த்து *kavir<sub>ttu</sub>* ‘derail; overturn or upset s.t., as a boat; turn over, turn upside down, upend, flip/tip/keel over’

Another transitivizer is the verb உஊட்டு *uuttu* ‘feed, nourish, imbue, instill, infuse, provide, nourish, inject or introduce new life or interest into s.t.’ which, in combination with certain verbs (or nouns) expressing emotions, makes new verbs that mean something like ‘propagate, contribute to, create or intensify an emotional state’. In the examples below, we either get a lexical noun such as உயிர் ‘life, life-breath’ compounded with *uuttu* or we get nouns that have been derived from verbs.<sup>29</sup>

- வெறுப்பு *veruppu* ‘hatred’ வெறுப்பூட்டு *veruppuu<sub>ttu</sub>* ‘fan the fires of hatred’. Note that வெறு *veru* is itself a verb; வெறுப்பு *veruppu* is a nominalization formed on the base of வெறு *veru*, which is a transitive verb (6 tr) meaning ‘hate.’
- எச்சரிப்பூட்டு *eccarippu<sub>ttu</sub>* ‘alert, call to a state of readiness, warn, bring to the attention of, make an issue of’
- உயிரூட்டு *uyiru<sub>ttu</sub>* ‘animate, breathe life into, enliven, spark, perk up, liven up, freshen (up)’
- நினைவூட்டு *nenevu<sub>ttu</sub>* ‘call forth/up; bring back to (the) mind; remind; recollect’
- மகிழ்ச்சி உஊட்டு *mahir<sub>cciyu<sub>ttu</sub></sub>* ‘cheer up; inspire or encourage with cheer; make happy; gladden; inject some life into s.o. or s.t., stimulate; (inf.) tickle s.o. pink’
- சோர்வூட்டு *soorvu<sub>ttu</sub>* ‘deject, dispirit; dishearten; disgust’
- அருவருப்பூட்டு *avuv<sub>aruppu<sub>ttu</sub></sub>* ‘cause loathing, aversion or nausea; cause s.o. to feel hate; offend the senses or sensibilities; make dizzy’
- மயக்கமூட்டு *mayakkamu<sub>ttu</sub>* ‘drive mad, crazy’
- கசப்பூட்டு *kasappu<sub>ttu</sub>* ‘embitter, make bitter; cause to feel disappointed, hostile or bitter’
- சக்தியூட்டு *saktiyu<sub>ttu</sub>* ‘energize, give energy to; make energetic’
- சலிப்பூட்டு *salippu<sub>ttu</sub>* ‘irk, weary or annoy; bother; irritate, gall, pique, nettle, exasperate, try s.o.’s patience; anger, infuriate, madden, incense, get on s.o.’s nerves; antagonize, provoke.’

### 0.2.6 A Tamil Thesaurus?

Another idea for future research that emerged during the preparation of this dictionary was that while Tamil lacks a thesaurus, i.e. a dictionary similar to Roget’s Thesaurus of English (Kipfer and Chapman 2001), which groups words by their similarity of meaning, into ‘fields of knowledge’, the database of this dictionary could be used to construct a first-step towards a Tamil Thesaurus.

This could be done by sorting words according to the general synonyms they are provided with. One of the main features of this dictionary is that most main entries are provided with one or more synonyms—verbs similar in meaning to the main entry. Originally many of these synonyms were separate entries but because of considerations of space and volume needed for sound files, we consolidated many examples into synonym files. But some general features were planned in advance. When I was in the early stages of planning this dictionary, and when studying the verbs in Fabricius’ *Tamil-English Dictionary*, I noticed that for every verb that had some kind of meaning associated with sound, he provided the synonym சத்தம் போடு *sattam po<sub>du</sub>*, i.e. ‘make a sound.’ We have continued this tradition, so every verb that involves making a sound is provided by the same synonym (சத்தம் போடு *sattam po<sub>du</sub>*). If the database were searched for this synonym, a large number of verbs having to do with ‘making a sound’ would emerge, and could be brought together under one rubric for the purpose of the thesaurus. Similar studies could be done for other verbs, by first calculating the frequency of certain synonyms, and then sorting by synonym, rather than main entry. Thus a rudimentary Thesaurus for Tamil would emerge, which could be enlarged by consulting other electronic resources for Tamil.

<sup>29</sup> For more on denominal derivation, see Schiffman 2005.

### 0.2.7 Photocomposition.

The final copy of the PDF version of this dictionary has been photocomposed and typeset using the electronic document preparation system  $\LaTeX$ , a subvariety of the called  $\TeX$  system, developed by Donald Knuth. The Tamil fonts used in this system were prepared by the PI together with the help of Thomas B. Ridgeway, then Director of the Humanities and Arts Computer Center (HACC) at the University of Washington, using the METAFONT font generation system, also devised by Knuth. This Tamil  $\TeX$  font is in the public domain and can be copied from those in the possession of any current user. It must be used with the  $\TeX$  system, e.g. with PC- $\TeX$ , Mac $\TeX$ , or any  $\TeX$  system that runs on a mainframe. It includes preview fonts for screen viewing of the  $\TeX$ 'ed document and laser fonts for printing with the usual laser printers. It comes with a pre-processor that converts a romanized transcription of Tamil into macros that can be recognized and processed by  $\TeX$ . Other users will either need to convert their romanized transliterations (in the case of MS-DOS systems) into our romanization system before  $\TeX$ ing, or will need to write their own preprocessor to convert directly from their romanization to  $\TeX$ .

Tamil $\TeX$  has fonts in sizes 7 pt, 10 pt, 12 pt, 14pt, 17pt, and 24 pt), all of these in **bold face**, and *slanted* (which is the default for Tamil.) Other sizes and styles are in preparation.

### 0.2.8 Acknowledgements.

Many people have been involved in various ways in the preparation of this dictionary. Funding for the work done in the U.S. was by the Research Tools Division of the National Endowment for the Humanities; for the work in India, the Smithsonian Institution provided substantial funding. The American Institute of Indian Studies supported the principal investigator on two separate Senior Fellowships for work in India, and the Director of the AHS in India, Pradeep Mehendiratta, provided administrative support for facilitating the disbursement and accounting of the Smithsonian funds for work done in India. The South Asia Studies Program at the University of Washington provided support in early phases of the work, and the U. of W. Graduate School Research Fund supported research assistants at two stages of the work. The Humanities and Arts Computer Center and its Director, Thomas B. Ridgeway, provided valuable support, not the least of which was the production of the METAFONT font for Tamil, as well as other help with computer processing of the data. Stacy Waters of HACC (now known as CARTAH) taught the principal investigator all he knows about  $\TeX$  and  $\LaTeX$  and pitched in numerous times to debug in problematical ways. Steven Graham and Bart Hamilton of the UW Office of Instructional Computing were very generous at various stages. The University of Pennsylvania computing facilities and the Department of South Asia Studies, as well as the Penn Language Center and the Language Resource Center have all been generous with time, expertise and help. More recently, the U.S. Department of Education's Research Division provided substantial support, as did the South Asia Language Resource Center (Chicago) and the Consortium for Language Teaching and Learning (Yale). Edward McDermott, IRS Senior Program Officer, International Research and Studies Program, a division of The International Education and Foreign Language Studies of the U.S. Department of Education has been particularly supportive.

We appreciate the diligent work of Thanam Baba, Charlton Clay, Norman Cutler, Claire Englebrecht, Ganesh Krishnamurthy, Cindy Metzler, Pradeep Murali, and Sandhya Sundhareshan in the tedious process of inputting the Tamil data. Various Tamil-speakers have assisted in evaluating Tamil data and providing the expertise of mother-tongue consultants. S. Ramakrishnan, E. Annamalai, Uma Subramoniam, Sam Sudhanandha, Prathima Christdas, K. Karunakaran, S. Radhakrishnan, S. Arokianathan, and R. Vasu are among the numerous people, named and unnamed, who helped to improve the fit between Tamil and English. R. Vasu, beyond providing deep insight into Tamil in its various forms, has been deeply involved in developing computer support for the preparation of the various versions of the dictionary. We are also deeply appreciative of the lexicographers who have gone before, from Johann Philip Fabricius to Chidambaranatha Chettiar, for their contribution to English-Tamil dictionary making.

# Bibliography

- [1] Anketell, Cyrus P. 1833. *A pronouncing English-Tamil dictionary with abbreviations*. Madras: Society for Promoting Christian Knowledge Press. Iv., 322 pp. (rev. 1898).
- [2] Arden, A. H. 1942 (rev. 1963). *A Progressive Grammar of the Tamil Language*. Madras: Christian Literature Society.
- [3] Brother, T. (ed.) 1852. *A dictionary, English and Tamil*. Vepery (Madras): Christian Knowledge Society Press. Part I: 242 pp. Part II: 272 pp.
- [4] Burrow, T. and M. B. Emeneau 1961. *A Dravidian Etymological Dictionary*. Oxford: the Clarendon Press.
- [5] Chidambaranatha Chettiar, A. (ed.) 1961 *English-Tamil Dictionary*. Madras: University of Madras. xxvii, 1223 pgs.
- [6] Dharmotharan, A. 1977. *Tamil Dictionaries: A Bibliography*. Wiesbaden: Franz Steiner Verlag. BEITRÄGE ZUR SÜDASIEN-FORSCHUNG, Südasien-Institut der Universität Heidelberg, Vol. 50.
- [7] Duverdiere, Gérald. 1978. “L’Imprimerie tamoule de Vépery et les deux dictionnaires de Fabricius.” Gutenberg-Jahrbuch 1978, pp. 189-197.
- [8] Fabricius, Johann Philip and Johann Christian Breithaupt. *A Malabar and English Dictionary, wherein the words and phrases of the Tamulian language, commonly called by Europeans the Malabar language, are explained in English*. Vepery (Madras): 1779. iv, 185. (Revised 1809, 1911, 1933, 1972.)
- [9] Ghosh, Pijush K. 1983. *An approach to Type Design and Text Composition in Indian Scripts*. (Research sponsored by National Science Foundation IST-8201926). Stanford: Computer Science Department, Rep. No. STAN-CS-83-965.
- [10] Hymes, Dell. 1971. *Pidginization and Creolization of Languages*. London and New York: Cambridge University
- [11] Kipfer, Ann and Robert L. Chapman (eds.) 2001. *Rogets international thesaurus*. New York: HarperResource Press.
- [12] Knight, Joseph and L. Spa[u]lding. 1842. *An English and Tamil dictionary or Manual lexicon for schools, giving in Tamil all important English words, and the use of many phrases*. Jaffna: American Mission Press, 970 pp. Revised by S. Hutchings, Jaffna and Madras, 1844, xxii, 831 pp. 3rd edition, American Mission Press 1852. 970 pp. Rev. 1888, repr. 1890.
- [13] Knuth, Donald. 1979. *TEX and METAFONT: New Directions in Typesetting*. American Mathematical Society: Digital Press.
- [14] Labov, William. 1971. “The notion of ‘System’ in Creole Studies.” in Hymes 1971.
- [15] Lisker, Leigh. 1958. “The Tamil Occlusives: short vs. long or voiced vs. voiceless,” IL, 19.
- [16] McAlpin, David. 1976. *A Core Vocabulary for Tamil*. Department of S. Asian Regional Studies, University of Pennsylvania. Final Report, USOE Contract 300-75-0314.
- [17] Muttu-Tambi Pillai, M.A. 1907. *A Dictionary of the English language containing all English words and phrases now used with their meanings, synonyms and Tamil equivalents*. Jaffna, Naval Press, 1907. iv, 1180 pp. Revised frequently, e.g. Jaffna 1911.

- [18] Ochterlony, John 1851. *A pocket dictionary of English and Tamil, giving the nomenclature employed as well by the lower as the higher classes of natives, the Tamil words being printed both in English and in the native character*. Vepery (Madras): Christian Knowledge Society Press.
- [19] Paramasivam, K. 1979. "Effectivity and Causativity in Tamil." *International Journal of Dravidian Linguistics* 8(1):71-151.
- [20] Percival, Peter. 1861 *A dictionary, English and Tamil*. Madras: Madras School Book and Literature Society, 1935. 441 pp.
- [21] da Proença, Antam. *Vocabulario Tamulico com a significacam Portuguesa. Na impressa Tamulica da Provincia do Malabar, por Ignacio Aichamoni impressor della*. 1679. 247 fo. New edition, Anto da Proença's Tamil-Portuguese dictionary, A.D. 1679. Kuala Lumpur: University of Malaya, 1966. 20 pp. 247 fo.
- [22] Pope, George U. 1906. *An English-Tamil dictionary. Being Part IV of 'A Handbook of the ordinary dialect of the Tamil language'*. Oxford: Clarendon Press, 1906. 108 pgs.
- [23] Rajaram, S. *English-Tamil Pedagogical Dictionary*. 1986. Thanjavur: Tamil University.
- [24] Sankaranarayana Chettiar, Paluri. *An English-Tamil dictionary with copious English synonyms and brief and accurate definitions*. Madras: P. Kameswara Row Bros., 1908. 1130 pp. 2nd ed., Madras 1911-12, 1378 pp.
- [25] Schiffman, Harold F. 1975. On the Ternary Contrast in Dravidian Coronal Stops." in H. Schiffman and C. M. Eastman (eds.) *Dravidian Phonological Systems*, Seattle, Institute for Comparative and Foreign Area Studies, pp. 69-85.
- [26] Schiffman, Harold F. 1980. The Tamil Liquids," in *Proceedings of the Berkeley Linguistics Society*, 6:100-110. (Emeneau Felicitation Volume). Berkeley: Department of South and South east Asian Languages.
- [27] Schiffman, Harold F. 1993. "Intervocalic V-deletion in Tamil: Its Domains and its Constraints." *Journal of the American Oriental Society* 113(4)513-528.
- [28] Schiffman, Harold F. 1998: "Standardization and Restandardization: the case of Spoken Tamil." *Language in Society*, Vol. 27 (3) 359-385.
- [29] Schiffman, Harold F. 1999. *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
- [30] Schiffman, Harold F. 2005. "Deverbal Nominal Derivation in Tamil" *International Journal of Dravidian Linguistics*, Vol. 34, No. 2, June 2005, pp. 159-166.
- [31] Narayan, Shrikanth, Dani Byrd, and Abigail Kaun. 1999 "Geometry, kinematics, and acoustics of Tamil liquid consonants." *Journal of the Acoustic Society of America*, Vol. 106, No. 4, Pt. 1.
- [32] Subramanian, P.R. (ed.) 1992. *Kriyavin tarklat Tamil akarti: Tamil-Tamil-Aangilam*. [Dictionary of contemporary Tamil-Tamil-English.] Madras: Kriyaa.
- [33] Visvanatha Pillai, V. 1888. *Dictionary, English and Tamil*. Revised ed., London, 676 pp. Revised 1963.
- [34] Zgusta, Ladislav. 1971. *Manual of Lexicography*. Mouton: The Hague. 360 pp.