Chapter 1

Phonology and Transliteration

1.1 Introduction

1.1.1 Background

1.1.1.1 Standard Spoken Tamil

This grammar assumes that there exists a variety of spoken Tamil that is ‘standard’ alongside the long-since standardized LT variety (LT). This is a somewhat problematic assumption. Many linguistic scholars have approached 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 ‘emerging’ 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, in situation comedies. It is the variety that is used when 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 to other Tamils from vastly different parts of Tamilnadu. An attempt to be comprehensible to the largest number of speakers 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 higher-caste, educated speech of non-Brahman groups in Tamilnadu; according to some it is neither from the far north (i.e. Madras) or from the southernmost reaches of Tamilnadu (e.g. Kanyakumari District), but rather from urban areas
CHAPTER 1. PHONOLOGY AND TRANSLITERATION

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 apparently tend to lean more toward LT, and may choose a form that is not actually found in any spoken regional or social dialect, but is known from LT. Since LT is the form that all educated speakers know, it can be a repository from which general forms can be chosen; this is another aspect of Labov’s maxim (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.1

For some, including both researchers and speakers of Tamil, Tamil is not ‘standardized’ because it has not been codified by a committee or a board or an eminent person, or because a standard has not been declared and disseminated by the school system or whatever; or because a ‘book’ has not been written called A Grammar of Spoken Tamil. In fact, I claim (Schiffman 1998) that Spoken Tamil (ST) has become standardized by a process of informal consensus, in the same way that other diglossic languages that possess ancient standard literary languages have evolved modern spoken koinés. It is in fact quite easy to get Tamil speakers to agree that certain forms are preferred and others are dispreferred; there is remarkable unanimity in this area, wherever Tamil is spoken, with the exception of Sri Lanka. The film, and spoken drama groups before it, have been responsible for the evolution and dissemination of this consensual standard.

For example, speakers may model their choice of the past neuter form of verbs on the LT past နோடு adu, e.g. வொடு—னோடு vand-adu, rather than the form found very commonly in many non-Brahman dialects, i.e. -cci or -cci, e.g. vand-cci ‘it came’ (which is not found per se in LT with this verb, but has spread from Class III verbs, or from the prototypical pasts in -cci of verbs like போ போ ‘go’ and செம்மை செம்மை ‘become’, which have spoken pasts pooccusu and aacusu (from LT போ and செம்மை, respectively). Other speakers may choose the cci/ti forms unequivocally, so that no hard and fast rules can be given for many forms.

In fact even though we conclude that some consensus may exist as to what ST entails, the situation must be described as being variable and fluid. Individual speakers may vary considerably in their own speech, depending upon whom they are talking to, their gender (or the gender of their interlocutor) or what the topic of conversation is. These phenomena have been noted by many linguists working in the field of sociolinguistics, and are not

1 It is interesting to note that though some writers deny that ST is standardized in any way, the variety they describe in their writings is extremely close to what is described here. For example, the variety Asher (1982) describes, though he claims it is not possible to say it represents a standard, happens, not by chance, to resemble closely what I would call standard.
1.2. PHONETICS OF ST

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, their age, their occupation, whom they are talking with, and any other social markers one may isolate.\(^2\)

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 they 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.

1.2 Phonetics of ST

There are a number of descriptions of the phonetics of ST available in various sources (Asher 1982, Rajaram 1972, Firth 1934 (in Arden 1934)); it is useful to point out where many of them lean in the direction of LT pronunciation, even when they profess to be describing ST. Thus, though we would like to concentrate on describing ST, we must often do it by contrasting it with LT, in order to emphasize the differences, which do not often get adequately described.

1.2.1 Tamil Vowels

ST has a ten-vowel system with long and short \( \text{\textipa{i}} \equiv \text{\textipa{I}}, \text{\textipa{e}} \equiv \text{\textipa{E}}, \text{\textipa{a}} \equiv \text{\textipa{A}}, \text{\textipa{a}a}, \text{\textipa{a}o}, \text{\textipa{a}o}, \text{\textipa{u}} \equiv \text{\textipa{U}}.\(\)\(^3\) We use throughout this grammar the double-letter representation of long vowels, except where we are being explicitly phonetic. That is, \( \text{\textipa{a}} \) will be transcribed as \( \text{\textipa{a}a} \) except when a purely phonetic representation is wanted; then it will be \( \text{\textipa{a}}.\)

\(^2\)Many people have contested the notion that the Tamil social film is in ‘standard’ spoken Tamil because of the variety of dialects, some of them deliberately used for humorous or other effect, found there. To this I would reply that in most of these films, the main characters (hero, heroine, perhaps other friends or kin) speak SST; other characters around them are ‘character actors’ and use the non-standard, rural, rustic, or other dispreferred varieties of speech, for deliberate effect of some sort. (In fact, many films deliberately lampoon the non-standard forms; certain character actors, such as the famous Nagesh, specialized in this.) Thus the film provides not only a model of standardness or correctness (the main characters) but also a model of speech to be avoided.

\(^3\)The diphthongs \( \text{\textipa{ai}} \) and \( \text{\textipa{au}} \) found in LT are not usual in ST; a few loan words contain \( \text{\textipa{au}} \), but often these can be represented by \text{\textipa{a}u}, as in \text{\textipa{a}u} pattrayu ‘pound.’
CHAPTER 1. PHONOLOGY AND TRANSLITERATION

1.2.1.1 Initial Position

All vowels may appear in initial position in both ST and LT, but there is one major difference between LT and ST in that high short vowels (இ i and ன u), when followed by a single consonant and the vowel ா a, are replaced by எ e and ன o respectively. That is, the high vowels are lowered to mid-vowel position if they occur in the first syllable. (They may be preceded by a consonant, but are not required to be.) This means that there are no words in ST that now begin with short high vowels followed by a single consonant and the vowel ா a or ன ai (which in ST usually is replaced by எ e).

- LT எ_ எம எம எம 'place'
- LT எ_வம எம 'body'
- LT எ_வெல 'leaf'
- LT எ_வெல கூரக்கு 'child'
- LT எ_வெல எய 'think'

Note that in these forms, LT final ன ai’s all become எ e.

1.2.1.2 Final Position

In ST all ‘words’ now end in a vowel (unlike LT where words could end in sonorants and glides) so a number of phonetic changes have taken place in final position.

1.2.1.3 Words Ending in a Vowel plus Nasal

Words of more than one syllable ending in a vowel plus the nasal consonants ஊ m and ஊ n change to nasalized vowels, and the nasal segment is deleted.\(^4\)

Generally, long vowels retain their same quality, but add nasalization. Short vowels may change, e.g. undergo rounding, fronting or some other phonetic process.

- Thus words ending in ஊ ஐம such as ஊமம maram ‘tree’ are pronounced [o] in final position, i.e. maram is phonetically [məɾə].
- Long ஊ ஑ before ஊ m retains the low-central quality: சொலமொல poohalaam ‘let it go’ is phonetically [poohalai].

\(^4\)This does not occur with final retroflex ஊ ஑ ; such items have an epenthetic ஊ added (cf. §1.2.5 below).
1.2. PHONETICS OF ST

- Long ə oo before ʋ ʋ retains the tense mid back position: இரக்கூம் ‘we are’ irukkroom is [irukrőː].
- Long ə ee before ʋ ʋ retains the tense mid front articulation: இரக்கிழ் irukkree ‘I am’ is [irukrē];
- Short ə u after ʋ ʋ remains high, back and rounded: பொகும் poohum ‘it will go’ is [poːhū];
- Short ə ə before ʋ ʋ is fronted in the dialects of many speakers, to [e] or [ɛ]: அவன் avan ‘he’ is [avɛ] or [avē];
- Short ə e before ʋ ʋ occurs mainly in the item ஆன் en ‘my’, where it is pronounced [ˈɛ].
- Short ə o before ʋ ʋ occurs mainly in the item ஓன் on ‘your’, where it is pronounced [ˈo];

- There are few if any occurrences of ஐ uu, or of short ə i before nasals in final position; LT ஓீ un ‘gum’ follows the pattern of adding ஐ.

Pronouns that end in nasal consonants, such as நான் naan, இன் en etc. behave differently from other words; they undergo nasalization (to [nāː], [ɛ], etc.) irrespective of syllable count or vowel length, but other words like உண்ம் miinu ‘fish’ do not; epenthetic ə u is added instead.

1.2.1.4 The Epenthetic Vowel ə u

When all other methods have been exhausted, Tamil can always make a word end in a vowel by adding the so-called ‘epenthetic’ (or ‘nunciative’) vowel ə u to any word that does not already have a final vowel. This ‘fleeting’ vowel is often present only in isolation, i.e. before a pause. If the word is joined to another, this vowel then disappears. Phonetically, this vowel is usually an unrounded high back lax vowel [uː], [u] or [i]. In fact, Tamil pronounces all orthographic ʋ’s as [u] after the first syllable of a word, except for final ʋ ʋ’s in some names, chiefly male nicknames. Furthermore, many ʋ’s are also pronounced [u] or [i] in similar positions, so words like சப்தத்திருக்கு soappidūkkirukku ‘it is eating’ is actually phonetically [sapidūkkukku], i.e. all the vowels after the first syllable are identical, or in

5 This is also a pattern in some Indo-Aryan languages, and may be borrowed. Thus றூலாலாத்தோம் baalaathu may be shortened to றூலா baalu, வாசஜன் raajendran to வாசஜ raaju, etc.
CHAPTER 1. PHONOLOGY AND TRANSLITERATION

some cases have been deleted. This vowel has been unrounded for so long that most speakers of Tamil pronounce it this way in LT as well as in ST, i.e. it is not an ST innovation. What is different in ST is the business of making ⟨ə⟩ i into [i] or [u] as well.

1.2.1.5 Words Ending in Sonorants, Liquids and Glides

In LT, words may not end in a consonant, but they may end in sonorants of various sorts, liquids and rhotics, and glides such as ⟨ѡ⟩ y. In ST, words that in LT end in liquids such as ⟨ѡ⟩ l and ⟨strings⟩ l usually double the liquid consonant if it is a monosyllabic word with a short vowel, or delete it in final position if it is polysyllabic. If it is monosyllabic but with a long vowel, epenthetic ⟨ə⟩ u is added. In some dialects, the liquid is deleted. Many pronouns end in ⟨strings⟩ l and their final laterals are always deleted before pause, but reappear if a suffix is added.

- LT ⟨இ⟩ nil ‘stand’ —— ST ⟨இ⟩ nillu
- LT ⟨இ⟩ aal ‘man’ —— ST ⟨இ⟩ aalu
- LT ⟨இ⟩ kal ‘toddy’ —— ST ⟨இ⟩ kalu
- LT ⟨இ⟩ naal ‘day’ —— ST ⟨இ⟩ nalu or ⟨இ⟩ naa (in some dialects)
- LT ⟨இ⟩ niingal ‘you (PL)’ —— ST ⟨இ⟩ niinga
- LT ⟨இ⟩ vayal ‘field’ —— ST ⟨இ⟩ vayalu
- LT ⟨இ⟩ vaasal ‘gate, door’ —— ST ⟨இ⟩ vaasalu

The variability of deletion or non-deletion of final laterals is perhaps greater than any other consonant-final situation in Tamil; no other final consonants display this amount of variation. One other kind of change seen in some dialects is that ⟨strings⟩ l may simply be replaced by ⟨ѡ⟩ l across the board. This model of neutralization is dispreferred in ST, so we will not give examples of it. It is, however, a pedagogical problem wherever Tamil literacy is taught, since some speakers simply have no contrast in their dialects.

6 But many of these ‘deleted’ consonants then reappear if something is added, especially since ⟨strings⟩ l is often used in pronouns of various sorts and as a PNg marker on verbs; cf. § 1.3 below.
1.2. PHONETICS OF ST

1.2.1.6 Rhotics, Final and Otherwise

The question of what is an r-like sound, and what is not, is a thorny question in Tamil. Tamil possesses, besides the laterals ɭ and ɭ, some other sounds that are phonetically related, but because of variability and inconsistency in their pronunciation in some dialects, and because of some across-the-board changes in ST as compared to LT, this area is fraught with sociolinguistic complexity.\(^7\)

In LT, there was originally a phonetic contrast between three r-like consonants:

- ğ r, a phonetically flapped or tapped r, more or less alveolar, phonetically [ɾ]. This sound could not occur in initial position in older forms of LT, but in ST occurs initially, medially, and finally (but usually finally is followed by epenthetic [u] before pause.) This sound does not undergo gemination (doubling).

- ɭ r was originally an alveolar stop in Proto-Dravidian and in older forms of Tamil. It did not occur initially or finally (just like retroflex stops) and when geminated had the value of a voiceless alveolar stop: ɭɭ was phonetically [ʈʈ] or [ʈ] with r-like offset: [ʈɾ]. Intervocically, ɭ was trilled: [ɾ], and in some dialects, mainly southern (Kanniyakumari etc.) a real phonetic contrast between this sound and the previous one is maintained. However, in the speech of most Tamil speakers, a phonetic contrast is not maintained, even if speakers claim that they do so. This sound is orthographically maintained when writers depict spoken Tamil in plays and novels, since it is the marker, among other things, of the present tense. We therefore maintain it for spelling contrast in our transcriptions, e.g. the two ‘r’s of வரூடம் varūram are not phonetically distinct, but we write them as ğ and ɭ (in Tamil script) because Tamil linguistic culture prefers this. This sound is only found in native-Dravidian lexica.

- Another sound that is sociolinguistically complex is the retroflex frictionless continuant ģ r which, under ideal conditions, is phonetically [ɭ].\(^8\) In contemporary Tamil, many speakers replace this sound totally with the retroflex lateral ɭ and a plain lateral [ɭ] has also been the symbol used by most Europeans for the name of the language: ‘Tamil’ is தமிழ் tamir; the ‘Chola’ Kingdom is சோழ coora, but there is also the item ‘Coromandel’ (from கோரமண்டல கோரமண்டலம் cooramandalam)

---

\(^7\)I have tried to deal with this in my 1980 paper, ‘The Tamil Liquids.’
\(^8\)This was the symbol used by Firth (1934) in his appendix to Arden’s grammar (Arden 1934:xvi).
CHAPTER 1. PHONOLOGY AND TRANSLITERATION

for the coast south of Madras, with an /tʃ/ instead of an /l/. Because of the ‘mystique’ surrounding this sound (Tamilis seem to believe it is ‘unique’ in Tamil) it is learned only through literacy by many speakers, and even then, some never master it. Therefore, if foreigners can learn it, it gets them good karma, so we use ṭ r wherever it occurs in LT, even though it is rare in ST. Tamil second-language learners should note, however, that forms like அரங்கு கோண்டே ‘child’, அரங்கு வள்ளச் சம்பூ ‘banana, plantain’ are more likely to be pronounced [kolande] and [valːtːpːe] than with corresponding ṭ r.9

The ṭ r sound is never geminated (another reason to consider it to be an r-sound) and does not undergo many of the morphophonemic rules that apply to ķ l, e.g. it does not become replaced by nasals, or become a stop the way ķ l does. Neither sound occurs in initial position, both because rhotics and liquids did not in LT, and because retroflex consonants never occur in initial position in native-Dravidian words. In some cases, intervocalic ṭ r may be deleted, with compensatory lengthening of the vowel: LT அரங்கு porudu ‘time’ — ST அரங்கு poodu.

Our solution, therefore, is always to distinguish these three sounds, and the two l’s, in all our transcriptions, even though many of them may be neutralized in many people’s speech.

1.2.2 Nasal Consonants

LT script distinguishes six different nasal consonants, but ST only has three phonemically distinct nasal sounds.10 LT also distinguishes a dental nasal ṭ from an alveolar nasal ķ, but even in LT these two are in complementary distribution: the ṭ occurs initially and before ķ t, while the alveolar nasal occurs finally and before ķ r. Thus in our transcription, we only distinguish between m, n and ŋ; the palatal and velar nasals we simply write as n before the appropriate consonant, and we transliterate both ṭ and ķ as n, since no Tamil speakers (despite claims otherwise) distinguish between these two sounds.

As noted above, if LT final ķ n occurs in final position, in ST an epenthetic ķ n is added to the word (and the consonant is doubled if the

9 I have heard, however, a hawk selling bananas on a railway station platform in Trichy, with clear [tʃ] in all the appropriate places, i.e. [vaːɾːpːe].
10 In LT as well, the nasal ṭ [ʃ] only occurs before velars, i.e. ķ k, and the palatal nasal ṭ ŋ generally only occurs before palatal ķ c. However, in very few words, such as அரங்கு சாஸ்பே ‘Sunday’ and in some borrowed words like அரங்கு காண்டி ‘sage’ the initial palatal nasal does occur without any conditioning.
1.2. PHONETICS OF ST

stem vowel is short), e.g. வெள்ள pen ‘woman’ — வெள்ளம் ponnu ‘ibid’.\textsuperscript{11}

All other sequences of vowel plus nasal in final position undergo nasalization of the vowel as described previously.

1.2.3 Glides

Glides, particularly ஏ y are also generally deleted in final position except in monosyllabic words, where they are instead doubled: LT கூ க்கை ‘do, make’ — கூக்கிச் seyyi, நாள் ruhbag ‘rupee’ — [ruːbaj] (or [ruːwaj] ‘ibid.’).

1.2.4 Stop Consonants

In LT, as in Proto-Dravidian, it seems clear that there was a series of six stop consonants:

- Velar: ṭ k
- Palatal: ṭ c or s
- Retroflex: ṭ t
- Alveolar: ṭ t
- Dental: ṭ t
- Labial: ṭ p

In initial position (except for ṭ and ṭ) and when geminated, the above LT stops are phonetically voiceless (and unaspirated). When they occur after a nasal, all are voiced, and somewhat more lax. Intervocally, they are laxed, and with the exception of ṭ and ṭ, voiced. This is shown in Table 1.1.

In ST, some things have changed, as shown in Table 1.2. Alveolar ṭ when geminated, i.e. ṭṭ tt, has now merged with ṭṭ tt: பாதி patti is [patti], etc. Its other (more r-like) phonetic realizations have merged phonetically with ṭ r, except in southern dialects. Initial ṭ which used to be unambiguously c, i.e. [ɕ], now varies widely; some speakers have [s] in initial position only, with the affricated pronunciation reserved for geminate [cc]. Other speakers have [ɕ]\textsuperscript{12} before certain vowels and [ɾ] before others, e.g. சிவன் cinna [cinna] ‘small’ but சராவி saari [sarvi] ‘key.’ Some speakers

\textsuperscript{11}Note that the vowel also undergoes rounding because it occurs between a labial and a retroflex consonant.

\textsuperscript{12}In all cases, the symbol [c] is an affricated stop [ʧ] similar to English ‘ch’ in ‘cheese.’
CHAPTER 1. PHONOLOGY AND TRANSLITERATION

Table 1.1: Phonetics of LT Stop Consonants

<table>
<thead>
<tr>
<th>Stops</th>
<th>Initially</th>
<th>Medially and Geminated</th>
<th>Between Vowels</th>
<th>After Nasal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar &amp; k</td>
<td>[k]</td>
<td>[kk]</td>
<td>[h]</td>
<td>[g]</td>
</tr>
<tr>
<td>Palatal &amp; c</td>
<td>[c]</td>
<td>[cc]</td>
<td>[s]</td>
<td>[j]</td>
</tr>
<tr>
<td>Retroflex ̂ t</td>
<td>...</td>
<td>[tt]</td>
<td>[d]</td>
<td>[d]</td>
</tr>
<tr>
<td>Alveolar ̃ t</td>
<td>...</td>
<td>[tt], [tr]</td>
<td>[r]</td>
<td>[r]</td>
</tr>
<tr>
<td>Dental ø t</td>
<td>[t]</td>
<td>[tt]</td>
<td>[Ñ]</td>
<td>[d]</td>
</tr>
<tr>
<td>Labial ω p</td>
<td>[p]</td>
<td>[pp]</td>
<td>[b], [y]</td>
<td>[b]</td>
</tr>
</tbody>
</table>

Table 1.2: Phonetics of ST Stop Consonants

<table>
<thead>
<tr>
<th>Stops</th>
<th>Initially</th>
<th>Geminated</th>
<th>Medially</th>
<th>Between Vowels</th>
<th>After Nasal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar &amp; k</td>
<td>[k]</td>
<td>[kk]</td>
<td>[h] or [...]</td>
<td>[g]</td>
<td></td>
</tr>
<tr>
<td>Palatal &amp; c</td>
<td>[c] or [s]</td>
<td>[cc]</td>
<td>[s]</td>
<td>[j]</td>
<td></td>
</tr>
<tr>
<td>Retroflex ̂ t</td>
<td>...</td>
<td>[tt]</td>
<td>[d]</td>
<td>[d]</td>
<td></td>
</tr>
<tr>
<td>Alveolar ̃ t</td>
<td>...</td>
<td>[tt]</td>
<td>[r]</td>
<td>[r]</td>
<td></td>
</tr>
<tr>
<td>Dental ø t</td>
<td>[t]</td>
<td>[tt]</td>
<td>[Ñ]</td>
<td>[d]</td>
<td></td>
</tr>
<tr>
<td>Labial ω p</td>
<td>[p]</td>
<td>[pp]</td>
<td>[b], [y], [w]</td>
<td>[b]</td>
<td></td>
</tr>
</tbody>
</table>

have only [c] except intervocally; for many speakers, other sounds are also merged with & c, i.e. they have no contrasts such as ϕä 5 or øi j.

Many speakers also have variation in their pronunciation of intervocalic ω b—sometimes we get [b], sometimes we get [v] or even [w]. This seems to depend on the degree of indigenousness of certain borrowed words, as with [ruwa] ‘rupee’ above.

1.2.4.1 Deletion of Intervocalic ø v and ø k

It must be noted that ST fairly systematically deletes intervocalic ø v and ø k (phonetically [h]) in certain environments. The rules are very complicated, but for ø v one can state that if it is not the expression of a future or causative marker of weak verbs, or the product of sandhi (cf. §1.3 below), ø v and the vowel that follows it will probably be deleted after the