

Figure 9-20. Tamil Ligatures with *ra*

$$\begin{aligned} \text{ர} + \text{◌} &\rightarrow \text{ர} \text{ } r \\ \text{ர} + \text{◌ி} &\rightarrow \text{ரி} \text{ } ri \\ \text{ர} + \text{◌ீ} &\rightarrow \text{ரீ} \text{ } rī \end{aligned}$$

However, various governmental bodies mandate that the basic shape of the consonant *ra* ர should be used for these ligatures as well, especially in school textbooks. Media and literary publications in Malaysia and Singapore mostly use the unchanged form of *ra* ர.

Ligatures with aa in Traditional Tamil Orthography. In traditional Tamil orthography, the vowel sign *aa* ◌ஶ optionally ligates with ஶ்ஶ, ஶ், or ஶ், as illustrated in Figure 9-21.

Figure 9-21. Tamil Ligatures with *aa*

$$\begin{aligned} \text{ஶ} + \text{◌ஶ} &\rightarrow \text{ஶ்ஶ} \text{ } nā \\ \text{ஶ} + \text{◌ஶ} &\rightarrow \text{ஶ்} \text{ } nā \\ \text{ற} + \text{◌ஶ} &\rightarrow \text{ற்ஶ} \text{ } rā \end{aligned}$$

These ligations also affect the right-hand part of two-part vowels, as shown in Figure 9-22.

Figure 9-22. Tamil Ligatures with *o*

$$\begin{aligned} \text{ஶ} + \text{◌ொ} &\rightarrow \text{ஶொ} \text{ } no \\ \text{ஶ} + \text{◌ோ} &\rightarrow \text{ஶோ} \text{ } nō \\ \text{ஶ} + \text{◌ொ} &\rightarrow \text{ஶொ} \text{ } no \\ \text{ஶ} + \text{◌ோ} &\rightarrow \text{ஶோ} \text{ } nō \\ \text{ற} + \text{◌ொ} &\rightarrow \text{றொ} \text{ } ro \\ \text{ற} + \text{◌ோ} &\rightarrow \text{றோ} \text{ } rō \end{aligned}$$

Ligatures with ai in Traditional Tamil Orthography. In traditional Tamil orthography, the left-side vowel sign *ai* ◌ஶ is also subject to a change in form. It is rendered as ◌ஶ when it occurs on the left side of ஶ்ஶ, ஶ், ஶ, or ஶ், as illustrated in Figure 9-23.

Figure 9-23. Tamil Ligatures with *ai*

ண + ை ◯ → னை *nai*
 ண + ை ◯ → னை *nai*
 ல + ை ◯ → லை *lai*
 ள + ை ◯ → ளை *lai*

By contrast, in modern Tamil orthography, this vowel does not change its shape, as shown in Figure 9-24.

Figure 9-24. Vowel *ai* in Modern Tamil

ண + ை ◯ → ணை *nai*

Tamil aytham. The character U+0B83 TAMIL SIGN VISARGA is normally called *aytham* in Tamil. It is historically related to the *visarga* in other Indic scripts, but has become an ordinary spacing letter in Tamil. It is used to modify the sound of other consonants and, in particular, to represent the spelling of words borrowed into Tamil from English or other languages.

Punctuation. Danda and double danda marks as well as some other unified punctuation used with Tamil are found in the Devanagari block; see Section 9.1, *Devanagari*.

Tamil Named Character Sequences

Tamil is less complex than some of the other Indic scripts, and both conceptually and in processing can be treated as an atomic set of elements: consonants, stand-alone vowels, and syllables. Table 9-21 shows these atomic elements, with the corresponding Unicode characters or sequences. In cases where the atomic elements for Tamil correspond to sequences of Unicode characters, those sequences have been added to the approved list of Unicode named character sequences. See NamedSequences.txt in the Unicode Character Database for details.

In implementations such as natural language processing, where it may be useful to treat such Tamil text elements as single code points for ease of processing, Tamil named character sequences could be mapped to code points in a contiguous segment of the Private Use Area.

In Table 9-21, the first row shows the transliterated representation of the Tamil vowels in abbreviated form, while the first column shows the transliterated representation of the Tamil consonants. Those row and column labels, together with identifying strings such as “TAMIL SYLLABLE” or “TAMIL CONSONANT” are concatenated to form formal names for these sequences. For example, the sequence shown in the table in the K row and the AA column, with the sequence <0B95, 0BBE>, gets the associated name TAMIL SYLLABLE KAA. The sequence shown in the table in the K row in the first column, with the sequence <0B95, 0BCD>, gets the associated name TAMIL CONSONANT K.

Details on the complete names for each element can be found in NamedSequences.txt.