Theoretical and Practical Aspects of Chillus in Malayalam

K. P. Mohanan

National University of Singapore

ellkpmoh@nus.edu.sg

This is a brief note triggered by Rajeev J. Sebastian's paper "Atomic Chillus Cause Spoofing" presented at the workshop on Problems of Malayalam encoding in Unicode held at the University of Kerala, 24-25 Jan 2007. I will deal with the theoretical aspects of chillus first, and then proceed to outline my views on the practical aspects in orthographic representations created by typing on a computer keyboard.

1 Phonology and Orthography

There are two important characteristics of Malayalam orthography that are central to the debate on chillus. First, the traditional Malayalam orthography is a syllabary, it is not alphabetic; second, Malayalam orthography is phonemic, not morphophonemic.

1.1 Alphabets and Syllabaries

In an alphabetic system of writing, as in English, an orthographic character typically represents a single "phoneme" or sound. Thus, the orthographic representation *best* has four letters (*b*, *e, s,* and *t*), each representing a sound. This corresponds to the linguist's phonological representation of the pronunciation of the word as /best/.

In a syllabary, such as in Malayalam, each letter typically represents a single syllable. Thus, the Malayalam word for frog, തവള pronounced as /tawaLa/ has six sounds, but three syllables (ta – wa – La), and it is represented by three orthographic characters, one for each syllable.

1.2 Morphophonemic and phonemic systems

In a morphophonemic system of writing, each morpheme has a single representation, regardless of its phonemic realization. Take, for instance, the English words *divine* and *divinity*, the second word being composed of the morphemes divine and -ity. In the word *divine*, the morpheme is pronounced as /divain/, but in the word *divinity* the same morpheme is pronounced as /divin/, with a different vowel in the second syllable. Regardless of this difference, the orthographic system of English uses the same character *i* in the second syllable of divine for both words. Likewise, the pronunciation of the morpheme hymn in the word *hymn* /him/ does not have a final /n/, but the same morpheme in the word *hymnal* /himnəl/ does have a final /n/ sound. The orthographic representation has the character *n* for the morphophonemic /n/ regardless of its phonemic realization.

In contrast, Malayalam orthography is based on the phonemic level of representation (with a few minor exceptions). To illustrate, take the compound word മഹോത്സവം /mahootsawam/ 'great festival', derived from the stems മഹാ /mahaa/ 'great' and ഉത്സവം /utsawam/ 'festival'. The phonemic representation of this word has four syllables (മ ma – ഹോ hoo – ത്സ tsa – വം wam). Traditional Malayalam orthography represents this word with four characters, with a vowel diacritic for the long vowel ോ /oo/ on the second character, and another diacritic for ം /m/ on the fourth. Had Malayalam orthography been morphophonemic, we would have represented the word with five characters, with a diacritic for /aa/ on the second syllable, and the character for /u/ as the third syllable: /മ ma – ഹാ

haa – ഉ u – ത്സ tsa – വം wam/. Likewise, the orthographic representation of മരക്കുതിര /marakkutira/ ʻwooden horse' derived from the stems മരം /maram/ ʻwood' and കുതിര /kutira/ ʻhorse' corresponds to the phonemic level /മ ma – ര ra – ക്കു kku – തി ti – ര ra/. Were it a morphophonemic system, the characters would correspond to /മ ma – രം ram – കു ku – തി ti – ര ra/ with a final ം /m/ in the second syllable, and no double consonant in the third.

2 What is a chillaksharam?

Without a diacritic, a consonant letter by itself is interpreted in Malayalam orthography as being accompanied by the vowel sound അ /a/. Thus, the three characters for the word തവള /tawaLa/ ʻfrog' each represents a /C+a/. When a consonant letter needs to be represented with a different vowel (as in the word കിളി /kiLi/ ʻbird', for instance), we use a vowel diacritic. When a consonant letter needs to be represented without any accompanying vowel (as in the final consonant of രാജേഷ് /raajeeS/ ʻa name') we use the chandrakkala as diacritic.

A chillaksharam is a special character that represents phonemes ൺ /N/, ൻ /n/, ൽ /l/ , ൾ /L/ or ർ /R/ when not accompanied by a vowel, as an alternative to the more common strategy of the chandrakkala used for the other phonemes. Take, for instance, the words അവൻ /awan/ ʻhe', അവൾ /awaL/ ʻshe' and അവർ /awaR/ ʻ they'. These words have two syllables each (അ a – വൻ wan; അ a – വൾ waL; അ a – വർ waR), and yet, each of them is traditionally represented with three instead of two characters, the last character being a chillu.

Notice that there is no chillu for the consonant sound ം /m/: the word മരം /maram/ ʻtree' (or ʻwood') is represented with two characters corresponding to two syllables (/മ ma – രം ram/); an *anuswaram* represents the final ം /m/ in the second syllable. Note also that there are no chillus for sounds other than /ൻ n, ൺ N, ൽ l, ൾ L, ർ R/: all the other sounds use chandrakkala for the representation of consonant sounds unaccompanied by vowels.

3. Chandrakkala with samvrithokaaram

In traditional orthography, consonant + chandrakkala when combined with the diacritic for /u/ (a small circle under the character) represents what Malayalam grammaricans call ʻsamvrithookaaram', which is the neutral vowel that linguists call ∂ ʻschwa'. Thus, when a word final consonant character has the diacritic for /u/ plus a chandrakkala, the word is pronounced with a schwa (as in the case of കാട് /kaaT∂/ ʻforest', but when there is no diacritic for /u/, there is no vowel after the consonant (as in the case of പാർട് /paaRT/ ʻpart'.)

Thus, when a word final consonant character has a chandrakkala + diacritic for /u/ (e.g, ₍uC^u സ്), the word is pronounced with a schwa (as in the case of അവന് /awan∂/ 'he-dative', രാജന് /raajan∂/ 'rajan-dative' but when there is no diacritic for /u/ (e.g. C^u സ്), there is no vowel after the consonant (as in the case of അവൻ /awan/ ʻhe-nominative', രാജൻ /raajan/ 'rajan-nominative' (just a name)) [1]

---

1 Another instance of the contrast between ₍uC^u and C^u is in the distinction between what I have called sub-componds and co-compounds in my book *Lexical Phonology*. A sub-compound has two stems the second of which is the head, while a compound can have more than two stems, all of them being heads. Thus, മരക്കുതിര /marakkutira/ ʻwooden horse' is a sub-compound, while ആനകുതിരമയിലൊട്ടകം /aanakutiramayiloTTkam/ ʻelephant, horse, peacock and camel' (from ആന /aana/ elephant', കുതിര /kutira/ ʻhorse', മയിൽ /mayil/ ʻ peacock' ഒട്ടകം /oTTakam/ ʻcamel') is a co-compound. In a sub-compound, final sonorant consonants (ൽ /l/, ൾ /L/, ൺ /N/, ൻ /n/, ർ /R/ etc.) of the first stem combine with the initial consonant of the second stem to form a single consonant sequence, but in a co-compound there is the option of breaking up the sequence with a schwa. Thus, there cannot be a schwa in the sub-compound വാൾപ്പരിചകൾ /waaLpparicakaL∂/ ʻsword-shields' from വാൾ / waaL/ ʻsword' and പരിച /parica/ ʻshield', but the corresponding co-compound can have a schwa വാള് പരിചകൾ /waaL@paricakaL∂/ ʻswords and shields'.

Except word final ന /na/, however, the choice between consonant alone (represented by C^u സ്) and consonant plus schwa (represented by ₙC^u സ്) is a matter of the style of speech: the use of schwa is more common in informal or colloquial styles than in formal styles.

I must add that the issue here is the use of chandrakkala vs. chillu, not which consonant should be used for the representation of the final consonant in അവർ /awaR/. Morphophonemically, this consonant is derived from ര /ra/, as shown by the appearance of /ra/ in അവരല്ല /awaralla/ 'not them' (in contrast to കാർ /kaaR/, derived from റ /R/ as shown by കാറല്ല /kaaRalla/). But since Malalayam orthography is phonemic, not morphophonemic, the appropriate consonant would be the surface one, namely റ /R/. But there might be practical considerations for the use of ര /r/ instead of റ /R/. What is important is the consistent use of the same base consonant character for ർ /R/ instead of using two different characters depending upon their morphophonemic source.

4 Practical matters

Given the points made above, a number of practical questions arise in the representation of consonant sounds unaccompanied by vowel sounds. The very first question is: do we need separate chillaksharams in a revised system of Malayalam orthography?

My answer is no. The final consonant of അവൻ /awan/ 'he' can be represented with first symbol of നായ /naaya/ 'dog', with a chandrakkala; the final consonant of അവൾ /awaL/ 'she' can be represented with the second consonant of വള /waLa/ 'bangle', with a chandrakkala, and the final consonant of അവർ /awaR/ 'they' can be represented with the first symbol of റാണി /RaaNi/ 'queen', with a chandrakkala.[2]

Will the use of chandrakkala to replace chillus create complications in the representation of forms like വൻ യവനിക /wan yawanika/ (represented traditionally with a chillu) and വന്യ വനിക /wanya wanika/ 'represented with a consonant with a semivowel diacritic? My answer coincides with that of Rajeev Sebastian: the distinction between the two forms is not in terms of the sounds themselves, but their syllabification: /വൻ wan – യ ya ... / vs. /വ wa – ന്യ nya ... /. If we wish to express this rare distinction in orthography, we may do so by using a hyphen (as Rajeev suggests), or a space between the relevant symbols. (The same remark applies to similar cases with sequences like /...സ്വ swa.../ vs. /...സ് s – വ wa.../, /...ഗ്വ gwa.../ vs. /...ഗ് g – വ wa.../ etc.)

Will the chandrakkala solution be affected by the so called 'multivalency" of chillus? Is the final consonant of അവർ /awaR/ 'they' for instance, different from the final consonant of കാർ /kaaR/ 'car'? The answer is that the two consonants are different morphophonemically (as shown by the ര /ra/ in /awaralla/ 'not they' vs. റ /Ra/ in /kaaRalla/ 'not car') but this distinction in the source is irrelevant for Malalayalam, since, as pointed out above, Malayalam orthography is phonemic. Similar remarks apply to the other alleged sources of ൽ /l/ and ൾ /L/.

Finally, how important is it to have distinct orthographic representations of consonant alone (represented by സ്) and consonant plus schwa (represented by സ്) ? If such contrasts are frequent, it would make sense to maintain the distinction. If not, nothing serious is going to be lost if we use the same orthographic representation both. After all,

---

2   I must add that the issue here is the use of chandrakkala vs. chillu, not which consonant should be used for the representation of the final consonant in അവർ /awaR/. Since Malayalam orthography is phonemic, the appropriate consonant would be the surface one, namely റ്/R/. But there might be practical considerations for the use of ര്/r/ instead of റ്/R/.

the distinction between the dental nasal and the alveolar nasal (as in the first and second consonants in നാന /na na/ is not represented in Malayalam orthography, and no one seems to be unduly worried by the potential loss of contrast.

The views expressed above are from the stand point of a theoretical approach to the relation between phonology and orthography. From this perspective, the three ways of representing ന്മ /nma/, one with a chillu ൻമ, one with a character with a chandrakkala ന്മ, and one with a conjunct letter ന്മ, are equivalent ways of doing the same thing, and hence at least two of them are redundant. From the point of view of engineering application, however, it would make eminent sense to treat all the three as equivalent but provide for the use of all the three with a minimal cost. Likewise, the phonologist in me is not convinced that there is a need to have anuswaaram in addition to a മ /ma/ with a chandrakkala, but for engineering purposes providing this extra option wouldn't hurt if its cost is not high[3].

5 Conclusion

Given the above considerations, it seems to me that the arguments for explicit representation for chillus in IDNA are probably based on a lack of understanding of the relation between phonology and orthography, and the nature of Malayalam orthography.

When committees assert their opinions without justification, and often without adequate information on the linguistic and technical aspects, it is important that their assertions be not taken at their face value, without questioning and without demanding justification.

I understand that the Unicode community includes the UTC, mailing list and other forums. If what a committee or some person makes unilateral decisions about Malayalam, and these decisions are accepted without examining the relevant considerations, then an eminent body as the UTC, the mailing list, and other forums become in effect redundant, which would be a pity.

I would therefore make an appeal to go through the relevant evidence and argumentation carefully, and select the most appropriate solution to represent Malayalam for use in computing applications.

---

3   Notice that the anuswaram of the final consonant of /maram/ 'tree' can also be represented with the same symbol as the first one, with a chandrakkala. However, the treatment of anuswaram is not what is at stake here. The parallel treatment of chillus and anuswarams may be justified from a purely theoretical approach to avoid redundancy, but a more prudent practical approach perhaps would be to retain anuswaram as it is entrenched in both the traditional and the modern systems. Replacing it with /m/+chandrakkala may not be 'cost effective' in terms of the disturbance it causes.