Tamil Script Reform and Glyph Rendering Approach in Unicode: Past and Present Attempts to Simplify Tamil Writing System

Vasu Renganathan
Harold F. Schiffman
University of Pennsylvania

> "… Until recently, the received opinions on these issues, in the west at least, have mainly been based on or at least strongly affected by their explication by Georg Bühler fully one century ago in his highly influential, if somewhat controversial monograph On the Origin of the Indian Brahma Alphabet (Indian Studies No.III) . Bühler argued for an early origin of writing in India and posited an extensive pre-history, going as far back as the 8th century BC, for the Brâhmî script, which he derived from the Phoenician script. Although more recent writers such as David Diringer have tended to doubt such an early date for Brâhmî and have looked to the Aramaic rather than the Phoenician script as its probable source, Bühler's materials and arguments have continued to guide the discussion long after many of them have become outdated." (Solomon, 1995: 271).

Tamil Script – Past

The designation for the most ancient script of Tamil remains controversial. However, a general consensus is that the Northern Brāhmī had a Southern version from which the Tamil script is evolved after a series of intermediate stages from 2nd Century B.C. to 18th Century A.D.[1]. The pictorial script form that became the main writing system in Harappa are compared to that of the graffiti on the megalithic pottery in South India, datable in the first millennium B.C. This attempt is particularly made by those who argue in favor of a strong connection between the Harappan and Dravidian culture. However, the pictographic writing system resembles much different from the much later syllabic writing system that is represented by the Brāhmī scripts, despite the fact that the latter is considered to be a derivative of the former[2]. The Southern version of Brāhmī script, invariably called Tamil-Brāhmī, Dravidi and Damili[3], is dated in the 3rd B.C. mostly based on the inscriptions survived in several distinct forms[4]. Substantiating these facts, a

[1] See Mahadevan (1970: 1), Zvelebil (1970: 11-19) and Velu Pillai(1980: 6).
[2] Langdon (1931: 433) assumes, "perhaps there were some intermediate links between the Indus script and the Brahmi, but the variations are solely on the side of the Brahmi, the Indus script remained almost constant through the centuries".
[3] Nagaswamy (1972) notes that the term Damili occurs in Samavayanga Sutta, a Prakrit Jaina work. He is of the opinion that Damili is an independent script that was used among the eighteen enumerated scripts during the pre-Christian era.
[4] Mahadevan (1970:3) presumes based on old Jaina tradition that Bhadrabahu migrated to Sravana Belgola in Mysore in the time of Chandragupta Maurya, and his disciples under the leadership of Visakhamuni came further south to preach the faith some time around the beginning of the Third Century B.C. He

standardized literary variety of the Tamil language is seemed to have gained currency in its well-developed form during this period. This is observed mainly from the poems of the Sangam collections: the kuṟuntokai, aiṅkuṟunūṟu, puṟanānūṟu, patiṟṟuppattu and in the grammar of tolkāppiyam, which are dated between the 2nd B.C. and 3rd Century A.D[5]. This fact may be taken to mean that Tamil must have adopted a well refined writing system during this time period, as is evident from the fact that the grammar of Tolkappiyam defines the word eṟuttu to be consisting of thirty letters in a sequence from *a* to *n* along with the three other secondary forms.

எழுத்தெனப்படுப
eṟuttenappaṭupa
அகரமுதல் னகர இறுவாய் முப்பஃதென்ப
akaramutal nkara iṟuvāy muppaxtenpa
சார்ந்துவரன் மரபின் மூன்றலங் கடையே
cārntuvaran marapin mūṉṟalaṅ kaṭaiyē

The Tamil word eṟuttu is a noun derived from the verb eṟutu meaning 'write', which designates the fact that there existed a writing system that represented the sounds as close as to the present. Further, Tolkappiyam defines the formation of syllables and the role of the dot on top of consonants in the following verse.

புள்ளி யில்லா எல்லா மெய்யும்
puḷḷi yillā ellā meyyum
உருவுரு வாகி அகரமொடு உயிர்த்தலும்
uruvuru vāki akaramoṭu uyirttalum
ஏனை உயிரொடு உருவுதிரிந்து உயிர்த்தலும்
ēnai uyiroṭu uruvutirintu uyirttalum
ஆயீ ரியல உயிர்த்த லாறே (17).
āyī riyala uyirtta lāṟē

According to Tolkappiyar, when a consonant occurs without a dot on the top, it is usually pronounced with the vowel sound 'a', but when it occurs with other vowels from the sound "ā" on, the shape of the consonant is modified to refer to the corresponding syllable sound. The other purpose of dot was to distinguish between short and long e and o. A dot placed above these two letters were intended to indicate the short e and o. This is illustrated in Tolkappiyar's sutra as follows: எகர ஒகரத் தியற்கையு மற்றே ekara okarat tiṟkaya maṟṟē (16). Mahadevan (1970: 5), who bases his opinion about the then Tamil scripts based on the Arikamedu inscriptions found near Pondichery dated between 1st A.D. and 2nd A.D.[6], suggests that the dot was introduced in Tamil script mainly to mark pure consonants as well as to distinguish between short *e* and *o*.

---

further points out that the rock-shelters with Jaina Buddhist associations contain the earliest specimens of the Tamil-Brahmi script.

[5] See Zvelebil (1970: 15).

[6] See Mahadevan (1966: 56-73) for samples of a number of inscriptions excavated during this period including the Arikamedu inscriptions.

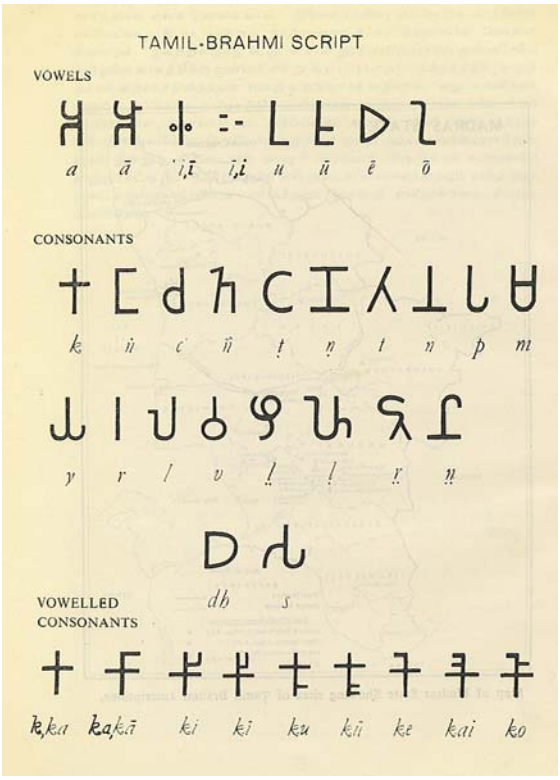


Fig. 1. Tamil- Brāhmī Alphabet reconstructed based on various inscriptions found in Tamil Nadu (Mahadevan 1968: 56)

Mahadevan (1966: 58), who presents an alphabet system of Tamil- Brāhmī based on a corpus of inscriptions excavated in Tamil Nadu in various periods (See Fig. 1), notes that the Tamil- Brāhmī script used during the first Millennium B.C. does not use dot, nor does it distinguish between short and long e and o by way of distinct symbols.

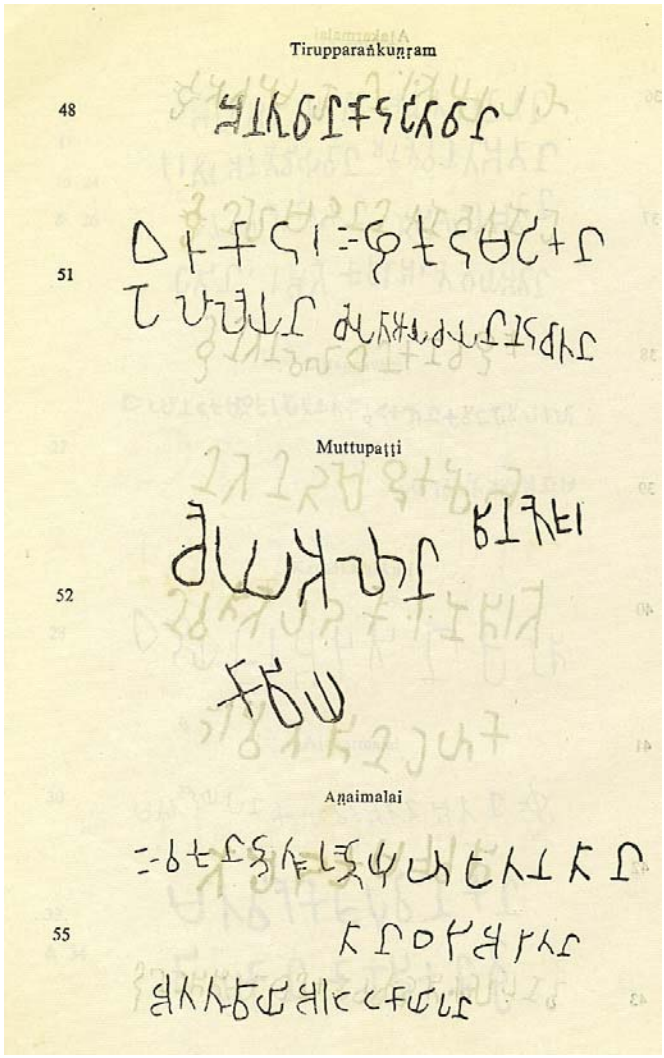


Fig. 2. Sample inscriptions found in pot-shreds and graffiti in Tamil Nadu (Mahadevan, 1968:)

Even though both short and long *e* and *o* sounds were recognized during the time of Tolkappiyar, the short forms must have been represented either with a dot or a line above. Absence of separate letters for short *e* and *o* was common in the Tamil writing system even until the time of Beschi (1680-1747), who first made a change in the system by introducing a separate symbol for long *e* with a slanting stroke added to the short *e* ெ as ே. Similarly, to the short ொ he added a loop at the bottom to make it the long letter ோ. The corresponding secondary symbols for these two vowels were introduced by him as ◌ெ and ◌ொ respectively. This change is in effect until the present day Tamil script. The older version of Tamil writing system prior to Beschi's alternations may be attested

in the Tamil Bible book printed in Colombo in the year 1741 by the Dutch East India Company[7] (see figs. 3 and 4).

Fig. 3. Title page of the first book (Tamil Bible) printed in Colombo in the year 1741. (Balasingham, 1968).

[7] Even though the first book in Tamil was printed in 1671, according to Balasingam (1968: 528), the Beschi's changes were not popular in Ceylon during 1741 when the first Tamil book was printed in Colombo.

Fig. 4. Sample page from the first printed book in Colombo in the year 1741 (Balasingham, 1968).

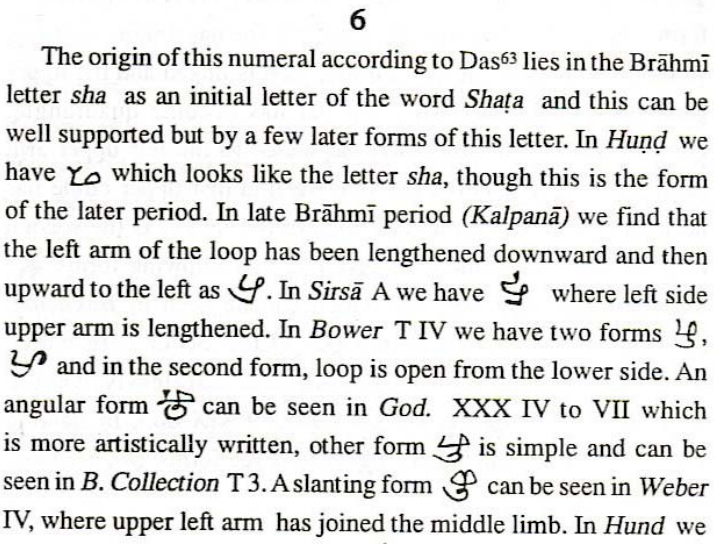
The major difference between the letters used during this time and the present day Tamil script may be summarized as follows. Long *e* and *o* are made with a line above the letter. Thus, the word **எலியுது** is to be read as **ஏலியுது**; **ஒவெத்** is to be read as **ஓவெத்.** The syllabic long e and o are distinguished in a similar manner. Thus, **சுவிசெஷும்** must be read as **சுவிசேஷும்**; **சகொதரர்** is to be read as **சகோதரர்.** Further, as the letter r and long syllabic symbol for a are marked with the letter **ா,** a line above the syllabic letter is used to distinguish it from the letter r as in **குமான்.** Trill r and flap r are marked with the symbols **ற** and **ா** respectively, as opposed to the present forms **ற** and **ர.**

The later development of Tamil- Brāhmī script was termed Vaṭṭeḻuttu (circular script) and it is distinct in many respects from the Brāhmī script in that it introduced circular shapes that look very much similar to the present day Tamil script. This scheme of writing was commonly used from the beginning of the Pallava rule in Tamil country which dates to C. 600 A.D. Degeneration of Vaṭṭeḻuttu led to the present form of Tamil

script, which was adopted during the post Pallava period and the early Chola period, which lasted between 9th Century A.D. and 12th Century A.D. According to Nagaswamy (1968: 414), a greater number of Vaṭṭeḻuttu inscriptions of the seventh century A.D. are found in the northern Kongu country and Tondaimandalam. The Grantha script was developed along these lines to write formerly the Prakrit language and later the Sanskrit language.

Evolution of the letter ழ in Tamil

The alveolar fricative sound represented by the letter ழ has been indigenous to the Tamil language, and the origin of this letter may shed some light on the history of Tamil script and its connection with Brāhmī script. This letter seemed to have occurred in Brāhmī both in the Southern version (see fig. 1) as well as in Northern version (fig. 5.). Mahadevan (1968) in his reconstruction of Tamil-Brāhmī alphabet includes the letter ழ as a lateral sound ḻ. Strikingly, however, Sharma (2002: 260), who analyzes the development of Brāhmī script in North-Western India shows a letter very similar to ழ as an equivalent of the palatal retroflex ṣ (see fig. 5). If the Tamil letter ழ is indeed same as the one identified in Brāhmī script and not a case of accidental similarity, the connection between Brāhmī and Tamil script can further be substantiated.

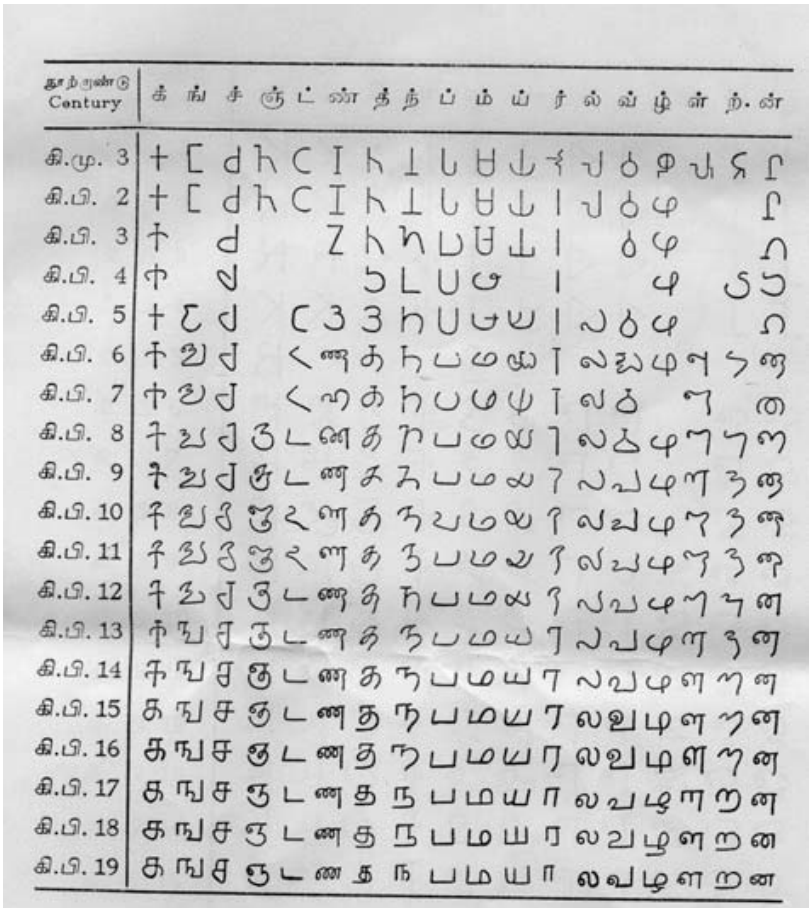


6

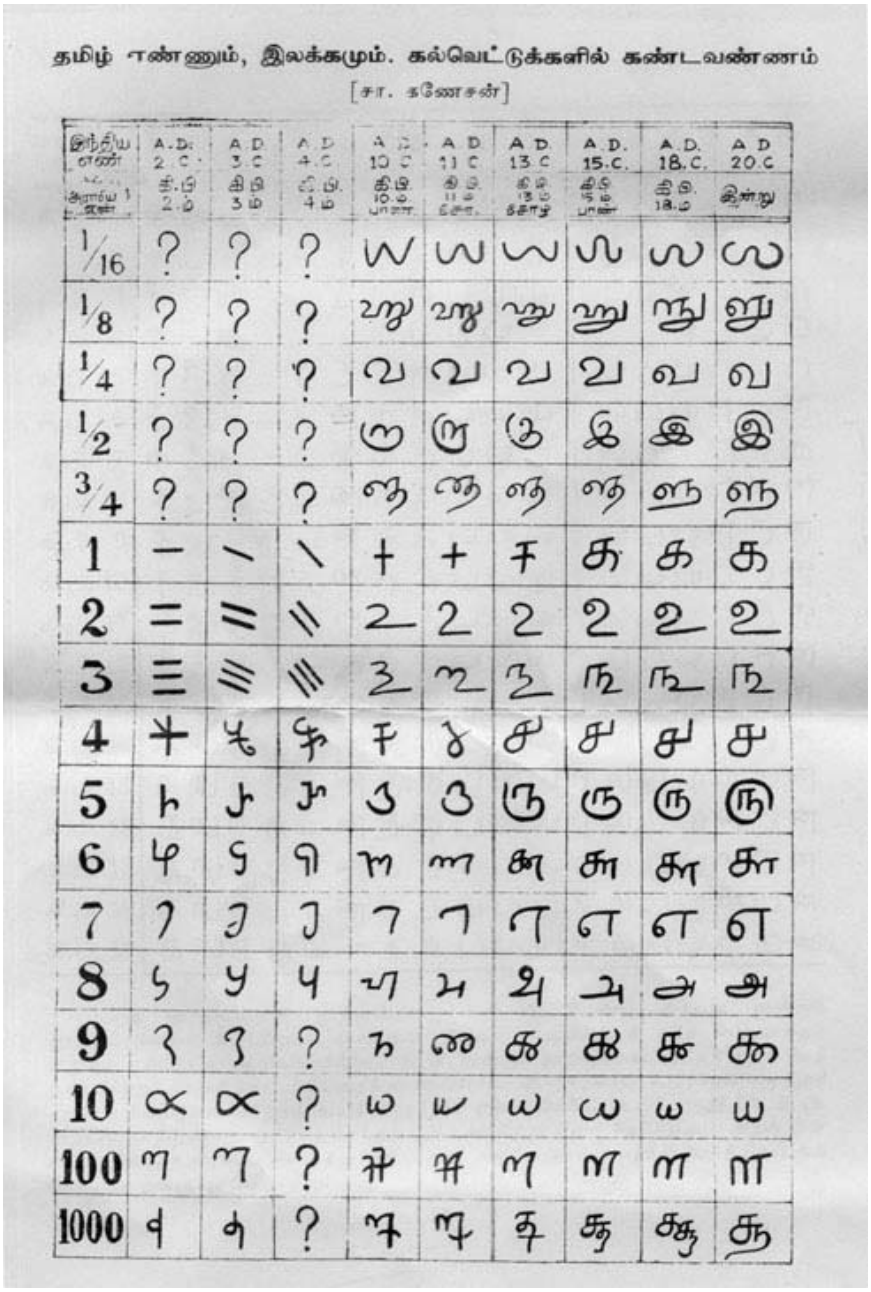
The origin of this numeral according to Das[63] lies in the Brāhmī letter *sha* as an initial letter of the word *Shaṭa* and this can be well supported but by a few later forms of this letter. In *Huṇḍ* we have ५० which looks like the letter *sha*, though this is the form of the later period. In late Brāhmī period *(Kalpanā)* we find that the left arm of the loop has been lengthened downward and then upward to the left as ५. In *Sirsā* A we have ५ where left side upper arm is lengthened. In *Bower* T IV we have two forms ५, ५ and in the second form, loop is open from the lower side. An angular form ५ can be seen in *God.* XXX IV to VII which is more artistically written, other form ५ is simple and can be seen in *B. Collection* T 3. A slanting form ५ can be seen in *Weber* IV, where upper left arm has joined the middle limb. In *Hund* we

Fig. 5. Use of the letter ழ (ṛ) in North-Western Brāhmī (Sharma, 2002: 260)

நூற்றாண்டு Century	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 3											
கி. பி. 2											
கி. பி. 3											
கி. பி. 4											
கி. பி. 5											
கி. பி. 6											
கி. பி. 7											
கி. பி. 8											
கி. பி. 9											
கி. பி. 10											
கி. பி. 11											
கி. பி. 12											
கி. பி. 13											
கி. பி. 14											
கி. பி. 15											
கி. பி. 16											
கி. பி. 17											
கி. பி. 18											
கி. பி. 19											



கி. மு. மூன்றும் நூற்றாண்டுக்கும் கி. பி. இரண்டு அல்லது மூன்றும் நூற்றாண்டுக்கும் இடைப்பட்ட காலத்தில் வரிவடிவில் குறிப்பிடத் தக்க பெருமாற்றம் எதுவும் ஏற்பட்டதாகத் தெரியவில்லை. கி. பி. மூன்றும் நூற்றாண்டை ஒட்டிக் குறிப்பிடத் தக்க மாற்றம் பெறத் தொடங்கியதைக் காணுகிறோம். ஆனால், மெய், உயிர்மெய் எழுத்துக்களில் ஏற்பட்டுள்ள மாற்றம் உயிர் எழுத்தில் ஏற்படவில்லை. உயிர் எழுத்தில் கி. பி. ஐந்தாம் நூற்றாண்டை ஒட்டியே மாற்றம் ஏற்படலாயிற்று எனத் தெரிகிறது. அதன் பின் உயிர், மெய், உயிர்மெய் யெழுத்துக்கள் யாவும் படிப்படியாய் வளர்ச்சி பெறத் தொடங்கியிருப்பது நன்கு புலப்படுகிறது.



Tamil Script – Present

ASCII uses 7 bits to represent characters. Some 7 bit systems do not clear the 8th bit, causing some characters to be displayed incorrectly. If some characters are not being displayed correctly, try enabling the **Strip 8th bit** option.

Input methods

In order to create a word processing document containing Unicode IPA, it is necessary to have an input method for Unicode IPA characters, in addition to a Unicode

IPA font and a word processing application that supports Unicode. Neither Windows 2000 nor Mac OS X includes an input method specifically designed for Unicode IPA, so other options must be explored.

References:

Asher, R. E. (1968). (ed.) *Proceedings of the Second International Confernces Seminar of Tamil Studies.* International Association of Tamil Research: Madras.

Balasingham, C. (1968). "Some Unusual Characters used in Tamil Printing in Ceylon During Dutch Times" (pp. 528-30), in R. E. Asher (ed.).

Langdon, S. (1931). "The Indus Script", in J. Marshall (ed.), *Mohenjodaro and the Indus Civilization*, London.

Mahadevan, Iravatham. (1970). *Tamil – Brahmi Inscriptions.* The State Department of Archaeology: Madras.

Nagaswamy, R. (1968). "The Origin and Evolution of the Tamil Vatteluttu and Granta Scripts", in R. E. Asher (ed.).

Solomon, Richard. (1995). "On The Origin Of The Early Indian Scripts: A Review Article of Der Beginn der Schrift und frühe Schriftlichkeit (1993) in Indien". Journal of the American Oriental Society:115.2 (1995), 271-279

By Oscar von Hinüber. Akademie der Wissenschaften und der Literatur, Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse, Jahrgang 1989, Nr.11. Mainz: Akademie Der Wissenschaften Und Der Literatur / Stuttgart: Franz Steiner Verlag Wiesbaden, 1990. 75 pp; and Schrift im alten Indien: Ein Forschungsbericht mit Anmerkungen. By Harry Falk. ScriptOralia 56. Tübingen: Gunter Narr Verlag, 1993. 355 pp. DM 136.

Sharma, Ram. (2002). *Brāhmī Script: Development in North-Western India and Central Asia.* B.R. Publishing Corporation: New Delhi.

Velu Pillai, A. (1980). *Epigraphical Evidences for Tamil Studies.* International Institute of Tamil Studies: Madras.

Zvelebil, Kamil. (1970). *Introduction to the Historical Grammar of the Tamil Language.* Academia, Prague.