47

Rationalisation in Tamiz ezuttu niral: An Agenda to Minimise Number of Glyphs Used in Tamiz Syllabary

N D Logasundaram

Chennai

The richness found in a language is the index of its civilization as it is an inseparable part of a culture of a land. Thousands of languages are in use in world today and majority of them do not have script. Hence most of them are at the verge of extinction. There are languages found in the history which were once very popular in some region for hundreds of years, and after reaching the apex lost their flavour and identity and gone into oblivion. Not only the history, the whole spectrum of human knowledge, its acceleration cum explosion was possible only out its accumulation with this tool the 'script'.

Tamiz is one of the progressive and classical languages of the world, which is in continuous and useful service for few thousand years, on the eastern meridians of this planet, demonstrating its richness and strength.

Without a script a mere spoken form of a language can quickly get corrupted and distorted within few transfers in the interfaces as mouth-ear-mouth. Therefore script is a must for maintaining uniformity in a spoken language. Script will fix the language to a standardized use and prevent it going astray without bounds by unavoidable phonetic variants creeping from not-so-committed users or from unwarranted fancy mongers.

We know well that there are scripts contrived even for reading by finger tips for blind people while non script cum non phonetic body language mode of communication were also devised and sported through out the globe from time immemorial. Even today it is very much alive and used not only by hearing-impaired people, but also on the turf of sports playgrounds to tarmac handling of supersonic aircraft.

Primitive humans painted simple graffiti and later devised and scribbled imprints out of necessity of passing on their feelings, thoughts or messages to others in their group separated by time, distance or both. A created record also serves as an aid for review / alteration by the author or by others. By nature it has to go always with a 3D medium.

In old ages they were static as rock panel glyphs, temple wall or nadukal stone inscriptions or portable as palm leaves, animal skins, clay tablets, metal plates etc. Paper, a modern avatar of palm leaves, facilitates easier multiplication and distribution. Scientific and industrial advancements found many fast gadgets as printing, telegraph, television, telephotocopying (fax), etc. and all of them were fully exploited by every one, every day.

The new and omnipotent medium cum tool covering the whole planet with lightning speed, the Computer and Internet, is in reign today. Realising that the communication is vital for growth and advancement, wise people in a rational society never lose great opportunities. Tamiz the language spoken by several hundred millions of population, always had wisdom flexibility and tenacity in the past and continue to strive to hold the unique position as one of the dynamic languages of the world.

Any language of a land is always subjected to inevitable changes springs from geographical and political variations, cultural mutations, fusion and diffusions out of war, trade, migration, scientific and industrial advancements etc.

The evolution of Tamiz script, over thousands of years, are well studied by epigraphists with age old stone inscriptions, copper plate grants, coins, palm leaf editions of sacred hymns and classic literature and found it to be natural, continuous and need based.

Tamiz script being written from left to right in a horizontal line formation while there are right to left and top to bottom progressing formats are in use in other languages of the world. Rare ancient scripts even reported to be written from bottom to top and even in opposing directions in alternate lines.

Tamiz is based on words framed out in a set of varying permutation combination of phonemes/letters from its alphabet and train of such words make a sentence to convey a message. Languages in this mode dominates the world today. But Chinese or Korean language uses indivisible pictures or complex glyph to represent whole word or a sense, just as icons in the computer screen. We can say, the common characters as $,%,@ etc. on our keyboard are akin to it. If you look closer into the lines *'eNN en2pa En2ai ezutten2pa ivviraNTum kaNNen2pa vAzum uyirkku' & 'eNNum ezuttum kaNNen2attakum', the word 'eNN' mean* not only mathematical figures, (which is actually a pictograph) but also, represent an *'eNNam'* an idea or a sense conveyed monolithically which is different from assemblage of phonemes as in *ezuttu*. The two different forms of the writing system, used being mentioned here as a set

History have recorded instances of institutionally inducted minor changes to Tamiz script form with rationale. For example few centuries ago oRRaiccuzik kompu - ி- the present modifier for kuril ekara uyirmey was then intended for nedil Ekaaram, an uyirmey of two nodi duration. For ekara uyirmey (kuRil) the same oRRaiccuzik kompu was written but with a puLLi - ி்- on top of it denoting shortened articulation time. In this case pulli denotes one noTi instead of two. The irraTTaiccuzik kompu - ெ- was devised anew for Ekaara uyirmey which is a nedil of two nodi.

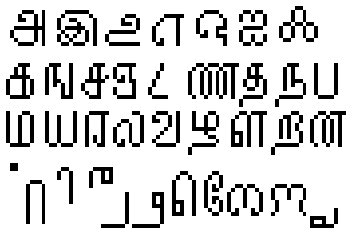
During third quarter of last century the aakaara uyirmey Naa, Raa, n2aa - ணாநா- and aikaara uyirmey Nai, lai, Lai, n2ai - ணைலாை- are redesigned and given the streamlined form as on today i.e. they were assigned with same modifier as that of other letters in their set eliminating different type of conjunctions.

But, till today ukara-uukaara uyirmey letters were not touched, in waiting, for much longer period, probably to comply with the emerging requirements of the modern medium, the IT, in one stroke, for a similar reform, i.e. use of a common modifier similar to akara, aakaara, ikara, iikaara, ekara, eekaara, okara, ookaara, aikaara, aukaara uyirmey letters. There are several instances of such an excise mooted out, even by influential educationalists, Tamiz Enthusiasts, but with little success, may be, due to high dose of radicalness.

Hence a very simple scheme easy to register even in the mind of an average person, with least psychological inertia, by reshaping ukara-uukaara uyirmey letters (18 + 18) along with other

minor plans for uyir kuRil to uyir nedil morphosis (5)and Takara ikram and iikaaram (2) is presented here as 'An Agenda to Reduce Number of Glyphs used in Tamil Syllabary'.

By this proposed blueprint the number of characters in the suite will drastically reduce to mere 36, which is even less than what is in English (52). Therefore unscaled full sized version of a Tamiz keyboard can now be on the miniaturised keypads of handhelds as organisers and palmtops.



The reduced number of characters when implemented will bring out changes to Tamiz language especially in the digital world towards an auspiciously liberated ambience. Here are the few.

* By virtue of lesser number of characters it will be easier to learn reading and writing.

* Learning methods, tools and time needed to learn will become easier, convenient and shorter.

* Foreign language speakers can even read and write standard Tamiz quickly in crash course time.

* In Indian languages, Tamiz is already having fewer characters, without any ligatures, can now with this reform will have still lesser characters[1].

The pace of progress made in the cyber world both in hardware and software points that computer is going to be the multipurpose ubiquitous tool in any human activity. Therefore reduction in no. of characters to be dealt in Tamiz digital world will pave way for:

* Fixing all Tamiz letters[2] even into the miniaturised keypads of handhelds.

* Accommodating spilled over special characters, numerals and punctuation marks into Tamiz keyboard.

* Easier and simpler software programming for (A) OCR devices (B) voice input / output applications / utilities (C) language translations transliterations and transcription programs (D) applications for handicap ped people in speech, sight and hearing[3] impaired (E) and the bunch of interlinking utilities between all the above.

What are the proposals ?

1. ukara uukaara uyirmey

For the ukara and uukaara uyirmey letters, a simple and new common ottezuttu (modifier)one for each set of 18 consonants is suggested replacing separate individual glyphs in the traditional syllabary in rhythm with other uyirmey transformations.

1. ukara uyirmey.


[4]

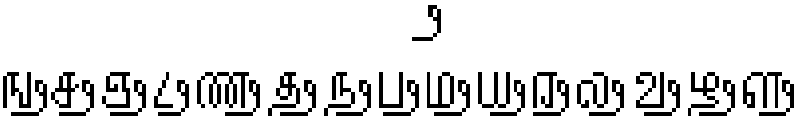
The last set has more number of letters (9) than others and hence a modifier (oTTezuttu) design was taken out from them to have more familiarness.

It is approximately a laterally inverted 'Ta' which has to be positioned after a akara consonant slightly lower to the horizontal rule of writing without touching it simulating the present ukaram.

The new rationalised shape differ not much from old one in Nu, tu, nu, lu, zu, Lu, Ru, n2u. In others though it is apparent, it will take less than a minute to learn even by an individual of average IQ.

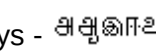
2. uukaara uyirmey

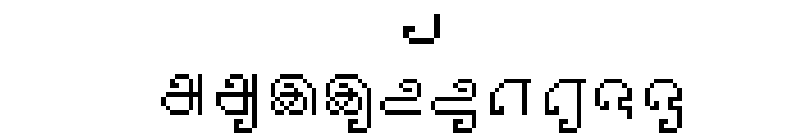
For uukaara uyirmey a modifier similar to the above but with a soonya koolaccuzi added at its upper end is proposed.



For this uukaara uyirmey several shapes as -  - were considered before deciding a suitable shape and finally the one as above was zeroed-in being a better choice. The first one was giving an illusion as aakaarakkaal as bottom horizontal line almost fusing with parent character. Second one is having too many curves and longer character length. Third was similar to the chosen one, but the vertical line in it, merges with parent consonant or if given with sufficient gap the character length is increasing. All these three shapes are devised for rhythm in design with new ukara modifier seen above. The last one was a shape already being used with vadamozi grantha letters found mixed in Tamiz fonts. But its counterpart in ukaram will be creating illusion of iikaara talaikkiiRRu. Hence the set was not considered.

3. uyir neTil

uyirkkuril letters changing into nedil in 5 different ways -  - By giving a common oTTezuttu (modifier) to all the 5 kuRil uyir, it is possible to modify them to neTil uyir as in uyirmey letter's transformation. The modifier which is already in aakaaram is sugested for use with other 4 uyir for continuity and to avoid strangeness. Use of common modifier for kuRil to neTil uyir is a new plan and this can reduce 4 characters in the suite.



Though variations are notable in iikaaram and uukaaram it has thread to follow sans a complete change. Hence it will not induce psychological resistance to switch.

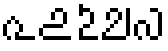
4. Takara ikaram and iikaaram

The Takaram can start with a slanted line instead of perfect vertical line -  - so that normal ikara Ikaara oTTezuttu (talakkIRRu)what is given to other 17 consonants can be used here touching at top right side starting point. This will dispense two special glyphs of takaram -  - in the suite. Though new shape will look odd initially the strangeness will disappear after little use.

These 4 proposals will bring uniformity and cadence into the complete set of letters. Each uyirmey lettersets in the uyir / mey matrix will have only one type of oTTezuttu and no where the formula is changed while the uyir will have a streamlined appearance. Visualise the easy to grasp, facile syllabary. How much simple and attractive the Tamiz ezuttu niral (see Table 1) will be for a new learner whether he/she be a child from tamil speaking land or users of foreign tongue.

Minor touchups, trimmings and cosmetic surgery in the shape of some characters to improve their distinctness both for human and machine reading.

The following are just suggestions which are not connected with reformation. But are proposal to improve the lucidity and zero-fault reading and programming of the characters which are nearly similar when written in hand or in small size font.

A. - 

The ukaram ('uyir')can be written with more of a left protruding belly i.e.the top part with a good recess towards right side. This will make it distinct from the numeral '2' (or the old Tamiz character once used for 'date' or piLLaiarcchuzi which is normally written with starting stroke on horizontal line)

The numeral '2' shall be given with a small horn like upward stroke to denote it as a numeric character and not the uyir 'u'.

The vakaram ('va') can start at a top level instead of from a position near horizontal line and let the lakaram ('la') to start at horizontal line and end at the top with a horizontal stroke towards left side. These will make one look much different from the other.

B. - 

The oRRaiccuzik kompu can start with a smaller cuzi and let it have usual semi circular top lingam while ekaram can start with bigger cuzi and stroked with minor flatness at the top right along with a forward directing horny projection in a normal way.

Though in 'nja' the top part is similar to ekaram because it is written in a smaller scale, up to half the height of a character and finished with the semi circular dish above the rule, it shall not give any illusion.

C. - 

The aaytam (/ae/ - erroneously taught and pronounced as 'ak') is normally written as three separated soonyaccuzi in a conical formation. In many fonts these felt as some special punctuation characters.(An instance in classification of a character by experts group as "Unicode" initially classified as puntuation mark and then reclassified as character-others) In some fonts it is not even noticeable as a character inside a word. Hence all three soonyaccuzi can be written in one continuous curved line maintaining its basic shape of triangle (a three dot koolam) This will make it a respectable language character. In hand written mode continuity is the easiest and quickest way to finish the shape.

D. -  The kakaram can be finished with the stroke touching well the main body or even if it penetrates into it and get closed at the start point (nearly as a kolam) it will be more purposeful in bringing difference between 'ta'.

Similar to the Tamiz numeral 4 the 'ca' can have an upward stroke at the finish in the horizontal line at half the height of the character. This will make it vary much different from kakaram.

E. - 

In the twin column vertical lines of the Rakaram, let the left side shall be slightly shorter while the right side a bit taller. This will make it much different from nakaram. Incorporating a small inside depression on the third vertical line it will enhance the difference further dramatically. This idea was already in use in several fonts and others can follow.

F. - 

Because the top strokes of zakaram is identical to makaram let it only up to half the character height in a smaller scale. The bottom half is a double folded tail like stroke ending with a downward appendix. The horizontal portion of it shall be coinciding with the horizontal rule similar to what is in ta, na, Ra.

G. - 

The Lakaram shall have a cuzi in the first column start (bottom) and it must be absent in first column of Nakaram and n2akarm. The second column of Lakaram must not have the cuzi. Nakaram and n2akaram 's middle columns will have the cuzi as in the present practice.

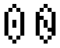
The aikaara oTTezuttu must not be given with any cuzi at all and it must be written with plain convexly curved lines while the whole character can be written with shorter character height. As the vadamozi 'sa' will have a extra vertically ending stroke and to a normal character height it will be distinct from the aikaara oTTezyttu.

H. - 

The Lakaram must be left solely as a uyirmey character. Identical glyph must not be used as modifier in aukaaram. Instead use a reshaped aukaara oTTezuttu as - ◌ൗ- and maintain the typical Tamiz characters one to one relation between phonetic and written forms. In the proposed rationlaised uukaaram the character La as a modifier is avoided to maintain virtue. Aukaara glyph same as Lakaram is provided with a code point 0170 in TSCII and the glyph of Tamiz numeral 'hundred' (code point which is now looks similar to this Lakaram can be with a horn like upward stroke as what is sugested for other numerals. (described below)

I. - 

In Japan many people write the numeral 5 with a horn like upward stroke in the top left corner. This makes it look more different from the English letter 'S'. This innovation will serve good purpose when alpha numeric codes such as part no. of commercial products intermixed with S and 5. As we use bilingual fonts (tsc) where English is already there along with international numerals, it is useful. We can have this feature incorporated in Tamiz numerals too because Tamiz numerals are similar to the alphabetic characters.

J. -  - The numeral zero (0) looks similar to the English alphabet 'O'. To distinguish one from the other, the zero is written with a dot in the middle or a slanted line stroke is given inside the zero to distinguish one from the other as found in many fonts. This feature or similar device in the numeral zero can be introduced.

These minor configurations are for enhancing both human as well as mechanised reading. We are aware of the specifically devised numerals in MCR cheques. Forth coming Tamiz font designers can become pioneers in delivering OCR-friendly fonts with infused features for easy, error free machine reading which is going to relieve large chunk of human activity in the near future. I hereby request all the institutions and individuals connected with Tamil Language, to consider these Rationalising proposals, recommand and act for an early adaptation.

Table 1

்	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஒ	ஓ	ஐ	ஔ	ஃ
க்	க	கா	கி	கீ	கு	கூ	கெ	கே	கொ	கோ	கை	கௌ	கஃ
ங்	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙொ	ஙோ	ஙை	ஙௌ	ஙஃ
ச்	ச	சா	சி	சீ	சு	சூ	செ	சே	சொ	சோ	சை	சௌ	சஃ
ஞ்	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞொ	ஞோ	ஞை	ஞௌ	ஞஃ
ட்	ட	டா	டி	டீ	டு	டூ	டெ	டே	டொ	டோ	டை	டௌ	டஃ
ண்	ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணொ	ணோ	ணை	ணௌ	ணஃ
த்	த	தா	தி	தீ	து	தூ	தெ	தே	தொ	தோ	தை	தௌ	தஃ
ந்	ந	நா	நி	நீ	நு	நூ	நெ	நே	நொ	நோ	நை	நௌ	நஃ
ப்	ப	பா	பி	பீ	பு	பூ	பெ	பே	பொ	போ	பை	பௌ	பஃ
ம்	ம	மா	மி	மீ	மு	மூ	மெ	மே	மொ	மோ	மை	மௌ	மஃ
ய்	ய	யா	யி	யீ	யு	யூ	யெ	யே	யொ	யோ	யை	யௌ	யஃ
ர்	ர	ரா	ரி	ரீ	ரு	ரூ	ரெ	ரே	ரொ	ரோ	ரை	ரௌ	ரஃ
ல்	ல	லா	லி	லீ	லு	லூ	லெ	லே	லொ	லோ	லை	லௌ	லஃ
வ்	வ	வா	வி	வீ	வு	வூ	வெ	வே	வொ	வோ	வை	வௌ	வஃ
ழ்	ழ	ழா	ழி	ழீ	ழு	ழூ	ழெ	ழே	ழொ	ழோ	ழை	ழௌ	ழஃ
ள்	ள	ளா	ளி	ளீ	ளு	ளூ	ளெ	ளே	ளொ	ளோ	ளை	ளௌ	ளஃ
ற்	ற	றா	றி	றீ	று	றூ	றெ	றே	றொ	றோ	றை	றௌ	றஃ
ன்	ன	னா	னி	னீ	னு	னூ	னெ	னே	னொ	னோ	னை	னௌ	னஃ
்		ா	ி	ீ	ு	ூ	ெ	ே	ொ	ோ	ை	ௌ	

[1] English is guilefully shown as easiest language to learn by having only 26 letters. This is a preemptive/brainwashing statement used by the English administrators. At first instance it has of

52 characters to be learned and there are polyphonics assigned with contextual relationship, including a null status between phonetic and written form. There are polygraphs too for many phonems which are hiden. Therefore each and every word has be learnt for its spelling and its assigned phonetics.

[2] Tamiz language was allotted with 128 code points in Unicode+ (h0B80-2944 to h0BFF-3071) on par with other Indian scripts as Devanagri, Bengali, Gurumuki, Gujarati, Oriya, Telugu, Kannada and Malaya lam. There is provision for ukara uukaara ottezuttu (modifiers) at slots h0BC1-3009 and h0BC2-3010.

[3] Tamiz software for these fields such as Braille and body language/hand signals are also to be developed in future an par with other languages.

[4] The letters zu Lu - ழ்ள- are shown here in a better shape. What we see in print and in computer fonts these days can confuse many readers. They have to infer the correct letter from contextual usage. People write the letter - ற- as - ர- a shortened form for urgent or for fancy. Just in the same way some have started writing za - ழ- as - ழ- which when changing into ukara uyirmey got transformed into forms which are creating ambiguity. See how similar the zu and mu - ழு- looks in this form. Even the za - ழ- itself is not distiguishable from zu. - ழ

In the case of La - ன- see the similarity in the letters Na, La, n2a - ண ன ன- In hand writing, rightside vertical strokes of these 3 letters are finished on the horizontal rule. Hence when extending into ukara uyirmey these letters have to change similarly as Nu, Lu, n2u - ணு ளு னு- But with La - ன- some have started writing Lu ளுas - ரு- for short or for fancy, creating non uniformity. Similar shortening laziness or false fancymania has not crept into Nu and n2u - ணு னு- because those are of older non-rationalised shape of Naa and n2aa - ணா னா- already. See now the non uniformity created by this practice resulting a condition of similar type of change process for different uyirmey character sets.

In the letter la - ல- though it belongs to the set of characters ending with an upward stroke as nga, ca, pa, va, ya - ங ச ப வ ய- it is not given the similar adding vertical downward stroke for ukara uyirmey transformation as it is very difficult to differentiate between letters vu and lu - வு- in manuscript.