=WHY TAMIL SCRIPT REQUIRES=
VUTAM LAYOUT FOR INPUTTING TEXT
============================

Dated at Aruppukottai, the First of March 2009.

Introduction:

   VUTAM (u thamiz) is a "Type As You Write " Typewriter Key Board Layout that allows the typists to type Tamil the way they learnt to write it! Pragmatists know why "Type As You Write" is better suited to type Ancient Tamil Works. The key board has no intelligence and, as the name suggests, the layout provides for consonants with u and U vowel marks accommodated in the SHIFT and Ctr+Alt baskets, respectively. Thus the layout has an International key board facility. While the coding scheme follows that of TSCII, coding is Unicode spread in three baskets like the default ones provided by Micro Soft in their Win 2k, XP and Vista Operating systems.

Historical Stuff:

   Historically, before the Unicode Consortium came into existence, Tamil had to take recourse to what is now considered as hacked encoding, using the code spaces of other scripts, mainly English. Using the English script's space, the engineers found it necessary to utilize diacritic marks to represent the u and U vowel marked consonants, accommodating the vowels and other consonants in the available two baskets. This was the time when even some consonants with A and ai vowel marks were written by the public, in one go, or differently, compared to the present day's way of writing them.

   As a later development, interested experts finalized on TSCII. But then their coding scheme, suitable for text and email purposes, could not be accommodated in the English two basket keyboard. It required at least four baskets, for which software was hard to come by. Finding that a simple, dumb key board can't accommodate the codes, they zeroed in on the English Transliteration Key Board, requiring a lot of built in intelligence for every script or language.

   To weed off the accusation of hacked encoding and taking advantage of the present way of writing the script, efforts were on to find a suitable key board driver using Unicode, if possible for free, or for an affordable price, affordable for even a group of philanthropists, who can offer their products for free use of individuals or educational institutions.

English Transliteration:

It may be commented here that English is least suited as a model of Transliteration, considering its idiosyncrasies in spelling or script vs. sound that is represented by it. For example, consider how put and but are pronounced! This is only an ice berg. Refer to what Dr. Bernard Shah has had to quip in detail about it, and his well documented essays on this issue. That is the reason why the International Organization had to lay down rules of English transliteration for every conceivable written script / language, using the scheme.

People with missionary zeal purchased such driver software using English Transliteration and offered it for free to the world, to type Tamil in Windows. Already, private enthusiasts had developed thiru, adami etc, using the same method, but with their own encodings, to write and read Tamil Script, in DOS.

Intelligent Key Boards:

The English Transliteration Key Board driver has to have a lot of intelligence built into it. As an example, consider: Tamil a, is obtained by pressing English a; Tamil A is obtained by pressing A or by pressing aa. Without intelligence, aa should have to give a, a and not A. There are many such built-in ones. The difficulty with such intelligent keyboards is the gap between these provisions and the knowledge about them, possessed by the user. This gap further aggravates due to the knowledge the typist has, about the English language and its conflicts with the International Standard for Transliteration.

While in the initial phase of learning, the learning curve is quite steep, in course of time, it tapers off. Thus, even after consistent and prolonged use of such key boards, unintended errors do occur while typing. Even if there is a brief discontinuity in use, these errors become a menace.

Thus it will not be wrong to say that there is enough dissent in the use of the English Transliteration Scheme of key boards, amongst those who know to write Tamil with pencil and paper, themselves. They are in an overwhelming majority amongst the users of Tamil Typing. To mitigate this problem, it is clear that a dumb keyboard, with no intelligence built in, will have the least of such errors, and should be the aim of any well meaning individuals or groups. When nothing better was available, it was alright to be satisfied with available ones; but when things are looking up, when it is possible to get something to fit as a T to our needs, it is time that we grab it. If such a one can be made suitable for even the mechanical type writers, which at present use diacritic marks, so far the better!! After all, inputting is the catchment's area of typing anything (document) already created and needs to be simple to use, clean, error free and natural.

Encoding & Accommodation:

Now it is time to look at such a possible key board and Unicode's encoding for Tamil. The Unicode Consortium has coded Tamil script within 128 code points (2944 - 3071), of which 54 are lying spare. They have accommodated all 12 vowels, Aitham, 23 consonants with a vowel mark, puLLi, kaal, ai vowel mark, vowel marks (e, E, i, I, o, O, u, U), au vowel mark and its length marker, Tamil numerals, Tamil abbreviations like ten, hundred, thousand, patRu, varavu, -do-, no., yr., month, dt., etc etc. OM is the latest addition. ksha and Sri are not given code points. Two alien characters also find a place in our area!

One basket of a normal keyboard accommodates 47 keys, of which 26 are occupied by English characters and 21 by numerals plus punctuation marks. Since Tamil also requires these numerals and punctuation marks, as far as possible, they should not be disturbed. Thus we have about 26 key spaces per basket, in the normally available English key board. In the International key board that has three baskets, we will be able to use all the 47 key spaces in basket three.

We require to accommodate, in basket one, the best basket, apart from numerals and punctuation marks, 6 short vowels, other than au, 18 consonants with a vowel mark and 3 vada mozi scripts, making a total of 27 key spaces. So we have to usurp one punctuation mark's space. Note that if we have to accommodate everything in basket one, then we have to have an intelligent key board, which is the least desired. Vowel marks of importance like ai, puLLi, kaal, e, i are also to be accommodated here, requiring 5 more usurpations, assuming vowel marks o, and O can use kaal with e and E vowel marks, and au vowel mark can be obtained with e vowel mark and La, as in mechanical, unintelligent, type writers.

In basket two, we can accommodate all the long vowels and the long vowel marks. Since vowel ai doesn't have a long form, perhaps its vowel mark can occupy its space in this basket, thus reducing the usurpation in basket one, by one. For minimizing these usurpations, au and its vowel mark need not be separately provided for, even though codes are available for these. Providing ai vowel mark in basket two will also reduce the strain of the human memory, if suitably positioned with reference to the vowel ai. However it should be noted that ai vowel mark is often required compared to vowel ai.

Non use of o and O vowel marks, like au and au vowel marks reduces more usurpation.

In basket two, a lot of key space is still available!

Tamil Phonetic:

Now let us look at the way we write Tamil, with pencil and paper. The 18 consonants with u and U vowel marks are written in one go as against the other consonants with kaal, puLLi, ai vowel mark, e and E vowel marks, o and O vowel marks.

For these 18 + 18 consonants, the Unicode consortium has not allotted separate code points; presently they are required to be typed, using the u and U vowel marks used in conjunction with the five vada mozi characters, which are then "transformed" by special display software, called Uniscribe, to look like the conventional script. However, in the concerned document, ku (say) is written as k followed by the u vowel mark used for the vada mozi script!! Is this cheating or anything short of it?

Again, in Tamil, vowel marks are either written before the base consonant or to its right or even in both sides of it. In Unicode, however, every vowel mark is to be written, following the base consonant only, which is then "set right" in display portion, using Uniscribe. In the document, the codes are still in the typed order, vowel marks following the base consonant! So it may not be wrong to say that the Unicode Consortium has taken upon them the responsibility of deciding how Tamil is to be written by Tamils!! The Unicode Consortium has made the fundamental mistake of dubbing Tamil Script as a complex one, requiring a complex display engine; while the fact is that the Tamil Script is as simple as the English one, not requiring anything special. Some like to complicate, otherwise simple things!! They only believe in, two negatives make a positive!

It may be noted, while passing, that whatever is displayed is printed. Remember "print screen".

Solution:

How to resolve this situation to facilitate the dumb type writer that doesn't require intelligent software in the key board or in the screen?

If the Unicode Consortium:-

1) Allots 18 + 18 code points for these consonants with u and U vowel marks,
2) Allots ksha and srii their own code points,
3) Vowel marks are allowed to be typed as they naturally occur in written Tamil, which will require at least three more code points and
4) Vowel au and its vowel mark are allowed to be typed as two / three characters, not requiring any new code points,

the problem is solved.

Do we have enough vacant space in the allotted space? Yes! Then where is the hitch?

Script Reform:

Script reform, even though may solve the situation without the Unicode Consortium's cooperation, is it acceptable? When a very recent body like the Unicode Consortium (mushroom of yesterday's rain!) is steadfast by its "stability" and "downward compatibility", will the users of the classic language budge?

Even though script reform to the extent of keeping ALL vowel marks simple, (with only one character and not two as required in o, O and au vowel marks surrounding the base consonant}, and to the right of the consonants concerned, including u and U vowel marks for all base consonants, as for vada mozi, has definite advantages in the very early stages of learning the script, and also in its retentivity in the background of early discontinuance of studies at the elementary level, the reformists do not have the courage or missionary zeal to propound it; they are also not sure of the extent of such reforms; viz, to go too deep like the Marathi script, where the vowels are also written in a simplistic manner, or where to cry a halt!! Since the script users will have their own rights of stability &c., we need only to cater to the present situation.

Whatever little we may gain in script reform, should not bring a very huge loss of forgetting our roots with our earlier scripts and our links with other dead or live languages!!

It may be in order, to point out to the reformists, that even today people write in documents, name boards, sign boards etc., in the old style for naa, Naa, Raa, elephant trunk like ai vowel mark for some consonants, one go for kku etc.

Grammar:

There is an opinion that only whatever is agreed upon in ancient Tamil grammar books, should be considered for coding. While the grammar books can at best be snapshots of the language at the time of their writing, they can't be movies of the ever evolving, live language! So it will be in order to consider the script in its present form.

Implementation:

Assuming the Unicode consortium is able to see eye to eye with us, how do we go further? Do we have enough vacant space in basket 2 to accommodate all these consonants with u and U vowel marks? No! Then how to go about it? Basket 3, even though not to everyone's liking, but is used in Europe as the International key board, and is available in Microsoft's Operating System as default International Key Board and for its Tamil "phonetic" key board, is the only solution, where we can accommodate these and some more!!

In basket two we can accommodate all the long vowels, the 3 left out vada mozi ezuththkkaL (2 + u hook) and the, to be single coded 18 consonants with their u vowel mark. In basket 3 we can accommodate all the left over characters and the usurped

punctuation marks. Thus we will be able to have Tamil numerals, srii, Aitham, the, to be single coded 18 Tamil consonants with U vowel mark, U vowel mark for vada mozi, Tamil abbreviations and the like. Thus, while English uses the two baskets for its small and capital letters, Tamil will use three baskets for consonants with a vowel mark, consonants with u vowel mark and consonants with U vowel mark, giving the name VUTAM.

Third basket should be with Ctrl + Alt, and not with Alt Gr., to facilitate easy typing with both hands, which will be already cumbersome in this basket. Fortunately, the provided characters in basket 3 are only occasionally, if not rarely, used.

Other issues:

There is a feeling that ALL the Tamil letters, around 300, should be given individual codes, doing away with the vowel marks, for easy searching, sorting and data integrity. This scheme will require at least seven baskets, for the "Type As You Write" key board, and hence not suitable for it. Intelligent key board becomes the only solution for this, with its own pranks, like remembering the grammatical rules of conjuncts (?!?) and the key sequences required to achieve these! It is better to be fore warned that this scheme is akin to the Transliteration and the Phonetic schemes!

Sorting and searching, required to be done on already created stuff, that too only occasionally, can be achieved by translating the available document, with a suitable internal and temporary encoding for the job on hand, doing the job in this temporary encoding and then reverse engineering the result in the normally encoded document.

All very important documents, like Government records and the like, can be stored in any of the Unicode's Transformed formats, for data integrity.

Layout:

General

Probably the English key board was the first to have its layout. It doesn't have a standard till date!! It has officially two "standards" one qwerty and the other dworek. Even today, freeware is made for changing its layout suitable to its user! Hence it is imperative that any freeware key board driver should have a built-in facility to change the layout for Tamil, and it is encouraging to say that such a one is presently available!!

VUTAM

The layout in all three baskets has to be user friendly, in that the user should not be required to memorize a lot. Those who resort to Tamil typing would have already passed English typing or will be using a key board with English scripts written on the key tops. To minimize memorizing, the five English vowels are to be allotted with corresponding Tamil ones; the sixth e is to be in y. For the 6 vadamozi scripts, keys s, h, and j to have sa, ha and ja scripts. Punctuation mark reverse solidus to be given nja. PuLLi is in apostrophe, kaal is in solidus, e vowel mark in left square bracket and i vowel mark in right square bracket.

Allotments of consonants are on the basis of commonality of sound, shape, convention, nothing, in that order. Thus, b is for the biggest character Na; x is for Ra; z is for za, g is for nga, f is for La since while the English f looks like a gracing sheep standing on its hind legs, the Tamil La looks like a gracing sheep standing on its fours (this facilitates association in human memory); a is for a; o is for o, i is for ai, y is for e; e is for i; w is for ya; and q being left out is for thannagaram.

The long vowels, the long vowel marks and the consonants with u vowel marks occupy corresponding positions in basket 2, corresponding to their counter parts in basket one. In this basket, S has sha; H has u vowel mark for vada mozi; J has ksha. Vowel mark ai is in I.

In basket three, Aitham in a; sri in s; vowel mark U for vada mozi in H; the 18 consonants with U vowel mark occupy their corresponding positions with respective to the consonants with u vowel mark in basket two. Tamil numerals occupy the corresponding locations of the normal ones in basket one. Vacant spaces in this basket may accommodate Tamil abbreviations. Braces may occupy - and +; square brackets their corresponding positions; apostrophe its corresponding position; \, | and / to occupy positions of , . and /. Positions of grave accent and semicolon can be occupied by Tamil Om and section marker.

The suggested lay out will have the least strain on the memory, even though may not confirm to "most used" character to occupy "best" position proposition!! Ease of use becomes a primary requirement, like the dumb key board that can also be used for the mechanical typewriter layout.

Conclusion:

The freeware, Vutam3.exe, is at http://sarovar.org/projects/keytrans

The layouts are in .map files. These files can be edited after they are opened with Notepad. It has facility for the conventional four baskets. For VUTAM, to get any character in the third basket, type space twice every time.

u and U vowel marked consonants are implemented in a round about manner, using ligature sequences. The pre positioned vowel marks are implemented, again, in a

round about manner, using ligature replacements. Thus the present layout uses intelligence. This could be done away with, if new codes are obtained for these, as mentioned above, from the Unicode consortium.

Those who were using the free transliteration key boards can be assured of unfrustrated and error free typing with VUTAM, if they can spend in VUTAM as much time and effort they spent in learning to type in the transliteration keyboards.

By,

V.Ramasami,
Aruppukottai.

****