**Requirement for numbering VarNam in Tamil**

**Basic Need:** Tamil does not have four VarNa letters. For Example, there is one ka. But Sanskrit has and Baraha coding uses K,g, G for second, third and fourth letters of ka varga. This is extened to all Consonants.

**What is done manually through Find and Replace in Baraha:**

1. We find letters K,g,G,C,j,J,Th,D,Dh, th,d,dh,P,b and B which are varna letters with all vowels and replace with respective letter +plus corresponding 2,3,4 after the vowel.
2. If Swara q,# or $ comes it reflects it wrongly when it reads as **ga3q.**

The anudattam come under the number and not vowel. So we replace all combinations of 2,3,4 for three combinations of Swaras q,# $ by placing swara before VarNa Number like **gaq3.**

1. Then we take all possible letter combinations of Conjunt Consonants and replace the VarNa for Conjunct Consonants – Ky as K2y, ddh as d3dh dByaH as d3B4yaH.
2. We take care of other issues like if dhai is replaced first dha combination will mark it as dha4i this creates wrong sound. So for ai and au, vowel sound we move the varNa number after a to after ‘i; and ‘u; by finding 2,3,4 between ai and au, resulting in ai3,au4 (examples)

**Limitations:**

1. **This is manual and subject to errors and missing any one or more combination that occurs.**
2. **Rare combination of conjuct consonants can be missed out which is specific to a PraSna, Subject or file.**

**Need for Python code to automate these steps:**

1. All these are simple rules which can be covered through Python.
2. The Program developed for Jatai and Ghanam will have some basic definitions for letter and swaram.
3. A separate program can be started which I will complete by adding more rules for every combination. After two/three trails it can be made perfect through coverage of all rules.
4. Input file will be a Word file with Baraha Code.
5. Output file will be a word file with Baraha Code with the VarNa numbers.

**Extension of this program to other languages:**

1. Since separate Baraha files are used presently for each language for considering language specific issues, Some find and replace steps are done specifically for Malayalam.
2. Such requirements exist in Kannada and Telugu which were manually taken care by Vounteering Groups.

**Extension of this Program for Validations:**

1. During find and replace we miss nasal symbols, dot for swarabhakti, dot after n for pause and certain conventions of representation which can be validated through one Program.
2. This will be actually automated in Jata and Ghana Program.
3. All existing files should go through these checks so they are made more accurate through actual replacement or giving output of an error file.
4. When we give Padam input in Excel, we would also like to give a few validations for a small program so that Padam input becomes more reliable and accurate. Manual/typing errors are weeded out before Output generation to the best extent.

**Way Forward:**

1. I need to download same version of Python, Front end and other tools and functions that are inbuilt as in your system.
2. Start my basics on simple syntax and adding to your code for additional rules.
3. Once a small framework is developed, we can add validations in the specific function, module or loop as suggested by you.
4. Jatai and Ghanam Baraha input could be converted easily to Malayalam and Tamil. This we plan to do only after Sanskirt version is made thorough to the best extent possible by us.
5. It has been confirmed by Shri Nurani Venkat that Malayalam involves three replacements only and no python will be required.   
   His funny remark was “no AK 47 to catch a hen !!!”

**(Addendum to Specs starts from this Page dated- 20/01/2020)**

**This part has not been programmed yet.**

**Conversion of ன to ந:** (Typed Tamil letter with Google Tamil input)

Baraha converts n to “ன” and n with a caret symbol (~n) correctly to “ந”

Beginning a statement with this hard na (**ன)** is not acceptable nor using soft “ந” everywhere.work annam typed with soft na looks odd for a Tamil readers. Please see the Baraha Trail typing file by opening in Tamil language.

1. So replace n with ~n (caret symbol before n) at the beginning of a statement, padam or if a space is available
2. We type brackets or insert “-“ hypen to distinguish sandhi,Korvai or for readability. So ‘n’ after a bracket or start of lines with any such character shall be replaced with ~n
3. Strings **‘nt’, ‘nd’ and ‘nR’** appearing in the middle of a Padam, word etc (other than at start shall be replaced with ~n respectively.
4. If ‘nn’ **is continuous** don’t replace.

**Conversion of anuswaram typed as m or M:**

1. Within a word like SaMgrAmam (in Sanskrit) looks odd in Tamil it is to be read as Sa~ggrAmam. This also works for the reading across words where nasal of the following letter class is replaced for the anuswaram. E.g. gaNAnAm tvA is read as gaNAnAn tvA.
2. Please do not combine this n and t together. Reader should know there is a conversion of anuswaram
3. So replacement Logic for anuswaram within and across word is

|  |  |  |
| --- | --- | --- |
| Ending in anuswaram | Following letter class | Replace with |
| M or m | K,k,g,G | ~g |
| M or m | C,c,j,J | ~j |
| M or m | T,Th,D,Dh | N |
| M or m | t,th,d,dh | n |
| M or m | ‘n’ | ‘n’ |

1. If the following letter is a single letter like te,no,ca combine them with the nasal of the previous word

dravinam ca hangs as dravina~j ca does not look good. Better be dravina~jca. ( we have not been 100% consistent in this due to manual editing or deleting the space and we are sleepy at night ☺ )

“SaM no” becomes “Sanno together” as in Sanno mitra Sam varunaH. “saM te” becomes “sante”. These look better presented and readable.