

Symmetries in Number Forms of Tamil and Dravidian Languages

Authors: Muthiah Annamalai[†] ezhillang@gmail.com

Abstract

We propose the Tamil number forms are equivalent by isomorphism (single rule over all numbers to corresponding number forms) in Telugu, Kannada and Malayalam. The latter being almost indistinguishable from Tamil except for prosody; this is based on intuition of digit forms [1]. We further contend that algorithm for generating numerals in any of the four languages are structurally identical due to the equivalence of numerals in a abstract way. We propose common algorithm for generating and parsing number forms [2] in these languages to/from text and into audio TTS generation. These can be used in various applications like token-queue systems, spoken calculators, etc.

1 Introduction

It is quite well known the digits of Dravidian languages and even number forms are roughly symmetric [1]; however the structure of this symmetry is the subject of this paper and we demonstrate the structural symmetry can be applied to a computer algorithm for generation and parsing of numbers in these language simultaneously. To the best of our knowledge this is a first effort in such linguistic motivated algorithm.

2 Evidence

Using a publicly available corpora [3,4] and using Google translate [5] we have established parallel evidences of the sample number forms across the 4 southern languages. By inspection one can ascertain a rough correspondence; however this symmetry goes beyond rough correspondence as authors prior work [2] can be extended by parameterization with suitable suffixes and modifications by language family we posit a single algorithmic routine can perform both parsing and generation of a numbers.

எண் மதிப்பு	தமிழ்	மலயாளம்	கன்னடம்	தெலுங்கு
1	ஒன்று	onn	Ondu	Okaṭi
2	இரண்டு	raṇṭ	eraḍu	reṇḍu
3	மூன்று	mūnn	mūru	mūḍu

[†]Corresponding Author:

Ezhil Language Foundation,
Hayward, CA, USA

4	நான்கு	nāl	nātku	nālugu
5	ஐந்து	añc	aidu	aidu
6	ஆறு	ār	āru	āru
7	ஏழு	ēḷ	ēḷu	ēḍu
8	எட்டு	eṭṭ	eṇṭu	enimidi
9	ஒன்பது	ompat	ombattu	tom'midi
10	பத்து	patt	hattu	padi
11	பதினொன்று	patineānn	hannond	padakoṇḍu
12	பனிரண்டு	pantraṇṭ	hanneraḍu	panneṇḍu
15	பதினைந்து	patinañc	hadinaidu	padihēnu
19	பத்தொன்பது	patteāmpat	hattombattu	pantom'midi
20	இருபது	irupat	ippattu	iravai
30	முப்பது	muppat	mūvattu	muppai
40	நாற்பது	nālpāt	nalavattu	nalabhai
50	ஐம்பது	ampat	aivattu	yābhai
75	எழுபத்தைந்து	eḷupatti añcu	eppattaidu	ḍebbai aidu
90	தொன்னூறு	teāḷḷāyiraṁ	ombainūru	tom'midi vandalu
99	தொன்னூற்றொன்பது	teāḷḷāyiratti ompat	ombainūra ombattu	tom'midi vandala tom'midi
100	நூறு	nūr	ondu nūru	vanda
200	இருநூறு	irunnūr	innūru	reṇḍu vandalu
500	ஐநூறு	aññūr	aidu nūru	aidu vandalu
1000	ஆயிரம்	āyiraṁ	sāvira	veyyi
5000	ஐந்து ஆயிரம்	ayyāyiraṁ	aidu sāvira	aidu vēlu
154999	ஒரு இலட்சத்து ஐம்பத்து நான்கு ஆயிரத்து தொள்ளாயிரத்து தொன்னூற்றொன்பது	oru lakṣatti ampattināḷāyiratti teāḷḷāyiratti ompat	nūra aivattanāḷku sāvira ombhainūra ombattu	nūṭa yābhai nālugu vēla tom'midi vandala tom'midi

Table 1: Parallel listing of number to words in Tamil, Malayalam, Kannada and Telugu

3 Algorithms

We modify the algorithm first presented in [2] for generating integral and floating point non-negative numbers for Tamil, but instead by view of symmetry in Tamil, Kannada, Malayalam and Telugu together called Dravidian languages (**DL**) we reorganize it as, follows,

applied piece-meal to each of **DL**. In simpler terms we say the algorithms of [2] are parameterized by suffixes and prefixes specific to each DL but the overall structure remains the same by property of isomorphism. We note the steps for joining sections are to be handled in language specific way yet overall algorithm being invariant to source language.

3.1 Algorithm for Generating Numbers

Input: floating point number **N**

Output: string of **DL** words **T**

Algorithm:

1. Load list of prefix and string suffix for all **DL** number words - 63 words in all.
2. Find the quotient **Q**, remainder **R** for **N** divided by 1 crore, lakh, thousand, hundreds, or tens
3. If **Q** is zero set **N=R** and continue to 2.
4. Convert the quotient to words **T**
 - a. Take special care to handle 90s, 900s, 9000s, correctly.
 - b. Take special care to handle number in 11-19.
5. Invoke same algorithm recursively for remainder **R**.
6. Concatenate results from 5 to **T**
7. Return **T**

3.2 Algorithm for Parsing Numbers

Input: string of **DL** list of words **T**

Output: floating point number **N**

Algorithm:

1. Load list of prefix and string suffix for all **DL** number words - 63 words in all.
2. Initialize **N** at 0
3. Create temporary stack **S**
4. FOR word **W** in **T**
5. IF **W** in stop words (crores, lakhs, thousands, hundreds, tens)
 - a. Convert words in stack **S** into value and scale temporary result **N** using a helper routine which handles input upto value 100,0000.
 - b. Empty stack **S**
6. ELSE: push **W** into **S**
7. END loop started at 4.
8. Stack **S** is mostly non-empty and you have to use a helper routine to get the final portion of the number using the same helper function in 5a.
9. Correctly parsed value is stored in **N**

4 Applications

Similar to the applications presented in [2] we can enable parameterized, by Dravidian Language (DL), a TTS generation, and calculator applications by reusing the algorithms of sec 3.1 and sec 3.2.

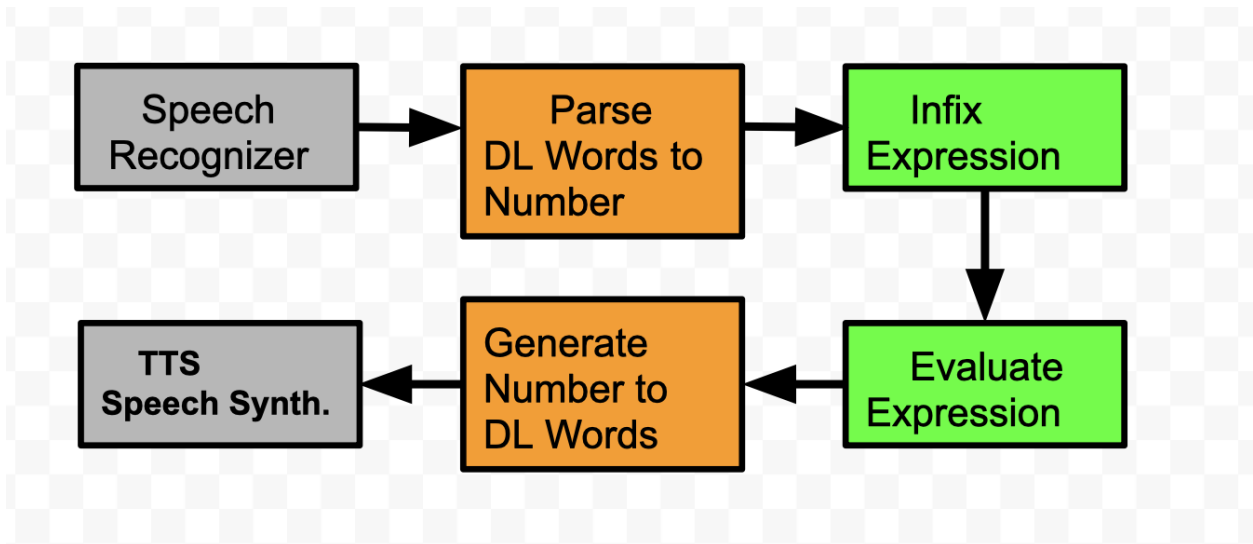


Fig 1: Organization of Speech input and Audio output calculator in each Dravidian Language (DL)

5 Summary and Conclusion

We have established algorithms to exploit the symmetry of number to words in Dravidian languages of Tamil, Malayalam, Kannada and Telugu for various applications; we demonstrated parameterized algorithm for generating and parsing number forms in these languages and provided a framework for applications like token-queue systems, spoken calculators, etc. leveraging this discoveries.

References:

1. Wikipedia on Tamil Numeral Influence,
https://en.wikipedia.org/wiki/Tamil_numerals#Influence (accessed Nov 14, 2022)
2. M. Annamalai, S. Mahadevan, "Generation and Parsing of Number to Words in Tamil," Tamil Internet Conference, 2020.

3. Rao Vemuri, Freshman Lecture notes from “Learn Telugu and Its Culture,” at UC Davis, Fall 2006. <https://www.cs.ucdavis.edu/~vemuri/classes/freshman/index.html>
 4. Malayalam Numbers, <https://www.learnentry.com/english-malayalam/vocabulary/numbers-in-malayalam/> (accessed Nov 22, 2022)
 5. Google Translate, <https://translate.google.com> (accessed Nov 22, 2022)
-