Comparitive Study between Today's Korean and Tamil Languages, their Lexicon and Grammar

Language and Society: CS2.203

Author: Harinie Sivaramasethu

11/17/22

Professor: Dr. Dipti Misra

Contents

Abstract

A genetic link between the Dravidian languages and Korean was first hypothesized by Homer B. Hulbert in 1905. Morgan E. Clippinger, a Korean studies scholar, gave a detailed comparison of Korean and Dravidian vocabulary in his article "Korean and Dravidian: Lexical Evidence for an Old Theory" (1984). He further proposed over 400 sets of similar words between the two languages.

This is a study between the spoken languages of Korean and Tamil. It will verify the long-proposed link between the two languages by comparing the present-day spoken languages in Tamil Nadu (through native language speaker competence) and Korea (through Korean TV shows, popularly called K-drama).

This project provides an opportunity to look further into these connections, to see if there are any social similarities, and a chance to explore why such a resemblance came into being.

Background

Historical Links

India's historical connection with Korea goes back to more than 2000 years when Korea was a collection of Tribal city states. A famous Tamil quote "Yaadhum Oore Yaavarum Kelir" by Kaniyan Poongunranar (600 BCE – 100 CE) which means, "To us all towns are one, all men our king". The quote has been followed by the Dravidians since ages. The historical record clearly shows that people of ancient Tamilakam had a good relationship with the rest of the world in terms of maritime trade and soft power diplomacy. (Aj)

Six million Koreans or almost 10% of the Korean population trace their ancestry to an Indian Princess. Heo (or Sembavalam in Tamil) was the princess of Ayuta Kingdom which was located in the southern part of India. Ayuta Kingdom here refers to "Ay Kingdom" that belonged to the Pandyan dynasty of ancient Tamilakam. This can be substantiated with the fact that princess Heo carried with her the symbol of Twin fish in trident. Pandyan Kingdom depicts two fishes on their flags, coins, and emblem. (Kannan)

Historical evidence clearly shows that the ancient Tamils were seafarers who traded across the seas with different parts of the world including the South East Asian countries. The Pandyan kingdom of Tamil Nadu had maritime trade connections with Korea 2000 years ago which could be a major reason behind the language and cultural similarities between India and Korea.

A unique similarity between Tamil Nadu and Korea is pearl hunting. The activities like pearl hunting in the industry for coral ornaments were familiar only in ancient Tamil Nadu about 2000 years ago when princess Heo got married to Kim Suro. Surprisingly, even today women in Tamil Nadu and South Korea are involved in deep-sea diving for the purpose of pearl hunting. Koreans are largely rice eaters and their paddy cultivation was said to have been taught them by the Tamils.

Mythology

- Dan-Gun
  The legend of "Dan-Gun, First King of Korea" is narrated in Samguk Yusa [14], a Korean text written by a monk, Iryon (1206 AD-1289 AD).

  "In ancient times Hwan-in (Heavenly King, **Chesok or Sakrodeveendra**) had a young son who wished to descend from heaven and live in the human world. His father, chose T'aebaek-san (the Myohyang Mountains in North Korea) as a suitable place for his heavenly son to bring

happiness to human beings. He gave his son three heavenly treasures, and commanded him to rule over his people. The son led his ministers of wind, rain and clouds in teaching the people more than 360 useful arts.

A person familiar with Indian mythology will easily find parallels here. The heavenly father is referred in this myth as Sakrodeveendra. The God with Chakra and the word 'Deveendra' applies to the primordial supreme God, often cited by Vedas as Narayanan.

The ministers such as Earl Wind, Chancellor Rain and Chancellor Cloud that are referred in the Korean myth are also the elemental powers that governed ancient Tamil landscape (Ainthinai).

Literature Background

- Since 1905 when, on the basis of syntactic and typological similarities, Homer Hulbert first suggested a genetic relationship between Korean and Dravidian. Such a relationship had been suggested by nineteenth century French missionaries who, coming to Korea from India, noted many similarities between the Dravidian languages and Korean. In the eighty years since Hulbert published his comparative grammar, however, no serious systematic attempt has been made to pursue his theory.
- Later, Morgan E. Clippinger publised a paper presenting over four hundred sets of Korean-Dravidian cognate pairs, including many basic vocabulary items, in an attempt to uncover links between the two language groups.

  He makes the connection that Dravidian and Korean must be closely tied to the history of migration of the peoples of Asia, a subject that is still not clearly understood. The connection could have resulted from direct or indirect genetic links, from intimate contacts between the two language groups and a third language group at an ancient period, or from contacts resulting from migration to the Korean peninsula at a fairly recent period.

  Some of the Korean-Dravidian cognates are so close that one is inclined to discount any kind of ancient relationship, while other cognates can be explained by well-documented sound changes in both languages, suggest ing a much earlier link. In short, the evidence suggests that at a very ancient period Dravidian and Korean shared a common heritage, and this heritage was reinforced much later by migrations to the Korean penin sula, perhaps in the later years of the first millennium B.C. (Clippinger)

Method: Potential Korean-Tamil Cognates

The theory proposed earlier looked at words in usage at earlier times. Our goal today is to see if the proposed link still exists in today's lexicon of both languages. Having native language speaking proficieny in Tamil, we shall compare that and try to recognize similar lexicon from Korean dramas, which will be our source for current spoken Korean language in the country of South Korea.

The following lexicon have been taken from different resourceful Korean dramas (K-dramas), each of which focus on different themes to enable us to get data from different domains. The reason we have taken them from K-dramas is because it is part of pop or current culture, meaning it always tries to stay most relevant to current times, be it in terms of food, clothes, fashion or language and speech. So, to study the current Korean societys' speech

pattern with the current Tamil spoken speech society's we are using K-dramas. Hence, the chosen K-dramas include

- Flower of Evil
  Themes: family, wife, husband, cop, serial killer, fighting, blood, hospital
  (So-hee)Themes: woman, boxing, mafia, father, boss, gun, police
- Vincenzo
  Themes: humor, mafia, money, fighting, lawyer, pharmacy, family, father, daughter, building
  *English TV series which depicts korean language usage (since main character and character's family are of korean ethnicity)
- Kim's Convenience
  Themes: grocery, humor, family, daughter, wife, husband, shop, sale, food
- Doctor John
  Themes: doctor, medicine, illness, patient, body parts, kindness, empathy
- Little woman
  Themes: lawyer, sister, flower, nature, journalism, art, money, love
- My secret romance
  Themes: dietician, nutritionist, food, cook, grocery, love, marriage, wife, parents, in laws, mother

Aside from K-dramas, there are other pop culture resouces which have also been considered for such as youtube korean street food videos (for food references), youtube videos (korean-tamil), etc.

Adding to these cognate sets, we shall also be going through the 400+ sets proposed by Morgan. E. Clippinger. (Clippinger)

Let's look at the cognates from the above given series.

1. Kinship

- Mother
  Tamil: அம்மா - amma
  Korean: 엄마 – umma

- Father
  Tamil: அப்பா - appa
  Korean: 아빠 – appa
  According to respective Wordnets, there is also common usage of the term 'abbeoji' in korean which can resonate to 'appaji' in Tamil.

- Aunt
  Tamil: அண்ணி - anni
  Korean: 언니 – unnee

- Wife
  Tamil: மனைவி - manaivi

Korean: 마님 – manim

2. Personal Pronouns

- Me
  Tamil: என்னை - ennai
  Korean: 나 – na

- I
  Tamil: நா - na
  Korean: 나 – na
  In Tamil speech, 'I' is commonly used as 'nan' or 'nanu' of which there is an equivalent variation in korean as well, which goes by 'naneun'.

- You
  Tamil: நீ – ni
  Korean: 너 – neo Source: My name
  In Tamil speech, 'you' is commonly used as 'neenga' of which there is an equivalent variation in korean as well which goes as 'neega'.

3. Body parts

- Stomach
  Tamil: வயிறு- Vairu
  Korean: 위- Vae

- Teeth
  Tamil: பல் - pal
  Korean: 이빨 - eepal

- Eye
  Tamil: கண் - kan
  Korean: 눈 - nun

- Hair
  Tamil: முடி - mudi
  Korean: 머리 - meoli

- Head
  Tamil: தலை - thalai
  Korean: 테카리 - tekari

- Nose

Tamil: மூக்கு - mhukhu
Korean: 코 - khu

- Heart
  Tamil: மனம் - manam
  Korean: 마음 - mam

- Knee
  Tamil: முட்டி - mutti
  Korean: 무릎 - murup

4. Cultural terms/ items

- Greeting
  Tamil: வணக்கம் - vanakkam
  Korean: 박까암따 - vakkaamtta
- Knife
  Tamil: கத்தி - katti
  Korean: 칼 - khal

- Stick
  Tamil: குச்சி - kutchi
  Korean: 큐 - kkuo

- Cleanliness
  Tamil: சுத்தம் - suttam
  Korean: 순수한- sutswam

- Day
  Tamil: நாள் - naal
  Korean: 낮 - naal

- Tommorrow
  Tamil: நாளை - naalai
  Korean: 내일 - Naeil

- This
  Tamil: இது - idhu
  Korean: 이것 - igo

5. Food

- Filled dumpling
  Tamil: கொழுக்கட்டை - kozhukattai (coconut, jaggery filled)

Korean: 만두 korea mandu (meat, vegetables and tofu filled)

- Rice porridge
  Tamil: கஞ்சி kanji
  Korean: 韓 Kanji

- Cooked Rice
  Tamil: சோறு - sourru
  Korean: 쌀 - saal

6. Flora, Fauna and Nature

- Grasshopper
  Tamil: வெட்டுக்கிளி - vettukilli
  Korean: 메뚜기 - maettugi

- Snake
  Tamil: பாம்பு - pambu
  Korean: 뱀 - bambu – baem

- Grass
  Tamil: புல் - pul
  Korean: 풀 - pul

- Manure
  Tamil: உரம் - oram
  Korean: 비료 - uraam

- Mountain / hill
  Tamil: மலை - malai
  Korean: 언덕 - malrangi

- Shore
  Tamil: கரை - karai
  Korean: 물가 - uka

7. Expressions

- Oh!
  Tamil: ஐயோ - aiyo
  Korean: 아이고 - aigo

- Little – little
  Tamil: கொஞ்சம் கொஞ்சம் - konjam konjam
  Korean: 조금 조금 - Jogeum Jogeum

- Maybe/ yes
  Tamil: ஆமாம் - aamaam
  Korean: 아마도 - aamaado

8. Verb

- Draw/ scribble
  Tamil: கிறுக்கு - kirukutal
  Korean: 그리다 - kirida

- Fight
  Tamil: சண்டை - sandai
  Korean: 싸움 – ssaw

- Knowledge / to know
  Tamil: அறிவு - ariwu
  Korean: 앎 – ar

*Romanization for the above words is done solely based on author's hearing of the lexicon directly from source (K-drama, Youtube).

Discussion

1. Observations from lexicon data collected

a. Phonology

- **First root**
  In general, the first sylable (root or stem) remains to be most persistent when we compare the data above.

- **ஜ (ja), ஷ (ṣa), ஸ (sa), ஹ (ha)**
  These letters were not originally used in Tamil script and we shall see how the script got by without using them in the next segment (convertability law). Now, we have borrowed these four letters from the sanskript script for usage purposes.

  So having established that pure Tamil rejects the sharp Sanscrit sibillants s and sh, a soft sh with as little sibillation or aspiration takes their place. When doubled, the sound ap proaches that of ch or tsh.

  If we carefully observe, we may observe a similar phenomenon in Korean as well. In Korean, similarly, s is not pronounced on the tip of the tongue as in English, but is made with a sort of lisp as if sh were being pronounced without protruding the lips. (Hulbert)

- **Sentence - Closing Ending**

In korean speech, all sentences end with 요 (yo sound). And that's how you know how a sentence ends or not. This special suffix 요 is called 'sentence-closing ending'. The Korean language uses it and it literally does what it's called. It concludes a sentence right away when you use it. So, basically, it's the same as the period mark in English.

저는 예뻐요 저는 미쳤어요

I'm pretty. I'm crazy.

So, if one said '저는 예뻐요 저는 미쳤어요', you can actually understand they are 2 different sentences even without period marks. This phenomenon is also very evidently noticeable when one watches any shot clips of Korean language conversation.

To a similar extent, Tamil speech also usually ends with sounds 'uh', 'a', etc.

Eg) அவன் என்ன செய்கிறான்

'Avan enna ceykiran' is the literal pronounciation of the above sentence but in casual speech this would be uttered as 'avan enna ceyrə'

b. Morphology

- **Similar Agglutination**

  The Tamil, padu means to lie down, but padu-ttu means to lay down; tar means to be low but tar-ttu means to lower.

  In Korean, there is a similar class of verbs. durujinda means to fall, but if it changes to durutturinda it means to let drop. In like manner, numujinda changes to numuttirinda.

  This change in form usually indicates a change from an intransitive to a transitive idea.

2. Other Observations

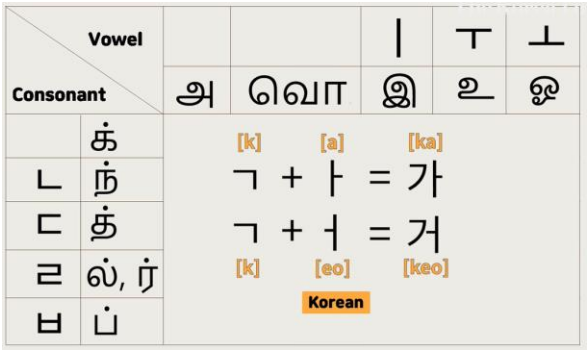
a. Vowel

Observing the lexicon, vowels tend to form the major basis of both languages. Tamil grammarians have recognized that the vowel is the life of the syllable while the consonant is the body. (Hulbert) In a similar manner, Koreans call the vowel – mother and the consonant – child.

We can compare this with an opposing variety like that in the case of Arabic. Here, the principal letters tend to be consonants and vowels merely help with the enunciation.

This is clearly visible when we look at vowel + consonant script formation. This form is also called long math form in Tamil. It is present similarly in many indian languages like hindi, etc.

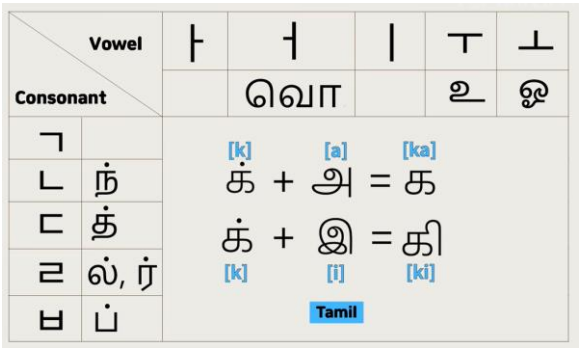
- **Hangeul (Korean script)**



- **Tamil**

  2nd row: அ(a) வொ(vo) இ(e) உ(u) ஒ(o)

  2nd column: க்(ik) ந்(in) த்(ith) ல்,ர்(il, ir) ப்(ip)



- **Arabic**

  Arabic alphabet systems include 28 consonants. There are also 6 vowels. These include three short vowels: a(Fatha), u(Damma) and I(Kasra) and three long vowels: aa(Double Fatha), ii(Double Damma) and uu(Double Kasra).

  Now words are strung together using the consonants and the vowels are added at different places for varied enunciations. Look at the example below where there are different vowels placed on the singular consonant 'b – ب ':

  'ba – بَ' (fatha) 'bu – بُ' (damma) 'bi – بِ' (kasra)

b. Phonology

- **Law of convertibility of Surds and sonants**

  The stops are "sonants" (or voiced in modern terminology) when they are intervocalic (that is, when they come between vowels) or when nasals precede them. This is why க is sounded like 'ga' in அகம் (agam) or அங்கம் (angam). When geminated (that is, when the consonant is duplicated), they are "surds" (or unvoiced). க sounds like 'ka' in அக்கம்.

  The system of representing more than one sound by a single letter is termed as the law of convertibility of surds and sonants. It is remarked as the Tamilian rule which requires the same consonant to be pronounced as k in one position and g in another, as t, th, p in one position and as d, dh, b in another is not a mere dialectic peculiarity. (Caldwell)

Let's look at this in further analysis.
If the word starts with this alphabet or has an ◌̇ (஼) over it, it is pronounced as [k,th,ch,t,p].
If it is in the middle of the word without the ◌̇ or following it's corresponding n it is pronounced [g,dh,s,d,b] exception being ச pronounced as j if after it's n.

Korean agrees perfectly with the Tamil in the use of the same letter for both surd and sonant. This applies not only to the written character but to the laws of sound which are inherent in the languages. In Korean, as in Tamil the consonants k, p, t, and ch are always surds at the beginning of words and whenever doubled.

- **Euphonic Permutation of Consonants**
  When two compounded words stand in a case relation to each other or when the first is governed by the second, the in intial surd of the second word is not softened but doubled and hardened in token of its activity. (Hulbert)
  Let's understand this in simpler terms by looking at examples:
  Tamil: பலா (pala) and பழம் (pazham) gives பலாப்பழம் pala-(ip)pazham not pala-pazham.
  Korean: ha (sun) and pit light do not combine to make ha-bit but ha-(p)pit.

- No clear demarcation in writing 'ga','ka' and 'pa' ,'bha'
  Eg) Koreans write Pusan or Busan alternatively as they can't exactly differentiate between 'pa' and 'ba' using their phoenitical system, as in Tamil.
  Coffee can't be written in Tamil and Hangul, instead only 'kappi - கஃபி' is possible as there is no letter 'f' in these languages

However, we cannot draw more substantial conclusions in the absence of quantitative data and a larger dataset, which we are lacking there of in the present study. Getting a larger dataset through comparing audio datasets can be done through further speech processing computational methods.

3. Computational

A further verification which links the two languages is modern computational linguistics.

- Similarity database
  Having selected a small number of cognates which have equivalent counterparts in both languages being studied, we compare the word meanings from core vocabularies and compare strictly on phonetic representation.

  CogNet is a large-scale database of cognate pairs: it contains **8.1 million cognates** in 338 languages, 38 writing systems, and 91285 concepts. It was automatically constructed from wordnets and dictionaries contained within the UKC resource.

  (Gábor Bella) is an already worked upon model which was established in 2021. It uses automated similarity computation. The identification of cognate pairs having already been done by CogNet. It computes the cognate-content-based similarity between the lexicons of languages A and B is as follows:

$$S_{AB} = \frac{\sum_{\forall <c_i^A, c_i^B>} \alpha + (1-\alpha)\text{sim}(c_i^A, c_i^B)}{\frac{2|L_A||L_B|}{|L_A|+|L_B|}}$$

Where $<C^A_i, C^B_l>$ is the i^{th} cognate pair retrieved from CogNet for the languages A and B and sim(C^A_i, C^B_l) is a string similarity value:

$$\text{sim}(w_1, w_2) = \frac{\max(l_{w_1}, l_{w_2}) - \text{LD}(w_1, w_2)}{\max(l_{w_1}, l_{w_2})}$$

Where LD is the Levenshtein distance and l_w is the length of word w.

This database is free to access and can be downloaded from http://ukc.disi.unitn.it/wp-content/uploads/2021/06/similarities_1.0.tsv.zip . In the database, the contents are of the form:

ISO_1	LangName_1	ISO_2	LangName_2	Similarity	Robustness
cat	Catalan	spa	Spanish	22.099	High
ces	Czech	tir	Tigrinya	0.183	Medium
dak	Dakota	lak	Lakota	89.250	Low

The korean – tamil relationships are as follows:

kor Korean tam Tamil 0.173 High

This above relation further acts as proof for the high similarities and link between both languages. The cognates used for this similarity can also be downloaded from github.

Summary

We first started off by talking about the historical connection between both countries. Having seen the connections in the past, we look at theories made in a more recent past. In the 20th century, many linguists hypothesized about the link between both languages.

Now our goal in this project was to see if this link still exists in today's societys' speech. To achieve that we had to collect data set from both parts of society and see if there exists and interlink between both circles in the venn diagram.

To get today's Tamil speaking society from main land Tamil Nadu's speech, I've used my native language speaker proficiency. On the other hand, to get South Korea Mainland Korean spoken society's speech dataset, I had to get my hands on a dataset which worked in a) spoken context b) present spoken language. The one thing that would efficiently cover both grounds is pop culture as it grows with the media. A media source where we can observe dialogue, proounciation, with proper english translation was easily K-dramas.

Having gone through a handful of dramas, I noted down words that stood out the most to me in terms of understanding without looking at the subtitles. Later I compared these with the Tamil counterparts (same meaning, similar pronounciation).

Having done so, we notice some curious similarities in terms of phonology, agglutination, and morphology. Further research also showed some similarity in the script and writing systems.

Conclusion

Having just summarized the project, we see how there are so many similarities between two languages which seem very unrelated. Two languages, which play huge roles in today's world in terms of entertainment be it K-pop, K-drama, Kollywood films, etc, but there has been little done to study both. One can only imagine the possibilities or heights we can reach.

From this project, one thing we can all conclude is that the link that was proposed centuries ago still exists. However, it is not easy to say if it very stable. Only firther research and study can prove that. What can be saif from this project though is that the link that was there years ago is still there.

Cross cultural understanding and bilateral relationship are pivotal in political and economical development among nations. A recent political slogan in India is "Look east" to capitalize the growth potential in Asia.

The similarities between both Tamil Nadu and Korea in terms of Language, cultural expressions, food and games have drawn the attention of many researchers to do a detailed comparative study. It is imperative for us to know and pass on to the next generation our historical events and exchange of cultures. Looking into the future requires an understanding of the past. Such an understanding can illuminate the present and enable one to think more meaningfully about the future.

Future Plans

- While researching for the project, I stumbled upon a non-profit organization, 'InKo Centre' which conducts consistent inter-cultural dialogue programmes that draws on the rich traditions of India and Korea. InKo Centre is a registered, non-profit society supported primarily by TVS Motor Company Limited and Hyundai Motor India Limited; the Korean Association in Chennai and a host of Indian and Korean companies based in Chennai.

  Nandini Menon, teaching and information personnel at the centre claims that there may be as many as 3000 such words in common between Tamil and Korean. This thought is well believed by all in the centre.

  Facing a time constraint, I was unable to contact them at the time of the project duration. However, in the future, it would be a wonderous oppurtunity to get in touch with the team at InKo to learn more about their research on this wide and interesting topic.

- An interesting discovery I made during my research for the project is the 'Similarity Database of Modern Lexicons'. It's been explained in the previous section in detail. In simple words tough, it takes in a limited dataset of lexicon in two languages and compares the similarity and gives a numerical output indicating similarity.

  They have provided enough data, and research papers for us to attempt to make a similar crude model. What we could work on is having automatic speech analysis which we can run on audio containing similar words in Korean and Tamil. This would be very useful in studying, say in our case K-dramas and spoken Tamil. Having the model directly get similar sounding

words from K-dramas, comparing them with an audio dataset of Tamil lexicon, and then providing us a direct similarity value is a project I hope to pursue.

Bibliography

Aj, Oviya. *Tamil Nadu and Korea: An untold tale of princess Sembavalam*. 29 October 2021.

Caldwell, Robert. *Comparative grammar of Dravidian Languages*. n.d.

Clippinger, Morgan E. "Korean and Dravidian: Lexical Evidence for an Old Theory." *Korean studies 8*. 1984. 1-57.

Doctor John. Perf. Lee Se-young Cha Yo-han. n.d.

Gábor Bella, Khuyagbaatar Batsuren, and Fausto Giunchiglia. "A Database and Visualization of the Similarity of Contemporary Lexicons." *24th International Conference on Text, Speech, and Dialogue*. Olomouc, Czech Republic, 2021.

Hulbert, Homer B. *The korean language and the dravidian dialects of india*. seoul, korea, 1906.

Kannan, Narayanan. *Tamil-Korean relationship*. n.d.

Kim's Convenience. Perf. Jean Yoon, Andrea Bang Paul Sun-Hyung Lee. n.d.

Little women. Perf. Nam Ji-hyun, Park Ji-hoo Kim Go-eun. n.d.

My name. Perf. Han So-hee. n.d.

My secret romance. Perf. Song Ji Eun Sung hoon. n.d.

The Flower of Evil. Perf. Moon Chae-won Lee Joon-gi. n.d.