

Language and Society (CL2.203)

Monsoon 2022, IIIT Hyderabad

**Comparitive study between today's Korean and Tamil language, their
grammar and lexicon**

Project Interim Report

Name: Harinie Sivaramasethu

Roll number: 2021114008

Native Language: Tamil

Abstract

This is a study between the spoken languages of Korean and Tamil. We will be comparing their shared lexicon 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).

We shall then use this to see if there is any link between the grammar of both languages.

Progress

Data Collection

I have reviewed two Korean TV shows (namely, Flower of Evil and) and have watched clips from other shows with Korean being spoken in them (namely, Kim's Convenience).

To date, I have collected over 20 such shared lexicon.

I have also started watching another popular K-drama, 'Hospital Playlist'. So far into this drama, I have picked out a handful of similar lexicon. Further analysis is required to correctly identify

Literature Review

I have also read through Morgan E. Clippinger's paper where he has over 400 sets of cognates. I have gone through all of them and have made a list of the cognates whose Tamil parts are used in today's spoken Tamil. The next step would be to analyse the Korean counterparts of these words.

I plan on using online translation portals between English and Korean, English and Tamil, Tamil and Korean, to analyse the counterparts. After attaining the word in the native scripts, the next step to analysing them would be to look up their relations through the respective language Wordnet.

I have also read the 'Tamil-Korean relationship' paper by N Kannan, mentioned in my Literature Review. There, he states that Hangeul (Korean language script) is similar to Tamil in agglutinated letters and combination of vowels and consonents. I plan on analysing this in the data I have collected so far.

Further Steps

I plan to proceed in the following way for the rest of my project.

1. Gathering a concrete dataset
2. Comparing the shared lexicon using respective wordnets
3. Classify lexicon based on different kinds: Nouns, Verbs, Onomatopoeic words, etc
4. Comparing the different contrasting stop positions, namely, those of labial, dental and velar
5. Gather etymologies of the words

From Clippinger's study, we realise that the first syllable, or the root or stem, is the most persistent part of the Korean and Dravidian lexical inventory. Abiding to this, I also plan to largely compare this part of each lexicon.

Clippinger further states that in Proto-Dravidian, the vowel system has five vowels with contrastive length—/a/i/e/o/u/—but no nasalization. Vowels in noninitial syllables are less stable and may fall into subphonemic vowel harmony.

In the Middle Korean, the vowel system had /a/ë/o/u/i/i/ and the diphthongs ëy > e; ay > â; and oy > ô. In addition, had a vowel ɔ, which might have been an unround midback vowel or a low-back rounded vowel, and a diphthong ɔy.

After focusing on the lexicon, I plan to move to a small study of both language's grammar in terms of syntax. Both Korean and Tamil are known to be primarily SOV (subject-object-verb) languages. This word order is similar in both languages with the modifiers preceding the modified word. A very basic analysis of both can be done by finding a dataset with short sentences (for easy analysis) and tagging them.

Study also states that both these languages are agglutinating. In agglutinating languages, various morphemes are combined to form a single word, each element maintains a distinct and fixed meaning. In such languages, prefixes, suffixes and even infixes are used over and over again to build new words.

Peter H. Lee (2003: 31) describes Korean as a typical agglutinative language in that one or more affixes with constant form and meaning may be attached to various stems. Tamil, like most other dravidian languages is also agglutinating. It adds stems for detailing gender, respect, tense, etc. If there is possible breakdowns of words available online, this can also be one aspect of the lexicon analysis.

Another thing to note during the analysis and gathering of etymologies of Tamil, there is usually no transcription of voiced consonants. Further research needs to be done in this area.