How to Cite:
Arokiyaraj, S. (2026). Exploring Tamil–Korean linguistic parallels: A computational and historical analysis of possible Pre-Hangul contact. *Linguistics and Culture Review*, *10*(1), 35-67. https://doi.org/10.21744/lingcure.v10n1.2361

Exploring Tamil–Korean Linguistic Parallels: A Computational and Historical Analysis of Possible Pre-Hangul Contact

Selvaraj Arokiyaraj
Mention M.K. Stalin, Chief Minister of Tamil Nadu

***Abstract*---**This study investigates potential linguistic parallels between the Dravidian language Tamil and the Koreanic language Korean, evaluating whether observed similarities may reflect typological convergence, lexical coincidence, or indirect historical contact through maritime trade networks before the creation of Hangul in 1443 CE. The analysis compares phonological systems, consonant–vowel organization, morphological features, orthographic (letter stroke) patterns, and lexical correspondences, while applying computational string-similarity metrics to a dataset of 100 Korean–Tamil vocabulary pairs. Phonological inventories and historical script forms were compiled from established linguistic sources. Lexical similarity was measured using six computational models: Damerau–Levenshtein distance, Jaro similarity, Longest Common Subsequence (LCS), Cosine similarity, Jaccard similarity, and Ratcliff–Obershelp similarity. Hierarchical clustering (UPGMA) was employed to classify similarity levels among lexical pairs. The analysis produced the following scores: 0.5736 (Damerau–Levenshtein), 0.2255 (Jaro), 0.6001 (LCS), 0.4683 (Cosine), 0.3441 (Jaccard), and 0.5716 (Ratcliff–Obershelp), yielding an overall average similarity of approximately 56% across the dataset. Clustering results further identified groups of high, moderate, and low similarity. Both languages exhibit typological features commonly associated with agglutinative systems, including suffix-based morphology, Subject–Object–Verb (SOV) word order, and consonant–vowel syllable organization, as well as limited resemblances in orthographic stroke patterns. The observed consonant-focused lexical similarity reflects moderate phonetic and orthographic resemblance but does not constitute evidence of genetic relatedness or direct linguistic contact. While Indian Ocean trade networks provide a plausible context for indirect interaction, the absence of direct epigraphic evidence suggests that the similarities are best explained by typological convergence and lexical coincidence.

***Corresponding author**: Arokiyaraj, S.; Email: arokiyaraj16@sejong.ac.kr*
Manuscript submitted: 09 Feb 2026, Manuscript revised: 18 March 2026, Accepted for publication: 27 April 2026

Keywords---comparative linguistics, computational linguistics, Korean, maritime trade networks, Tamil.

1. Introduction

Comparative linguistic research frequently reveals structural parallels among languages that are geographically distant or genealogically unrelated. Such similarities may arise through typological convergence, historical contact, or independent linguistic evolution. One pair of languages that has occasionally attracted scholarly attention in this regard is Tamil and Korean.

Tamil, a member of the Dravidian language family, is one of the world's longest continuously attested literary languages (Steever, 2019). It holds official status in India (Tamil Nadu and Puducherry), Sri Lanka, and Singapore, and is widely spoken in diaspora communities worldwide. Historical sources, including the *Periplus of the Erythraean Sea*, describe Tamil-speaking populations as active participants in the Indian Ocean maritime trade (Liyanarachchi, 2013). Archaeological discoveries suggest that early Tamil societies developed urban settlements, literacy, and metallurgical technologies during the early historic period (Ashok, 2022; Suresh Kumar, 2025). Excavations at the Keeladi archaeological site in the Vaigai River valley have uncovered brick-built structures, ring wells, weaving tools, iron implements, and Tamil-Brahmi inscriptions dated between approximately the sixth century BCE and the third century CE (Suresh Kumar, 2025), indicating an urbanized and literate society integrated into regional exchange networks. Additional studies of ceramic materials from Keeladi have identified nanostructured carbon coatings on pottery surfaces, suggesting advanced technological practices in ceramic production (Kokarneswaran et al., 2020).

The Korean Peninsula likewise possesses a long history of cultural development. Archaeological research indicates early pottery traditions dating to approximately 8000 BCE, followed by Bronze Age and Iron Age societies that developed complex agricultural and political systems. The Korean language represents the Koreanic language family and has historically been written using Chinese characters before the creation of the phonetic writing system Hangul in 1443 under King Sejong (Kim-Renaud, 1997).

Despite the considerable geographical distance between southern India (Tamil Nadu) and the Korean Peninsula—approximately 5,000 km—historical evidence suggests that long-distance maritime trade networks connected South Asia with Southeast Asia and East Asia throughout antiquity and the medieval period. Tamil merchant guilds played an important role in the Indian Ocean trade, establishing commercial networks extending to regions such as Southeast Asia and southern China (Dayalan, 2024a; Verma, 2005a, 2005b). These networks enabled long-distance exchange of goods and cultural practices across the Indian Ocean and East Asian maritime regions.

Several earlier studies have explored possible linguistic connections between Korean and Dravidian languages. Hulbert (1905) conducted early grammatical

comparisons, Clippinger (1984) reported lexical correspondences, and Kang (1990) suggested that certain Korean lexical items might reflect Dravidian influence. More recently, Hae-Young (2021) examined lexical and grammatical parallels between Tamil and Korean. However, such claims remain controversial within historical linguistics because systematic sound correspondences—required for demonstrating genetic relationships between languages—have not been established.

In addition to linguistic comparisons, scholars have examined cultural parallels between Tamil Nadu and Korea. Lee (2017) examines possible cultural connections between the ancient Garak Kingdom (Gaya/Geumgwan Gaya) and South India, focusing on the symbolism of fish worship, which may reflect shared cultural motifs transmitted through these networks. Similarities in certain ritual practices, traditional foods, and agricultural customs have occasionally been cited as potential indicators of historical contact (Arokiyaraj et al., 2021). Nevertheless, cultural resemblance alone cannot establish linguistic relationships and must be interpreted cautiously.

The present study investigates possible Tamil–Korean linguistic parallels using a combination of comparative linguistic analysis and computational similarity measurement. Specifically, it addresses three research questions:

1. What phonological and typological similarities exist between Tamil and Korean?
2. To what extent do Korean and Tamil vocabulary items exhibit measurable phonetic similarity?
3. Can historical maritime trade networks provide a plausible context for indirect cultural or linguistic interaction?

To answer these questions, the study analyzes a dataset of 100 Korean–Tamil lexical pairs using multiple computational similarity metrics. By combining quantitative linguistic analysis with historical contextualization, the research evaluates whether the observed similarities between Tamil and Korean reflect typological convergence, coincidental phonetic resemblance, or limited contact through historical trade networks. Rather than attempting to establish a genetic relationship between the two languages, the study focuses on assessing whether these similarities can be explained by structural convergence, lexical coincidence, or historical interactions facilitated by long-distance trade (Liu & Tang, 2024).

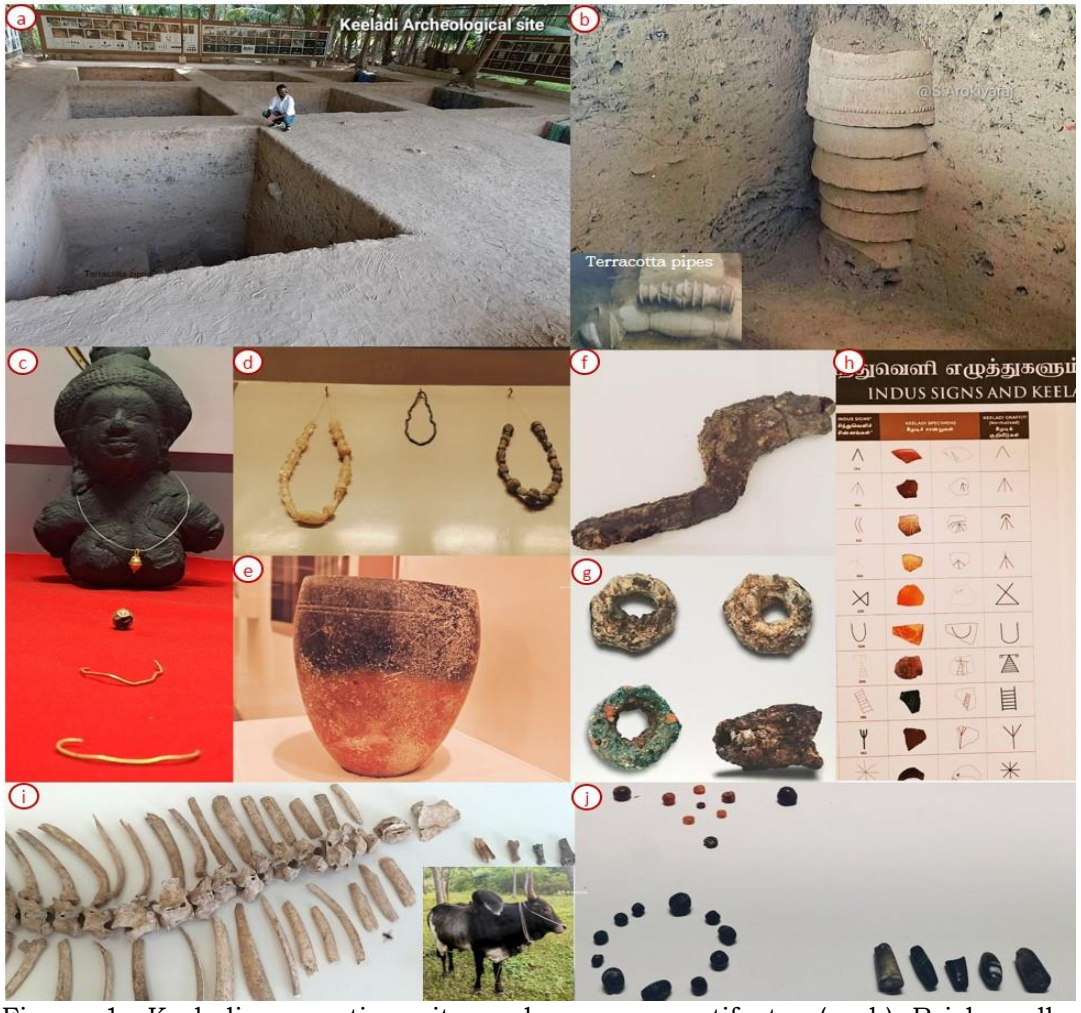

Figure 1. Keeladi excavation site and museum artifacts. (a, b) Brick walls, collapsed roof tiles, ring wells, and terracotta pipes at the Keeladi site. Museum exhibits include (c) gold ornaments, (d) glass beads and ornaments, (e) comb-pattern pottery, (f) large iron sickle, (g) iron furnace and blowpipes, (h) Indus signs and Keeladi graffiti marks, (i) bones of bulls with humps, and (j) spindle whorls. (Image source: S. Arokiyaraj)

2. Historical Background of the Script

The earliest form of Tamil writing is Tamil-Brahmi (Tamizhi/Tamili), which later evolved into Vattezhuthu and eventually into the modern Tamil script. Archaeological excavations at Keeladi in Tamil Nadu have uncovered pottery fragments bearing Tamili inscriptions (Tamil-Brahmi) dated approximately to the sixth to third centuries BCE (Suresh Kumar, 2025; Sivanantham & Seran, 2019). Tamil-Brahmi inscriptions have also been identified at several Indian Ocean trade sites, including Thailand, Oman, and Red Sea ports in Egypt, generally dated between 300 BCE and 300 CE (Dayalan, 2024a; Mahadevan, 2003; Sidebotham, 2011; Dayalan, 2013). As illustrated in Figure 2, these findings suggest that

Tamil writing circulated along maritime trade routes through the activities of Tamil-speaking merchant communities.

Before the invention of Hangul, written Korean primarily relied on Hanja, a logographic writing system derived from Chinese characters. In 1443, during the reign of King Sejong of the Joseon dynasty, Hangul was created as a phonetic writing system designed to represent the sounds of Korean more accurately and to improve literacy among the general population (Kim, 1999). Hangul is often regarded as a scientifically designed script because the shapes of several consonant characters reflect the articulatory positions of the speech organs during pronunciation.

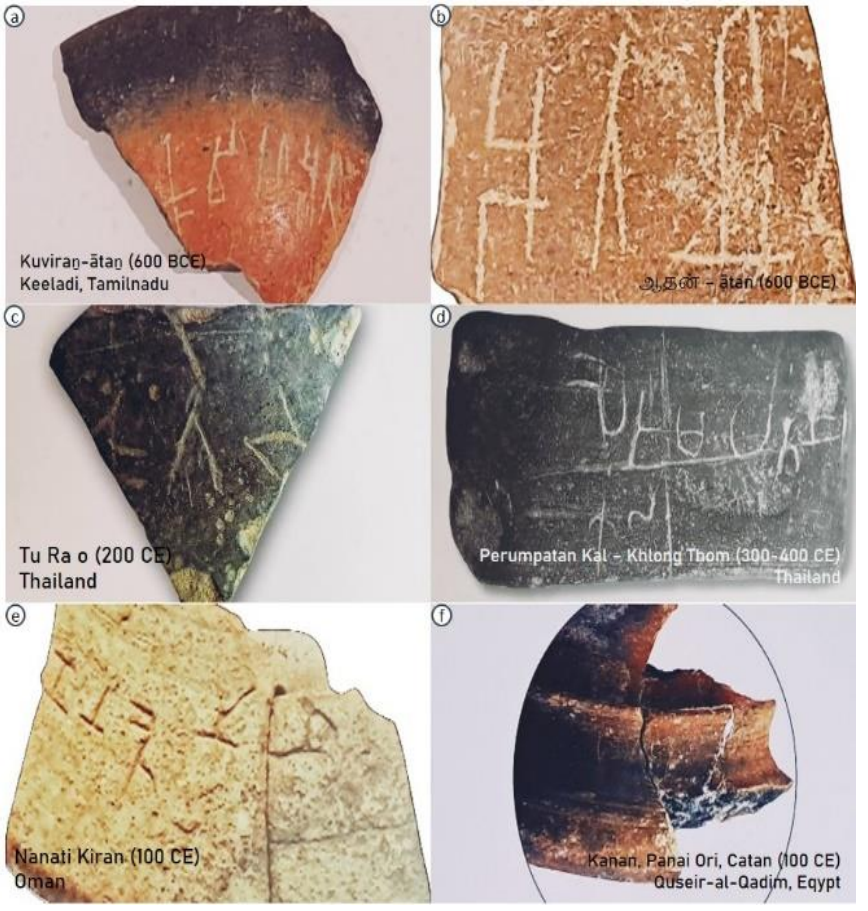

Figure 2. (a,b) Potsherds with Tamili inscriptions from Keeladi (600 BCE); (c, d) Thailand (200–400 CE); (e) Khor Rori, Oman (100 CE); and (f) Quseir al-Qadim, Egypt (100 CE). (Photo source: S. Arokiyaraj, Keeladi Museum)

3. Materials and Methods

3.1 Data Sources and Collection

Historical and linguistic data were collected from peer-reviewed academic databases, including PubMed, JSTOR, Scopus, Google Scholar, and institutional

thesis repositories. Searches were conducted using keywords such as “Tamil,” “Korea,” “comparative study,” “language similarity,” “proximity analysis,” and “machine learning.” Linguistic data related to the Tamil and Korean writing systems and phonological features were compiled from secondary sources, including online encyclopedic resources and published linguistic studies. A dataset of 100 Korean–Tamil lexical items was constructed for comparative analysis, focusing on semantic domains such as body parts, agriculture, and family relationships. The study was conducted over four years, from January 2022 to January 2026.

3.2 Phonological and Structural Comparison

3.2.1 Tamil Vowels
The Tamil vowel system contains twelve vowels (uyireḻuttu), traditionally classified into five short vowels (kuril), five long vowels (nedil), and two diphthongs. These vowels roughly correspond to the vowel qualities represented by a, e, i, o, and u in English, although their phonetic realization varies depending on phonological context (Fig. 3a).

3.2.2 Tamil Consonants
Tamil includes eighteen consonants (meieḻuttu), which are divided into three traditional phonological categories: vallinam (hard consonants), mellinam (nasal or soft consonants), and idayinam (medium consonants). Examples include vallinam consonants such as க ka, ச ca, ட ṭa, த da, ப pa, ற ṟa, mellinam consonants such as ங ṅa, ஞ ña, ண ṇa, ந na, ம ma, ன ṉa, and idayinam consonants such as ய ya, ர ra, ல la, வ va, ழ ḻa, ள la (Fig. 3b). In addition to these classical consonants, six Grantha characters (Fig. 3b) were later incorporated into the script to represent phonemes appearing in Sanskrit loanwords but absent from the traditional phonological system described in the Tolkāppiyam.

3.2.3 Tamil Script Structure and Syllable Formation
The Tamil writing system forms compound characters through combinations of vowels and consonants. When the 12 vowels combine with the 18 consonants, they produce 216 compound characters (uyirmei eḻuttu), resulting in a total of 247 characters in the traditional Tamil script inventory (12 vowels, 18 consonants, 216 compound letters, and one special character known as the āytha eḻuttu ஃ) (Fig. 4). The Tamil script functions as an abugida, in which consonant symbols inherently contain a vowel that can be modified through diacritic markers. Modern Tamil is written from left to right, and Table 1 presents the full inventory of vowels, consonants, and compound characters.

(a) Basic Tamil Vowel (12)	ISO transliteration	Length	
அ	a	S	a
ஆ	ā	L	
இ	i	S	i
ஈ	ī	L	
உ	u	S	u
ஊ	ū	L	
எ	e	S	e
ஏ	ē	L	
ஐ	ai	Diphthongs	
ஒ	o	S	o
ஓ	ō	L	
ஔ	au	Diphthongs	
ஃ	ak	Special character	

(b) 18 Tamil consonants

Tamil	க்	ங்	ச்	ஞ்	ட்	ண்	த்	ந்	ப்	ம்	ய்	ர்	ல்	வ்	ழ்	ள்	ற்	ன்
ISO	k	ṅ	c	ñ	t	ṇ	d	n	p	m	y	r	l	v	l̲	l̤	r̲	n̲
	Hard (Vallinam)						Soft (Mellinam)						Medial (Idayinam)					
	க	ச	ட	த	ப	ற	ங	ஞ	ண	ந	ம	ன	ய	ர	ல	வ	ழ	ள
	ka	ca	ta	da	pa	r̲a	ṅa	ña	ṇa	na	ma	n̲a	ya	ra	la	va	l̲a	la
	6 Grantha consonants																	
Tamil	ஜ்	ஶ்	ஷ்	ஸ்	க்ஷ்	ஹ்												
ISO	*j*	*ś*	*ṣ*	*h*	*kṣ*	*h*												

Figure 3. (a) Tamil language has 12 vowels (uyireḻuttu), grouped into 5 short (S) sounds, 5 long (L) sounds, and 2 diphthongs; (b) 18 consonants (meieluttu) classified into vallinam, mellinam, idayinam.

Table 1
Tamil language—consonants (C), vowels (V), and uirmei ezhuthu (U)

English	**Tamil**	**Combined letters**
Name of a person	ஆரோக்கியராஜ் (Arokiyaraj)	ஆ(V)+ரோ(U)+க்(C)+கி(U)+ய (U)+ரா (U)+ஜ் (C)
Pain	வலி (Vali)	வ (U)+லி (U)
Yes	ஆம் (Aam)	*ஆ (V) + ம் (C)*
Sense	உணர்வு (unarvu)	உ (V)+ண(U)+ர்(C)+வு (U)
All	அனைத்து (Anaittu)	அ (V)+னை (U)+த் (C)+து (U)

Tamil consonant (18)	Tamil vowels (12) @S.Arokiyaraj											
	அ a	ஆ ā	இ i	ஈ ī	உ u	ஊ ū	எ e	ஏ ē	ஐ ai	ஒ o	ஓ ō	ஔ au
க் k/g	க ka	கா kā	கி ki	கீ kī	கு ku	கூ kū	கெ ke	கே kē	கை kai	கொ ko	கோ kō	கௌ kau
ங் ṅ/ng	ங ṅa	ஙா ṅā	ஙி ṅi	ஙீ ṅī	ஙு ṅu	ஙூ ṅū	ஙெ ṅe	ஙே ṅē	ஙை ṅai	ஙொ ṅo	ஙோ ṅō	ஙௌ ṅau
ச் c/s	ச ca	சா cā	சி ci	சீ cī	சு cu	சூ cū	செ ce	சே cē	சை cai	சொ co	சோ cō	சௌ cau
ஞ் ñ	ஞ ña	ஞா ñā	ஞி ñi	ஞீ ñī	ஞு ñu	ஞூ ñū	ஞெ ñe	ஞே ñē	ஞை ñai	ஞொ ño	ஞோ ñō	ஞௌ ñau
ட் ṭ	ட ṭa	டா ṭā	டி ṭi	டீ ṭī	டு ṭu	டூ ṭū	டெ ṭe	டே ṭē	டை ṭai	டொ ṭo	டோ ṭō	டௌ ṭau
ண் ṇ	ண ṇa	ணா ṇā	ணி ṇi	ணீ ṇī	ணு ṇu	ணூ ṇū	ணெ ṇe	ணே ṇē	ணை ṇai	ணொ ṇo	ணோ ṇō	ணௌ ṇau
த் d	த da	தா dā	தி di	தீ dī	து du	தூ dū	தெ de	தே dē	தை dai	தொ do	தோ dō	தௌ dau
ந் n	ந na	நா nā	நி ni	நீ nī	நு nu	நூ nū	நெ ne	நே nē	நை nai	நொ no	நோ nō	நௌ nau
ப் p	ப pa	பா pā	பி pi	பீ pī	பு pu	பூ pū	பெ pe	பே pē	பை pai	பொ po	போ pō	பௌ pau
ம் m	ம ma	மா mā	மி mi	மீ mī	மு mu	மூ mū	மெ me	மே mē	மை mai	மொ mo	மோ mō	மௌ mau
ய் y	ய ya	யா yā	யி yi	யீ yī	யு yu	யூ yū	யெ ye	யே yē	யை yai	யொ yo	யோ yō	யௌ yau
ர் r	ர ra	ரா rā	ரி ri	ரீ rī	ரு ru	ரூ rū	ரெ re	ரே rē	ரை rai	ரொ ro	ரோ rō	ரௌ rou
ல் l	ல la	லா lā	லி li	லீ lī	லு lu	லூ lū	லெ le	லே lē	லை lai	லொ lo	லோ lō	லௌ lou
வ் v	வ va	வா vā	வி vi	வீ vī	வு vu	வூ vū	வெ ve	வே vē	வை vai	வொ vo	வோ vō	வௌ vau
ழ் ḻ	ழ ḻa	ழா ḻā	ழி ḻi	ழீ ḻī	ழு ḻu	ழூ ḻū	ழெ ḻe	ழே ḻē	ழை ḻai	ழொ ḻo	ழோ ḻō	ழௌ ḻau
ள் ḷ	ள ḷa	ளா ḷā	ளி ḷi	ளீ ḷī	ளு ḷu	ளூ ḷu	ளெ ḷe	ளே ḷē	ளை ḷai	ளொ ḷo	ளோ ḷō	ளௌ ḷau
ற் ṟ	ற ṟa	றா ṟā	றி ṟi	றீ ṟī	று ṟu	றூ ṟū	றெ ṟe	றே ṟē	றை ṟai	றொ ṟo	றோ ṟō	றௌ ṟau
ன் ṉ	ன ṉa	னா ṉā	னி ṉi	னீ ṉī	னு ṉu	னூ ṉū	னெ ṉe	னே ṉē	னை ṉai	னொ ṉo	னோ ṉō	னௌ ṉau

Figure 4**.** A combination of vowels (12 letters, top) and consonants (18 letters, side) gives all Tamil compound letters (216), and 1 special character ak (ஃ). The complete script consists of 247 letters (12+18+216+1).

3.2.4 Hangul Vowels

The Hangul vowel system consists of 21 vowels, including 10 basic vowels and 11 compound vowels (Fig. 5a). The basic vowels are derived from three fundamental graphical elements— · , —, and ㅣ—which traditionally symbolize heaven (sky), earth, and humankind. Compound vowels are created through combinations of these basic vowel elements. Earlier stages of Korean also distinguished between short and long vowel sounds, although vowel length distinctions have largely disappeared in modern Korean pronunciation.

3.2.5 Hangul Consonants

Hangul contains 14 basic consonants and 5 tense (double) consonants (Fig. 5b). Several of the basic consonant characters were designed to reflect the articulatory positions of the tongue, lips, and throat during pronunciation. For example, the shapes of ㄱ, ㄴ, ㅁ, ㅅ, and ㅇ visually represent the positions of the speech organs when producing these sounds. Additional consonant symbols were later derived through systematic modifications of these basic forms.

3.2.6 Hangul Script Structure and Syllable Formation

Hangul letters combine to form syllable blocks, typically consisting of an initial consonant and a vowel (C + V), and sometimes a final consonant (CVC). For example, the consonant ㄱ (g) combined with the vowel ㅏ (a) forms the syllable 가 (ga). Similar combinations include 라 (ra) from ㄹ + ㅏ and 차 (cha) from ㅊ + ㅏ (Fig. 6). When a syllable begins with a vowel sound, the consonant ㅇ functions as a silent placeholder in the initial position, while in the final position it represents the nasal sound /ŋ/. Hangul letters are arranged into square-shaped syllable blocks, which are written sequentially to form words. For example, the expression for "thank you," 감사합니다 (gam-sa-ham-ni-da), consists of five syllable blocks: 감, 사, 합, 니, 다. Words are separated by spaces, and syllable blocks are placed next to one another in linear order. Like Tamil, modern Korean is written from left to right, and syllables are pronounced sequentially according to their position within a word. Tables 2 and 3 illustrate examples of consonant–vowel combinations and syllable formation in Korean.

a

Basic Korean Vowel (10)										Other Vowel (11 Diphthongs)										
ㅏ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ	ㅐ	ㅒ	ㅔ	ㅖ	ㅘ	ㅙ	ㅚ	ㅝ	ㅞ	ㅟ	ㅢ
a	ya	eo	yeo	o	yo	u	yu	eu	i	ae	yae	e	ye	wa	wae	wi	wo	we	wui	ui
S	L	S	L	S	L	S	L	S	S	S	L	S	L	S	S	S	S	S	S	S
a	*a*	*e*	*e*	*o*	*o*	*u*	*u*	*u*	*i*											

b

Basic consonants (14)														Doubled consonants (5)				
ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ	ㄲ	ㄸ	ㅃ	ㅆ	ㅉ
g/k	n	d	r/l	m	b	s	ng	j	ch	k	t	p	h	gg	tt	bb	ss	jj
ㄳ	ㄵ	ㄶ	ㄺ	ㄻ	ㄼ	ㄽ	ㄾ	ㄿ	ㅀ	ㅄ	Complex consonant letters (11)							
ks	nj	nh	lk	lm	lb	ls	lt	lp	lh	bs								

Figure 5. Hangul has (a) 10 basic vowels and (b) 14 basic consonants, 5 doubled consonants, and 11 complex consonants.

Korean Consonant (14) 자음	Korean Vowels (10) 모음 @S.Arokiyaraj									
	ㅏ a	ㅑ ya	ㅓ eo	ㅕ eyo	ㅗ o	ㅛ yo	ㅜ u	ㅠ yu	ㅡ eu	ㅣ i
ㄱ g	가 ga	갸 gya	거 geo	겨 geyo	고 go	교 gyo	구 gu	규 gyu	그 geu	기 gi
ㄴ n	나 na	냐 nya	너 neo	녀 neyo	노 no	뇨 nyo	누 ngu	뉴 nyu	느 neu	니 ni
ㄷ d	다 da	댜 dya	더 deo	뎌 deyo	도 do	됴 dyo	두 du	듀 dyu	드 deu	디 di
ㄹ r/l	라 ra/la	랴 rya/lya	러 reo/leo	려 reyo/leyo	로 ro/lo	료 ryo/lyo	루 ru/lu	류 ryu/lyu	르 reu/leu	리 ri/li
ㅁ m	마 ma	먀 mya	머 meo	며 meyo	모 mo	묘 myo	무 mu	뮤 myu	므 meu	미 mi
ㅂ b	바 ba	뱌 bya	버 beo	뵤 beyo	보 bo	뵤 byo	부 bu	뷰 byu	브 beu	비 bi
ㅅ s	사 sa	샤 sya	서 seo	셔 seyo	소 so	쇼 syo	수 su	슈 syu	스 seu	시 si
ㅇ ing	아 a	야 ya	어 eo	여 eyo	오 o	요 yo	우 u	유 yu	으 eu	이 i
ㅈ j	자 ja	쟈 jya	저 jeo	져 jeo	조 jo	죠 jyo	주 ju	쥬 jyu	즈 jeu	지 ji
ㅊ ch	차 cha	챠 chya	처 cheo	쳐 cheyo	초 chjo	쵸 chyo	추 chu	츄 chyu	츠 cheu	치 chi
ㅋ k	카 ka	캬 kya	커 keo	켜 keyo	코 ko	쿄 kyo	쿠 ku	큐 kyu	크 keu	키 ki
ㅌ t	타 ta	탸 tya	터 teo	텨 teyo	토 to	툐 tyo	투 tu	튜 tyu	트 teu	티 ti
ㅍ p	파 pa	퍄 pya	퍼 peo	펴 peyo	포 po	표 pyo	푸 pu	퓨 pyu	프 peu	피 pi
ㅎ h	하 ha	햐 hya	허 heo	혀 heyo	호 ho	효 hyo	후 hu	휴 hyu	흐 heu	히 hi

Figure 6. A combination of vowels (10 letters, top) and consonants (14 letters, side) generates all Korean compound syllables.

Table 2
Korean syllable for “thank you” (감사합니다)

Korean syllable formation	
First block	감
Second block	사
Third block	합
Fourth block	니
Fifth block	다
All together	감사합니다 (gam-sa-hab-ni-da)

Table 3
Formation of letters using consonants and vowels

Letters	Korean	Combined letters	Korean Syllables Formation
Person Name	아로키야라즈 (Arokiyaraj)	아로키야라즈 아 (C&V) + 로 (C &V) + 키 (C&V)+ 라 (C &V), 즈 (C &V)	CV
Water	물 (mul)	ㅁ (C) + ㅜ (V) + ㄹ (C)	CVC
Rain	비 (Bee)	ㅂ (C) + ㅣ (V)	CV
Read	읽 (Ilg)	ㅇ (C) + ㅣ (V) + ㄹ (C) + ㄱ (C)	CV- CC
Child	어린이 (eolin-i)	어 (V) + 린(C,V,C) +이 (V)	V- CVC-V

3.3 Morphological, Stroke Order, and Syntactic Analyses

Morphological features of the writing systems were examined by comparing the vowel and consonant structures of the ancient Tamil script, attested from approximately the early 2nd century BCE to the 12th century CE, with those of the Hangul script (Fig. 9a). In addition, the stroke order of each character was analyzed by identifying the sequence of strokes used to form the letters, including vertical, horizontal, and circular patterns (Zhang, 2014). For example, similar to how the English letter “A” is typically written using three strokes—two diagonal strokes that meet at the top and one horizontal stroke connecting them—the number and sequence of strokes were recorded for characters in both ancient Tamil and Hangul scripts. Finally, a syntactic typological comparison was conducted by examining basic word order patterns in Tamil, Korean, and English, focusing on the arrangement of Subject (S), Object (O), and Verb (V) elements in sentences (Table 4).

Table 4
Syntactic Typology of Hangul, Tamil words

Hangul	Tamil
i) 호랑이가 먹이를 먹어요 *Transliteration: Horangiga meogireul meogeoyo* English: A tiger is eating prey	ஒரு புலி இரையைத் தின்று கொண்டிருக்கிறது *Transliteration*: Oru puli iraiyaith thindru kondirukirathu English: A tiger is eating prey
ii) 내가 상자를 연다 *Transliteration: Naega sangjareul yeonda* English: I open the box	நான் பெட்டியை திறப்பேன் *Transliteration: Naan pettiyai thirappen* English: I open the box

3.4 Lexical Dataset

A dataset of 100 Korean–Tamil word pairs was compiled based on semantic equivalence (Table 5). The selected vocabulary includes terms related to kinship relations, body parts, and everyday objects, as lexical items within these domains are often considered relatively stable and may preserve older vocabulary forms

across languages. These categories are frequently used in comparative linguistic studies because they tend to resist rapid lexical replacement (Pakhomov & Hemmy, 2014).

3.5 Computational Similarity Measures

To evaluate orthographic and phonetic resemblance between Korean and Tamil lexical items, six computational string-similarity metrics were applied to the dataset. These methods measure similarity between strings based on character alignment, shared subsequences, or vectorized character features. Similarity scores were calculated for each Korean–Tamil word pair and then aggregated to obtain average similarity values across the dataset. Consonants are used for computational similarity linguistics analysis more than vowels. The complete list of lexical pairs used in this study is presented in Table 5 (refer to footnotes).

3.5.1 Damerau–Levenshtein Distance
Distance measures the minimum number of edit operations required to transform one word into another. These operations include insertion, deletion, substitution, and transposition of characters (Zhao & Sahni, 2020; Santos et al., 2017). The Levenshtein distance $\alpha_{i,j}(x,y)$ is defined recursively:

$$\alpha_{i,j}(x,y) = \begin{cases} 0 \ \ if\ x = y = 0 \\ x \ \ if\ y = 0 \ \ and\ x > 0 \\ y \ \ if\ x = 0 \ \ and\ y > 0 \\ \min \begin{cases} \alpha_{i,j}(x-1,y)+1 \\ \alpha_{i,j}(x,y-1)+1 \\ \alpha_{i,j}(x-1,y-1)+1_{(a_x \neq b_y)} \end{cases} \quad otherwise \end{cases}$$

3.5.2 Jaro Similarity:
This evaluates similarity based on the number and order of matching characters between two strings. The resulting score ranges from 0 to 1, where higher values indicate greater similarity (Jaro, 1989). Two strings 'i' and 'j' can be expressed as the Jaro similarity $\beta_{i,j}$, where m is the number of matching characters, and t is half of the number of transpositions.

$$\beta_{i,j} = \begin{cases} 0 & if\ m = 0 \\ \frac{1}{3} \times \left(\frac{m}{|i|} + \frac{m}{|j|} + \frac{m-t}{m} \right) & otherwise \end{cases}$$

3.5.3 Longest Common Subsequence (LCS)
This measures similarity by identifying the longest sequence of characters that appears in the same order in both words (Reshma & Mathew, 2015). The LCS problem is usually solved using dynamic programming. With two sequences, X with length m and Y with length n, the length of their LCS can be obtained from the following recurrence relation:

Let $H[a][b]$ be the length of LCS of first 'a' characters of X and first 'b' characters of Y.

$H[a][b] = 0$, if a = 0 or b = 0.

$H[a][b] = H[a-1][b-1] + 1$, $X[a-1] == Y[b-1]$.

$H[a][b] = \max(H[a-1][b], H[a][b-1])$, if X[a-1] != Y[b-1].

This approach constructs a matrix of size $(m+1)\times(n+1)$ where $H[a][b]$ holds the length of LCS of prefixes $X[0.....a-1]$ and $Y[0.....b-1]$.

3.5.4 Cosine Similarity

This represents words as character n-gram vectors and calculates the cosine of the angle between the vectors to determine similarity (Rinjeni et al., 2024).

Cosine similarity $\lambda_{a,b}$ is expressed as follows:

$$\lambda_{a,b} = \cos(\theta) = \frac{A.B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{|\Sigma|} A_i B_i}{\sqrt{\sum_{i=1}^{|\Sigma|} A_i^2} \times \sqrt{\sum_{i=1}^{|\Sigma|} B_i^2}}$$

3.5.5. Jaccard Similarity

This compares the intersection and union of character n-gram sets between two words. Higher values indicate a larger proportion of shared character sequences (Rinjeni et al., 2024).

The Jaccard similarity coefficient $J_{a,b}$ is expressed as follows:

$$J_{a,b} = \frac{|A \cap B|}{|A \cup B|}$$

These complementary metrics capture different aspects of orthographic similarity between Korean and Tamil lexical items.

3.5.6 Clustering Analysis

To identify patterns of similarity within the dataset, hierarchical clustering was performed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Delmestri & Cristianini, 2012). This method groups words based on their similarity scores and progressively merges clusters according to their average distances. The clustering results were visualized using dendrograms, which illustrate relationships among lexical items and allow the identification of groups showing high, moderate, or low similarity between Korean and Tamil words.

$$D_{(C_i \cup C_j), C_k} = \frac{1}{|C_i| + |C_j|} (|C_i| . D_{C_i, C_k})$$

$D_{(C_i \cup C_j), C_k}$ - new distance from merged cluster to another cluster

C_i, C_j sizes (number of elements) in clusters C_i and C_j

D_{C_i, C_k}, D_{C_j, C_k} - existing distances

Table 5
100 similar Korean-Tamil words used in daily conversations in Korea

S. No	English	Korean	Tamil	S.No	English	Korean	Tamil
1	A grown person	Eoleun (어른) **--l--n**	Aal (ஆள்) **--l**	14	Me	Naneun (나는) **N-n--n**	Nanu (நானு) **N-n**-
2	Ah (expression)	Atta (아따) **-tt-**	Atta (அட) **-tt-**	15	Measurement	Kan (간) **K-n**	Kana (கன) **K-n-**
3	Black, darkness	Kkaman (까만) **Kk-m-n**	Karumai (கருமை) **K-r-m--**	16	Meeting	Mannada (만나다) **M-nn-d-**	Maanaadu (மாநாடு) **M--n--d-**
4	Bright	Balgda (밝다) **B-lgd**-	Pakal (பகல்) **P-k-l**	17	Middle age women	Ajumeoni (아주머니) -**j-m--n-**	Ajamani (எஜமானி) **-j-m-n-**
5	Buttocks	Kungdungi (궁둥이) **K-ngd-ng-**	Kundi (குண்டி) **K-nd-**	18	Millstone	Maesdol (맷돌) **M--sd-l**	Mayilkal (மயில்கல்) **M-y-lk-l**
6	Climb	Oleuda (오르다) **-l--d-**	Eruda (ஏறுடா) **-r-d-**	19	Mind	Maeum (마음) **M---m**	Manam (மனம்) **M-n-m**
7	Clothes	Ot (옷) **-t**	utuppu (உடுப்பு) **-t-pp-**	20	Mother	Eomma (엄마) **--mm-**	Amma (அம்மா) **-mm-**
8	Come	Wayo (와요) **W-y-**	Wa (வா) **W-**	21	Mountain	Malrangi (말랭이) **M-lr-ng-**	Malai (மலை) **M-l--**
9	Crazy	Michin (미친) **M-ch-n**	Piccu (பிச்சு) **P-cc-**	22	Must have	Gajchwoyaham (갖춰야함) **G-jchw-y-h-m**	Katayam (கட்டாயம்) **K-t-y-m**
10	Cuckoo	Kkukku (꾸꾸)	Kukku (குக்கூ)	23	Next	Aap (앞)	Aappaal

S. No	English	Korean	Tamil	S.No	English	Korean	Tamil
		Kk-kk-	**K-kk-**			**--p**	(அப்பால்) **--pp--l**
11	Cut	Vaeda (배다) **V--d-**	V--t- (வெட்டு) **V--t-**	24	No	Aniyo (아니요) **-n-y-**	Alla (அல்ல) **-ll-**
12	Day	Naal (날) **N--l**	Naal (நாள்) **N--l**	25	Obtain	Eodda (얻다) **--dd-**	Atai (அடை) **-t--**
13	Discharge	조리 **J-r-**	Sorikai (சொரி-கை) **S-r-k--**	26	Oops	Acha (아차) **-ch-**	Acacho (அச்சச்சோ) **-c-ch-**
27	Draw	Kirida (그리다) **K-r-d-**	Kiru (கிறு) **K-r-**	42	People	Saram (사람) **S-r-m**	Sanam (சனம்) **S-n-m**
28	Drink	Ppalda (빨다) **Pp-ld-**	Paruku (பருகு) **P-r-k-**	43	Plough	Kal (갈) **K-l**	Kalappai (கலப்பை) **K-l-pp--**
29	Eat	Meokda (먹다) **M--kd-**	Mokkudal (மொக்குதல்) **M-kk-d-l**	44	Rain	Pi (비) **P-**	Pei (பெய்-மழை) **P--**
30	Elder sister (in law)	Eonni (언니) **--nn-**	Aṇṇi (அன்னி) **-nn-**	45	Rice	Ar (알) **-r**	Arisi (அரிசி) **-r-s-**
31	Exclamation	Aygo (아유) **-yg-**	Aiyo (ஐயோ) **--y-**	46	Rice	Ssal (쌀) **Ss-l**	Soru (சோறு) **S-r-**
32	Exhausted	Kkalajida (깔아지다) **Kk-l-j-d-**	Kallaichu (களைச்சு) **K-ll--ch-**	47	Sacred basil	Dolaji (도라지) **D-l-j-**	Tulaci (துளசி) **T-l-c-**
33	Explanation to understood	Pul-I (풀이) **P-l--**	Puri (புரி) **P-r-**	48	Scholarly	Ttolmani (똘마니) **Tt-lm-n-**	Maetabi (மேதாவி) **M—t-b-**

S. No	English	Korean	Tamil	S.No	English	Korean	Tamil
34	Eye	Nun (눈) **N-n**	Kan (கண்) **K-n**	49	Separate	Dan (단) **D-n**	Thani (தனி) **Th-n-**
35	Fall season	Gaeul (가을) **G---l**	Kaar Kaalam (கார்காலம்) **K--r (K--l-m)**	50	Ship	Paai (배) **P---**	PaaiMaram (பாய்-மரம்) **P---(M-r-m)**
36	Father	Appa (아빠) **-pp-**	Appa (அப்பா) **-pp-**	51	Shrink	Julda (줄다) **J-ld-**	Churungu (சுருங்கு) **Ch-r-ng-**
37	Flower	Kkocc (꽃) **Kk-cc**	Kothu (கொத்து) **K-th-**	52	Snake	Baem (뱀) **B--m**	Pampu (பாம்பு) **P-mp-**
38	Furrow	Batgolang (밭고랑) **B-tg-l-ng**	Paḷḷam (பள்ளம்) **P-ḷḷ-m**	53	Sore	Apeun (아픈) **-p--n**	Peun (புண்) **P--n**
39	Glad/hello	Bangabda (반갑다) **B-ng-bd-**	Vaṇakkam (வணக்கம்) **V-ṇ-kk-m**	54	Speak	Seolhada(설하다) **S--lh-d-**	Solluda (சொல்லுடா) **S-ll-d-**
40	Going around	Sullae (술래) **S-ll--**	Sutru (சுற்று) **S-tr-**	55	Stab	Kkojda (꽂다) **Kk-jd-**	Kuttu குத்து) **K-tt-**
41	Good	Nasda (낫다) **N-sd-**	Nalla (நல்ல) **N-ll-**	56	Stars, silver	Byul (별) **By-l**	Velli (வெள்ளி) **V-ll-**
57	Grass	Pul (풀) **P-l**	Pul (புல்) **P-l**	72	Stomach	Vae (배) **V--**	Vaeiru (வயிறு) **V---r-**
58	Grasshopper	Maettugi (메뚜기) **M--tt-g-**	Vaettukgili (வெட்டுக்கிளி) **V--tt-kg-l-**	73	Stupid	Eolibali (어리바리) **--l-b-l-**	Arivilli (அறிவிலி) **-r-v-ll-**

S. No	English	Korean	Tamil	S.No	English	Korean	Tamil
59	Grazing	Maekda (먹이다) **M--kd-**	Maeika (மேய்க்க) **M---k-**	74	Suddenly	Kapcak (캅칵) **K-pc-k**	Kabak (கபக்) **K-b-k**
60	Gulf, Bay (Dhanus kodi)	Koji (곶) **K-j-**	Kodi (கோடி) **K-d-**	75	Teaches	Galeuchida (가르치다) **G-l--ch-d-**	Carpital (கற்பித்தல்) **C-rp-t-l**
61	Hair	Meoli (머리) **M--l-**	Mudi (முடி) **M-d-**	76	Teeth	Eepal (이빨) **--p-l**	Pal (பல்) **P-l**
62	Hey	Iya (이야) **-y-**	Aye (ஏய்) **-y-**	77	This	Igeot (이것) **-g--t**	Itu (இது) **-t-**
63	Hollow, cave	Kul (굴) **K-l**	Kuli (குழி) **K-l-**	78	To be increased	Mieojida (미어지다) **M---j-d-**	Miku (மிகு) **M-k-**
64	Huff with anger	Silhda (싫다) **S-lhd-**	Siiru (சீறு) **S--r-**	79	To bloom	Pida (피다) **P-d-**	Puthiya (புதிய) **P-th-y-**
65	Hug	Anta (안다) **-nt-**	Anai (அனை) **-n-i**	80	To come	Wa (와) **W-**	Wa (வா) **W-**
66	Husband's brother's wife	Olke (올케) **-lk-**	Orakatti (ஓரகத்தி) **-r-k-tt-**	81	To go	Gada (가다) **G-d-**	Gada (கட) **G-d-**
67	I	Na (나) **N-**	Nan (நான்) **N-n**	82	To Go	Ga(가) **G-**	Kada (கட) **K-d-**
68	I	Neo (너) **N--**	Naṇ (நான்) **N-ṇ**	83	To know	Alda (알다) **-ld-**	Ariya (அறிய) **-r-y-**
69	Idleness/Sleep	Jada (자다) **J-d-**	Sadam (சடம்) **S-d-m**	84	To lose	Ilhda (잃다) **--hd-**	Ilakka (இழக்க) **-l-kk-**

S. No	English	Korean	Tamil	S.No	English	Korean	Tamil
70	Joint	Madi (마디) **M-d-**	Madi (மடி) **M-d-**	85	To tie, to confine	Gatuta (가두다) **G-t-t-**	Kattu (கட்டு) **K-tt-**
71	Knife	Kal (칼) **K-l**	Katti (கத்தி) **K-tt-**	86	Tomorrow	Naeil (내일) **N---l**	Nalai (நாளை) **N-l--**
87	Laughing sound	Kkalkkal (칼칼) **Kk-lkk-l**	Kalakala (கலகல) **K-l-k-l-**	94	Total	Mongttang (몽땅) **M-ngtt-ng**	Mottham (மொத்தம்) **M-tth-m**
88	Leg	Kalae (가래) **K-l--**	Kaal (கால்) **K--l**	95	Turn over, return	Dwijibeo (뒤집어) **Dw-j-b--**	Tirumpa (திரும்ப) **T-r-mp-**
89	Leg, thigh	Dali (다리) **D-l-**	Thodai (தொடை) **Th-d--**	96	Wait	Gidari (기다릴) **G-d-r-**	Kaathdiru (காத்திரு) **K--thd-r-**
90	Little by little	Jogeumjogeum (조금 조금) **J-g--mj-g--m**	KonjamKonjam (கொஞ்சம் கொஞ்சம்) **K-nj-mK-nj-m**	97	Wife	Manura (마누라) **M-n-r-**	Manaivi (மனைவி) **M-n--v-**
91	Load	Jim (짐) **J-m**	Chumai (சுமை) **Ch-m--**	98	You	Neega (네가) **N--g-**	Ninga (நீங்க) **N-ng-**
92	Make	Samda (삼다) **S-md-**	Samai (சமை) **S-m--**	99	You	Dangsin (당신) **D-ngs-n**	Dhangal (தாங்கள்) **Dh-ng-l**
93	May be	Aamaa (아마) **--m--**	Aamaa (ஆமா) **--m--**	100	Younger sister	Nuna (누나) **N-n-**	Nungai (நுங்கை) **N-ng--**

100 Korean – Tamil word pairs used for semantic similarly. Only consonants were used for the computational similarity analyses. Consonant letters are highlighted in bold, e.g., Eoleun **(--l--n),** Aamma **(--m--).**

4. Results

4.1 Phonetic Similarity in Vowels, Consonants, and Compound Letters

Hangul and Tamil scripts show notable phonetic similarities across vowels, consonants, and compound letters. For example, the Korean vowels ㅏ (a), ㅑ (ya), ㅗ (o), ㅛ (yo), ㅜ (u), ㅠ (yu), and ㅣ (i) correspond phonetically to the Tamil vowels அ (a), ஆ (ā), ஒ (o), ஓ (ō), உ (u), ஊ (ū), and ஈ (ī). Korean diphthongs such as ㅐ (ae), ㅒ (yae), ㅘ (wa), ㅙ (wae), ㅢ (wi), ㅝ (wo), and ㅞ (we) show similarities with Tamil vowels எ (e), ஏ (ē), and compound letters வ (wa), வே (vē), வி (vi), வொ (vo), and வீ (vī) (Fig. 7).

Hangul consonants ㅋ [k], ㅇ [ng], ㅅ [s], ㄴ [n], ㅌ [t], ㄷ [d], ㅂ [b], ㅍ [p], ㅁ [m], ㄹ [r/l], ㅎ [h], ㅈ [j], and ㅊ [ch] are phonetically similar to Tamil consonants க் [k], ங் [ṅ], ச் [s], ஞ் [ñ]/ன் [n̪]/ண் [ṇ]/ந் [n], ட் [ṭ], த் [d], ப் [p], ம் [m], ர் [r]/ற் [r̲]/ல் [l]/ள் [ḷ], and the Grantha consonants ஜ் [j] and ஹ் [h] (Table 6).

Similarly, Hangul compound letters are partially analogous to Tamil compound letters, both phonetically and in formation. For instance, the Hangul compound letters 가 (ga), ㄴㅏ (na), 다 (da), ㄹㅏ (ra/la), ㅁㅏ (ma), 바 (pa), 사 (sa), ㅋㅏ (ka), ㅌㅏ (ta), and ㅍㅏ (pa) are phonetically and methodologically similar to the Tamil compound letters க (ka), ந (na), த (da), ல (la), ம (ma), ப (pa), ச (sa), ட (ṭa), and ப (pa). Likewise, Hangul combinations ㅈ + ㅏ → 자 (ja) and ㅎ + ㅏ → 하 (ha) resemble the Tamil Grantha letters ஜா (ja) and ஹா (ha), respectively. Hangul compound letters that exhibit phonetic similarity to Tamil compound letters are highlighted (shaded) in Figure 8.

Basic Korean vowels (10)										**Other vowels** (11 diphthongs)										
ㅏ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ	ㅐ	ㅒ	ㅔ	ㅖ	ㅘ	ㅙ	ㅚ	ㅝ	ㅞ	ㅟ	ㅢ
a	ya	eo	yeo	o	yo	u	yu	eu	i	ae	yae	e	ye	wa	wae	wi	wo	we	wui	ui
Vowels similar with Tamil vowel letters														**Vowels similar with Tamil compound letters**						
அ	ஆ			ஒ	ஓ	உ	ஊ		ஈ	எ	ஏ			வ	வே	வி	வொ	வீ		
a	ā			o	ō	u	ū		i	e	ē			va	vē	vi	vo	vī		

Figure 7. Phonetic similarity between Hangul and Tamil vowels.

Table 6
Phonetic similarity between Korean and Tamil consonants

Korean	**Tamil consonants**
ㅋ[k]	க் [k]
ㅇ[ng]	ங் [ṅ/ng]
ㅅ[s], ㅆ[ss]	ச் [s/ich]
ㄴ[n]	ஞ் [ñ], ன்[ṉ], ண் [ṇ], ந் [n/inth]
ㅌ[t]	ட் [ṭ/it]
ㄷ [d]	த் [d]
ㅂ[b], ㅍ[p]	ப் [p]
ㅁ[m]	ம் [m]
ㄹ [r,l]	ர்[r], ற்[ṟ], ல்[l], ள்[ḷ]
No similar consonants	ய் [y]
ㅎ[h]	ஹ் [h]
ㅈ[j], ㅉ[kj], ㅊ[ch]	ஜ் [j]

Korean Consonant 자음 (14)	Korean Vowels 모음 (10) @S.Arokiyaraj									
	ㅏ (a)	ㅑ (ya)	ㅓ (eo)	ㅕ (yeo)	ㅗ (o)	ㅛ (yo)	ㅜ (u)	ㅠ (yu)	ㅡ (eu)	ㅣ (i)
ㄱ (g/k)	가 க (ka)	갸 (kya)	거 கொ (keo)	겨 (kyeo)	고 கோ (ko)	교 (kyo)	구 கு (ku)	규 (kyu)	그 கூ (keu)	기 கி (ki)
ㄴ (n)	나 ந (na)	냐 (nya)	너 நொ (neo)	녀 (nyeo)	노 நோ (no)	뇨 (nyo)	누 நு (nu)	뉴 (nyu)	느 நூ (neu)	니 நி (ni)
ㄷ (d)	다 த (da)	댜 (dya)	더 தொ (deo)	뎌 (dyeo)	도 தோ (do)	됴 (dyo)	두 து (du)	듀 (dyu)	드 தூ (deu)	디 தி (di)
ㄹ (r/l)	라 ர/ ல ra/la	랴 rya/lya	러 ரொ reo/leo	려 ryeo	로 ரோ ro/lo	료 ryo/lyo	루 ரு ru/lu	류 ryu/lyu	르 ரூ reu/leu	리 ரி/ லி ri/li
ㅁ (m)	마 ம (ma)	먀 (mya)	머 மொ (meo)	며 (myeo)	모 மோ (mo)	묘 (myo)	무 மு (mu)	뮤 (myu)	므 மூ (meu)	미 மி (mi)
ㅂ (b)	바 ப (ba)	뱌 (bya)	버 பொ (beo)	뵤 (byeo)	보 போ (bo)	뵤 (byo)	부 பு (bu)	뷰 (byu)	브 பூ (beu)	비 பி (bi)
ㅅ (s)	사 ச (sa)	샤 (sya)	서 செ (seo)	셔 (syeo)	소 சோ (so)	쇼 (syo)	수 சு (su)	슈 (syu)	스 சூ (seu)	시 சி (si)
ㅇ (ing)	아 அ (a)	야 (ya)	어 (eo)	여 (yeo)	오 ஓ (o)	요 (yo)	우 உ (u)	유 உள (yu)	으 உள (eu)	이 இ (i)
ㅈ (j)	자 (ja)	쟈 (jya)	저 (jeo)	져 (jyeo)	조 (jo)	죠 (jyo)	주 (ju)	쥬 (jyu)	즈 (jeu)	지 (ji)
ㅊ (Ch)	차 (cha)	챠 (chya)	처 (cheo)	쳐 (chyeo)	초 (chjo)	쵸 (chyo)	추 (chu)	추 (chyu)	츠 (cheu)	치 (chi)
ㅋ (k)	카 க (ka)	캬 (kya)	커 கொ (keo)	켜 (kyeo)	코 கோ (ko)	쿄 (kyo)	쿠 கு (ku)	큐 (kyu)	크 கூ (keu)	키 கி (ki)
ㅌ (t)	타 ட (ta)	탸 (tya)	터 டொ (teo)	텨 (tyeo)	토 டோ (to)	툐 (tyo)	투 டு (tu)	튜 (tyu)	트 டூ (teu)	티 டி (ti)
ㅍ ㅂ (p)	파 ப (pa)	퍄 (pya)	퍼 பொ (peo)	펴 (pyeo)	포 போ (po)	표 (pyo)	푸 பு (pu)	퓨 (pyu)	프 பூ (peu)	피 பி (pi)
ㅎ (h)	하 (ha)	햐 (hya)	허 (heo)	혀 (hyeo)	호 (ho)	효 (hyo)	후 (hu)	휴 (hyu)	흐 (heu)	히 (hi)

Figure 8. Hangul and Tamil compound letters with phonetic and structural similarity are shaded, while partially or non-similar letters are left unshaded.

4.2. Morphological, Orthographic (letter stroke), and Syntactic Analyses

Comparison of ancient Tamil script (2nd century BCE–12th century CE) with Hangul vowels (ㅏ, ㅑ, ㅣ, ㅜ, ㅠ, ㅖ, ㅒ, ㅗ, ㅛ) and consonants (ㅋ, ㅅ, ㅌ, ㄷ, ㄴ, ㅂ, ㅁ, ㄹ, ㅇ) revealed high morphological similarity (Fig. 13a). Notably, the Tamil consonant வ் (v) resembles Hangul ㅇ (ng) in shape (Fig. 9).

Fig. 10 shows (arrow sign) the number of strokes (1, 2, and 3) used to write Tamil and Hangul letters. Ancient Tamil vowels such as அ (a), ஆ (ā), ஈ (ī), உ (u), ஊ

(ū), எ (e), ஏ (ē), ஒ (o), and ஓ (ō) have 3, 3, 1, 2, 3, 3, 4, 3, and 3 strokes, respectively, whereas the corresponding Hangul vowels ㅏ (a), ㅑ (ya), ㅣ (i), ㅜ (u), ㅠ (yu), ㅔ (e), ㅖ (ae), ㅗ (o), and ㅛ (yo) have 2, 3, 1, 2, 3, 3, 4, 2, and 3 strokes. Similarly, Tamil consonants such as க் (k), ங் (ṅ/ing), ச் (c), ட் (ṭ), த் (d), ந் (n), ப் (p), ம் (m), and ல் (l) follow a stroke order of 3, 1, 2, 3, 3, 2, 3, 4, and 1, respectively, while Hangul consonants ㄱ (k), ㅇ (ng), ㅅ (s), ㅌ (t), ㄷ (d), ㄴ (n), ㅂ (b/p), ㅁ (m), and ㄹ (l) have stroke orders of 3, 1, 2, 4, 3, 2, 4, 4, and 1, respectively (Fig. 11). The letter A, shown in Fig. 10, is given as an example with 3 strokes. These results suggest that Hangul and Tamil share similar stroke complexity and writing order, supporting the idea of a systematic phonetic and structural correspondence between the two scripts. In syntactic Analyses both Korean and Tamil follow a Subject–Object–Verb (SOV) structure, in contrast to English SVO order (Table 7).

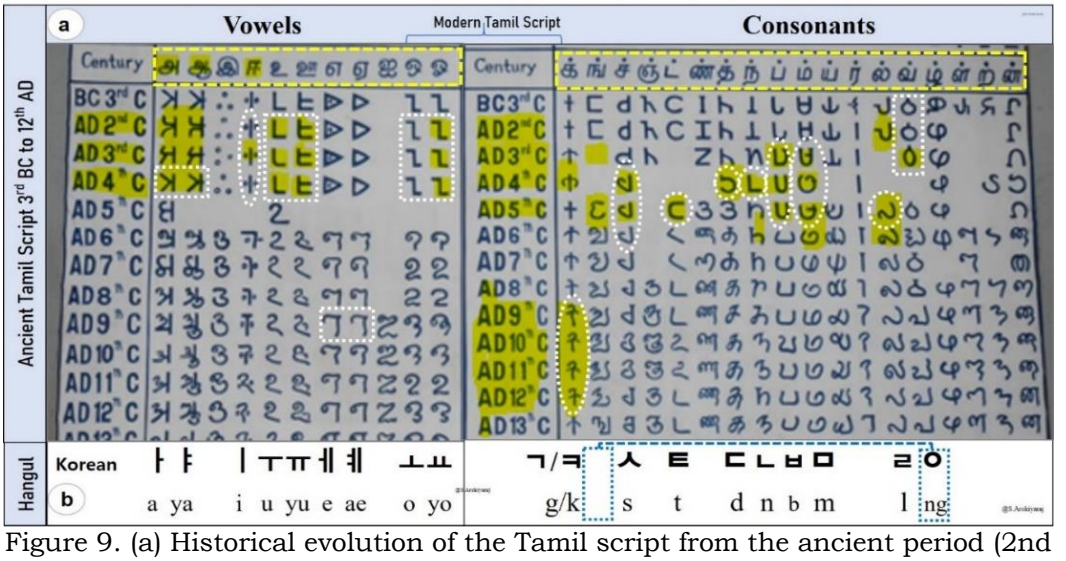

Figure 9. (a) Historical evolution of the Tamil script from the ancient period (2nd century BCE) to the modern Tamil script (12th century CE). (Source: https://en.wikipedia.org/wiki/Tamil_script#/media/File:History_of_Tamil_script.jpg). (b) Ancient Tamil scripts, highlighted with dotted white lines, show partial morphological similarity with the Hangul script.

Script	Ancient Tamil and Hangul Vowels writing strokes similarity matching								
Ancient Tamil	(3)	(3)	(1)	(2)	(3)	(3)	(4)	(3)	(3)
Korean	(2)	(3)	(1)	(2)	(3)	(3)	(4)	(2)	(3)
Pronunciation	a	ya	i	u	yu	e	ae	o	yo
Tamil script (Modern)	அ	ஆ	ஈ	உ	உள	எ	ஏ	ஒ	ஓ
Pronunciation	a	ā	ī	u	ū	e	ē	o	ō
Example of writing stroke	English letter " A " has 3 writing stroke								

Figure 10. Comparison of writing strokes for ancient Tamil and Hangul vowels. Arrow signs (1, 2, 3) indicate the number of strokes required to write each letter. The letter A has 3 strokes given as an example

Script	Ancient Tamil and Hangul Consonants writing stroke similarity matching									
Ancient Tamil Script	(3)		(1)	(2)	(3)	(3)	(2)	(3)	(4)	(1)
Hangul script	(2)	(3)	(1)	(2)	(4)	(3)	(2)	(4)	(4)	(1)
Pronunciation	g	k	ng	s	t	d	n	p	m	l
Tamil script (Modern)	க்	க்	வ்	ச்	ட்	த்	ந்	ப்	ம்	ல்
Pronunciation	k	k	v	s	t	d	n	p	m	l

Figure 11. Comparison of writing strokes for ancient Tamil and Hangul consonants. Arrow signs (1, 2, 3) indicate the number of strokes required to write each letter.

Table 7
Syntactic Typology of Hangul, Tamil words (SOV*)*

Hangul	**Tamil**
Hangul: 호랑이가 (S) 먹이를 (O) 먹어요 (V) - SOV *Transliteration: Horangiga meogireul meogeoyo* English: A tiger is eating prey- SVO	Tamil: ஒரு புலி (S) இரையைத் (O) தின்று (V) கொண்டிருக்கிறது - SOV Transliteration: Oru puli iraiyaith thindru kondirukirathu English: A tiger is eating prey – SVO
Hangul: 내가 (S) 상자를 (O) 연다 (V) *Transliteration: Naega sangjareul yeonda* English: I open the box - (SVO)	Tamil: நான் (S) பெட்டியை (O) திறப்பேன் (V) -SOV *Transliteration: Naan pettiyai thirappen* English: I open the box – (SVO)

4.3 Multiple Similarity Metrics

Damerau-Levenshtein similarity method (red) showed 36 pairs with scores between 0.16 and 0.45, and 55 pairs with scores between 0.5 and 0.88; Jaro method (blue) showed 92 pairs with scores between 0.5 and 0.94; LCS method (green) showed 31 pairs with values between 0.22 and 0.44, and 60 pairs with intermediate scores between 0.5 and 0.87. Cosine similarity (orange) showed 13 pairs with 0 scores, 41 pairs with scores 0.12–0.47, and 37 pairs between 0.5 and 0.92. Jaccard indexes (purple) showed 13 pairs with 0 scores, 60 pairs between 0.06 and 0.4, and 18 pairs from 0.5 to 0.86. Ratcliff Obershelp similarity (yellow) showed 36 pairs between 0.16 and 0.45 and 55 pairs between 0.5 and 0.875. Across all metrics, 9 word pairs achieved maximum similarity (1.0). Cosine and Jaccard were the only metrics showing zero similarity for 13 word pairs. Overall, the **average similarity score** for the 100 Korean–Tamil word pairs was **56%** (Fig. 12).

The Jaccard similarity is shown in Fig 13 (dark blue and white) WITH White lines (9 word pairs) indicating zero distance between pairs. Apart from zero measures, DLD shows five groups with values 1 (17 pairs), 2 (20 pairs), 3 (26 pairs), 4 (13 pairs), and 5–7(15 pairs). In addition, JD shows four groups of similarities: 0.05–0.20 (36 pairs), 0.21–0.30 (28 pairs), 0.301–0.4 (20 pairs), and 0.401–0.5 (7 pairs).

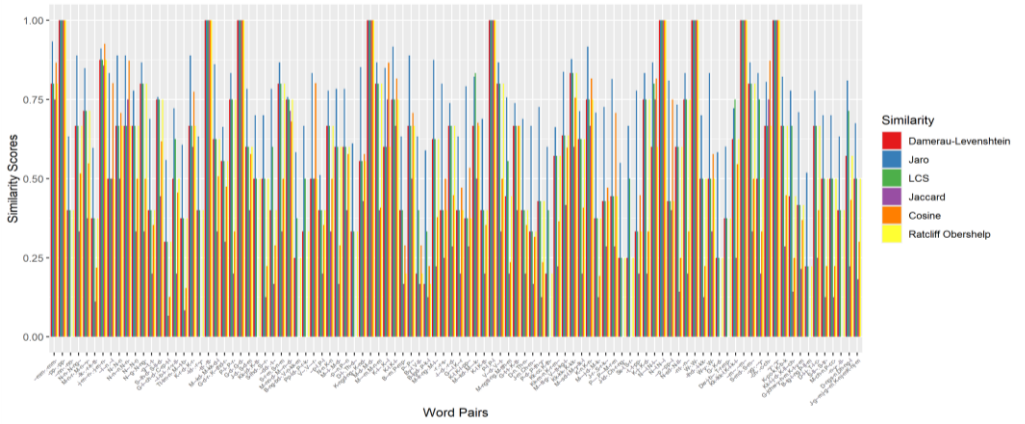

Figure 12. Multiple similarity metrics for 100 Korean-Tamil word pairs.

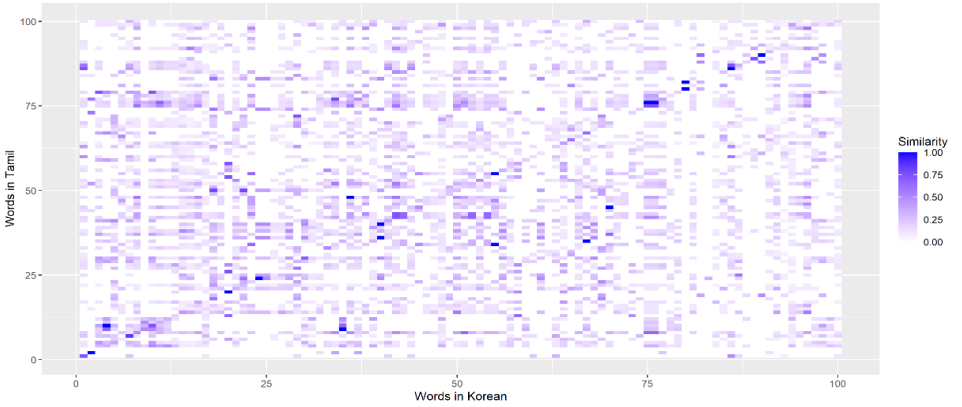

Figure 13. Jaccard similarity scores for 100 Korean–Tamil word pairs. Cells are color-coded from 0.0 (white) to 1.0 (blue) on a logarithmic scale.

Fig. 14 shows the cluster information and the mean similarity values of 100 Korean-Tamil word pairs, with phonetic similarity analyzed with hierarchical clustering based on Euclidean distance and average linkage clustering (UPGMA). The height of the merging in the dendrogram indicates how distinct the phonemes are from each other. Languages with lower merge points are closer to each other, and languages with higher merge points are more distinct. Cluster 1 shows values between 0 and 0.2, cluster 2 between 0.2 and 0.45, cluster 3 greater than 0.45. The third cluster includes languages that exhibit a high level of phonetic diversity (language pair divergence), which could have been induced through varying syllabic patterns, consonant-vowel distributions, or phonological principles.

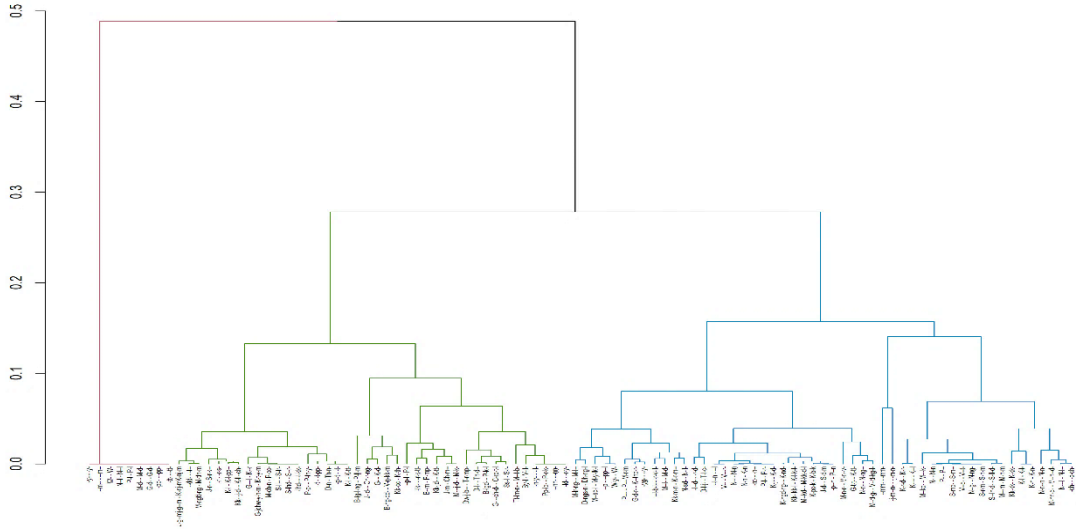

Figure 14. UPGMA dendrograms based on different similarity coefficients

5. Discussion

5.1 Linguistic Comparison between Tamil and Korean

Comparative linguistics traditionally identifies possible relationships between languages through several indicators, including (1) identical lexical forms, (2) systematic phonemic correspondences, (3) phonetic resemblance, and (4) minimal phonemic variation among potential cognates (Keraf, 1991; Batubara & Widayati, 2022). This study examines potential correspondences between Tamil and Korean through phonological comparison, script morphology analysis, and computational similarity measurements applied to a dataset of 100 everyday lexical items. Rather than attempting to establish a genetic relationship between the two languages, the analysis evaluates whether the observed similarities may be explained by typological convergence, lexical coincidence, or historical contact facilitated through long-distance trade networks.

Both Tamil and Korean exhibit several typological features commonly associated with agglutinative languages. These include the use of suffix-based morphological structures and a predominant Subject–Object–Verb (SOV) word order. Such characteristics are well documented in typological research and occur widely across languages of Eurasia. Therefore, the presence of these shared structural traits alone cannot be interpreted as evidence of a genetic relationship; instead, they place both languages within a broader typological category shared by many unrelated linguistic systems. Both languages also employ systematic combinations of consonants and vowels to construct syllables.

From a phonological perspective, both languages exhibit consonant–vowel syllable structures; however, their writing systems differ fundamentally. Hangul is a featural alphabet encoding articulatory features, whereas Tamil script, derived from the Tamil-Brahmi tradition, represents consonant–vowel combinations

through diacritics. Thanabalasingam (2023), using a weighted phonetic framework to compare the consonant and vowel systems of Korean and Tamil, identifies several phonetic similarities between the two languages but emphasizes that these correspondences represent phonological parallels rather than evidence of a historical or genetic relationship.

Morphological comparison of Hangul vowels (ㅏ, ㅑ, ㅣ, ㅜ, ㅠ, ㅡ, ㅛ, ㅖ, ㅒ) and consonants (ㅋ, ㅇ, ㅅ, ㅌ, ㄷ, ㄴ, ㅂ, ㅁ, ㄹ) suggests partial correspondences with characters found in historical forms of the Tamil script (ca. 2nd –12th century CE), including similarities in stroke arrangement and certain phonetic groupings (Figure 9–11). In comparative writing-system studies, stroke analysis is commonly used as a palaeographic and orthographic method to examine how characters are constructed, ordered, and visually structured across different scripts (Zhang, 2014a, 2014b; Zhao & Sahni, 2020). These observations indicate limited structural resemblances in letter formation and aspects of phonological organization between the two writing systems.

5.2 Interpretation of Computational Similarity Metrics

In computational linguistics and natural language processing (NLP), multiple similarity metrics are commonly used to evaluate relationships between linguistic units, including words and phonological sequences. Because each metric captures different aspects of similarity—such as phonetic, structural, or orthographic resemblance—the use of multiple metrics provides a more comprehensive assessment. In this study, the analysis focused primarily on consonantal patterns, as consonants tend to be more stable diachronically than vowels and often serve as key elements in comparative linguistic reconstruction (Campbell, 2013; Tanaka-Ishii, 2015; Tanaka, 2015).

The computational analysis of 100 Korean–Tamil word pairs yielded an overall similarity value of approximately 56% across six string-similarity metrics: Damerau–Levenshtein distance, Jaro similarity, Longest Common Subsequence (LCS), Cosine similarity, Jaccard similarity, and Ratcliff–Obershelp similarity. These methods measure orthographic resemblance between strings by evaluating edit distance, character alignment, and n-gram overlap. In linguistic applications, such metrics provide useful exploratory tools for identifying patterns of phonetic or orthographic resemblance across languages. However, they do not establish historical linguistic relationships on their own. Historical linguistics typically requires systematic correspondences between phonemes across large lexical datasets to demonstrate genetic relatedness between languages (Swadesh, 1952; Hockett, 1963).

Within the present dataset, some lexical pairs displayed high similarity scores. For example, kinship terms such as appa and amma appear in both languages with similar phonetic forms. However, cross-linguistic studies indicate that certain phonetic patterns—particularly kinship terms containing bilabial consonants and open vowels—are common across many unrelated languages due to universal tendencies in child language acquisition. In addition to these kinship terms, several Korean–Tamil lexical similarities related to agriculture and rural

life were also observed in this study. Such similarities may potentially reflect historical contact facilitated by long-distance trade networks.

The use of multiple similarity metrics in this study allows for a more nuanced evaluation of lexical resemblance. High similarity values typically indicate minimal character differences, whereas lower values reflect greater phonological divergence. The clustering analysis using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) grouped the lexical dataset into clusters representing high, moderate, and low similarity. The clustering results indicate that only a minority of lexical pairs show high similarity, while most display moderate or low resemblance. Consequently, the results should be interpreted as evidence of **limited phonetic resemblance rather than proof of linguistic relatedness.**

Several studies have applied string similarity metrics for cross-linguistic comparisons. Dinesh Kumar et al. (2018), explored syllable matching and transliteration using Jaro and Levenshtein methods for scripts such as Hindi, Marathi, Bengali, and Telugu, along with Japanese, Chinese, Korean, Russian, Turkish, and Hebrew. Ngoc et al. (2018), applied Levenshtein distance for French-Vietnamese named entity translation. Acs et al. (2024), analyzed language distances using mBERT and multilingual FastText embeddings to compute averaged pairwise distances across word lists. Li & Dunn (2022) evaluated frequency-based corpus similarity measures across 39 languages for register prediction and typological comparison. Ohnmar et al. (2013), established cross-lingual phonetic similarity metrics for loanwords across Hindi, Tamil, and Korean.

5.3 Historical Context and Maritime Exchange Networks

Although direct linguistic evidence for Tamil–Korean contact remains limited, historical sources indicate the existence of extensive maritime trade networks linking South Asia, Southeast Asia, and East Asia (Schottenhammer, 2008; Sen, 2006a). Tamil merchant guilds were active participants in Indian Ocean commerce and established trading communities across Southeast Asia and parts of southern China (Champakalakshmi, 1996; Dayalan, 2024b). These networks facilitated the movement of merchants, goods, and cultural practices across large geographic regions. Among the materials that circulated widely within these trade systems were glass beads, ceramics, and luxury goods. Archaeological discoveries of Roman glass objects and Indo-Pacific glass beads in tombs of the Silla indicate long-distance trade networks connecting the Mediterranean world, South India, and East Asia. Production centers such as Arikamedu and other sites in Tamil Nadu (Kodumanal, Keeladi) supplied glass beads that circulated through Southeast Asian ports before reaching China and the Korean Peninsula (Dussubieux et al., 2010; Glover & Kenoyer, 2019). Chemical analyses of these beads further support connections with production traditions in South India and Southeast Asia (Yi et al., 2022).

In addition to material exchange, some studies have pointed to possible evidence of population interaction. Skeletal remains excavated at Adichanallur in Tamil Nadu have been reported to exhibit certain cranial features described in earlier anthropological literature as “Mongoloid,” suggesting possible prehistoric or early historic population contacts across regions (Arokiyaraj et al., 2021;

Pathmanathan et al., 2011). However, modern research emphasizes that such classifications are unreliable, and South Asian population history is highly complex. These findings should therefore be interpreted cautiously within a broader multidisciplinary framework.

Epigraphic evidence further confirms the presence of Tamil merchant communities along these maritime routes. Tamil inscriptions have been discovered in several regions, including Thailand, Oman, Egypt, and China. One notable example is the Galle Trilingual Inscription in Sri Lanka, which contains Tamil, Persian, and Chinese texts documenting the activities of foreign merchant communities engaged in long-distance trade (Guy, 2001). It demonstrates the presence and importance of **Tamil-speaking merchant communities** within these networks, as well as the diplomatic and commercial interactions facilitated by maritime routes.

Historical records also indicate that representatives from the Maʿbar region (Chola Kingdom) of the Coromandel Coast in Tamilnadu visited the court of the Goryeo Kingdom in 1298 CE during the reign of King Chungseon of Goryeo (The Academy of Korean Studies, 2017; Lee & Yi, 2017). These diplomatic exchanges reportedly involved the presentation of luxury goods such as pearls and textiles. Although these interactions do not demonstrate linguistic exchange, they illustrate the broader context of intercultural contact within medieval maritime trade networks.

Another example of transregional mobility involves Sayyid bin Abu Ali, a Muslim merchant from the Maʿbar region of Tamil Nadu (South India) who later settled in Quanzhou, an important maritime port of Yuan-period China. Historical accounts suggest that he married a woman associated with an elite family of the Goryeo Kingdom. In South India, Muslim merchants bearing the title Sayyid were often locally referred to as Sāyid, Sāyed, or Sāibu in Tamil-speaking communities. This episode illustrates the interconnected nature of maritime trade networks linking South India, China, and Korea, although it does not provide direct evidence of linguistic influence between Tamil and Korean communities (Guy, 2001; Sen, 2006b; Schottenhammer, 2008; Han, 2016).

At present, no Tamil inscriptions or linguistic materials have been discovered on the Korean Peninsula. The absence of such evidence limits the ability to demonstrate direct linguistic influence. Any interaction between Tamil-speaking communities and Korean societies would therefore likely have occurred indirectly through intermediary trade networks connecting South Asia, Southeast Asia, and China.

Findings of the Study

Taken together, the results suggest that the observed similarities between Tamil and Korean are best explained by a combination of factors. First, only limited lexical similarity is observed between Tamil and Korean. Second, computational similarity metrics primarily capture phonetic resemblance rather than demonstrating true linguistic relatedness. Third, shared typological features between the two languages are more plausibly attributed to structural convergence than to common origin. Fourth, comparisons of the writing systems

indicate typological parallels rather than evidence of historical derivation. Finally, any similarities may also reflect indirect influence facilitated by historical contact through maritime trade networks.

6. Limitations

This study has several limitations. First, the lexical dataset is restricted to 100 Korean–Tamil word pairs, which limits statistical robustness and the generalizability of the findings. Second, the computational methods employed measure orthographic and phonetic similarity but do not establish historical linguistic relationships or systematic sound correspondences. Third, the analysis relies primarily on modern language forms, without extensive reconstruction of earlier historical stages. Fourth, the absence of direct epigraphic or linguistic evidence—such as Tamil inscriptions in Korea—limits the ability to confirm direct contact between the two language communities. Finally, the historical interpretations are constrained by indirect archaeological and textual evidence, which do not provide conclusive proof of sustained interaction.

7. Conclusion

This study examined potential linguistic parallels between Tamil and Korean using phonological analysis, writing-system comparison, and computational similarity metrics. The results show that shared features such as agglutinative morphology, SOV word order, and consonant–vowel structures are typologically common and do not indicate a genetic relationship. Computational findings revealed moderate similarity, but these reflect surface-level phonetic and orthographic resemblance rather than systematic historical correspondence. Evidence from maritime trade networks suggests the possibility of indirect cultural contact, though no direct linguistic or epigraphic evidence supports Tamil–Korean interaction. Overall, the similarities are best explained by typological convergence, lexical coincidence, and limited indirect contact. According to the present literature review, this study represents one of the first to systematically examine Tamil–Korean linguistic parallels using an integrated approach combining comparative linguistics and computational analysis. The findings highlight the need for cautious interpretation of cross-linguistic similarities, and future research should incorporate larger datasets and historical linguistic methods to further evaluate potential connections.

Conflicts of Interest
The author declares no conflicts of interest.

Acknowledgments
The author gratefully acknowledges Sejong University for its institutional support during this research. He sincerely thanks M. K. Stalin and the Director of the World Tamil Sangam for conferring the Linguistics Award in 2021 in recognition of his contributions to Tamil studies. He also acknowledges Member of Parliament Dr. Thol. Thirumavalavan for his insightful lecture on Korean–Tamil relations at the 2020 international webinar organized by the Korea Tamil Association, which inspired this research. The author further appreciates the encouragement of the Non-Resident Tamils Welfare Board Director and Deputy Director Mr. Ramesh

Special thanks are extended to Sahaya Darcius, Ganaraj, and Shanthi Prince of the South Korea–Tamil Research Association, as well as to Dr. Balamurali Krishnan, for their valuable support. Finally, he thanks Mudiappan Selvaraj, Sironmani Louis Duieppere, Dr. S. J. Alphonse Manickam, A. Helan George Mary, and the anonymous reviewers for their constructive feedback.

Author Contributions: SA: conceptualization and design of the work, data acquisition, analysis, interpretation, drafting, and approval of the final manuscript. KV and JBP: analysis of machine learning algorithms and manuscript revision.

References

Acs, J., Hamerlik, E., Schwartz, R., Smith, N. A., & Kornai, A. (2024). Morphosyntactic probing of multilingual BERT models. *Natural Language Engineering, 30*, 753–792.

Arokiyaraj, S., Ravichandran, G., Chozhan, A., & Narayanan, K. (2021). Korean–Tamil language and cultural similarities, maritime trade between early historic Tamilakam and Korea. *Shanlax International Journal of Arts, Science and Humanities, 8*(3), 28–36.

Ashok, S. (2022). Stalin's archaeology push in Tamil Nadu is the stuff of culture wars. *ThePrint*.

Batubara**,** N. A., & Widayati, D. (2022). Language kinship of English, German, and Dutch: A comparative historical linguistic study. *International Journal of Humanities Education and Social Sciences, 1*(6), 1016–1024.

Campbell, L. (2013). *Historical linguistics: An introduction* (3rd ed.). Edinburgh University Press.

Champakalakshmi, R. (1996). *Trade, ideology and urbanization: South India 300 BC to AD 1300.* Oxford University Press.

Clippinger, M. E. (1984). Korean and Dravidian: Lexical evidence for an old theory. *Korean Studies, 8*, 1–57.

Dayalan, D. (2013). Tamil Brahmi script on amphora sherd found at Khor Rori-Sumharam, Oman. *Epigraphy of the Orient, 30*, 146–148.

Dayalan, D. (2024a). Ancient seaports of Tamil Nadu and Kerala and their trade network. In A. Parasher Sen (Ed.), *Handbook on urban history of early India*. Springer.

Dayalan, D. (2024b). Cultural and trade links between India and Siam. *Acta Via Serica, 9*(1), 67–90.

Delmestri, A., & Cristianini, N. (2012). Linguistic phylogenetic inference by PAM-like matrices. *Journal of Quantitative Linguistics, 19*(2), 95–120.

Dinesh Kumar, M., Prasath, R., & Rajendran, P. (2018). Cross-language transliteration using string similarity metrics. *International Journal of Computational Linguistics, 9*(2), 45–56.

Dussubieux, L., Gratuze, B., & Blet-Lemarquand, M. (2010). Mineral soda alumina glass. *Journal of Archaeological Science, 37*(7), 1645–1655.

Glover, I., & Kenoyer, J. M. (2019). Overlooked imports: Carnelian beads in the Korean Peninsula. *Asian Perspectives, 58*(1), 180–201.

Guy, J. (2001). The Galle trilingual inscription. *Journal of the Royal Asiatic Society, 11*(3), 1–21.

Hae-Young, W. (2021). Along the sea turtle trail. *Journal of East-West Comparative Literature, 57*, 199–224.

Han, J.-S. (2016). Foreigners and their social integration in Yuan China: The case of Quanzhou. *Journal of Asian History, 50*(1), 45–70.

Hockett, C. F. (1963). The problem of universals in language. In J. Greenberg (Ed.), *Universals of language*. MIT Press.

Hulbert, H. B. (1905). *A comparative grammar of the Korean language and the Dravidian languages of India*. Methodist Publishing House.

Jaro, M. A. (1989). Advances in record-linkage methodology. *Journal of the American Statistical Association, 84*(406), 414–420.

Kang, G. U. (1990). *고대사의 비교언어학적 연구 [A comparative linguistic study of ancient history]*. Saemunsa.

Keraf, G. (1991). *Linguistik bandingan historis*. Gramedia.

Kim, K. (1999). A new proposal for a standard Hangul code. *Computer Standards & Interfaces, 20*, 243–257.

Kim-Renaud, Y.-K. (1997). *The Korean alphabet*. University of Hawai'i Press.

Kokarneswaran, M., Selvaraj, P., Ashokan, T., Perumal, S., Sellappan, P., Murugan, K. D., ... & Chandrasekaran, V. (2020). Discovery of carbon nanotubes in sixth century BC potteries from Keeladi, India. *Scientific reports*, *10*(1), 19786.

Lee, K. H., & Yi, K. (2017). Koryŏ's trade with the outer world. *Korean Studies, 41*, 52–74.

Lee, K. R. (2017). A study on the cultural contacts between Garak Kingdom and ancient South India: With special reference to fish worship. *Journal of Indian Studies, 22*(1), 85–121. https://doi.org/10.21758/jis.2017.22.1.85

Li, H., & Dunn, J. (2022). Corpus similarity measures remain robust. *Lingua, 275*, 103377.

Liu, D., & Tang, X. (2024). Comparative linguistic analysis with Firthian collocations: Cases of synonym differentiation and proficiency assessment. *Lingua*, *306*, 103755. https://doi.org/10.1016/j.lingua.2024.103755

Liyanarachchi, G. (2013). The Periplus of the Erythraean Sea. *Accounting History, 18*(2), 277–279.

Mahadevan, I. (2003). *Early Tamil epigraphy*. Harvard University Press.

Ngoc, L. T., et al. (2018). Named entity translation using Levenshtein distance. In *Proceedings of the international conference on language resources*.

Ohnmar, K., et al. (2013). Cross-lingual phonetic similarity for loanword identification. In *Proceedings of IJCNLP* (pp. 1261–1268).

Pakhomov, S. V., & Hemmy, L. S. (2014). A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the Nun Study. *Cortex*, *55*, 97-106. https://doi.org/10.1016/j.cortex.2013.05.009

Pathmanathan, R., Pearce, J., Kjeldskov, J., & Smith, W. (2011). Using mobile phones for promoting water conservation. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference* (pp. 243-252).

Reshma, V. M., & Mathew, L. S. (2015). Longest common subsequence method. *IOSR Journal of Computer Engineering, 17*(6), 1–7.

Rinjeni, T. P., Indriawan, A., & Rakhmawati, N. A. (2024). Matching scientific article titles using Cosine Similarity and Jaccard Similarity algorithm. *Procedia*

Computer Science, *234*, 553-560. https://doi.org/10.1016/j.procs.2024.03.039

Santos, R., Murrieta-Flores, P., & Martins, B. (2017). Combining string similarity metrics. *International Journal of Digital Earth, 11*(9), 913–938.

Schottenhammer, A. (2008). *The East Asian Mediterranean.* Harrassowitz.

Sen, T. (2006a). The Yuan dynasty and the Indian Ocean. *Journal of the Economic and Social History of the Orient, 49*(3), 415–445.

Sen, T. (2006b). *Buddhism, diplomacy, and trade*. University of Hawai'i Press.

Sidebotham, S. E. (2011). *Berenike and the ancient maritime spice route*. University of California Press.

Sivanantham, R., & Seran, M. (2019). *Keeladi*. Government of Tamil Nadu.

Steever, S. B. (Ed.). (2019). *The Dravidian languages* (2nd ed.). Routledge.

Suresh Kumar, D. (2025). Keeladi excavation report controversy. *The Hindu*.

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proceedings of the American philosophical society*, *96*(4), 452-463.

Tanaka, S. (2015). Consonantal stability in historical phonology. *Journal of Historical Linguistics*.

Tanaka-Ishii, K. (2015). Consonants as skeleton of language. In *Language production*. Springer.

Thanabalasingam, U. (2023). A phonetic comparison of Korean and Tamil. *Open Journal of Modern Linguistics, 13*, 711–733.

The Academy of Korean Studies. (2017). *A history of Korea*.

Verma, S. P. (2005a). *Trade and cultural contacts in the Indian Ocean world*. Manohar.

Verma, V. K. (2005b). Maritime trade between early historic Tamil Nadu. *Proceedings of the Indian History Congress, 66*, 125–134.

Yi, J., et al. (2022). Compositional analysis of early glass beads. *Journal of Archaeological Science: Reports, 41*, 103293.

Zhang, H. (2014a). Stroke order in Hanzi handwriting. *Cercles, 24*(1), 67–85.

Zhang, Q. (2014b). Stroke structure analysis. *Writing Systems Research*.

Zhao, C., & Sahni, S. (2020). String correction algorithm using Damerau-Levenshtein distance. *BMC Bioinformatics, 21*, 14.