EJTAS
ISSN 2786-7447
EUROPEAN JOURNAL OF THEORETICAL AND APPLIED SCIENCES
OPEN ACCESS
PEER REVIEWED

Morphologic, Syntactic, and Phonologic Distance Between Japanese and Altaic, Dravidian, Austronesian, and Korean Languages

Wenchao Li*
Department of Japanese Studies, Zhejiang University, China

Article Information

Suggested Citation:
Li, W. (2023). Morphologic, Syntactic, and Phonologic Distance between Japanese and Altaic, Dravidian, Austronesian, and Korean Languages. *European Journal of Theoretical and Applied Sciences, 1*(2), 14-36.
DOI: 10.59324/ejtas.2023.1(2).02

*** Corresponding author:**
Wenchao Li
e-mail: widelia@zju.edu.cn

Abstract:

The present study measures the resemblances of Japanese with Altaic languages (Turkic; Tungstic; Mongolic; Nivkh); the Dravidian language Tamil; Austronesian languages (Western Malayo-Polynesian; Malayo-Sumbawan; Central Luzon; Central Malayo-Polynesian), and Korean, in an effort to pin down the genealogy of Japanese. Morphologic, syntactic, and phonologic distance are calculated using data from corpora. The chi-square homogeneity test and Euclidean distances are used for statistical analysis. The finding brings to light, morphologically, in the light of preferences of causative/inchoative verb alternation patterning and morphemes that convey the alternation, that Japanese and Korean are close for the most part. Syntactically, Altaics and Tamil convey case via suffixes; case in Austronesian languages is marked by prefixes. Japanese and Korean share a similarity in rendering case with particles. Phonologically, the Tamil and Austronesian languages share a resemblance in the harmony of vowel height. The Korean, Altaic languages, and Austronesian languages show similarities in the harmony of vowel backness. Japanese, the Altaic languages, and the Austronesian language Madurese display vowel-consonant harmony. Pulling these strands together, a conclusion is thus drawn that Japanese is most closely related to Korean.

Keywords: *Japanese, genealogy, morphology, phonology, case system.*

Introduction

Previous studies on the genealogy of the Japanese language have considered six streams: (a) that it originated within the Korean language (Aston, 1879); (b) that it originated within the Polynesian language (Ohno, 1957); (c) that it was an Altaic language the typological approach (Hattori, 1959); (d) that it was a member of the Dravidian family, and resembled the Tamil language (the historical approach) (Ohno, 1981); (e) that it was a mixture of Tungus and Austronesian (the linguistic-geographic approach) (Matsumoto, 2007; Sasakiyama 2012); and (f) as Robbeets et al. (2021) have argued, that it was of Transeurasian origin – it developed in the West Liao River region in the Early Neolithic and dispersed into two branches in the Late Neolithic and Bronze Age, with the northward branch expanding to the Mongolian Plateau, with Proto-Turkic, and the eastward branch dispersing to the Korean peninsula and to the Japanese islands. The present study looks at the correspondences of Japanese with the Altaic languages, the Austronesian languages, the Dravidian language Tamil, and Korean, by measuring distances of morphology (causative/inchoative verb alternation patterning; the morphemes that render verb alternation); syntax (case marking system; word

order); and phonology (assimilation in vowels and consonants). It aims to arrive at a conclusion about the genealogy of Japanese.

Japanese

Modern Japanese is deemed to be phonologically moraic and morphologically agglutinative, cf. (1).

(1) a. **moraic**

L H L

| | |

ku ji ra

b. **agglutinative**

食べ-させ-られ-まし-た-か

tabe-sase-rare-mashi-ta-ka.

eat (stem)-causative-passive voice-honorification-tense. past-question marker

Agglutination refers to: one or more suffixes are added to a verb/adjective stem to give rise to complex predicates. It differs from (a) inflection, e.g. Germanic language: walk, walk-s, walk-ed; (b) internal inflection, e.g. African language Tchad: yĕ (telic)/yè (atelic); (c) vowel alternation, e.g. Arabic language: katab (telic, active voice)/kattab (causative, telic, active voice)/kutib (telic, passive voice); (d). fusion, e.g. Latin: stella 'star-singular.nominative'; (e) isolation, e.g. Chinese: wǒ yǐ jīng zhī dào nà gè gù shì; Thai: rûaŋ nán phŏm rúu lέεw 'I already know that story'; (f) incorporation, e.g. African language Chichewa: Mtsikana ana-chit-its-a kuti mtsuko u-gw-e. [girl AGR-do-make-ASP that waterpot AGR-fall-ASP] (Baker, 1988).

Timeline	*Joomon period*		*Yayoi period*
	Ten thousand – B.C. 3	→	B.C. 3 –3 A.D.
Influenced by	Polynesian		Korean

Figure 1. The origin of the Japanese language (Ohno, 1957)

There has been an extensive study regarding the genealogy of Japanese in the literature. In earlier times, the mainstream view took a historical-linguistic approach. Ohno (1957) argues that Japanese originated in Polynesian in the Joomon period (10,000 years ago–3 BC) and was influenced by Korean in the Yayoi period (3 BC–

Ohno (1981) conducted a field investigation in Tamil (the Dravidian family of languages, spoken in the southeast of India and Sri Lanka) and assumed Japanese originated in the Austronesian family of languages during the Joomon period and was connected to the Tamil language of the Yayoi period (see Figure 2).

Timeline	*Joomon period*		*Yayoi period*
	Ten thousand – B.C. 3	→	B.C. 3 –3 A.D.
Influenced by	Austronesian		Tamil (Dravidian family)

Figure 2. The origin of the Japanese language (Ohno, 1981)

A different view comes from typological linguists. Hattori (1959) deems Japanese an Altaic language. Matsumoto (2001) argues that the Altaic people migrated to Japan, along with their languages, between the end of the Joomon period and the beginning of the Yayoi period. The 'Altaic-related' view is also championed outside of Japan, with Miler (1971) being the main scholar propagating this view. The evidence that Miler (1971) defends is the resemblance of word order, i.e. SOV (Turkic, Tungstic, and Mongolic).

Another pathway comes from a linguistic-geographic view, alleging that Japanese is a mixed language made up of the Northern Tungstic and Southern Austronesian languages (Matsumoto, 2007). A similar conclusion is reached by Sasakiyama (2012), who studied the sound and lexicons in Old Japanese from ethnology and discovered that: the Ka(h) particle is used both in Indonesian (kah) and Japanese (ka) as a question marker; 八 (eight) behaves as a prefix, meaning 'many' instead of a quantifier; and 親oya refers to an ancestor, which resembles Malaysian usage. The similarities in lexicons have inspired linguists to deduce that the

Austronesian languages have a genetic relationship with Japanese.

There is another notable work to mention: Kobashi and Tanaka (2011) created unrooted trees from 12 languages via: (a) neighbor joining; (b) maximum parsimony; and (c) Bayesian. The three tests all indicate that the distance between Japanese and the Altaic languages is the shortest, which indicates that Japanese is closest mostly to the Altaic languages. The approaches that have contributed to the Japanese genealogy so far are summarised in Table 1.

Table 1. Previous hypotheses on Japanese genealogy

Hypothesis 1	Korean originated	Aston (1879)	historical linguistic approach
Hypothesis 2	Polynesian originated	Ohno (1957)	historical linguistic approach
Hypothesis 3	Altaic originated	Hattori (1959)	typological approach
Hypothesis 4	Dravidian related	Ohno (1981)	historical linguistic approach
Hypothesis 5	Mixed language (Northern Tungusic & Southern Austronesian)	Matsumoto (2007)	linguistic geographic approach
Hypothesis 6	West Liao River region	Robbeets et al. (2021)	language dispersal, agriculture expansions, population movements

Even if we take similarities between Japanese and potentially related languages (e.g. being agglutinative, displaying SOV word order) into consideration, there are data indicating distinctions. In particular, the phonological feature of vowel harmony is salient in the Altaics, but does not exist in Modern Japanese. Certainly, it can be argued that in Old Japanese (AD 700–800; a dead language spoken in Asuka and Nara periods), the eight vowels, i.e. /a/, /e_1/, /e_2/, /i_1/, /i_2/, /o_1/, /o_2/, /u/, may suggest harmony. However, no evidence so far has confirmed the existence of harmonic assimilation thousand years ago. Undeniable distinctions are further tied to case marking and affixation. In an Austronesian language, case is marked by prefixes, while Japanese case is conveyed by particles. It appears then that neither the historical nor the typological approach gives us the most precise account of Japanese affiliation. In this paper, we take a more rigorous approach by making use of insights from mathematical linguistics.

The present study calculates the resemblances of Japanese with the Altaic languages; Tamil; the Austronesian languages, and the nearest continental neighbor, Korean, with a focus upon:

(I) **Morphologic distance**: causative/inchoative verb alternation patterning; the morphemes that render verb alternation;

(II) **Syntactic distance**: the case marking system; word order ((a) order of subject, verb, and object; (b) order of adjective and noun; and (c) order of adposition and noun); and

(III) **Phonologic distance**: assimilation (vowel harmony and vowel-consonant harmony).

This paper is organised as follows. Section 2 introduces the corpora. It also provides an insight into the chi-square homogeneity test and Euclidean distance. Sections 3 measures the morphologic distance, i.e. the patterning of causative/inchoative verb alternation and the distribution of morphemes that convey verb alternation. Section 4 moves on to a comparison of syntax: the case marking system and word order. Section 5 delves into a phonologic issue, i.e. assimilation in vowels and consonants. Section 6 highlights the results and concludes the paper.

Methodology

Data

The central goal of this study is to establish whether Japanese is morphologically, syntactically, and phonologically related to the Altaic, Austronesian, Dravidian, and Korean languages. To this end, distances of morphology, syntax, and phonology are calculated.

The morphological distance is measured with data from *The World Atlas of Transitivity Pairs* (Dryer, 2005), focusing particularly on two issues: (a) alternation patterning frequency; and (b) morphemes that invite causative/inchoative verb alternation.

The syntactic distance is measured with data from *The World Atlas of Language Structures* (Dryer, 2005).

Analysis

To decide whether a morphologic relationship exists between two languages, the feature values and patterning frequencies are calculated. Building on this, the chi-square homogeneity test is employed to identify the distance between them. The morphological distance is determined through the formula:

$$x^2 = \sum \frac{(O-E)2}{E} \sim x^2(r-1)(c-1) \qquad (1)$$

Among them, the number of degrees of freedom of the χ2 independence test statistics:

$$d.f. = (rows - 1) * (columns - 1) \qquad (2)$$

The E in the formula stands for the expected frequency accounts, determined via:

$$E = \frac{\text{Row variable Total} \times \text{Column variable Total}}{\text{Grand Total}} \qquad (3)$$

The syntactic distance is determined via Euclidean distance. Assume L_1 and L_2 as the vectors representing the compared languages, the syntactic distance between L1 $(L1,1, \ldots, L_{1,n})$ and L_2 $(L_{2,1}, \ldots, L_{2,n})$ would be calculated through the formula:

$$d(L1, L2) = \sqrt{(L1,1 - L2,1)^2 + (L1,2 - L2,2)^2 + \cdots + (L1,n - L2,n)^2} = \sqrt{\sum_{i=1}^{n}(L1,n - L2,n)^2} \qquad (4)$$

In addition, given that Old Japanese is a dead language, it seems difficult to segment and tag the documents that are written in different writing systems. In a pilot study, the MeCab and Unidic Libraries[1] are tested. Neither of them, however, seems to return proper results.

```
4       '       '       PUNCT   補助記号-句点   _       3       punct   _       SpaceAfter=No
5       '       '       PUNCT   補助記号-句点   _       3       punct   _       SpaceAfter=No
6       伊豆    イズ    PROPN   名詞-固有名詞-地名一般  _       8       compound        _       SpaceAfter=No
7       毛夜    毛夜    NOUN    接尾辞-名詞的一般       _       8       compound        _       SpaceAfter=No
8       弊賀岐  弊賀岐  NOUN    名詞-普通名詞-サ変可能  _       0       root    _       SpaceAfter=No
9       '       '       PUNCT   補助記号-句点   _       8       punct   _       SpaceAfter=No
10      '       '       PUNCT   補助記号-句点   _       8       punct   _       SpacesAfter=\n
```

Figure 3. Kojiki kayō, hentai-kanbun data (MeCab)

MeCab did not segment the sample data**伊豆毛夜弊賀岐** idumwo ya-pye-gaki properly. **伊豆毛**idumwo is a name of a place and therefore ought to be segmented as follows.

(6) **伊豆** **毛夜弊賀岐**
idumwo ya-pye-gaki
Idumo many-fenced palace

Though Unidic2ud successfully segmented **夜久毛**ya-kumwo and**多都** tatsu, it fails to tag them properly, i.e. **夜** ya is a prefix in **夜久毛**, meaning 'many'; **多都** tatsu is an unergative verb, corresponding to tatsu (rise).

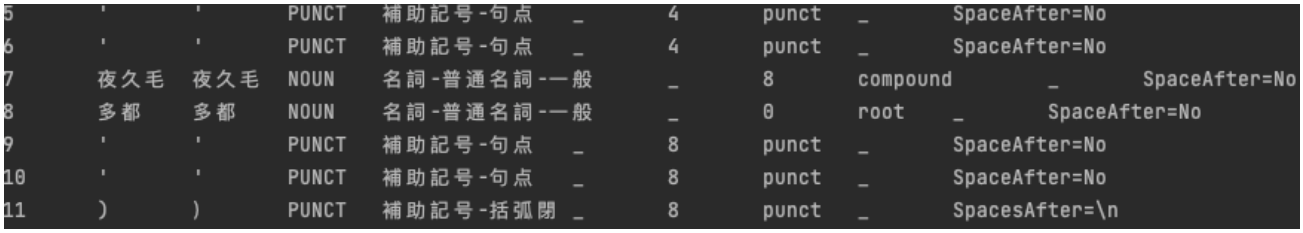

Figure 4. Nihonshoki kayō, junsee-kanbun data (Unidic2ud)

The present study would therefore handedly mark the part of speech of Old Japanese data. (7) provides an illustration of segmentation as well as the tagging.

(7) 加是布加牟登須 [kaze puka-mu to su] 'say that the wind would blow' (Kojiki kayō. 20).

[kaze puka-mu to su]

| | | | |

S V C V

登to is a complementiser, indicating a subclause rendered by the verb 須su (say), which indicates an SOCV (SOV) word order.

Results and Discussion

Morphologic distances between Japanese and the Altaics, the Austronesians, Tamil, and Korean

Drawing on the methodology highlighted above, this section proceeds to examine the correspondences in Japanese and the potentially related languages. Our starting point is morphology. Two matters are posed: (a) distance on causative/inchoative verb alternation patterning; and (b) distance on morphemes that convey the alternation.

To begin with, causative/inchoative verb alternation is observed in about 16 language families and 80 languages. Haspelmath (1993) generalized five alternation patterns: (a) the anticausative type; (b) the causative type; (c) the equipollent type; (d) the labile type; and (e) the suppletive. Table 2 presents the morphological relationship between the inchoative and causative verbs of each pattern.

Table 2. Morphological relationships between inchoative and causative verbs

Causative/inchoative alternation pattern	Formal relationship between causative/inchoative verb
Anticausative (A)	Causative verb based; Inchoative verb derived via anticausativisation (e.g. kir-0/-(r)u 'cut [transitive]'→ kir-e-ru 'cut [intransitive]')
Causative (C)	Inchoative verb based; Causative verb derived via causativisation (e.g. or-e-ru 'break [intransitive]'→or-ø-(r)u) 'break [transitive]')
Equipollent (E)	Causative/inchoative verb derive from the same root (e.g. hiroi (Adj)→hirom-ar-u 'broaden' [unaccusative]; hiroi (Adj)→hirom-e-ru 'broaden' [transitive])
Labile (L)	Causative/inchoative verb share the same word form (e.g. ma-ku 'roll'; tojiru 'close')
Suppletive (S)	Causative/inchoative verb are formally distinct and underived (e.g. die/kill)

Alternation patterns in the Altaics, Austronesians, Tamil, Japanese, and Korean

The Altaic language family has the following genera:

(a) Turkic genus: Kyrgyz, Khakas, Kazakh, Azerbaijani, Uzbek, Turkish, and Turkmen

(b) Tungusic genus: Ewen, Nanai, Udihe, and Manchu

(c) Mongolic genus: Mongolian

(d) Nivkh genus: Sakha

Causative/inchoative alternation patterning in each genus is presented in Table 3.

Table 3. Causative/inchoative verb alternation patterns in Altaic languages

Genera	Languages	Causative/inchoative alternation (main)	Causative/inchoative alternation (other option)
Tungstic genus	Ewen	A (13 tokens)	O (8 tokens); S (8 tokens); C (4 tokens); L (4 tokens)
	Nanai	C (11 tokens)	A (9 tokens); E (9 tokens); L (5 tokens); S (7 tokens)
	Udihe	S (19 tokens)	C (11 tokens); O (9 tokens); L (3 tokens); A (2 tokens); E (2 tokens)
	Manchu	C (26 tokens)	S (11 tokens); O (5 tokens); L (3 tokens); E (1 tokens); A (1 tokens)
Turkic genus	Kyrgyz	C (19 tokens)	A (14 tokens); E (5 tokens); S (1 token)
	Khakas	C (17 tokens)	A (12tokens); E (2 tokens); S (1 token)
	Kazakh	C (22 tokens)	A (13 tokens); E (2 tokens); S (1 token)
	Azerbaijani	C (21 tokens)	A (16 tokens); E (4 tokens); L (1 token)
	Uzbek	C (18 tokens)	A (10 tokens); E (4 tokens)
	Turkish	C (18 tokens)	A (10 tokens); E (3 tokens); S (1 tokens)
	Turkmen	C (15 tokens)	A (12 tokens); E (2 tokens); S (1 token); L (1 token)
Mongolic genus	Mongolian	C (26 tokens)	A (8 tokens); E (1 token); S (1 token) and L (1 token)
Nivkh genus	Sakha	C (23 tokens)	A (7 tokens) and E (5 tokens)

Altaic languages are more likely to derive a causative verb via causativization (i.e. pattern C). The second largest number of tokens is attributed to anticausativization (i.e. pattern A: deriving an inchoative verb by adding a morpheme to a causative verb root). The equipollent pattern had the third largest applicability.

Turning to Austronesian languages, four genera are examined.

(a) **The Western Malayo-Polynesian genus**

Ilokano and Tagalog appear to favor pattern E (18 and 31 tokens, respectively), with C the second option (15 and 6 tokens).

(b) **The Malayo-Sumbawan genus**

In Indonesian, pattern E (18 tokens) is preferred, with pattern C the second likely option (15 tokens).

(c) **The Central Luzon genus**

Kapampangan is more in favor of pattern E (28 tokens), with pattern A the second option (8 tokens).

(d) **The Central Malayo-Polynesian genus**

Lamaholot apparently prefers pattern C (16 tokens) to the other options (e.g. L pattern: 6 tokens).

Table 4. Causative/inchoative verb alternation patterns in Austronesian languages

Austronesian Family	Languages	Causative/inchoative alternation (main option)	Causative/inchoative alternation (other option)
Western Malayo-Polynesian genus	Ilokano	E (18 tokens)	C (5 tokens); A (11 tokens)
	Tagalog	E (31 tokens)	C (6tokens)
Malayo-Sumbawan genus	Indonesian	E (18 tokens)	C (15 tokens)
Central Luzon genus	Kapampangan	E (28 tokens)	A (8 tokens); C (5 tokens); S (1token); L (1 token)
Central Malayo-Polynesian genus	Lamaholot	C (16 tokens)	L (6 token); E (1 token); S (1 token)

Austronesian languages' causative/inchoative verb alternation patterning is summarised in Table 4.

In Japanese, pattern A (20 tokens) is the most favored, with pattern C (16 tokens); E (4 tokens); S (3 tokens); and L (2 tokens), as the second, third, fourth, and fifth preference. The most favored option in Tamil and Korean would be pattern C (Tamil: 75 tokens; Korean: 15 tokens). A summary of the discussion is provided in Table 5.

Table 5. Causative/inchoative verb alternation pattern in Japanese, Tamil and Korean

Languages	Causative/inchoative alternation pattern (main)	Causative/inchoative alternation pattern (others)
Japanese	A (20 tokens)	C (16 tokens); E (4 tokens); S (3 tokens); L (2 tokens)
Tamil	C (75 tokens)	L (12 tokens); S (2 tokens)
Korean	C (15 tokens)	A (13 tokens); L (3 tokens); E (3 tokens); S (1 token)

With the data established, a thorough frequency analysis could provide clearer insights into how these languages are related morphologically. A chi-square homogeneity test was carried out. The results are summarized in Table 6.

Table 6. Morphological distance via Chi-square homogeneity

Language genera	Language	x^2
Altaic, Tungstic genus	Ewen	4.06
	Nanai	5.4
	Udihe	6.2
	Manchu	19.8
Altaic, Turkic genus	Kyrgyz	1.3
	Khakas	1.58
	Kazakh	2.98
	Azerbaijani	1.1
	Uzbek	2.59
	Turkish	2.49
	Turkmen	0.96
Altaic, Mongolic genus	Mongolian	9.03
Altaic, Nivkh genus	Sakha	7.2
Unknown	Korean	0.52
Austronesian, Western Malayo-Polynesian genus	Ilokano	20.9
	Tagalog	45.3
Austronesian, Malayo-Sumbawan genus	Indonesian	28.6
Austronesian, Central Luzon genus	Kapampangan	28.8
Austronesian, Central Malayo-Polynesian genus	Lamaholot	14.9
Dravidian family	Tamil	50.8

It appears:

(a) Korean has the lowest x^2 and thus is deemed mostly close to Japanese.

(b) The Altaic language family, particularly the Turkic genus, the Turkmen language, seems to bear a close morphological link to Japanese.

(c) The four genera in the Austronesian family and the Dravidian family of languages do not occur to facilitate a morphological relationship with Japanese.

Perhaps we can pause and draw a preliminary conclusion here: the Dravidian language Tamil and Austronesian languages are unexpectedly

remote from Japanese. This results do not fit into Ohno's hypothesis (1981). Korean shows a morphological closeness to Japanese. The Altaic language Turkmen appears to have the second closest likely kinship to Japanese. This is to be elaborated from a morphological point of view, to which we now turn.

Morphemes that convey causativization / anticausativization

Before getting started, it would be appropriate to give a rough classification of morphemes that render causative/inchoative verb alternation:

(8)

Added morpheme: a morpheme that carries a causative/inchoative property is added to an inchoative/causative verb root.

Paired morpheme: a morpheme that indicates a causative/inchoative property is paired with another morpheme that bears inchoative/causative property.

Table 7 presents a picture of morpheme types that convey causative/inchoative verb alternation morphemes in Japanese, the Altaics, the Austronesians, Tamil, and Korean.

Table 7. Causative/inchoative verb alternation morphemes in Japanese, Altaics, Austronesians, Tamil and Korean

Language family	Languages	Causativisation morpheme	Morpheme types	Inchoative morpheme	Morpheme types
Altaic Turkic	Turkish	/-et/, /-dur/, /-ir/ (/-tir/, /-dir/)	added suffix	/-il/, /-öl/ (-ol/), /-ül/ (/-ul/)	added suffix
		/-t/	paired suffix	/-n/	paired suffix
				/öl -/	prefix
	Turkmen	/-dur/	added suffix	/-ül/, /-yl/	added suffix
		/-t/	paired suffix	/-n/	paired suffix
	Khakas	/-dir/	added suffix	/-il/	added suffix
		/-t/	paired suffix	/-n-/	paired suffix
	Uzbek	/-dir/	added suffix	/-il/	added suffix
		/-t/	paired suffix	/-n-/	paired suffix
	Azerbaijani	/-t/, /-dur/, /-ir/	added suffix	/-il/, /-n/, /-ül/	added suffix
	Kyrgyz	/-t/, /-tür/	added suffix	/-il/, /-ül/	added suffix
	Kazakh	/-dir/, /-ir/	added suffix	/-il/	added suffix
		/-t/	paired suffix	/-n-/	paired suffix
Tungstic	Ewen	/-u/, /-n/	added suffix	/-p/	added suffix
	Nanai	/-wəən/, /-waan/	added suffix	/-biə/, /-p/	added suffix
		/-li/	paired suffix	/-ə/	paired suffix
	Udihe	/-wənə/, /-waan/	added suffix	null	
	Manchu	/-bu/	added suffix	null	
Mongolic	Mongolian	/-g/, /-ö/ (/-o/), /-a/, /-e/, /-uul/ (/-üül/)	added suffix	/-gd/	added suffix
		/-l/	paired suffix	/-r/	paired suffix
Nivkh	Sakha	/-ar/, /-t/, /-or/, /-er/, /-üt/, /-nör/	added suffix	/-un/, /-n/	added suffix
Indonesian Western Malayo-Polynesian	Ilokano	/-en/	added prefix		
		/pa-/	added prefix	/ma-/, /ag-/, /mang-/	added prefix
	Tagalog	/pa-/, /i-/	added prefix	/ma-/	added prefix

		/-in/	added suffix		
Indonesian Central Luzon	Kapampangan	/pa-/, /i-/	added prefix	/ma-/, /mi-/	added prefix
		/-p/	paired suffix	/-m/	paired prefix
Indonesian Malayo-Sumbawan	Indonesian	/me-/, /ke-/	added prefix	/di-/, /me-/, /ter-/, /ber-/, /ke-/	added prefix
		/-kan/	added suffix	null	
Indonesian Central Malayo-Polynesian	Lamaholot	/nə̃ʔə̃-/	added prefix	null	
Unknown	Japanese	/-e-/, /-as-/, /-os-/, /-s-/	paired suffix	/-ar-/, /-e-/, /-i-/	paired suffix
Dravidian	Tamil	/-kkavai/, /-kkaccey/, /-yavai/, /-yaccey/, /-ttu/, /-ppu/, /-Rpi/, /-ppi/	added suffix	null	null
		/-avai/, /-accey/	paired suffix	/-u/	paired suffix
Unknown	Korean	/-i-/, /li-/, /-wu-/, /-y-/, /-hi-/	added suffix	/-ci-/, /li-/, /-y-/, /-hi-/	added suffix

In Altaic languages, suffixes play an essential role, specifically:

(a) In the Turkic genus, Turkish suffixes /-et/, /-t/, /-dur/, /-ir/ (/-tir/, /-dir/) render causativization; /-il/, /-n/, /-öl/, /-ül/ (/-ul/) convey anticausativization; /-t/ and /-n/ are a pair of

causative/inchoative verb suffix. Turkmen suffixes /-dur-/ indicate causativization, while

suffixes /-ül/, /-yl/ render anticausativization; /-t-/ and /-n/ are a pair of causative/inchoative

verb suffixes.

(b) The Tungstic genus distinguishes the Turkic genus in that there are only added suffixes, e.g. /-wənə/ and /-bu/ for causativization in Udihe and Manchu.

Tamil is a language where causativization plays the main part (pattern C attributes 75 tokens among a total of 89 tokens), leading to a rich inventory of causative suffixes. Among the ten types of causative suffixes, eight are added causative suffixes, which presents the following derivation regulation.

(9). **Regulation of causativisation via suffix in Tamil**

(a). /-kkavai/ *attached to a root that ends with vowel /a/

(`a). /-kkaccey/ *attached to a root that ends with vowel /i/

(b). /-yavai/ *attached to a root that ends with vowel /i/

(`b). /-yaccey/ *attached to a root that ends with vowel /i/

(e). /-ttu/ *attached to a root that ends with consonant /r/

(f). /-ppu/

(g). /-Rpi/

(h). /-ppi/

There are additionally two paired suffixes in Tamil:

(10) **Two pairs of suffixes rendering causativization in Tamil**

(a) /-u/ → /-avai/ *attached to a root that ends with vowel /u/

(`a) /-u/ → /-accey/ *attached to a root that ends with vowel /u/

Like Tamil, Altaics, Japanese, and Korean both employ a suffix for causative/inchoative alternations. Japanese suffixes /-e-/, /-as-/, /-

os-/, /-s-/ invite causativization, while /-ar-/, /-e-/, /-i-/ render anticausativization. In Korean, /-i-/ (/-hi/, /-li/, /-ki/) and /wu-/ are causative suffixes, while /ha-/ and /ci-/ are inchoative.

Turning to the Austronesians, unlike the aforementioned languages, prefixes play an essential role in causative/inchoative verb alternation:

(11) **Prefixes in Austronesian verb alternation**

(a) In Ilokano, /ma-/ and /pa-/ are a pair of prefixes that bear causative/inchoative characters.

/ma-/ (/mang-/), /ag-/ are prefixes of anticausativization.

(b) Indonesian employs the prefix and suffix: /me-/ and /di-/ are a pair of prefixes rendering causativization/anticausativization; /ke-/ is a prefix that may indicate the causative as well as the inchoative; /me-/ and /ber-/ are inchoative prefixes and /-kan/ is a causativization suffix.

(c) In Kapampangan, prefixes /pa-/ and /i-/ are causative, and /ma-/, /mi-/ are inchoative; /-p/ and /-m/ are a pair of causative/inchoative prefixes.

(d) Lamaholot's alternation displays only two patterns: causativization and labile. The prefix /nə̃ʔə̃-/ renders all instances of causativization.

In light of the preferences of verb alternation patterning and based on morphemes that contribute to the alternation, we assume it is Japanese and Korean that have the shortest morphological distance[2]; the second closest languages to Japanese would be the Altaic languages.

A further look into syntactic distance might shed more light on this position.

Syntactic Distances Between Japanese and the Altaics, the Austronesians, Tamil, and Korean

Case Marking

Cross-linguistically, there are nine ways of marking a case across languages (Dryer, 2005).

(12) **Case marking typologically**

(a) case is rendered by suffixes

(b) case is rendered by prefixes

(c) case is coded by tone

(d) case is conveyed via changing a stem

(e) case is rendered by postpositional clitics

(f) case is rendered by prepositional clitics

(g) case is rendered by inpositional clitics

(h) mixed morphological case

(i) no case affixes or appositional clitics

Case in Modern Japanese is marked by particles (postpositional clitics in Dryer's (2005) terminology).

(13) **Case marking in Modern Japanese**

a. nominative case particle:	が	ga
b. accusative case particle	を	o
genitive case particle	の	no
dative case particle	に	ni
instrumental case particle	で	de
ablative case particle	から	kara
directional case particle	へ	e
comitative case particle	と	to

Case marking in Old Japanese offers rather a complex picture. Note that the writing system in Old Japanese is a mixture: hentai-kanbun, 'variant Chinese' (Kojiki kayō AD 712); junsei-kanbun, 'classical Chinese' (Nihonshoki kayō AD 720); and man'yōgana (Man'yōshū AD.759). Here, case is marked by particles, via a variety of characters, as shown in Figure 5–Figure 7.

行标签	登	迩	賀	能	袁	总计
Accusative					42	42
Comitative	8					8
Complementiser	23					23
Dative		17				17
Genitive				12		12
Nominative			10			10
总计	31	17	10	12	42	112

Figure 5. Case particles in Kojiki kayō (AD.712, hentai-kanbun 'variant Chinese')

行标签	Accusative	Comitative	Complementiser	Dative	Genitive	Nominative
登		1	5			
等		3				
餓					4	
餓						63
珥				66		
爾				67		
廻	1					
廻					2	
能					87	
苔			2			
騰			2			
烏	2					
塢	23					
总计	26	4	9	133	93	63

Figure 6. Case particles in Nihonshoki kayō (AD.720, junsei-kanbun 'classical Chinese')

行标签	Ablative	Accusative	Comitative	Complementiser	Dative	Genitive
0			0			
二					11	
我						
従	3					
登				3		
等				1		
尓					90	
荷					1	
乎		61				
跡				7		
乃						118
能						19
庭					1	
吾						
雄		1				
意					1	
与				2		
之						122
自	1					
总计	4	62	0	13	104	259

Figure 7. Case particles in Man'yōshū (AD.759, man'yōgana)

It was discovered that, in Old Japanese:

(a) The most frequent particle indicates the genitive case (five types, 256 tokens), rendered by能 (133 tokens) and 之 (123 tokens). The Chinese character能 is borrowed for its phonetic value, while 之 is borrowed for its semantic value.

(b) The second largest number of tokens is attributed to dative case particles (eight types, 270 tokens), e.g. 尓 (90 tokens), 爾 (83 tokens), 珥 (66 tokens), 迩 (17 tokens), 二 (11 tokens), 庭 (one token), 荷 (one token), and 意 (one token).

(c) Seven types, 140 tokens, are applied to accusative case particle, i.e. 乎 (65 tokens), 袁 (43 tokens), 遠 (three tokens), 塢 (25 tokens), 廻 (one token), 雄 (one token), and 烏 (two tokens).

(d) The complementizer case particle has the fourth largest applicability. Among them are the following lexemes: 登 (31 tokens); 跡 (seven tokens); 等 (five tokens); 苔 (two tokens); and 騰 (two tokens) – these are borrowed for phonetic value while与 (two tokens) is for semantic value.

(e) There are six types, with 86 tokens in nominative case particles. Among them, 我 (eight tokens) and 吾 (two tokens) are borrowed for semantic value, 何 (one token), 加 (one token), 賀 (11 token) and 餓 (63 tokens) are borrowed for phonetic value.

Case particles in Old Japanese is summarised in Figure 8:

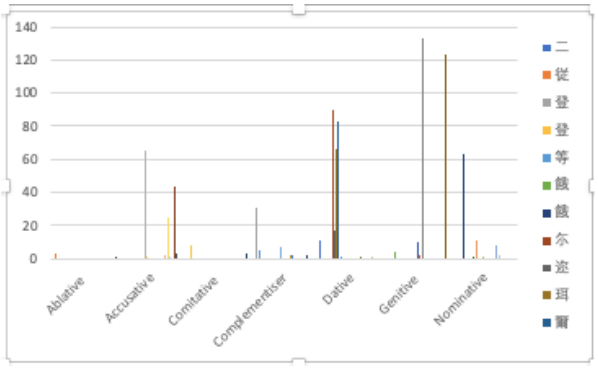

Figure 8. Case particles in Old Japanese

The Korean case system is rendered by suffixes (corresponding to Japanese case particles). The case suffixes split into two groups: (a) a case suffix attaches to a lexeme of a closed syllable

(i.e. CVC, consonant-vowel-consonant); and (b) a case suffix attaches to a lexeme that is an open syllable (i.e. CV, consonant-vowel).

(14) **Case marking in Modern Korean and Japanese** (Table 8).

Table 8. Case marking in Modern Korean and Japanese

Case	Korean case suffix CVC	Korean case suffix CV	Japanese case particles
nominative case	이	가	が
accusative case	을	를	を
genitive case	의		の
dative case	에		に
instrumental case	으로	로	で
locative case	에서		で
comitative case	과	와	と

Furthermore, both Japanese and Korean feature an accusative case alignment system.

(15) a. Taroo wa kabin o watta.
Taroo top vase acc break-past

b. Kabin ga wareta.
Vase nom broke

(16). a. 타로 는 꽃병 을 깨뜨렸다.
Taroo top vase acc break-past

b. 꽃병 이 깨졌다.
Vase nom broke

In (15a) and (16a), the transitive verb watta, 깨뜨렸다 (broke), has two arguments: the external argument 'Taroo' (agent A) and the internal argument 'vase' (direct object O). In (15b) and (16b), the intransitive verb wareta, 깨졌다 (broke), has only one argument, i.e. 'vase' (subject S). The subject of intransitive verb behaves like the agent of a transitive verb but differently from the object of a transitive verb: A=S≠O (Tsunoda, 2009).

Moving on to the Altaic languages, these are highly agglutinative, a leading case being indicated by suffixes. Essentially, the case suffixes obey vowel harmony and vowel-consonant harmony; (17) and (18) are illustrations from Turkish.

(17) **Turkish dative case affixes (vowel harmony)**

(i) /-a/: attach to the lexemes that end with vowels /a/, /ɪ/, /o/, /u/, e.g. okula (go to).

(ii) /-e/: attached to the lexemes that end with vowels /e/, /i/, /ö/, /ü/, e.g. işe (drive to).

(18) **Turkish accusative case affixes (vowel-consonant harmony)**[3]

(i). /-yɪ/, /-yi/, /-yu/, /-yü/: attach to lexemes that end in vowels, e.g. gazeteyi (being reading).

(ii). /-ɪ/, /-i/, /-u/, /-ü/: attach to lexemes that end in consonants, e.g. sandviçi (want).

Like the Altaics, case in Tamil is also marked by suffixes. The Austronesian languages display an isolation character, in that no affixes or adpositional clitics are employed to render a case. To put it another way, the Austronesian languages have no case marking. Table 9 summarizes the case marking systems in Modern Japanese, Old Japanese, the Altaics, the Austronesians, Tamil, and the Korean languages.

Table 9. Case marking in Japanese, Altaics, Austronesians, Tamil and Korean

Family	Genera	Languages	Case marking
Altaic	Tungstic genus	Ewen	Suffix
		Nanai	Suffix
		Udihe	Suffix
		Manchu	Suffix
	Turkic genus	Kyrgyz	Suffix
		Khakas	Suffix
		Kazakh	Suffix
		Azerbaijani	Suffix
		Uzbek	Suffix
		Turkish	Suffix
		Turkmen	Suffix
	Mongolic genus	Mongolian	Suffix
	Nivkh genus	Sakha	Suffix
Austronesian	Western Malayo-Polynesian	Tagalog	No case marking
	Malayo-Sumbawan genus	Indonesian	No case marking
	Central Luzon genus	Kapampangan	No case marking
	Central Malayo-Polynesian genus	Lamaholot	No case marking
Dravidian		Tamil	Suffix
unknown		Modern Japanese	Postpositional clitics (particles)
		Old Japanese	Postpositional clitics
unknown		Korean	Postpositional clitics

The conclusion emerging from our examination is that Japanese shares a similar marking system with Korean. Altaic case marking differs from Japanese in two ways: (a) Japanese conveys case via particles, while the Altaic languages employ suffixes; (b) Modern Japanese case markers do not involve a harmonic rule, while the case markers of the Altaic languages follow vowel assimilation. The distinctions further extend to delicate matters. For instance, the Turkish comitative and instrumental case share the same marking, i.e. suffix -le, while in Japanese, the two are rendered by different particles. The Mongolian dative and locative case share the same marking, i.e. the suffixes -ᴀ (d), -aᴀ (ad), -иᴀ (id), -т (t), while in Japanese, they are differentiated by different particles. In Indonesian, passive voices are employed extensively. In sentences with active and passive voices, only the agents of passive sentences are marked, which has prompted linguists to assume that Indonesian is an ergative language (Cartier, 1979). Japanese, an accusative language, does not facilitate a syntactic correspondence with Austronesian languages.

Table 10. Word order in Japanese, Altaics, Austronesians, Tamil and Korean

Family	Genera	Languages	Word order	Adjective Noun	Preposition / postposition
	Tungstic genus	Udihe	SOV; SV; OV	Adjective-Noun	Postposition
		Manchu	SOV; SV; OV	Adjective-Noun	Postposition
	Turkic genus	Azerbaijani	SOV; SV; OV	Adjective-Noun	Postposition
		Uzbek	SOV; SV; OV	Adjective-Noun	Postposition
		Turkish	SOV; SV; OV	Adjective-Noun	Postposition
		Turkmen	SOV; SV; OV	Adjective-Noun	Postposition
	Malayo-Sumbawan genus	Indonesian	SVO; SV; VOS	Noun-Adjective	Preposition
	Central Luzon genus	Kapampangan	VS; VO; VOX	Noun-Adjective	Preposition
	Central Malayo-Polynesian genus	Lamaholot	SVO; SV; VOS	Noun-Adjective	Preposition

Dravidian		Tamil	SOV; SV; OV	Adjective-Noun	Postposition
unknown		Modern Japanese	SOV; SV; OV	Adjective-Noun	Postposition
		Old Japanese	SOV; SV; OV; XVO[4]	Adjective-Noun	Postposition; Preposition
unknown		Korean	SOV; SV; OV	Adjective-Noun	Postposition

Word order

To confirm our findings, this section looks at word order, specifically: (a) the order of subject, verb, and object; (b) the order of adjective and noun; and (c) the order of adposition and noun.

Incorporating the information from the World Atlas of Language Structure, we arrive at the following comparative results among the languages.

The Altaic languages, Tamil, Modern Japanese, and Korean are verb final, i.e. the adjective comes before the noun and the adpositional phrase comes after the noun (i.e. postpositional). The Austronesian languages, however, are verb initial, with the adjective coming after the noun and the adpositional phrase coming before the noun (i.e. prepositional). This study calculated the Euclidean distance via python with a focus upon: (a) the order of subject-object-verb; (b) the order of adjective-noun; and (c) the order of adposition-noun. Assuming that L_1 and L_2 as the vectors representing the compared languages, the syntactic distance between L_1 ($L_{1,1}$, . . ., $L_{1,n}$) and L_2 ($L_{2,1}$, . . ., $L_{2,n}$) would be determined through the formula:

(19)

$$d(L_{1,}L_{2}) = \sqrt{(L_{1,}1 - L_{2,}1)^2 + (L_{1,}2 - L_{2,}2)^2 + \cdots + (L_{1,}n - L_{2,}n)^2} = \sqrt{\sum_{i=1}^{n}(L_{1,}n - L_{2,}n)^2} \quad (5)$$

The following results are returned.

(20) **Distance regarding word order in Japanese and the languages in focus**:

(a). Japanese and Altaics: dJapanese-Altaic ≈ 0

(b). Japanese and Tamil: dJapanese-Tamil ≈ 0

(c). Japanese and Korean: dJapanese-Korean ≈ 0

(d). Japanese and Indonesian: dJapanese-Indonesian & Lamaholot ≈2.44

(e). Japanese and Kapampangan: dJapanese-Kapampangan ≈ 2.44

(f). Japanese and Lamaholot: dJapanese-Lamaholot ≈2.44

The syntactic distance between Japanese, Tamil, Korean, and the Altaic languages are equally short, which in turn suggests that Japanese, Tamil, Korean, and the Altaic languages share similarities in regard to word order.

This section has drawn a picture of the morphologic distances between Japanese and potentially related languages. Drawing the strands together, we can conclude that Japanese and Korean are, for the most part, closely related syntactically.

Phonologic Distances Between Japanese and Altaic, Austronesian, Tamil, and Korean

Another matter that serves to reveal the genealogy of Japanese involves assimilation. The discussion starts with vowel harmony.

Vowel harmony is a phonological process where all the vowels are required to be of the same class: VaCVbCVbC → VaCVaCVaC (e.g. front or back, rounded or unrounded, alveolar or postalveolar) in a stem, and continue to harmonize the affixes in derivation, inflection, or conveying a case (e.g. case in the Altaic languages is marked by suffixes). To be precise, vowel harmony entails the following variations (Rose & Walker, 2011):

(21) **Types of vowel harmony**

Backness harmony

Round harmony

Height harmony

Tongue root harmony

Vowel harmony is extensive in the Altaic languages. In the Turkic genus, two harmonic types are presented: front/back vowels and rounded/unrounded vowels (Kurtuluş, 2000). (22) is an illustration taken from Azerbaijani.

(22) **Turkic genus, Azerbaijani vowel harmony** (Backness Type & Round Type)

[+front] & [-rounded] vowels: / e, ə, i/

[+front] & [+rounded] vowels: / ö, ü/

[+back] & [-rounded] vowels: / a, ı/

[+back] & [+rounded] vowels: / o, u/

Mongolian has seven vowels, /i/ [i], /ü/ [u], /u/ [ʊ], /e/ [ɤ], /ö/ [o], /o/ [ɔ], /a/ [ɑ], and presents a tongue root harmony type: /a, ʊ, ɔ/ are pronounced with a retracted tongue root (RTR), are masculine, and tend to render a negative reading; /u, e, o/ are pronounced with an advanced tongue root (ATR), are feminine, and tend to convey a positive meaning. The remaining /i/ is neutral.

(23) **Mongolian vowel harmony** (Tongue Root Type)

RTR vowels: /a, ʊ, ɔ/, masculine, negative

ATR vowels: /u, e, o/, feminin, positive

Korean has 21 vowels (10 monophthongs; 11 diphthongs) and 19 consonants. The vowel harmony falls into the 'backness harmony' type, i.e. front vowels tend to be positive (ㅏ [a], ㅗ [o]), while middle vowels render a rough meaning (ㅓ [eo], ㅜ [u]).

(24) **Korean vowel harmony** (Backness Harmony Type)

Front vowels: ㅏ [a], ㅗ [o], positive

Middle vowels: ㅓ [eo], ㅜ [u], negative

Tamil has ten to 14 vowels (25). Among them, ten are short/long pairs; four are diphthongs.

(25) **Tamil vowels** (Table 11)

Table 11. Tamil vowels

	/a/	/e/	/i/	/o/	/u/
Short vowel	[ʌ]	எ [e]	இ [i]	ஒ [o]	உ [u]
Long vowel	ஆ [a:]	ஈ [e:]	ஐ [i:]	ஔ [o:]	[u:]

On the basis of Beckman's (1998) insights, Tamil middle vowels and round vowels appear only in initial syllables. Moreover, in spoken Tamil, it appears that /o/ and /e/ are lowered to [ɔ̩] and [ɛ̩] when they are followed by /a/. When long vowel ஈ [e:] is followed by a low vowel, it would be lowered to [æ][5]. Given this 'vowel lowering', perhaps we can arrive at an initial conclusion that spoken Tamil displays a harmony of vowel height.

In the Austronesian languages, vowel harmonization displays two patterns: height harmony and backness harmony. Robinson (1968), Kroeger (1992), Hurlbut (1981), Harris (1993), and Boutin (1993) show that a coastal dialect spoken in Malaysia, i.e. Tindal, has four vowels: /a/, /i/, /o/, /u/; /o/ changes to /a/ if it is followed by a syllable containing /a/. It should be noted that adjacent distance is not obligatory; to put it another way, consonants may intervene between /o/ and /a/. 'Backness harmony' is suggested. Moreover, Tindal shares a striking similarity with the Kagoshima dialect, of the southwest of Japan, in that the stress in Tindal and Kagoshima always falls on the penultimate syllable. As Hirayama (1936) puts it, in the Kagoshima dialect, the rule of penultimate stress is always followed, i.e. even if it is followed by a case particle, or forms a compound noun with multiple noun components. This finding is solid evidence for the Austronesian-origin hypothesis regarding the genealogy of Japanese.

The backness harmony type is seen in Chamorro, another Austronesian language, spoken in Guam and the Mariana islands. The progressive vowel assimilation is as follows: for a back vowel (/a/, /o/, /u/) in a root, when proceeded by a front vowel, the back vowel

would be replaced by the front vowel (/ä/, /e/, /i/) so as to match to the same tongue height; the proceeding front vowel can be rendered by articles, affixes (mark case), or a modifier, as specified in (26).

(26). **Vowel harmony in Chamorro** (on the basis of Topping's (1968) insights) (Table 12).

Table 12. Vowel harmony in Chamorro

Condition	(backness harmony)
Definite article i 'the'	/u/ → /i/
Definite, personal article si	/a/ → /ä/
Indefinite article ni	/o/ → /e/
Goal case marker -in- (infix)	/o/ → /e/
Directional case marker san- (prefix)	/a/ → /ä/
Prefix mi- 'lots of'	/o/ → /e/

Chamorro further bears height harmony: in morphemes with the structure CV_1CV_2, the height of the stressed vowel restricts the range of the height of the final vowel (Topping, 1968). The phenomenon of vowel harmony also extends to other Austronesian languages. For example, in Selayarese, /e/ and /o/ are lowered when followed by a syllable with /a/; essentially, adjacent distance is required between /e/, /o/, and /a/. In Kadazan (a language spoken in Malaysia), for vowel /a/ in a root, when followed by a syllable with /o/, /a/ would become /o/ (Hurlbut, 1988). This in turns suggests a harmony of height.

An important matter we have pointed out earlier, but which it is necessary to come back to at this point, is whether the eight vowels in Old Japanese, i.e. /a/, /e_1/, /e_2/, /i_1/, /i_2/, /o_1/, /o_2/, /u/[6], assign assimilation. We would like to recapitulate here that Old Japanese had three writing systems, resulting in vowels being conveyed by a variety of characters. (27)–(29) are vowels and corresponding characters based on Bjarke Frellesvig (2010).

(27) **/i/ in Old Japanese**

a. 伊i, 為wi, 知ti, 遅di, 之si, 自zi, 爾ni, 利ri

b. 比 pi_1, 鼻 bi_1, 美 mi_1, 支 ki_1, 祇 gi_1

c. 肥 pi_2, 備 bi_2, 未 mi_2, 貴 ki_2, 疑 gi_2

(28) **/e/ in Old Japanese**

a. 天te, 田de, 勢se, 是ze, 尼ne

b. 平 pe_1, 弁 be_1, 売 me_1, 家 ke_1, 牙 ge_1

c. 戸 pe_2, 倍 be_2, 米 me_2, 気 ke_2, 義 ge_2

(29) **/o/ in Old Japanese**

a. 於 o, 乎 wo, 富 po

b. 毛 mo_1, 古 ko_1, 刀 do_1, 度 do_1, 俗 zo_1

c. 母 mo_2, 其 go_2, 止to_2, 特do_2, 乃no_2

Arisaka (1934) puts forward a constraint on the combination of the vowels, known as Arisaka's law, cf. (30).

(30) **The law of vowel's occurrences in Old Japanese**

a. /o_1/ and /o_2/ cannot appear in a monological unit

b. /u/ seldom occur with /o_2/ in a monological unit

c. /a/ seldom occur with /o_2/ in a monological unit

Arisaka's law is considered a solid support to the hypothesis that Japanese originated in the Altaic languages. Another important work to mention is Ohno (1980), where Japanese is deemed to originate in four vowels: /a/, /i/, /u/, /ö/ (o_2). It was [/u/+/i_1/] and [/i_1/+/o_2/] that merged into /i_2/; [/i_1/+/a/] merged into /e_1/; [/a/+/i_1/] merged into /e_2/; and /u/+/a/ merged into /o_1/.

(31) **Old Japanese vowels (Ohno's regard)**

/u/+/i_1/ → /i_2/

/i_1/+/o_2/ → /i_2/

/i_1/+/a/ → /e_1/

/a/ +/i_1/ → /e_2/

/u/+/a/ → /o_1/

We are obliged to point out that such mergings are fossils and thus cannot be alleged as examples of harmonization. It is at this point

where our greatest difficulty surfaces: does Old Japanese truly bear vowel harmony? A corpus-based investigation might help us to answer the question. As harmony processes can be adjacent as well as widely separated, this study selects the mono noun and the compound noun as candidates. The data is drawn from The Japanese Lexicon: A Rendaku Encyclopedia. Fifty tokens of each group are extracted randomly. The findings are summarized in Table 13.

Table 13. Cooccurrences of vowels in mono noun and compound noun in Old Japanese

Mono noun (N_1)		**Compound noun (N_1-N_2)**	
Combination of vowels	Tokens	Combination of vowels	Tokens
/a/+/o_2/	4	/a/+/o_2/	1
/i_1/+/e_1/	1	/a/+/i_1/+/e_1/	8
/a/+/a/	9	/a/+/i_1/+/o_2/	2
/i_1/+/o_2/	1	/a/+/e_1/+/i_1/+/o_2/	2
/a/+/u/	8	/a/+/u/	5
/o_1/+/o_1/	3	/a/+/e_1/+/u/	3
/a/+/i_1/	7	/a/+/i_1/	10
/u/+/i_1/	1	/a/+/i_1/+/o_1/	1
/u/+/o_2/	1	/a/+/e_1/+/o_2/	1
/u/+/e_2/	1	/a/+/i_1/+/u/	6
/u_1/+/i_1/	1	/a/+/u/+/i_1/+/o_2/	1
/o_1/+/i_1/	1	/a/+/o_1/	3
/a/+/e_1/	1	/a/+/u/+/i_2/+/e_2/	1
/a/+/i_1/+/o_1/	1	/a/+/e_1/	3
/a/+/i_1/+/o_2/	1	/a/+/a/	2
		/a/+/u/+/e_1/+/e_2/	1
		/a/+/o_1/+/i_1/	1
		/a/+/o_1/+/e_1/+/i_1/	2
		/a/+/i_1/+/e_1/+/e_2/	1
		/a/+/o_2/+/e_1/+/i_1/	1
		/e_1/+/i_1/	2

The finding brings to light that the combination of [O_1 + O_2] in either a mono noun or a compound noun is ruled out, which supports Arisaka's law (1934). However, in regard to Arisaka, the ill-formed [/u/+/o_2/] and [/a/+/o_2/] are well accepted in a mono noun

and a compound noun. This brings out the idea that harmony in Old Japanese was not salient, as in the Altaic languages and Korean; maybe the hypothesis of vowel harmony in Old Japanese ought to be rejected.

Having said this, vowel-consonant harmony is detected in the sequential voicing phenomenon in Old Japanese. Sequential voicing refers to when, during a process of forming a nominal compound, the second constituents (N_2) rendered by aspirated consonants become voiced. This study shows that there is an assimilation between Old Japanese vowels, /a/, /e_1/, /e_2/, /i_1/, /i_2/, /o_1/, /o_2/, and /u/, and the aspirated consonants, /k/, /s/, /t/, and /h/.

(32) **Vowel-consonant harmony in sequential voicing in Old Japanese**

/a/ is most likely to form with N_2 where the initial consonant is /k/, /p/ or /t/

o_1/ and /o_2/ are both likely to combine with a voiced consonant /k/

It is unlikely that vowel /a/ will invite a voiced '/s/-initial' N_1'

/e_1/ is more likely to invite a voiced consonant than /e_2/

/e_1/ does not invite a voiced /p/

/e_2/ does not yield a voiced /s/

/i_1/ is likely to take a voiced consonant than /i_2/

/k/ and /t/ are most likely to be voiced when combining with N_1 ending with /i_1/

/i_2/ does not invite a voiced /p/, /s/ or /t/

/u/ never results in a voiced /s/

The vowel-consonant harmony outside Japanese is seen in Turkish accusative case affixes (illustration (18)) and Madurese (Austronesian family). Madurese has 27 consonants and eight vowels: (a) front unrounded: /i/, /ɛ/; (b) back rounded: /u/, /ɔ/; (c) high back unrounded: /i̱/, /ə/; and (d) low back unrounded: /ɤ/, /a/. The following vowel-consonant harmony is confirmed: (a) consonants of voiced stops, voiceless aspirated stops tend to cooccur with vowels /i/, /ɤ/, /u/; and (b) consonants of voiceless unaspirated stops, nasals, and voiceless fricatives are likely to occur with vowels /a/,/ɛ/, /ɔ/ (Anderson, 1991). It should be pointed out that the harmonization differs in that Japanese and Madurese vowel-consonant harmony arrives at a lexical level, while Altaic vowel-consonant harmony is facilitated at a morph-syntactic level, suffixation. Table 14 outlines the patterns of vowel harmony in the Altaics, the Austronesians, Tamil, Japanese, and the Korean language

Table 14. Phonologic distances between Japanese, Altaics, Austronesians, Tamil and Korean

Family	Genera	Languages	Vowel harmony	Vowel-consonant harmony
Altaic	Turkic genus	Azerbaijani	Backness harmony & Round harmony	○, morph-syntactic level
	Mongolic	Mongolian	Tongue Root harmony	
Dravidian		Tamil	Height harmony	
unknown		Modern Japanese	×	○, lexicon level
		Old Japanese	×	○, lexicon level
unknown		Korean	Backness Harmony	
Austronesian	Malayo-Sumbawan	Madurese	×	○, lexicon level
	Malayo-Polynesian	Selayarese	Height harmony	×
	North Borneo	Kadazan	Height harmony	×
	Chamorro	Chamorro	Height harmony & backness harmony	×
	North Borneo	Tindal	Backness harmony	×

It turns out that Tamil shares similarities with the Austronesian languages in displaying a 'height harmony'. Korean shares a similarity with the Altaic languages and the Austronesian languages in presenting 'backness harmony'. Japanese shares similarities with the Altaic languages in 'vowel-consonant harmony'. However, it ought to be pointed that the phenomenon is seen in a different part of speech: in Japanese, vowel-consonant harmony is detected in the compound noun; in the Altaic languages, it is observed in case marking.

The phonology in Modern Japanese and Modern Korean is worthy of attention. The two bear different phonological structures. Japanese is CV (consonant-vowel). Korean has two types: CVC (consonant-vowel-consonant) and CV. The vowel system in Old Japanese, however, comes to resemble Korean's monophthong. A corresponding list is provided in (29).

(29) **Korean monophthong corresponds to Old Japanese vowels** (Table 15)

Table 15. Korean monophthong corresponds to Old Japanese vowels

	[a]	[ɔ]	[o]	[u]	[ɯ]	[i]	[ɛ]	[e]
Old Japanese	阿a	母 mo_2, 其 go_2, 止to_2, 特do_2, 乃 no_2	毛 mo_1, 古 ko_1, 刀 do_1, 度 do_1, 俗 zo_1	宇u	肥 pi_2, 備 bi_2, 未 mi_2, 貴 ki_2, 疑 gi_2	比 pi_1, 鼻 bi_1, 美 mi_1, 支 ki_1, 祇 gi_1	戸 pe_2, 倍 be_2, 米 me_2, 気 ke_2, 義 ge_2	平 pe_1, 弁 be_1, 売 me_1, 家 ke_1, 牙 ge_1
Modern Korean	/a/ ㅏ	/ɔ/ ㅓ	/o/ ㅗ	/u/ ㅜ	/ɯ/ ㅡ	/i/ ㅣ	/ɛ/ ㅐ	/e/ ㅔ

The similarity of vowel systems thousands of years ago suggests the idea that the Japanese language derived from the language spoken on the Korean peninsula.

With this in mind, we assume that Japanese does not have salient similarities with the Austronesian languages in regard to phonology. The phonologic distinctions between Japanese and the Austronesian languages further tie to the fact that the Austronesian languages are stress languages, while Japanese is moraic.

Conclusion

The main aim of this study was to uncover how Japanese is related to the Altaic languages, the Austronesian languages, the Dravidian language Tamil, and Korean. The chi-square homogeneity test and Euclidean distance were a great help in achieving this goal, enabling us to calculate the morphologic, syntactic, and phonologic distances between the languages.

Morphologically, in light of preferences of causative/inchoative verb alternation patterning and morphemes that convey the alternation, it is postulated that Japanese and Korean are for the most part close; the language with the second closest features to Japanese is the Altaic language Turkmen.

Syntactically, Japanese and Korean share a similarity in rendering case with particles; the Altaics and Tamil convey case via suffixes; case in the Austronesian languages is conveyed by prefixes. Furthermore, Altaic language case markers follow vowel harmony and vowel-consonant harmony, while Japanese case markers do not involve a harmonic rule. Japanese, Tamil, Korean, and the Altaic languages share the same word order: 'verb final' (SOV). Austronesian languages are mostly verb initial (VOS). Morpho-syntactic distinctions between the Austronesian languages and Japanese further extend to language typology, i.e. Indonesian has an ergative language character (Cartier, 1979), while Japanese is an accusative language.

Phonologically, Tamil and the Austronesian languages share a resemblance in harmony of vowel height. Korean, the Altaic languages, and the Austronesian languages have similarities in the harmony of vowel backness. Japanese, the Altaic languages, and the Austronesian language Madurese display vowel-consonant harmony. It should be pointed out that the harmonization in Japanese and Madurese reaches the lexical level (nominal compounding), while in the Altaic languages, vowel-consonant harmony is facilitated by morpho-syntax (suffixation).

Pulling these strands together, we can conclude that Japanese is most closely related to Korean. Perhaps it would not be unsound to note that the Austronesian language Tindal and the Kagoshima dialect (Japanese) bear a penultimate stress.

This study may not be able to account for all linguistic resources in all languages, but we hope that the conclusion reached by this study can be tested by other grammatical elements.

References

Anderson, S. (1991). Proceedings of Arizona Phonology Conference: *Vowel–consonant interaction in Madurese*. In Ann, Jean and Yoshimura, Kyoko (eds.). Tuscon, AZ. Department of Linguistics, University of Arizona.

Arisaka, H. (1934). Kodai Nihongo ni okeru onsetsuketsugoo no hoosoku [The law of combining vowels and consonants in Old Japanese]. *Kokugo to kokubungaku [National language and literature], 11*(1), 80–92.

Baker, M.C. (1988). *Incorporation: A Theory of Grammatical Function Changing.* Chicago: University of Chicago Press.

Beckman, J.N. (1998). *Positional Faithfulness.* Doctoral dissertation, University of Massachusetts, Amherst.

Bjarke F. (2010). *A History of the Japanese Language.* Cambridge University Press.

Boutin, M.E. & Pekkaneri, I. (1993). *Phonological Descriptions of Sabah Languages.* The Rosetta Project: A Long Now Foundation Library of Human Language.

Cartier, A. (1979). De-voiced transitive verb sentences in formal Indonesian. In F. Plank (ED.), *Ergativity: Towards a theory of grammatical relations* (pp. 87-161). New York: Academic Press.

Dryer, M.S. (2005). Position of Case Affixes. In M. Haspelmath, M.S. Dryer, D. Gil & B. Comrie (Eds.), *The World Atlas of Language Structures.* Munich: Max Planck Digital Library.

Hashimoto, S. (1950). *Kokugo Onin no hensen [The transformation of Japanese phonology].* Hashimoto, Shinkichi Hakase Cyosakusyuu [Collection of Hashimoto, Shinkichi Doctor's extraordinary work].

Haspelmath, M. (1993). *A Grammar of Lezgian.* (Mouton Grammar Library, 9.) Berlin: Mouton de Gruyter.

Hattori, S. (1959). *Nihongo no Keitoo.* Iwanami Press.

Hattori, S. (1975a). Joodai Nihongo no boin taikei to boin cyoowa [The vowel system and vowel harmony in Old Japanese]. *Gengo [Language]*, 5-6.

Hattori, S. (1975b). Joodai Nihongo no boinonso wa muttsu deatte, yattsu dewanai [There is only six vowels in Old Japanese, not eight]. *Gengo [Language],* 5-12.

Hirayama, T. (1936). Minami Kyuushuu akusento no kenkyuu [On the accent of South

Kyuusyuu]. *Hoogen [Dialectology]* 6.4; 6.5.

Hurlbut, H.M. (1981). Morphophonemics of Labuk Kadazan. In Aɪo Gonzalez & D. Thomas (Eds.), *Linguistics across continents: Studies in honor of Richard S. Pittman* (pp. 46-53). Manila: Summer Institute of Linguistics.

Hurlbut, H.M. (1988). *Verb Morphology in Eastern Kadazan. Pacific Linguistics Series B – No. 97.* Canberra: The Australian National University.

Kanno, H. (1988). *Bunpogaisetsu [An introduction to Korean grammar]*. Kosumosu Choowa Jiten [Kosumosu Korean-Japanese Dictionary]. Hakusuisha Press.

Kim, M. (1970). Kokugogaku ni tsuite [On Korean case system]. *Kokugokokubungaku [National Language and Linguistics]. 49*, 50.

Kröger, F. (1992). *Buli - English Dictionary: With an Introductory Grammar and an Index English - Buli.* Münster / Hamburg: LIT Verlag.

Kurtuluş, Ö. (2000). *Elementary Azerbaijani. Türk Dilleri Arastirmalari Dizisi.* Santa Monica, California – Istanbul.

Matsumoto, K. (1975). Kodai Nihongo Boinsoshikikoo - Naiteki Saiken no kokoromi – [The vowel system in Old Japanese: an attempt to internal reconstruction]. *Kanazawa Daigaku Hoobungakubu Ronshuu Bungakuhen [Studies and essays by the Faculty of Law and Literature, Kanazawa University. Literature], 22.*

Matsumoto, K. (1975a). Nihongo no boin soshiki [Japanese vowel system]. *Gengo [Language], 5-6.*

Matsumoto, K. (1975b). Manyogana no オretsu Koo/Otsu ni tsuite [On Manyogana's オ Line: the distinction of Koo and Otsu type]. *Gengo [Language],* 5-11.

Matsumoto, K. (2007). *Sekaigengo no naka no nihongo: nihongo keetooron no aratana chihee [Japanese among cross linguistics: a new view on Japanese genealogy]*. Sanseedo Press.

Miler, R.A. (1971). *Japanese language and other Altaic languages.* Chicago university Press.

Obashi, M. & Tanaka, K. (2011). Proceedings of the 17th Meeting of Natural language processing of Japan*: Suuriteki shuhoo o mochiita Nihongo no keitooni kansuru koosatsu [A mathematical linguistic approach to Japanese genealogy]*. Gengoshorigakkai.

Ohno, S. (1957). *Nihongo no kigen [The origin of Japanese language]*. Iwanami Press.

Ohno, S. (1980). *Nihongo no keetoo [Japanese genealogy]*. Shinbundoo Press.

Ohno, S. (1981). *Nihongo to Tamirugo [Japanese and Tamil]*. Shinchoo Press.

Robbeets, M., Bouckaert, R., Conte, M., Savelyev, A., Li, T., An, D.I., Shinoda, K.I., Cui, Y., Kawashima, T., Kim, G., Uchiyama, J., Dolińska, J., Oskolskaya, S., Yamano, K.Y., Seguchi, N., Tomita, H., Takamiya, H., Kanzawa-Kiriyama, H., Oota, H., … Ning, C. (2021). Triangulation supports agricultural spread of the Transeurasian languages. *Nature, 599*(7886), 616-621. https://doi.org/10.1038/s41586-021-04108-8.

Robinson, L.C. (2006). Proceedings from Tenth international conferences on Austronesian linguistics. Linguistic Society of the Philippines and SIL International: *Vowel harmony in Borneo: an examination of vowel changes in Tindal Dusun.* Palawan, Philippines.

Rose, S. & Walker, R. (2011). Harmony Systems. In J. Goldsmith, J. Riggle, A. Yu (Eds.). *Handbook of Phonological Theory* (2nd ed.). Blackwell.

Sakiyama, O. (2012). Japanese as a mixed language: sound law and semantic change from Proto Austronesian to Ancient Japanese. *Research report of National Museum of Ethnology, 36*(3): 353-393.

Topping, D.M. (1968). Chamorro vowel harmony. *Oceanic Linguistics, 7*(1), 67-79.

Old Japanese resources

Kojiki kayō: 112 poems; approx. 2527 words, written in hentai-kanbun 'variant Chinese',

Completed year: AD. 712, genre: historical book that supports the emperor's sovereignty.

Nihon shoki kayō: 133 poems; approx. 2444 words, written in junsei-kanbun 'classical Chinese', Completed year: AD. 720, genre: Authentic historical book.

Man'yōshū: 4685 poems; approx. 83706 words, written in man'yōgana, Completed year: AD. 759, genre: poetry.

Appendix

[1] Japanese morphological analysers further include Janome, Chasen, KyTea, and Juman.

[2] One may argue that the similarity in morphology in Korean and Japanese perhaps resulted in the Japanese annexation (1910–1945). Although Japanese was made an official language, leading to a mixed Hanja-Hangul script during the annexation, the effects are limited to the lexicon (lexical roots are written in Hanja), with the Korean grammatical forms retained. The morphology-driven feature in Korean was established long before the annexation.

[3] Vowel harmony is worth commenting on. We will return to it in Section 5.

[4] X refers to oblique.

[5] This information is drawn from https://www.reddit.com/r/linguistics/comments/nrg00d/vowel_harmony_in_tamil/

[6] In Early Middle Japanese, when the pure phonetic script kana was invented, vowels merged into: /a/, /i/, /u/, /e/, and /o/.