

## A FORMAÇÃO LINGUÍSTICA DO NORDESTE ASIÁTICO: múltiplas origens das famílias coreânica e japônica<sup>1</sup>

*Marcus Tanaka de Lira*

Universidade de Brasília - UnB, Brasília/DF, Brasil

<https://orcid.org/0000-0003-1750-8727>

tanakadelira@unb.br

### RESUMO

O Nordeste Asiático, como região, hospeda um número de línguas cujas relações com outras famílias linguísticas do globo continuam obscuras até hoje. Apesar dessa descrição ser frequentemente usada para se referir às ditas línguas Paleossiberianas, um conjunto heterogêneo de línguas isoladas e famílias linguísticas rasas do Extremo Leste Russo, poderíamos bem dizer o mesmo de seus vizinhos mais amplamente difundidos ao sul – a dizer, as famílias linguísticas coreânica e japônica (das quais o coreano e o japonês são seus membros mais famosos, respectivamente). Geralmente ligadas à família linguística altaica, tentativas frustradas de se reconstruir uma protolíngua convincente e resolver questões relacionadas às ramificações internas, somadas às várias dúvidas sobre as origens das duas primeiras línguas, levantam dúvidas quanto a essa conexão. As páginas a seguir apresentam um resumo multidisciplinar, coletando dados de áreas como a genética e a arqueologia, do que (não) sabemos até agora sobre a formação linguística da península coreana e do arquipélago japonês.

**PALAVRAS-CHAVE:** Coreano; Família coreana; Hipótese altaica; Linguística; Linguística histórica; Nordeste asiático; Japonês.

## THE LINGUISTIC FORMATION OF NORTHEAST ASIA: Multiple Origins of the Koreanic and Japonic Families<sup>2</sup>

### ABSTRACT

Northeast Asia, as a region, is home to a number of languages whose relationships to other language families of the world remain murky to this day. Although this description is often used to describe the so-called Paleosiberian languages, a motley assortment of language isolates and shallow language families from the Russian Far East, it could well be applied to their more widely spoken neighbours to the South – namely, the Koreanic and the Japonic language families (of which Korean and Japanese are their most famous members, respectively). Often linked to the Altaic language family, failed attempts to reconstruct a well-accepted protolanguage and resolve issues related to internal branching, compounded with the many doubts regarding the origins of the Koreanic and Japonic families, raise doubts regarding this connection. The following pages present a multidisciplinary summary, trawling for data in fields such as genetics and archaeology, of what we (do not) know so far about the linguistic formation of the Korean peninsula and the Japanese archipelago.

**KEYWORDS:** Altaic hypothesis; Historical linguistics; Japanese; Japonic family; Korean; Koreanic family; Linguistics; Northeast Asia.

---

<sup>1</sup> Esta publicação é uma tradução abreviada, com ligeiras modificações quando consideradas necessárias para um público de língua inglesa, de Tanaka de Lira (no prelo) publicada pelo Hon No Mushi Journal.

<sup>2</sup> This publication is an abridged translation, with slight modifications where deemed necessary for an English speaking audience, of Tanaka de Lira (in press) published by the Hon No Mushi Journal.

## LA FORMACIÓN LINGÜÍSTICA DEL NORESTE DE ASIA: múltiples orígenes de las familias coreana y japonesa<sup>3</sup>

### RESUMEN

El noreste de Asia, como región, alberga una serie de idiomas cuyas relaciones con otras familias lingüísticas en todo el mundo siguen sin estar claras hasta el día de hoy. Aunque esta descripción se usa a menudo para referirse a las llamadas lenguas paleosiberianas, una colección heterogénea de lenguas aisladas y familias de lenguas superficiales del Lejano Oriente ruso, bien podríamos decir lo mismo para sus vecinos más extendidos al sur: a saber, las familias lingüísticas coreana y japonesa (de las cuales el coreano y el japonés son sus miembros más famosos, respectivamente). Generalmente vinculado a la familia de lenguas altaicas, los intentos frustrados de reconstruir una protolengua convincente y resolver cuestiones relacionadas con las ramificaciones internas, sumados a las diversas dudas sobre los orígenes de las dos primeras lenguas, plantean dudas sobre esta conexión. Las siguientes páginas presentan un resumen multidisciplinario, recopilando datos de áreas como la genética y la arqueología, de lo que (no sabemos) hasta ahora sobre la formación lingüística de la península de Corea y el archipiélago japonés.

**PALABRAS CLAVE:** Coreano; Familia coreana; Hipótesis altaica; Lingüística; Lingüística histórica; Noreste de Asia; Japonés.

Artigo submetido ao sistema de similaridade

Submetido em: 24/10/2022 – Aprovado em: 25/02/2023 – Publicado em: 31/03/2023

Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional.

---

<sup>3</sup> Esta publicación es una traducción abreviada, con ligeras modificaciones cuando se considera necesario para un público anglófono, de Tanaka de Lira (en prensa) publicado por la revista Hon No Mushi Journal.

## 1 - Introduction

Linguistically, Northeast Asia presents a quite peculiar scenario: despite being home to two of the most widely spoken languages in the world, the Korean and Japanese languages, there are many lingering doubts about the origins of these languages and their relation to other language families in the continent and beyond (SHIBATANI, 1990, p. 94; TRANTER, 2012; LEE e RAMSEY, 2011, p. 13-30). These difficulties in finding genetic relations with other language families, not an uncommon situation for threatened languages, are much rarer among more widespread languages, there being no analogous situation among other world languages with more than 10 million speakers (ETHNOLOGUE, 1992; UNESCO, 2010).

The inconclusive status of these languages becomes even more relevant in the regional context wherein they are found, being neighbour to the so-called "Paleosiberian Languages" spoken further north: a motley assortment of isolated languages and shallow language families without any recognised genetic relationship among themselves and whose commonality is their location (in east and northeast Siberia and, occasionally, northern Japan when the Ainu language is included) and do not belong to larger and more consolidated language families that are relative newcomers in the region (COMRIE, 1981, p. 238-279). Rather than exceptional cases in the region, the Korean and Japanese languages are emblematic reminders of the hardships in understanding the origins and relationships of the languages in Northeast Asia.

Owing to the difficulties present in the identification of the relationships and origins of the languages in the region, more and more linguists and other researchers have resorted to a multidisciplinary approach to try and elucidate the relationship of the Korean and Japanese languages to other language families, such as Robbeets, Bouckaert, et al. (2021), which sought to cross language data with information about population genetics and archaeological artefacts. In spite of the lack of consensus about the interpretation of the facts, this partnership among different fields of knowledge has been beneficial to a better understanding of how the language formation of this corner of the world took place.

The lack of material in Portuguese<sup>4</sup>, for its part, makes the insertion of incipient lusophone researchers in the fields of Korean and Japanese studies harder in studies about the origins of these languages. Particularly at the University of Brasilia there has been growing interest in diachronic Asian linguistics, with works such as Araújo (2018) and Resende (2019) representing the first pulses of research in the field. With the broadening of access to materials about Korean and Japanese languages in the university, it is hoped that a brief panorama about the linguistic history of the region may serve as tool and inspiration for future investigations in the field.

Far from proposing a solution to the problem of the origin of the Korean and Japanese languages, therefore, the following pages seek to offer a summary of ongoing research about the topic and some paths it may take in the near future. (...)

---

<sup>4</sup> This work was originally written in Portuguese, catering to Brazilian undergraduate students interested in diachronic Asian linguistics. It is here reproduced in English in an effort to showcase the current research carried out at the Centre for Asian Studies (NEASIA/CEAM) for wider audiences.

The first problem language research about the linguistic formation of Northeast Asia runs into is the relative dearth of data. Both in Korean and in Japanese, to focus on the most well documented languages in the region, the first documents range from about a thousand and three hundred years ago (in Japan) to eight hundred and seventy years ago (in Korea). The fact that both the Korean peninsula and the Japanese archipelago have been nearly monolingual for a while have only decreased the variety of data which can be used in the comparative method for reconstructions.

Worse still, the first documents in both languages were written in Classical Chinese, a typologically and genetically unrelated language from the languages of the region, whose script not always make clear to contemporary researchers the phonetic value its ideograms may have had when used phonologically.

In order to sketch a portrait of how the linguistic formation of the region took place, we may focus essentially on two difficulties:

1. What are the origins and genetic relations of the languages in the region? Do they show any sort of mutual relationship, if any?
2. What is the role contact between different peoples may have exerted in the formation of the region? Did the relative proximity lead only to loanwords and convergence areas, or could any of the languages be the result of a contact phenomenon between two different language families?

These two difficulties complement each other in a way, with the first one treating mainly the interlinguistic relations and the second one focusing on matters referring to the internal formation of these languages, hereby considered intralinguistic relations.

## 2 - Intralinguistic Relations

As mentioned in the previous section, the Korean peninsula and the Japanese archipelago have approached a scenario of complete monolingualism with very few exceptions: the languages of Jeju, in the Republic of Korea (South), and the so-called Yukjin dialect, spoken in the People's Democratic Republic of Korea (North) and in China (Vovin, 2013), which may be considered independent languages, albeit close, from Korean; and, in the Japanese archipelago, besides Ainu which has no proven relationship to the Japanese language and is severely threatened, there are the Ryukyu languages, belonging to the Japonic family. All these less spoken languages, however, are in an extremely vulnerable situation facing imminent extinction (ETHNOLOGUE, 1992; UNESCO, 2010).

It is possible that these regions may have had a wider array of languages in the past, as will be seen below, and the discussions about language variation are usually centred in the period prior to the Silla Unification, in Korea, and the Jomon, Yayoi and Kofun migrations, in Japan.

### 3 - Settling of the Korean Peninsula

Little is known about the beginnings of human settlement in the Korean peninsula, with the first traces pointing to the presence of early hominids, such as *homo erectus*, dating around hundreds of millions of years (SETH, 2011, p. 11; KIM, 2012, p. 1) and the first signs of *homo sapiens sapiens* presence dating from the palaeolithic period (from around 100 thousand to 40 thousand years ago) (PRATT, 2006, p. 29). Naturally, nothing is known about the language spoken by these early settlers, due to the lack of methodological tools that allow us to reconstruct such an ancient language in a satisfactory manner.

The origins of a local culture seen as being continuous with contemporary history are generally dated from 6000 to 2000 AC, either with the arrival of migrants from southern Siberia which supplanted the original populations throughout the peninsula, or due to the submersion of what could have been evidences of previous occupation due to rising sea levels and poor preservation (YI, 2015, p. 588).

A first explanation to the appearance of this culture is the hypothesis of a “double wave”, in which Paleosiberian populations could have been replaced by a more numerous Tungusic migratory wave with knowledge of bronze working. The caveats to this hypothesis is that there are archaeological artifacts, such as projectiles and pottery found in Jeju Island which, upon showing the continuous presence of the first arrivals to the peninsula, may go against this possibility (YI, 2015, p. 587-9). The question would be, therefore, to understand up to what point previous migrations affected the constitution of the Korean population and how it may have affected the language.

Undoubtedly, from 5000 AC onwards, pottery artefacts along the Cheongcheon and Han rivers adorned with parallel lines as if combed, as well as middens, appeared in the Korean coast (SETH, 2011, p. 11; KIM, 2012, p. 4; PRATT, 2006, p. 30). This period, Jeulmun, is named after the pottery, and is frequently compared to the pottery of neighbouring regions, such as Siberia and Japan, reflecting a continuous contact with other peoples, but without offering many clues regarding the relevance of such interactions in the formation of the population and their language. All that is known is that the population in the peninsula does not seem to have increased to justify the belief that there was the arrival of a new migratory wave (YI, 2015, p. 588).

One may add to the list of contacts of the local population with neighbouring peoples the transition to a more sedentary lifestyle with the appearance of agriculture as the basis for subsistence, initiated in the region with the domestication of millet around 4000 years ago, possibly as the result of interactions with other neolithic cultures from Northeast China and Southeast Russia (LEE, 2011, p. S307; SETH, 2011, p. 13). It is possible to say that, starting from 1500 AC in North Korea and 1000 in South Korea, up to 300 AC, there was the emergence of the period of simple pottery, or Mumun (PEARSON, 1977, p. 1243; SETH, 2011, p. 13), coinciding the adoption of rice in agriculture.

This shift mentioned in the previous paragraph seems to have happened thanks to the emergence of the first recognisably Korean tribal federation, Gojoseon, whose history is still somewhat obscure, but which was located to the north and to the east of the closest Chinese territories, with whom hostilities would often break out from their initial defeat to Chinese troops in 109 BC (SETH, 2011, p. 17; PRATT, 2006, p. 32-4) until the invaders were driven out of the peninsula in the early 2<sup>nd</sup> century (PRATT, 2006, p. 37; SETH, 2011, p. 24). It is through the reports of these Chinese forces that often clashed with the locals that we have a description of what the Korean peninsula was like.

The Chinese officials from the region reported that there was not a cohesive culture spoken throughout the peninsula, and they would lump them all and their Japanese neighbours as “Oriental Barbarians” (SETH, 2011, p. 20). Among the Korean peoples, the Chinese would recognise the existence of different barbarians and there were reports of at least three different language groups: Suksin (肅慎) and Buyeo (夫餘), scattered along Manchuria, South Siberia and the northern part of the peninsula, and Han (韓), who occupied its southern tip (LEE e RAMSEY, 2011, p. 34). Each group was divided into different languages (Suksin: Suksin, Emnu, Mulgil e Malgal; Buyeo: Buyeo, Goguryeo, Okjeo e Ye; Han: Mahan, Jinhan, Byeonhan), which will be herein considered languages due to the fact that, apart from the Mahan language within the Han group, the Chinese would say that these languages were mutually intelligible varieties among the member of each group (LEE e RAMSEY, 2011, p. 34-6).

This cultural plurality probably remained in place during the most part of the period from 57 BC to 668 AD, when the Korean peninsula found itself divided into three greater kingdoms, Goguryeo to the north, Baekje to the southwest, and Silla to the southeast, and by somewhat smaller entities, such as the Gaya Confederation to the south. Goguryeo belonged to the Buyeo language group, as seen above, whereas Baekje, Silla, and Gaya rose from the Han territories where the Mahan, Jinhan and Byeonhan previously resided. The Suksin group did not establish any unified chiefdom of which we are aware of, with the caveat that our sources all come from foreigners or from Koreans that lived much later, such as the Samguk Sagi, from 1145 (LEE e RAMSEY, 2011, p. 37; SETH, 2011, p. 93-4).

And thus, two kinds of problem arise: one with the Samguk Sagi, and another one with descriptions handed down to us by other peoples.

The first one is that, despite of Samguk Sagi's author's citation of previous works, these sources Pu-sik Kim quoted have all been lost. This is not exactly unexpected as the Three Kingdom Period had ended almost five hundred years prior, after a renewed Chinese attack to the peninsula then repelled by Silla, which drove invaders away and incorporated Goguryeo and Baekje, leading to the political and linguistic unification of the peninsula (SETH, 2011, p. 45; LEE e RAMSEY, 2011, p. 47). We cannot, therefore, compare its references with older local sources.



Furthermore, still related to this first problem, the texts were usually written in classical Chinese, as previously mentioned. Sometimes the Chinese ideograms were used by their phonological value, the so-called *hun* reading (LEE e RAMSEY, 2011, p. 38) - which naturally varied among different languages. And, if there was a more detailed description about the diverging pronunciations of the languages spoken by the three interacting peninsular groups with the Silla language, they were not included in the reports.

Nor can the foreign sources be taken as true at face value. Of immediate importance to this topic is the affirmation found in the *Nihon Shoki* that the Japanese Jingu empress had conquered Mimana, a part of the Korean peninsula usually identified as the Gaya Confederacy, which was later on allegedly handed to Baekje and then subdued by Silla (SETH, 2011, p. 32; LEE e RAMSEY, 2011, p. 36). What could have been an important record of Nipponic presence in the peninsula does not seem credible, as there are no other facts to corroborate this Japanese conquest in the continent.

This does not mean there is no good reason to believe that the Japonic language family was not present until then in the Korean peninsula, being necessary to make a side note about the Buyeo language spoken by the Goguryeo.

Despite all problems with the *Samguk Sagi*, a long section of the document describes how king Gyeonggeok in 757 carried out naming changes of toponyms (LEE e RAMSEY, 2011, p. 38), explaining how the old Goguryeo names were changed.

Many old toponyms from Goguryeo were written in the local *hun* reading of the Buyeo language, and in order to know how these names were pronounced, we would have to know how the locals read these Chinese ideograms. As previously mentioned, no such information is available. We can resort, however, to the different *hun* readings used for the same concept, thus comparing these cross-references and suppose how they might have been pronounced.

Places with names related to water (水) were usually written in Goguryeo names as 買, 美 or 彌, used only for their phonological value, allowing for reconstructions such as \*mæ:j, \*mi or \*mji. Valleys (谷), for their part, were usually written as 旦, 頓 or 吞, which in turn may have been read as \*tan, \*twon or \*thwon (LEE e RAMSEY, 2011, p. 39). These readings are immediately recognisable to Japanese speakers, as the native reading of these ideograms in the archipelago are nowadays /mizu/ and /tani/. The numbers 3, 5, 7 and 10 were also usually written as 蜜, reconstructed as \*mil, 干次 resulting in \*wucha, 難隱 giving us \*nanun, and 德 which could have been read as \*tek. The numbers 3, 5, 7 and 10 in Old Japanese were mi, itu, nana and topo, suggesting a similarity between the languages. It is possible that a Japonic language was not only spoken in the peninsula as we may hypothesise it belonged to the Buyeo group, putting this potential member of the family in the north of the Korean peninsula and in parts of Manchuria and Mongolia.

There are many difficulties involved in this reconstruction, and similar arguments can be made for different language families of the region, as there are toponyms of apparent Tungusic origin in the old provinces of Hamgyeong and Pyeongan, today located in North Korea (TOH, 2005, p. 19). There is also a great ongoing discussion regarding the relationship of the languages of the Goguryeo, Baekje, and Silla kingdoms with both language families, among others.

Beckwith (2005), for example, argues that the language of Goguryeo (as well as the language spoken by the court in Baekje, stemming from Goguryeo) and the Japanese language present a common ancestor, leading to what the author dubs the Nipo-Goguryeolic; meanwhile, the contemporary Korean language, derived from a Han language, as well as the language from Baekje commoners, would not present genetic relationship to the languages of the archipelago.

Janhunen (2005), proposes that Goguryeo, Baekje, and Silla were multi-ethnic and multilingual states, and that the presence of a Japonic language in the kingdoms of Goguryeo and Baekje was the result of a migration of proto-Japonic speakers originating from the southeast coast of China in route to Japan (the Yayoi migration which will be referred to below). As the majoritarian language of Silla would become the language we now know as the Korean language, the dominant language of Goguryeo would belong to the Tungusic family.

Unger (2005) and Vovin (2005) defend that the languages of Goguryeo, Baekje, and Silla were just dialects of old Korean, but that the Silla language had not arrived to the southern part of the peninsula before the 3<sup>rd</sup> century. This is owed to the fact that the Yayoi migration to Japan would have originated there, with Japonic toponyms in the Korean peninsula being vestiges of a linguistic substratum in the region.

That is, the history of the formation of the Korean language is still full of gaps and ongoing debates. The status of the Japanese language and the debates about the settlement of the archipelago are not much different.

## 4 - Settlement of the Japanese Archipelago

The questioning of the settlement in the Japanese archipelago shows some similarity to the questions shown above. If, on the one hand, Japan does not seem to have had multilingual kingdoms as Goguryeo and Baekje may have been, the Japanese archipelago presents not just difficulties relating to the origins of the Japanese language (which would eventually include the languages of Okinawa), but also the mysteries related to the genetic relationship with the Ainu language, today nearing extinction.

In a manner parallel to the Korean peninsula, we do not know for certain when the first hominids arrived to what we recognise today as the Japanese territory. However, we know that the oldest human fossil found in the archipelago is dated of around 30 thousand years ago, and most artefacts that could have



told us more about the period were probably lost to the rising sea levels (HENSHALL, 2004, p. 8).

Our knowledge about the settlement of the archipelago starts to improve at about 15 thousand years ago, when the first ceramic pottery appears in the Japanese territory, a first in world's history (HENSHALL, 2004, p. 9), giving rise to what we now know as the Jomon Period, name given to the cord-marked pattern that adorned the pottery.

It is supposed that the Jomon people arrived in Japan from Siberia (with some scholars pointing to a possible origin from Southeast Asia), before the rising sea levels, when the country was still connected to the continent (SOKAL and THOMSON, 1998, p. 2; HONG, 2005, p. 2-3; COOKE, MATTIANGELI, et al., 2021, p. 10), scattering throughout the country reaching the Ryukyu, stopping at the island of Okinawa (HUDSON, 2015, p. 573).

The link between the Ainu people and the Jomon migration is known at least since the latter part of the 19<sup>th</sup> century, when Yoshikiyo Koganei, after digging archaeological sites in Hokkaido and comparing Jomon and Ainu skeletons, showed that there were similarities in order to show a relationship between the two peoples (LOW, 2012, p. 557).

At around 400 BC, then, there was another change to the demographic make-up of Japan: a new migratory wave came about with the Yayoi people, which besides physical differences (such as fairer skin and higher stature), brought along a rice-based diet and knowledge about bronze and iron working (HENSHALL, 2004, p. 12). This new wave originated mostly in the Korean peninsula to the Japanese island of Kyushu (HUDSON, 2015, p. 575-6), extending rapidly to the centre of the Honshu Island, leaving the north of Japan mostly to the Jomon influence (HENSHALL, 2004, p. 13). Genetically, the individuals of the Yayoi wave present characteristics stemming from the basin of the Amur River (East of Russia and Northeast of China) and the East side of the Liao River (Northeast of China), presenting a great affinity with the modern Tungusic peoples and old populations of the Amur River (COOKE, MATTIANGELI, et al., 2021, p. 10).

The Yayoi period was followed by the Kofun period, when great tombs appeared in Japan, at around the 3<sup>rd</sup> century. According to recent research about the genetic make-up of the Japanese population throughout the ages, this cultural shift seems to have coincided with a third (and last) migratory influx from the continent (COOKE, MATTIANGELI, et al., 2021), possibly resulted from a drought in the Han basic that led Baekje farmers to look for new lands in Kyushu (Hong, 2005). The individuals of the Kofun wave, for their part, present genetic characteristics originating from East Asia, which were added to the genetic make-up from the previous two waves – forming the current profile of the Japanese people (COOKE, MATTIANGELI, et al., 2021, p. 9-10).

The difficulty, to the historical linguist, is that very little is known about the languages spoken by the migrants in each of these waves, obscuring the origins of the Japanese language and its closest relatives, such as the languages of the

Ryukyu Islands. They offer, at most, suggestions of shared histories, such as the ones between the Japanese and the Ainu, in the Jomon wave, or between the Japanese and Koreans in the migrations possibly from Baekje to Kyushu in the Yayoi language. But these are, for now, just speculation.

It is not possible, therefore, to establish how exactly the main languages of Northeast Asia were formed, even in possession of all this information of historical and genetic nature. The enterprise of searching for genetic relationships with other language families and reconstructing a shared protolanguage for the languages of Northeast Asia and its possible relatives also present important gaps.

## 5 - Interlinguistic Relations

There is no certainty or scientific consensus regarding the origins of the languages in the region and the genetic relationship between themselves and other language families of the world – which does not mean there is not a long tradition in historical linguistics trying to elucidate these links.

In order to better comprehend the hypotheses and the criticism aimed at them, the hypotheses will be mentioned in increasing order of language involved. Firstly, the Ainu-Japanese hypothesis will be mentioned, as well as the comparisons between the Ainu and the Japonic languages, connected by the Jomon migratory wave. Afterwards, the Japanese-Korean hypothesis will be mentioned, looking for genetic relationship between the Japonic and the Koreanic families and the reconstruction of the protolanguage that may have given rise to both families. Up next, different versions of the Altaic hypothesis will be presented, beginning with the Micro-Altaic hypothesis, which exclude the Japonic and Koreanic families, and then the Macro-Altaic and Uralic-Altaic, which include these families.

## 6 - The Ainu-Japanese Hypothesis

As mentioned before, the Ainu language is usually considered a Paleosiberian language (COMRIE, 1981; PATRIE, 1982, p. 6), due to the difficulty which is typical of the region in establishing genetic relations with other language families. As has also been already mentioned, the Paleosiberian languages do not constitute a language family in the strict sense of the world, being rather a group of languages from Northeast Asia whose relationship to other language families is still unknown.

However, due to the geographical proximity between the Japanese language and Ainu, and the connection between their speakers with the Jomon wave mentioned above, sometimes appear some attempts to look for genetic relationships between the Ainu language and the Japonic family. A usual problem

are the many differences between the languages, more than their similarities, which have been frequently portrayed in the literature about the topic.

Chamberlain (1887), for example, already mentioned that many characteristics of the Ainu language are so different from the Japanese language that the hypotheses of the related languages, such as the Ainu personal affixes, the polysynthetic morphology of ancient Ainu, the lack of verbal inflection, and, at last, the suffixes and verb suppletive forms for subject and object in the plural number (SHIBATANI, 1990, p. 6).

But, in the proponents' camp of the debate, Hattori (1964, p. 27-8), besides mentioning the typological similarities (which are either common to the region or to verb-final languages), mentions that Ainu radicals such as {kur-} meaning shadow and present in words such as "niskur" (cloud) and "kunne" (black), seems to present cognates in other languages of the ultra-altaic hypothesis (which will be discussed later), as "kurasu" (dark), "kuro" (black) and "kumo" (cloud) in Japanese; "gureum" (soot), "geom" (black), and "geurimja" (shadow) in Korean; "kurunyuk" (soot) and "komnomo" (black) in Evenki, referred to by the author as Tungus; and even "korom" (soot) in Hungarian. Unfortunately, there was no attempt to make sound correspondences or a broader list of possible cognates.

This enterprise was undertaken by Patrie (1983) who, through several sound change rules, besides grammatical comparisons, tried to not only establish a connection between the Ainu language and the Japonic language, but also expanded the attempt to include Ainu in the Macro-Altaic hypothesis. Despite some positive criticism to the endeavour, such as Miller (1983), the work is usually seen as a valid work that, despite not being convincing enough to attest a genetic relationship between the Ainu and Japanese languages, showed possible paths and hypotheses not only pertaining to the Japonic family, but also to neighbouring languages, such as Korean, due to the quantity of possibly shared lexical material (BYNON, 1983; DETTMER, 1983; STREET, 1982).

Besides the Ainu language, these inevitable relationships between the languages of the archipelago and the languages of the continent would also extend to the two most widely spoken ones: Japanese and Korean.

## 7 - The Japanese-Korean Hypothesis

It would be impossible to present an outline to the formation of northeast Asia without mentioning if there is any relationship between the two most spoken languages in the region. In fact, there is a long research tradition about the relationship between the Koreanic family and the Japonic family, which will be hereafter called the Japanese-Korean hypothesis, which is still plagued by several gaps if true (SHIBATANI, 1990, p. 113; SOHN, 1999, p. 36).

Due to their geographical and cultural proximity, it is not exactly unexpected that the two languages are usually the object of comparison. But, in spite of the comparison between the two languages being undertaken for a few

centuries now, such as Arai (1717) being an example of attempts to establish etymologies common to both languages, a number of these first attempts to show a connection between the two languages had flawed methodologies and presented as cognates pairs that did not exhibit sound correspondences, Kanazawa (1910) being one such example of the latter case. This was owed, at least in some part, to political questions (as Japan had just invaded and annexed the Korean peninsula at the time (HENSHALL, 2004, p. 95)).

Maybe it would be ideal to consider Martin (1966) as the starting point for a broader and more systematic analysis of the two languages due to its reconstruction of 320 possible cognates and their respective sound correspondences. Since then, that is, a stricter methodological rigor was applied to the topic, even though some of the correspondences were inconsistent (the Japanese /a/ phoneme at some point could correspond to at least six different Korean vowels) and the constructions were not done taking into account the oldest available versions in either language, but with either modern or pre-modern vocabulary (SOHN, 1999, p. 33).

Since then, there have been other attempts, such as Vovin (2010) and Francis-Ratter (2016), and systematizations, such as Whitman (2012) which propose interesting sound correspondences between proto-Korean and what the latter author calls Proto-Japanese-Ryukyuan. It is common to all these instances that the protolanguages are reconstructed with 6 or 7 phonemic vowels, with quite simple phonological systems (with the aspirations of the Korean language being explained as later processes through the result of syncope, and the distinction between voiceless and voiced occlusives in Japanese as the result of pre-nasalisation) (WHITMAN, 2012, p. 27-8).

Some of the systematizations seek to explain the differences between the phonological systems of the languages, such as the lack of \*h and \*tʃ which are not found in proto-Japanese-Ryukyuan. An often-fancied hypothesis is that the \*h in proto-Korean corresponds to \*s before \*i and \*j but \*k in other environments, whereas \*tʃ corresponds to \*s before high vowels and \*t in other environments, as shown in the table below.

pNR	pK	pNK
*se- “Do”	*hjə- idem	*hjə-
*siro- “White”	*hjə- idem	*hjə-
*kasa “Mass”	*ha “much, big”	*ha-
*kəsi “back”	*heli idem	*həti
*kunsu “Pueraria Lobata”	*hicirk idem	*hincu
*kusi “espeto”	*kotʃ idem	*kotʃ-
*puta- “two”	*pətʃək “pair”	*pətʃə
*mi(t)- “three”	*mjetʃ “some”	*mietʃ

Table 1. Reconstruction for \*h and \*tʃ of Proto-Japanese-Korean by Whitman (2012, p. 30-1)

There are explanations for other sound changes found in the table which, for length constraints, will be suppressed here, with the supposition that \*ti in the reconstruction \* of proto-Japanese-Korean has also undergone a process of lenition in proto-Korean (resulting in \*l) and through a process of palatalisation in proto-Japanese Ryukyuan (leading to \*s). But it is already possible to have an idea of how the hypothesis and systematizations are carried out by the authors in order to reconstruct a protolanguage common to both languages.

The difficulties begin to pile up when comparisons are done a little further beyond a small number of possible cognates and comparisons are carried out in a more systematic manner, such as when comparing the numbers of both languages:

Glosa	pNR	pK
1	*pitə	*p̥iris “primeiro”
1	*kata “um de um par”	*hət(V)- “um”
2	*puta	*p̥t̥jāk < *p̥t̥jāk “dobro”
2		*tupir
3	*mit	*mjet̥n “alguns, quantos”
3		*se-
4	*jə	*ne
5	*itu	*tasə
6	*mu(t)	*jəsəs
7	*nana	*nilko/up
8	*jǎ	*jətərp
9	*kəkənə	*ahop
10	*təwə	*jer

Table 2. Reconstruction of numbers in Proto-Japanese-Ryukyuan e Proto-Korean according to Whitman (2012, p. 33)

The systems do not seem to have one single match, unless we suppose at least one of the languages present innovations for all the numbers between 1 and 3, and run out of answers from 4 onwards. In this case, proponents of the Altaic hypothesis have tackled these difficulties with references to innovations and loans to explain the apparent non-corelation between the cognates (MILLER, 1969; HAMP, 1970; BLAŽEK, 1999). Naturally, it is a valid tool and such phenomena happen frequently in the natural languages, but taken together with other systematisation problems in the Altaic hypotheses, it renders the proposals somewhat problematic. Pronouns, to cite one other fragility with the Altaic hypotheses, do not present a more promising case, showcasing the difficulties with more systematic comparisons.

Therefore, if on the one hand, the Koreanic and Japonic families are considered close to one another, it has been extremely difficult to reconstruct a common protolanguage to both language families. And, if there are still gaps in the confirmation of a genetic relationship between both language families, the situation becomes even more complicated if we introduce more variables – and this is exactly what happens once we take into account the different Altaic hypotheses and mixed hypotheses.

## 8 - The Altaic (Transeurasian) Hypotheses

There are many hypotheses connecting the Koreanic and Japonic families to other natural languages, such as the Indo-European, Afro-Asiatic, Sino-Tibetan, Dravidian and even Amerindian families (SOHN, 1999, p. 18; LEE e RAMSEY, 2011, p. 15; SHIBATANI, 1990, p. 94-5). However, generally, when a hypothesis is formulated regarding the relationship of the Koreanic and Japonic families with other families, besides the ones mentioned above, there is usually at least a passing reference to the Altaic family (also referred to as Transeurasian in more recent publications).

According to this hypothesis, the Koreanic and Japonic families shared a common ancestor with Tungusic, Mongolic and Turkic families, corresponding approximately to a large swathe of land that crosses the central part of the Asian continent from Turkey to Japan. A counterhypothesis, which states that the similarities are due to a history of prolonged contact rather than to a common origin, is also frequently mentioned in these discussions. It is important, therefore, to offer a more detailed sketch of the hypotheses and their relationships to the families from Northeast Asia.

Both the Altaic hypothesis and the counterhypothesis that the languages are similar due to contact share a centuries long history. Maybe the first mentioned to a common origin to Turkic and Mongolic peoples dates from the 17<sup>th</sup> century, by the historian Abu 'I-Gazi Bahadur Khan, from the Khanate of Khivan; and, the first mention to the similarities being a result of continuous contact came the following century by the Prussian naturalist and researcher Peter Simon Pallas. Since then, different versions of the Altaic hypotheses have been formulated, being necessary to distinguish the different meanings behind the term so as to avoid any misunderstanding.

Originally, the label “Altaic” referred to a possible language stock that combined three language families from Central Asia and surrounding areas: the Turkic family to the West, and the Mongolic and Tungusic families further to the East.

The Turkic family, the most widespread of the families often spoken of as “Altaic”, is present from the Bosphorus Strait, in Turkey, to the Sakha Republic, or Yakutia, in Eastern Russia. Undoubtedly, it is a language family with a common origin, being relatively simple to demonstrate a relationship between its various



languages, the cultural differences of their very varied speakers notwithstanding (BOESCHOTEN, 1998, p. 1). It was possibly originated in Southern Siberia, or neighbouring areas, where they came into contact with Indo-European nomads who had already domesticated horses (GOLDEN, 1998, p. 16). The Turkic languages typically have agglutinative morphology, that is, they are characterised by the juxtaposition of several clearly identifiable morphemes bounded to the root of a word, vowel harmony and a CV(C)(C) syllable structure (JOHANSON, 1998, p. 31-5).

The Mongolic family, for its part, is frequently compared to the Turkic family, even for researchers that dismiss the Altaic hypothesis, showing a great number of lexical and morphosyntactic correspondences (RÓNA-TAS, 1998, p. 78; SCHÖNIG, 2003, p. 403). Outside Mongolia, the Mongolic languages are also spoken in Siberia, among other parts of Russia, and in China (RYBATZKI, 2020, p. 24). As with the previous case, the relationship between its different members is quite clear, suggesting a fairly recent dispersal history (JANHUNEN, 2003, p. 1). Also in a similar fashion, the languages of the Mongolic family tend to present vowel harmony, a (C)V(C) syllable structure and agglutinative morphology, being possible to reconstruct these characteristics in a protolanguage (JANHUNEN, 2003, p. 1-10).

And, finally, there are also the Tungusic languages, spoken in Eastern Siberia, Northern Manchuria, and parts of Xinjiang and Mongolia, and whose number of speakers has dwindled dramatically due to the adoption of the Mandarin and Russian languages by the local peoples (RYBATZKI, 2020, p. 22-3). It is also possible to find in Tungusic languages, such as Evenki, vowel harmony and agglutinative morphology, with (C)V(C) syllable structure (BULATOVA e GRENOBLE, 1999; NEDJALKOV, 1997).

These three language families represent what could be considered a core of a language stock and which is present in all the different manifestations of the Altaic hypothesis. And, to this core, the Koreanic, Japonic and, less often, the Uralic families tend to be added. This last family, which spans from Scandinavia to Siberia, is constituted by family such as Finnish and Hungarian, and just as the previous families, these languages exhibit agglutinative morphology and often show vowel harmony and a (C)V(C) syllable structure, all traceable to their protolanguage (AIKIO, 2022).

Following Comrie, we will call “Micro-Altaic” the hypotheses that only take into consideration the core of this language stock and “Macro-Altaic” the hypotheses that also include the Koreanic and/or Japonic families – originally, this is the context in which authors such as Robbeets (2020) use the moniker “Transeurasian”. And, at last, there is also the hypothesis that the languages belonging to the “Macro-Altaic” group might be related, albeit distantly, to the Uralic family. Exceptionally, we will call this hypothesis “Uralic-Altaic” in order to differentiate it from the other two hypothesis. Systematically, we then have the following scenario:

Family/Hypothesis	Micro-Altaic	Macro-Altaic	Uralic-Altaic
Mongolic	✓	✓	✓
Tungusic	✓	✓	✓
Turkic	✓	✓	✓
Koreanic	✗	✓	✓
Japonic	✗	Sometimes	✓
Uralic	✗	✗	✓
Total	3 families	4 to 5 families	6 families

Table 3. Systematisation of the different sorts of Altaic hypotheses

Because its members do not belong to Northeast Asia, strictly speaking, the validity of the Micro-Altaic hypothesis will not be seen with much depth. It is necessary to point out, however, that in spite of the long research tradition, dating from the early 20<sup>th</sup> century with the comparisons of the Turkic and Mongolic language families by Gustaf John Ramstedt, all hypotheses that these families share a common ancestor have been met with the possibility that the similarities all stem from prolonged contact (BLAŽEK, 2019, p. 43-6).

Such is the present state of affairs after more than a century, with authors that once defended a common origin, such as Vovin (2017), having switched sides due to the focus of Altaic literature on lexical comparisons at the expense of morphological comparisons, and the excessive mention to semantic drift to explain away potential cognates and sound correspondences that also seem to have too many exceptions and variations. Naturally, proponents such as Dybo and Sarostin (2017) seek to respond to such criticism with methodological explanations, arguing that it is expected from Altaicists a set of standards that is not applied to other well-established families, such as Indo-European and Austronesian.

In short, when discussing the relationship of the languages from Northeast Asia with other families of the Altaic hypotheses, it is necessary to keep in mind that the very validity of the Micro-Altaic hypothesis still lacks a firm foundation, so the following words do not refer to the addition of the Japanese and Korean languages to a widely accepted language stock.

And yet, the addition of North-eastern Asian families to the discussion did not take too long to come about, with Ramstedt himself drawing comparisons in a cautious manner with the Japanese language in 1924 and then writing a Korean language grammar 15 years later (LEWIN, 1976, p. 392). This does not mean he was the first to see this possibility, although the previous works relied on an extremely faulty methodology.

One such example, and something of a cornerstone for the research about the genetic relationship of the Japanese language with the Uralic-Altaic hypothesis was Katsuji Fujioka's 1908 publication, which mentioned that the languages belonging to this group shared, at least fourteen common typological traits (SHIBATANI, 1990, p. 96), such as:

- 1) Consonant clusters not allowed at the beginning of words;
- 2) The /r/ phoneme being inexistent in the beginning of words;
- 3) Presence of vowel harmony;
- 4) Lack of articles;
- 5) Lack of grammatical gender;
- 6) Verb inflection expressed by suffix;
- 7) Several verb suffixes;
- 8) Grammatical relations being expressed by grammatical particles;
- 9) Preference for postpositions instead of prepositions;
- 10) The use of verbs to express possession;
- 11) Comparison with the use of ablative constructions;
- 12) Presence of question particles at the end of interrogative sentences;
- 13) Relative low use of conjunctions;
- 14) Modifiers preceding modified heads, objects preceding verbs.

The problem with the methodology used by Fujioka becomes evident for two reasons.

For starters, a methodology that seeks to establish genetic relationship based solely on typological traits is bound to yield unwanted result, thus forcing us to realise its fragility. With the development of Greenbergian universals in the latter half of the 20<sup>th</sup> century, we have known that verb-final languages show a strong preference for postpositions, and it is extremely common among these languages for modifier words to precede the nouns they modify<sup>5</sup>. These traits, therefore, are not exactly something that sets the region apart, let alone suffice to present a convincing case for a shared origin.

On top of that, Northeast Asia, where many of these so-called Altaic languages are spoken, is a known area of language convergence. In these areas, also known as Sprachbund, languages that do not show a common origin present structural convergence (be it in their phonology or morphosyntax, among two possibilities). Just like Europe (HEINE e KUTEVA, 2006) and Continental Southeast Asia (ENFIELD, 2017), Northeast Asia is a vast area that presents characteristics of a Sprachbund (TANAKA DE LIRA, 2021), leading precisely to the sort of typological similarities Fujioka took as a sign of a common origin. In

---

<sup>5</sup> If we cross the surveys <[https://wals.info/combinations/81A\\_87A\\_85A\\_26A#4/0.53/286.74](https://wals.info/combinations/81A_87A_85A_26A#4/0.53/286.74)>, we find out that at least a third of all the 279 verb-final languages included in the World Atlas of Language Structures present all these traits. Inevitably, many of these languages are "false positives", with no possible relation with the languages of the Uralic-Altaic hypothesis, such as the Quechua and Ayamara languages of South America, and the Indo-European and Dravidian languages spoken in the Indian subcontinent.

this case, despite the languages showing structures with similar functions, such as the verbal inflections of Mongolian and Korean (SONG, 2011), the structures do not seem to be cognates, but the result of parallel developments.

Japanese linguists from early 20<sup>th</sup> century did not always try to connect the Japanese language to the Altaic hypotheses, often times defending instead that the Korean language was the one with Altaic links, with no mention whatsoever to their native tongue. An example was Kurakichi Shiratori, who published in 1914 a vocabulary comparing 595 words to propose a relationship of the Korean language with the Uralic-Altaic hypothesis, recognising the importance of factors such as vowel harmony, without including the Japanese language (LEWIN, 1976, p. 392; SOHN, 1999, p. 18).

It was from the latter half of the 20<sup>th</sup> century, with the contributions of works by Ramstedt, more systematic comparisons began to appear, such as Poppe (1960). An example of the comparisons made is the sound correspondence in which \*p in Proto-Altaic, through the processes of lenition and debuccalisation, hypothesised in the evolution of “p > f > h > ZERO” which would have remained as it originally was in Korean, become /f/ in Manchu, /h/ in other Tungusic languages such as Evenki and Even, and elided in Turkish and Mongolian (in the latter case being retraceable to /p/ in Classical Mongolian and /h/ in Medieval Mongolian). The Japanese words are added below for the sake of comparison:

	p	F	H	∅	
	Korean	Manchu	Evenki / Even	Mongolian	Japanese
Pray	pil-	firu-	hiruge- (Evenki)	iryge-	inoru
Village, plain	pəl	falga		ail	hara
Blow	pul-	fulgije	hu- (Even)	ulije	huk-
Season, year, spring	pom	fon		on	haru

Table 4. Comparison of cognates for the verbs “to pray” and “to blow” and for the nouns “village / plain” and “season / year / spring” in Poppe (1960) with added suggestion of Japanese cognates by (1999, p. 19)

The tradition to consider both the Koreanic family as well as the Japonic family as part of the Macro-Altaic hypothesis went on with Miller (1971), which offered to some lexical and sound correspondences to justify the insertion of both families in the language stock. Since then, both the inclusion as well as the position of both language within the stock has been a fruitful research programme.

Nowadays, it is not rare that the Koreanic family is considered the closest to the Tungusic family within the Macro-Altaic hypothesis, with proposals of creation of a Macro-Tungusic separated from the other two families. On the other hand, it is also common for Altaicists to consider the Japonic family, for its turn, to be considered the closest to the Koreanic family, even if in this latter case it is deemed necessary to establish more proof of relationship between the two languages. There are even those who, like Poppe (1971), suggest that Proto-(Ainu-)Japanese-Korean had a common ancestor with Proto-Altaic, placing them outside the stock but within a larger North Asian phylum.

Another possibility mentioned in research publications about the origin of languages from Northeast Asia, especially when it comes to the Japanese language, is that the difficulties stem from the search of a single origin – when it is well possible that the Japonic family may have arisen out of the contact of two different language families, such as one from the Altaic stock and another one from the Austronesian family. This kind of possibility will be herein dubbed “Mixed Hypotheses”.

## 9 - The Mixed Hypotheses

Due to the diverging phonology of the Japonic family when compared to the continental languages, the difficulties with language comparison and the reconstruction of a shared common protolanguage, as well because of the several migratory waves that altered the populational make-up of the Japanese people, there is also a history of considering the family as the result of a contact between an Altaic language and a language from an altogether different origin – usually from South or Southeast Asia. An example is Polivanov (1974) who, when commenting about prosody elements and pitch accent in Japanese language, draws parallels with Austronesian languages.

The Austronesian family, hitherto unmentioned, is the second most widely spoken language family in extension (behind only Indo-European) and possibly the language family with the largest absolute number of members, with around 1200 languages, or around 20% of all natural languages (ADELAAR, 2005, p. 1). Typologically, the family presents a very wide variability (HIMMELMANN, 2005, p. 110), especially outside Oceania (that is, in Asia and in Madagascar), but its members generally present three to five vocalic phonemes and something like 16 to 20 consonantal phonemes, (C)V or (C)V(C) syllable structure, generally with restrictions regarding which consonants can appear in coda position and clusters in syllable boundaries (HIMMELMANN, 2005, p. 115). These are the phonological traits that draw the attention of Japanese language researchers.

Polivanov (1974, p. 139, 146) presents some characteristics of the Japanese language that may hark back to a contact with an Austronesian language:

- 1) The use of the {ma-} prefix to indicate intensity together with reduplication process in Malayo-Polynesian and lengthening of the first consonant in Japanese (/mat:aira/ “very plain”, /mak:uro/ “very black”);
- 2) Preference for two-syllable lexical morphemes (such as in kata “shape / person” and naka “inside / centre”), but with monosyllabic grammatical morphemes;
- 3) Presence of prefixes in Japanese, which may have had Austronesian origin, in opposition to the absolute preference for suffixation in so-called Altaic languages;
- 4) Functions of morphological reduplication, either partial or total, which the author dubs “the most archaic layer of Japanese morphology”;
- 5) Simplicity of its vowel inventory, without vowel harmony and with pitch accent;
- 6) What the author calls “musical Wortakzent”, that is, pitch accent instead of stress accent typical of languages such as Portuguese;
- 7) Prominence of open syllables;
- 8) Compatibility of old Japanese consonant inventory and that of the Polynesian languages, without voiced and voiceless pairs, but with the presence of three nasal consonants;
- 9) Debuccalisation process regarding \*p, a parallel change in Japanese (pi > fi > hi “fire”) and Polynesian (\*apui > api > ahi IDEM);
- 10) Evolution of voiceless and voiced opposition through prenasalisation (“<sup>m</sup>b > b” e “<sup>m</sup>d > d”) both in Japanese and Melanesian (an opposition that was not always present, as mentioned in #7).

Naturally, some caveats can be made regarding some of these points, one of them being that there is also pitch accent present in the Korean peninsula (#4) at the very least since Mediaeval Korean and which, despite its disappearance in standard language, remains a distinctive trait in dialects from the Eastern part of the peninsula, such as in the Hamgyeong dialect (North Korean Eastern coast) and Gyeongsang (South Korean Eastern coast) (SOHN, 1999, p. 66, 71; YEON, 2012, p. 169). And, as mentioned above regarding Katsuji Fujioka’s work, this sort of comparison so wide without the identification of cognates and sound correspondences may yield unwanted results. And yet, this is a research programme that attracted some support in Japanese historical linguistics.

Hisanosuke Izui, in the 1950’s, was the first to try to formulate in a more systematic manner how this mixed origin may have come about. He posits that an Austronesian language may have been a substratum in the formation of the Japanese language combined with an Altaic superstratum (MURAYAMA, 1976, p. 420; SHIBATANI, 1990, p. 104). Other systematisations offered by Izui are the following:



	Tagalog	Japanese
MP *n – J n	nam-nam “experiment”	namu “lick”
MP *ŋ - J n	bunja “fruit”	hana (< *pana) “flower”
MP *p – J *p (p, φ, h)	pusod “belly button”	hoso (idem)
MP *d - J t, d	dakip “keep”	daku “hug”

*Table 5. Sound correspondences proposed by Izui between Malayo-Polynesian languages and Old Japanese and examples in contemporary Tagalog and Japanese according to Shibatani (1990, p. 105)*

Other linguists after Izui continued to propose variations of this mixed origin, with Ono and Murayama (to mention just two cases). In the former case, Ono supposed the Japanese language sprang from three waves:

- 1) Austronesian/Papuan: a first language could have entered Japan some 10 thousand years ago with a simple phonology consisted of 4 vowels and open-syllables;
- 2) Dravidian: a second language, proto-Tamil, would then have arrived in Japan sometimes during the Jomon period, around 3500 BC, bringing along a few cognates;
- 3) Altaic: at last, a third language of Altaic origin, coming from Goguryeo, may have shown up approximately in the beginning of the Yayoi Period (300 BC), introducing elements such as vowel harmony that were lost later on (SHIBATANI, 1990, p. 106).

Shichiro Murayama (1976, p. 420) offered something less complex than Ono’s tripartite origin, but not less radical: instead of suggesting just a Malayo-Polynesian substratum in the Japanese language, the defence was that it was genuinely a full-on mixed language. Others, such as Kawamoto (1980), considered that the Altaic part was actually the substratum, with the Japanese language being for the most part Austronesian in origin.

And this is just a brief list of Japanese and foreign linguists who have sought other explanations for the origins of the Japonic family, with the variety being extended to former Altaicists who have come to believe the Japanese language has a stronger bond with Papuan languages, such as Minoru Goh (KAMIMURA, 2015; SHIBATANI, 1990, p. 109) and even researchers who see a Tibeto-Burman origin, thus linking it to the Sino-Tibetan language, such as Charles Parker and Tatsuo Nishida (PARKER, 1940; UMEDA, 2012, p. 122; SHIBATANI, 1990, p. 110). These comparisons have been even more speculative, obtaining less support from the scientific community than the attempts to trace the origins of the Japanese language with any version of the Altaic hypotheses. And, leaving the archipelago for the continent, the situation is not much different.

As told in the section about the formation of the Korean language, issues regarding a multiple origin are not exclusive to the Japonic family in Northeast Asia. Kim (1981), upon mentioning the formation of the Korean language, and the difficulties in the identification of the language spoken in the three Kingdoms, mentions cognates with the Japanese language, the Tungusic family and even the neighbouring isolate Nivkh, which complicates the search for a unique origin. Furthermore, there is also the realisation that if linguists try and use toponyms as a proxy for the search of a substratum or a previous stage, there is no way to tell these languages did not in fact come from a previous unrelated language (KIM, 1981, p. 177).

Part of the problem emanates from the fact that few cognates in either language have been found for comparisons with language families outside Northeast Asia, which ends up making the use of the comparative method, essential to historical linguistics, all the more difficult (SHIBATANI, 1990, p. 113). As mentioned in the first pages, even if some tentative sound correspondences have been formulated for the identification of cognates in Japanese, Korean, and so many other languages mentioned in the Altaic hypotheses, it has not been possible to formulate rules that explain and predict cognates which have not already been mentioned.

## 10 - CONCLUSION

Northeast Asia has a very long history of migrations and extinction of several peoples of different origins. Much information that could have been useful to us has been lost, such as language descriptions of the Jomon and/or Yayoi populations as well as about the language spoken by each of the kingdoms that have existed in the Korean peninsula. Besides that, as history marches on, other local languages have disappeared, not always leaving behind enough data so others can be aware of their existence.

In these cases, the use of multidisciplinary tools as genetics and archaeology has sought to complement the information provided by historical linguistics, but there is still a long way to go.

Naturally, this is just a fragment of the contemporary state of ongoing research about the origin of languages from Northeast Asia. Due to length constraints, less attention to Paleosiberian language has been given than would have been ideal, and not enough data was offered to hypotheses that lack wider scientific support (such as the ones that link the Japanese language to the Dravidian and Tibeto-Burman families). It is hoped, however, that this summary may come in handy for incipient researchers, still without working command of the several languages in which these debates are carried out (English, Japanese, Korean or Russian), so that they can be aware of some basic facts about the investigations already made about the origins of the Koreanic and Japonic language families.

## Referências

ADELAAR, A. **The Austronesian Language of Asia and Madagascar: a Historical Perspective**. In: ADELAAR, A.; HIMMELMANN, N. P. *The Austronesian Languages of Asia and Madagascar*. Abingdon, Oxon, UK: Routledge, 2005. p. 1-42.

AIKIO, A. Proto-Uralic. In: BAKRÓ-NAGY, M.; LAAKSO, J.; SKRIBNIK, E. **The Oxford Guide to the Uralic Languages**. Oxford, UK: Oxford University Press, 2022. p. 3-27.

ARAI, H. Tōga. **Internet Archive**, 1717. Disponível em: <<https://archive.org/details/JAPAF.1219.1/page/n7/mode/2up>>. Acesso em: 20 jun. 2022.

ARAÚJO, G. H. D. S. **A teoria altaica e a língua japonesa: relações fonológicas**. Brasília: Trabalho de Conclusão de Curso (Licenciatura em Língua e Literatura Japonesa) — Universidade de Brasília, 2018.

BECKWITH, C. I. **The Ethnolinguistic History of the Early Korean Peninsula Region: Japanese-Koguryōic and other Languages in the Koguryō, Paekche, and Silla kingdoms**. *Journal of Inner and East Asian Studies*, Seoul, v. 2, n. 2, p. 33-64, December 2005.

BLAŽEK, V. **Numerals : comparative-etymological analyses of numeral systems and their implications (Saharan, Nubian, Egyptian, Berber, Kartvelian, Uralic, Altaic and Indo-European languages)**. In: BLAŽEK, V. *Altaic Numerals*. Brno, CZ: Masarykova Univerzita, v. 1, 1999. p. 102-140.

BLAŽEK, V. **Altaic Languages: History of Research, Survey, Classification and a Sketch of Comparative Grammar**. Brno, Czechia: Masaryk University Press, 2019.

BOESCHOTEN, H. **The Speakers of Turkic Languages**. In: JOHANSON, L.; CSATÓ, É. *The Turkic Languages (Routledge Language Family Series)*. London, UK: Routledge, 1998. p. 1-15.

BULATOVA, N.; GRENOBLE, L. **Evenki (Languages of the World/Materials 141)**. Munich, GER: Lincom Europa, 1999.

BYNON, T. **The Genetic Relationship of the Ainu Language by James Patrie**. *The Journal of the Royal Asiatic Society of Great Britain and Ireland*, Cambridge, UK, v. 1, p. 167-168, 1983.

CAMPBELL, L. **Historical Linguistics: An Introduction**. Cambridge, MA: MIT Press, 1999.

CHAMBERLAIN, B. H. **The Language, Mythology, and Geographical Nomenclature of Japan viewed in the Light of Aino Studies**. Tokyo, JP: Imperial University, 1887.

- COMRIE, B. **The Languages of the Soviet Union**. Cambridge, UK: Cambridge University Press, 1981.
- COOKE, N. P. et al. **Ancient Genomics Reveals Tripartite Origins of Japanese Populations**. *Science Advances*, Washington, DC, v. 7, n. 38, p. 1-15, September 2021.
- CROWLEY, T.; BOWERN, C. **And Introduction to Historical Linguistics**. 4th. ed. Oxford, UK: Oxford University Press, 2010.
- DETTMER, H. A. **The Genetic Relationship of the Ainu Language by James Patrie**. *Monumenta Nipponica*, Tokyo, JP, v. 38, n. 3, p. 331-333, Autumn 1983.
- DYBO, A. V.; STAROSTIN, G. S. **In Defense of the Comparative Method, or the End of the Vovin Controversy**. In: ROBBEETS, M. *Transeuratian Linguistics (Critical Concepts in Linguistics)*. Abingdon, Oxon, UK: Routledge, v. 4, 2017. p. 338-455.
- ENFIELD, N. J. **Language in Mainland Southeast Asia Area**. In: DIXON, R. M. W.; AIKHENVALD, A. Y. *The Cambridge Handbook of Linguistic Typology*. Cambridge, UK: Cambridge University Press, 2017. p. 601-623.
- ETHNOLOGUE. **Languages of the World**. 12th. ed. Dallas, TX: Summer Institute of Linguistics, 1992.
- FRANCIS-RATTE, A. T. **Proto-Korean-Japanese: A New-Reconstruction of the Common Origin of the Japanese and Korean Languages**. Columbus, OH: Ohio State University. PhD Dissertation., 2016.
- GOLDEN, P. B. **The Turkic Peoples: A Historical Sketch**. In: JOHANSON, L.; CSATÓ, É. Á. *The Turkic Languages (Routledge Language Family Series)*. London, UK: Routledge, 1998. p. 16-29.
- HAMP, E. P. **On The Altaic Numerals**. In: JAKOBNSON, R.; KAWAMOTO, S. *Studies in General and Oriental Linguistics Presented to Shiro Hattori on the Occasion of His Sixtieth Birthday*. Tokyo, JP: TEC, 1970. p. 188-197.
- HARRISON, S. P. **On The Limits of the Comparative Method**. In: JOSEPH, B. D.; JANDA, R. D. *The Handbook of Historical Linguistics*. Malden, MA: Blackwell Publishing, 2003. p. 213-243.
- HATTORI, S. **アイヌ語方言辞典**. Tokyo, JP: Iwanami Shoten, 1964.
- HEINE, B.; KUTEVA, T. **The Changing Languages of Europe**. Oxford, UK: Oxford University Press, 2006.
- HENSHALL, K. G. **A History of Japan: From Stone Age to Superpower**. 2nd. ed. London, UK: Palgrave MacMillan, 2004.

HIMMELMANN, N. P. **The Austronesian Languages of Asia and Madagascar: Typological Characteristics**. In: ADELAAR, A.; HIMMELMANN, N. P. *The Austronesian Languages of Asia and Madagascar*. Abingdon, Oxon, UK: Routledge, 2005. p. 110-181.

HONG, W. **Yayoi Wave, Kofun Wave, and Timing: The Formation of the Japanese People and Japanese Language**. *Korean Studies*, Honolulu, HI, v. 29, p. 1-29, 2005.

HUDSON, M. J. **28 - Japan: Archaeology**. In: NESS, I. *The Global Prehistory of Human Migration*. Chichester, West Sussex, UK: John Wiley & Sons, 2015. p. 572-585.

JANHUNEN, J. **Proto-Mongolic**. In: JANHUNEN, J. *The Mongolic Languages (Routledge Language Family Series)*. London, UK: Routledge, 2003. p. 1-29.

JANHUNEN, J. **The Lost Languages of Koguryō**. *Journal of Inner and East Asian Studies*, Seoul, v. 2, n. 2, p. 65-86, December 2005.

JOHANSON, L. **The Structure of Turkic**. In: JOHANSON, L.; CSATÓ, É. Á. *The Turkic Languages (Routledge Language Family Series)*. London, UK: Routledge, 1998. p. 30-65.

KAMIMURA, T. **Notes on the History of the Studies of Papuan Languages in Japan**. *Language and Linguistics in Oceania*, Komatsu, Ishikawa, JP, v. 7, p. 19-28, April 2015.

KANAZAWA, S. **日韓両国語同系論**. 東京: 三省堂書店, 1910.

KAWAMOTO, T. **日本語の源流**. 東京: 講談社, 1980.

KIM, B.-H. **The Relationship between the Korean and Japanese Languages**. *京都産業大学国際言語科学研究所所報*, Kyoto, JP, v. 2, n. 4, p. 173-184, Dec 1981.

KIM, J. **A History of Korea: from "Land of the Morning Calm" to States in Conflict**. Bloomington, IN: Indiana University Press, 2012.

LEE, G.-A. **The Transition to Foraging to Farming in Prehistoric Korea**. *Current Anthropology*, Chigago, IL, v. 52, n. S4, p. S307-S329, October 2011.

LEE, K. M.; RAMSEY, S. R. **A History of the Korean Language**. Cambridge, UK: Cambridge University Press, 2011.

LEWIN, B. **Japanese and Korean: The Problems and History of a Linguistic Comparison**. *The Journal of Japanese Studies*, Seattle, WA, v. 2, n. 2, p. 389-412, Summer 1976.

LOW, M. **Physical Anthropology in Japan: The Ainu and the Search for the Origins of the Japanese**. *Current Anthropology*, Chicago, IL, v. 53, n. S5, p. S57-S68, April 2012.

MARTIN, S. E. **Lexical Evidence Relating Korean to Japanese**. *Language*, Washington, DC, v. 42, n. 2, p. 185-251, April-June 1966.

MILLER, R. A. **The Altaic Numerals and Japanese**. *The Journal-Newsletter of the Association of Teachers of Japanese*, Boulder, CO, v. 6, n. 2, p. 14-29, 1969.

MILLER, R. A. **Japanese and the Other Altaic Languages**. Chicago, IL: The University of Chicago Press, 1971.

MILLER, R. A. **The Genetic Relationship of the Ainu Language by James Patrie**. *Language*, Washington, DC, v. 59, n. 2, p. 447-448, June 1983.

MURAYAMA, S. **The Malayo-Polynesian Component in the Japanese Language**. *The Journal of Japanese Studies*, Seattle, WA, v. 2, n. 2, p. 413-326, Summer 1976.

NEDJALKOV, I. **Evenki (Descriptive Grammars)**. London, UK: Routledge, 1997.

PARKER, C. K. **Early Japanese Systems of Counting**. *Monumenta Nipponica*, Tokyo, JP, v. 3, n. 1, p. 274-280, January 1940.

PATRIE, J. **The Genetic Relationship of the Ainu Language**. *Oceanic Linguistics Special Publications*, Honolulu, HA, v. 17, p. 1-174, 1982.

PEARSON, R. **Paleoenvironment and Human Settlement in Japan and Korea**. *Science*, Washington, DC, v. 197, n. 4310, p. 1239-1246, September 1977.

POLIVANOV, E. D. **Selected Works - Articles on General Linguistics**. The Hague, NL: Mouton & Co N. V. Publishers, 1974.

POPPE, N. Foreword. In: MILLER, R. A. **Japanese and Other Altaic Languages**. Chicago, IL: University of Chicago Press, 1971. p. ix-xi.

POPPE, N. N. **Vergleichende grammatik der altaischen sprachen (Vol 1)**. Wiesbaden, DE: Otto Harrassowitz, 1960.

PRATT, K. **Everlasting Flower: A History of Korea**. London, UK: Reaktion Books Ltd, 2006.

RANKIN, R. L. **The Comparative Method**. In: JANDA, R. D.; JOSEPH, B. D. *The Handbook of Historical Linguistics*. Malden, MA: Blackwell Publishing, 2003. p. 183-212.



RESENDE, J. V. T. L. **Proto nipo-coreano: uma análise de algumas hipóteses sobre o parentesco entre as línguas japonesa e coreana.** Brasília: Trabalho de Conclusão de Curso (Licenciatura em Língua e Literatura Japonesa) — Universidade de Brasília, 2019.

ROBBEETS, M. **The Classification of the Transeurasian Languages.** In: ROBBEETS, M.; SAVELYEV, A. *The Oxford Guide to the Transeurasian Languages.* Oxford, UK: Oxford University Press, 2020. p. 31-39.

ROBBEETS, M. et al. **Triangulation Supports Agricultural Spread of the Transeurasian Languages.** *Nature*, London, v. 599, p. 616-621, 10 November 2021.

RÓNA-TAS, A. **The Reconstruction of Proto-Turkic and the Genetic Question.** In: CSATÓ, É. Á.; JOHANSON, L. *The Turkic Languages (Routledge Language Family Series).* London, UK: Routledge, 1998. p. 67-80.

RYBATZKI, V. **The Altaic Languages: Tungusic, Mongolic, Turkic.** In: ROBBEETS, M.; SAVELYEV, A. *The Oxford Guide to the Transeurasian Languages.* Oxford, UK: Oxford University Press, 2020. p. 22-30.

SCHÖNIG, C. **Turko-Mongolic Relations.** In: JANHUNEN, J. *The Mongolic Languages (Routledge Language Family Series).* London, UK: Routledge, 2003. p. 403-420.

SETH, M. J. **A History of Korea: From Antiquity to the Present.** Lanham, MD: Rowman & Littlefield Publishers, 2011.

SHIBATANI, M. **The Languages of Japan.** Cambridge, UK: Cambridge University Press, 1990.

SOHN, H. M. **The Korean Language.** Cambridge, UK: Cambridge University Press, 1999.

SOKAL, R. R.; THOMSON, B. A. **Spatial Genetic Structure of Human Populations in Japan.** *Human Biology*, Detroit, MI, v. 70, n. 1, p. 1-22, February 1998.

SONG, J. **Verbal Inflections in Koreand and Mongolian: a Contrastive Analysis.** *The Journal of Studies in Language*, Gongju, KR, v. 27, n. 1, p. 99-116, 2011.

STREET, J. **The Genetic Relationship of the Ainu Language by James Patrie.** *The Journal of the Association of Teachers of Japanese*, Boulder, CO, v. 17, n. 2, p. 192-204, November 1982.

TANAKA DE LIRA, M. **O Nordeste Asiático como Área de Convergência Linguística: a Língua Japonesa em seu Contexto Regional.** In: JOKO, A. T., et al. *Diálogo Linguístico: Ocidente e Oriente.* Brasília, DF: Editora Universidade de Brasília, 2021. p. 315-336.

TANAKA DE LIRA, M. **Formação Linguística do Nordeste Asiático: Múltiplas Origens das Famílias Coreânica e Japônica.** Hon No Mushi, Manaus, AM. In Press.

TOH, S. H. **About Early Paekche Language Mistaken as Being Koguryō Language.** *Journal of Inner and East Asian Studies*, Seoul, v. 2, n. 2, p. 9-32, December 2005.

TRANTER, N. **Introduction: typology and area in Japan and Korea.** In: TRANTER, N. *The Languages of Japan and Korea.* Oxon, UK: Routledge, 2012. p. 3-23.

UMEDA, H. 追悼 西田龍雄先生. 日本言語学会, 2012. Disponível em: <[http://www.ls-japan.org/modules/documents/LSJpapers/journals/143\\_obituary.pdf](http://www.ls-japan.org/modules/documents/LSJpapers/journals/143_obituary.pdf)>. Acesso em: 06 July 22.

UNESCO. **North-east Asia.** In: UNESCO Atlas of the World's Languages in Danger. 3rd. ed. Paris: UNESCO Publishing, 2010. p. 48-58.

UNGER, J. M. **Summary Report of the Altaic Panel.** In: BALDI, P. *Linguistic Change and Reconstruction Methodology - Part 1 (Trends in Linguistics, Studies and Monographs 45).* Berlin, DE: Mouton de Gruyter, 1990. p. 479-482.

UNGER, J. M. **When was Korean First Poken in Southeastern Korea?** *Journal of Inner and East Asian Studies*, Seoul, v. 2, n. 2, p. 87-107, December 2005.

VOVIN, A. **Japanese, Korean, and Tungusic: Evidence for Genetic Relationship from Verbal Morphology.** In: HONEY, D. B.; WRIGHT, D. C. *Altaic Affinities: Proceedings of the 40th Meeting of the Permanent International Altaistic Conference (PIAC).* Bloomington, IN: Indiana University: Research Institute for Inner Asian Studies, v. 168, 2001. p. 183-202.

VOVIN, A. **Koguryō and Paekche: Different Languages or Dialects of Old Korean?** *Journal of Inner and East Asian Studies*, Seoul, v. 2, n. 2, p. 107-140, December 2005.

VOVIN, A. **Koreo-Japonica: A Re-Evaluation of a Common Genetic Origin (Hawai'i Studies on Korea).** Honolulu, HA: University of Hawai'i Press, 2010.

VOVIN, A. **Northeastern and Central Asia: "Altaic" Linguistic History.** In: BELLWOOD, P. *The Global Prehistory of Human Migration.* Oxford: John Wiley & Sons, 2013. p. 507-521.

VOVIN, A. **The End of the Altaic Controversy.** In: ROBBEETS, M. *Transeurasian Linguistics.* Abingdon, Oxon, UK: Routledge, v. 4, 2017. p. 288-335.

WHITMAN, J. **The Relationship between Japanese and Korean**. In: TRANTER, N. *The Languages of Japan and Korea* (Routledge Language Family Series). Abingdon, Oxon, UK: Routledge, 2012. p. 24-38.

YEON, J. **Korean Dialects: a General Survey**. In: TRANTER, N. *The Languages of Japan and Korea*. Abingdon, Oxon, UK: Routledge, 2012. p. 168-186.

YI, S. 29 - **Korea: Archaeology**. In: BELLWOOD, P. *The Global Prehistory of Human Migration*. Chichester, West Sussex, UK: John Wiley & Sons, 2015. p. 586-597.