Proceedings of the fifth International Conference on Asian Geolinguistics

EDITED BY
TRỊNH CẨM LAN, TRẦN THỊ HỒNG HẠNH,
HIROYUKI SUZUKI, AND MITSUAKI ENDO

Geolinguistic Society of Japan

ISSN 2436-6471

Proceedings of the fifth International Conference on Asian Geolinguistics

EDITED BY
TRỊNH CẨM LAN,
TRẦN THỊ HỒNG HẠNH,
HIROYUKI SUZUKI,
AND MITSUAKI ENDO

Geolinguistic Society of Japan
2023

Studies in Geolinguistics, Monograph series, No. 6
ISSN 2436-6471

Proceedings of the fifth International Conference on Asian Geolinguistics, edited by
Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki, and Mitsuaki Endo, 2023

Cover photo: Sunshine in the schoolyard © 2019 Đỗ Văn Hùng

First published 2023
doi: https://doi.org/10.5281/zenodo.8382130

Published by: Geolinguistic Society of Japan
Website: https://geolinguistics.sakura.ne.jp/
Office address: Room G1305, Aoyama Gakuin Univeristy, 4-4-25 Shibuya, Shibuya-ku, Tokyo

Preface

This volume contains papers presented at the fifth International Conference on Asian Geolinguistics (ICAG) held at the University of Social Sciences and Humanities, VNU, Ha Noi, Vietnam, from 4 to 5 May, 2023. The previous ICAGs were held as follows: 1st, at Aoyama Gakuin University, Tokyo, 2012; 2nd, at Chulalongkorn University, Bangkok, 2014; 3rd, at Royal University of Phnom Penh, Phnom Penh, 2016; and 4th, at Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia, Jakarta, 2018.

The proceedings of each ICAG (PICAG) have been edited and published as open access documents; see below for bibliographical information. Before PICAG-5, the titles are denoted as *Papers from...*; for PICAG-5, *Proceedings of...* is used.

PICAG-1: Endo, Mitsuaki (ed.) (2012) *Papers from the First International Conference on Asian Geolinguistics*. https://doi.org/10.5281/zenodo.6423581

PICAG-2: Endo, Mitsuaki (ed.) (2014) *Papers from the Second International Conference on Asian Geolinguistics*. https://doi.org/10.5281/zenodo.6423601

PICAG-3: Endo, Mitsuaki (ed.) (2016) *Papers from the Third International Conference on Asian Geolinguistics*. Fuchu: Research Institute for Languages and Cultures of Asian and Africa. https://publication.aa-ken.jp/papers_3IC_Asian_geolinguistics_2016.pdf

PICAG-4: Suzuki, Hiroyuki and Mitsuaki Endo (eds.) (2018) *Papers from the Fourth International Conference of Asian Geolinguistics*. Fuchu: Research Institute for Languages and Cultures of Asia and Africa. https://publication.aa-ken.jp/papers_4IC_Asian_geolinguistics_2018.pdf

The editors

TABLE OF CONTENTS

CONFERENCE PROGRAM

The Fifth International Conference on Asian Geolinguistics

Venue: Room Room 307, Building E, University of Social Sciences and Humanities, VNU 336, Nguyen Trai, Thanh Xuan, Hanoi, Vietnam
Online: Zoom Meeting

1st day 4th May 2023

Opening ceremony 8:45 – 9:00 Vietnam time
Chair: *Trịnh Cẩm Lan (University of Social Sciences and Humanities, VNU, Hanoi, Vietnam)*

Assoc. Prof. Đào Thanh Trường (Vice Rector, USSH, Vietnam National University Hanoi),
Welcome address

First session 9:00 – 10:00
Chair: *Trịnh Cẩm Lan (University of Social Sciences and Humanities, VNU, Hanoi, Vietnam)*

1. *Trần Trí Dõi, translated by Trần Thị Hồng Hạnh (University of Social Sciences and Humanities, VNU, Hanoi, Vietnam)*,
   Geographical distribution of autonyms for "person/people" in Mon-Khmer languages in South East Asia in contribution to the understanding of the original location (homeland) of Luòyuè residents
2. *Mitsuaki Endo (Aoyama Gakuin University, Tokyo, Japan)*,
   Geographical distribution of some Korean family names

Second session 10:15 - 11:45
Chair: *Atsuko Kanda Utsumi (Meisei University, Hino, Japan)*

3. *Truong Nhat Vinh (University of Social Sciences and Humanities, VNU, Hanoi, Vietnam),*
   Geographical distribution of placenames with the element "kẻ" in Red River Delta (Vietnam)
4. *Satoko SHIRAI (The University of Tokyo, Tokyo, Japan),*
   Basic roots varying among dialects in nDrapa: A geolinguistic analysis in the Qiangic context

5. *Mika Fukazawa (National Ainu Museum, Shiraoi, Japan),*
   Tracking basic Ainu vocabulary: Updates and changes of geographical distributions

Third session 14:00 – 15:30
Chair: [online] *SHEN Ruiqing (National University of Singapore)*

6. [online] *Yan Zhuo (Nankai University, Tianjin, China),*
   Linguistic Map of "Breast" in Zhuang and Its Interpretations
7. [online] *Hiroyuki SUZUKI (Kyoto University, Kyoto, Japan),*
   Geolinguistic analysis of the word form derived from *red* in Tibetic languages in Khams and Amdo
8. *Atsuko Kanda Utsumi (Meisei University, Hino, Japan),*
   The distribution of different types of voice system in the Austronesian world

Fourth session 16:00 – 17:30
Chair: *Trần Thị Hồng Hạnh (University of Social Sciences and Humanities, VNU, Hanoi, Vietnam)*

9. [online] *Zeng Zhichao (National University of Singapore),*
   Reconstructing Proto-Southern Mĭn *ãi, uãi: A Case of ABAB Distribution
10. *Kohei NAKAZAWA (Shinshu University, Matsumoto, Japan),*
    Regional Differences in Syllable Fusion in Japonic
11. *Chitsuko Fukushima (University of Niigata Prefecture, Niigata, Japan),*
    How Rip Weaving Spread in Japan: Interpreting Maps of Words and Referents

2nd day 5th May 2023

Sixth session 9:00 – 10:00
Chair: *Chitsuko Fukushima (University of Niigata Prefecture, Niigata, Japan)*

12. [online] *Zeng Xiaoyu (Nankai University, Tianjin, China),*
    Classification and interpretation of tone sandhi in three-tone Jiaoliao Mandarin
13. *Sầm Công Danh (University of Social Sciences and Humanities, VNU, Hanoi, Vietnam)*, Tone as a criterion for identifying a lect: the case of Tai Yo language via geographical distribution of tone

Seventh session 10:15 – 11:45
Chair: [online] *Zeng Xiaoyu (Nankai University, Tianjin, China)*

14. [online] *Liu Jinghan (Nankai University, Tianjin, China),*
    The Underlying Tonal Characteristics of Northeast Mandarin in Liaoning
15. [online] *Xiao, Nengping (Nankai University, Tianjin, China),*
    On the Disyllabic Sandhi Tone Pattern in Four-tone Middle Area Mandarin

16. [online] *SHEN Ruiqing (National University of Singapore),*
    A preliminary study of the geographic patterns of tonal development in Northern Min

Eighth session 14:00 – 15:00
Chair: *Mitsuaki Endo (Aoyama Gakuin University, Tokyo, Japan)*

17. [online] *Khairul Ashraaf Saari, Nor Hashimah Jalaluddin, and Fazal Mohamed Mohamed Sultan (National University of Malaysia, Kuala Lumpur, Malaysia),*
    Homorganic Liquid in Downstream and Upstream Dialects of Perak; Geolinguistic Analysis
18. [online] *Huang He (Fudan University, Shanghai, China),*
    The Border Effect of Historical Boundaries on Dialect Distribution: A quantitative approach by applying the spatial autoregressive model

Ninth session 16:00 – 17:00
Chair: [online] *Nor Hashimah Jalaluddin (National University of Malaysia, Kuala Lumpur, Malaysia)*

19. *Kazue Iwasa (Nagoya University of Foreign Studies),*
    Geolinguistic Analysis of the Yi Script written in *Huayi Yiyu*
20. *Trịnh Cẩm Lan (University of Social Sciences and Humanities, VNU, Hanoi, Vietnam),*
    Characteristics and Geographic Distribution of the Variants of Some Kinships of Mother's Side in Son Tay Dialect (Hanoi)

The geographical distribution of autonyms for 'person/people' in the Mon-Khmer languages of South East Asia and its contribution to the understanding of the original location (homeland) of the Luòyuè residents

TRẦN, Trí Dõi
(Faculty of Linguistics, USSH Vietnam National University Hanoi)

Abstract: The Sino-Vietnamese phonetic cluster Lạc Việt (雒 越) first appeared in the Hanshu (漢書, Hàn shū), a first-century CE Chinese work. At that time, this cluster was used to refer to a group of non-Sinitic residents in the Bách Việt (百越, Băi yuè) area, but there were no specific indications about the exact location of this group or the owners of this name. From the perspective of geolinguistics, this paper investigates autonyms that refer to 'person/people' in the Mon-Khmer languages of Southeast Asia. The aim of this paper is to prove that Lạc Việt, used by the Chinese in the 1st century, was used to record the autonym of the Mon-Khmer residents of Southeast Asia.
Key words: homeland, Lạc Việt (Luòyuè), Mon-Khmer, Hànshū, autonyms for 'person/people,' geolinguistic.

1. Introduction

When discussing the etymology of the name Lạc Việt (雒 越, Luòyuè) recorded in the Hanshu (漢書, Hàn shū), a Chinese book from the first century CE, Ferlus (2011) supposed that at that time, the Chinese people used these characters to record the cluster [*p.rak *wat]. The syllable [*p.rak] is the phonetic form of the Mon-Khmer autonym which etymologically means 'person/people.' His etymological explanation of the components of the Sino-Vietnamese phonetic cluster Lạc Việt was based on an analysis of its relations to other languages in southern China and Southeast Asia. He also believed that Southeast Asia cannot be constrained within today's administrative limits, and the geo-cultural region of Lạc Việt, including the Vietnamese language and the culture of the Viet (Nam) people, as recorded in ancient Chinese history books, must encompass the south of China, i.e., 'south of the Yangtze River' and present Southeast

TRẦN, Trí Dõi. 2023. The geographical distribution of autonyms for 'person/people' in the Mon-Khmer languages of South East Asia and its contribution to the understanding of the original location (homeland) of the Luòyuè residents. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 1–11. doi: https://doi.org/10.5281/zenodo.8374552

Asia. The author's works, published in 2017 and 2022, were not only in strong agreement with Ferlus's explanation but developed his idea by offering an additional discussion on the concept of Lạc Việt (雒越) in prehistory. As pointed out in these two studies, several Austroasiatic languages in Southeast Asia still retain the etymological meaning of that ancient name, deriving the ethnonym from when the Chinese used the character Lạc (Luò 雒) to phonetically record that non-Chinese name (Trần Trí Dõi, 2017, 2022). Recently, several anthropological studies (Đinh Hồng Hải, 2018; Dinh & Kelley, 2021) have also considered this etymological interpretation of the name Lạc Việt proposed by linguists.

Following previously published studies, the author presented another study (Trần Trí Dõi, 2023) on what names are currently used to refer to 'person/people' in some languages of Southeast Asia and the southern part of China. There are various different phonetic forms in the Mon-Khmer languages themselves and other forms in the languages of the Tai group, the Tai-Kadai family. However, in the case of the Tai languages, there is a distinction between the phonetic forms borrowed from the Chinese language (or the Sinotic group) when the Chinese language recorded the Mon-Khmer self-names, and on the other hand, the phonetic forms that the Tai languages themselves borrowed directly from the Mon-Khmer languages when the Tai people migrated southward to Southeast Asia.

From the perspective of geolinguistics, this paper examines the geographical distribution of those autonyms for 'person/people' in the Mon-Khmer and Tai languages. An interpretation of geographical maps will be to provide further discussion on the 'homeland' or 'cradle' of the people referred to by the phonetic cluster [*p.rak *wat] 雒 越 (Luòyuè) by the Chinese in the 1st century CE. The findings are expected to reveal that the real owners of that cluster are probably inhabitants of the Mon-Khmer linguistic region.

2. Geographic distribution of autonyms for 'person/people' in the languages of Southeast Asia

2.1. In the Mon-Khmer languages

As stated in the author's latest research (2023), at present, the autonyms for 'person/people' in the Mon-Khmer languages can be divided into three different types based on the structure of the second syllable in the disyllabic or sesquisyllabic words.

The phonetic forms corresponding to the reconstructed form *p.rak can be found in the Palaung-Waic, Khmuic, Vietic, Katuic, Bahnaric, and maybe Khmeric groups.[1]

In the Palaung-Waic group, which is located in the northernmost part of the Mon-Khmer linguistic area (currently southern China), as Ferlus (2011) indicated, the existing phonetic forms of the autonyms are *pəʐaək*, *parauk*, and *pʰalok,* depending on the dialects (Ferlus, 2011, p. 7). He also suggested that the phonetic form in the Khmuic group is *rɔːk*. However, as indicated by Nguyễn Văn Lợi (2009, p. 393), the Mang language, a member of the Khmuic group, currently uses the form [ha^3]. Thus, in these two language groups from the northern part of the Mon-Khmer linguistic area, the existing forms are either a disyllabic form with a closed main syllable (*parauk*) or a monosyllabic form, which is closed or open (*rɔːk*, ha^3).

In terms of the Vietic group, the author's fieldwork data show that the form used in the Arem language is [cmrawˀ], a disyllabic word in which the main syllable is open (Trần Trí Dõi, 2013, p. 121). Mã Liềng, another Vietic language, uses either [ɓrʊˀ] or [kari] depending on the dialects (Ferlus, 1995, p. 55) and also [kri] (Chamberlain, 2018). Those are all sesquisyllabic words in which the main syllables are all open. In the Katuic group, the form used in the Bru-Van Kieu language is [bru].

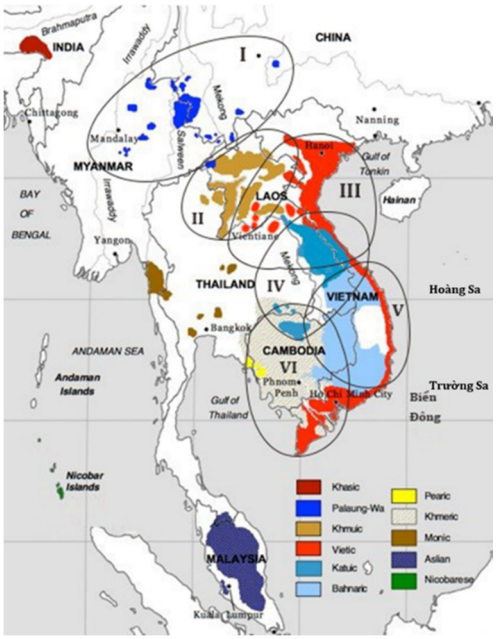
Meanwhile, in the Bahnaric languages, the Mạ ethnic group has another name, *châu Mạ* (Đặng Nghiêm Vạn, 2003, p. 223), which is most likely a syllabification of a phonetic cluster as in the case of the Arem language [cmrawˀ]: chau < [*cmrawˀ]. Sidwell and Jacq (2003, p.65) studied the Western Bahnaric subgroup and reconstructed the Proto-West Bahnaric word for 'person' as [*raa].

In the Khmer language, the form [mənuh] for 'person/people' is likely to be a Sanskrit loanword, but the form [proh], which means 'man, male human' could be an Austroasiatic-originated word.[2] Similar forms are used as autonyms for 'person/people' in six groups in the Khmer branch of the Austroasiatic family, and the phonetic correspondence between them and the aforementioned form [*p.rak] can be considered as strong proof that they originated from the same source. Specifically, the first consonant of monosyllabic autonyms or the first consonant of the main syllable in disyllabic autonyms must always be either [r] or [h, l]. The distribution of autonyms

[1] I has not had the opportunity to process data in other groups of the Mon-Khmer branch. However, the data of those six of nine groups of this branch can be considered as strong proof to the Mon - Khmer origin of these autonyms.

[2] I must thank Dr Frédéric Pain at the Laboratoire Langues et Civilisations à Tradition Orale (LACITO, Paris) for providing me with this data when we carried out a Khmer language fieldwork study in Southern Vietnam.

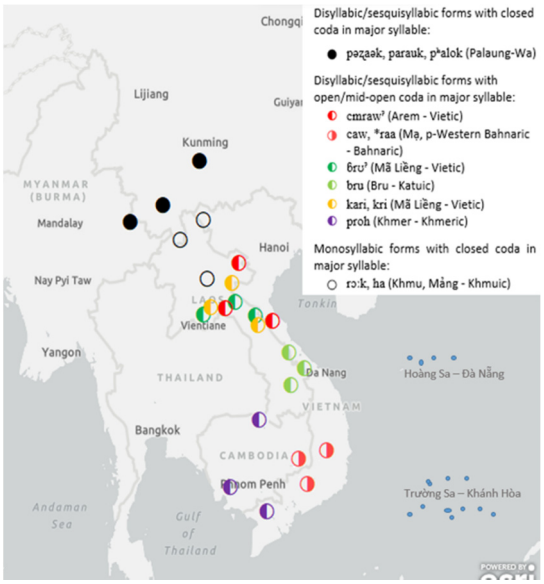
for 'person/people' in the six mentioned Mon-Khmer languages can be seen on the map below.



Map 1: Current location of six language groups in the Mon-Khmer branch
(Source: Gutman & Avanzati, 2013)

Using Gutman and Avanzati's (2013) Austroasiatic language family map as a reference, the author has circled the six above-mentioned language groups and shown the expansion of the Vietic inhabitants to the west: (I) Palaung-Wa (Palaung-Waic); (II) Khmu (Khmuic); (III) Viet language group (Vietic); (IV) Katu (Katuic); (V) Bahnar (Bahnaric); (IV) Khmer (Khmeric) languages.

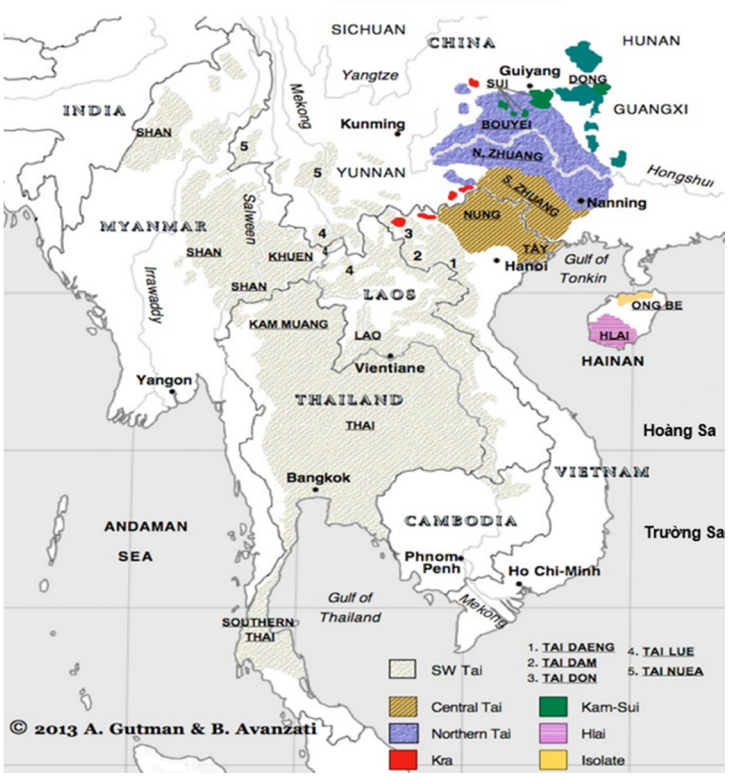
The next map shows the geographical distribution of the phonetic forms of autonyms for 'person/people' in each group. In this map as well as in the following maps, each form in each language group will be displayed in at least three representative locations.



Map 2: Geographical distribution of autonyms for 'person/people' among six language groups in the Mon-Khmer branch

2.2. Borrowed forms in the Tai and Sinitic groups

The distribution of loanwords for 'person/people' in the Tai group is shown in Map 3 below. Like Map 1, this map is by Gutman and Avanzati (2013); along with the Kam-Sui languages, they have also clearly mapped the boundary between the Southwestern Tai (SWT), the Central Tai, and the Northern Tai (NT) subgroups of the Tai group of the Tai-Kadai family. Among them, as Haudricourt (1953) indicated, the Southwestern Tai subgroup is part of the Tai population that migrated southward from the beginning of the first millennium CE until around the 10th century.



Map 3: Tai languages in Southeast Asia and Southern China

(Source: A. Gutman và B. Avanzati 2013)

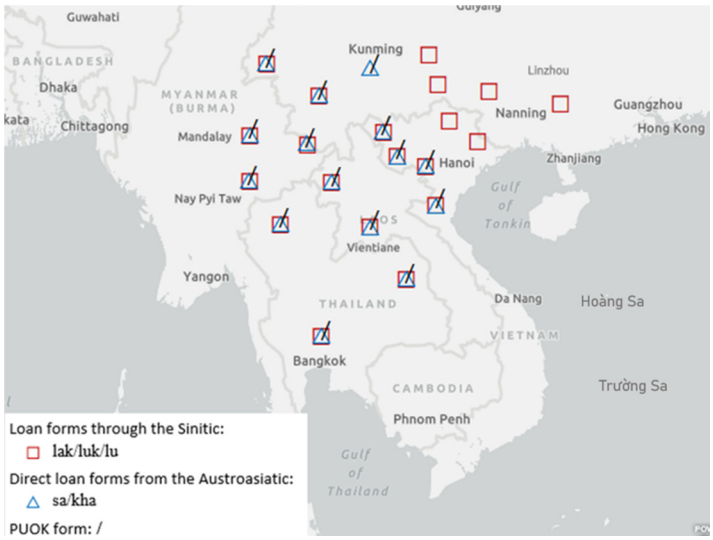
As shown in the author's 2023 paper, there are now two different phonetic forms of the autonym 'person/people' within the current Tai languages. One is directly borrowed from the Chinese cluster 雒 越, which was used to record the Mon-Khmer cluster [*p.rak *wat]. Most of the Kam-Sui and Tai languages have borrowed forms in which the first consonant of the main syllable is [l]. This can be found in the case of [lak] in the Maonam or Lakkia languages (Ferlus, 2011, p. 7); [lak] in the Zhuang language; or [luk, lu] in the Tày, Nùng, White Tai, Black Tai, Lao, and Thai languages (Trần Trí Dõi 2022, pg.520; Trần Trí Dõi 2023, pg. 6–7). The other, [puok], was also borrowed from the Chinese language, according to Chamberlain[3] (Chamberlain, 2016), but it is only preserved in the Southwestern Tai subgroup. Thus, the two forms of the

[3] Chamberlain (2016) wrote that the Ksing Mul people were called *Puak,* but in Vietnam, they can be called *Puok*, so I chose this spelling.

autonym for 'person/people' (lak/lu/lu and puok) that the Tai languages borrowed from the Chinese language are now preserved in two different geographical regions.

In addition, three different forms, *xá*, *kha,* and *kha,* related to the Austroasiatic autonym for 'person/people,' are currently used in the Southwestern Tai subgroup. As explained by Chamberlain (2016), this is the phonetic form that the Southwestern Tai languages borrowed directly from the Mon-Khmer languages when the Tai speakers migrated from southern China to Southeast Asia during the Proto-Southwestern Tai (PSWT) period. Chamberlain assumed that these two phonetic forms /*sa*/ and /*kha*/ that the inhabitants of the Southwestern Tai subgroup used to name the Mon-Khmer inhabitants are a consequence of the phonological change of the PSWT **khraaA1*: /*sa*/ or /*kha*/ < **khraaA1* PSWT. The form **khraaA1* in PSWT means 'Austroasiatic ethnic,' though it is used in other related languages as the general autonym for 'person.'

The distribution of the forms /puok/, /lak, luk, lu/, and /sa, kha/ can be seen in Map 4 below.



Map 4. Geographical distribution of loan words in the Tai languages

Regarding the Sinitic language, as aforementioned, from the perspective of historical linguistics, the Sino-Vietnamese cluster 雒 越 (Lạc Việt) was used by the Chinese people in the Old Chinese period (OC) and was recorded in the Hanshu (1st century CE) (Trần Trí Dõi 2017; 2022). The character 雒, vocalised in OC as [*p.rak], was an autonym referring to 'persons/people' who spoke the Mon-Khmer language. The form [puok], as Chamberlain (2016) explained, is also very likely the autonym for 'person/people' that the Chinese used to refer to the Austroasiatic inhabitants of the

Dian kingdom, which was pacified by the Chinese later than the 1st century CE (this area is now Yunnan province, China).

Map 5 shows the general geography of the Sino-Tibetan language family, and the geographical distribution of the Sinitic forms borrowed from Austroasiatic languages is shown in Map 6 below.

Map 5: General geography of the Sino-Tibetan language family

Map 6. Distribution of autonyms for 'person/people' in the Sinitic languages

3. Discussion

The distribution of autonyms for 'person/people' in the Mon-Khmer, Tai-Kadai, and Sino-Tibetan languages in Southeast Asia and southern China, as shown in Maps 2, 4, and 6, could contribute greatly to further discussion regarding the homeland and ethnic group of the people who used these autonyms during the Old Chinese period. First, consider the six language groups belonging to the Mon-Khmer branch examined in this paper. Taking into account the simultaneous existence of both monosyllabic and disyllabic or sequisyllabic autonyms in all six linguistic groups examined, it can be argued that [*p.rak], the reconstructed Old Chinese phonetic rendering of 雒, is related to the autonyms for 'person/people' in the Mon-Khmer languages. Additionally, the current distribution of those autonyms in those Mon-Khmer language groups which are distant from southern China (the Palaung-Waic, Khmuic, Vietic, Katuic, Bahnaric, and Khmeric groups) suggests that the autonyms must have originated in the Mon-Khmer languages. Furthermore, this argument seems to be strengthened by considering the geographical distribution of names for 'people' in the Tai-Kadai languages. While the forms [lak, luk, lu] now exist in the Tai-Kadai language family, including both the Kam-Sui group in the north and the Tai group in the south, the form /sa, kha/ is found only to a limited extent in the Southwestern Tai subgroup of the Tai group. In other words, the forms currently used in the Mon-Khmer language groups are essentially similar, while there are distinctions between the Southwestern Tai subgroups and the rest of the Tai-Kadai family. A possible explanation for this might be that those two forms which did not appear in Kam-Sui and two other subgroups of Tai were directly borrowed from Mon-Khmer languages while the form [lak, luk, lu] was a borrowed form of [*p.rak] from the Old Chinese language. These findings are further supported by the geographical distribution of the two forms [*p.rak] in OC and [puok].

These results taken together suggest that the historical linguistic proof does not support the argument of some Chinese researchers, including the authors Liang and Li (2017), that Zhuang-speaking members of the Tai group were the people referred to by the autonym Lạc Việt (雒 越). There are no Mon-Khmer or Tai historical documents identifying which group was called 'lạc việt/luòyuè' during the last centuries BCE and first centuries CE; thus, in this case, linguistic data can be considered as the most reliable basis for making the identification. Historical linguistics, as well as geographical distribution, provides evidence to support the notion that Lạc Việt is an autonym for Mon-Khmer speakers of the Austroasiatic language family, not members of the Tai language group.

References

Chamberlain, James R. (2016) Kra-Dai and proto-history of South China and Vietnam. *Journal of the Siam Society* 104: 27–76.

Chamberlain, James R. (2018) A Kri-Mol (Vietic) bestiary: Prolegomena to the study of ethnozoology in the Northern Annamites. *Kyoto Working Papers on Area Studies*, 133: 1–171. URI: http://hdl.handle.net/2433/231260

Đặng, Nghiêm Vạn (2003) *Cộng đồng quốc gia dân tộc Việt Nam* [*The Vietnamese national community*]. Ho Chi Minh City: Publishing House of Vietnam National University.

Đinh, Hồng Hải (2018) Lạc Việt: Từ tộc danh đến căn tính tộc người [Lạc Việt: From the ethnic name to ethnic identity]. *Journal of Social Sciences and Humanities* 6: 710–726.

Dinh, Hong Hai and Kelley, Liam C. (2021) Competing Imagined Ancestries: The Lạc Việt, the Vietnamese, and the Zhuang. In J. Gillen, L. C. Kelley, & L. H. Phan (Eds.), *Vietnam at the Vanguard: New Perspectives Across Time, Space, and Community* (pp. 89–107). Singapore: Springer.

Ferlus, Michel (1997) Le maleng brô et le vietnamien. *Mon-Khmer Studies* 27: 55–66.

Ferlus, Michel (2011, June 30–July 1) *Les Băiyuè, ou les "pays des (horticulteurs/mangeurs de) tubercules"* [*The Băiyuè, or the "country of (horticulturists/eaters of) tubers"*] [Conference presentation]. 24th Day of Linguistics of East Asia, East Asian Linguistic Research Center (EHESS CNRS), Paris, France.

Ferlus, Michel (2014) Arem, a Vietic language. *Mon-Khmer Studies* 43.1: 1–15 (ICAAL5 special issue).

Gutman, Alejandro, and Avanzati, Beatriz (2013) *Austroasiatic languages*. The Language Gulper. URI: http://www.languagesgulper.com/eng /Austroasiatic.html

Gutman, Alejandro, and Avanzati, Beatriz (2013) *Tai-Kadai languages*. The Language Gulper. URI: http://www.languagesgulper.com/eng/Taikadai.html

Gutman, Alejandro, and Avanzati, Beatriz (2013) *Tibeto-Burman languages*. The Language Gulper. URI: http://www.languagesgulper.com/eng/Tibeto.html

Haudricourt, André-Georges (1953) La place du vietnamien dans les langues austroasiatiques [The place of Vietnamese in the Austroasiatic languages]. *Bulletin de la Société de Linguistique de Paris* 1: 122–128.

Liang, Tingwang and Li, Sheng (Eds.) [梁庭望、厉声（主编）] (2017) 骆越方国研究（上下册）[A study of Luoyue Fangguo, Vol 1 & 2]. Beijing: The Ethnic Publishing House].

Nguyễn, Văn Lợi, Nguyễn, Hữu Hoành, Tạ, Văn Thông (2008) *Tiếng Mảng* [*The Mang language*]. Hanoi: Social Science Publishing House.

Sidwell, Paul, and Jacq, Pascale (2003) *A handbook of comparative Bahnaric*, *Vol.1: West Bahnaric*. Canberra: Pacific Linguistics.

Trần, Trí Dõi (2017) Trao đổi thêm về từ nguyên của yếu tố "lạc (luò 雒/駱)" trong tổ hợp "lạc việt (luòyuè 雒越)" [A further discussion on the etymology of "lạc (luò 雒/駱)" in the cluster "lạc việt (luòyuè 雒越)"]. *Journal of Museum and Anthropology* 18:41–53.

Trần, Trí Dõi (2022) *Lịch sử ngôn ngữ người Việt góp phần tìm hiểu văn hóa Việt Nam* [*The history of Vietnamese language in contributing to understanding Vietnamese culture*]. Hanoi: Vietnam National University Publishing House].

Trần, Trí Dõi (2023) Những dạng thức lưu giữ tên tự gọi (autonym) chỉ "người (person/people)" của những ngôn ngữ Môn-Khmer ở Đông Nam Á [Retainers of autonyms referring to people in the Mon-Khmer languages of Southeast Asia]. *Journal of Language and Life* 335: 5–13.

A Study of Three Korean Family Names: *Kim*, *Seo*, and *Baek*

Mitsuaki Endo
(Aoyama Gakuin University)

Abstract: The geographical distribution patterns of three Korean family names have been studied. *Kim* was found to have uniform distribution, *Seo* southern dominant distribution, and *Baek* northern dominant distribution. *Seo* may have originated in the ancient *Xu* country in China and have a common origin with certain toponyms, and *Baek* may have common etymology with the names of the ancient countries *Goguryeo* and *Baekje*. The geographical distribution of *Baek* and the findings of some ancient bronze mirrors with multiple loops attached in certain geographical areas seem to coincide.
Key words: family name, Korean peninsular, toponym, archaeology, Old Chinese

1. Introduction

This paper aims at examining the patterns of human groups' migration based on Korean family names. In this study, the most common family name *Kim*, a southern dominant family name *Seo*, and a northern dominant name *Baek* are examined.

The primary resource used for the study was published in 1934 under the name of Chosen Sotokufu (1934), of which real name was Zensho Eisuke. It is a collection of charts showing a number of family names in each administrative unit throughout Korea. The charts reflect the initial situation in Korea before World War II. After the war, Korea was divided into two parts, which resulted in population distribution changes as a result of the Korean War (1950–1953). Zensho (1943) analyzed the census and described the relevant villages. This paper is an attempt to analyze the name distribution via a geolinguistic method. The data were input by Ms. Tomoko Kurosawa.

According to Takeda (1984), family names of kings first appeared in 2nd c. BC, and ordinary people who dwelled in big cities began to acquire family names at the end of 10th c. AD on the Korean Peninsula. In fact, the information related to ordinary people was first recorded in a chapter on geography in the *Sejong Chronicle* in 15th c. AD. It is possible that the geographical distribution of family names at the time

ENDO, Mitsuaki. 2023. A Study of Three Korean Family Names: *Kim*, *Seo*, and *Baek*. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 12–20. doi: https://doi.org/10.5281/zenodo.8374562
* This work was supported by JSPS KAKENHI Grant Number JP23H00625.

reflected only the situation later. However, Suzuki (2022) revealed that the geographical patterns of Japanese family names correlated with the geographical distribution of the Y chromosome DNA haplogroups as well as that of the genomes, even though family names were inherited from fathers. Although the time stretch of the family names' occurrence was similar to that of the Korean name groups. In this study, it will be shown that the geographical patterns of different family names reflect ancient origins.

2. The name Kim

Kim is the most common family name of the 250 existing ones in Korea. It was the name of 3,821,496 families in Korea on October 1, 1930.

Figure 1 shows the geographical distribution of *Kim*. This map displays the absolute frequency as reported in each location. The frequencies are divided into 5 groups, the biggest being over 12,137 families, and the smallest just under 406. The frequency of the name occurrence is proportional to the size of each location.

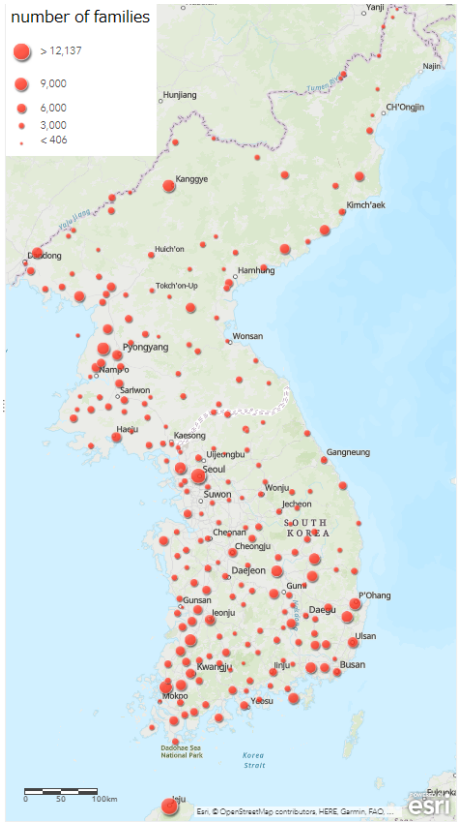


Figure 1: Map of *Kim*.

Figure 2 shows the relative ratio of the frequency to the size of each location. The statistics come from the national census of 1985, 2000, and 2015 in South Korea and have been provided by the Bureau of Statistics of the Korean government: https://sgis.kostat.go.kr/statbd/family_01.vw?topBottom_onoff=OFF, accessed on July 9, 2023. The ratio in *Kimhae*, which is the assumed place of origin of this family name (depicted with a gray border in the southernmost corner of the map from 2015), equates to around 23%, although the frequency based on the three censuses slightly differs.

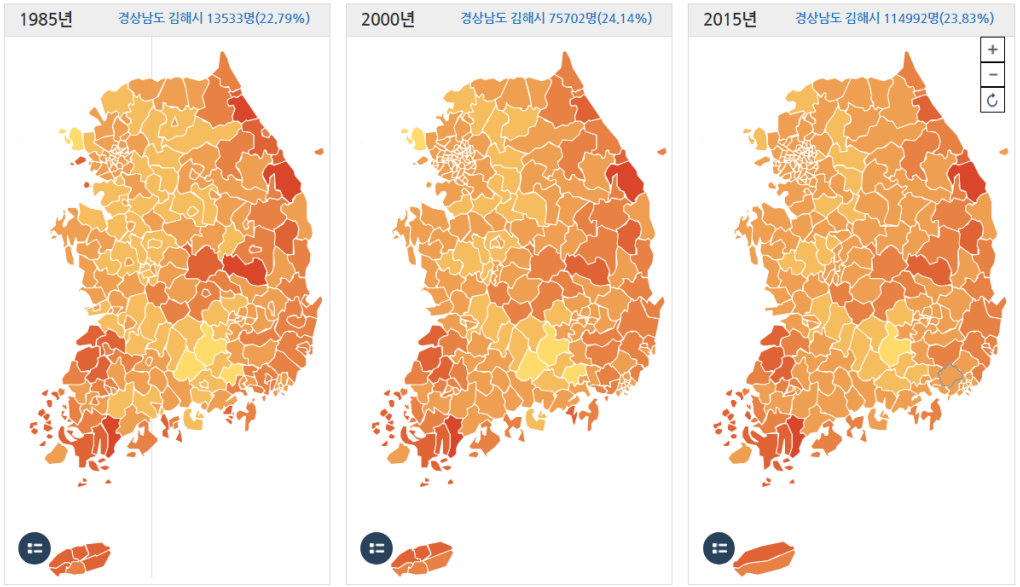


Figure 2: Relative ratio of frequency of *Kim* in each administrative unit.

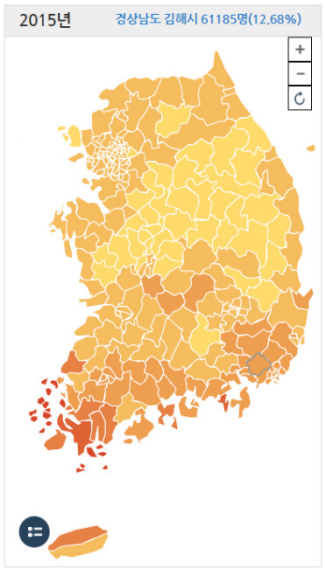


Figure 3: Kimhae *Kim*.

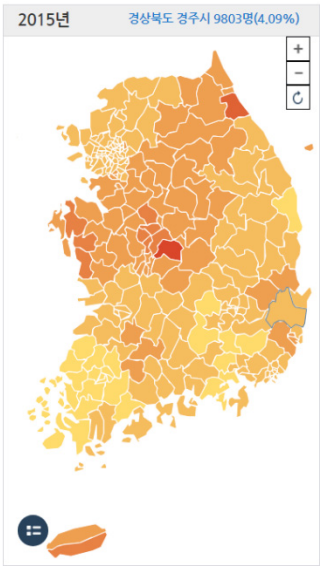


Figure 4: Kyungju *Kim*.

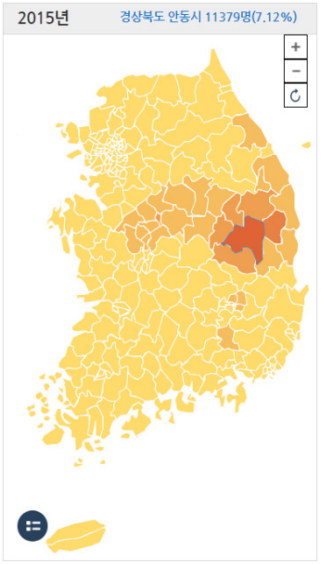


Figure 5: Andong *Kim*.

It is noteworthy that the areas (*bongwan* 本貫) from which the big clans originated are linked to family names. The Korean Bureau of Statistics provided maps of family names endemic to these places.

Figure 3 shows the relative ratio of the name *Kim* from *Kimhae*. It is dense in the southwestern part, while relatively low in the middle area. In contrast, the geographical distribution of *Kim* from *Kyungju*, as shown in Figure 4, seems to exhibit the opposite: The ratio is dense in the middle area and low in the southwestern part. This situation may reveal the migration direction of two human groups: *Kimhae*, once known as the biggest city in the Gaya kingdoms, is located in the southeast, and people migrated westward from there; while *Kyungju*, the capital of Silla, occupied an area more to the north. People migrated to the northwest, where the core area of *Baekje* used to be. Both *Kimhae* and *Kyungju*, the areas of origin of *Kim*, do not have the highest ratios of the name's occurrence.

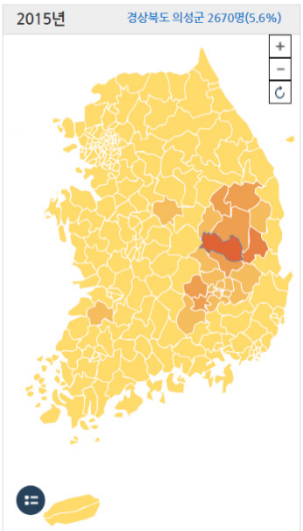


Figure 6: Uiseong *Kim*.

The distribution of *Kim* in *Andong* and *Uiseong* is shown in Figures 5 and 6, respectively. The centers are the areas of origin. Since these areas are much smaller and had less political and economic power than *Kimhae* and *Kyungju*, the areas to which the names spread from there were also smaller and easier to get hold of. In case of *Andong*, people migrated westward, possibly along the Nakdong River; while for *Uiseong*, the direction of migration seems to be northward and southward along the Geumho River, a branch of the Nakdong River.

3. The name Seo

Figure 7 shows three different types of classification based on the same data for the family name *Seo*. The first figure shows five groups with the name held by fewer than 19, 90, 150, 250, and more than 310 families, respectively, and it represents the southwestern dominant distribution. The second figure shows the areas with the name held by fewer than 19, 200, 350, 500, and more than 684 families, respectively, and it represents the southern dominant distribution. The third shows the areas with the name held by fewer than 19, 300, 600, 900, and more than 1,235 families, respectively, and it shows less of a difference in terms of the area size. Figure 8 is the relative ratio map of *Seo* provided by the Korean Bureau of Statistics. It is similar to the second map of

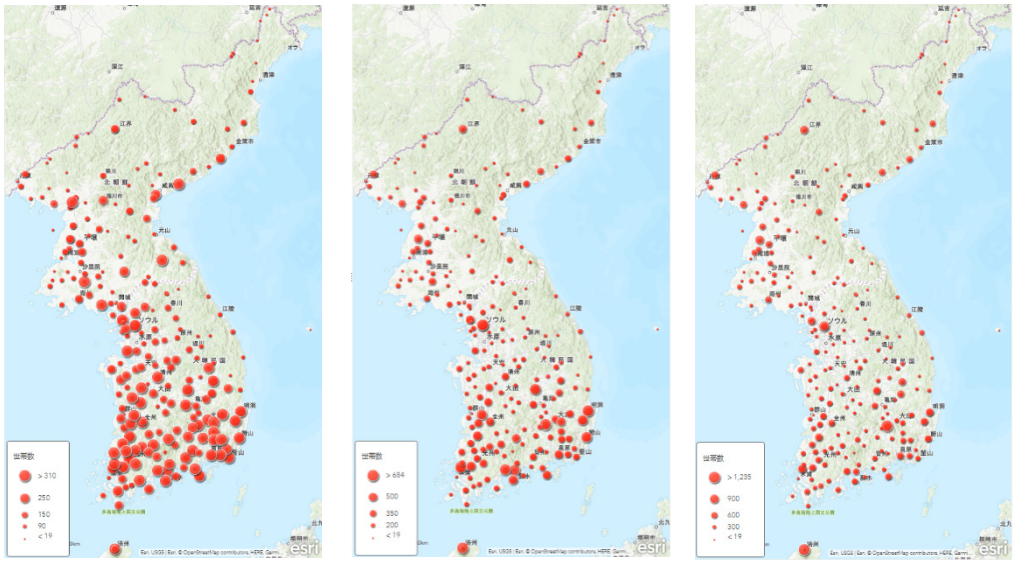


Figure 7: Map of *Seo*.

Figure 7. Here, we see the difficulty in accessing the real tendency of the name occurrence.

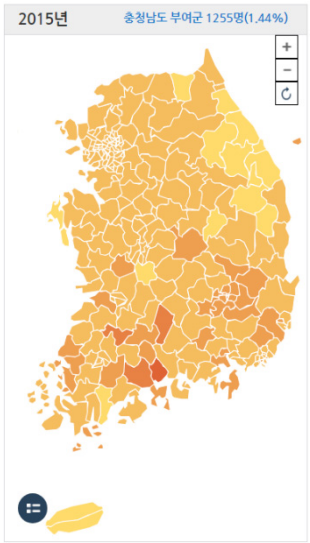


Figure 8: Relative ratio of *Seo*.

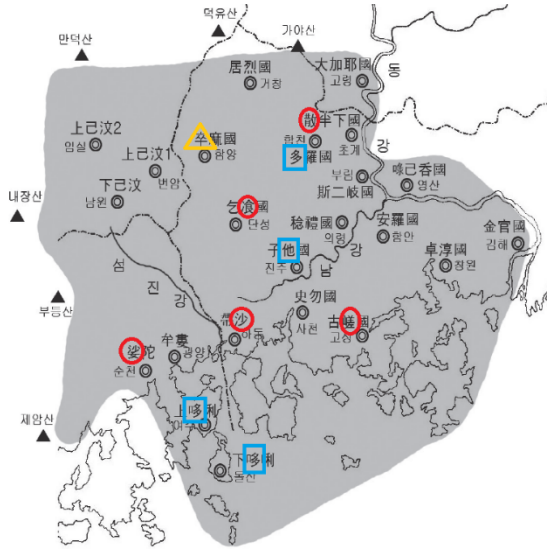


Figure 9: *Sa, So*, and *Ta* in Gaya.

Seo 徐 also appears in toponyms. The most famous one is 徐羅伐 *Seorabor,* which was the capital of Silla (57 BC–AD 935). *Bor* means "village" or "city" in Korean. The meaning of *ra* is "place," or even, possibly, the second element of the Old

Chinese *sə.la for 徐, according to Baxter and Sagart (2014). Here, it served as the meaning of "capital." The present capital name Seoul came from Middle Korean *Seoßeur*, in which the previous *ra* disappeared, at the same time when the intervocalic [b] underwent lenition to [β], and further disappeared .

Although the Chinese characters used to spell that name were different, the word could be seen frequently among ancient toponyms in the southeastern part of the Korean Peninsula. The base map in Figure 9 has been adopted from Kim (2010: 172). It shows ancient toponyms and their estimated positions in 5th to 6th c. AD, which came from *the Chronicles of Japan* 日本書紀. A red circle denotes the sound *sa, a blue square denotes the sound *ta, and a yellow triangle denotes the sound *so of the same word 徐. *sa reflects its Old Chinese pronunciation, and *so is the Middle Chinese pronunciation, while *ta reflects the result of the sound change of *sa > *a. The etymology of the word *Seo* 徐 may have come from the *Xu* 徐 country (circa 20 BC–512 BC) of China. Its migration to other places is to be discussed in a future paper.

4. The name Baek

Figure 10 shows the geographical distribution of the name *Baek* 白. It is dense near the northwestern border with China, and there are three more areas with frequent occurrence of the name from north to south: along the Taedong River, the Geum River, and the middle basin of the Nakdong River. Its distribution is similar to the geographical distribution of the bronze mirrors with multiple loops attached (see Figure 11, cited from Jeon 1991:453). Jeon (1991) inferred that this reflects the migration of King Jun 準王 from 194 BC to 180 BC.

Figure 12 has been cited from Kobayashi (2019:275). According to his chronological analysis, the original type of the mirror was found in Xiaoheishi 小黒石 in the western Liaoning area and dated back to the Upper Xiajiadian culture in the last half of the Western Zhou dynasty (until 8th c. BC). It spread to East Liaoning around 7th c. to 6th c. BC. Further, it spread to Northwest Korea before 5th c. BC. The type found in Southwest Korea spread directly from West Liaoning. Kobayashi (2009:43) infers that it arrived in the Geum River area before the first half of 4th c. BC, if not sooner, and it spread to North Kyushu almost at the same time. According to Kobayashi's chronological analysis, the spreading does not have anything to do with the historical migration of King Jun.

The Chinese character 白 is used to represent a part of the sound of the word *maek* 貊, an ancient tribe of the ancient Korean Peninsula. 貊 can be considered as the

etymology of the first syllable of *Baekje* 百済. There is a myth about the name *Baekje* that it came from *Goguryeo* 高句麗, which was spelled as *bökli* in the Orkhon inscription of 8th c. AD and was written in the Old Turkic alphabet 突厥文字. *Bökli* may be interpreted as composed of *Baek* 貊 and 夷 (barbarian), *gogu* stands for *bök*, while *ryeo* stands for *li*. *Baek* 貊 is also written as 貉, in which the phonetic part is 各, with the initial consonants *kr-. 百, corresponds to the Tibetan b-r-gya, hence it is thought to come from the Old Chinese or common Sino-Tibetan *pˁrkˁak, rather than the Old Chinese *pˁrak as claimed by Baxter and Sagart (2014). Thus, the family name *Baek* may stand for the ruler class originating from West Liaoning, which spread to the Korean Peninsula, and has a similar sound as that of the ancient countries of *Goguryeo* and *Baekje*.

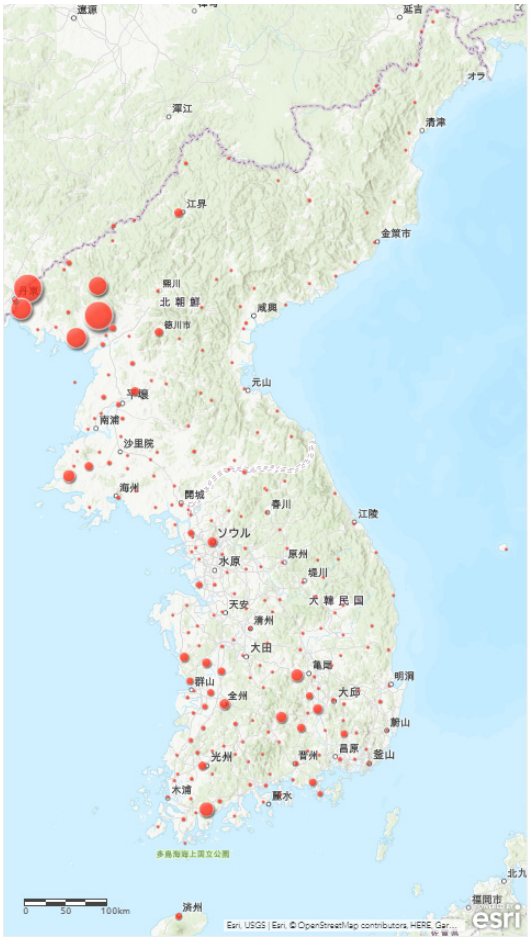


Figure 10: Map of Baek.

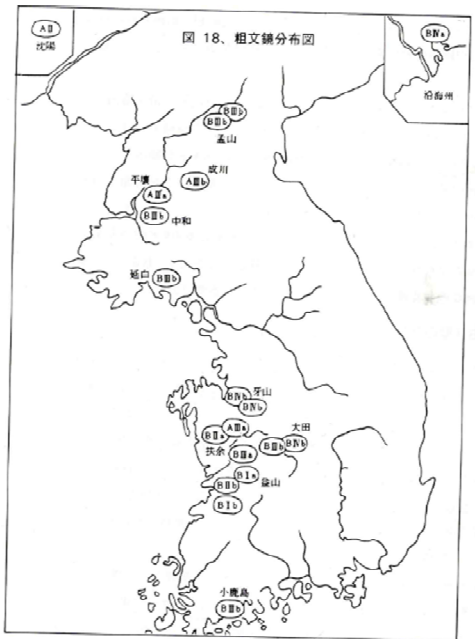


Figure 11: Geographical distribution of bronze mirrors with multiple attachmentloops (adopted from Jeon 1991)

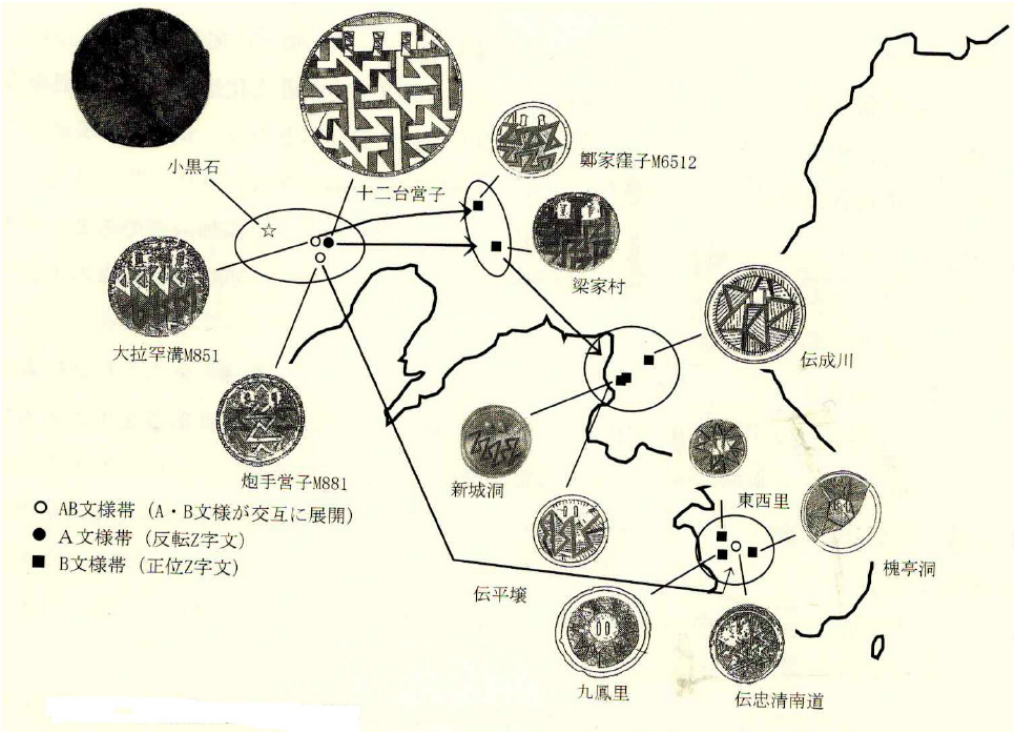


Figure 12: Geographical distribution of bronze mirrors with multiple loops attached (adopted from Kobayashi 2019).

5. Final remarks

Above is a rough sketch that shows the possible reflection of family names on human groups' migration patterns. There are many other family names to be analyzed, which may become the subject of future studies.

References

Baxter, William H., and Laurent Sagart (2014) The Baxter-Sagart reconstruction of Old Chinese (Version 1.1, 20 September 2014). URL: https://ocbaxtersagart.lsait.lsa.umich.edu/

Chosen Sotokufu [朝鮮総督府] (1934) *Chosen no sei*『朝鮮の姓』 [Family names in Korea]. Reprinted in 1977. Tokyo: Daiichi Shobo.

Jeon, Yeonglae [全栄来] (1991) *Kankoku seidoki jidai bunka kenkyu* 『韓国青銅器時代文化研究』 [Studies on culture during bronze age in Korea]. Jeonju: Sin-a Publishing.

Kim, Taesik [김태식] (2010)「古代 王權의 成長과 韓日關係－任那問題를 包含하여－」[Development of Ancient Royal Authorities and the Relationship between Korea and Japan: Including the Nimna Problem] 『제2기 역사 공동 연구 보고서』[Second Report of Cooperative Studies on History] Vol. 1: 125-254. Seoul: Committee of the Korea-Japan Cooperative Studies on History. URI: https://www.jkcf.or.jp/wordpress/wp-content/uploads/2010/10/1-allk.pdf

Kobayashi, Seiji [小林青樹] (2009) Yayoi seidouki saishi no kigen to Ryonei seidouki bunka 弥生青銅器祭祀の起源と遼寧青銅器文化 [Several issues about the bronze culture of Liaoning and the origin of the bronze ritual object of Yayoi culture]. In *Bulletin of the Research Center for Traditional Culture, Kokugakuin University* 『國學院大學伝統文化リサーチセンター研究紀要』1: 35-45. DOI: 10.57529/00001963

Kobayashi, Seiji [小林青樹] (2019) *Yayoi bunka no kigen to higashi Ajiya kinzokuki bunka* 『弥生文化の起源と東アジア金属器文化』[The origin of Yayoi culture and the culture of metal objects in Eastern Asia]. Tokyo: Hanawa Shobo.

Suzuki, Rumiko [鈴木留美子] (2022) Nihonjin no myoji no chiikisei to idengaku 日本人の名字の地域性と遺伝学 [Locality of Japanese surnames related with genetics]. *Studies in Geolinguistics* 2: 1-28. DOI: 10.5281/zenodo.7121484

Takeda, Yukio [武田幸男] (1984) Chosen no seishi 朝鮮の姓氏 [Family names in Korea]. In: (2022) *Shiragi seiji shakaishi kenkyu*『新羅政治社会史研究』 [Studies on the history of politics and society in Silla], 339-388. Tokyo: Bensei Shuppan.

Torigoe, Kenzaburo [鳥越憲三郎] (1992) Kodai Chosen to Wazoku 『古代朝鮮と倭族』 [Ancient Korea and the Wa tribe]. Tokyo: Chuo Koronsha.

Zensho, Eisuke [善生永助] (1943) *Chosen no seishi to dozoku buraku*『朝鮮の姓氏と同族部落』 [Family names and Con-sanguineal Villages in Korea]. Tokyo: Toko Shoin.

Variation in basic roots across dialects in nDrapa: Geolinguistic analysis in the Qiangic context

Satoko Shirai
(The University of Tokyo)

Abstract: This paper examines the geographical distribution of three vocabulary items in the Qiangic area that have with divergent roots among nDrapa dialects: 'sand', 'leaf', and 'knee'. The local diversity of word forms in the nDrapa dialects is considered in the broader perspective, including the Qiangic languages. Linguistic maps of the three items illustrate three different histories. Non-Qiangic word forms for 'sand' and 'leaf' are distributed in Central nDrapa, whereas loanwords have diffused from the northern and southern peripheries. As for 'knee', two types of word forms have cognates in Qiangic languages. This paper demonstrates the need for caution in conducting comparative linguistic research on the nDrapa language.
Key words: basic word, dialectology, geolinguistics, language contact, nDrapa, Qiangic

1. Introduction

The Qiangic languages are spoken in a multilingual area called the Western Sichuan Ethnic corridor or the Tibet-Qiang-Yi corridor in Southwest China. The languages of this area, share many typological features, although their systematic relationship has not been identified using the methods of comparative linguistics. nDrapa (ISO 639-3: zhb) is one of the languages spoken here.

This study was conducted to clarify the relationship between nDrapa and nearby Qiangic languages. In a previous study of nDrapa dialects (Shirai and Huang forthcoming), it was found that some items of basic vocabulary exhibit diverse word forms that could be traced back to different roots. This paper seeks to find corresponding roots in Qiangic languages and examines their distribution to identify their historical formation. Then, using this distribution as a cue, we discuss the relationship between nDrapa and Qiangic languages.

SHIRAI, Satoko. 2023. Variation in basic roots across dialects in nDrapa: Geolinguistic analysis in the Qiangic context. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 21–33. doi: https://doi.org/10.5281/zenodo.8374568

* The work reported in this presentation was supported by JSPS KAKENHI Grants No. JP18H05219 and No. JP19K00543.

1.1. nDrapa and the Qiangic linguistic area

nDrapa belongs to the Sino-Tibetan language family, and it has about 10,000 speakers, living in Southwestern China. Figure 1 presents a map of nDrapa villages surveyed by Shirai and Huang (forthcoming). We can divide nDrapa dialects into three main groups: Southern, Central, and Northern, but there is general intelligibility across all dialects. nDrapa has no orthography. It has experienced relatively heavy language contact with Tibetan, and Khams Tibetan has been the traditional lingua franca of this area. In the modern age, Southwest Mandarin is the superstratum.

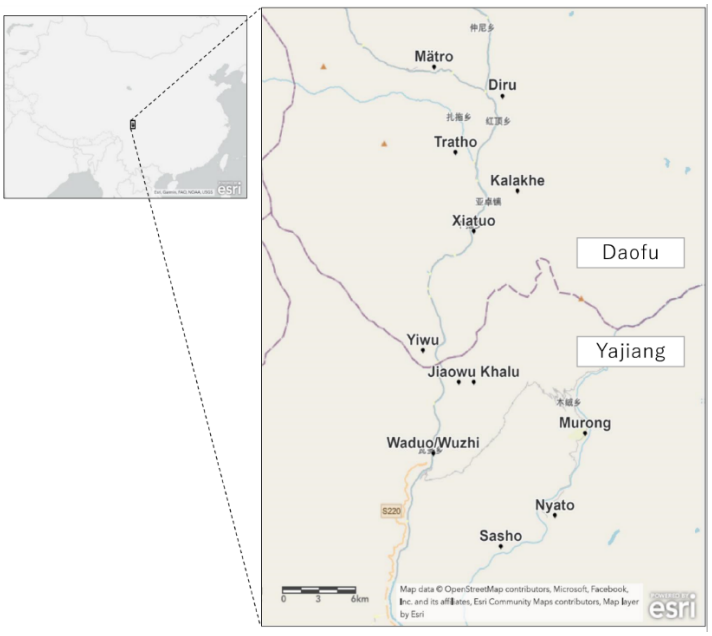


Figure 1: nDrapa dialects (Shirai and Huang forthcoming)

Figure 2 (taken from Shirai 2020) illustrates the distribution of the Qiangic languages and nearby languages. The nDrapa language is spoken in the areas marked with red square outlines. In their proposal of a genetic tree for Qiangic, Jacques and Michaud (2011: Appendix) excluded the southern subgroup of languages, which were included in Qiangic by Sun (2001), and they divided the language group into the rGyalrongic subbranch and other languages. In Figure 2, the languages of the rGyalrongic subbranch are marked with blue triangles, while the languages that were not included in their Qiangic group are marked with brown seals. More recent studies divided rGyalrongic into rGyalrong proper and Western rGyalrongic, to which Tangut,

an extinct language with Qiangic features, belongs (Lai, Gong, Gates, and Jacques 2020). However, the attribution of nDrapa remained unclear.

Conversely, other studies, including Chirkova (2012) and Shirai (2020), cast doubt on the genetic unification of the Qiangic languages, and they proposed that the languages form a linguistic area. In this study, we tentatively use the term Qiangic languages without intending to answer the question of their relationship.

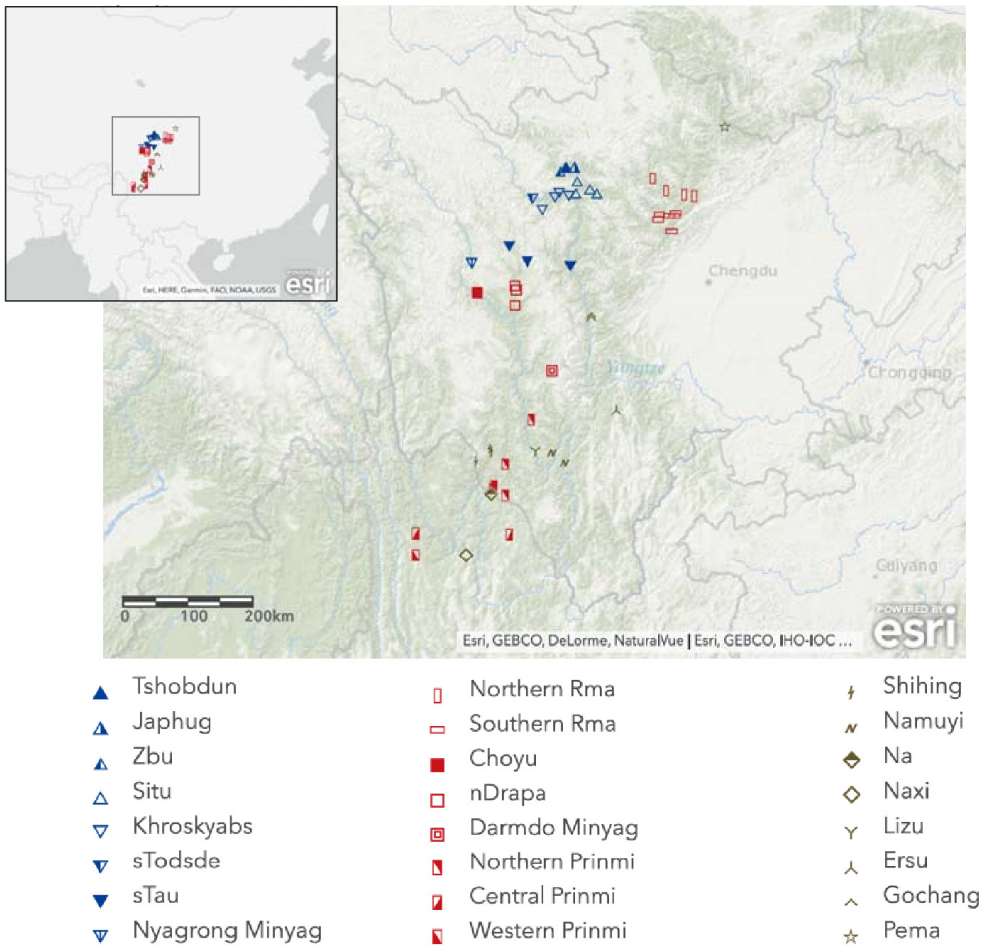


Figure 2: Languages of the Qiangic linguistic area (Shirai 2020: 369)

1.2. Previous studies

This study uses nDrapa dialect data collected and analyzed in a previous study (Shirai and Huang forthcoming). There, a dialect-geographical study was conducted in 13

villages based on the Swadesh 100 basic word list. From these data, we proposed three dialect groups. However, some items, such as 'knee', 'leaf', and 'sand' showed different roots in a contrast between central and peripheral dialects (Table 1). It is well known that in Japanese dialects, for instance, peripheral regions tend to preserve older forms. However, in the case of nDrapa, this may not be an appropriate assumption because roads to the north and south of the nDrapa region connect Tibetan areas to Han Chinese areas. Therefore, Shirai and Huang (forthcoming) hypothesized that the central dialects are conservative and that the peripheral regions have received influence from other languages. However, here, we only focused on the distributions of nDrapa dialect forms, without regarding other Qiangic languages.

Table 1: Example word forms in nDrapa dialects

	Northern		Central		Southern	
	Mätro	Kalakhe	Yiwu	Jiaowu	Murong	Nyato
'sand'	tsemá	$ts\imath^{55}ma^{33}$	$ŋ_e^{55}lɯ^{33}$	$ɲe^{55}lə^{33}$	$hts\eta^{55}ma^{33}$	$^{p}ts\eta^{55}ma^{33}$
'leaf'	lomá	$sɛ^{11h}pa^{33}la^{11}$	$ŋ_a^{33}ŋ_a^{53}$	$ɲə^{55}ɲa^{33}pa^{33}la^{33}$	$ɲə^{55}ɲa^{33}pa^{33}la^{33}$	$lo^{55}ma^{55}$
'knee'	pŭ	pu^{33}	pu^{24}	$mɛ^{33}ŋgi^{33}$	$mɛ^{33}ŋgi^{33}$	pu^{35}

In this study, I further develop the relative chronology of these diverse roots from the viewpoint of the geolinguistic study of the Qiangic languages. Table 1 shows examples of nDrapa dialectal word forms of the three vocabulary items that this paper uses up to analyze geographical distribution in the Qiangic linguistic area. Table 2 illustrates the languages and dialects that this study refers to with the classification based on Jacques and Michaud (2011) in the left columns and data sources in the right column.

Table 2: Language classification and data sources

Classification			Languages and dialects	Data sources
Na-Qiangic	Qiangic	rGyalrongic	Yelong Khroskyabs, Guanyinqiao Khroskyabs	B. Huang (2007)
			Daofu Stau	Huang (ed.) (1992)
			Geshitsa	Duoerji (1998)
			Zbu, Tshobdun, Japhug, Situ, sTodsde, Wobzi Khroskyabs, Erkai, Nyagrong Minyag	Nagano & Prins (2013)
			Yoci bTsanlha, Bola Situ	Shirai's fieldnotes
		Other Qiangic	Ronghong N. Rma	LaPolla with Huang (2003)
			Puxi S. Rma	C. Huang (2007)
			Ekou N. Rma, Mawo N. Rma, Longxi S. Rma, Taoping S. Rma, Mianchi S. Rma	Evans (2001)
			Darmdo Minyag, Youlaxi Choyu, Tratho nDrapa	Huang (ed.) (1992)
			Lhagang Choyu	Suzuki & Sonam Wangmo (2018)
			Kara Choyu	Nagano & Prins (2013)
			Diru nDrapa, Kala nDrapa	by courtesy of Hiroyuki Suzuki
			Mätro nDrapa, Kalakhe nDrapa, Nyato nDrapa	Shirai's fieldnotes
			Xiatuo nDrapa, Yiwu nDrapa, Jiaowu nDrapa, Khalu nDrapa, Murong nDrapa, Waduo nDrapa	by courtesy of Yang Huang
			Sasho nDrapa	Y. Huang (forthcoming)
			Waduo-Wuzhi nDrapa	Gong (2007)
			Xinyinpan Central Prinmi	Ding (2014)
			Other Prinmi dialects	Lu (2001)
	Other Na-Qiangic		Maibeng Gochang, Upper Shihing, Lizu, Luobo Namuzi	Huang (ed.) (1992)
			Shihing	Sun et al. (2014)

2. A geolinguistic analysis of three lexical items

This section provides a geolinguistic analysis of the lexical items 'sand', 'leaf', and 'knee' which are included in Swadesh's 100 basic words. However, nDrapa dialects show various roots for them. This suggests that they are not stable words (Matisoff 2009) in nDrapa dialects or perhaps in Qiangic languages. A geolinguistic examination of such non-stable basic words may shed light on linguistic history of the area.

Shirai and Huang (forthcoming) discussed the distribution of the word forms for 'sand' and 'leaf' in nDrapa dialects. However, when these are examined in the broader Qiangic linguistic area, we may draw other conclusions.

The hash mark (#) is used in the following discussion to indicate earlier word forms tentatively extrapolated from dialectal forms in the absence of sufficient clarification of a phonological correspondence.

2.1. 'Sand'

Shirai and Huang (forthcoming) described the distribution of word forms denoting 'sand' in nDrapa dialects as follows: the northern and southern areas show similar forms, but the central area has completely different ones (as seen in Table 1 above). The northern and southern forms are based on Tibetan loanwords from Written Tibetan *bye ma*. Therefore, Shirai and Huang (forthcoming) concluded that indigenous vocabulary is retained in the central area, while Tibetan loans are seen in both the south and the north.

Figure 3 presents the geographical distribution of the word forms for 'sand' in the Qiangic linguistic area. Red circles indicate word forms with bilabial initials, e.g., Bola Situ *kə-mbjek*, Northern Prinmi *bi*^{13} *pə*^{13}, and Shihing *mei*^{35}. The green boxes indicate the palatal nasal type, as seen in Central nDrapa. Brown lines indicate the type of Tibetan loan, e.g., Puxi sTodsde *pjə'ma*. Blue triangles in the eastern regions indicate the type of Chinese loan, e.g., Ronghong Northern Rma *ʂats* (from 沙子 *shāzi*) and Lizu *xuo*^{33}*ʂa*^{53} (from 河沙 *héshā*). In the cases of these loans, both syllables of a disyllabic word correspond to the original Tibetan or Chinese forms. This fact is evidence that the word is not indigenous but a loan, as the Sino-Tibetan languages tend to be strongly monosyllabic, and disyllabic compounds generally do not show a correspondence between Qiangic languages and Tibetan or Chinese.

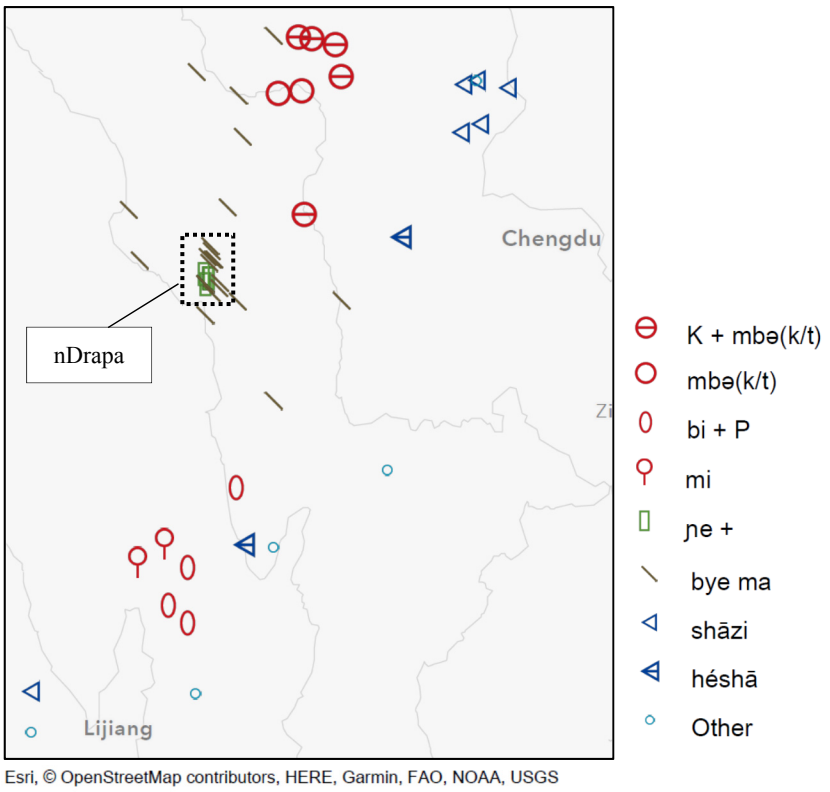


Figure 3: ‘Sand’ in the Qiangic area.

The geographical distribution shown in Figure 3 suggests that the bilabial type is the oldest in Qiangic languages. Alternatively, we cannot identify a palatal nasal type outside of the central nDrapa region; that is, it has no cognates in the Qiangic languages. Therefore, from a geolinguistic viewpoint, the palatal nasal type in Central nDrapa could be later compounds created in the region.

2.2. ‘Leaf’

Shirai and Huang (forthcoming) concluded that the $^{(h)}$*pala* type, which is found in Sasho Southern nDrapa and Diru Northern nDrapa, is the oldest dialectal variant. Dialects such as Kalakhe, Jiaowu, and Murong in Table 1 feature a compound involving this form. However, it was lost in the western area, replaced with word forms that might have been derived from the word for ‘green’, e.g., *ŋa*^{33}*ŋa*^{53} in Yiwu Central nDrapa (Table 1). Moreover, the Tibetan loanword (Written Tibetan *lo ma*) is found in the northernmost and the southernmost parts of the area, showing the same pattern as ‘sand’.

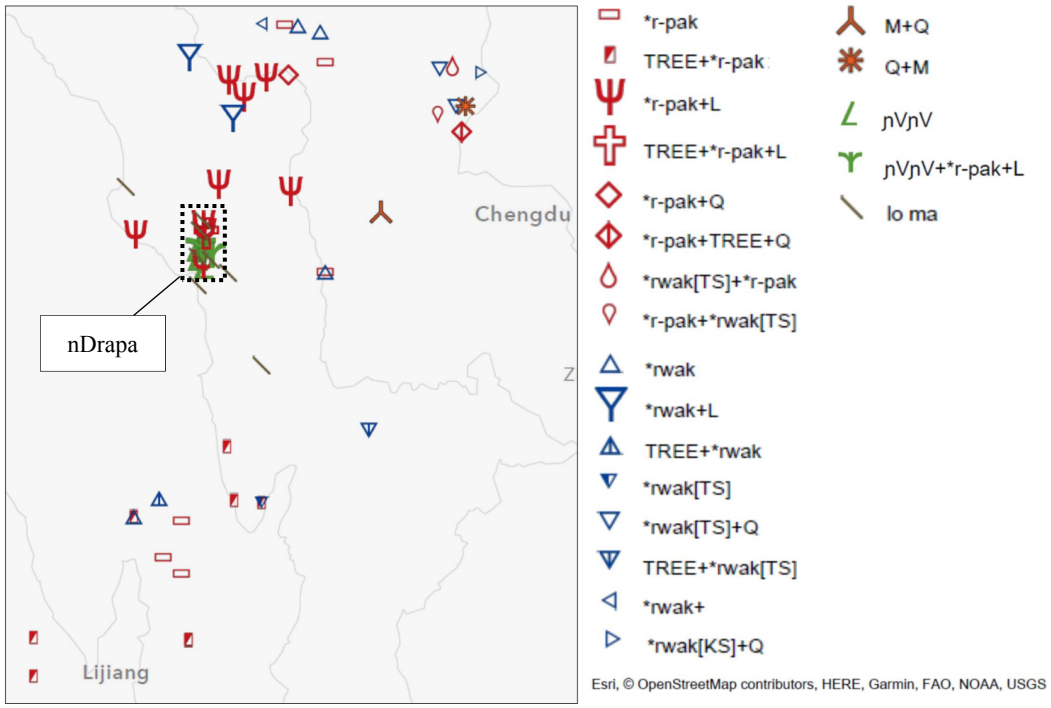


Figure 4: 'Leaf' in the Qiangic area.

Figure 4 presents the distribution of word forms denoting 'leaf' in the Qiangic area. In this map, the red marks indicate word forms featuring Proto-Tibeto-Burman (PTB) *r-pak.[1] The *[h]pala* type in nDrapa can be classified as this type, although the etymology of the second syllable remains unclear. The blue marks indicate word forms derived from another PTB root, *rwak. Orange marks indicate other types. Green marks indicate a palatal nasal type, while brown lines indicate a Tibetan loan. The geographical distribution of these suggests that the *r-pak type is the oldest form in the Qiangic languages. Conversely, the palatal nasal initial type is only found in nDrapa. We conclude that this type was formed later.

Moreover, we find disyllabic or polysyllabic word forms with second *l*-initial morphemes in a relatively broad area across the northwest. Many rGyalrongic languages have a similar morpheme to nDrapa, e.g., Puxi sTodsde *'rbala*, Wobzi Khroskyabs *rp[h]álá*, and Daofu Stau *lba lə*. Among the non-rGyalrongic languages, Youlaxi Choyu, which is spoken in the area adjacent to nDrapa also has a correspondent form, $ba^{13}la^{55}$. A possible etymon for this morpheme is PTB *s-la / *s-lap 'LEAF/TEA/FLAT THING'. Qiangic languages in other regions, such as Prinmi, do

[1] The PTB forms in this paper are cited from the STEDT database (see Matisoff 2015).

not have the second *l*-initial morpheme, but they do have a root derived from PTB *r-pak. Therefore, in this case, *r-pak + L type is a regionally diffused word form.

2.3. 'Knee'

Now we move to a vocabulary item that has not been discussed in detail previously: 'knee'. We can classify word forms denoting 'knee' into two types across in nDrapa dialects: Type A, which involves an unaspirated bilabial plosive initial and rounded closed back vowel: #pu, and Type B, which consists of two syllables, of which the first has a bilabial nasal initial and a mid-front vowel, while the second has a prenasalized velar plosive and a high-front vowel: #mengi. Among the examples in Table 1, Mätro, Kalake, Yiwu, and Nyato have Type A, while Jiaowu and Murong have Type B. There are also subtypes of Type A that have a second syllable with lateral initial: the P+L type, e.g., *pu*^{33}*lo*^{55} in Tratho, and the P+L+ type, e.g., *ˊpu lo wu ŋa* in Diru. In Figure 5, Type A (P, P+L, P+L+) is marked with red boxes whereas Type B (#mengi) is marked with blue triangles.

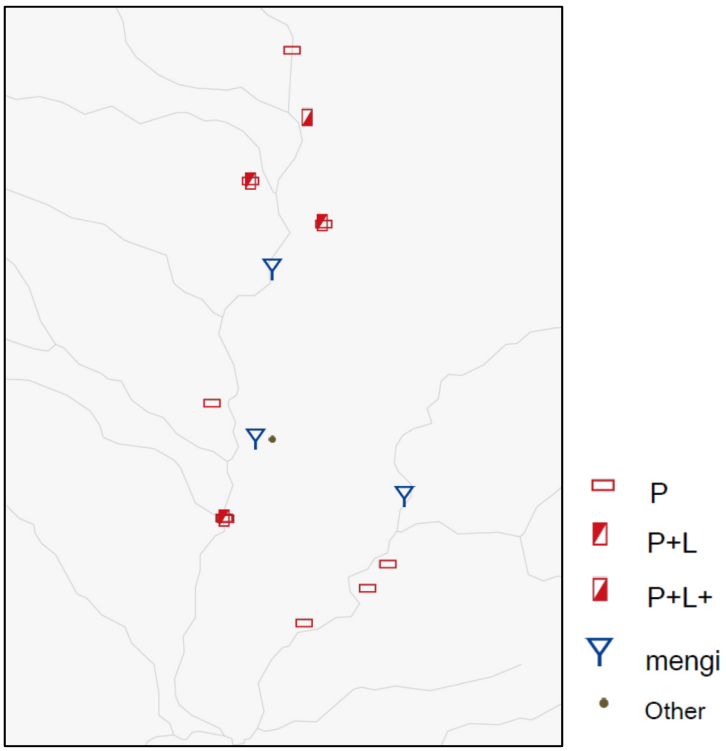


Figure 5: 'Knee' in nDrapa dialects.

For the geographical distribution, we find the monosyllabic Type A in the peripheral regions, Type B in the central area, and the compounded subtypes of Type A occurring between these two types. If we adopt the previous hypothesis that is, conservative central and interfered peripheral (Shirai and Huang forthcoming), we can conclude that Type B is the oldest. Moreover, the root #pu for Type A is like the root for the Written Tibetan form, *pus-mo*. However, the second morpheme, *lo* in nDrapa and *mo* in Tibetan, does not correspond. Therefore, it is difficult to identify it as a borrowing, and consequently, the relative chronology of Type A and B is unclear.

Let us examine the Qiangic word forms. Figure 6 presents the geographical distribution of word forms for 'knee' across the Qiangic linguistic area. The red marks indicate the distribution of Type A, the *p*-initial type, while the blue triangles show the distribution of Type B, with velar or uvular initial roots. While only Central nDrapa has the disyllabic #mengi type, its second syllable has a characteristic in common with Type B in Qiangic. Both types are widespread in the Qiangic area. For example, Type A: *tə'xpom* in Zbu, $py^{55}mu^{55}$ in Southern Prinmi; Type B: *ʐguəq* in Yadu Northern Rma, $ʂŋi^{55}$ in Youlaxi Choyu. Moreover, Type A is not found in the area adjacent to the distribution of Type A in the nDrapa dialects. In this case, therefore, we must interpret the distribution in a way other than through language contact.

I tentatively assume that there were two earlier forms in nDrapa: #me-ŋgi and #pu(-lo). Some Qiangic languages feature different words for 'knee hollow' and 'kneecap': for example, Qinghua Southern Prinmi $ka^{55}pu^{55}$ 'knee hollow' vs. $py^{55}mu^{55}$ 'kneecap.' Therefore, we can assume that at an earlier stage, the two words, #me-ŋgi and #pu(-lo), referred to two different notions but were closely related. Then, nDrapa dialects lost the distinction and adopted one of the two forms to denote 'knee' as a whole. This can only be a tentative conclusion for nDrapa words for 'knee.'

We can find a possible cognate of the second morpheme of the first type, #-ŋgi, in Tangut, an extinct Western rGyalrongic language of the 11th to 13th centuries: $ngẹ^{2}$~$rngẹ^{2}$ 'knee' (Sofronov 1978 via STEDT) or *ŋɣɛ* 'knee' (Nishida 1964 via STEDT). Moreover, some other Qiangic languages show possible cognate forms: Yelong Khroskyabs from Western rGyalrongic $rŋe^{55}$ and Taoping Southern Rma $χŋu^{55}$. However, the origin of #me- is unknown.

For the second type, #pu(-lo), we can find possible etyma, such as PTB *pu KNEE; *put-s ⩙ *pit-s KNEE, and PLB *put 'knee.' These also correspond to the Written Tibetan *pus-mo* 'knee.' However, the origin of #-lo is unknown.

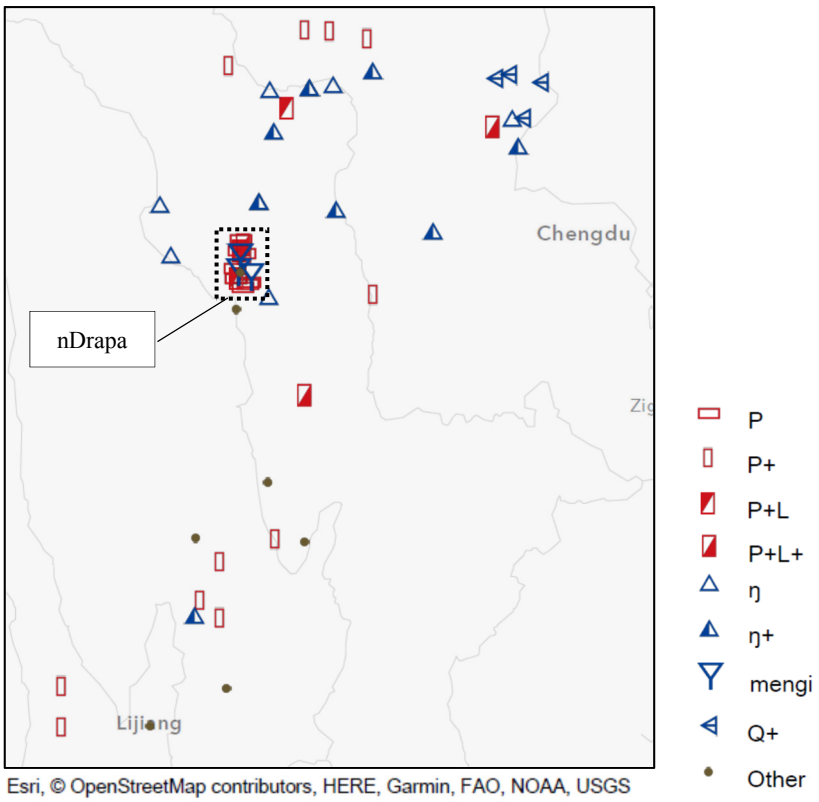


Figure 6: 'Knee' in the Qiangic area.

3. Conclusion

This paper examined the geographical distribution of three vocabulary items across the Qiangic area that have with different roots among nDrapa dialects. The local diversity of word forms in the nDrapa dialects can be explained using a broader perspective, including Qiangic languages.

The linguistic maps for the three items illustrate three different histories. For 'sand', a non-Qiangic form is found in Central nDrapa, whereas Tibetan loans are found in Northern and Southern nDrapa. I tentatively conclude that the Central nDrapa forms such as *ŋe*^{55}*lɯ*^{33} are relative novelties among Qiangic languages; in other words, the indigenous root for 'sand' was lost in nDrapa. However, it might also be that the Central nDrapa forms reflect a substratum.

For 'leaf', there are three types of word forms: #^{h}pala, which is shared with nearby languages, #ɲaɲa, which is found in Central nDrapa, and the Tibetan loan #loma, which

is distributed in southern, northern, and western areas. The oldest form is possibly #^{h}pala, although it might have been diffused from nearby languages if we consider it to be an areal word form.

Two different types of word form are used to denote 'knee' in nDrapa dialects. Their relative chronology is difficult to ascertain, as both types have cognates in other Qiangic languages. The two roots might originally have had a distinction, and one of them may have been inherited in each dialect.

The discussion in this paper demonstrates the need for caution in conducting comparative linguistic research on the nDrapa language. Some conclusions of the previous study were supported: loan words diffuse from the northern and southern peripheral areas. Alternatively, it remains a question whether the central word forms are older indigenous forms. Further examination is needed to resolve this issue and to clarify the linguistic history of nDrapa.

References

Chirkova, Katia (2012) The Qiangic subgroup from an areal perspective: A case study of Languages of Muli. *Language and Linguistics* 13(1): 133–170. https://www.ling.sinica.edu.tw/item/en?act=journal&code=download&article_id=382

Ding, Picus S. (2014) *A grammar of Prinmi: Based on the central dialect of northwest Yunnan, China*. Leiden/Boston: Brill. (Languages of the Greater Himalayan Region) https://doi.org/10.1163/9789004279773

Evans, Jonathan. P. (2001) *Introduction to Qiang phonology and lexicon: Synchrony and diachrony.* Tokyo: ILCAA, Tokyo University of Foreign Studies.

Gong, Qunhu [龚群虎] (2007) *Zhabayu yanjiu* 《扎巴语研究》[*A study of the nDrapa language*]. Beijing: Minzu Chubanshe.

Lai, Yunfan, Xun Gong, Jesse P. Gates, and Guillaume Jacques (2020) Tangut as a West Gyalrongic language. *Folia Linguistica Historica* 41(1): 171–203. https://doi.org/10.1515/flih-2020-0006

Huang, Bufan [黄布凡] (2007) *Lawurongyu Yanjiu*《拉坞戎语研究》[*A study of Khroskyabs*]. Beijing: Minzu Chubanshe.

Huang, Bufan (ed). [黄布凡主编] (1992) *Zangmianyuzu yuyan cihui* 《藏缅语族语言词汇》[*A Tibeto-Burman lexicon*]. Beijing: The Minzu Press.

Huang, Chenglong [黄成龙] (2007) *Puxi Qiangyu yanjiu*《蒲西羌语研究》[*A study of Puxi Qiang*]. Beijing: Minzu Chubanshe.

Huang, Yang [黄阳] (forthcoming) *Sichuan Yajiang Zhabayu*《四川雅江扎坝语》[*The nDrapa Language in Yajiang, Sichuan*]. Beijing: The Commercial Press.

LaPolla, Randy J. with Chenglong Huang (2003) *A grammar of Qiang*. Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110197273

Lu, Shaozun [陆绍尊] (2001) *Pumiyu fangyan yanjiu*《普米语方言研究》[*A study of Prinmi dialects*]. Beijing: Minzu Chubanshe.

Matisoff, James A. (2009) Stable roots in Sino-Tibetan/Tibeto-Burman. In: Yasuhiko Nagano (ed.) *Issues in Tibeto-Burman historical linguistics*, 291–318. Suita: National Museum of Ethnology. (Senri Ethnological Studies 75) https://doi.org/10.15021/00002570

Matisoff, James A. (2015) *The Sino-Tibetan etymological dictionary and thesaurus*. Berkeley: The Regents of the University of California. Data are also available as online STEDT Database. http://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl

Nagano, Yasuhiko and Marielle Prins (2013) *rGyalrongic languages database.* http://htq.minpaku.ac.jp/databases/rGyalrong/

Shirai, Satoko (2020) A geolinguistic study of directional prefixes in the Qiangic language area. *Himalayan Linguistics* 19.1: 365–392. https://doi.org/10.5070/H919142521

Shirai, Satoko and Yang Huang (forthcoming) A geolinguistic approach to nDrapa dialectology.

Sun, Hongkai [孙宏开] (2001) Lun Zangmianyuzu zhong de Qiangyuzhi yuyan〈论藏缅语族中的羌语支语言〉 [On the languages of the Qiangic branch in Tibeto-Burman]. *Language and Linguistics* 2(1): 157–181. https://www.ling.sinica.edu.tw/item/en?act=journal&code=download&article_id=25

Sun, Hongkai, Dan Xu, Guankun Liu, and Lurongduoding [孙宏开、徐丹、刘光坤、鲁绒多丁] (2014) *Shixingyu yanjiu*《史兴语研究》[A study of the Shihing language]. Beijing: Minzu Chubanshe.

Suzuki, Hiroyuki and Sonam Wangmo (2018) Lhagang Choyu wordlist with the Thamkhas Dialect of Minyag Rabgang Khams (Lhagang, Dartsendo). *Asian and African Languages and Linguistics* 12: 133–160. https://doi.org/10.15026/91144

Tracking basic Ainu vocabulary: Updates and changes of geographical distributions

Mika Fukazawa
(National Ainu Museum)

Abstract: This paper provides the geographical distributions and updates of Hattori and Chiri (1960), as well as the preliminary data. The basic vocabulary of the Ainu dialects collected by Hattori and Chiri (1960) is the most significant vis-à-vis Ainu dialectology. The National Ainu Museum houses the preliminary vocabulary list of Hattori and Chiri (1960). In this study, we investigated the differentiation between the published version of Hattori and Chiri (1960) and the preliminary data. The results of this study show that the dialects on dialectal boundaries often vary between the two sets of data. This not only suggests that elicitation in linguistic fieldwork has occurred but also argues that it is necessary to reconsider the use and treatment of existing materials, such as the data of Hattori and Chiri (1960).
Keywords: Basic vocabulary, Ainu, Elicitation, Dialectal boundary

1. Introduction

This study aims to track the Ainu 200-basic word list in Hattori and Chiri (1960), the most significant work on the Ainu dialects. Before our discussion, we will outline the Ainu language and dialects and their study.

1.1. Ainu language and dialects

The Ainu language is a language isolate and is typologically different from Japanese and other Northeast Asian languages (Bugaeva 2022). The major subgrouping of the Ainu language into the three groups of the Sakhalin, Hokkaido, and northern Kuril dialects is generally accepted in previous studies (Hattori and Chiri 1960, Asai 1974, Tamura 2000, Nakagawa and Fukazawa 2022). The Hokkaido dialect can be grouped into eastern and western dialects. The southern Kuril dialect can be involved in the eastern Hokkaido dialect (Hayashi 1973 [1940]). The dialects in and around Saru and

FUKAZAWA, Mika. 2023. Tracking basic Ainu vocabulary: Updates and changes of geographical distributions. In Cẩm Lan Trịnh, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 34–48. doi: https://doi.org/10.5281/zenodo.8374573

* This work was supported by National Ainu Museum's Research Project, Grant Number 2023C01.

Chitose in the western Hokkaido area often show special patterns in vocabulary, including functional words, which may be similar to those of the Sakhalin dialect. In this study, we refer to them as the dialects of Central Hokkaido. In addition to this, Hattori and Chiri (1960) and Asai (1974) suggested the minor subgrouping of the northernmost (southernmost) dialect of Sōya (Samani) in Hokkaido.

1.2. The study of the Ainu dialects

Hattori, Chiri, and their collaborators conducted crucial research on the Ainu dialects from 1955 to 1956. In the introduction of Hattori and Chiri (1960: 307), Hattori reported that some informants were the last (native) speakers of the Ainu dialect. The speakers were very old. Some could speak the Ainu language fluently, whereas others knew only a few words. Hattori and Chiri investigated the Ainu dialects of Hokkaido and Sakhalin, all in Hokkaido. Because the Soviet Union occupied the southern Sakhalin region at that time, the informants of the Sakhalin dialects lived in Hokkaido as "repatriates." Hattori and Chiri's (1960) study provided the data of the 200-basic word list, and Hattori (1964) edited a dictionary of the Ainu dialects.

Following Hattori and Chiri's (1960) works, Asai (1974) attempted to perform a cluster analysis of the Ainu dialects. To the data of Hattori and Chiri's list, Asai (1974) added the Chitose (Hokkaido) dialect through his fieldwork and the Kuril dialects from written materials (Torii 1903, Murayama 1971, Pinart 1872). He also modified Hattori and Chiri's (1960) data on the Asahikawa, Obihiro, and Kushiro (Hokkaido) dialects. Asai (1974) proposed the "major division" between Hokkaido, Sakhalin, and the Kuril Islands in the cluster analysis and scientifically established the dialectal groups of Ainu. A recent study by Ono and Fukazawa (2022) partly revealed how Asai (1974) treated the data of Hattori and Chiri (1960), although this was not included in his paper.

2. Materials and Methods

In this section, we will overview the materials and methods of this study.

2.1. Materials: The Ainu dialect dataset (Hattori and Chiri 1960)

The data of the Ainu dialects in Hattori and Chiri (1960) consist of the following investigations: In April 1955, Hattori and Chiri investigated the Ainu dialects, and in the summer of 1955, Hattori gathered data on one Hokkaido and three Sakhalin dialects, while his collaborators researched some Hokkaido dialects. In addition, Hattori and his collaborators surveyed the dialects again in the summer of 1956 for a dictionary of the Ainu dialects. In September 1956, Hattori added the investigation of a Sakhalin dialect.

Figure 1 shows the geographical points of the 19 Ainu dialects in Hattori and Chiri's (1960) dataset: 13 dialects in Hokkaido and 6 from Sakhalin. Table 1 presents the information on the investigators.

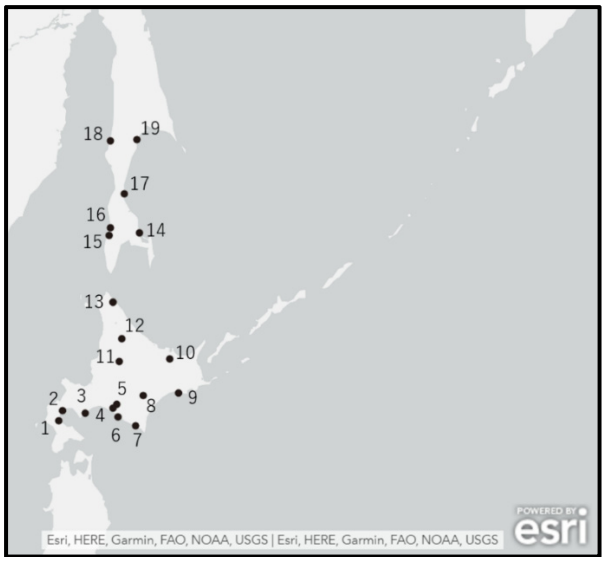


Figure 1: Geographical points of the Ainu dialects in Hattori and Chiri (1960)

Table 1: Investigators of the Ainu dialect and geographical points (Hattori and Chiri 1960)

Point	Dialect	April 1955	Summer 1955	Summer 1956	September 1956
1	Yakumo	Hattori and Chiri (Date: 30)	Kitamura	Kitamura[1]	
2	Oshamambe	Hattori and Chiri (Date: 30)			
3	Horobetsu	Hattori and Chiri[2] (Date: 29)			
4	Biratori, Fukumitsu	Hattori and Chiri (Date: 28)	Fukuda	Fukuda	
5	Nukibetsu		Hattori		
6	Niikappu (Ukegoi)	Hattori and Chiri (Date: 28)			
7	Samani	Hattori and Chiri (Date: 26, 27)			
8	Obihiro	Hattori and Chiri (Date: 24)		Fukuda	
9	Kushiro	Hattori and Chiri (Date: 23)			

[1] Kitamura investigated again the Yakumo dialect in February and March 1957 (Hattori and Chiri 1960).
[2] Later, Chiri investigated the Horobetsu dialect again (Hattori and Chiri 1960).

10	Bihoro	Hattori and Chiri (Date: 22)	Kitamura[3]		
11	Asahikawa	Hattori and Chiri (Date: 19, 20)		Yamamoto	
12	Nayoro	Hattori and Chiri (Date: 19)		Kimura	
13	Sōya	Hattori and Chiri (Date: 13, 14)	Hattori	Mineya	
14	Ochiho	Hattori and Chiri (Date: 16)			
15-1	Tarantomari	Hattori and Chiri (Date: 15-18)			
15-2	Tarantomari	Hattori and Chiri (Date: 15-18)			
16	Maoka		Hattori		
17	Shiraura				Hattori
18	Raichishika		Hattori	Hattori	
19	Nairo		Hattori		
		Preliminary data			

The National Ainu Museum housed the carbon-copied material of the Ainu dialects, same as those used in the investigation of April 1955. Presumably, Chiri wrote and owned this carbon-copied version, and he gave the original one to Hattori. The number of dialects in the carbon-copied material is less than that in Hattori and Chiri (1960), while the two speakers' vocabularies of the Tarantomari dialect are recorded separately: Mr. Nishizaki's vocabulary is 15-1, and Mr. Kawamura's vocabulary is 15-2 as in Table 1. Hattori and Chiri's (1960) paper indicates the data of each speaker together as one Tarantomari dialect. In this study, the carbon-copied material of the National Ainu Museum is referred to as the preliminary material or preliminary data.

Hattori and Chiri (1960) and the preliminary data arranged a list of 200 basic word items from Swadesh word lists. The first 100 items are the same as in Swadesh's later 100-word list, although the second half includes several different words from the remaining 107 words in Swadesh's original 200-word list (Fukazawa 2017, 2018, Nakagawa and Fukazawa 2022). Thereafter, Hattori (1964) edited a dictionary of the Ainu dialects, the data of which were based on the field investigations of 1955–1956 by Hattori, Chiri, and some collaborators, as mentioned above.

2.2. Methods: USED or/and UNDERSTOOD word forms

In the next section (Section 3), we will map each vocabulary item listed in Hattori and Chiri (1960) and the preliminary material and compare their geographic distribution.

[3] Later, Tamura investigated the Bihoro dialect in the winter of 1956–1957 (Hattori and Chiri 1960).

Here, we would like to introduce the distinction between USED and UNDERSTOOD word forms (Tokugawa 1966, Shibata 1969).

- USED word forms: The word forms "USED" for the speaker to express the meaning of the item.
- UNDERSTOOD word forms: The word forms, which are "UNDERSTOOD" but (may) not be used for the speaker to express the meaning of the item.

Shibata (1969) suggests that in practice, investigators cannot obtain UNDERSTOOD word forms in only the first survey because they cannot ask about the same survey item from the perspectives of both the USED and UNDERSTOOD word forms. If they know the likely UNDERSTOOD word form in advance, they will ask for it separately from the USED word form. However, in principle, the first survey should be conducted to determine the geographical distribution of the word form. The second survey should include the UNDERSTOOD word form as an entry of the survey form (Shibata 1969: 42).

According to Tokugawa (1966), the USED and UNDERSTOOD word forms are distinctive as answers to the type of questions. On the one hand, USED word forms appear in response to the WH question: "What is this item called?" — "The word form *X-1*." On the other hand, UNDERSTOOD word forms appear in answer to the Yes-No question: "Is the word form *X-2* called for this item?" — "Yes." Although USED word forms are theoretically part of UNDERSTOOD word forms, let us differentiate them for simplicity. Here, *X-1* is a USED word form and *X-2* is an UNDERSTOOD word form.

Thus, we will suggest that preliminary data of Hattori and Chiri's first survey can identify the USED word forms. We will also interpret the added word forms in Hattori and Chiri (1960) as the UNDERSTOOD word forms against the USED word forms. This article will unify the word phonological notations and accent markers in these materials with phonemic forms between slashes / /.

3. Geographical distributions

In this section, we compare the geographical distribution of the word forms found in the preliminary data and the Hattori and Chiri's (1960) data. In the following subsections, we will show the differences between the two sources at the well-known boundaries of the Ainu dialects since the UNDERSTOOD word forms in one dialect group are often pervasive to/from another.

3.1. Boundary between Hokkaido and Sakhalin dialects

Here, we will observe the northern Hokkaido dialects of Nayoro and Sōya at the boundary between Hokkaido and Sakhalin.

3.1.1. Sakhalin > Hokkaido and Sakhalin

The older form for 'tongue' is considered to be the same form */aw/* as the Sakhalin form (Fukazawa 2017). Figure 2 shows the maps for 'tongue.' In the preliminary data, the Nayoro dialect has the same form as the Sakhalin dialects, */aw/*, but in Hattori and Chiri (1960), the other Hokkaido form, */parunpe/*, was also collected. Table 2 shows the other items in the same Sakhalin > Hokkaido and Sakhalin type as the item of 'tongue.'

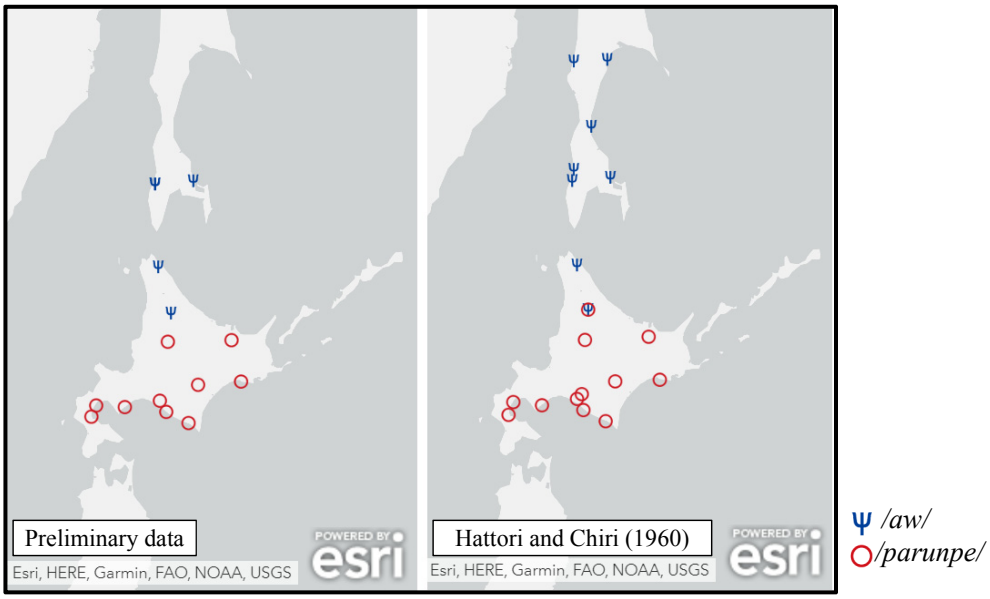


Figure 2: Maps for 'tongue'

Table 2: Sakhalin > Hokkaido and Sakhalin

Word item	Dialect	Preliminary data	Hattori and Chiri (1960)
what (No. 7)	Sōya	*hemáta /hemata/*	*hemáta /hemata/* *nép /nep/*
tongue (No. 44)	Nayoro	*'aw /aw/*	*'áw /aw/* *parúnpe /parunpe/*
smoke (No. 81)	Sōya	*pa /pa/*	*pa /pa/* *sipúya /sipuya/*
ice (No. 148)	Nayoro	*rup /rup/*	*rúp /rup/* *kónru /konru/*
pull (No. 174)	Nayoro	*'ehékem /ehekem/*	*'etáye /etaye/* *'ehékem /ehekem/*
push (No. 175)	Sōya	*'opítuye /opituye/*	*'opítuye /opituye/* *'e'áciw /eaciw/*

3.1.2. Hokkaido > Hokkaido and Sakhalin

The geographical distributions of 'fire' is shown in Figure 3. The Sōya dialect has the same form as the other Hokkaido dialects, /ape/, in the preliminary data, but the Sakhalin form, /unci/, was also collected in Hattori and Chiri (1960). Table 3 shows the other items in the same Hokkaido > Hokkaido and Sakhalin type as the item of 'fire.'

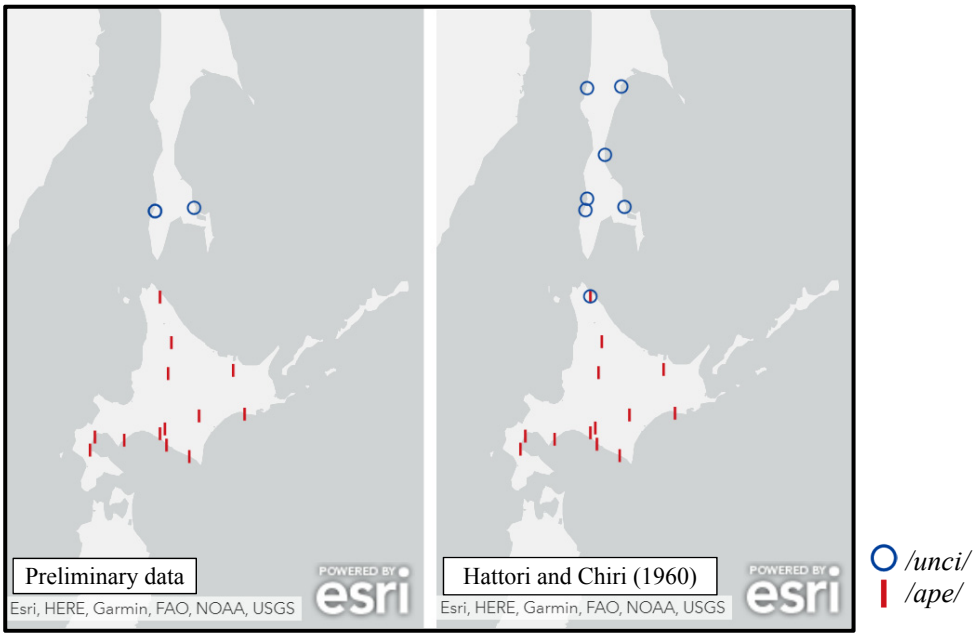


Figure 3: Maps for 'fire'

Table 3: Hokkaido > Hokkaido and Sakhalin

Word item	Dialect	Preliminary data	Hattori and Chiri (1960)
root (No. 26)	Sōya	*sínrit /sinrit/*	*sínrit /sinrit/* *cínkew /cinkew/*
fire (No. 82)	Sōya	*'apé /ape/*	*'apé /ape/* *'únci /unci/*
lip (No. 130)	Nayoro	*pátoy /patoy/*	*pápus /papus/* *cápus /capus/* *pátoy /patoy/*

3.1.3. Sakhalin > (Central) Hokkaido

In the preliminary data of the item of 'head,' the Sōya dialect has the same form as the Sakhalin dialects, */sapa/*, but in Hattori and Chiri (1960), the (Central) Hokkaido form, */pake/*, was only recorded, as shown in Figure 4. In the subsequent dictionary, Hattori (1964) recorded both forms; therefore, Hattori and Chiri (1960) would miss the form of */sapa/*. The item of 'bone' and 'river' in Table 4 shows the same Sakhalin > (Central) Hokkaido type. However, because the later dictionary of Hattori (1964) also collected

only the (Central) Hokkaido form, the Sakhalin form of 'bone' and 'river' in the preliminary data may only be mistaken in the recording (See also Section 4.1).

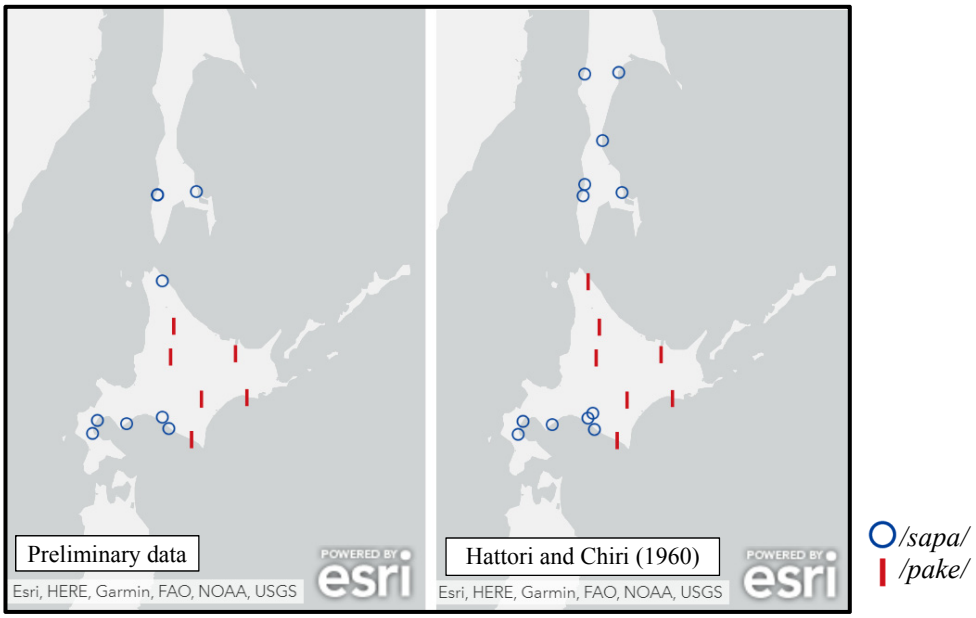


Figure 4: Maps for 'head'

Table 4: Sakhalin > (Central) Hokkaido

Word item	Dialect	Preliminary data	Hattori and Chiri (1960)
bone (No. 31)	Sōya	*poní /poni/*	*poné /pone/*
head (No. 38)	Sōya	*sapá /sapa/*	*paké /pake/*
river (No. 121)	Sōya	*nay /nay/*	*pét /pet/*

3.2. Boundary between the Southwestern and Central dialects of Hokkaido

Here, we will observe the southwestern Hokkaido dialects of Yakumo, Oshamambe, and Horobetsu at the boundary between the southwestern and central dialects of Hokkaido.

3.2.1. Central Hokkaido > Central and Southwestern Hokkaido

In the preliminary data of the item of 'fish,' the Horobetsu dialect has the same form as the Central Hokkaido dialects, */cep/*. In Hattori and Chiri (1960), the southwestern Hokkaido forms of Yakumo and Oshamambe, */ciep/*[4], was also collected in Horobetsu, as shown in Figure 5. The item of 'all' in Table 5 shows the same type as the item of 'fish.'

[4] The form of */ciep/* and */cep/* for 'fish' can be analyzed as *c(i)-e-p* 1PL.A-eat-thing 'Lit. the thing which we eat.'

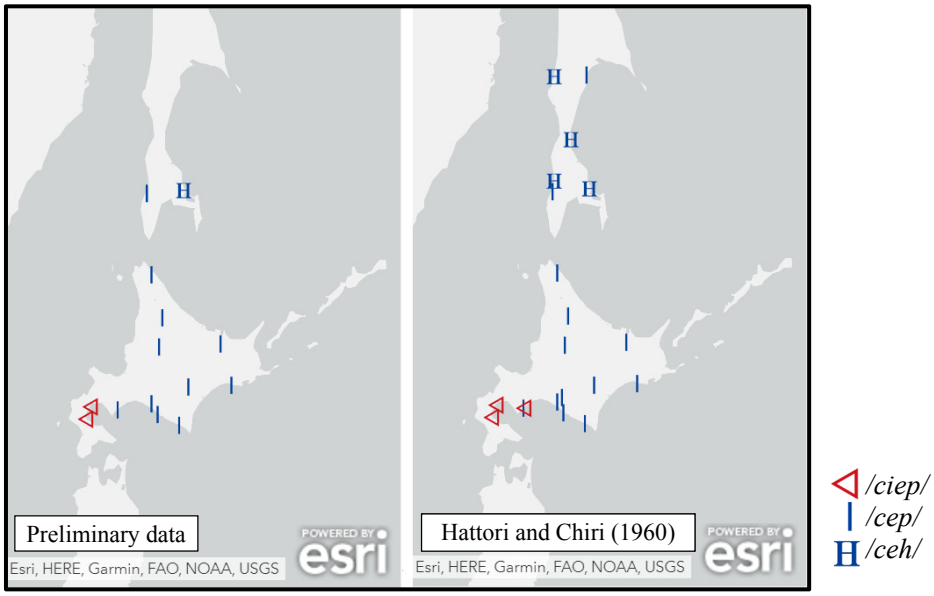


Figure 5: Maps for 'fish'

Table 5: Central Hokkaido > Central and Southwestern Hokkaido

Word item	Dialect	Preliminary data	Hattori and Chiri (1960)
all (No. 9)	Yakumo	*'opítta /opitta/*	*'opítta /opitta/* *'epítta /epitta/*
fish (No. 19)	Horobetsu	*cep /cep/*	*cép /cep/* *ci'ép /ciep/*

3.2.2. Southwestern Hokkaido > Central Hokkaido

In the preliminary data, the Yakumo and Oshamambe dialects have the original form for 'three,' */reppis/*, as a form of southwestern dialects. However, in Hattori and Chiri (1960), the same form as the other Hokkaido form, */rep/*, was also collected in these dialects, as shown in Figure 6. This type does not exist for the other items, as shown in Table 6.

Table 6: Western Hokkaido > Central Hokkaido

Word item	Dialect	Preliminary data	Hattori and Chiri (1960)
three (No. 110)	Yakumo, Oshamanbe	*réppis* /reppis/	*rép* /rep/

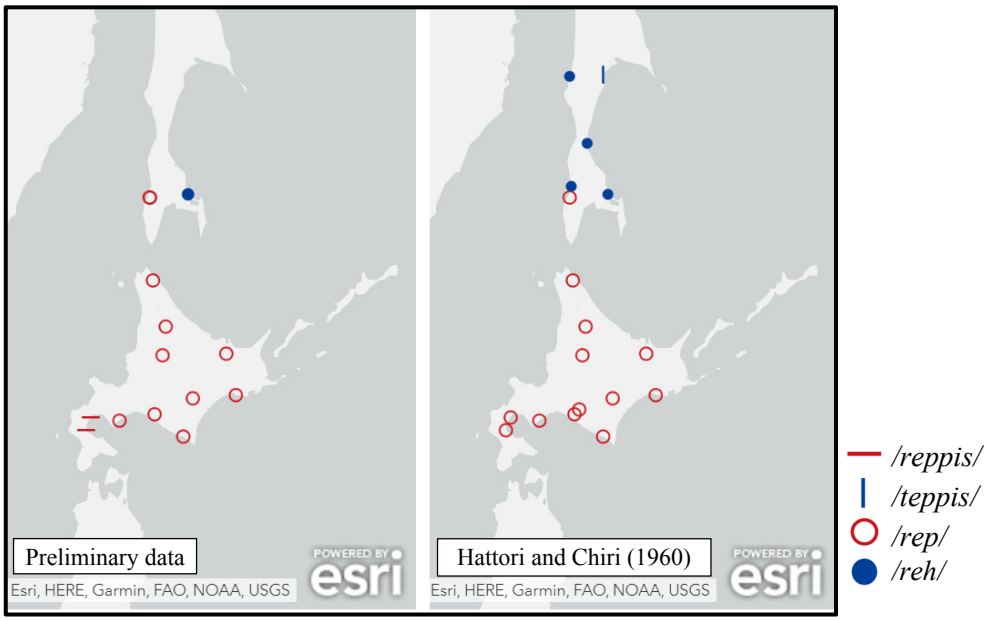


Figure 6: Maps for 'three'

3.3. Contact pattern: Tarantomari < Tarantomari and Maoka

Here, we will observe another contact pattern of geographical distributions between the Tarantomari and Maoka dialects. The preliminary data show the word form for 'freeze' */rupkoro/* in Tarantomari, while Hattori and Chiri (1960) also show the form */rupus/*, which is the same as the Maoka dialect (see Figure 7 and Table 7). Table 7 shows other similar patterns as the item of 'freeze.' Note that the preliminary data have separate lists for the two Tarantomari speakers.

Table 7: Tarantomari < Tarantomari and Maoka

Word item	Dialect[5]	Preliminary data	Hattori and Chiri (1960)
all (No. 9)	Tarantomari (N)	*'anpáhno* /anpahno/	*'anpahno* /anpahno/ *'imiki* /imiki/
	Tarantomari (K)	*'ánpahno* /anpahno/	
many (No. 10)	Tarantomari (N)	*'okáyno* /okayno/	*'okayno* /okayno/ *renkayne* /renkayne/ *poronno* /poronno/
	Tarantomari (K)	*'okáy* /okay/	
freeze (No. 150)	Tarantomari (N)	*rúpkoró* /rupkoro/	*rupkoro* /rupkoro/ *rupus* /rupus/
	Tarantomari (K)	*rúpkoró* /rupkoro/	

[5] The speaker of Tarantomari (N) is Mr. Nishizaki and Tarantomari (K) is Mr. Kawamura.

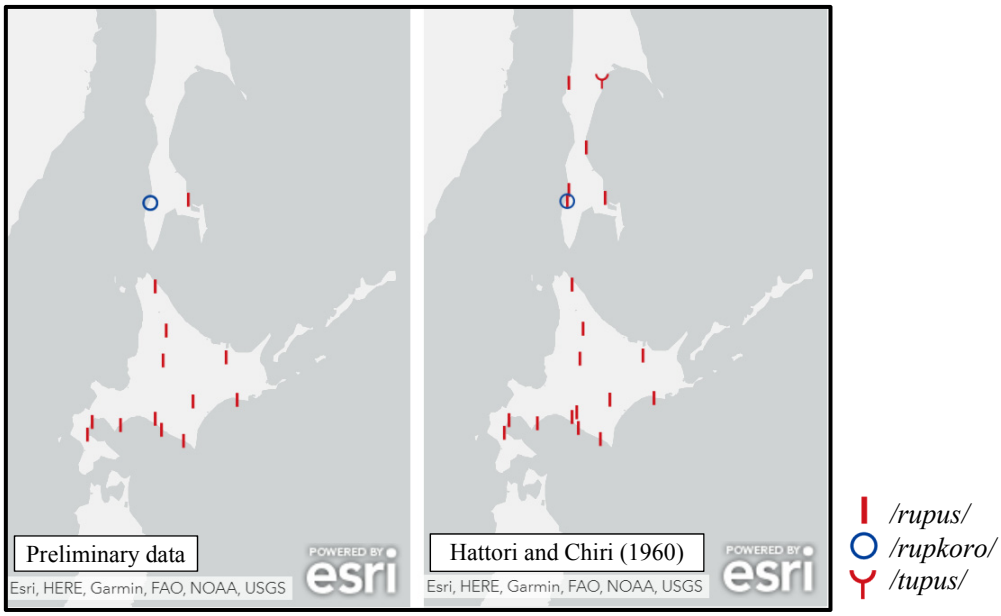


Figure 7: Maps for 'freeze'

4. Discussion

This section will examine the patterns of differences between the preliminary data and Hattori and Chiri (1960), in addition to how these data can contribute to the historical interpretations of the geographical distribution.

4.1. Updates and changes from the preliminary data

The updated and changed patterns from the preliminary data to Hattori and Chiri (1960) are as follows:

a) Improvement from confusion between transitive and intransitive verbs: 'say (No. 71)' and 'fear (No. 191)'
b) Improvement from affixation form: 'hear (No. 58)' and 'all (No. 9)'
c) Removal of unnecessary verbs: 'down (No. 194)'
d) Unification of linguistic plurality: 'fly (No. 64)'
e) Corrections of phonological notations: 'dirty (No. 164)'
f) Other

However, this situation is more complicated for items such as body parts. For example, in the vocabulary item of 'foot (No. 46),' each dialect refers to different parts

of feet and legs, and Hattori and Chiri (1960) also could not correct this confusion until an Ainu dialect dictionary was supplied in Hattori (1964). In addition, when the word forms of the preliminary data were lost and changed in Hattori and Chiri (1960), as shown in Table 8, Hattori (1964) may later have provided the key to treating them. If Hattori's (1964) data match the preliminary data, Hattori and Chiri (1960) would forget to collect the word form in the preliminary data. If Hattori's (1964) data match Hattori and Chiri's (1960) data, we would consider the preliminary data to be mistaken, or Hattori (1964) and Hattori and Chiri (1960) forgot to collect the word form in the preliminary data.

Table 8: The word form /sapa/ lost in Hattori and Chiri (1960)

Word item	Dialect	Preliminary data	Hattori and Chiri (1960)	Hattori (1964)
head (No. 38)	Sōya	*sapá /sapa/*	*paké /pake/*	*sapá /sapa/* *paké /pake/*

4.2. Analysis of elicitation in fieldwork

In this section, we will consider the issue of elicitation in fieldwork as an answer to the Yes-No question. The preliminary material tends to record USED word data, while Hattori and Chiri (1960) tend to record UNDERSTOOD word data. According to Tokugawa (1966), UNDERSTOOD words may be newer or older than USED words. If an UNDERSTOOD word is in a cultural center, it may be about to expand into a new region.

Hattori (1964) wrote "*parúnpe; 'awéhe*[6]" for the item of 'tongue' in the Soya dialect. The description indicates that *'parúnpe /parunpe/* is the UNDERSTOOD word in Nayoro (see also Figure 8). Soya and Nayoro, the boundary dialects between Hokkaido and Sakhalin, have been recognized as part of the Hokkaido dialect. However, when these dialects selected the same word form as the Sakhalin dialect in the first survey of the preliminary data, it suggested that the dialect had more Sakhalin-like dialectal features than recognized in current Ainu dialectal studies. In addition, Hattori (1964) collects */rehpis/* in the Raichishika dialect, which is of the same type as */reppis/* and */teppis/* in the preliminary data and Hattori and Chiri (1960). This */reppis/* type may be older than */rep/* and */reh/* because it is distributed in the peripheral area (See Figures 9 and 10). Figure 10 shows a superimposed map of Figure 9 for 'three.'

[6] The form of *awéhe* is a possessive form of *áw*.

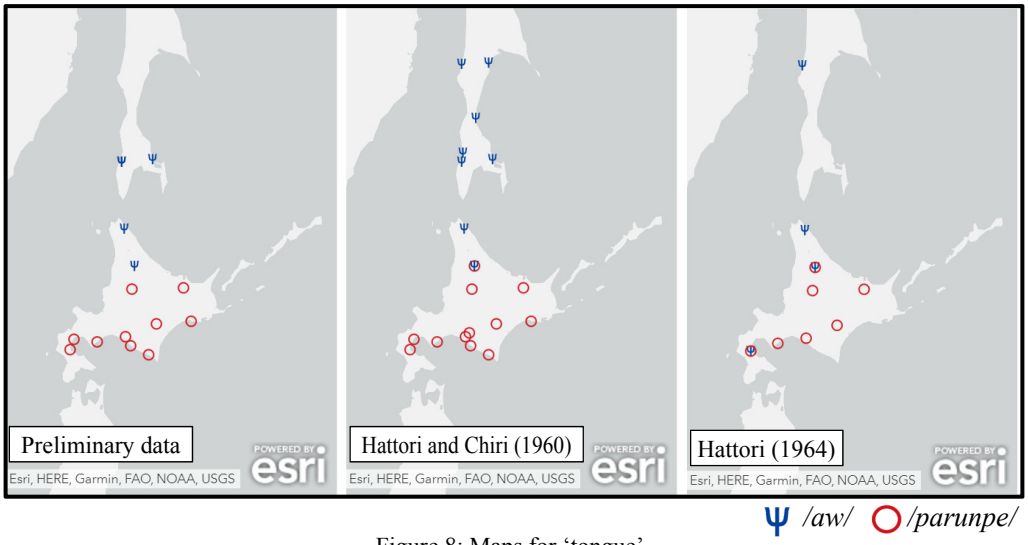


Figure 8: Maps for 'tongue'

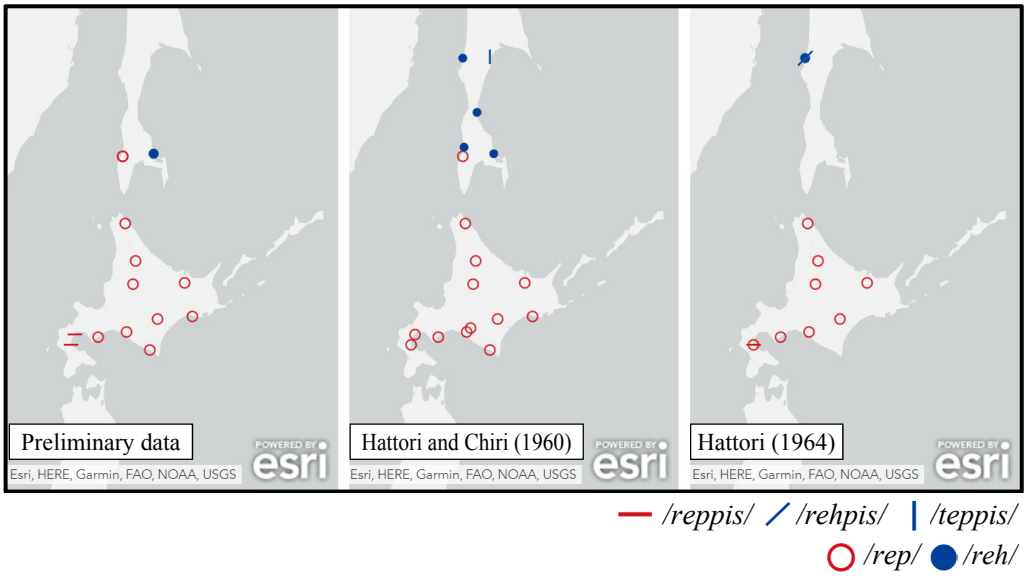


Figure 9: Maps for 'three'

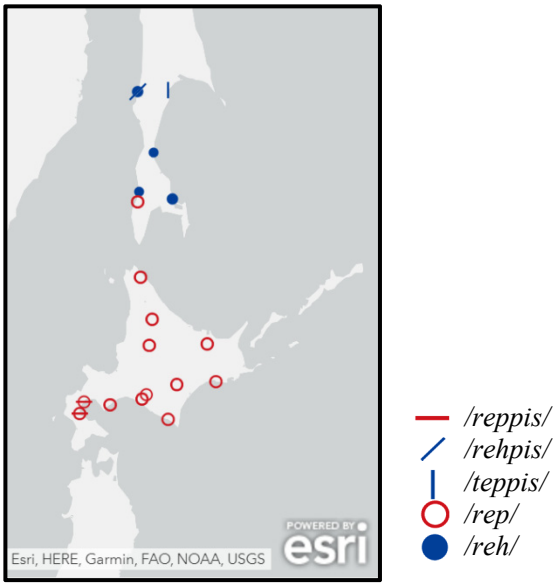


Figure 10: Superimposed map of Figure 9 for 'three'

5. Concluding remarks

This study suggests that the elicitation data from the second survey should be considered as UNDERSTOOD word forms and distinguished from the USED word forms in the first survey. The USED and UNDERSTOOD word forms can create a historical layer in the vocabulary data of every survey. We can then further dynamically analyze the geographical distribution of each item using historical layers. This idea can be adopted not only for the Ainu language but also for other languages. A historical discussion of individual vocabulary items will be left for future research.

References

Asai, Toru (1974) Classification of dialects: Cluster analysis of Ainu dialects. *Bulletin of the Institute for the Study of North Eurasian Cultures, Hokkaido University* 8: 45–136.

Bugaeva, Anna (2022) Introduction. In: Anna Bugaeva (ed.) *Handbook of the Ainu language,* 1–19. Berlin: Mouton De Gruyter. doi: https://doi.org/10.1515/9781501502859-001

Chiri, Mashiho [知里真志保] (1955–1960) Preliminary Material of the Ainu 200-basic word list (Manuscript). Shiraoi: National Ainu Museum.

Fukazawa, Mika [深澤美香] (2017) *Kagake monjo ni okeru Ainugo no kenkyū*『加賀家文書におけるアイヌ語の研究』. Chiba: Chiba University. URI: https://opac.ll.chiba-u.jp/da/curator/103627/

Fukazawa, Mika (2018) Geographical distribution and patterns of basic Ainu vocabulary in Hattori and Chiri (1960). In: Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Papers from the fourth international conference on Asian geolinguistics*, 76–88. Jakarta: Universitas Indonesia. Online: https://publication.aa-ken.jp/papers_4IC_Asian_geolinguistics_2018.pdf

Hattori, Shirō [服部四郎] (ed.) (1964) *Ainugo hōgen jiten*『アイヌ語方言辞典』[Ainu dialect dictionary]. Tokyo: Iwanami Shoten.

Hattori, Shirō [服部四郎] and Mashiho Chiri [知里真志保] (1960) Ainugo shohōgen no kisogoi tōkeigakuteki kenkyū「アイヌ語諸方言の基礎語彙統計学的研究」[A lexicostatistic study on the Ainu dialects]. *Japanese Journal of Ethnology* 24(4): 307–342. doi: https://doi.org/10.14890/minkennewseries.24.4_307

Hayashi, Kingo [林欽吾] (1973 [1940]) Shikotan-tō no Ainu-zoku「色丹島のアイヌ語族」[The Ainu people on Shikotan Island]. In: Nihon Jōmin Bunka Kenkyūjo [日本常民文化研究所] (ed.) *Nihon jōmin seikatsu shiryō sōsho*『日本常民生活資料叢書』[Series of materials on the lives of the people of Japan] 7, 552–505. Tokyo: Sanichi Shobō.

Murayama, Shichirō [村山七郎] (1971) *Kitachishima Ainugo*『北千島アイヌ語』[Northern Kuril Ainu]. Tokyo: Yoshikawa Kobunkan.

Nakagawa, Hiroshi and Mika Fukazawa (2022) Hokkaido Ainu dialects: Towards a classification of Ainu dialects. In: Anna Bugaeva (ed.) *Handbook of the Ainu language,* 253–328. Berlin: Mouton De Gruyter. doi: https://doi.org/10.1515/9781501502859-009

Ono, Yōhei and Mika Fukazawa (2022) Ainugo shohōgen no gokei no ruiji ni kansuru kiso dēta no fukugen「アイヌ語諸方言の語形の類似に関する基礎データの復元：論文に書きれなかった研究者の判断・思考に迫る」[Reconstruction of Original Data on Similarity between Word Forms in Ainu Dialects: Approaching the Researcher's Unwritten Judgment and Thoughts]. *Hoppō gengo kenkyū*『北方言語研究』[Northern Language Studies] 13, 213-246.

Pinart Alphonse Louis (1872) Vocabulary in Russian and Ainu dialect of the Paramushir and Limushir Islands in the Kuriles. Transcribed and Translated: Asai, Toru (1974) Classification of dialects: Cluster analysis of Ainu dialects. *Bulletin of the Institute for the Study of North Eurasian Cultures, Hokkaido University* 8: 101–136.

Shibata, Takeshi [柴田武] (1969) gengochirigaku no hōhō『言語地理学の方法』[Methods of Geolinguistics]. Tokyo: Chikuma Shobo.

Tamura, Suzuko (2000) *The Ainu language*. Tokyo: Sanseido.

Tokugawa, Munemasa [徳川宗賢] (1966) Rikai no Tōgosen [理解の等語線] Dialectal areas of passive knowledge and of active usage. *Kokugogaku*『国語学』[Studies in the Japanese Language] 64: 11–21.

Torii, Ryūzō [鳥居龍藏] (1903) *Chishima Ainu*『千島アイヌ』[Kuril Ainu]. Tokyo: Yoshikawa Kobunkan.

Linguistic Map of "Breast" in Zhuang and Its Interpretations

Yan Zhuo
(Nankai University)

Abstract: The term "breast" is one of the basic words in a language. It has numerous lexical forms in Zhuang dialects, which can be divided into 13 groups. The lexical forms of the term "breast" can also be divided into five types, including n-, m-, p-, tɕ-, and ʔ-, based on initials. Geographical distribution and etymological research showed that different forms emerged because of influences such as internal innovation, borrowing, and tonal derivation. Additionally, the word "breast" is intimate to women and is inappropriate for direct address. Therefore, people frequently use borrowed or invented forms to avoid this taboo, which contributes to the overlap of many forms.
Keywords: Zhuang language, "breast", geographical distribution, diachrony

1. Introduction

1.1. The motivation for the research question

The word "breast" is one of the core words for body parts, appearing at position 51 on Swadesh's list of 100 words. Generally, it is believed that terms for body parts are among the most stable core words. Monosyllabic forms of "breast" are used in the Zhuang language, however, there are complicated lexical differences within dialects and unclear correspondences within Tai-Kadai.

Why do Zhuang dialects have many complicated expressions for a single monosyllabic core word? How can its category distributions and their hierarchical connections be explained? This is the question that underlies this work.

YAN, Zhuo. 2023. Linguistic Map of "Breast" in Zhuang and Its Interpretations. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 49–61. doi: https://doi.org/10.5281/zenodo.8374575

*This research was funded by the National Social Science Foundation Project of China (22CYY048) and the fellowship of China Postdoctoral Science Foundation (2022M711714). This paper is rewritten from a section of the author's PhD thesis and presented at the fifth International Conference on Asian Geolinguistics, thanks to the guidance of Prof. Zeng Xiaoyu and valuable comments from Prof. Mitsuaki Endo.

1.2. Literature review

Wu (2017:752) reported that synonyms like "chest" "nipple" and "milk" have etymological ties to the term "breast" in Eurasian languages. In the Zhuang language, "breast" and "milk" have the same forms in most dialects, however in certain dialects, "milk" is expressed as "water + breast". Endo (2016) discussed the types of "milk" in Tai-Kadai and analyzed the chronological sequence of each category based on their distribution characteristics, including different synonyms of "breast" in the Zhuang language. However, due to the limited length of the article, the discussion cannot be fully explored, and there is still an opportunity for further investigation by combining the geographical distribution with etymological information and performing a comparative analysis.

Therefore, this article aimed to create a linguistic map of the word "breast" in Zhuang dialects by examining its distribution patterns, investigating the origin of various synonyms, and drawing a conclusion regarding their diachronic sequence.

1.3. Source of data

Most of the information regarding the word "breast" in the Zhuang language in the article came from a study of 61 Zhuang dialects conducted in 2019 and 2020. The works *A Study on Zhuang Dialects* (1999) and *Vocabulary of Minority Languages in Guangxi* (2008) were also used as additional sources. To give a comprehensive description of the word "breast" in the Zhuang language, different expressions for the same dialect, different pronunciations in different towns for the same site, and results different from findings of previous research are presented on the map.

2. Lexical forms of "breast" in the Zhuang language and their geographical distribution

The synchronic distribution of the lexical forms for "breast" in the Zhuang language is shown in Figure 1[1]:

[1] The language mapping software developed by Nankai University and Tianjin Xinhui Network Technology Service Center is used in all the language maps listed in this paper; the base map is based on Gaode map (©2023AutoNavi—GS (2022)1061).

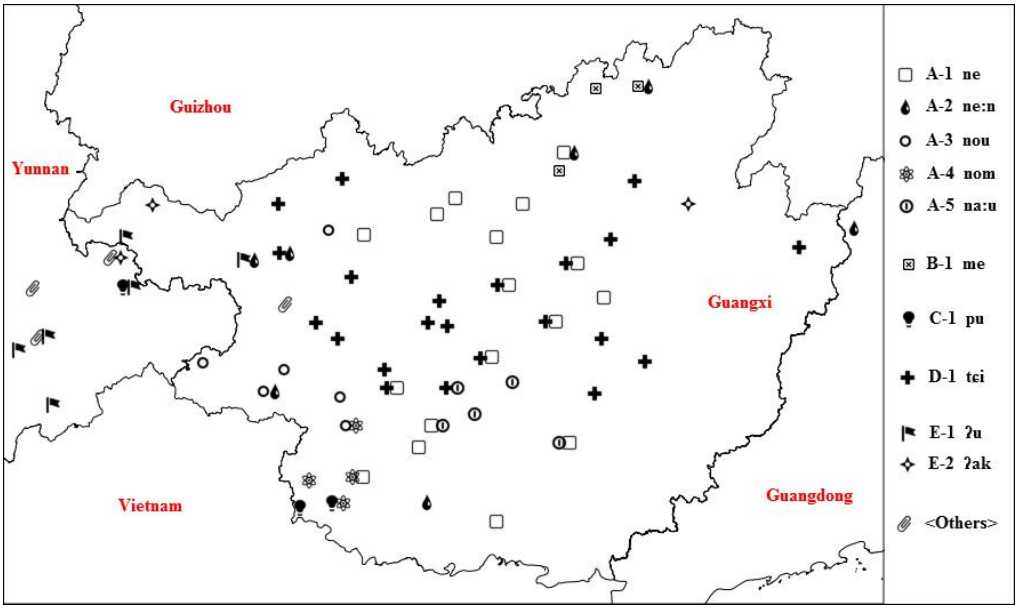


Figure 1: Geographical distribution of "breast" in Zhuang dialect.

From Figure 1, it is evident that the lexical forms for "breast" in modern Zhuang dialects are complex. The word for "breast" in various Zhuang dialects is monosyllabic, and different lexical forms reflect different origins. Based on the initials, they could be classified into five categories (Table 1).

Table 1: Lexical forms of "breast" in the Zhuang language.

Type	Legend	Variation	Number of types and the proportion
Type A *n-*	A-1 *ne*	ne$^{1/3/4/5}$; nɛ$^{3/5}$; ni^{1}	17, 21.3%
	A-2 *ne:n*	ne:n^{5}; nən^{5}; nɛ:n^{3}; ni:n^{3}	7, 8.7%
	A-3 *nou*	nou^{4}; nu$^{4/5}$	6, 7.5%
	A-4 *nom*	nom^{2}; num^{2}	4, 7.5%
	A-5 *na:u*	na:u^{5}; ne:u$^{1/5}$	5, 6.2%
Type B *m-*	B-1 *me*	me^{5}; mɛ^{5}; mei^{1}	3, 3.8%
	<Others>	mɒ^{1}	1, 1.2%
Type C *p-*	C-1 *pu*	pu$^{1/3/5}$	3, 3.8%
	<Others>	pe^{6}	1，1.2%
Type D *tɕ-*	D-1 *tɕi*	tɕi$^{3/4}$; ɕi^{3}; tɕei^{3}; tɕe^{4}; ɕe^{4}	22, 27.5%
Type E *ʔ-*	E-1 *ʔu*	ʔu^{3}	6, 7.5%
	E-2 *ʔak*	ʔak^{7}; ʔa^{5}	3, 3.8%
	<Others>	ʔɛm^{2}	2, 2.5%

The division into the five aforementioned types is based on phonetic forms, hence the links between the subcategories within each category are not particularly close. The relationships among these forms and surrounding languages will be analyzed in the following text to explain their origins, determine the nature of different lexical forms, and make assumptions about the ancient forms of the Zhuang language.

2.1. Type A: *n-*

2.1.1 A-1 *ne*

There are mainly five forms of *n-* type, and the A-1 *ne* type distribution is relatively widespread, found in both the southern and northern dialects of the Zhuang language. However, there are differences in tones among different dialects. Similar nɛ^{6} form can also be found in the Mulam language. Liang and Zhang (1996: 288, 595) reconstructed the Proto Kam-Tai form *mlɛ based on ne^{5} in the Liujiang Zhuang dialect, mɛ^{3} in Bouyei, nɛ^{6} in Mulam, and me^{1} in the northern Kam dialect. Endo (2016) reported that the *ne* type might be caused by the drop of the final consonant of A-2 *neːn* type. However, within the Zhuang language itself, there are as many as four tone categories for the *ne* type. If it is a native word, there should not be as many tone categories associated with it.

According to Cao (2008: 76), a considerable number of Chinese dialects in Guangxi refer to the breast as *"nai" (奶)*. Yu (2016:212) reported that in the Ping dialect of northern Guangxi, the term for "breast" is pronounced as [ni5][ne3], which is very similar to the pronunciation in the Zhuang language. The Chinese dialects commonly use *"nai" (奶)* or *"naizi" (奶子)* to refer to the breast. In the *Guangyun* dictionary, *"nai"* has two entries: "*milk, nu xie qie" (乳也，奴蟹切)*and "*Chu people call mother, nu li qie*" *(楚人呼母，奴礼切)*. The pronunciation "*nu li qie*" (*奴礼切*), derived from "mother", can explain various pronunciations such as ni^{5}/ne^{3} in Guangxi Chinese dialects. Due to the various dialect locations and times after borrowing into Zhuang language, different tones and vowels were produced.

Therefore, we speculate that A-1 *ne* might not be an inherent form but rather a newly borrowed word.

2.1.2 A-2 *neːn*

The distribution of A-2 *neːn* is relatively scattered and often co-occurs with other lexical forms. Jin (2012:193) reported that *neːn* is a variant form of *ne*, which is the result of merging the overlapping forms. Endo (2016) reported that *neːn* type is located in the most peripheral places in isolated districts and it should be the oldest form in this

area. It is worth noting that *neːn* type is similar to the Sinitic word *nin* which is dominant in Min, Cantonese, and Hakka dialects. They probably share a common origin. Since there is no Chinese character for *nin* in Sinitic, the origin may be Tai-Kadai. Liu (1998), Tang (2009), and Takashi (2016) also reported that the *nin* and *nɛn* in Southern Chinese dialects are borrowed from Tai-Kadai. We agree with Prof. Endo's viewpoint on the relationship between the Zhuang language's *neːn* and the Min, Cantonese, and Hakka dialects of Chinese. However, further research is needed to investigate whether they share the same origin, the direction of borrowing, and their strata in Tai-Kadai.

The Vocabulary of Chinese Dialects (1995:254) records *nɛn*, *nin*, and *nĩ* as phonetic characters for the word *"nai"(奶)*, which are also written as "孻" and "嫩" and are recorded as "腩" in the Min dialect. However, these colloquial characters are not included in *Guangyun*, *Jiyun*, or *Chinese Dictionary*. It can be assumed that these characters were created by locals. Therefore, we put forward a question, Do these characters (孻、嫩、腩) used in dialects represent underlying words of the Tai-Kadai? If so, why do they occur sporadically within the Tai-Kadai proper but have a very widespread distribution in southern Chinese dialects? Chinese dialects were compared with the Zhuang language *neːn*, Thai language *nom*, Zhuang language *naːu*, and Be language *no?* by Liu (1998) and Tang (2009), respectively. However, it is challenging to pinpoint the path of borrowing because *neːn* is only sporadically spread across the Zhuang language in Guangxi. The forms *nɛn*, *nin*, *nĩ* in Chinese dialects and *nom*, *naːu* in the Zhuang language, apart from having nasal initials in common, it difficult to prove a genetic relationship.

Xie (2003:2-3, 101-108) reported that *nɛn* in Hakka is not directly related to the pronunciation of "milk" in southern minority dialects. He thought it was an er-final formed by merging *nɛ* and *-n*, which developed from *"nu li qie"(奴礼切),* which can be represented as *nɛ+n→ nɛn*. In the Min, Gan, Hui, and Cantonese dialects, the pronunciation of "milk" is also derived from er-final of the character *"奶" (nai)*. In this study, we agree with Xie's argument, which postulates that the word *"奶" (nai)* in southern Chinese dialects has undergone a widespread fusion of er-final.

The Zhuang's *neːn* is distributed in Rong'an, Shangsi, and Lianshan in Guangdong Province, where the Hakka or Cantonese dialects are used. The pronunciation forms of nɛːn and niːn are also used in Lingle, Tianlin, and Jingxi, but it is noteworthy that these three places surround Baise, where Cantonese has a considerable influence. Therefore, based on the geographical distribution, we found that the distribution of *neːn* and its surroundings often overlap with the distribution of the Cantonese, Hakka, and other

Chinese dialects. Chinese dialects may have influenced the *neːn* in Zhuang, therefore borrowing should be from Chinese to Zhuang.

Furthermore, in the map provided by Endo (2016), there is also a distribution of *neːn* in Hlai dialect of Hainan, but only one point. In the Hainan region, there is also a distribution of the Min and Cantonese dialects, thus it cannot be ruled out that the Tai-Kadai form *neːn* in the border areas of Hainan is influenced by Chinese dialects. If the borrowing direction of *neːn* is indeed from Chinese to Zhuang, then *neːn* is not the oldest form in Tai-Kadai. Let us continue exploring the early forms of the Tai-Kadai language.

2.1.3 A-3 *nou* and A-5 *naːu*

The A-3 *nou* and A-5 *naːu* are two different forms. The *Zhuang-Han Vocabulary* also recorded them as *noux* and *nauq* in different Zhuang scripts. The nou^4 is relatively concentrated in the Dejing dialect, while naːu^5 is only distributed in the border area between the northern and southern dialects of the Zhuang language.

Xing (1999: 457) used the Chinese character *"㱩"* to correspond to the Tai ʔu^3. The term *"㱩"* in *Shuowen Jiezi* means *"gou, milk"(㱩,乳也)*. It is pronounced as *"gu hou qie"(古候切)*. In *Yupian*, it is divided into two pronunciations, *"nu dou qie"(奴豆切)* and *"gong dou qie"(公豆切)*. Prof. Xing pointed out that the correspondence between Chinese *"㱩"* and Tai is prescient, but the specific form corresponding to Tai is open to discussion. Meng (2010: 44) pointed out that in the Wuming Zhuang dialect, the term for "breast" is naːu^5, and he believed that naːu^5 can correspond to the pronunciation of "nou"*(㱩)* and they have a similar meaning. While, according to the annotations in *Guangyun*, the pronunciation of *"㱩"* is divided into two entries *"gu hou qie"(古候切)* and *"nai hou qie"(乃后切)*, which both mean "milk". The form *"nai hou"* is pronounced very similar to "nou^4/nu^4" in Zhuang, the initials, finals, and tones are all in a corresponding relationship. Therefore, it is speculated that Zhuang's nou^4/nu^4 may be borrowed from the Chinese character *"㱩"*.

The variations of A-5 *naːu* include neːu^5/ ŋ̍eːu^1. Endo (2016) suggested that *neːu* may be the result of A-1 *ne* with an added sound -u. Cao (2008: 76) marked it as *nɛːu* on the Chinese dialects map, and it only appears within the Guangxi region without a corresponding character. The form naːu^5 is only found near Guangxi Yongning with a narrow range. It has a primary variant of *naːu*, which cannot be explained by paragoge. Therefore, we assume that it is an innovation that emerged at the line separating dialects.

2.1.4 A-4 *nom*

The A-4 *nom* is mainly found in the Zuojiang dialect and it is an important form in the Southwest and Central branches of Tai. In Saek, this form is used to refer to "milk".

nom^{2} Thai nom^{2} Lao num^{2} Dai nom^{2} Shan nom^{2} Nung nom^{2} $_{(milk)}$ Saek

Zhengzhang (2014) reported that Chinese [njo] *"乳"* is related to Thai /nom/, Saek /nɔm/, Dai /-num/ for "breast" or "milk". He also compared it to Burmese /nou/ and Tibetan /nu-ma/ for "milk", "breast" or "nipple". However, the formal comparison is not strict enough, and the similarities between these words are only limited to having nasal initials. Additional proof of phonetic correlation should be offered if there is a comparable relationship.

Liang and Zhang (1996: 322) pointed out that in Yongning, "breast" is called ne:u^{1}, but "fig" is called nəm^{2} tsi^{2}, which means "the breast of a cow", indicating that nəm^{2} was used to mean "breast" in Yongning, which can be represented as nəm^{2} → ne:u^{1}.

Based on the relationship between the Central and Southwest branches of Tai, It is assumed that the development of *nom* should have been an early invention between two different linguistic branches. Therefore, it is speculated that its stratum is relatively early.

2.2. Type B: *m-*

There are only two forms of *m-* type: *me* and *mɒ*.

Yu (2016: 225) proposed two main origins of "breast/milk" words in the Chinese dialects in Guangxi province: "mother" and "grandmother". The Binyang dialect uses the same word, *弥* [mɛ^{1}] [mi^{1}], to express both "breast" and "milk", in the Cantonese of Nanning, the word is *弥弥* [mɛ^{2} mɛ^{3}], while in Cantonese of Wuzhou, the word is *nai(奶)*. Wang (2018: 218) pointed out that the "*nai*"*(奶)* and the "*ma"(妈)* are two important forms of the breast in the Chinese dialects, and they are respectively associated with the taboo nature of "*nai*" and the fact that breast is a characteristic of mothers. There are two possible origins from which B-1 *me*^{5} in Zhuang derived: first, by borrowing from neighboring Chinese dialects; second, by tonal change of the word [me^{6}] , which means "mother" in Zhuang language. Further evidence from Bouyei: [me^{4}] "breast" may come from the tonal change of the word [me^{6}] "mother". Although different languages and dialects employ the same tonal modification approach to create new words, the kind of tones that are altered may differ. The second source is more reliable in our opinion.

The <Others> [mɒ^{1}] in Yanshan Zhuang is thought to be borrowed from the Chinese word *(妈),* which means "mother".

2.3. Type C: *p-*

There are also two forms of *p-* type, *pu*, and *pe*.

The C-1 *pu* type forms have the same consonant and vowel but different tones, indicating a consistent origin. Wang (2018:218) proposed that the reduplicated form *nainai(奶奶)* in Chinese dialects is an important term for "breast" and is widely distributed. However, the real pronunciation often differs from the term *nainai(奶奶)* which means "grandmother". In Zhuang dialects, pu^{2} is used to mean "grandmother", and we speculate that $pu^{1/3/5}$ may be a tonal change derivation of pu^{2}. Additionally, in the Bouyei language, both "breast" and "maternal grandmother" are pronounced as po^{4}, suggesting a similar strategy of word formation. Considering Chinese and Tai languages, we suspect that there might be some interlinguistic similarities in the sound variation of "grandmother/maternal grandmother" indicating "breast". The "breast" is an important organ for female offspring nurturing, and the B-1 *me* mentioned above and the C-1 *pu* can be interpreted as tonal change derivation of kinship terms for female close relatives.

According to Cao (2008:76), in many locations in western Guangxi, the term "breast" is represented as $pɛ^{35}$, but there is no Chinese character for it. Xie (2007:1473) documented that in the Baise Nabi Ping and the Fusui Longtou Ping dialects, the pronunciation of "breast" is $pɛ^{55/45}$, also without a native character. The striking similarity in the Ping dialects across different locations, combined with the fact that only Qiubei Zhuang has the form *pe*^{6}, leads us to speculate that <Others> *pe*^{6} in Qiubei Zhuang is borrowed from the Ping dialect.

2.4. Type D: *tɕ-*

There are numerous *tɕ-* type word variants, which are particularly common in the northern dialects of the Zhuang language, also *tɕ-* type accounts for the largest percentage among 5 types.

There are many cognates of [$tɕi^{3}$] in other Tai-Kadai languages:

$ɕi^{3}$ Zhuang$_{Du'an}$ $tɕi^{3}$ Zhuang$_{Tiandong}$ $tɕiu^{3}$ Zhuang$_{Guangnan}$ $tʃɛ^{4}$ Zhuang$_{Leye}$ $ʃe^{4}$ Zhuang$_{Lingyun}$ $tɕei^{3}$ Zhuang$_{Hezhou}$ $ɕu^{4}$ Jin cu^{3} Saek tju^{4} Sui $sɛ^{1}$;$tsɛ^{5}$ Maonan tsi^{5} Hlai$_{Tongshi}$ $tsei^{1}$ Hlai$_{Baoding}$ tsu^{4} Then co^{33} Lachi $tɕəu^{35}$ Gelao $tsɯ^{33}$ Paha

Jin (2012:193) suggested [tsi^{4}] to be the inherent form of Tai-Kadai, considering its extensive distribution. Later, it was replaced by new forms in many languages and dialects.

For Zhuang dialects, first, [$tɕi^{3}$] is an unmotivated form, which means "breast, milk", and has no connection with Chinese dialects. Second, phonetically, the initials

are affricates or fricatives having similar places of articulation, the vowels are [-low], and the tones are of B-type. Third, geographically, *tɕ-* type shows a slightly symmetrical distribution in the northwest-southeast direction. To some degree, it can be seen as an ABA distribution with *n-* type, as shown in Figure 2.

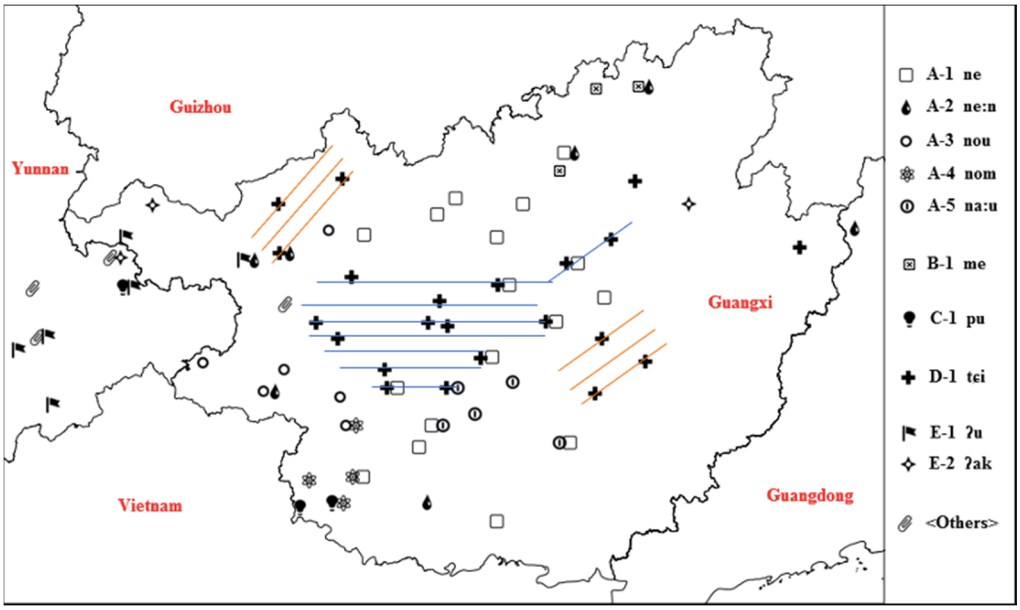


Figure 2: ABA distribution of *tɕ-* and *n-* types.

Therefore, we agree that the lexical forms of *tɕ-* type are native. Also, we suggest that their historical strata are earlier than *n-* type.

2.5. Type E: *ʔ-*

There are three forms of *ʔ-* type: *ʔu*, *ʔak*, and *ʔɛm*.

The E-1 $ʔu^3$ is a common form to represent "breast" in Dai language. Endo (2016) pointed out that the *ʔu* type is mainly found in Yunnan Province. Also, we found that the dialects using *ʔu* in the Zhuang language are mostly located in the border area between Yunnan and Guangxi. According to the records of four books in the *Baiyi Yiyu[百夷译语]*, including *Mangshi Yiyu[芒市译语]*, *Lujiang Yiyu[潞江译语]*, *Wandian Yiyu[湾甸译语]*, and *Gengma Yiyu[耿马译语]*, the word "breast" had several forms in Dehong Dai during the Qing dynasty, as illustrated in Table 2.

Table 2: The "breast" in the *Baiyi Yiyu* of the Qing Dynasty.

	Dai Scripts	Dai Transcription	Zhuyin Scripts	Modern Dai
Mangshi Yiyu		ʔu^{1}	五	ʔu^{1} tau^{3}
Lujiang Yiyu		ʔu^{1}	呅	ʔu^{1}
Wandian Yiyu		lom^{2}	淪	ʔu^{1} tau^{3}
Gengma Yiyu		nom^{2}	浓	ʔu^{1} tau^{3}

As is shown in Table 2, there are two forms to express "breast" in Dehong Dai in the Qing dynasty, ʔu^{1}, and nom^{2}. However, the modern Dai has only one form ʔu^{1} (tau^{3} means gourd). Historically, ʔu^{1} replaced nom^{2} in Dehong Dai. The A-4 *nom^{2}* is a relatively earlier form in Tai languages. Therefore, we suggest that the *ʔu* form replaced the *nom^{2}* in Yunnan province.

Therefore, where does the *ʔu^{3}* form originate? From Figure 1, it can be noticed that the dialects using the form *ʔu^{3}* are mostly located near the Sino-Vietnamese border. The common expression for "breast" in Vietnamese is vú/vu^{45}/, moreover, the Qabiao language which is spoken in the Sino-Vietnamese border area also uses the form ʔu to represent breast (according to Wei 2019). Consequently, it can be speculated that *ʔu* in Zhuang and Dai languages might have been borrowed from Vietnamese. It is thought that the borrowing most likely took place no earlier than the Qing Dynasty based on the alternation of terms in the *Baiyi Yiyu*.

In the Zhuang language, the term E-2 *ʔak^{7}* originally means "chest, thorax", but in some dialects, the distinction between "chest of a man" and "breast of a woman" is no longer made, and both are represented by *ʔak^{7}*. This can be explained as semantic broadening. Using *ʔak^{7}* to refer to the "breast" is a euphemistic way of replacing the body organ with a body part.

ʔɛm^{2} is found in Youjiang district in Baise and northern Guangnan in Wenshan, Yunnan province. It could also be found in the Buyang language, pronounced as ʔem^{55}. They are similar to ʔɛm and mɛːm in Austroasiatic languages, suggesting that they have connections.

What mentioned above (from 2.1 to 2.5) provides the geographical distribution and lexical explanations of the term "breast" in the modern Zhuang dialects. Therefore, our question is as to what the hierarchical relationship between these lexical forms is.

3. Historical strata for "breast" in Zhuang

The term "breast" in modern Zhuang dialects exhibits complex lexical forms and has deep relationships with neighboring languages. The hierarchical relationships can be analyzed based on two different sources, including inherent words and borrowed words.

Regarding correspondences within the same linguistic family, *tɕ-* type has relatively clear cognates in Tai-Kadai, which is widely spread in Tai, Hlai, Kam-Sui, and even Kra branch. Combining the ABA distribution pattern of the *tɕ-* type and *n-* type within the Zhuang language, it is believed that *tɕ-* type is the oldest form, despite it is lacking in southern Zhuang dialects. The nom^{2} is the second oldest form and has a quite restricted distribution among the modern Zhuang dialects. We believe that there is an innovation in Central and Southwestern Tai branches after comparing with Dai and Thai languages. The $na{:}u^{5}$ is a spontaneous innovation in Zhuang and only appears along the border of southern and northern dialects, suggesting that it is relatively new. The inherent forms in Zhuang are speculated to have a hierarchical order of $tɕi^{B}$ → nom^{2} → $na{:}u^{5}$.

The word nou^{4} corresponds to the Chinese character *"敦"* in terms of Chinese loanwords and conforms to the correspondence law of borrowing words in Middle Chinese. The $ne{:}n^{5}$ form is related to the er-final of the word *"奶"* in southern Chinese dialects, possibly borrowed from the Cantonese or Hakka dialect. According to the time when Cantonese entered Guangxi, it is speculated that $ne{:}n^{5}$ was not earlier than the Ming Dynasty. The forms *ne* and $mɒ$ are newer loanwords, which borrowed from *nai ("奶")*, *ma ("妈"),* respectively. The *pe* type is borrowed from the Ping dialect. These are the forms of Chinese loanwords, and their relative chronology is believed to be nou^{4} → $ne{:}n^{5}$ → $ne^{5}/mɒ^{1}/pe^{6}$.

It is speculated that $ʔu^{3}$ is a loanword from Vietnamese. According to records of *Baiyi Yiyu*, there is an alternation between nom^{2} and $ʔu^{3}$ in the Dehong Dai language during the Qing Dynasty. As we mentioned above, nom^{2} is the main form of the Central and Southwestern branches of the Tai language, suggesting that nom^{2} was replaced with $ʔu^{1}$ along the border of China and Vietnam. Based on the alternation, it is likely that the borrowing of *ʔu* has not taken place before the Qing Dynasty.

4. Conclusion

(1) The lexical forms of "breast" in Zhuang dialects are complex because of a variety of factors such as internal innovation, borrowing, and tonal derivation.

(2) We could figure out that the strata of inherent forms are in the order *tɕi*^{B}→ *nom*^{2} → *na:u*^{5}. Besides, the strata of loanwords are: *nou*^{4} → *ne:n*^{5} → *ne*^{5}/*mɒ*^{1}/*pe*^{6}. Additionally, it is speculated that *ʔu*^{3} is a loanword from Vietnamese. In the Zhuang and Dai languages along the Sino-Vietnamese border, nom^{2} has been replaced by *ʔu*^{1},which can be represented as *nom*^{2} → *ʔu*^{1}.

(3) Furthermore, it is hypothesized that the term "breast" is thought to be unsuitable for direct addressing because of its intimate nature to women, which explains why Zhuang dialects have numerous forms and strata for the term "breast". People frequently use borrowed or invented forms of the term "breast" to avoid this taboo, making them more euphemistic and discreet.

References

Cao, Zhiyun[曹志耘] (2008) Hanyu Fangyan Dituji(Cihui Juan)《汉语方言地图集（词汇卷）》[Atlas of Chinese Dialects (Vocabulary Volume)]. Beijing: Commercial Press.

Endo, Mitsuaki (2016) Geolinguistic analysis of 'milk' in Tai-Kadai, *Studies in Asian Geolinguistics III*,18-19.Tokyo: Research Institute for Languages and Cultures of Asia and Africa. URI: https://publication.aa-ken.jp/sag3_milk_2016.pdf

Jin, Lixin[金理新] (2012) *Hanzang Yuxi Hexinci* 《汉藏语系核心词》[*Sino-Tibetan's Basic Words Languages*]. Beijing: Minzu Chubanshe.

Liang, Min[梁敏], Junru Zhang[张均如] (1996) *Dongtai Yuzu Gailun*《侗台语族概论》[*Introduction to the Kam-Tai Languages*], Beijing: Zhongguo Shehui Kexue Chubanshe.

Linguistics Teaching and Research Office of the Department of Chinese Language and Literature at Peking University[北京大学中国语言文学系语言学教研室编](1995) *Hanyu Fangyan Cihui (Di'er Ban)*《汉语方言词汇（第二版）》[*Vocabulary of Chinese Dialects (2nd edition)*]. Beijing: Language and Literature Press.

Liu, Zhenfa[刘镇发] (1998) Keyueyu Linjin Minzu Yuyan de Gongtongci 客粤语临近民族语言的共同词 [Common Words between Hakka, Cantonese and Minority Languages]. *Linguistic Research* 4: 44-50.

Meng, Yuanyao[蒙元耀] (1998) *Zhuanghanyu Tongyuanci Yanjiu*《壮汉语同源词研究》 [*A Study on Cognate Words between Zhuang and Chinese*]. Beijing: Minzu Chubanshe.

Swadesh, Morris (1955) Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics* 21(2): 121–137.

Takashi Kosaka[小坂隆一], Guoyan Zhou[周国炎], Jinfang Li[李锦芳] (1998) *Geyang Yuyan Cihuiji* 《仡央语言词汇集》 [*Vocabulary Collection of Gelao Language*]. Guiyang: Guizhou Minzu Chubanshe.

Takashi Ueya, Kenji Yagi (2016) Geolinguistic analysis of 'milk' in Sinitic language, *Studies in Asian Geolinguistics III*, 12-13.Tokyo: Research Institute for Languages and Cultures of Asia and Africa. https://publication.aa-ken.jp/sag3_milk_2016.pdf

Tang, Qiyuan[唐七元] (2009) *Hanyu Fangyan Tongyuanci Yanjiu*《汉语方言同源词研究》[*A Study on Cognate Words in Chinese Dialects*]. PhD Dissertation of Fudan University.

The Working Committee of Minority Language and Script of Guangxi[广西少数民族语言文字工作委员会] (2018) *Zhuanghan Cihui*《壮汉词汇》[*Zhuang-Chinese Vocabulary*]. Nanning: Guangxi Minzu Chubanshe.

The Working Committee of Minority Language and Script of Guangxi[广西少数民族语言文字工作委员会] (2008) *Guangxi Minzu Yuyan Fangyin Cihui*《广西民族语言方音词汇》[*Vocabulary of Minority Languages in Guangxi*]. Beijing: Minzu Chubanshe.

Wang, Weihui [汪维辉] (2018)*Hanyu Hexinci de Lishi yu Xianzhuang Yanjiu*《汉语核心词的历史与现状研究》[*A Study on the History and Current Situation of Chinese Core Words*]. Beijing: Shangwu Yinshuguan.

Wei, Shuguan[韦树关], Haiyun Yan[颜海云], Sha Li[黎莎] (2019) *Guowai Zhuangdong Yuzu Yuyan Cihuiji*《国外壮侗语族语言词汇集》[*Foreign Kam-Tai Language Vocabulary Collection*] . Beijing: Shijie Tushu Chuban Gongsi.

Wu, Anqi[吴安其] (2017) *Ya'ou Yuyan Jibenci Bijiao Yanjiu*《亚欧语言基本词比较研究》[*A Comparative Study of Basic Words in Asian and European Languages*]. Beijing: Zhongguo Shehui Kexue Chubanshe.

Xie, Liuwen[谢留文] (2003) *Kejia Fangyan Yuyin Yanjiu*《客家方言语音研究》[*Phonetic Study of Hakka Dialects*]. Beijing: Zhongguo Shehui Kexue Chubanshe.

Xie, Jianyou [谢建猷] (2007) *Guangxi Hanyu Fangyan Yanjiu*《广西汉语方言研究》[*Research on Chinese Dialects in Guangxi*]. Nanning: Guangxi Renmin Chubanshe.

Xing, Gongwan[邢公畹] (1999) *Hantaiyu Bijiao Shouce*《汉台语比较手册》[*Comparative Handbook of Chinese and Tai*]. Beijing: Shangwu Yinshuguan.

Yu, Jin[余瑾] (2016) *Guangxi Pinghua Yanjiu*《广西平话研究》[*A Study of Ping Dialect in Guangxi*]. Beijing: Zhongguo Shehui Kexue Chubanshe.

Zhang, Fugang[章富刚] (2018) *Jiyu Ming Qing "Baiyi Yiyu" de Dehong Daiyu ji Yunnan Guanhua Yuyin Lishi Yanjiu*《基于明清"百夷译语"的德宏傣语及云南官话语音历史研究》[*A Study on Phonetic Evolution of Dehong Dai and Yunnan Mandarin Based on Baiyi-Yiyu of Ming and Qing Dynasty*]. PhD Dissertation of Nankai University.

Zhang, Junru[张均如], Min Liang[梁敏], Jueya Ouyang[欧阳觉亚], Yiqing Zheng[郑贻青], Xulian Li[李旭练], Jianyou Xie[谢建猷] (1999) *Zhuangyu Fangyan Yanjiu*《壮语方言研究》[*A Study on Zhuang Dialects*]. Chengdu: Sichuan Minzu Chubanshe.

Zhengzhang, Shangfang[郑张尚芳] (2014) Hantai Shentici Tongyuan Bijiao Wushi Cili 汉泰身体词同源比较五十词例 [A Comparative Study of 50 Body Terms between Han and Tai Languages]. *Minority Languages of China* 3: 22-27.

Zhou, Guoyan[周国炎] (2011) *Buyi-Han Cidian*《布依-汉词典》[*Bouyei-Chinese Dictionary*]. Guiyang: Guizhou Minzu Chubanshe.

Geolinguistic analysis of the word form derived from *red* in Tibetic languages in Khams and Amdo

Hiroyuki Suzuki
(Kyoto University)

Abstract: This article describes how the word form derived from Literary Tibetan *red* is used in varieties of Tibetic languages in the Khams and Amdo regions. Most varieties use these forms as copulative verb stems, and in most cases, it functions as a factual or statemental (+sensory) evidential form. However, we observe slight differences in its usage, and we can classify its functions across different languages as follows: (A) copulative verb stem, (B) stative verb, and (C) no corresponding words used. Varieties with Types B and C do not use *red* as a copulative verb stem for the factual evidential; however, in most cases, they also have the factual evidential category, for which a stem derived from another Literary Tibetan word form is used in place of *red*. The article draws maps of the distribution of the types listed above and interprets their features. The results of this can be used to support an analysis of the development of the use of *red* in the target region. In addition, we discuss how the factual evidential category developed in Tibetic languages.

Key words: factual evidential, lexical development, grammaticalisation, copulative verb, eastern Tibetosphere

1. Introduction

This article describes how the word form derived from Literary Tibetan (LT) *red* is used in varieties of Tibetic languages in the Khams and Amdo regions. Linguistic features and the origin of LT *red* are discussed (Takeuchi 1990, 2015; Shao 2016; Hoshi 2021; Zeisler 2022). To characterise *red*, we focus on its verb classification, especially with respect to verb types and the evidential system. First, the verbs of the Tibetic languages (see Tournadre and Suzuki 2023) can be classified into three

SUZUKI, Hiroyuki. 2023. Geolinguistic analysis of the word form derived from *red* in Tibetic languages in Khams and Amdo. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 62–73. doi: https://doi.org/10.5281/zenodo.8374577

* This work was supported by two grants-in-aid of Japan Society of the Promotion of Sciences 'Investigation of undescribed languages in the eastern Tibetosphere and their geolinguistic research' (headed by Hiroyuki Suzuki; No. JP17H04774) and 'Historical geolinguistics over the whole periods of the Sinitic language formation with the migration and language contact perspectives' (headed by Mitsuaki Endo; No. JP23H00625).

broader types: copulative, existential, and lexical verbs, where *red* appears is a copulative verb. Second, most Tibetic languages exhibit a complex evidential system, regarding which various analyses and descriptions are provided.[1] Here, I follow a system based on the arguments of Tournadre and LaPolla (2014) and Tournadre (2017) that is formulated as access-type evidentials, consisting of at least six categories (egophoric, factual/statemental, visual sensory, nonvisual sensory, sensory inferential, and logical inferential), as used by Dawa Drolma and Suzuki (forthcoming). A formulation of the copulative verbs using evidential categories of Lhagang Tibetan (Minyag Rabgang Khams) and rGyalthang Tibetan (Sems-kyi-nyila Khams) is shown in Table 1 as an example, with an emphasis on the position of the stem *red* in bold.[2] The statemental category is often called factual, and both terms denote access to general knowledge; we thus present this as factual/statemental.[3]

Table 1: Copulative verb system of Lhagang Tibetan and rGyalthang Tibetan.

	egophoric	statemental	visual sensory	nonvisual sensory	sensory inferential	logical inferential
Lhagang	*yin*	***red***			*yin sa red*	*yin le red*
rGyalthang	*zin*	***red***	*sni* / *zin snang*	*grags* / *zin grag*	*zin* + various suffixes	

As shown in Table 1, the form *red* is closely related to the development of the evidential system(s) in the Tibetic languages, as the morphological encoding of evidentiality was acquired later in their evolution from Old Tibetan (Hoshi 2016; Tournadre 2017; Takeuchi 2021). Takeuchi (1990, 2015) suggests that the development of *red* started begin somewhere in the eastern Tibetosphere.

Examples (1–3) show the use of *red* in modern languages. The forms are presented in the LT transliteration, following the method of de Nebesky-Wojkowitz (1956).

(1) *kho* *dge rgan* *red*
3 teacher CPV.STM
'He is a teacher.' (Lhagang; Minyag Rabgang group)

[1] See Suzuki et al. (2021) for a summary of relevant studies.
[2] We also find *red* in the inferential categories; however, we exclude it from the discussion.
[3] We find, in fact, pragmatic differences between them (Zhou and Suzuki 2022). However, this article deals with both in the same category.

(2) *de* *phag rgan* *red*
that pig CPV.STM
'That is a pig.' (Lithang; Southern Route group)

(3) *yi'i* *dgun chang* *red*
this wine CPV.STM
'This is wine.' (gYanggril; sDerong-nJol group)

Most varieties use *red* as a copulative verb stem, and in most cases it serves as a factual or statemental (+sensory, as in Lhagang) evidential stem. A sensory evidential refers to sensory access to information, including visual and auditory evidence. As seen in Table 1, some Tibetic languages distinguish between visual sensory and nonvisual sensory, and *red* is apparently used in the sensory category in those varieties that do not distinguish visual from nonvisual sensory evidence.

However, we find other usages of *red*, such as in an interjection to indicate agreement ('yes', 'you are right', etc.) or backchannel, as in (4b) in a conversation between A and B.

(4) a. *kho* *bod* *red* *ya*
3 Tibetan CPV.STM SFT
A: 'Oh, he is Tibetan.'
b. *red* *red* *red* *red*
INTJ INTJ INTJ INTJ
B: 'Yeah, yeah, you're right!' (nDawpa; sPomborgang group)

Marginally, we also find another use of *red* as an adjective in (5).

(5) *ma de* *red-snang*
this be good-STA.VSEN
'This is good.' (Zhollam; Sems-kyi-nyila group)

To summarise, we find three principal usages of *red* in modern Tibetic languages spoken in the eastern Tibetosphere: copulative factual/statemental verb stems, interjections, and adjectives meaning 'be good'.[4] From this, this article maps the use

[4] The use as an adjective is limited to a predicate and cannot function as an epithet.

and function of *red* to determine how the form *red* has evolved in the given linguistic region.

2. Classifying the function of *red*

As noted, three principal usages are attested in the Tibetic languages of the target region. These usages are independent of each other, but data from fieldwork show that *red* is used in at least two of the three ways in many varieties, and it is not used at all in a few varieties.

On the basis of these observations, I classify the uses to be distinguished for the present analysis into three major types A, B, and C, with two subclassifications for types A and B.[5]

A: copulative verb stem
  A-1: factual (+sensory) evidential only
  A-2: factual (+sensory) evidential + interjection meaning 'yes, alright'
B: stative verb
  B-1: stative verb meaning 'be good'
  B-2: stative verb meaning 'be good' + interjection meaning 'yes, alright'
C: no corresponding words used

Varieties showing Types B and C do not use *red* as a copulative verb stem for the factual evidential; however, in most cases, they also have a factual evidential category, where a stem derived from another LT word form (e.g., *snang*; see Suzuki 2012) is used instead of *red*.

3. Mapping *red* in the eastern Tibetosphere

This section provides maps of the distribution of the types classified in Section 2 and interprets their features. Doing so can support the analysis of the development of *red*

[5] It does not matter for the purposes of the present article whether *red* is also used as a copulative sensory form, as this discussion relates to how the evidential system works in a given Tibetic language. See Suzuki et al. (2021) and Dawa Drolma and Suzuki (forthcoming) for the evidential system.

in the target region. In addition, we discuss the way that the factual/statemental evidential category has developed in Tibetic languages.

3.1. Mapping and geolinguistic analysis

The map in Figure 1 reflects the use of *red* in the target region, according to the classification provided in Section 2.

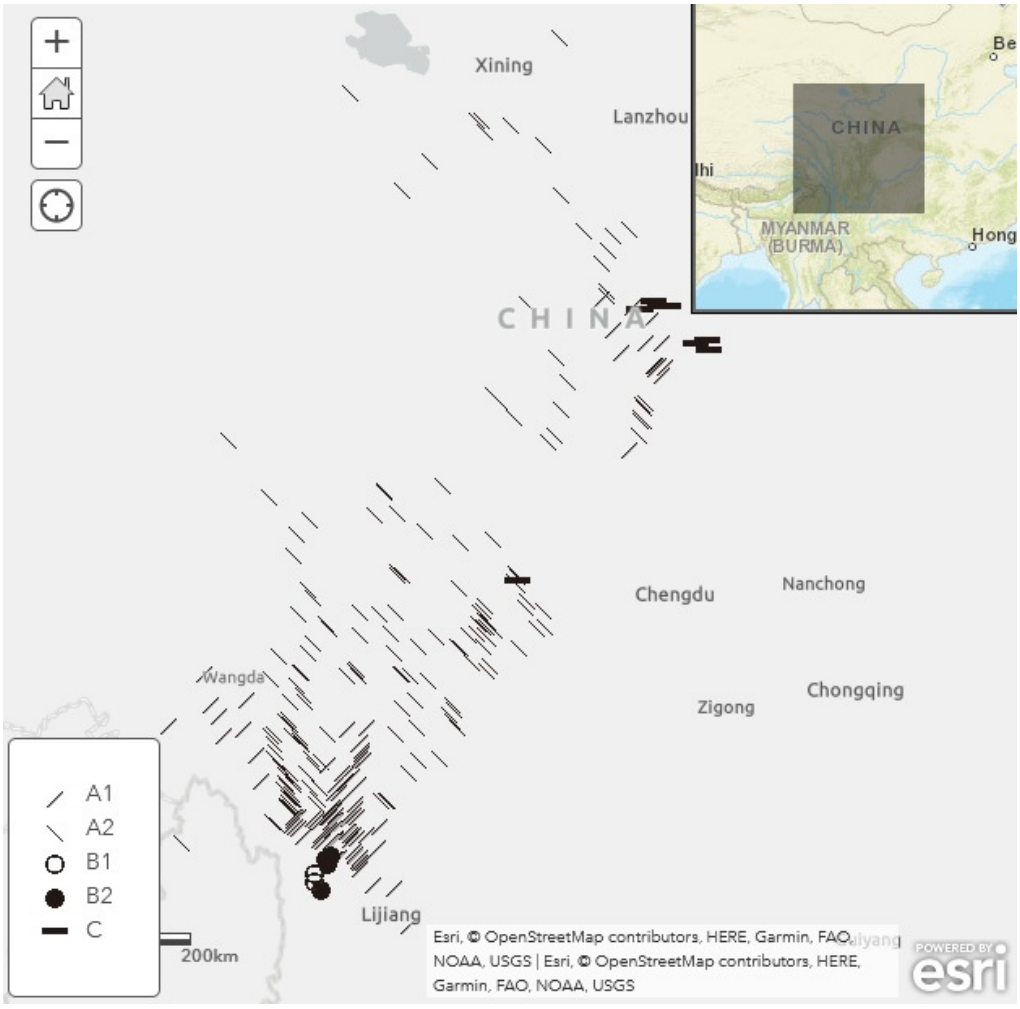


Figure 1: Use of *red* in the Tibetic varieties in the eastern Tibetosphere.

Most varieties are of Type A. Type B is found in the southernmost area. Type C is found in the northeasternmost area, with the exception of a variety spoken on the eastern side of the middle of Figure 1. Types B and C are minority types.

Type A's subclassifications, A1 and A2, are distributed everywhere. Type A1 is found on the eastern and southern peripheries, where it has contact with Types B and C, respectively; it is unclear whether this is coincidental or conditional.

Figure 2 provides an enlarged map of the southern part of Figure 1. The base map contains main river flows and traffic routes to provide an interpretation of Type B.

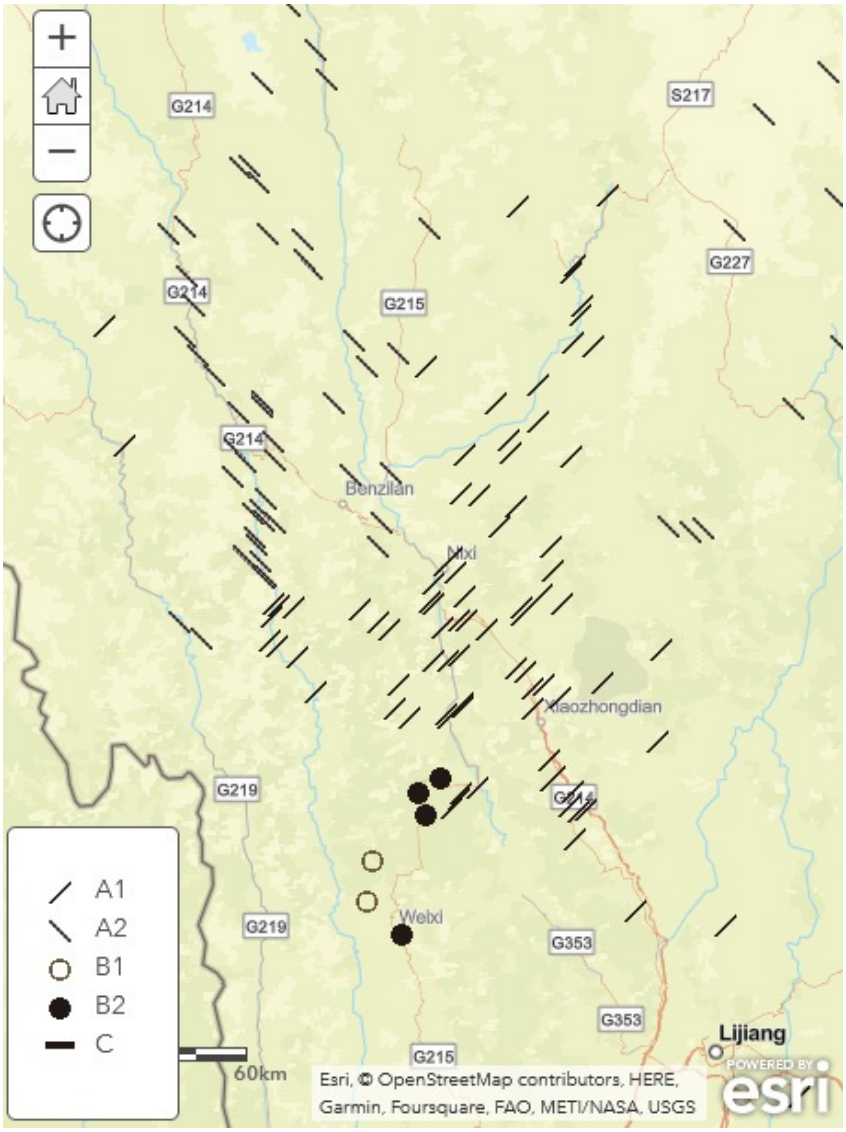


Figure 2: Use of *red* in the Tibetic varieties in the south-eastern Tibetosphere.

Type B only appears in the southernmost area of Figure 2. Varieties with Type B belong to the Melung subgroup of the Sems-kyi-nyila group. These varieties use a word form that is derived from LT *snang* 'appear' in the factual category, and *red* is used as an adjectival predicate, as in (5).

The distribution of B2 is along the large traffic road, whereas that of B1 is in a mountainous area. Because the southernmost point of B2 was populated by recent migrants from Tacheng township in the north (three places that exhibit B2), it is natural that linguistic features between them are similar. Both B1 and B2 are surrounded by non-Tibetic languages, such as Naxi, Malimasa, and Lisu. Although the varieties that have B2 appear to have a connexion with those with A1 in Tacheng Town, the two groups of varieties are separated by hamlets populated by Naxi-speaking people (Suzuki 2017). Hence, contact with other Tibetic languages possibly may have influenced of the distribution of Type B.

Figure 3 presents an enlarged version of the north-eastern part of Figure 1. The base map shows the main river flows and traffic routes to interpret the distribution of Type C.

Figure 3: Use of *red* in the Tibetic varieties in the north-eastern Tibetosphere.

Most varieties with Type C appear in the northeasternmost area of Figure 1, an enlargement of which is given in Figure 3. The given varieties exhibit the Thewo-smad and mBrugchu dialects. These varieties use a word form derived from LT *'gyur* 'change' in the factual category. In Tibetic, *'gyur* is a conspicuous form.

One exception is attested in the easternmost area, namely, in the Sogpho dialect (Rongbrag Tibetan; Khams). Here, the compound form *yin snang* is used instead. This form is not shared with other surrounding varieties of Rongbrag Tibetan, in which we find *red* as a copulative statemental verb. Because the distribution of

Rongbrag Tibetan is surrounded by Geshitsa (to the north-west), bTsanlha rGyalrong (to the north-east), and nGochang (to the south), and is thus independent from other Tibetic languages, direct influence from Tibetic is assumed for an evolution of the copulative statemental verb. However, we should take into account the cultural (especially religious) influences and intercommunication within the Tibetosphere, which could trigger language contact between a local vernacular and a culturally superstratum variety. This allows us to assume a change from Type C to Type A2, which would mean that Sogpho retains an older form.

3.2. Discussion

I will briefly discuss the expansion of the use of the form *red* as a copulative verb (Type A) relative to other cases (Types B and C). This expansion is related to the development of the evidential system, especially the marking of the factual/statemental evidential.

Although Type A is most commonly found in the majority of the eastern Tibetosphere (as well as across the central area; see Tournadre and Jiatso 2001), it is regarded as a more recent feature. Instead, Types B and C would be the older cases if we apply a concentric circle distribution (a.k.a. ABA distribution) from the centre (here, provisionally Lhasa) to the eastern Tibetosphere. However, Tibetic languages spoken in the rest of the Tibetosphere do not use *red* in the way that Type A does, and we should discuss what triggered this inequal development.

Takeuchi (1990) proposed that the development of *red* began in the north-eastern Tibetosphere, namely, from Amdo. His argument is based on the appearance of *red* in a document reflecting oral Amdo edited in the eighteenth century (1990:13). However, as he also states, it is difficult to identify the time and region for the beginning of its use.[6]

It appears that *red* has gone through grammaticalisation, that is, a lexical verb changing into copulative verb, and further into interjection. The lexical verb stem *red* is used as an adjective predicate meaning ‘be good’, ‘be well’, ‘work well’, etc. in the varieties with Type C. We find another use as a verb in other varieties, where *red* follows an adjective to denote ‘get to be’ and ‘become’. This usage is attested in Medieval Tibetan (Hoshi 2021:173). These meanings are considered to be essential uses of *red* as a lexical verb, and its factual/statemental function was derived from them, possibly intended to denote ‘to declare something good/right/testimonial of

[6] See Zeisler (2022) for a philological trait of *red*.

changes'. The 'archaic' meaning is found in the south-eastern region. The development from 'declaration' to 'factual' took place in another verb stem (*'gyur*) in the eastern region. The limited distribution of Types B and C suggests the archaicity of the given function of *red*.

The process of grammaticalisation suggests that Tibetic languages, at least those spoken in the eastern and central areas, acquired means of encoding factual/statemental evidentiality using a specific word form. Egophoric access has received significant attention in the literature on evidentiality in Tibetic languages; however, it seems that the distinctive feature is a phenomenon of the factual/statemental evidential before establishing the egophoric evidential. This question is left for further investigation.

4. Conclusion

This article describes the uses of the word *red* in the Tibetic varieties in the eastern Tibetosphere, focusing on its development. A linguistic map shows that the function of *red* as a copulative verb is prominent, and in only some minority varieties is it used as an adjective or not at all.

Although the eastern Tibetosphere appears to be the site of the origin of the development of *red* as a copulative verb, the fact that this use of *red* is pervasive does not suggest a direction of development from the east to the central region. The following process is more plausible: the origin of *red* is somewhere in the eastern Tibetosphere, where it was first acquired by influential varieties such as Lhasa Tibetan and then expanded to the east in the period when the connexion between Lhasa and the eastern Tibetosphere was closer than that with other areas in cultural and religious aspects.

Abbreviations

-	morpheme boundary
3	third person
CPV	copulative
INTJ	interjection
SFT	sentence final tag
STM	statemental
VSEN	visual sensory

References

Dawa Drolma & Hiroyuki Suzuki (forthcoming) The paradigmaticity of evidentials in the Tibetic languages of Khams. *Studies in Language*. https://doi.org/10.1075/sl.23006.dro

Hoshi, Izumi [星泉] (2016) *Koten tibettogo bunpoo: Ootoomeikyoosi (14 seiki) ni motoduite* 古典チベット語文法：『王統明鏡史』（14世紀）に基づいて [Grammar of Classical Tibetan: Based on the *Royal Genealogy* (14th Century)]. Fuchu: Research Institute for Languages and Cultures of Asia and Africa. https://doi.org/10.15026/94480

Hoshi, Izumi [星泉] (2021) Tibettogo no bunpoo wo dooteki ni miru チベット語の文法を動的に見る[To look at the Tibetan grammar in a dynamic way]. In Kazushi Iwao & Takumi Ikeda (eds.) *Tibetto no rekisi to syakai, gekan [syakai hen, gengo hen]*, 161–195. Kyoto: Rinsen Syoten.

de Nebesky-Wojkowitz, René (1956) *Oracles and demons of Tibet: The cult and iconography of the Tibetan protective deities*. 's-Gravenhage: Mouton.

Shao, Mingyuan (2016) Zangyuzu xidongci red de yufahua. *Language and Linguistics* 17(5): 679–715. https://doi.org/10.1177/1606822X16645742

Suzuki, Hiroyuki (2012) Multiple usages of the verb *snang* in Gagatang Tibetan (Weixi, Yunnan). *Himalayan Linguistics* 11(1): 1–16. http://doi.org/10.5070/H911123711

Suzuki, Hiroyuki (2017) The vitality of Khams Tibetan varieties in Weixi County. *Asian Highlands Perspectives* 44: 256–284. http://himalaya.socanth.cam.ac.uk/collections/journals/ahp/pdf/AHP_44.pdf

Suzuki, Hiroyuki (2022) *Geolinguistics in the eastern Tibetosphere: An introduction*. Geolinguistic Society of Japan. https://doi.org/10.5281/zenodo.5989176

Suzuki, Hiroyuki, Sonam Wangmo, & Tsering Samdrup (2021) A contrastive approach to the evidential system in Tibetic languages: Examining five varieties from Khams and Amdo. *Gengo Kenkyu* 159: 69–101. https://doi.org/10.11435/gengo.159.0_69

Takeuchi, Tsuguhito [武内紹人] (1990) Tibettogo no zyutubu ni okeru zyodoosi no kinoo to sono hattatu katei チベット語の述部における助動詞の機能とその発達過程 [The semantic function of auxiliary verbs in Tibetan and their historical development]. In Osamu Sakiyama & Akihiro Sato (eds) *Azia no gengo to ippan gengogaku*, 6–16. Tokyo: Sanseidoo.

Takeuchi, Tsuguhito (2015) The function of auxiliary verbs in Tibetan predicates and their historical development. *Revue d'études tibétaines* 31: 401–415. http://himalaya.socanth.cam.ac.uk/collections/journals/ret/pdf/ret_31_25.pdf

Takeuchi, Tsuguhito (2021) History of the Tibetan language. In Yasuhiko Nagano & Takumi Ikeda (eds.) *Grammatical phenomena of Sino-Tibetan languages 4: Link languages and archetypes in Tibeto-Burman*, 303–323. Kyoto: Research Institute for Humanities, Kyoto University. http://hdl.handle.net/2433/263983

Tournadre, Nicolas & Randy J. LaPolla (2014) Towards a new approach to evidentiality: Issues and directions for research. *Linguistics of the Tibeto-Burman Area* 37(2): 240–263. https://doi.org/10.1075/ltba.37.2.04tou

Tournadre, Nicolas (2017) A typological sketch of evidential/epistemic categories in the Tibetic languages. In: Lauren Gawne & Nathan W. Hill (eds.) *Evidential systems in Tibetan languages*, 95–129. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110473742-004

Tournadre, Nicolas & Hiroyuki Suzuki (2023) *The Tibetic language: An introduction to the family of languages derived from Old Tibetan*. Villejuif: LACITO Publications. https://lacito.cnrs.fr/the-tibetic-languages/

Zeisler, Bettina (2022) To be or not to be: On the Modern Tibetan auxiliary verb *red* in classical texts. *Himalayan Linguistics* 21(3): 50–84. https://doi.org/10.5070/H921352173

Zhou, Yang & Hiroyuki Suzuki (2022) Evidentiality in Selibu: A contact-induced emergence. *Diachronica* 39(2): 268–309. https://doi.org/10.1075/dia.19055.zho

Regional Differences in Syllable Fusion in Japanese Dialects

Kohei NAKAZAWA
(Shinshu University)

Abstract: When two syllables are fused into one, three types of changes occur in Japanese: 1. a change resulting in a long vowel (VV type); 2. a change resulting in a geminate consonant (CC type); and 3. a change resulting in neither a long vowel nor a geminate consonant (X type). Syllable fusion, which is the fusion of two syllables, is observed mainly in the following three environments in Japonic: 1. CV(S)V, 2. CVrV, and 3. $C_{[\alpha\ place]}VC_{[\alpha\ place]}V$. Of these, the CV(S)V environment is widely observed in Japanese. Syllable fusion in CV(S)V can be found in all languages, while all three syllable-fusion types (the VV, CC, and X types) occur under the condition of $CV_{[+high]}(S)V_{[-high]}$. Under this condition, regional differences are observed within Japanese. This study reviews the geographic distribution of these types and discusses diachronic relationships.

Key words: Japanese, sound change, syllable fusion, historical phonology

1. Introduction

According to the *Hōgen Bumpō Zenkoku Chizu* [Grammar Atlas of Japanese Dialects] (GAJ), various forms of the progressive *cirijoru* '(flowers) be falling' exist in Japonic (Japanese and Ryukyuan) languages, including forms with long vowels such as *cirj**oo**ru* and those with geminates (or NC clusters) such as *ci**rr**joru* (*ci**nr**joru*).

Figure 1 is a map of the GAJ, and Figure 2 is redrawn from Figure 1, extracting the cognate forms as *cirijoru*.[1] The areas using *cirj**oo**ru* and those using *ci**rr**joru* (or *ci**nr**joru*) roughly constitute geographical clusters. This suggests that there may be a geographical pattern of long vowel and geminate consonant forms throughout Japan. Unfortunately, Eastern Japanese dialects do not use *-joru* for the progressive; thus, there is no information on Eastern Japanese dialectal forms for the progressive.

NAKAZAWA, Kohei. 2023. Regional Differences in Syllable Fusion in Japanese Dialects. In Cẩm Lan Trịnh, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference of Asian Geolinguistics*, 74–85. doi: https://doi.org/10.5281/zenodo.8374589

* This work was supported by JSPS KAKENHI Grant Number JP21K12993.

[1] The data available at NINJAL (方言文法全国地図全データ, https://www2.ninjal.ac.jp/hogen/dp/gaj_all/gaj_all.html) were used to draw Figure 2.

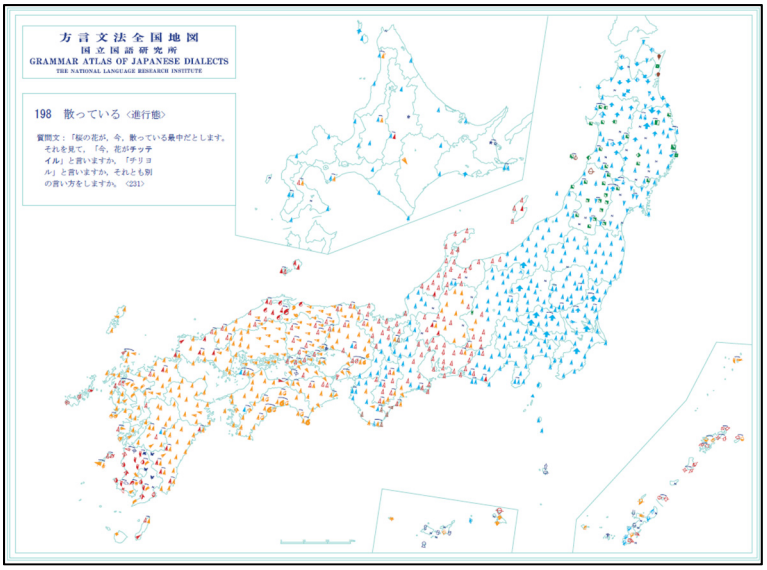


Figure 1: GAJ 198 *citteiru* '(flowers) be falling' (progressive)

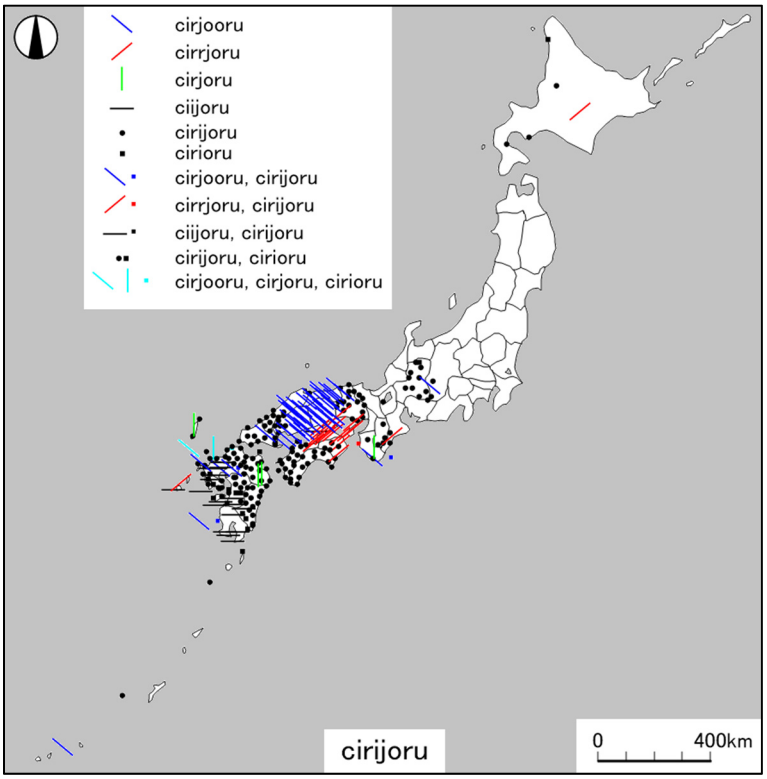


Figure 2: Distribution of *CIRYOORU*, *CIRRYORU*, and *CIRYORU* from GAJ 198

When two syllables are fused into one, three types of changes occur: 1. a change resulting in a long vowel, as seen in ***teo****no* > ***cjoo****no* 'hand axe' (VV type); 2. a change resulting in a geminate consonant, as seen in *ha****sira*** > *ha****ssja*** 'column' (CC type); and 3. a change resulting in neither a long vowel nor a geminate consonant, as seen in ***seo****u* > ***sjo****u* 'to carry' (X type) (cf. Nakazawa 2021).

The distribution of these variations is partially shown in Uwano et al. (1989); however, it is not exhaustive. Therefore, this study analyzes the regional differences in syllable fusion in Japonic and considers the historical relationships between these changes.

2. Japonic and Japanese archipelago

Japonic language is spoken in the Japanese archipelago, which is located on the eastern side of the Eurasian continent.

Japonic is broadly divided into Japanese and Ryukyuan. Because Ryukyuan languages have undergone more drastic sound changes than Japanese, this study focuses only on Japanese dialects.

Japan is divided into the following regions, and Japanese languages or dialects are also generally divided according to these regions: Hokkaido, Tōhoku, Kantō, Chubu, Hokuriku, Kinki, Chūgoku, Shikoku, and Kyūshū (see Figure 3)[2].

3. Method

Syllable fusion is observed under the following conditions: 1. CV(S)V, 2. $C_{[\alpha\ place]}VC_{[\alpha\ place]}V$, and 3. CVrV. Among these, syllable fusion in CV(S)V can be found in all Japanese dialects, and all types of syllable fusion (the VV, CC, and X types) occur under the condition of $CV_{[+high]}(S)V_{[-high]}$. Consequently, by limiting the condition to $CV_{[+high]}(S)V_{[-high]}$, dialectal forms are gathered from *namari* ([corrupted forms], see Figure 4) in *Nihon Kokugo Daijiten* [Shogakukan's Japanese Dictionary] second edition[3]. Among the 200 words with the phonological condition of $CV_{[+high]}(S)V_{[-high]}$, 94 have dialectal forms with syllable fusion.

[2] Free materials from Craft MAP (http://www.craftmap.box-i.net/map.php) were used to create Figure 3.

[3] JapanKnowledge Lib (https://japanknowledge.com/library/) was used for data collection.

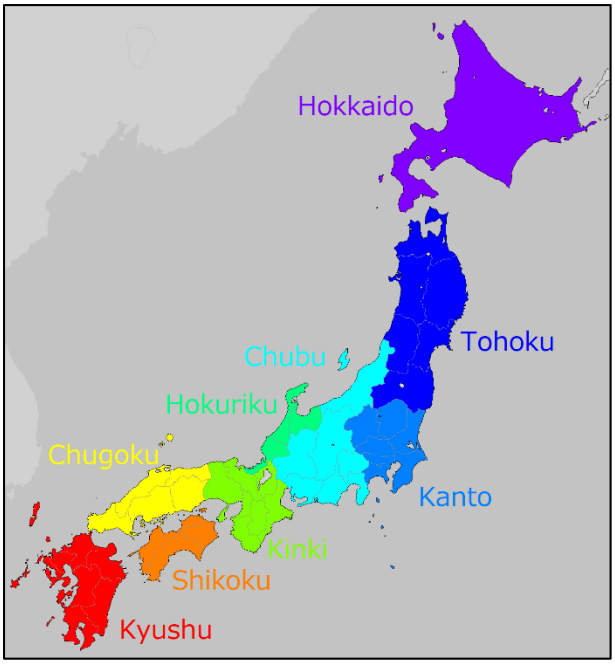


Figure 3: Regions of Japan

〈なまり〉イー〔ＮＨＫ（茨城）〕 イイ〔山梨〕 イウエ・イベ〔富山県〕 イエ〔瀬戸内〕 イエー〔ＮＨＫ（兵庫）〕 イゲ〔讃岐〕 ウエ〔岐阜・飛騨・愛知〕 エ〔青森・岩手・福島・栃木・千葉・岐阜・飛騨・淡路・播磨・ＮＨＫ（鳥取）・島根・広島県・佐賀・島原方言・鹿児島〕 エイ〔岩手・仙台音韻〕 エイヘ〔秋田〕 エウェ〔石川・島根〕 エエ〔播磨〕 エー〔岩手・栃木・千葉・山梨・ＮＨＫ（滋賀・兵庫）・淡路・和歌山県・紀州・島根〕 エッコ〔津軽ことば・秋田〕 エベ〔富山県・飛騨〕 ジェー〔ＮＨＫ（福島）・福島〕 ゼエ〔福島〕 ユエ〔山形〕

Figure 4: *namari* of いえ【家】in *Nihon Kokugo Daijiten* 2nd edition (JapanKnowledge Lib)

As shown in (1), 14 instances of syllable fusion are identified within bisyllabic words. There are 24 occurrences of syllable fusion at the beginning of a word. Seventeen instances of syllable fusion are observed at the end of a word. Additionally, 39 words exhibited syllable fusion in the middle of a word. A list of words for which no syllable fusion is found or cannot be classified is provided in (2).

Because syllable fusions in bisyllabic words behave phonologically close to those at the beginning of a word and those at the end of a word behave phonologically close to those in the middle of a word, the syllable fusion of two-syllable words is included in that at the beginning of a word, while the end of a word is included in the middle of the word.

(1) a. /ie/: いえ【家】, ひえ【稗】, /io/: しお【塩】, しお【潮】, にお【堆】, にお【鳰】, /iwa/: にわ【庭】, /ue/: うえ【上】, すえ【末】, つえ【杖】, ふえ【笛】, /uo/: うお【魚】, /uwa/: くわ【桑】, くわ【鍬】(14 words)

b. /ia/: しあさって, しあわせ, /ie/: きえる, にえる, みえる, /io/: しおからい, しおれる, /ija/: いやしい, いやらしい, ひやかす, /ijo/: ひよこ, ひよめき, ひより, /iwa/: しわぶき, しわぶる, /ua/: ぐあい, /ue/: うえる【植】, すえる【据】, すえる【饐】, /uo/: うおのめ, /uwa/: うわさ, くわい, くわのみ, すわる【座る】(24 words)

c. /ie/: なまにえ, /io/: うしお, てしお, わらにお, /ija/: うしや, おじや, かじや, しちや, /ijo/: つきよ, /iwa/: うちわ, ゆびわ, /ue/: つくえ, /uo/: かつお, /uja/: しゃくや, /uwa/: うつわ, まくわ, まぐわ (17 words)

d. /ia/: あきあげ, あしあと, うちあい, おきあがる, こしあげ, さしあげる, つきあい, つきあたり, とびあがる, とりあげ, とりあげる, のしあげる, ひきあう, ひきあげる, もちあげる, ゆきあう, わりあい, /ie/: おしえる, おびえる, /io/: いきおい, うちおく, こしおび, とりおく, へしおる, /ija/: うちやる, おみやげ, つきやま, にぎやか, /ijo/: としより, にちよう, /iwa/: いしわら, にぎわう, ひきわり, むぎわら, /uo/: かつおぶし, /uja/: まつやに, /ujo/: げつよう, /uwa/: でくわす, まくわうり (39 words)

(2) あきや, あしおと, あじわい, あによめ, いあい, いえる, いおう, いしずえ, いじわる, いつわる, いやす, いりあい, いわ, いわう, いわし, いわや, うえる【飢】, うきわ, うちよせる, うやまう, うわぎ, うわざ, うわつく, うわなり, うわべ, おきあい, かぐわしい, かしわ, かりや, きあい, きおい, きおち, きやすい, きやすめ, きよめる, きわ, きわどい, きわまる, くちぶえ, くつわ, くびわ, くみあい, くやしい, くわえる, くわしい, くわわる, こずえ, さしあい, しあけ, しあげ, しあげる, しあるく, しおり, しわ, しわけ, しわざ, しわよせ, しわらくさい, たくわえる, たちあい, つや, つよい, つりあい, ときわ, とびうお, なりわい, にあう, におい, におう, にぎやしい, にやける, にわつとり, にわとり, のしあがる, のりあがる, はにわ, はにわり, ひあがる, ひあぶり, ひえる, ひやけ, ひやこい, ひやじる, ひやす, びわ【枇杷】, ふえる, ふやす, まぎわ, まじえる, まじわる, まちや, まつわる, みあげる, みぎわ, みずおとし, みつあい, みつあし, みつあみ, みやげ, みやこ, みやすい, みやぶる, みわける, もちあげ, ゆくえ, よつあし (106 words)

4. Results

Among the three types of syllable fusion, the X type (the light syllable type) is the most frequent. We would like to focus on regional differences between the VV and CC types.

4.1. Syllable fusion at the beginning of a word

The VV type is often observed at the beginning of a word.

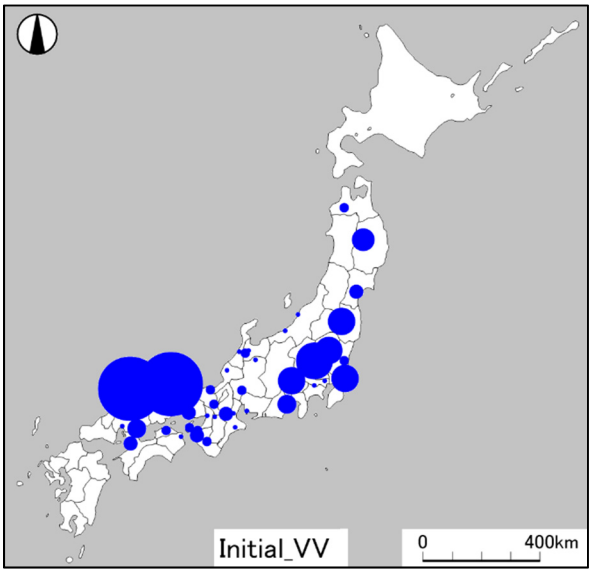


Figure 4: The frequency of the VV type at the beginning of a word

Figure 4 shows the frequency of the VV type at the beginning of a word. The size of the circle corresponds to the number of dialectal forms. The larger the circle, the greater the number of dialectal forms. The VV type is generally found throughout Japan; however, it is somewhat rare in the Chubu and Kinki regions, and it is not found in Kyūshū.

In contrast to the VV type, the CC type is rarely observed at the beginning of a word. The CC type is found only infrequently in the Hokuriku and other areas.

(3) a. *nio* > *nnjo* 'haystack,' *sio* > *ssjo* 'salt,' *kuwa* > *kkwa* 'hoe'
b. *mieru* > *mmeru*, *siwaburu* > *ssaburu*, *suwaru* > *ssaru*

Consequently, significant regional variations can be found in the changes that occur within words.

The X type is found throughout Japan and is especially common in the Tōhoku region.

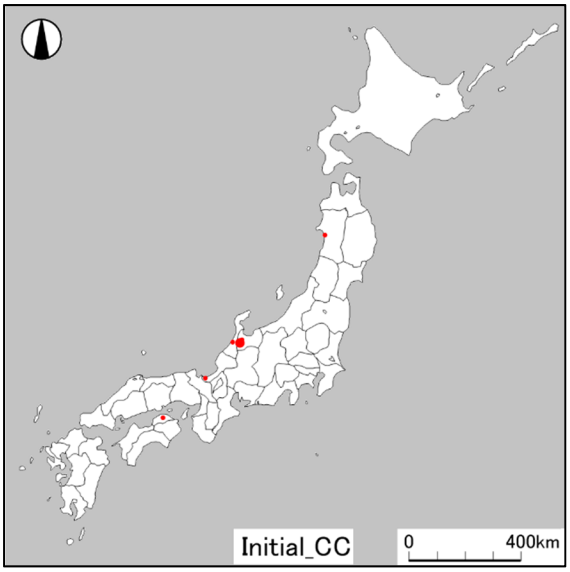


Figure 5: The frequency of the CC type at the beginning of a word

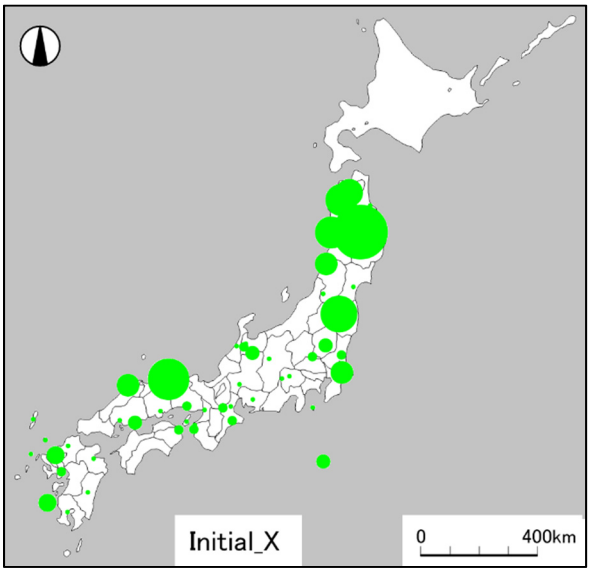


Figure 6: The frequency of the X type at the beginning of a word

4.2. Syllable fusion in the middle of a word

In the middle of the word, the VV type is found mostly in the Kantō region, and it can also be found in the Chūgoku region.

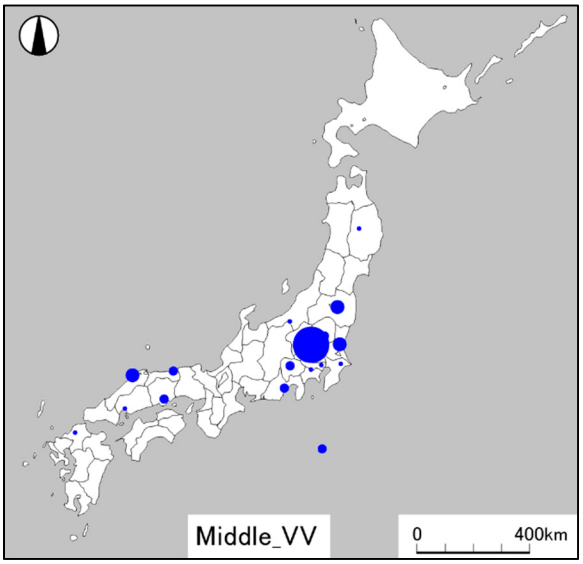


Figure 7: The frequency of the VV type in the middle of a word

The CC type is found in the middle of a word in a wide range of areas, and it is especially common in Shikoku.

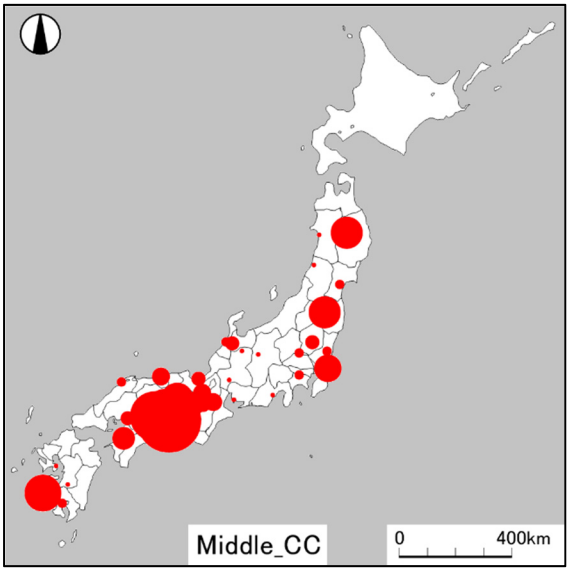


Figure 8: The frequency of the CC type in the middle of a word

The X type is also widely distributed in the medial position of a word.

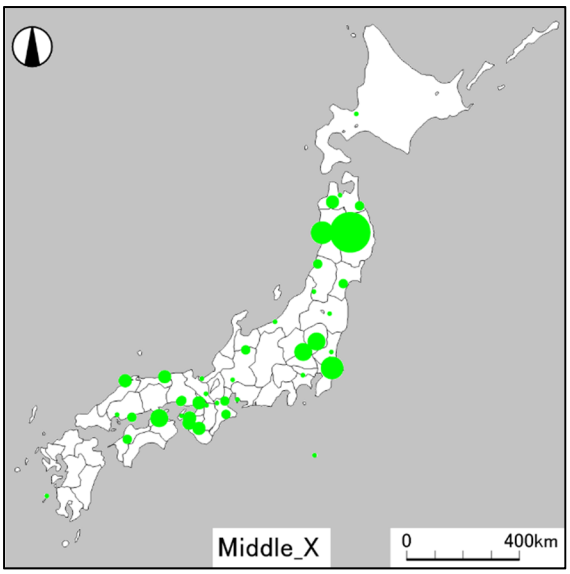


Figure 9: The frequency of the X type in the middle of a word

Figure 10 is a quantitative comparison of the VV and CC types in the medial position of a word.

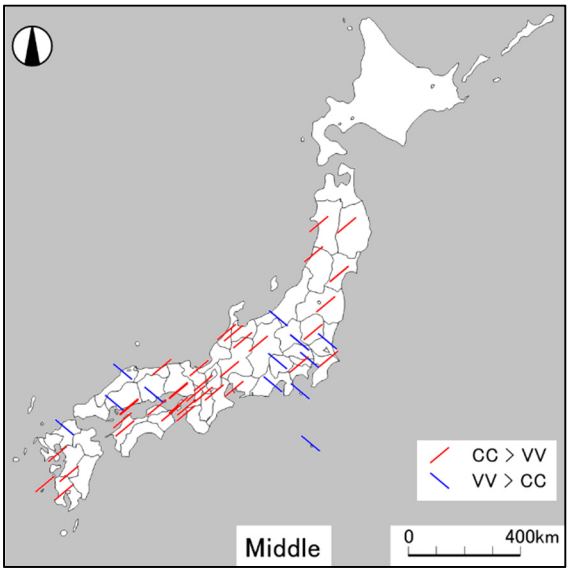


Figure 10: VV type vs. CC type in the middle of a word

The VV type is prevalent in the Kantō, Chūbu, and Chūgoku regions, while the CC type is predominant in other regions.

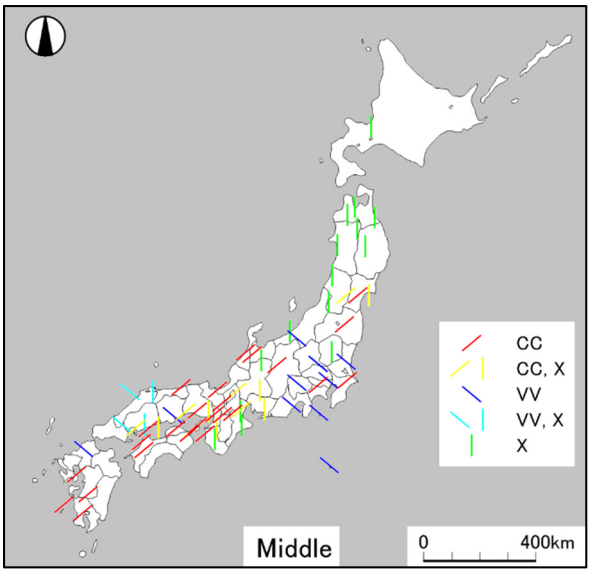


Figure 11: The distribution of VV, CC, and X types in the middle of a word

Figure 11 shows the distribution of the VV, CC, and X types. In the Tōhoku region, the X type is the most common. The addition of the X type to the map does not significantly change the dominance of the VV and CC types; however, it shows that the numbers of the VV and X types are almost the same in the Chūgoku region.

5. Discussion

5.1. VV type distribution

The VV type is prevalent in the Kantō, Chūbu, and Chūgoku regions, while the CC type is predominant in the Hokuriku, Kinki, and Shikoku regions. In the Tōhoku region, the X type predominates. Because the distinction between long and short vowels is obscure in Tōhoku dialects (cf. Sibata 1968), the X type in this region may be derived from the *VV type.

Comparing Figures 4 and 7, the VV type is more widely found at the beginning of a word; however, in both cases, the distribution is concentrated in the Kantō and

Chūgoku regions. Although the VV type was once dominant in Tōhoku dialects, the VV type had a wider geographic distribution, and it is presumed to be older.

5.2. Historical phonology from syllable fusion

In the Tōhoku and Chūgoku regions, both the (*)VV and CC types are observed. The reason for the presence of both types in these regions is not entirely clear; however, it may be attributed to the existence of the central vowel [ɨ] in these dialects and the frequent loss of /r/. In the CVrV condition, the change to the CC type occurs in various dialects, including those in the Tōhoku and Chūgoku(-San'in) regions, such as *hasira* > *hassja*, *usiro* > *ussjo* 'back' and *musiro* > *mussjo* 'mat.' Thus, /r/ exhibits a very unique behavior in Japonic. This change, along with the loss of /r/, has also led to CC-type changes in the CVSV condition.

It is unclear whether the central vowel [ɨ] or /r/ directly causes the sound change; however, because the CC type is also found in Kyūshū, it is probably the /r/ that is directly related.

The central vowel [ɨ] seems to cause the drop (or assimilation) of /r/ (cf. Hirako 2018), as, phonologically, /r/ seems to behave as a semi-vowel corresponding to the central vowel [ɨ].

The phonological characteristic of /r/ in Japonic is disputed (see Pellard 2016 and Labrune 2017); however, syllable fusion suggests that /r/ is a unique consonant in Japanese.

6. Conclusion

In syllable fusion, regional variation is observed in Japanese dialects. The VV type is dominant at the beginning of a word, whereas both the CC and VV types are found in the medial position of a word. Based on geographical distribution, the VV type is considered to be older than the CC type.

7. Further Investigation

There are two issues to be addressed in the future: First, syllable fusions other than $CV_{[+high]}(S)V_{[-high]}$ should be analyzed, especially those involving /r/. The second is to analyze syllable fusions in Ryukyuan languages. Because Ryukyuan forms are not listed in the *namari* of *Nihon Kokugo Daijiten* 2nd edition, another method must be considered.

Note

The free software, MANDARA version 10ex (https://ktgis.net/mandara/), was used to draw the maps.

Symbols

C: a consonant. V: a vowel. S: a semivowel. /c/: [ts] ~ [tʃ]

References

Hirako, Tatsuya [平子達也] (2018) Izumo Nita hōgen no boin o meguru onhenka nitsuite「出雲仁多方言の母音をめぐる音変化について」[Historical study on the vowels in Izumo-Nita Japanese]. The 157th Meeting of the Linguistic Society of Japan.

Labrune, Laurence (2017) More on Japanese /r/: A response to Pellard (2016). *Journal of East Asian Linguistics* 26(3): 301–321. https://doi.org/10.1007/s10831-017-9157-x

Nakazawa, Kohei [中澤光平] (2021) Awaji hōgen no onsetsu yūgō no kyōjiteki, tsūjiteki bunseki「淡路方言の音節融合の共時的・通時的分析」[A Synchronic and Diachronic Analysis of Syllable Fusion in the Awaji Dialects]. *University linguistic papers* 43: 159–182. https://doi.org/10.15083/0002002771

Pellard, Thomas (2016) Why /r/ is not a special, empty consonant in Japanese. *Journal of East Asian Linguistics* 25(4): 351–383. https://doi.org/10.1007/s10831-016-9147-4

Sibata, Takesi [柴田武] (1968) On'in「音韻」[Phonology]. In: Kokugogakkai (ed.) *Hōgengaku gaisetsu*『方言学概説』[*Overview of dialectology*]. 137–161. Tokyo: Musashino Shobo.

Uwano, Zendo [上野善道] (ed.) (1989) Nihon hōgen on'in sōran「日本方言音韻総覧」[Phonological overview of Japanese dialects] In: Tokugawa Munemasa [徳川宗賢] (supervision) *Nihon hōgen daijiten*『日本方言大辞典』[*Great Dictionary of Japanese Dialects*] 下巻: 1–77. Tokyo: Shogakukan.

How rip weaving spread in Japan: Interpreting maps of words and referents

Chitsuko Fukushima
(University of Niigata Prefecture)

Abstract: Rip weaving is a way of weaving a fabric using rip cloth as crosswise yarn on a loom. The textile spread during the Edo Period (1603-1868) after cotton was introduced to Japan. The old cotton cloths were used to make a typical *sakiori* fabric. There are three groups of words denoting rip weaving: *sakiori*, *cuzure*, and *boro*. *Sakiori* literally means rip weaving (*saki* "rip" + *ori* "weaving"), while *cuzure* and *boro* mean "rags or tattered clothes". Interestingly the referents of *sakiori* extended to 1) traditional fabric which is woven using hemp (*asa*), Japanese wisteria (*fuji*), false nettle (*karamushi*), etc., 2) newer fabric which is woven using cotton threads as crosswise and lengthwise threads, and 3) work clothes in general, which are not always made of *sakiori* fabric. This might be caused by the phonetic changes of *sakiori* words (*sakkyori*, *sakkori*, *sakkuri*, *zakkuri* etc.). On the other hand, *cuzure* is used for clothes made by rip weaving in some areas while it is used for clothes made by quilting (*sashiko*) in other areas. *Boro* is mainly used for sashes made by rip weaving. The distributions of the forms and the referents show the variation and change of words related with rip weaving.
Key words: geographical distributions, rip weaving, workwear, referents

1. Introduction: what is *sakiori* 'rip weaving'?

Sakiori "rip weaving" is a way of weaving a fabric using ripped cloths (usually cotton) as crosswise yarn (weft) on a loom. Before cotton was introduced to Japan, traditional plants such as *asa* "hemp", *karamusi* "false nettle", *fuji* "Japanese wisteria", *koozo* "a kind of papyrifera", etc. were used for weaving a fabric.

Cotton-growing was established in the 15th to 16th century in some specific areas of Japan. Cotton was planted only in the southern warmer area and not planted in the northern colder area. Thus, cotton wool and old cotton cloths were transported to the

FUKUSHIMA, Chitsuko. 2023. How rip weaving spread in Japan: Interpreting maps of words and referents. In Cẩm Lan Trịnh, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference of Asian Geolinguistics*, 86–96. doi: https://doi.org/10.5281/zenodo.8374597

* This work was supported by JSPS KAKENHI Grant Number JP19K00555. I am also grateful to the NINJAL Project, Spatial Information Addition to Language Resource, which gave me an opportunity to discuss the related research results.

latter areas by ship during the Edo Period (1603-1868). Around the same time, *sakiori*, the textile made by rip weaving, also spread.

Cotton cloths were valuable as they were warm and strong. Old cotton cloths were useful since they were used to remake fabric and clothes. One way of making use of old cotton cloths is rip weaving (to cut cloths into rips and weave a fabric by using the rips as crosswise yarn on a loom), and another way is quilting (to lay cloth on top of another and stitch them). You can also patch clothes using cotton cloths. In any of these ways, reproduced clothes are warmer and stronger.

2.Three words denoting *sakiori*

It has been found out that there are three words which denote *sakiori* 'rip weaving' (Nakamura 1984; Fukushima 2023a). First, the word *sakiori*, literally meaning "rip weaving", is a term for the fabric woven using ripped cloths. Second, the word *cuzure*, which originally meant "patchwork" or "patchworked clothes", is not a term for the fabric originally. Third, the word *boro*, whose meaning is "rag" or "tatters", is not a term for the fabric at all.

Fig. 1 shows the geographical distributions of these three words, which include cases in which the referent of the words is not the fabric woven using ripped cloths. Here I examine how the distributions of the forms and the referents have been produced.

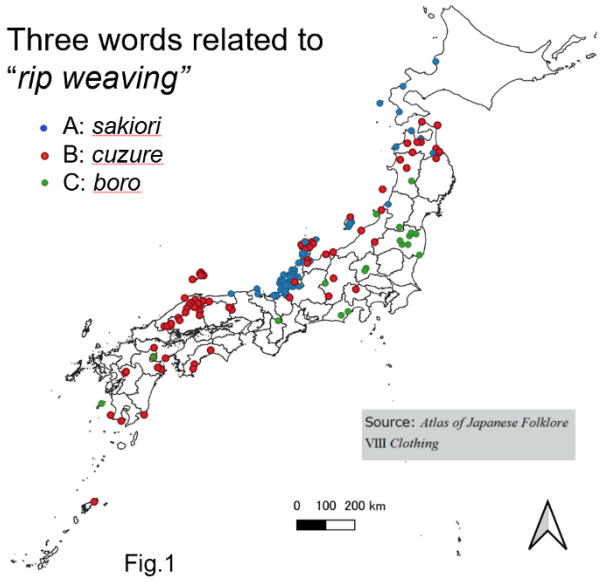


Figure 1: Geographical distributions of three words related to *sakiori* "rip weaving"

All the data for maps in this paper is taken from the *Atlas of Japanese Folklore VIII Clothing* (except the last map **Fig. 7**). **Figs. 2-3** are made using ArcGIS online, and the other figures, **Figs. 1** and **4-6**, using QGIS. See the photos at the end of this paper to see samples of *saguri,* a workwear and *boroobi*, a sash made by rip weaving.

2.1. The word *sakiori*

2.1.1. The geographical distributions of the forms of *sakiori*

Fig. 2 is a map of the forms of *sakiori*. The forms are distributed on the coast of the Sea of Japan. There are two groups of forms: the -o- group such as *sakkiyori* and *sakkori*, the -u- group such as *sakkuri,* and the -i- group such as *sakkiri*. Obviously, the form *sakiori* changed to the -o- forms, and to the -u- forms, and then to the -i- forms. Each group has a variation caused by the changes such as *sa* > *sha,* sa > za > zja, and k > g. Thus, the following changes are inferred.

sakiori > *sakkiyori* > *sakkori* > *sakkuri* > *sakkiri*
> *zakkuri*
> *sakuri* > *shakuri* > *zjakuri*
> *saguri* > *zaguri*
> *shakkuri* > *shakkiri*

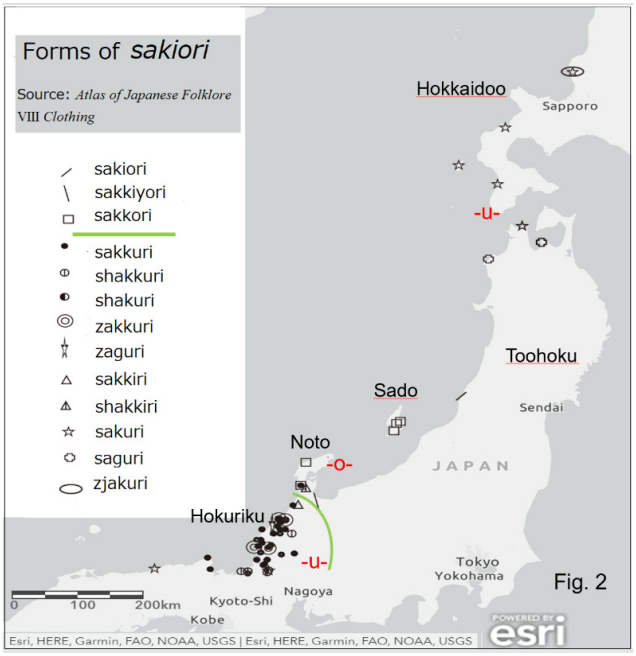


Figure 2: Geographical distributions of the forms of *sakiori*

These changes occurred in the Hokuriku area basically. The -u- forms such as *sakuri*, *saguri*, and *zjakuri* distributed in the northern Japan were transported from Hokuriku with the shipping route along the Sea of Japan.

2.1.2. The geographical distributions of the referents of *sakiori*

Fig. 3 is a map of the referents of *sakiori*. The original referent should be the textile, rip weaving, but it was expanded to the workwear (not specified to that made of rip weaving), men's workwear, or women's workwear. Again, these changes occurred in the Hokuriku area and spread to the northern Japan with the same route. Clothes made by rip weaving are warm, strong, and waterproof; the term was maybe expanded by people involved in farming, fishery, and forestry.

Figure 3: Geographical distributions of the referents of *sakiori*

2.2. The word *cuzure*

2.2.1. The geographical distributions of the forms of *cuzure*

Fig. 4 is a map of the forms of *cuzure*. The forms are distributed more broadly compared with *sakiori*. In the northern Japan, they are found along the coast of the Sea

of Japan. In the center of Japan, they are found in the mountainous area and also densely distributed in Noto. In the western Japan, the distributions are dense along the Chuugoku mountainous region, scattered in Shikoku and Kyuushuu. The variation of forms is mostly phonological; the forms in Noto such as *cuuri* or *cuure* are peculiar.

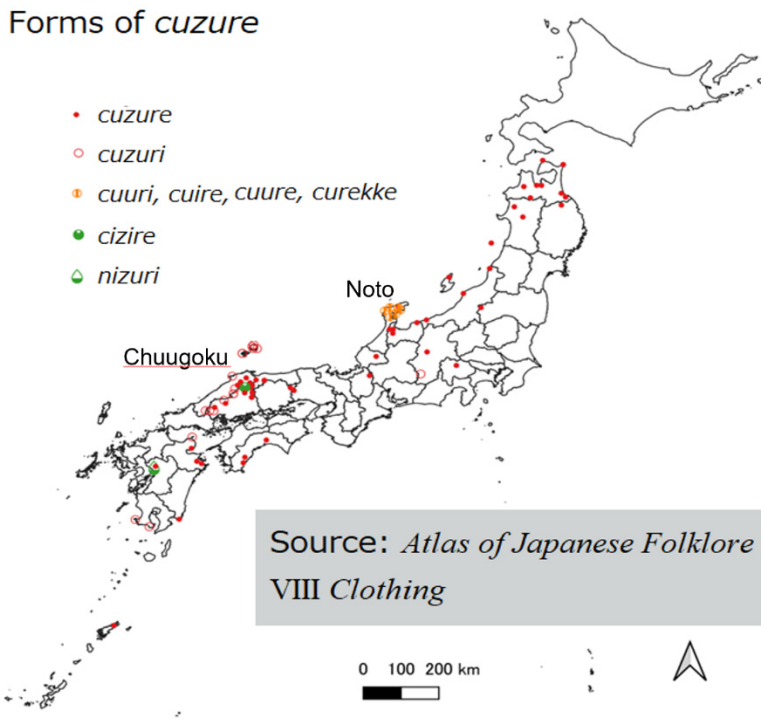


Figure 4: Geographical distributions of the forms of *cuzure*

2.2.2. The geographical distributions of the referents of *cuzure*

Fig. 5 is a map of the referents of *cuzure*. While the original referent, "patched clothes", is found peripherally, "clothes made by quilting" is found in the broad area from the north to the south, and "clothes by rip weaving" is distributed in Echigo, Noto, and Chuugoku surrounded by "clothes made by quilting"; therefore, it is considered that "clothes by rip weaving" is newer than "clothes made by quilting". As a variation, "sash made by rip weaving" and "clothes made by paper as weft" is found in some localities. The referent of *cuzure* has also expanded to the workwear where the word is used to denote some kind of clothes, which is similar to *sakiori*.

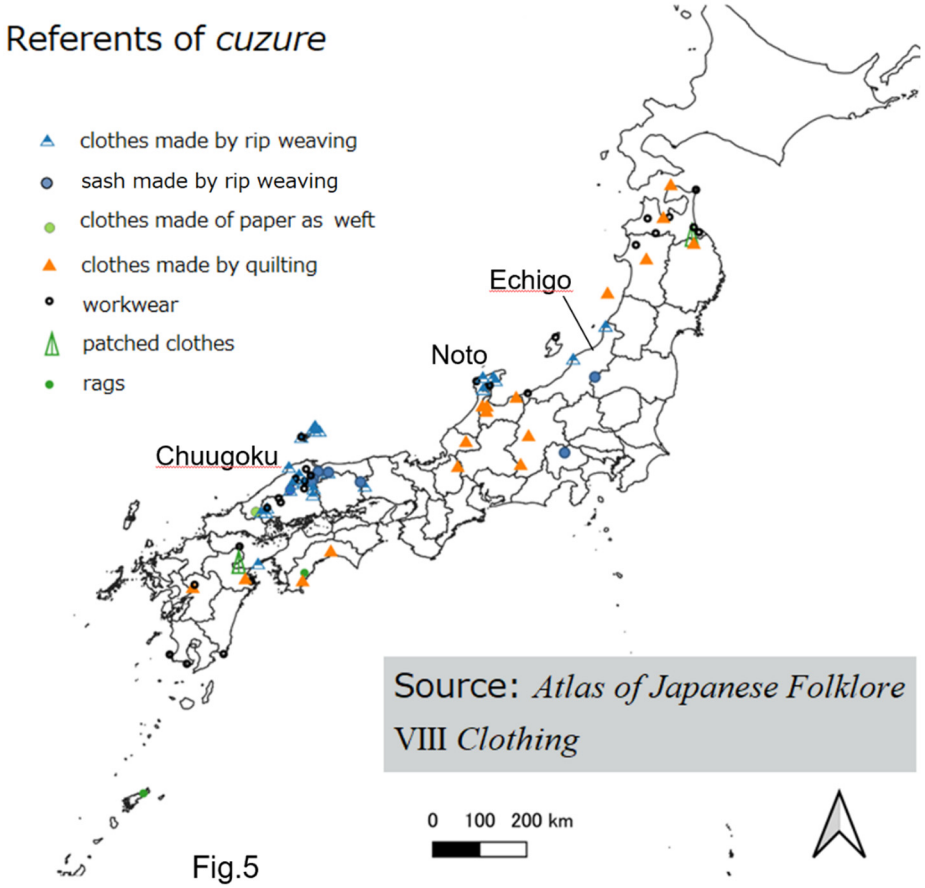


Figure 5: Geographical distributions of the referents of *cuzure*

2.3. The word *boro*

See **Fig. 1**. *Boro* is often used as *boroobi*, "sash made by rip weaving", in the Eastern Japan. The distribution of *boro* is close to the distribution of *cuzure* as "sash made by rip weaving" in **Fig. 5**. Rip weaving is used for making sash in this area using tattered old clothes.

3. The varieties of *sakiori*

Nakamura (1984) classified *sakiori* "rip weaving" into three categories by focusing on the materials. See **Table 1**. In addition to two typical categories using ripped cotton cloth as weft, "*sakiori* made of cotton cloth" and "*sakiori* made of cotton thread", there is another category whose warp and weft are traditional plants such as hemp, thus "*sakiori* made of hemp". This category is called *sakiori* although it is not rip weaving

at all. Otsuka et al. (2003) called these categories as *sakiori* B, C and A, respectively. I added one more category, which is made of cotton thread as both warp and weft (I call this *sakiori* D). Using cotton thread is considered new since it is manually made from cotton wool or it is made at the factory. I also added two more categories as varieties of *sakiori* A. One is *sakiori* A' whose weft is paper. Another is *sakiori* A" whose warp is cotton thread.

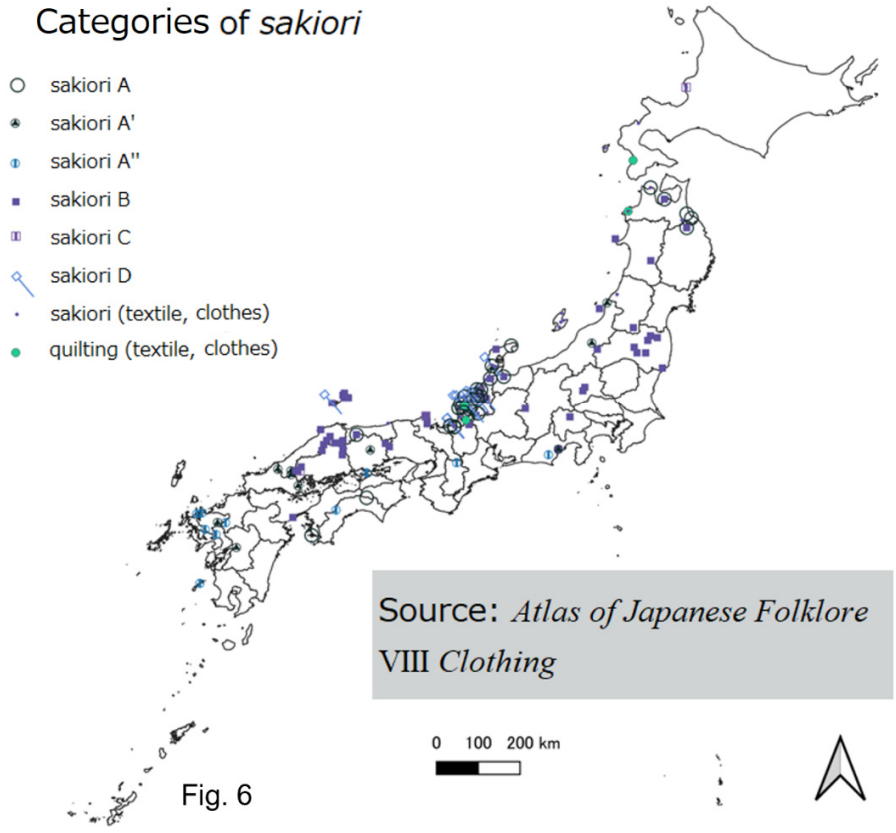
Table 1: Categories of *sakiori*

Nakamura 1984	lengthwise thread/yarn (warp)	crosswise thread/yarn (weft)	**Otsuka 2003**	**Fukushima 2023**
sakiori made of hemp	hemp, Japanese wisteria etc.	hemp, hemp cotton, Japanese wisteria, nettle, kudzu vine etc.	*sakiori* A	*sakiori* A
				sakiori A' (weft: paper)
				sakiori A" (warp: cotton thread)
sakiori made of cotton cloth	hemp, Japanese wisteria etc.	rips cut from cotton cloths	*sakiori* B	*sakiori* B
sakiori made of cotton thread	cotton thread	rips cut from cotton cloths	*sakiori* C	*sakiori* C
-	cotton thread	cotton thread	-	*sakiori* D

I made a map of *sakiori* categories based on my classification. See **Fig. 6**. *Sakiori* A is found in peripheries such as northern Tohoku and Chuugoku, but also in Hokuriku. *Sakiori* A' is rare but considered as traditional. *Sakiori* A'' is found in the southern area and considered as new. *Sakiori* B is found in the broad area including Tohoku, Hokuriku, and Chuugoku. *Sakiori* D is rare but found in Hokuriku and Oki. It is interesting that both new and traditional varieties are found in Hokuriku.

I found a map of *sakiori* materials in the Hokuriku area where there are varieties of *sakiori* words and referents. See **Fig. 7**. This is a map of materials of *sakkuri* (<*sakiori*) reproduced from Fig. 4-2 in Yamazaki 1995. Ripped cloth is used in two parts, Wakasa and Noto. In between, in Kaga and Echizen, both traditional plants and cotton threads are used. In the latter area, cotton was not been planted and cotton cloths were transported from outside (cf. my research on the data of the *Atlas of Japanese Folklore VIII Folklore*). The making of the typical rip weaving spread from Wakasa to Noto, but not to Kaga and Echizen. In the latter areas, old materials (hemp, Japanese

wisteria, papyrifera, etc.) and new materials (cotton cloth and cotton thread) are both used.



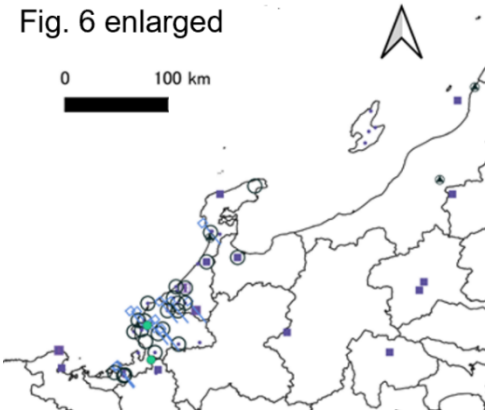


Figure 6: Categories of *sakiori*

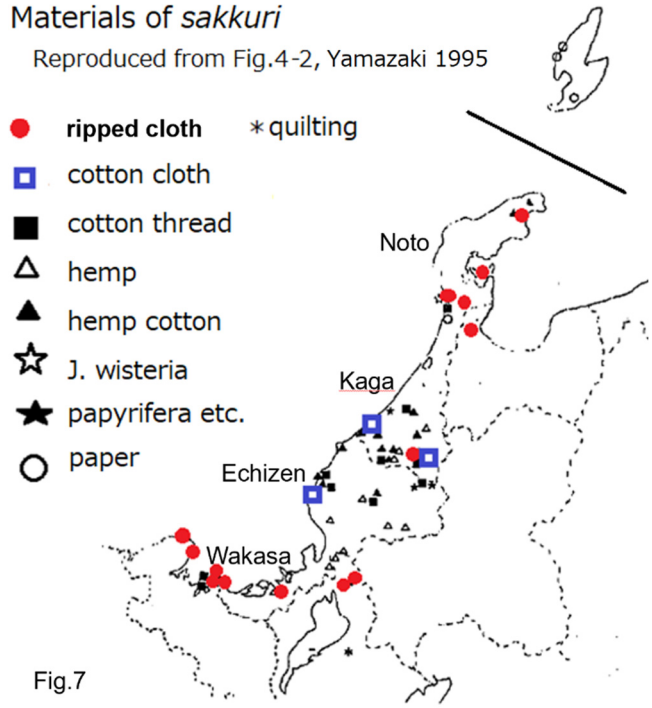


Figure 7: Materials of *sakkuri* in the Hokuriku area

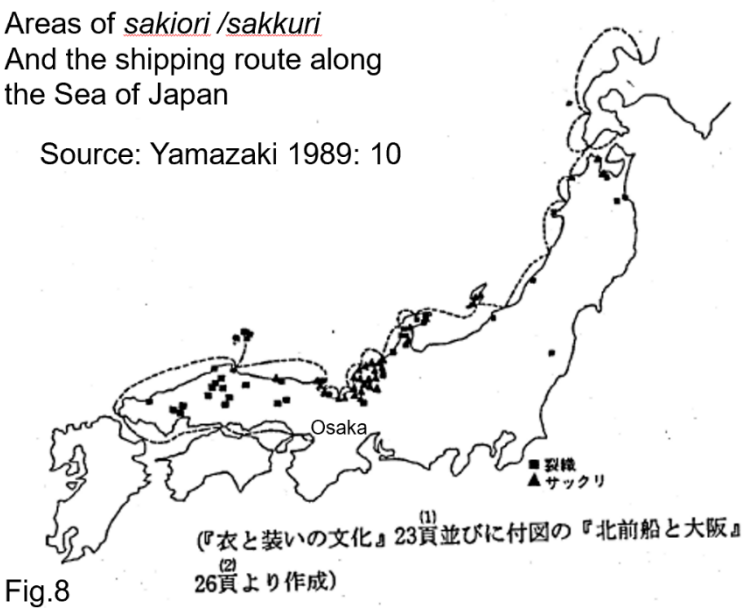


Figure 8: Areas of *sakiori/sakkuri* and the shipping route along the Sea of Japan

4. Conclusion

Rip weaving has been especially useful for people living in the colder area. The making of the textile spread with old cotton cloths transported on the shipping route along the coast of the Sea of Japan shown in **Fig. 8** (Yamazaki 1989).

The brief history of words related with *sakiori* is summarized as below. First, in the Hokuriku area, the referent of the word *sakiori* expanded to 1) traditional fabric which is woven using hemp (*asa*), Japanese wisteria (*fuji*), false nettle (*karamushi*), etc., that is, sakiori A, 2) newer fabric which is woven using cotton threads as crosswise and lengthwise threads, that is, sakiori C or D, and 3) workwear in general maybe because the wordforms changed to *sakkuri* etc. and were no more transparent. As a result, the words were used to denote various referents, and then the words spread to the North along the shipping route.

As for *cuzure*, in the Chuugoku area, Noto, and Sado, the word *cuzure* was used to denote the workwear made of rip weaving. Maybe the textile, rip weaving, replaced the older material and inherited the older name for the textile or workwear. This change should have occurred before the words originating from *sakiori* spread.

As for *boro*, in part of the Eastern Japan, rip weaving was applied to make sashes and *boro* was used to denote it. *Boro* was a term for the old clothes used to make sashes. This usage could have occurred independently in separate areas.

References

Bunka-cho [文化庁] (ed.) (1992) *Nihon Minzoku Chizu*『日本民俗地図』[*Atlas of Japanese Folklore*] VIII *Iseikatsu* 衣生活 [*Clothing*]. Bunka-cho.

Fukushima, Chitsuko [福嶋秩子] (2022) Sakkuri no bunpuzu o yomu「サックリの分布図を読む」[Reading maps of *sakkuri*]. Presented at the 2022 annual meeting of the Society for Life Culture of Niigata.

Fukushima, Chitsuko [福嶋秩子] (2023a) Sakiori / sakkuri no bunpuzu o yomu「裂き織り・サックリの分布図を読む」[Reading maps of rip weaving & *sakkuri*]. Presented at the meeting of the NINJAL Project: Spatial Information Addition to Language Resource.

Fukushima, Chitsuko [福嶋秩子] (2023b) Sakkuri no bunpuzu o yomu「サックリの分布図を読む」[Reading maps of *sakkuri*]. *Niigata no seikatsu bunka*『新潟の生活文化』[*Bulletin of Society for Life Culture of Niigata*] 29, 7-10.

Nakamura, Hiroko [中村ひろこ] (1984). Sakiori no bunpu to bunrui narabini sono keifu「サキオリの分布と分類ならびにその系譜」[Distributions, classification, and genealogy of *sakiori*] *Gekkan Senshoku α*『月刊染織α』[*Monthly Textiles α*] 34, 15-17.

Otsuka, Kohei et al. [大塚康平・植田憲・宮崎清・朴燦一] (2003) Shizen junkan-gata bunka—sakiori ni shiyoo sareru kofu no ryuutsuu kioo: shigen junkan-gata zookei bunka to shiteno sakiori (1)「自然循環型文化・裂き織りに使用される古布の流通機構ー資源循環型造形文化としての裂き織り(1)」[Distribution system of cotton cloths used for *sakiori*, resource recycling culture: *sakiori* as resource recycling art culture]' *Dezain-gaku kenkyuu*『デザイン学研究』 [*Study of Design*] 50(2), 53-62.

Rias Ark Museum of Art [リアスアーク美術館] (ed.) (1999) *Tami no matoishi koromo: kurashi ga unda toohoku no nuno*『民の纏いし衣 : 暮らしが生んだ東北の布』[*Clothes Worn by People: Cloth of Tohoku Born in the Life of Tohoku*]. Rias Ark Museum of Art.

Yamazaki, Mitsuko [山崎光子] (1989) Sakiori no tootta nihonkai no michi: echigo no cuzure to noto no cuzure 「サックリの通った日本海の道：越後のツヅレと能登のツヅレ」[The way of sakiori in the Sea of Japan: cuzure in Echigo and cuzure in Noto]' in Toyama-shi/Toyama-shi kyooiku iinkai/Nihon bunka o kangaeru Toyama shinpojiumu jikkoo iinkai [日本海を考える富山シンポジウム実行委員会] (eds.) *Nihonkai bunka kenkyuu ronbunshuu*『日本海文化研究』[*Collected Papers on Studies of Culture along the Sea of Japan*], 9-35.

Yamazaki, Mitsuko [山崎光子] (1995) Sakkuri bunka no hen'yō: mikuni o chūshin to suru sakkuri no bunpu to keifu kara mita「サックリ文化の変容：三国を中心としたサックリの分布と系譜からみた」 [Changing *sakkuri* culture: Viewed from distributions and genealogy of *sakkuri* around Mikuni]. *Niigata no seikacu bunka*『新潟の生活文化』[*Bulletin of Society for Culture in Niigata*] 2: 20-25.

Photos:

Saguri, a workwear
Source: Rias Ark Museum of Art 1999: 27

Boroobi, a sash made by rip weaving
Source: Rias Ark Museum of Art 1999: 30

Classification and interpretation of tone sandhi in three-tone Jiaoliao Mandarin

ZENG Xiaoyu
(Nankai University, China)

Abstract: Jiaoliao Mandarin is one of the dialects of Mandarin Chinese that coexists with three-tone and four-tone systems. The three-tone system was created by combining the main four traditional tone classes of Mandarin Chinese. This study focuses on the representative points of the three-tone dialect in Jiaoliao Mandarin and analyzes three categories of tone sandhi in the three-tone system. Additionally, the investigation suggests that these three categories reflect the different time stages of tone formation and development in the three-tone system.

Keywords: Jiaoliao Mandarin, three-tone system, four-tone system, tone sandhi, the categories of tone sandhi

1. Introduction

Jiaoliao Mandarin Chinese is primarily spoken in the Jiaodong Peninsula of Shandong Province and the Liaodong Peninsula of Liaoning Province (*Atlas of Chinese Languages (2nd Edition): Volume on Chinese Dialects*, 2012: B-14). Jiaoliao Mandarin in the Liaodong Peninsula originated from the Jiaodong Peninsula (Zhang, 2007; Luo, 2010), as shown in Figure 1.[1]

The Jiaoliao Mandarin dialects have between four (the mainstream) to three tones, with each of its subdialects (Qinglai, Denglian, and Gaihuan) having its three-tone varieties. Jiaoliao Mandarin has been shifting from a four-tone to a three-tone system in recent decades, and it is interesting to investigate the time when the three-tone dialects were created.

ZENG, Xiaoyu. 2023. Classification and interpretation of tone sandhi in three-tone Jiaoliao Mandarin. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 97–112. doi: https://doi.org/10.5281/zenodo.8374609

* This research was funded by the National Social Science Foundation Project of China (20AYY020). Thank you to the PhD student, Zeng Zhichao, for assisting with the English translation.

[1] The language mapping software developed by Nankai University and Tianjin Xinhui Network Technology Service Center is used in all the language maps listed in this paper; the base map is based on Gaode map (©2023AutoNavi—GS (2022)1061).

Based on the investigations of Qian (2001), Zhang (2000), Gao and Sun (2021), Liu (2021), and this article, the distribution of three-tone dialect points in Jiaoliao Mandarin is shown in Figure 2.

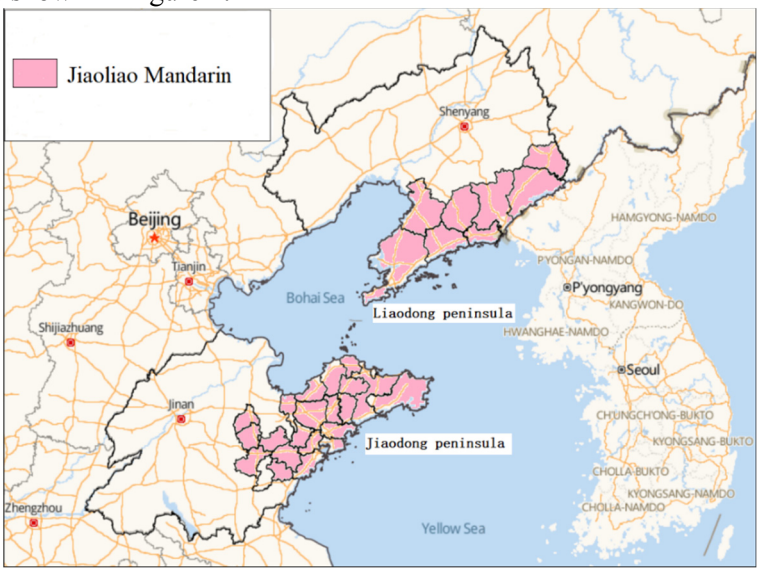


Figure 1: The distribution of Jiaoliao Mandarin.

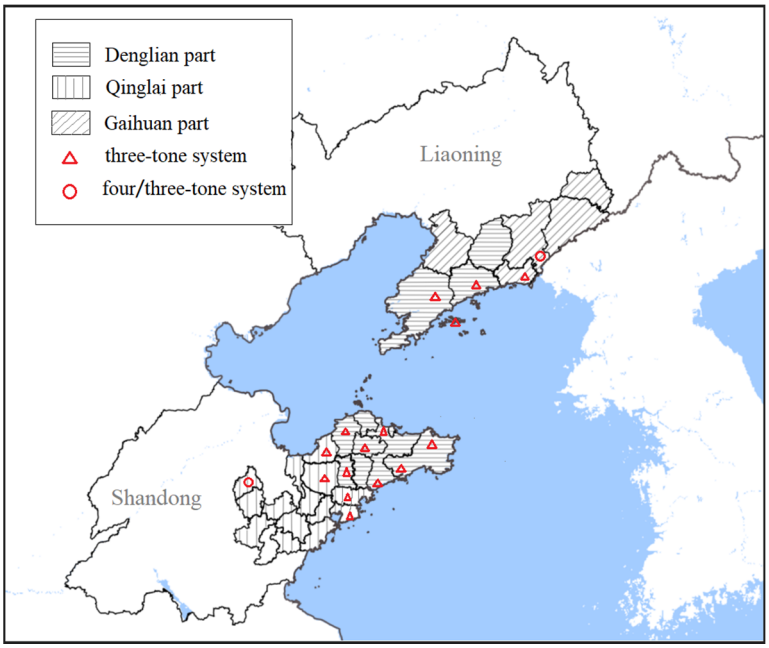


Figure 2: Jiaoliao classification and the number of tones.

2. The single-character tones and compound-character tones of the representative points of dialects in Jiaoliao Mandarin

According to Li (2004), there are two categories of disyllabic tones "tone changes due to phonetic variation" and "tone changes due to semantic variation (including neutral tones, and so on).

There is a difference between coarticulation and tone sandhi in disyllabic tones. The main difference between these two categories is that the tone sandhi can be recognized by the human ear, can traverse tonal categories, and is a manifestation of language peculiarity. Coarticulation often can not be recognized by the human ear due to physiological mechanisms of pronunciation and is universal to all languages (Wang, 2002).

Chinese has a dominant number of disyllabic words, which is the core domain of tone sandhi. This study focuses on the tone sandhi of disyllabic words in the representative dialects of Jiaoliao Mandarin, including Qingzhou, Qingdao, Laizhou, Yantai, Dalian, and Dandong.

2.1. Qingzhou dialect

2.1.1 Qingzhou urban dialect

The urban dialect of Qingzhou uses a four-tone system. Table 1 displays the tonal variations for the single-character tones and combined character tones (the boxed tone values in the table represent tone sandhi, the same applies to subsequent cases).

Table 1: Tones of the urban dialect of Qingzhou.

Citation tones	Disyllabic tones	
	Tone sandhi without neutral tone	Tone sandhi with neutral tone
①213 ②42 ③55 ④21 （Qian, 2001:100）	213+213→[24]+213 55+213→[24]+213 55+42→[435]+42 55+55→[42]+55	（No record）
①213 ②53 ③55 ④41 （Liu, 2021）	213+213→[35]+213 airplane 飞机 53+41→[45]+31 farm tools 农具 55+53→[435]+54 emotion 感情 55+55→[43]+45 factory manager 厂长	①+0→[31]·[11] knife 刀子 ②+0→[35]·[44] bug 虫子 ③+0→[213]·[44] plank 板子 ④+0→[45]·[21] trousers 裤子

Note: ①Yin; ②Yang; ③Shang; ④Qu; “0” refers to a neutral tone.

Although the two articles of Qian (200) and Liu (2021) were written twenty years apart, the patterns of single-character tones and the rules for tone sandhi in combined characters are largely comparable.

2.1.2 Qiujiayu dialect of Qingzhou (four/three-tone system)

Liu (2021) reported that there are differences in tone types amongst various age groups in Qiujiayu Village, which is located in the southwest of Qingzhou City. The single-character tones and combined-character tones of the Qiujiayu dialect for each age group are shown in Table 2.

Table 2: Tone variations in the Qiujiayu dialect of Qingzhou according to age.

	Citation tones	Disyllabic tones	
		Tone sandhi without neutral tone	Tone sandhi with neutral tone
Elderly group	①214 ②53 ③55 ④31	214+214→55+214 airplane 飞机 55+55→54+55 factory manager 厂长	①+0→31+22 ideas 心思 ②+0→24+54 mustache 胡子 ③+0→213+44 plank 板子 ④+0→55+31 case 案子
Middle-aged group	① 214 ② 54/55 ③55 ④31	214+214→55+214 airplane 飞机 55+55→54+55 factory manager 厂长	
Young group	① 214 ② ③ 55 ④31	214+214→55+214 airplane 飞机 55+55→54+55 the end of the yea 年底 55+55→44+55 factory manager 厂长	

Table 2 demonstrates that:

(1) Citation tones are in the transitional period from the four-tone to the three-tone system;

(2) The reason for merging ②Yang into ③Shang is that their pitch values are similar;

(3) There is a distinct difference in tonal categories on the syllable before the neutral tone regardless of whether the system is four or three tones, in all age groups.

2.2 Laizhou dialect

The Laizhou dialect is classified under the Qinglai subdialects of Jiaoliao Mandarin, and its tone system belongs to the three-tone type (the Qu tone merging with the Yang tone)[2]. The single-character tones and combined-character tones of the Laizhou dialect are shown in Table 3.

[2] In the Laizhou dialect, a small portion of the Qu tone characters are pronounced as Yin tone, which is mainly a result of language contact. Laizhou is closely adjacent to the Donglai dialect

Table 3: Tones of the Laizhou dialect.

Citation tones	Disyllabic tones	
	Tone sandhi without neutral tone	Tone sandhi with neutral tone
①213 ②④42 ③55 (Qian, Ota 2005:6)	213+213→45+213 cold 伤风 42+42→213+42 brown sugar 红糖 42+55→213+55 first 第一 55+42→213+42 buffalo 水牛 55+55→213+55 watches 手表 55+55→42+55 lotus root starch 藕粉	①+0→213+42 clothes 衣裳 ①+0→42+2 lively 热闹 ②+0→55+3 stone 石头 ④+0→42+2 uncle 大爷 ③+0→45+3 ear 耳朵 ③+0→42+2 dress up 打扮
①323 ②④51 ③44 (Yang, 2012)	323+323→44+323 airplane 飞机 51+51→323+51 wool 羊毛 51+55→323+44 government 政府 55+55→323+44 entertain guests 请客 55+55→51+44 factory manager 厂长	①+0→323+43 brick 砖头 ①+0→51+21 date 日子 ②+0→44+31 stone 石头 ④+0→51+21 can 罐头 ③+0→34+21 inside 里头

Table 3 indicates that:

(1) For more than a decade, the Laizhou dialect's citation tone pattern and tone sandhi rules have stayed constant;

(2) There is a distinct difference between merged Yang and Qu tones before the neutral tone;

(3) The original ④Qu tone (now pronounced as Yang and Yin) has the same pronunciation as the neutral tone, and its pitch value reflects the original feature falling.

2.3 Qingdao dialect

The Qingdao dialect is classified under the Qinglai subdialects of Jiaoliao Mandarin, and its tone system belongs to the three-tone type (the Qu tone merging with the Yang tone). The single-character tones and combined-character tones of the Qingdao dialect are shown in Table 4.

of Shandong, where approximately 42% of the Middle Chinese voiced sonorant characters in the three-tone dialect are pronounced as Yin tone, while the rest are pronounced as Qu tone (see Yang 2012 for more details).

Table 4: The tones of the Qingdao dialect.

Citation tones	Disyllabic tones	
	Tone sandhi without neutral tone	Tone sandhi with neutral tone
①213 ②④42 ③55 (Li, 1999:9-11)	213+213→55+213 airplane 飞机 42+42→31+42 moneygrubber 财迷 be bullied 受气 42+55→31+55 Tuan island 团岛 diving 跳水 55+55→31+55 island 海岛	①+0→31+2 knife 刀子 ②+0→55+4 pliers 钳子 ④+0→42+2 wood 木头 ③+0→55+4 dress up 打扮 ③+0→45+5 deal 买卖
①213 ②④42 ③55 (Qian, 2001:91, 98, 109)	213+213→55+213 42+42→31+42 42+55→31+55 55+55→31+55	①+0→31+· ②+0→55+· ④+0→42+· ③+0→55+· ③+0→434+·

Table 4 reveals that:

(1) There is a distinct difference between merged ②Yang and ④Qu tones before the neutral tone;

(2) The original ④Qu tone reflects the original falling feature before the neutral tone.

2.4 Yantai dialect

The Yantai dialect is classified under the Denglian subdialects of Jiaoliao Mandarin, and its tone system belongs to the three-tone type (the Yang tone merging with the Qu tone). The single-character tones and combined-character tones of the Yantai dialect are shown in Table 5.

Table 5: The tones of the Yantai dialect.

Citation tones	Disyllabic tones	
	Tone sandhi without neutral tone	Tone sandhi with neutral tone
①31 ③214 ②④去 55 (Qian, 1982:15, 66-80)	31+31→35+31 hillside 山坡 214+31→35+31 palm 手心 214+214→55+214 wheat 小麦 55+55→31+55 bark 树皮 labour 劳力	①+0→31+21 box 箱子 ③+0→214+55 tail 尾巴 ②+0→55+21 snake 长虫 ④+0→55+21 joke 笑话 occasionally: ③+0→55+31 dimple 酒窝

①31 ③214 ②④去 55 （Qian, 2001:91, 95, 108）	31+31→35+31 214+31→35+31 214+214→55+214 55+55→31+55	①+0→31+21 ③+轻→214+55 ②④+0→55+31 ③+0→55+31 (occasionally)

Data in Table 5 show that in Yantai dialect, the original ②Yang and ④Qu citation tones have merged, and there is no difference between them before the neutral tone (they are both high-level tones), which is different from Laizhou and Qingdao dialects.

2.5 Dalian dialect

The Dalian dialect is classified under the Denglian subdialects of Jiaoliao Mandarin, and the internal single-syllable tones of the Dalian dialect vary between the four-tone and three-tone systems. Song (1963) reported that the four-tone types in the urban area of Dalian are "Yin 312, Yang 34, Shang 213, and Qu 53", while in the Zhuanghe area of Dalian, the three-tone types are "Yin 312, Shang 213, and Qu (including Yang) 53".

2.5.1 Dalian urban dialect

According to historical survey records, the urban dialect of Dalian belongs to the four-tone system[3]. The single-character tones and compound-character tones of the Dalian urban dialect are shown in Table 6.

Table 6: The tones of the Dalian urban dialect.

Citation tones	Disyllabic tones	
	Tone sandhi without neutral tone	Tone sandhi with neutral tone
①312 ②34 ③213 ④53 （Song, 1963）	（No record）	（No record）
①312 ②34 ③213 ④53 （Li, 1996:733-735）	312+312→13+22 Jinzhou 金州 213+312→13+22 Jinzhou 锦州 34+312→34+22 53+312→53+22	（No record）

[3] Recently, it was found that there is also a three-tone system phenomenon in Dalian urban dialect (Liu 2023), which is not listed because the relevant investigation materials have not been officially published.

①312 ②35 ③213 ④53 （Dong, 2008:67-68）	312+312→13+312 drive 开车 312+35→31+35 arrange 安排 312+213→31+213 peak 山顶 312+53→31+53 cook 烧饭 213+312→13+312 Mars 火星 213+312→21+35 thunder 打雷 213+213→13+213 fruits 水果 213+53→21+53 vegetation 草木	①+0→312+□ doctor 医生 ②+0→35+□ knowledge 良心 ③+0→213+□ shoulder pole 扁担 ④+0→53+□ honesty and kind 厚道

Table 6 demonstrates that the citation tones of the Dalian urban dialect only have slight coarticulation and no tone sandhi.

2.5.2 Zhuanghe dialect of Dalian

The Yang tone and Qu tone have been combined to create the Zhuanghe dialect of Dalian, which uses the three-tone system. Some are pronounced as Yin tone (e.g., "鹅 goose"), while others are merged into the Qu tone (e.g., "麻 flax"). The pronunciation of the Middle Chinese voiced sonorant characters varies, and they are irregularly allocated to the Yin, Yang, and Qu tones. This phenomenon is quite common in the Denglian subdialects of Jiaoliao Mandarin.

Table 7: Tones of the Dalian Zhuanghe dialect.

Citation tones	Disyllabic tones	
	Tone sandhi without neutral tone	Tone sandhi with neutral tone
①312 ③213 ②④53 （Song, 1963）	（No record）	（No record）
①312 ③213 ②④53 （Dong, 2008:68-69）	312+312→13+312 sidehill 山边 312+213→31+213 hard 辛苦 312+53→31+53 high-rise 高楼 Peking Opera 京剧 213+312→13+312 Beijing 北京 213+213→13+213 seawater 海水 213+53→21+53 sprinkle 小雨	①+0→312+□ watermelon 西瓜 ③+0→213+□ eight 八个 ②④+0→53+□ nice and coo 凉快 nephew 外甥

Table 7 reveals that in the Dalian Zhuanghe dialect, both citation tones and disyllabic tones are three-tone types, and there is no tone sandhi.

2.6 Dandong dialect

The Dandong dialect is classified under the Gaihuan subdialects of Jiaoliao Mandarin, and there are distinct differences in its internal single-syllable tones between the four-tone and three-tone systems. Song (1963) reported that the four-tone types in the urban area of Dandong are "Yin 312, Yang 34, Shang 213, Qu 53", while in the Donggang (Dong'an) area of Dandong, the three-tone types are "Yin 312, Shang 213, Qu (including Yang) 53".

2.6.1 Dandong urban dialect (four/three-tone system)

A survey on the urban dialect of Dandong was carried out in August 2021. Mr. Sun, 66 years old, and Ms. He, 62 years old, who are both retired employees and have completed junior high school, were the pronunciation collaborators. Both of the collaborators believe that their speech is the same. However, Mr. Sun pronounces single-syllable tones in the four-tone system, while Ms. He pronounces single-syllable tones in the three-tone system. The analysis of recorded materials using the Praat speech software produced the descriptions of the single-syllable and connected-speech tones in the Dandong dialect in this article.

Table 8: Tones of the Dandong urban dialect.

Citation tones	Disyllabic tones	
	Tone sandhi without neutral tone	Tone sandhi with neutral tone
①312 ②34 ③213 ④53 （Song, 1963）	（No record）	（No record）
①312 ②24 ③214 ④52 （Liu, 1996:691-693）	312+312→35+312 hometown 家乡 214+312→25+312 Noon 午间 214+214→25+214 Little Li 小李	（无记录）
①43 ②24 ③13 ④51 （Zeng, 2023）	43+43→44+33 cloudy sky 阴天 43+24→44+24 flowerpot 花盆 13+13→24+21 fruits 水果	①+0→44+23 brick 砖头 ②+0→24+21 stone 石头 ③+0→22+34 wharf 码头 ④+0→53+21 can 罐头
①41 ②③13 ④52 （Zeng, 2023）	13+41→24+42 tomorrow 明天 13+13→24+23 bank 银行 13+13→24+21 milk 牛奶 13+52→24+52 culture 文化 13+41→22+42 train 火车 13+13→22+24 match 火柴 13+13→24+21 fruits 水果 13+52→22+52 glasses 眼镜	①+0→42+23 brick 砖头 ②+0→24+2 stone 石头 ③+0→22+34 wharf 码头 ④+0→53+2 can 罐头

From Table 8, it can be seen that:

(1) The pitch values of the ①Yin citation tone have been changed in recent decades from a low-falling tone to a high-falling tone (312 > 43/41);

(2) Tonal classifications have been simplified in recent years, such as merging ② Yang with ③Shang or ①Yin with ④Qu (Zheng, 2018; Liu, 2023);

(3) There is a noticeable division in tonal categories before the neutral tone in disyllabic tones, regardless of whether the citation tone is a four-tone or three-tone system.

2.6.2 Donggang dialect of Dandong

Local Dandong residents are aware of the linguistic difference between the Donggang dialect and the Dandong urban dialect. The Donggang dialect belongs to the three-tone system (with the merging of Yang tone and Qu tone). The specific situations of single-character tones and compound-character tones of Dandong Donggang dialect are shown in Table 9.

Table 9: Tones of the Donggang dialect of Dandong.

Citation tones	Disyllabic tones	
	Tone sandhi without neutral tone	Tone sandhi with neutral tone
①312 ③213 ②④53 (Song, 1963)	(No record)	(No record)
①51 ③213 ②④441 (Ma, 2007)	51+51→34+42 airplane 飞机 213+51→34+42 Mars 火星 213+21→34+323 fruits 水果 441+51→34+42 menu 菜单 long gown 长衫	①+轻→53+1 clothes 衣服 ③+轻→33+4 bone 骨头 ②④+轻→55+3 go in 进去 house 房子

Table 9 reveals that:

(1) In the Donggang dialect, the pitch value of the Yin tone has shifted from low-falling to high-falling (312 > 51) in recent decades;

(2) Both citation and disyllabic tones are three-tone types, and there is no tone sandhi.

3. From tone sandhi types to the formation and development of the three-tone system in Jiaoliao Mandarin

3.1 Types of tone sandhi within the three-tone system of Jiaoliao Mandarin

The various tone sandhi patterns in representative Jiaoliao Mandarin can be summarized and categorized as illustrated in Table 10 based on earlier citations (Tables 3-9).

Table 10: Comparison of tone sandhi categories in the Jiaoliao Mandarin three-tone dialects.

Category	Four -tone → three -tone system	The characteristics of tone sandhi		Dialects
		without neutral tone	with neutral tone	
I	Merger of ② and ④ tones	Merged tone categories that are not distinguished	Merged tone categories that are not distinguished	Yantai, Zhuanghe of Dalian, Donggang of Dandong
II	Merger of ② and ④ tones	Merged tone categories that are not distinguished	Merged tone categories that are distinguished	Laizhou, Qingdao
III	Merger of ② and ③ tones	Merged tone categories that are distinguished	Merged tone categories that are distinguished	Qiujiayu of Qingzhou, Dandong (urban area)

3.2 Timing of the three-tone Mandarin dialects formation as reflected by the different types of tone sandhi

Formation times of the three-tone Mandarin dialects in Jiaoliao are different.

For example, in the 1950s, dialects such as Yantai, Weihai, Fushan, Qixia, and Jimo in Shandong Province had become three-tone systems, while Qingdao, Haiyang, Rushan, and Zhaoyuan remained four-tone systems until the early 21st century (Qian, 2001, 2002).

Similarly, in the 1960s, Zhuanghe and Donggang in Liaoning Province were three-tone systems, while Dalian and Dandong urban areas remained four-tone systems until recent years (Song, 1963; Zeng, 2023; Liu, 2023).

Therefore, it is possible to infer the sequence of the three-tone Mandarin dialects formation in Jiaoliao through the differences in tone sandhi of the three-tone system, as shown in Table 11.

Table 11: Sequence of three-tone formation times in Jiaoliao Mandarin.

Category	I	II	III
Time of the formation	Before 1960	1960-2000	After 2000
The development stages	the stable stage	the formative stage	the initial stage
The characteristics of tone sandhi	Tone sandhi do not distinguish already merged tone categories	Tone changes before the neutral tone distinguish already merged tone categories	Tone sandhi distinguish already merged tone categories
Dialects	Yantai, Zhuanghe of Dalian, Donggang of Dandong	Laizhou, Qingdao	Qiujiayu of Qingzhou, Dandong (urban area)

3.3 The earliest region of the three-tone system formatting in Jiaoliao and its time

As depicted in Figure 3, Shandong Province's administrative boundaries under the Ming Dynasty included the majority of the Jiaodong and Liaodong Peninsulas and crossed the Bohai Strait. Therefore, at least 600 years ago, there was tight connectivity between the two peninsulas. In modern Jiaoliao variants with a three-tone system, the distribution of the three types of tone sandhi is shown in Figure 4.

The early three-tone system of Jiaoliao Mandarin (Category I) is distributed in both Shandong and Liaoning provinces (Figure 3 and Figure 4). As the Jiaoliao Mandarin in the Liaodong Peninsula came from immigrant languages in Jiaodong Peninsula, it can be considered that the Yantai dialect in Shandong is the representative variety of Jiaoliao Mandarin that first changed from a four-tone to a three-tone system.

Edkins recorded the Yantai dialect as a four-tone system in 1862: ①13, ②31, ③35, ④53 (quoted from Endo, 2015:204). Giet's recording of the Yantai dialect in 1946 had already become a three-tone system : ① a high-falling tone, ②④ a high-level tone, ③ a low-falling tone, with the merge of Tones ②and ④ (quoted from Endo, 2021). Therefore, based on that the three-tone system of the Yantai dialect was formed between 1862 and 1946, it can be assumed that the oldest instance of the transition from four to three tones in Jiaoliao Mandarin occurred no more than a century ago.

Figure 3: The area under Shandong's jurisdiction in the Ming dynasty.

Figure 4: Distribution of tone sandhi categories in the Jiaoliao Mandarin three-tone dialects.

4. Conclusion

Over the past century, some of the Jiaoliao Mandarin dialects have undergone a sound change from a four-tone to a three-tone system in monosyllabic words. This is mainly due to the merging of tonal categories with similar pitch values, and the fundamental reason is the polysyllabification of the lexicon (Qi, 2010; Wei & Zhu, 2021). The major findings of this study can be concluded as follows.

(1) There are three types of tone sandhi in disyllable words of the Jiaoliao Mandarin dialect with a three-tone system:

Category I—Tone sandhi do not distinguish merged tone categories;

Category II—Only the pre-neutral tone sandhi distinguishes merged tone categories;

Category III—Normal disyllabic and pre-neutral tone sandhi distinguish merged tone categories.

(2) The formation time of the Jiaoliao Mandarin dialect with a three-tone system can be inferred from early to late based on these three types of tone changes: Category I > Category II > Category III.

(3) The Yantai dialect in Shandong is likely the representative of Jiaoliao Mandarin that first underwent a change from a four-tone to a three-tone system, with the three-tone system diffused from this center.

(4) Hirayama (1998) pointed out that "the values of pre-neutral tone sandhi is diachronically one step before the citation tones". However, it is unclear how much earlier pre-neutral tones actually are, and even if they are, it might not have been that long ago.

References

Dong, Yan [董岩] (2008) *Dalian Fangyan Yuyin Yanjiu*《大连方言语音研究》[*Research on the phonetics of the Dalian dialect*]. Renmin University of China Master's thesis.

Endo, Mitsuaki (2015) Jin 150 Nian lai Hanyu Gezhong Fangyan Li de Shengdiao Yanbian Guocheng——Yi Edkins de Miaoxie Wei Chufa Dian 《近150年来汉语各种方言里的声调演变过程——以艾约瑟的描写为出发点》[The evolution of tones in various Chinese dialects in the past 150 years: A starting point based on Edkins's descriptions]，Endo Mitsuaki & Ishizaki Hiroshi (Ed.) *Historical Research on Modern Chinese*, 199-228. Hangzhou: Zhejiang University Press.

Endo, Mitsuaki (2021) Geographical distribution of the tone sandhi before the neutral tone in Shandong dialects. *Keizai kenkyū* 13: 1-13. doi: https://doi.org/10.34321/21864

Gao, Yujuan [高玉娟] (2018) Dalian Fangyan Yinping Diao Bianyi de Shehui yuyanxue fenxi 《大连方言阴平调变异的社会语言学分析》[Sociolinguistic analysis of the variation in yinping tone in the Dalian dialect]. *Journal of Liaoning Normal University (Social Science Edition)* 41: 86-93.

Gao, Yujuan [高玉娟] & Sun Qiuyan [孙艳秋] (2021) Liaodong Bandao Sandiao Xing Fangyan Qu Yuyan Yanjiu Jinzhan yu Qianzhan《辽东半岛三调方言区语言研究进展与前瞻》[Research progress and prospects of the three-tone dialects in the Liaodong Peninsula]，*Journal of Liaoning University*（*Philosophy and Social Sciences*）49: 124-132.

Hirayama, Hisao (1998) Cong Shengdiao Diaozhi Yanbian Shi de Guandian Lun Shandong Fangyan de Qingsheng Gian Biandiao《从声调调值演变史的观点论山东方言的轻声前变调》[From the perspective of the historical evolution of tone values to discuss the tone sandhi before the neutral tone in Shandong dialects]. *Fangyan* 1998(1): 7-13.

Institute of Linguistics, Chinese Academy of Social Sciences [中国社科院语言研究所] & Language Information Sciences Research Centre, City University of Hong Kong [香港城市大学语言信息科学研究中心] (2012) *Zhongguo Yuyan Ditu Ji (Di 2 Ban)*·*Hanyu Fangyan Juan* 《中国语言地图集（第2版）·汉语方言卷》[*Atlas of Chinese Languages (2nd Edition): Volume on Chinese Dialects*]. Beijing: The Commercial Press.

Li, Xiaofan [李小凡] (2004) Hanyu Fangyan Liandu Biaodiao de Cengji he Leixing《汉语方言连读变调的层级和类型》[The levels and types of tone sandhi in Chinese dialects with connected speech]. *Fangyan* 2004(1): 16-33.

Li, Xingjie [李行杰] (1999) *Qingdao Hua Yindang*《青岛话音档》[*Qingdao dialect audio files*]. Shaanghai: Shanghai Education Publishing House.

Liu, Jinghan [刘婧涵] (2023) *Liaoning Fangyan Shengdiao Yanjiu* 《辽宁方言声调研究》[*Research on the tones in the Liaoning dialects*]. Nankai University Master's thesis.

Liu, Minggang [刘明刚] (1996) Dandong Yinxi《丹东音系》 [The phonological system of Dandong dialect]，Edited by Chen Zhangtai [陈章太] and Li Xingjian [李行健] *Basic Vocabulary Collection of Mandarin Dialects*, 691-711. Beijing: Yuwen Press.

Liu, Xin [刘欣] (2021) *Shandong Weifang Fangyan Shengdiao Wenti Yanjiu*《山东潍坊方言声调问题研究》 [*Research on tones in the Weifang dialect of Shandong*]. Nankai University Master's thesis.

Luo, Futeng [罗福腾] (2010) Jiaoliao Guanhua《胶辽官话》[Jiaoliao Mandarin Chinese]，Edited by Qian Zengyi [钱曾怡] *Research on Mandarin Chinese Dialects*, 94-126. Jinan: Qilu Publishing House.

Ma, Xiaofang [麻晓芳] (2007) *Donggang Fangyan Yinxi Yanjiu*《东港方言音系研究》[*Research on the phonological system of Donggang dialect*]. Nankai University Master's thesis.

Qi, Haifeng [亓海峰] (2010) *Laiwu Fangyan Yuyin Yanjiu*《莱芜方言语音研究》[*Research on the phonetics of the Laiwu dialect*]. Beijing: Zhongguo Wenshi Press.

Qian, Zengyi [钱曾怡] (1982) *Yantai Fangyan Baogao*《烟台方言报告》 [*Report on the Yantai dialect*]. Jinan: Qilu Publishing House.

Qian, Zengyi [钱曾怡] (Ed.) (2001) *Shandong Fangyan Yanjiu*《山东方言研究》[*Research on the Shandong dialect*]. Jinan: Qilu Publishing House.

Qian, Zengyi [钱曾怡] (2002) *Hanyu Fangyan Yanjiu de Fangfa yu Shijian*《汉语方言研究的方法与实践》[*Methods and Practices in the Study of Chinese Dialects*]. Beijing: The Commercial Press.

Qian, Zengyi [钱曾怡], Ota Itsuku, Chen Hongxin [陈洪昕] & Yang Qiuze [杨秋泽] (2005) *Laizhou Fangyan Zhi*《莱州方言志》[*Laizhou Dialect Chronicle*]. Jinan: Qilu Publishing House.

Song, Xue [宋学] (1963) Liaoning Yuyin Shuolüe《辽宁语音说略》[An Overview of Liaoning Phonetics]. *Zhongguo Yuwen* 1963(2): 104-114.

Tan, Qixiang [谭其骧] (Ed.) (1996) *Zhongguo Lishi Ditu Ji* (Di 7 Ce) 《中国历史地图集》（第七册）[*Historical Atlas of China* (Volume 7)]. Beijing: Zhongguo ditu Press.

Wang, Jialing [王嘉龄] (2002) Beijing Hua Sisheng de Yingao Bianhua《北京话四声的音高变化》[The pitch changes of the four tones in Beijing dialect].*Nankai Linguistics* 1: 37-43.

Wei, Yang [魏阳] & Zhu Xiaonong [朱晓农] (2021) Yanhua Shiye Zhong de Jiaoliao Shengdiao Leixing 《演化视野中的胶辽声调类型》[Tone Types in Jiaoliao from an Evolutionary Perspective]. *Research on Folklore Texts*, Volume 27: 107-137. Beijing: The Commercial Press.

Yang, Jie [杨杰] (2012) *Shandong Laizhou Fangyan Shengdiao Yanjiu*《山东莱州方言声调研究》 [*Research on tones in the Laizhou dialect of Shandong*]. A Nankai University Master's thesis.

Zeng, Xiaoyu [曾晓渝] (2023) Dandong Hua Liandu Biandiao Yanjiu 《丹东话连读变调个案研究》[A case study of tone sandhi in Dandong Mandarin]. *Admiring the High Mountains: Felicitations on the 90th Birthday of Professor Wang Shi-yuan*, 185-198. Hong Kong: City University of Hong Kong Press.

Zhang, Shifang [张世方] (2000) Hanyu Fangyan Sandiao Xianxiang Chutan 《汉语方言三调现象初探》[An initial exploration of the three-tone phenomenon in Chinese dialects]. *Yuyan Yanjiu* 2000(4): 48-61.

Zhang, Shuzheng [张树铮] (2007) Jiaoliao Guanhua de Fenqu (Gao)《胶辽官话的分区（稿）》[Regional classification of Jiaoliao Mandarin (Draft)]. *Fangyan* (4): 363-371.

Zheng, Yafeng [郑雅凤] (2018) *Dandong Shiqu Fangyan Danzi Diao Shiyan Yanjiu* 《丹东市区方言单字调实验研究》[*Experimental Study on Single-Syllable Tones in Dandong Urban Dialect*]. Liaoning Normal University Master's thesis.

Tone as a criterion for identifying a lect: the case of Tai Yo language via geographical distribution of tone

Sầm Công Danh
(USSH, Hanoi VNU)

Abstract: In this paper, through the geographical distribution of tone, in particular, the tone patterns of Tai Yo – a Southwestern Tai language which is located in the North-Central area of Vietnam and Central Laos, we support the idea of tone value and tone category are important indications helping the native speakers realize their dialect firstly, and secondly their dialect community.
Keywords: Tai Yo, tone shape, tone pattern, dialect, tone

1. Introduction

Tai Yo is a Tai language that belongs to the Taic (Daic) branch of the Kam-Tai subfamily, of the Kra-Dai (Tai-Kadai) language family, according to the classification of Anthony Diller (2008). The language is spoken by Tai Yo people (the synonyms are Tai Muong, Tai Hàng Tổng, Tai Xiêng, and Tai Pao) with their distribution in the two provinces of Thanh Hoá and Nghệ An (Vietnam) and some varieties in Borikhamxay province of Laos named Tai Maen, Tai Pao. All ethnonyms of Tai Yo, Tai Muong, and Tai Hàng Tổng are autonyms of this speech community. Yet Tay Xiêng is an exonym used by Tai Moey (*Tay Mười* in Vietnamese) people in also Nghệ An.

Additionally, in Thanh Hoá province (Thường Xuân and Như Xuân districts), only the ethnonym "Tai Yo" exists in daily communication between native speakers, but in recent news media or public advertising, the term "Thái Trắng" (literally White Tai) is wielded even though it is imprecise information. On the contrary, in Nghệ An, we are faced with a difficult situation in applying the ethnonyms: *Tai Yo* is the oldest and most traditional autonym of the community ethnologically. And the autonym Tai Muong is either used in parallel or. Although *Tai Muong* just reflects the social status of this Tai group to distinguish them from other ones in the area. *Tai Muong* means "people in the city, people are not the residents." The name *Tai Hàng Tổng*, a special ethnonym, is

SẦM, Công Danh. 2023. Tone as a criterion for identifying a lect: the case of Tai Yo language via geographical distribution of tone. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 113–124. doi: https://doi.org/10.5281/zenodo.8374626

applied to mention Tai Yo who lived in the administrative division called *tổng* (總), in feudal society before. Honestly, the name *Tai Hàng Tổng* has never been used as a valid ethnonym. The last one, *Tai Pao*. This name appears to be mentioned among researchers rather than native speakers. Indeed, Tai Yo people in Nặm-Pao River basin just call themselves Tai Muong or Tai Yo. That is some ethnological information I have gathered on the field trips.

Besides, for the name *Tai Maen* in Laos, Chamberlain (1999) furnishes that *Maen* /mɛ:n^{A4}/ is derived from the name of the original homeland of this community - *Meuang of Xiêng Men* in Nghệ An province of Vietnam.

1.1. Language classification

Tai Yo is currently classified as a Southwestern Tai language (Ferlus, 2008). Another opinion grouped it into Northern Tai instead (Chamberlain, 1999). The classification of this language is needed to be continually considered in the future. In this paper, we accepted the first view, treating it like a member of the Southwestern subgroup.

Despite the fact that the linguistic surface of Tai Yo differs from the language of Nyaw/Nyo (ไทญ้อ - sometimes written as Yor, Lao Nyo, etc.) in Laos and Thailand, they share the common cognate of ethnonym: *jɔ:^{C} > jɔ:^{C4}/ɲɔ:^{C4}.

In Vietnam, the Tai Yo language is written in two traditional scripts known as Lai Tay and Lai Pao (or Lai Liệp Nặm). These scripts belong to the Abugida writing system, and they relate closely to other Southwestern Tai scripts (Sukhothai script, Fakkham script, etc.).

1.2. The materials

In this paper, most the data of geographical distribution of tone are collected from my personal fieldnotes. Also, some sources of available tonal data included:

- Tai Maen, Tai Pao and Tai Yo data from Chamberlain, James R. (1983) The Tai dialects of Khammouan Province: their diversity and origins. *Science of Language Papers*, 4, 62–95.

- Tai Maen data from Chamberlain, James R. (1991) Mène: A Tai dialect originally spoken in Nghê An (Nghê Tinh), Vietnam – Preliminary Linguistic Observations and Historical Implications. *Journal of the Siam Society*, 79(2), 103–123.

- Tai Maen data from the fieldnote of Jean Pacquement (2022), in personal communication.

- The corpus of recordings of Tai Yo (https://pangloss.cnrs.fr/corpus/Tai_Yo?lang=en&seeMore=true) and Tai Pao (https://pangloss.cnrs.fr/corpus/Tai_Paw?lang=en) in Pangloss' website.

2. The methods

2.1. Tone box models

Tonebox investigation is a suitable tool for collecting and analyzing tonal data. The basics of this approach are based on two criteria: *proto-tone categories* (p-T, the columns), and *proto-onset types* (p-O, the rows), or *proto-consonant types*.

In particular, there are three models of tone box as shown below.

Li Fangkuei's tone box (1977)

	A	B	C	D	
				DS	DL
Voiceless consonants	A1	B1	C1	DS1	DL1
Voiced consonants	A2	B2	C2	DS2	DL2

Gedney's tone box (1972)

		A	B	C	DS	DL
Voiceless consonants	***Voiceless frictions***	A1	B1	C1	DS1	DL1
	Voiceless unas. stops	A2	B2	C2	DS2	DL2
	Glottal	A3	B3	C3	DS3	DL3
Voiced consonants		A4	B4	C4	DS4	DL4
		Smooth syllables			***Checked syllables***	

Liao Hanbo 's tone box (2022)

		A	C	DS	B	DL
Voiceless consonants	***Aspirated 1A***	A-1A	C-1A	DS-1A	B-1A	DL-1A
	Continuant 1C	A-1C	C-1C	DS-1C	B-1C	DL-1C
	Unaspirated 1U	A-1U	C-1U	DS-1U	B-1U	DL-1U

	Glottal 1G	A-1G	C-1G	DS-1G	B-1G	DL-1G
Voiced consonants	***Plosive + continuant***	A2	C2	DS2	B2	DL2

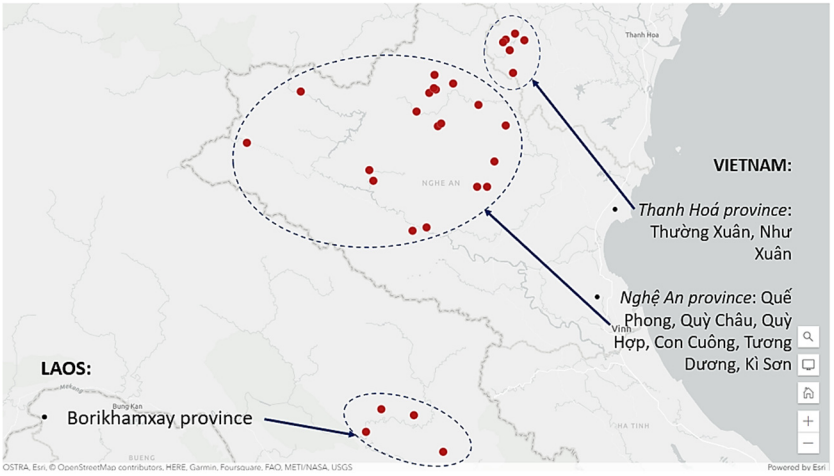
Among these three models, Gedney's tone box is considered the most suitable for Southwestern Tai. Hence, for Tai Yo, I applied to Gedney's. Moreover, Liao's tone box is the newest and most applicable model to the whole Tai languages, especially the Central Tai and Northern Tai languages, because of the complexity of splitting-merging in their tonal system.

2.2. Research questions

For the collected data, we have two questions in this study:

1. How about the tonal system of the Tai Yo language via its varieties?
2. What is the main factor in identifying a lect, in the case of Tai Yo?

3. Geographical distribution of tone in Tai Yo



The map of investigated places.

The investigated places of this paper (29 locations in total), include:

In Vietnam:

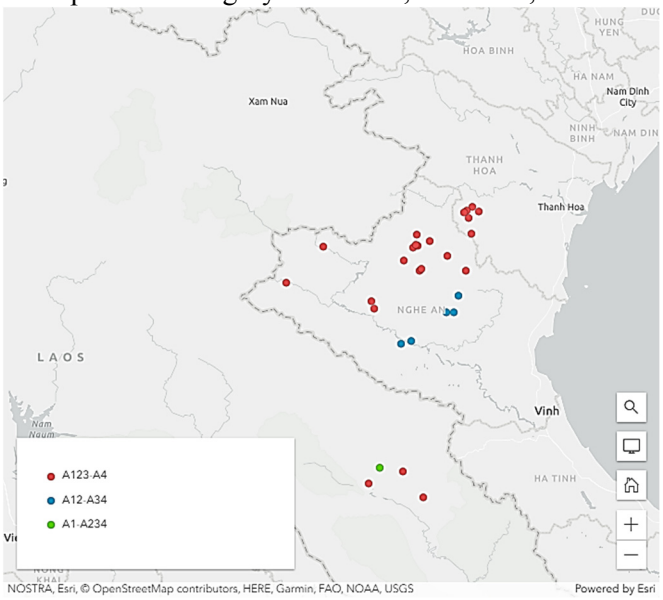
- Thanh Hoá province (the districts of Thường Xuân and Như Xuân): 6 locations.
- Nghệ An province (the districts of Quế Phong, Quỳ Châu, Quỳ Hợp, Con Cuông, Tương Dương, Kì Sơn): 19 locations.

In Laos:

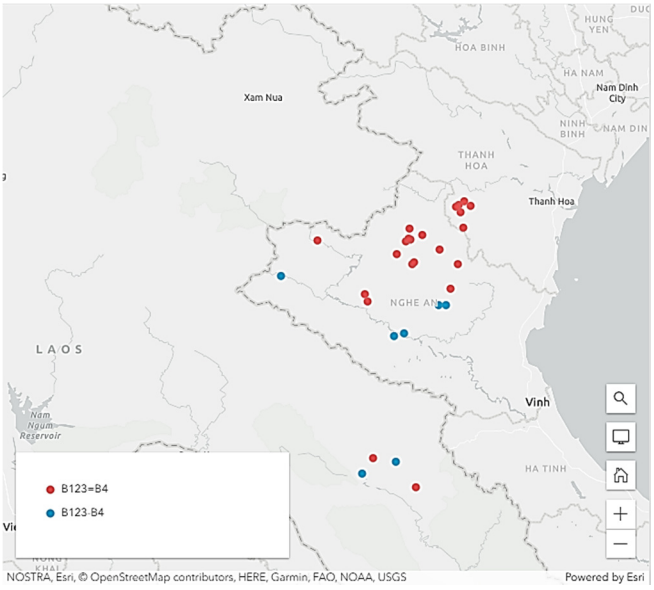
- Borikhamxay province: 4 locations.

3.1. Tone split of Tai Yo language

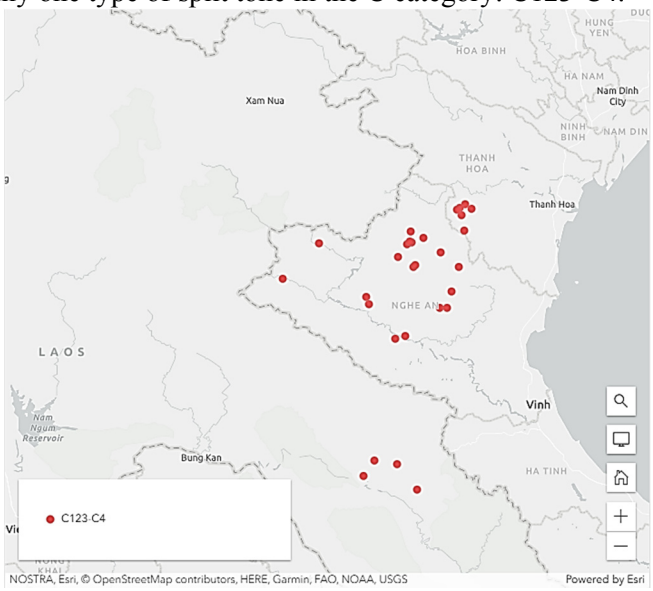
Three types of tone split in A category: A123-A4, A12-A34, and A1-A234.



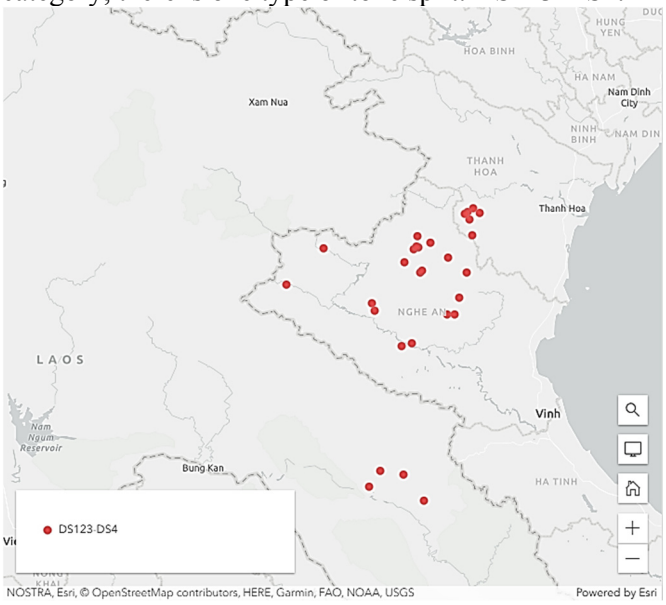
Two types of tone split in B category: B123=B4 (merged all in B) and B123-B4.



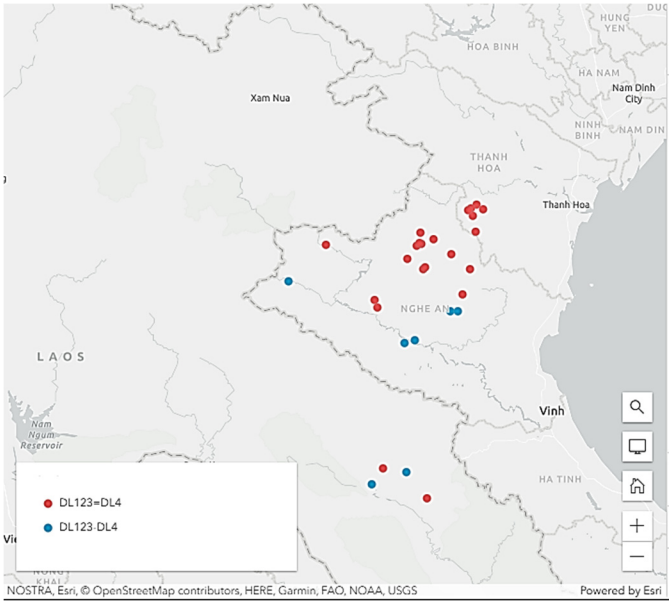
There is only one type of split tone in the C category: C123-C4.



In the DS category, there is one type of tone split: DS123-DS4.



Respectively B category, there are two types of tone split in the DL category: DL123=DL4 and DL123-DL4.



3.2. Tone shapes of Tai Yo language

Alternately, we will show the tone shapes in Tai Yo's tone, by proto-tone category as A, B, C, DS, and DL. The tone values are represented by the tone letters of Chao Renyuen and the numerical value.

The tone shapes in the A category:

- Proto-voiceless consonants (A1; A123; A23): ꩻ 21, ꩵ 22, ꩻꩻ 211, ꩻ 31 / ꩻ 312; ꩵ 13, ꩵ 113, ꩵ 114, ꩻꩵ 213, ꩵ 25, ꩵ 24, ꩵ 325, ꩵ 334
- Proto-voiced consonants (A4; A34; A234): ꩵ 35, ꩵ 224, ꩵ 225; ꩵ 445, ꩵ 455, ꩵ 355; ꩵ 44, ꩵ 55

The tone shapes in the B category:

- Proto-voiceless consonants (B123; B123=B4): ꩻ 21, ꩻ 21 / ꩻ 11, ꩻ 21 / ꩻ 31, ꩻ 31; ꩻ 221, ꩻ 331; ꩻ 11, ꩵ 22; ꩻ*ꩵ 214*
- Proto-voiced consonants (B123=B4; B4): ꩻ 21, ꩻ 21 / ꩻ 11, ꩻ 21 / ꩻ 31, ꩻ 31; ꩻ 221, ꩻ 331; ꩻ 11, ꩵ 22; ꩻ *32,* ꩻ *332,* ꩻ *331*

The tone shapes in C category:

- Proto-voiceless consonants (C123): ꩵˀ 33ˀ, ꩵ 33, ꩵ 44, ꩵˀ 44ˀ; ꩻ 332, ꩻˀ 322ˀ, ꩻˀ 32ˀ, ꩵˀ 545ˀ
- Proto-voiced consonants (C4): ꩻ 152, ꩻˀ 153ˀ, ꩻˀ 154, ꩻˀ 155ˀ; ꩻ 553, ꩻ 552, ꩻ 541, ꩻˀ 452ˀ; ꩻ 52, ꩻ 53, ꩻˀ 53ˀ, ꩻ 54

The tone shapes in DS category:

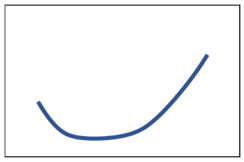
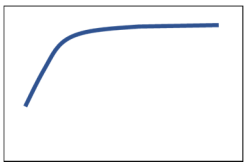
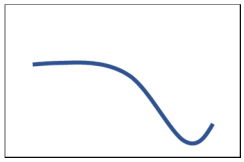
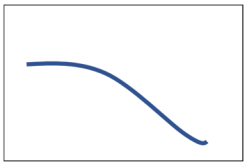
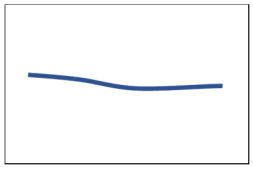
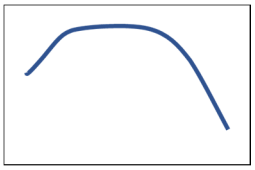
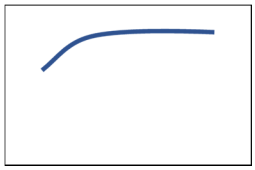
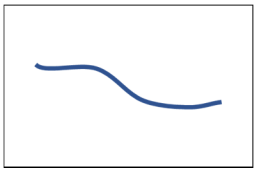
- Proto-voiceless consonants (DS123): ꩵ 34, ꩵ 35, ꩵ 45, ꩵ 45 / ꩵ 55, ꩵ 455; ꩵ 55, ꩵ 55 / ꩵ 45

- Proto-voiced consonants (DS4): ꜒ 22, ꜒ 22 / ꜓ 21, ꜒ 33, ꜒ 33 / ꜓ 21, ꜒ 33 / ꜖ 31, ꜒ 33 / ꜖ 32, ꜒ 44, ꜒ 44 / ꜖ 42, ꜒ 44 / ꜖ 41; ꜖ 332, ꜖ 31, ꜓ 21, ꜖ 332 / ꜒ 33

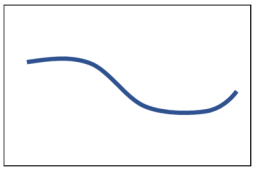
The tone shapes in the DL category:

- Proto-voiceless consonants (DL123=DL4; DL123): ꜓ 11, ꜓ 22, ꜓ 22 / ꜓ 21, ꜒ 33, ꜒ 33 / ꜖ 32, ꜒ 33 / ꜓ 21, ꜒ 44 / ꜖ 42, ꜒ 44 / ꜖ 41; ꜖ 42, ꜓ 21 / ꜒ 33, ꜖ 31, ꜖ 331, ꜓ 221; ꜓꜔ *214*
- Proto-voiced consonants (DL123=DL4; DL4): ꜓ 11, ꜓ 22, ꜓ 22 / ꜓ 21, ꜒ 33, ꜒ 33 / ꜖ 32, ꜒ 33 / ꜓ 21, ꜒ 44 / ꜖ 42, ꜒ 44 / ꜖ 41; ꜖ 42, ꜓ 21 / ꜒ 33, ꜖ 31, ꜖ 331, ꜓ 221

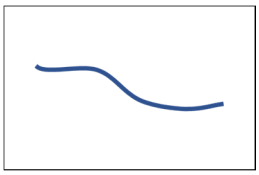
In summary, we have an overview of *generalized tone shapes* of Tai Yo:

	Proto-voiceless consonants	**Proto-voiced consonants**
A	 *Concave contour*	 *High-rising (+ level) contour*
B	 *Mid-falling (+ rising) contour*	 *Mid-falling contour*
C	 *Level contour*	 *Convex contour*
DS	 *High-rising (+level) contour*	 *Low-mid-falling contour*

DL



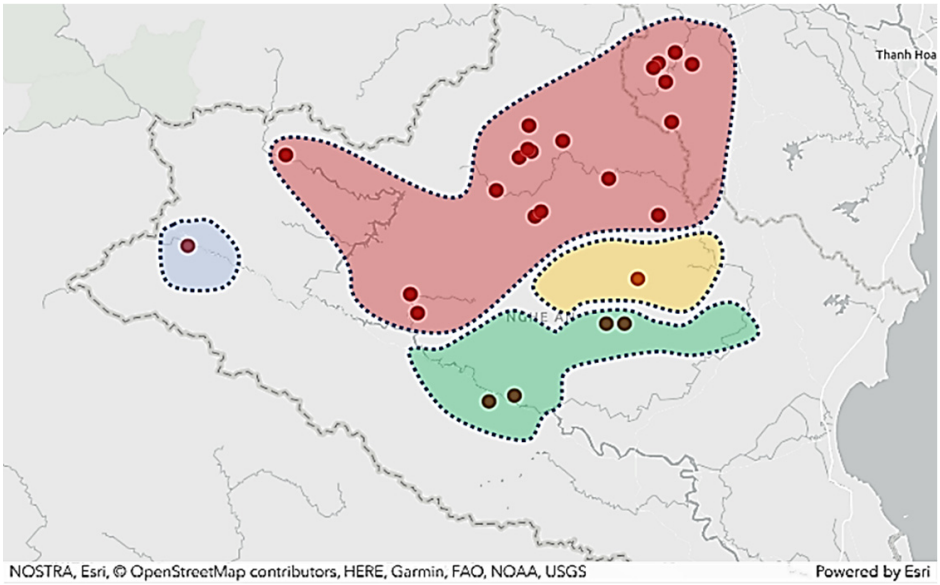
Low-mid-falling (+rising) contour



Low-mid-falling contour

3.3. Tone patterns of Tai Yo lects

Tone patterns of Tai Yo lects in Vietnam:



Pattern 1 (red part) – Thường Xuân, Như Xuân, Quế Phong, Quỳ Châu, Tương Dương:

A1	**B1**	**C1**	DS1	DL1
A2	**B2**	**C2**	DS2	DL2
A3	**B3**	**C3**	DS3	DL3
A4	**B4**	**C4**	DS4	DL4

Pattern 2 (yellow part) – Quỳ Hợp:

A1	**B1**	**C1**	DS1	DL1
A2	**B2**	**C2**	DS2	DL2
A3	**B3**	**C3**	DS3	DL3
A4	**B4**	**C4**	DS4	DL4

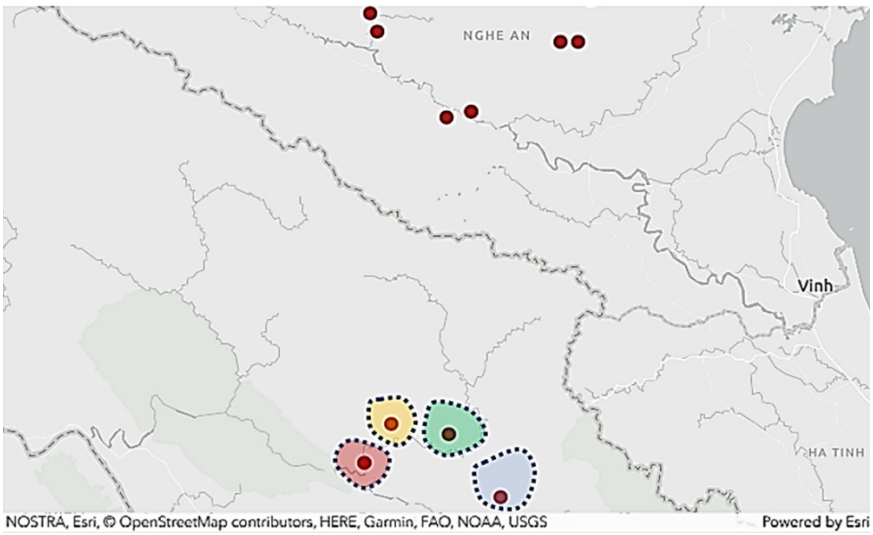
Pattern 3 (green part) – Quỳ Hợp, Con Cuông:

A1	**B1**	**C1**	DS1	DL1
A2	**B2**	**C2**	DS2	DL2
A3	**B3**	**C3**	DS3	DL3
A4	**B4**	**C4**	DS4	DL4

Pattern 4 (blue part) – Kì Sơn:

A1	**B1**	**C1**	DS1	DL1
A2	**B2**	**C2**	DS2	DL2
A3	**B3**	**C3**	DS3	DL3
A4	**B4**	**C4**	DS4	DL4

Tone patterns of Tai Yo lects in Laos:



Pattern 1 (red part) – Tai Pao (Chamberlain, 1983):

A1	**B1**	**C1**	DS1	DL1
A2	**B2**	**C2**	DS2	DL2
A3	**B3**	**C3**	DS3	DL3
A4	**B4**	**C4**	DS4	DL4

Pattern 2 (yellow part) – Tai Maen (Chamberlain, 1999):

A1	**B1**	**C1**	DS1	DL1
A2	**B2**	**C2**	DS2	DL2
A3	**B3**	**C3**	DS3	DL3
A4	**B4**	**C4**	DS4	DL4

Pattern 3 (green part) Tai Maen (Chamberlain, 1983):

A1	**B1**	**C1**	DS1	DL1
A2	**B2**	**C2**	DS2	DL2
A3	**B3**	**C3**	DS3	DL3
A4	**B4**	**C4**	DS4	DL4

Pattern 4 (blue part) – Tai Maen (Pacquement, 2022):

A1	**B1**	**C1**	DS1	DL1
A2	**B2**	**C2**	DS2	DL2
A3	**B3**	**C3**	DS3	DL3
A4	**B4**	**C4**	DS4	DL4

4. A criterion for identifying a lect

In self-identifying the lect (such as dialect, subdialect, etc.) of a tonal language, from the point of view of a native speaker, we recognized two important features at the surface of language:

The former is, regular lexicon or daily vocabulary, especially for subdialects. For example, in Tai Yo, "to chase, to run after" in this lect A is /tuət^{DL4}/, but other lects are /hɔ:m^{C123}/. The interesting point is, there is still /hɔ:m^{C123}/ in the lexicon of the lect A, but they do not use it in daily life. Therefore, in lect A, /tuət^{DL4}/ becomes a word of the regular lexicon/daily vocabulary. The latter is, tonal system, including tone patterns and tone shapes.

Hoàng Thị Châu (2009, in Vietnamese) said: “In Vietnamese, tone is one of the features used to distinguish the dialect and subdialect. **People usually base on the tonal features of the speaker but not on other ones, to recognize what a dialect, subdialect is and where it is distributed**.... Each special dialect, subdialect has its own tonal system. When somebody hears someone speak a strange dialect, he has to verify the code of the tones before he can understand the content.”

In the case of Tai Yo, a tonal language, of course, like Vietnamese. This statement of Hoàng Thị Châu is still relevant. Thus, we attested to the tone patterns (including tone split) in sections 3.1 and 3.3, as well as the tone shapes in section 3.2. Through the investigated data of Tai Yo.

5. Final remarks

We have two observations that will assist us in answering the research questions of this paper:

- By *tone patterns*: firstly, it can divide the tone patterns of Tai Yo into two groups 5-tone type and 6-tone type. According to the tonemes in tone categories of A, B and C (smooth syllable). The 5-tone type in fact are more common and typical. Secondly, the most noticeable distinction between these tone patterns of Tai Yo's varieties is the splitting-merging in the tone categories A, and B (resulting in category DL).
- By *tone shapes*: in my opinion, this is the main factor in identifying a lect. The tone shapes of varieties of Tai Yo are approximately similar to each other. Not only are tone contours, tone glottal important but so are tone values. For a native speaker, this is the primary thing his auditory sense feels.

References

Chamberlain, Robert James (1983) "The Tai Dialects of Khammouan Province: Their Diversity and Origins." *Science of Language Papers*, no. 4: 62–95.

Chamberlain, Robert James (1991) "Mène: A Tai Dialect Originally Spoken in Nghê An (Nghê Tinh), Vietnam – Preliminary Linguistic Observations and Historical Implications." *Journal of the Siam Society* 79 (2): 103–23.

Diller, Anthony V.N.; Edmonson, Jerold A.; Luo, Yongxian (ed.) (2008) *The Tai-Kadai languages*, Routledge.

Ferlus, Michel (2008) "Les Dialectes Tai Du Nghệ An, Vietnam (Tay Daeng, Tay Yo, Tay Muong)." In *The Tai-Kadai Languages*, edited by Anthony V N Diller, Jerold A Edmondson, and Yongxian Luo, 298–315. Routledge. https://halshs.archives-ouvertes.fr/halshs-01617791

Gedney, William J (1972) A Checklist for Determining Tones in Tai Dialects. In *Studies in Linguistics in Honor of George L. Trager*, 423–37. The Hague: Mouton.

Hoàng, Thị Châu (2009) *Phương Ngữ Học Tiếng Việt* [The dialectology of Vietnamese]. NXB ĐHQGHN.

Li, Fang-kuei (1977) *A Handbook of Comparative Tai*. Manoa: University of Hawaii Press.

Liao, Hanbo (2016) *Tonal Development of Tai Languages*. Payap University, Chiang Mai. https://www.academia.edu/29464397/TONAL_DEVELOPMENT_OF_TAI_LANGUAGES.

Liao, Hanbo (2022) An Integrated Tone Box Scheme for Determining Tones in Tai Varieties beyond Southwestern Tai: Diachronic and Synchronic Concerns. *Folia Linguistica*, November. https://doi.org/10.1515/flin-2022-2048.

The Underlying Tonal Characteristics of Northeast Mandarin in Liaoning

Liu Jinghan
(Nankai University)

Abstract: Northeast Mandarin's tonal value is quite comparable to that of Standard Mandarin, with T1 (Yinping) being only a little bit lower. Additionally, the number of Qingru morphemes merged with T3 (Shangsheng) in terms of tonal categories is higher than that of Standard Mandarin. However, in contrast to earlier claims, fieldwork reveals that there are various dialect points in Central Liaoning where T1(Yinping) has a low falling tone. This work suggests that the tonal system of Northeast Mandarin in Liaoning is highly biased towards the low register because T1 is a low-falling tone and T2 (Yangping) is a low-rising tone, according to the characteristics of sandhi tone patterns and historical documents. The causes of specific T1 manifestations are further analyzed in this work, and it suggests that language contact is the reason that some dialects have level form of T1.
Key words: Liaoning province, Northeast Mandarin, tone, underlying form

1. Introduction

1.1. Geographical and Linguistic Situation of Liaoning

The northeastern Chinese province of Liaoning faces the Jiaodong Peninsula across the Bohai Sea Strait, with the Bohai Sea and the Huanghai Sea to the south. The Liaoning region's Mandarin dialects are predominately Northeast Mandarin, which is found in the center part of Liaoning and covers the largest area; Beijing Mandarin is distributed in the western part and Jiaoliao Mandarin is predominately distributed in the southeast coastline area, as shown in Figure 1[1]:

LIU, Jinghan. 2023. The underlying tonal characteristics of Northeast Mandarin in Liaoning. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 125–142. doi: https://doi.org/10.5281/zenodo.8374630

* This research was funded by the National Social Science Foundation Project of China (20AYY020). The project is entitled "Research on the pattern of compound-character tones in Mandarin Chinese dialects". This paper is rewritten from a section of the author's master's thesis and read at the fifth International Conference on Asian Geolinguistics, thanks to the guidance of professor ZengXiaoyu and valuable comments from professer Mitsuaki Endo and professer Shen Ruiqing.

[1] The language mapping software developed by Nankai University and Tianjin Xinhui Network Technology Service Center is used in all the language maps listed in this paper; the base map is based on Gaode map (©2023AutoNavi—GS (2022)1061).

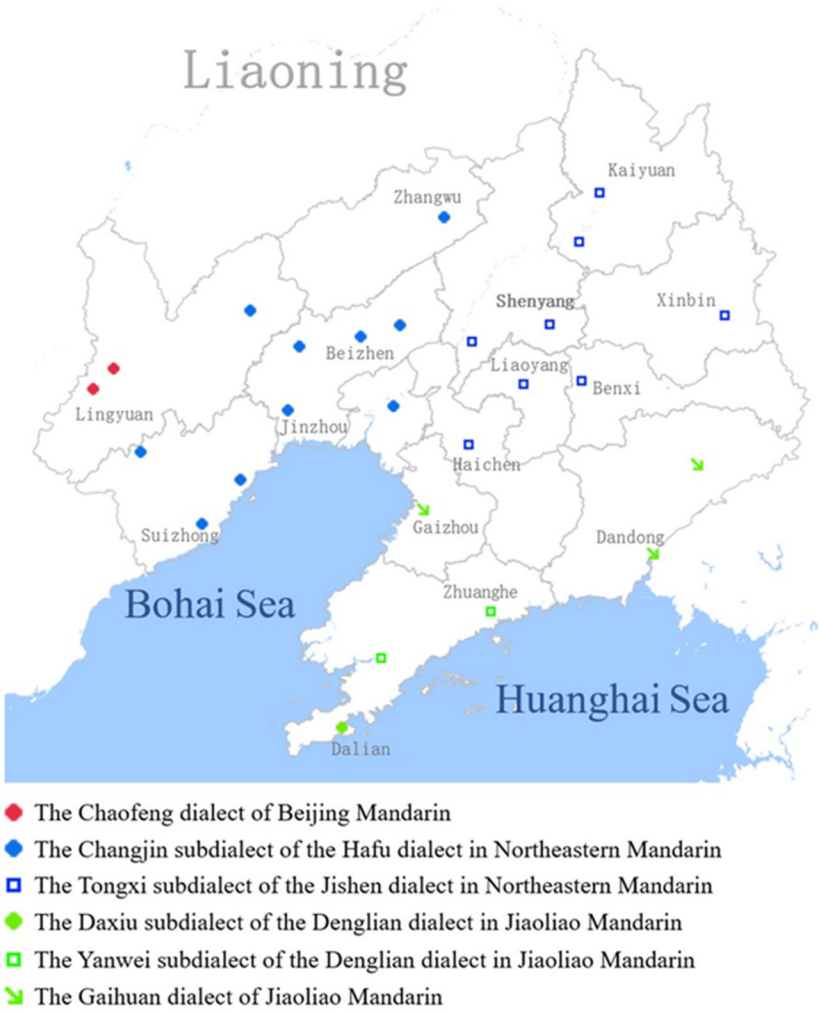


Figure 1: The regional distribution of Mandarin dialects in Liaoning.

The phonology of Northeast Mandarin has four tones, including T1(Yinping), T2 (Yangping), T3(Shangsheng), and T4(Qusheng). The merging of tones is consistent with that of Standard Mandarin; the only difference is that more Qingru morphemes merge with T3 than Standard Mandarin.

1.2. Background of the research problem

The data source of this article consists primarily of first-hand survey materials from fieldwork in 2021 and 2022, in addition to findings of previous studies, mainly <中国语言资源集·辽宁>, the data of Panjin from Li Nanyi (2019), and Kaiyuan from Liu

Xin (2021). The data of the first-hand survey materials were obtained through recording and pitch values were obtained through phonetic experiments. The fieldwork investigated the citation tone and tone sandhi of disyllable.

Studies have shown that the tone system of Northeastern Mandarin is relatively simple and similar to that of Standard Mandarin; only the register of T1 is slightly lower (He Wei 1986, Zhang Zhimin 2005), representing 33 or 44. However, the fieldwork showes that there are some survey locations in the Central Liaoning where T1 is a falling tone, contrary to previous reports.

Level tone and falling tone are two of the synchronic expressions of T1, and level tone is the diachronic expression, wheather falling tone or level tone is the earlier form? There are three possible explanations for the existing situation. First explanation is that falling tone has become more popular in recent years, and T1 in Liaoning's Northeast Mandarin is undergoing the "level tone > falling tone" transition. The second explanation is that T1 does not experience sound change, and the discrepancy between previous reports and fieldwork is due to different definitions of pitch decline. The decline of T1 should be regarded as a natural attenuation of pronunciation, and the results of the phonetic experiment should be regarded as level tone when summing up the phonology. However, the third explanation is that the falling tone is the older form, and in earlier research, the contour of T1 was incorrectly recorded due to a flaw in the standard notation system, and the level tone is a more recent development because of language contact.

This article supports the third explanation, and we believe that the underlying characteristic of the Northeast dialect in Liaoning is T1 expressing a falling tone. In the following text of this article, this point of view is revealed using historical records and the typical and distinguishing characteristics of Northeast Mandarin. Besides, this article will cover the linguistic influences that have affected T1 development over time, as well as the general tonal characteristics of the Northeast Mandarin in Liaoning.

2. Falling tone is the underlying form of T1

2.1. The synchronic performance of T1 of the Northeast Mandarin in Liaoning

As shown in Figure 2, T1 in central Liaoning is low falling tone or slightly falling tone. The blue points represent a slightly falling tone, and the red star represents a low falling tone. They are distributed in the red circle, which has a different contour than Standard Mandarin, where T1 is a level tone, and the register is always low. In the western region near Beijing Mandarin, T1 in most survey points represents a level tone.

Figure 2: The performance of T1 in Liaoning.

The phonetic experiment results show that T1 in related survey points has a clear downward trend. Figure 3 shows the tone pattern drawn based on the results of phonetic experiments, the black line represents T1, which always appears in the low register.

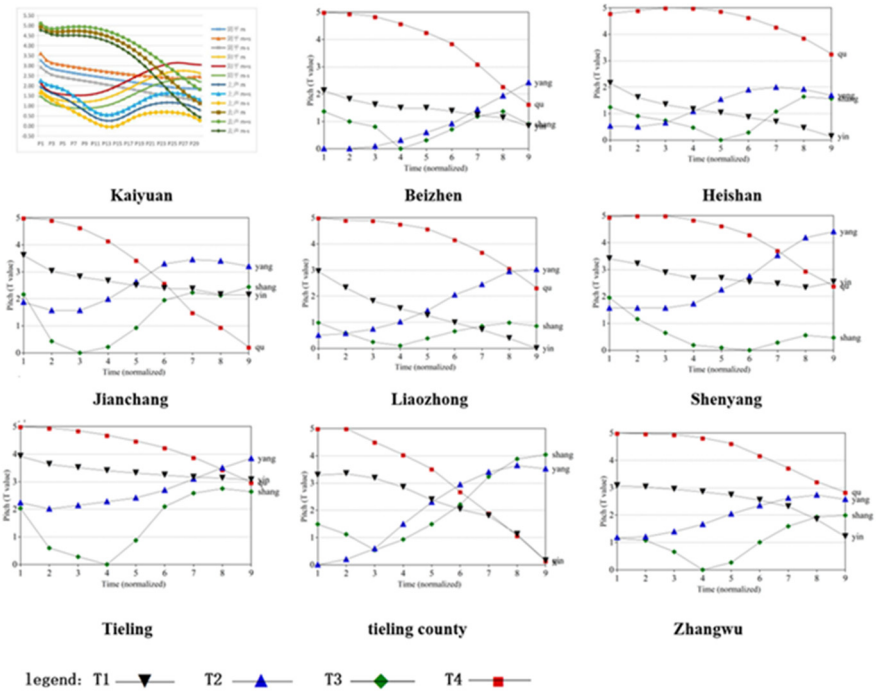


Figure 3: Tone pattern of survey points in central Liaoning.

T1 in these survey points is described as level tone in previous studies, as shown in Table 1. For example, the tone pitch of T1 in Heishan dialect is 31 measured by phonetic experiment, while Song Xue (1963) and Han Tianyu (2014) both recorded 44, similar situations exist in other survey points. The tone pitch in previous studies was recorded by traditional measures, relying on subjective judgment.

Table 1: Comparison between recording of phonetic experiment and traditional notation method.

Survey points	Experimental results	Traditional measure	
		Tone pitch	Source
Kaiyuan	32	44	Na Lin (2019)
Beizhen	32	44	Song Xue (1963)
Heishan	31	44	Song Xue (1963)
		44	Han Tianyu (2014)
Jianchang	43	44	Song Xue (1963)
Liaozhong	31	44	Song Xue (1963)
		33	Wang Qi (2016)
Shenyang	43	33	Song Xue (1963)
Tieling	43	44	Song Xue (1963)
Zhangwu	31	44	Song Xue (1963)

Two reasons for why falling tone is the early form of T1 will be presented in 2.2 and 2.3.

2.2. The T1 falling tone form remains in the last word of disyllable

Except for Tieling, the dialect points where T1 represents a falling tone in citation pattern, the sandhi tone pattern often retains T1's falling form, especially in the last word of disyllable. Taking Liaozhong as an example, in the citation tone pattern, T1 is 31, T2 is 13, T3 is 212, and T4 is 53. The tone sandhi in disyllable is shown in Table 2.

Table 2: Tone sandhi in disyllable in the Liaozhong dialect.

Last word / First word	T1 (31)	T2 (13)	T3 (212)	T4 (53)	Netual tone (0)
T1 (31)	33+31 飞机 "plane"	33+12 欢迎 "welcome"	33+211 真理 "truth"	33+53 公路 "highway"	44+0 公公 "father in law"
T2 (13)	24+31 平安"safety"	24+12 农民"farmer"	24+211 成语"idiom"	13+53 肥皂"soap"	24+0 娃娃"doll"
T3 (212)	21+32 火车"train"	11+13 火柴"match"	24+211 水果"fruit"	21+53 考试"exam"	21+0 祖宗"ancestors" 24+0 想想"to think"
T4 (53)	53+31 面包 "bread"	52+12 杏仁 "almond"	53+211 电脑 "computer"	53+53 运动 "motion"	52+0 弟弟 "younger brother"

Data in Table 1 show that recent studies continue to record T1 as a level tone, ruling out the possibility that T1's falling tone is a later form of sound change. This is because, in such a short period, T1's "level tone > falling tone" change was widely observed in central Liaoning and had a significant impact on both citation and sandhi tone pattern, which is a lack of motivation. The likelihood of significant change in tone contour occurring in a short period is also low. Therefore, how should the downward trend of T1 be defined? Is it a coarticulation brought on by energy attenuation and should be considered a level tone, or is it a falling tone and is recognized as a tonal characteristic of Northeastern Mandarin? To answer this question, we should investigate the situation of Northeast Mandarin in the whole Mandarin region.

Table 3: T1 in citation and sandhi tone pattern of the Mandarin dialects.

Mandarin area	Dialect point	Citation tone	Sandhi tone		The last word is a falling tone in sandhi tone pattern
			First word	Last word	
	Kaiyuan	32	33	32	+

Northeastern Mandarin	Beizhen	32	33	33/32	+
	Heishan	31	33/32	33/31/21	+
	Jianchang	43	43/33	33/31	+
	Liaozhong	31	33	32/31	+
	Shenyang	43	33	33/43/32/31	+
	Tieling	43	44	44/33	-
	Tieling county	41	44/33	31	+
	Zhangwu	31	33	32/31	+
	Baicheng	31	33	31	+
	Jiutai	33	33	33/32	+
	Tonghua	32	33	32	+
	Qiqihaer	33	33	33/32	+
	Mudanjiang	31	33	33/31	+
	Daqing	31	33	31	+
Beijing Mandarin	Beijing	55	55	55	-
	Lingyuan	33	44	44/33	-
	Jianping	44	44	44	-
	Chengde	55	55	55	-
	Chifeng	55	55	55	-
Jiaoliao Mandarin	Yantai	31	35/31	31	+
	Qingdao	213	55/31/213	213	-
	Zhucheng	214	24/214	214	-
Jilu Mandarin	Tangshan	55	24/55	55	-
	Tianjin	21	113/21	21	+
	Baoding	45	45	45	-
	Shijiazhuang	23	23/44	23	-
	Lijin	213	23/212	213	-
	Jinan	213	23/213	213	-
Zhongyuan Mandarin	Linfen	31	33	31	+
	Lingbao	31	31	31	+
	Xinyang	33	33	33	-
	Hanzhong	41	41	41	+
	Xi'an	21	24/21	21	+
	Baoji	21	24/21	21	+
	Jining	213	13/213	213	-
	Zhengzhou	24	24	24	-
	Baihe	24	24	24	-
	Fuyang	213	213	213	-
	Xuzhou	213	24/21/213	213	-
Lanyin Mandarin	Lanzhou	31	31	31/21	+
	Yinchuan	44	44	44	-
Southwest Mandarin	Chengdu	44	44	44	-
	Xichang	33	33	33	-
	Chongqing	55	55	55/44	-
	Zhaotong	44	44	44	-
	Mengzi	44	44	44	-
	Guiyang	55	55	55	-
	Wuhan	55	55	55	-
	Xiangfan	34	34	34	-
	Zigong	45	45	45	-
	Anqing	31	35/45/31	31	+

Jianghuai Mandarin	Hefei	31	35/31	31	+
	Nanjing	31	33/13/31	31	+
	Lianyungang	214	21/214	214	-

Data in Table 3 demonstrate that, except for Northeast Mandarin, where T1 is a falling tone in the final syllable, T1 is likewise a falling tone in its citation form, as in Yantai, Tianjin, Linfen, Xi'an, Hanzhong, and Lanzhou. However, in the regions of Beijing Mandarin and Southwest Mandarin where the citation form of T1 is a level tone, it was not possible to detect a clear downward trend when T1 is in the position of the last word of a disyllable. Therefore, coarticulation elements like "fall pitch" or "pitch declination" can not explain the downward tendency in Northeast Mandarin's final word of a disyllable because they do not significantly alter tone contour.

Figure 4: The contour of T1 in the position of the last word of disyllable in Mandarin dialects.

From Figure 4, it can be seen that in the region of Northeast Mandarin, although in citation form T1 occasionally represents a mid level tone or a slightly falling tone, the falling tone features in the last word of the sandhi tone pattern cannot be ignored. The adjacent Beijing Mandarin region and the Southwest Mandarin region, where T1 represents a level tone in the citation tone pattern, do not exhibit an obvious declination in the last word of the sandhi tone pattern as a result of "saving energy", proving that "saving energy" cannot account for this regional commonality, which is unique in the entire Mandarin region. Therefore, in early Northeast Mandarin, T1 in both citation tone pattern and the last word of disyllable is a falling tone, like the situation of Tianjin and Xi'an. However, in recent years, Northeast Mandarin is undergoing drastic changes. The speed and degree of change in T1 contour in citation and sandhi tone pattern are different. Therefore, the contour of T1 in some dialect points exhibits inconsistency between citation form and the last word of disyllable. Although the citation form and the first word of disyllable are influenced by Standard Mandarin and tend to become high or mid level tones, the common downward trend of the last word of disyllable can only be explained by "containing ancient ingredient " and reflecting the early form.

2.3. Records of historical documents

The falling tone form of T1 was kept in the early Beijing dialect. It is well known that the rulers of the Qing Dynasty brought many Northeast immigrants to Beijing. Therefore, the early Beijing dialect was deeply influenced by the Northeast dialect.

Professor Chen Xiao (2018) pointed out that T1 in the Beijing dialect was recorded to have both high level tone and high falling tone forms before 1900. Most people were used to the high level tone, while there was also a considerable number of people who were used to the high falling tone. The high level tone is a variant of the high falling tone. It was reported that "But it becomes the upper even monotone, in combination with another word following". Professor Endo (1986,2015)[2] and Bruce (1930)[3] also pointed out that there is a tendency for T1 to become T4 before the neutral tone according to Lao She's recording. Historical records showes that the tone sandhi of T1 is comparable to that of central Liaoning dialects, maintaining the earliest occurrence of Northeast Mandarin. Manchus from Northeast China are responsible for bringing the falling tone T1 form to the early Beijing dialect. Most of the Manchus brought to Beijing with the emperor spoke Northeast Mandarin, particularly those from the Shenyang region. The Eight Banner Soldiers learned a more authoritative common language when entering Beijing, but they also kept their dialect characteristics.

[2] Quoting from Chen Xiao(2018).
[3] Quoting from Chen Xiao(2018).

3. Source of the falling tone form of T1 in Northeast Mandarin in Liaoning

Historically, it is well known that the Han population in Liaoning was mainly composed of immigrants, and the native people were a minority ethnic group. After Liao Dynasty, minority rulers plundered people through wars and brought them to Liaoning, the main sources were Shandong, Shanxi(山西), and Hebei province. Also, many people migrated from Shandong to Liaoning due to the convenient geographical location. As a result, before Qing Dynasty, the Chinese dialects in Liaoning were developed under the dual influence of Youyan dialect and Jiaodong dialect.

From Figure 5, it can be seen that the falling tone form of T1 in the Liaoning dialect is consistent with the dialect on the Jiaodong Peninsula. Shanxi (山西) also has the situation that T1 pronounces as a falling tone. However, the current contour of T1 in Beijing, Tianjin, Henan, Hebei, and Shandong province are not falling form. Most of these regions belong to Zhili Prefecture in Qing Dynasty and they are important sources of immigrants to Liaoning. If the falling tone form of T1 in Northeast Mandarin in Liaoning is completely shaped by immigrant dialects, it is hard to explain why the falling tone form won in the competition, because the number of immigrants from Shanxi (山西) is not dominant, while immigrants from Jiaodong Peninsula mainly settled in the eastern coastal areas of Liaoning. Therefore, there is another source of the falling tone form of T1, which shaped the underlying form of T1 before the arrival of a larger number of immigrants.

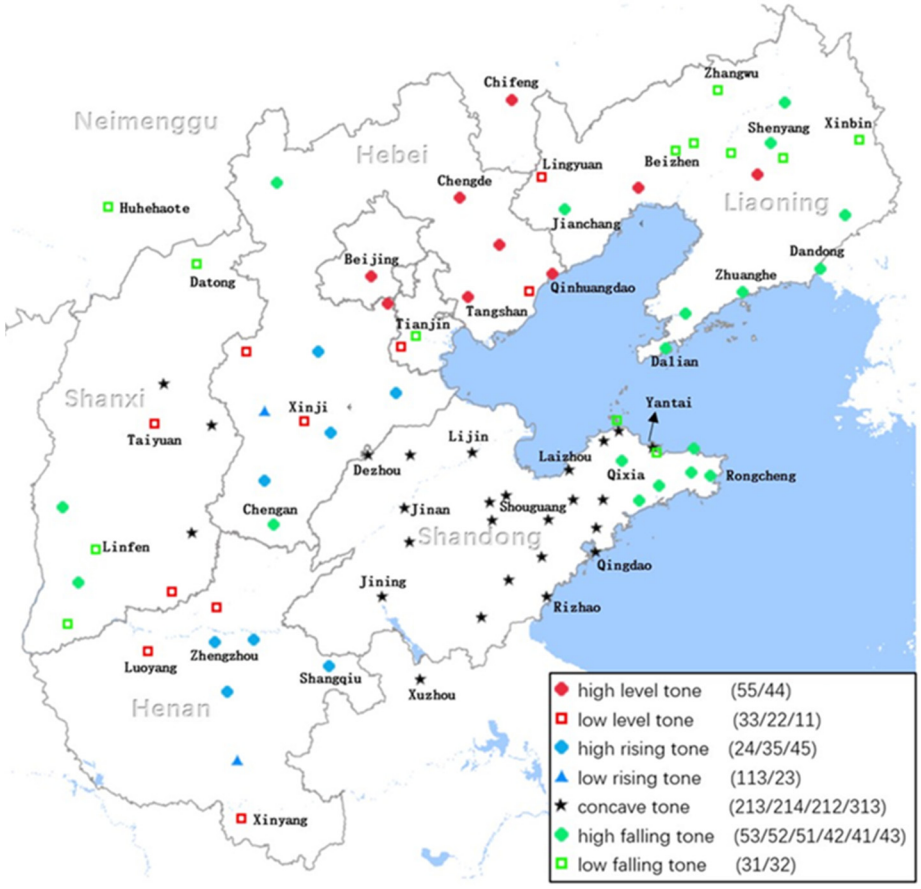


Figure 5: The contour of T1 in the surrounding regions of Liaoning.

The main source is the error made by ethnic minorities when they learned the standard language. To consolidate their dominant position, the ethnic minorities in the Northeast region started to strengthen their connection with Han people in the very early time. They learned Han culture as well as Chinese; therefore, the underlying form of the Liaoning dialect was Mandarin learned by ethnic minorities. Historically, the Khitan, Mohe, Jurchen, and eventually the Manchu were the minority groups who resided in and dominated the Northeast region. The languages they spoke belonged to the Altaic language family, which were stressful and toneless languages. Therefore, the tone was the most difficult for them in Chinese learning. At that time, ethnic minorities tended to learn the standard language, particularly Northern Mandarin and later the Beijing dialect, which was comparable to Putonghua today. For a very long period, T1 in Northern Mandarin maintained its level tone shape despite the tonal system being reconfigured by previous scholars.

As shown in Table 4, Zhongyuan Yinyun reflects the Northern Mandarin in the Yuan dynasty. Professor Wang Li (1957), Yang Naisi[4], and Ji Fu (1986) reconfigured Yinping as a mid level tone. Considering different literatures and different dialects, Professor Zhang Yulai (2017) reconfigured the Mandarian in the Ming dynasty, which is likely used in both north and south of China, in this pattern, T1 is a level tone as well.

Table 4: The tonal system of Northern Mandarin as reconfigured by former scholars.

		T1	T2	T3	T4
Zhongyuan Yinyun (Yuan Dynasty)	Wang Li	*Mid-level tone	*Mid-rising tone	*High-level tone	*Low falling tone
	Yang Naisi	*22	*55	*34	*51
	Ji Fu	*33	*45	*315	*51
Mandarin in Ming Dynasty	Zhang Yulai	*33	*24	*315	*51

The practice of pronouncing T1 as a falling tone is an error made by ethnic minorities when they learned Northern Mandarin. This is because a falling tone is relatively simpler in pronunciation than a level tone. The native speakers of the Altaic language family were most likely to make errors between T2 and T3 when learning Putonghua. The error in the pronunciation of T1 does not affect communication(Ayinuer Y& Liu Xiuming 2016, Hegezhule 2022). Native speakers of a stressful language can distinguish between the two falling tones in tonal pattern, one is the low falling tone T1, and another is the high falling tone T4.

Therefore, the falling tone of T1 was the underlying form in the early Liaoning dialect, which was caused by errors generated by learning and imitation. The underlying form was further reinforced by immigration factors, immigrants from Shanxi (山西) and the Jiaodong Peninsula also pronounced T1 as a falling tone. As a result, the falling tone form remained for a long time in the central Liaoning dialect, which can be observed in some survey points today.

4. The process of "falling tone > level tone" change of T1

Although the downward trend in the last word of disyllable is a regional commonality of Northeast Mandarin, T1 in citation form and the last word of sandhi form sometimes represent a slightly falling tone. The downward trend is unobvious, in some dialect points, T1 even represents a mid level tone in citation form, as shown in Figure 6. These

[4] Quoting from Lin Duan (1992).

phenomena suggest that T1 in Northeast Mandarin is undergoing a contour change, which first appeared in citation form and the first word of sandhi form.

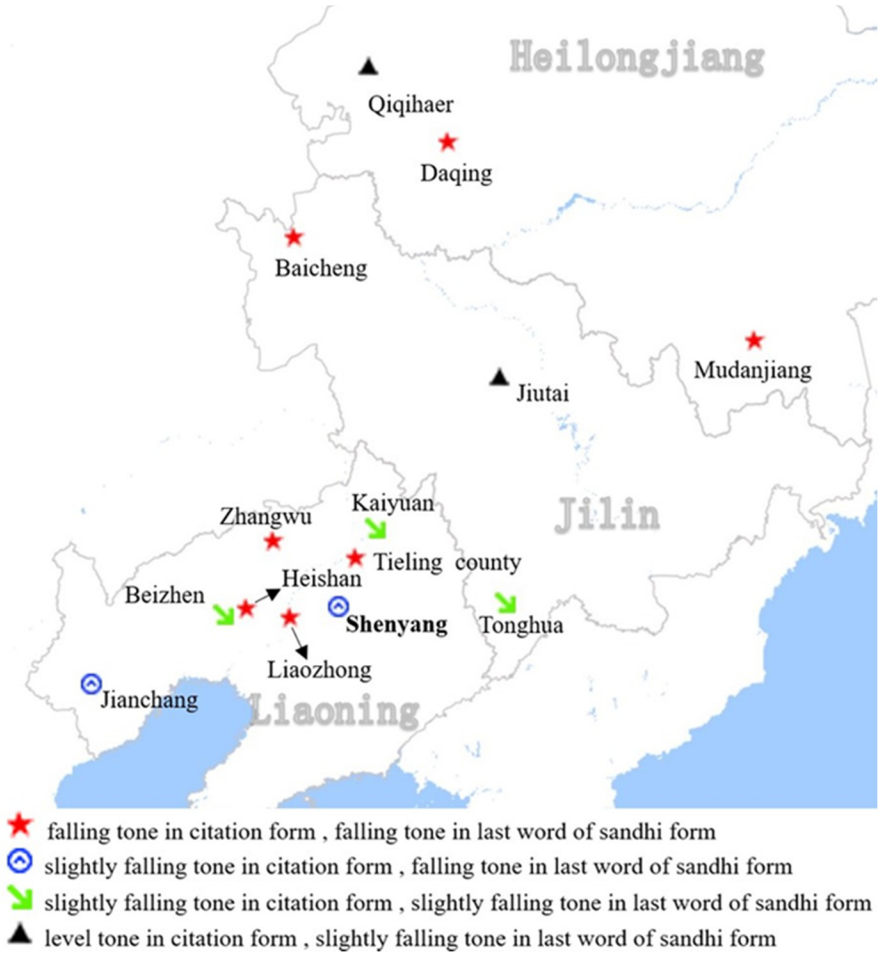


Figure 6: The contour of T1 in citation form and last word of sandhi form.

In citation form, the contour change experiences the "falling tone > slightly falling tone > level tone" process, for example, "31 > 43 > 44", slightly falling tone is a transitional phase. In sandhi form, after being influenced by Standard Mandarin, T1 represents as slightly falling tone or level tone in some environments, the falling tone form only retains in some tone combinations.

Table 5: Tone sandhi in disyllable in the Shenyang dialect.

Last word / First word	T1 (43)	T2 (25)	T3 (212)	T4 (53)	Netual tone (0)
T1 (43)	33+32 飞机 "plane"	33+13 欢迎 "welcome"	33+211 真理 "truth"	33+53 公路 "highway"	44+0 公公 "father in law"
T2 (25)	35+43 平安"safety"	25+14 农民"farmer"	35+211 成语"idiom"	24+53 肥皂"soap"	24+0 娃娃"doll"
T3 (212)	32+33 火车"train"	31+14 火柴"match"	35+211 水果"fruit"	21+53 考试"exam"	311+0 祖宗"ancestors" 35+0 想想"to think"
T4 (53)	53+31 面包 "bread"	53+13 杏仁 "almond"	53+211 电脑 "computer"	42+42 运动 "motion"	53+0 弟弟 "younger brother"

For example, the citation form of T1 in the Shenyang dialect has changed to a slightly falling tone 43, but in the Sandhi form, the falling tone form only persists after T1, T2, and T4, and the earliest form 31 only persists after T4. The retention of falling forms varied in different environments. According to Table 6, in disyllabic words, T1 turns into a level tone when it appears in the first word position. When T1 appears in the last word position, the "falling tone > level tone" transition is more likely to happen after T3 and T4, than after T2. The falling tone form is most stable after T1.

Table 6: The tone sandhi of T1 in different environments of disyllable.[5]

	First word	Last word			
		After T3	After T4	After T2	After T1
Tieling	+	+	+	+	+
Beizhen	+	+	+	+	-
Jianchang	+-	+	+	+	-
Kaiyuan	+	-	+	-	-
Heishan	+-	+	-	-	-
Shenyang	+	+	-	-	-
Liaozhong	+	-	-	-	-
Tieling county	+	-	-	-	-
Zhangwu	+	-	-	-	-

Each survey point is at a different stage of falling tone to level tone shift when the citation form and sandhi form are combined. As shown in Figure 7, Tieling changes the fastest, with T1 becoming a level tone in each position of the disyllable, while

[5] "+" means falling tone changes into level tone in sandhi form, "-" means Yinping's contour doesn't change .

Liaozhong, Tieling county, and Zhangwu change the slowest, with T1 becoming a level tone only in the first position of the word. The impact of Standard Mandarin, or Putonghua, is the primary cause of this transformation. In this process, there are two pulling forces, one is the influence of Beijing Mandarin in the west, whose pronunciation is similar to Putonghua, and the other is the radiation impact of the central city of Shenyang. Therefore, Jianchang and Beizhen located in the west near the Beijing Mandarin region change relatively faster. In the central city of Shenyang, the acceptance of Standard Mandarin is higher and influences nearby survey points. Zhangwu is relatively far from Shenyang, Liaozhong and Tieling county are suburban of Shenyang and Tieling. As a result, these three survey points change the slowest, complying with the peripheral distribution principle, and this explains why older forms often remain in peripheral regions.

Figure 7: The tone sandhi of T1 in the central Liaoning.

5. The characteristic of T2 in Northeast Mandarin in Liaoning

From the overall tone pattern, the special contour of T1 is not the only characteristic of Northeast Mandarin in Liaoning, T2 also has subtle differences compared with Standard Mandarin. From Figure 8, it can be seen that in some central survey points, T2 is a low raising tone, while it is a high raising tone in Standard Mandarin. These survey points almost remain the falling form of T1. Therefore, it is assumed that this

region reserves the underlying characteristics of Northeast Mandarin, in which T2 is likely in a low register.

The characteristic of T2 in Northeast Mandarin in Liaoning is less obvious than that of T1. It is difficult to be noticed in terms of auditory perception and is only described through phonetic experiments. This article suggests that because the characteristic of T2 does not change the contour, and T2 is always the only raising tone in the whole tone pattern, the performance of the register would not affect the recognition of T2.

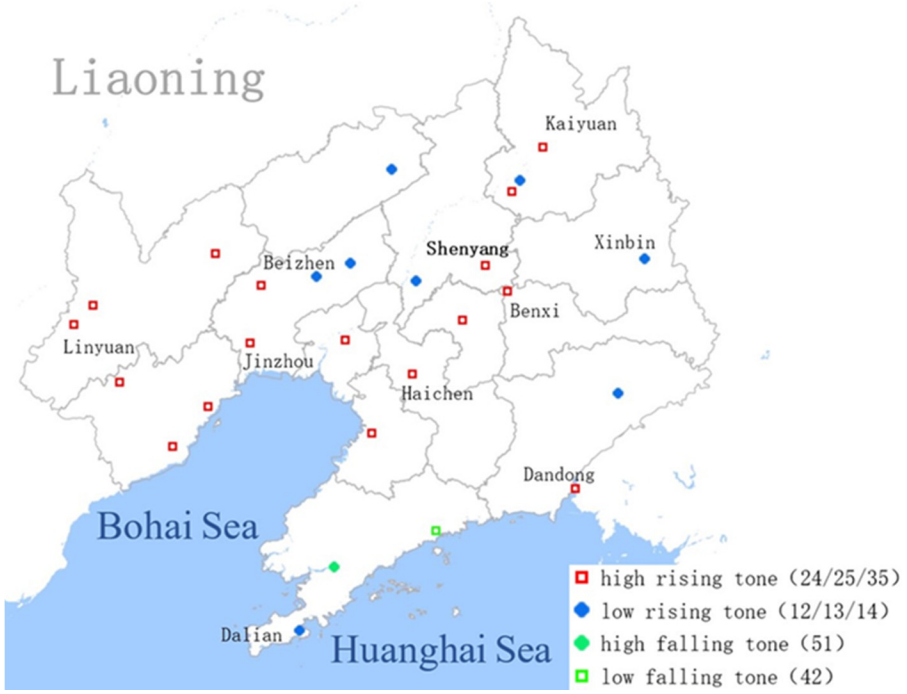


Figure 8. The performance of T2 in Liaoning.

6. Conclusion

This study demonstrates that the most prominent characteristic of Northeast Mandarin in Liaoning is the special performance of T1, which is undergoing an evolution from the older falling tone form to level tone. The major cause of the falling form was an error that ethnic minorities made when studying Northern Mandarin, and later immigrant dialects increased the unique contour. Changes first occurred in certain

combinations of sandhi form. The main reason for changes is the influence of Standard Mandarin. The contact with the western Beijing Mandarin region also has an impact.

Besides, T2 is underlyingly a low rising tone, thus the tonal system is highly biased towards the low register, which is different from Standard Mandarin. The tone pattern in the Liaozhong dialect is like the reconfigured underlying dialect in Liaoning. The left side of Figure 9 shows the tone pattern of Liaozhong, which retained the characteristics of underlying dialects in Liaoning, while the right side shows the tone pattern of the Beijing dialect. Although the contour of each tone is similar, the register pattern and pronunciation of T1 give the Liaoning dialect its distinctive sound. In Liaoning's underlying dialects, T1, T2, and T3 are in a low register, while only T4 is in a high register, and in the Beijing dialect, T1, T2, and T4 are in a high register, only T3 is in a low register.

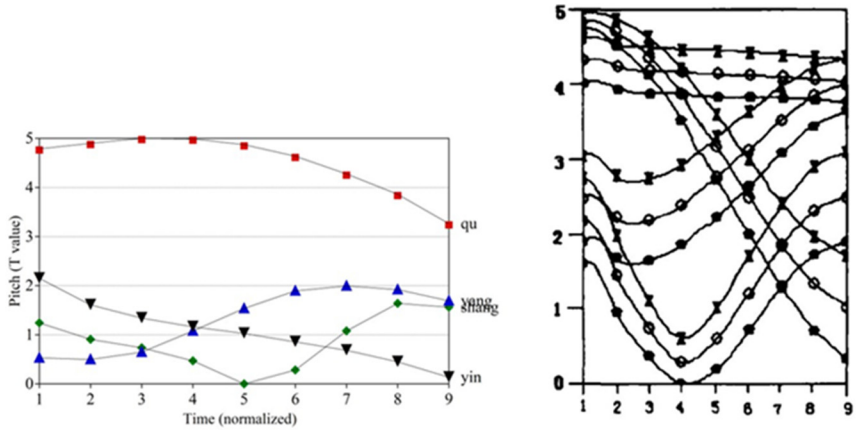


Figure 9: The comparision between underlying dialect in Liaoning and Beijing dialect

References

Ayinuer Yasheng [阿依努尔・亚生], Liu, Xiuming[刘秀明] (2016) Weiwuerzu Daxuesheng Hanyu Danzidiao Xide Pianwu de Shengxue Shiyan Yanjiu 《维吾尔族大学生汉语单字调习得偏误的声学实验研究》 [*An Acoustic Study on Errors Tone Acquisition Among Uyghur Learners of the Chinese Language*].*Research on Bilingual education* 3(01):64-68. doi:10.13953/j.cnki.syjyyj.2016.01.012

Bao, Zhiming (1990) *On the nature of tone*. Cambridge: Massachusetts Institute of Technology.

Bao, Zhiming (1999) *The structure of tone*. New York: Oxford University Press.

Chen, Xiao [陈晓] (2018) Qing Zhonghouqi zhi Minguo Chuqi Beijinghua de Shengdiao Bianhua 《清中后期至民国初期北京话的声调变化》 [*Tonal Variations of the Beijing dialect in Late-Qing Times and the early years of the Republic of China*]. *Studies of the Chinese Language* 4: 435-445+511.

Han, Tianyu [韩天宇] (2018) Heishan Fangyan Yinxi Yanjiu 《黑山方言音系研究》 [*Study on the Phonetic System of Heishan Dialect*]. Dalian: Liaoning Normal University.

Hegezhule [赫格珠乐] (2022) Zhongji Hanyu Shuiping Mengguguo Xuesheng Hanyu Shengdiao Xide Shiyan Yanjiu 《中级汉语水平蒙古国学生汉语声调习得实验研究》 [*Experimental Study on the Acquisition of Chinese Tones by Mongolian Students with Intermediate Chinese Proficiency*]. Tianjin: Tianjin Normal University.

He, Wei [贺巍] (1986) Dongbei Guanhua de Fenqu (Gao) 《东北官话的分区(稿)》 [*Division of Northeast Mandarin (draft)*]. *Dialect* 3: 172-181.

Li, Nanyi [李南依] (2019) Liaoning Panjin Fangyan Yuyin Yanjiu 《辽宁盘锦方言语音研究》 [*The Phonetic Study of Panjin Dialect in Liaoning Province*]. Jinan: Shandong University.

Lin, Duan [林端] (1992) Yuandai Dadu Kouyu de Diaowei Xitong 《元代大都口语的调位系统》 [*The Tone System of Dadu Spoken Language in the Yuan Dynasty*]. *Journal of Xinjiang University(Philosophy and Social Sciences)* 3: 79-86. doi: 10.13568/j.cnki.issn1000-2820.1992.03.014

Liu, Xin [刘欣] (2021) Kaiyuan Fangyan Shuangzidiao Biandiao de Shiyan Yanjiu 《开原方言双字调变调的实验研究》 [*The Experimental Study on the Tone sandhi of Two Characters in Kaiyuan Dialect*]. Dalian: Liaoning Normal University.

Na, Lin [那琳] (2019) Kaiyuan Fangyan Yuyin Yanjiu 《开原方言语音研究》 [*The Study on the Phonetics of Kaiyuan Dialect*]. Jinzhou: Bohai University.

Song, Xue [宋学] (1963) Liaoning Yuyin Shuolue《辽宁语音说略》[*Introduction to Liaoning Phonetics*]. *Studies of the Chinese Language* 2: 104-114.

Wang, Dan [汪丹] (2022) Hasakesitan Xuesheng Hanyu Shengdiao Xide de Yuyin Shiyan Yanjiu《哈萨克斯坦学生汉语声调习得的语音实验研究》[*The Phonetic Experimental Study on the Acquisition of Chinese Tones by Kazakh Students*]. Haikou: Bohai University.

Wang, Qi [王琪] (2016) Liaozhongxian Fangyan Yuyin Yanjiu《辽中县方言语音研究》[*The Phonetic Study of Liaozhong County Dialect*]. Dalian: Liaoning Normal University.

Zhang, Zhimin [张志敏] (2005) Dongbei Guanhua de Fenqu (Gao) 《东北官话的分区(稿)》 [*Division of Northeast Mandarin (draft)*]. *Dialect* 2: 141-148.

Study on the disyllabic third−tone sandhi in Shaanxi Middle Area Mandarin

Xiao, Nengping
(Nankai University, China)

Abstract: This study investigates the disyllabic third-tone sandhi in Shaanxi Middle Area Mandarin. It is a common phenomenon that T3 changes to another tone before T3 in two-syllable words in Mandarin dialects. Shaanxi is the primary region where T3 changes to T1 before T3. In this study, 10 geographical points of Middle Area Mandarin in Shaanxi are explored. Additionally, combining it with closely related types such as Middle Area Mandarin in other places, Jin and Hakka, three types of the disyllabic third-tone sandhi in Shaanxi Middle Area Mandarin are analyzed. We propose that these types reflect different historical strata. The tone change rule, which states that a T3 changes to a T1 before T3, was established during the Tang Dynasty and spread throughout the surrounding regions. And now, this rule is disappearing.
Keywords: Middle Area Mandarin, third-tone sandhi, the neutralization of tone sandhi

1. Introduction

In two-syllable words, it is common for T3 (Shǎngshēng 上声) to shift into a different tone before T3 in Mandarin. Also, different dialects use a variety of tonal changes, such as changing the initial character to T1 (Yīnpíng 阴平), T2 (Yángpíng 阳平), or T4 (Qùshēng 去声), a new tone, or no tonal change at all. The Middle Mandarin region of Shaanxi is where the phenomenon of the preceding syllable's T3 becoming T1 most frequently occurs.[1] Previous studies (Qi & Shi, 1998; Hirayama, 1999; Qi, 2008/2010:160–165; Wu, 2020; Endo, 2021) have mainly focused on the synchronic phonetics and the diachronic tonal value evolution to understand the causes of the disyllabic third-tone sandhi, while little attention has been paid to the geographic

Xiao, Nengping. 2023. Study on the disyllabic third-tone sandhi in Shaanxi Middle Area Mandarin. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 143–153. doi: https://doi.org/10.5281/zenodo.8374642

* This work was funded by the National Social Science Foundation Project of China (20AYY020). The project is entitled "Research on the pattern of compound-character tones in Mandarin Chinese dialects". I would also like to thank the participating experts, including Professor Mitsuaki ENDO, for the valuable advice they gave me at the conference.

[1] Shaanxi is a province in central China, and the main languages spoken within its borders are Middle Area Mandarin, Jin, and Southwest Mandarin.

distribution and their historical causes of the pre-syllable that changes to T1. In this study, the disyllabic third-tone sandhi in Shaanxi Middle Area Mandarin is investigated in terms of its typology and historical strata using a dialect geographic method.

2. Data sources

Most of the content in this essay is sandhi from 10 dialectal locations in the Shaanxi Middle Area of Mandarin. A few of the Xi'an materials were obtained from fieldwork performed in April 2023. The informants were Ms. Zhang and Mr. Zhang, who are a pair of siblings. Ms. Zhang, is a 60-year-old retired employee with a junior high school diploma. Mr. Zhang, a 58-year-old worker with a high school diploma, is currently employed. Their grandparents are residents of Lianhu District, Xi'an, where they both were born and raised. Other data and information were mostly obtained from Chen & Li (1996) and other dialect survey works. Specific sources are listed in the Appendix.

It should be noticed that there are a lot of weak stress cases in the Xi'an dialect. The latter syllable can either be pronounced in its original tone or with weak stress and when it is pronounced with weak stress, the rhyming vowel is easily centralized or even displaced. Although weak stress and neutral tone in the Xi'an dialect are difficult to identify phonetically, they can be identified by the following: neutral tone distinguishes semantics and syntax, while weak stress does not; neutral tone is obligatory, while weak stress is not. The neutral tone in the Xi'an dialect was also reported by Sun (2007) as being a component of tone sandhi. We think that neutral tone and weak stress should both be excluded from tone sandhi in the Xi'an dialect. Several earlier recordings of dialects in the Xi'an region that consider the neutral tone and weak stress as tone sandhi have been omitted from references (Zhang, 1990; Sun, 2017).

3. The geographical distributions of disyllabic third-tone sandhi

According to the tonal category, the disyllabic third-tone sandhi in Shaanxi can be divided into three types as follows:

A. T3+T3→T1+T3

C. T3+T3→T3+T3

AC. T3+T3→T1/T3+T3

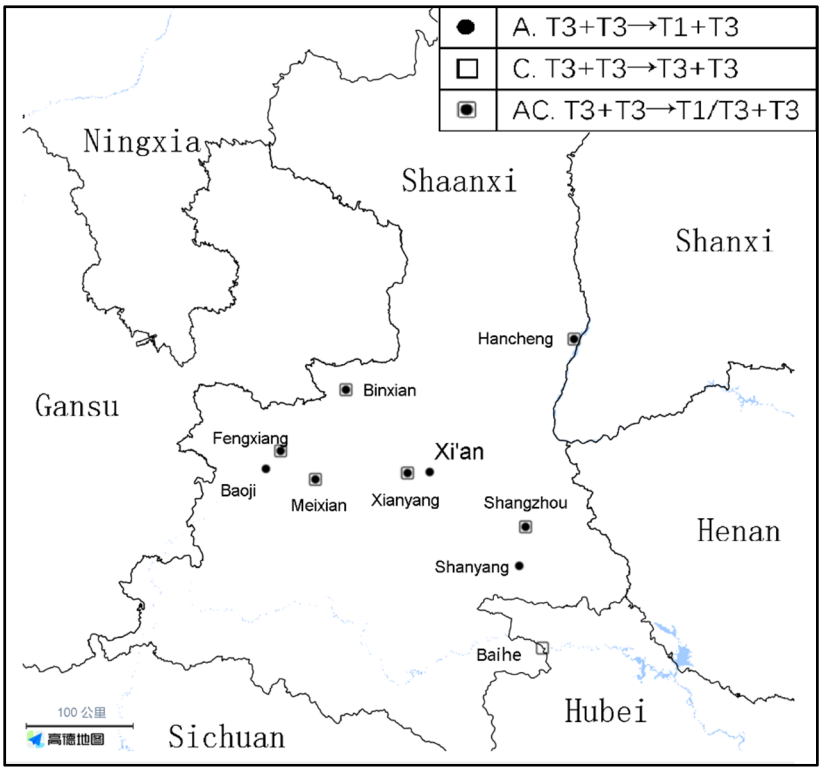


Figure 1: The geographical distributions of disyllabic third tone sandhi in Shaanxi Middle Area Mandarin.[2]

The three types of Shaanxi Middle Area Mandarin's disyllabic third-tone sandhi are distributed geographically in Figure 1.[3] The AC-type is the most prevalent form, followed by the A-type, and only a very small number of dialect points are of the C-type. The main distribution centers for the A-type are in Xi'an, Baoji, and Shanyang. The districts and counties near Xi'an are distributed with the AC-type. Additionally, the C-type is exclusively found in Baihe. Can they assist distinguish between the old

[2] Wu & Zhang (1996:1079-1081) found that in the Xi'an dialect, T3 becomes T1 before T3, however our most recent survey found that T3 before T3 no longer undergoes a tone change. Because T3 is a high descending tone, the coarticulation affects a few words and causes them to become high flat, which is tonally the same as T4. A comparable situation can be observed in the newly notated data of Shaanxi's Shanyang dialect (Zhao, 2022). This should be a recent change, and for historical comparison with other dialects, we will continue to classify the Xi'an and Shanyang dialects as A-type in this Figure.

[3] The map in this research was created utilizing The Linguistic of GIS, which was created by Nankai University and Tianjin Xinhui Network Technology Service Center, and the basis map was taken from Gaode Map (2023AutoNavi-GS (2021)6375).

and new forms based on their geographic distribution? In the following sections, we attempt to explain the causes and historical strata of each of these types separately.

3.1. The AC-type

The AC-type is a combination of A and C types. It results from the coexistence of old and modern strata, with the A-type denoting the former and C-type denoting the latter. Geographically, the A and AC types in Xi'an have a concentric distribution; the A type is in the center, while the AC-type is on the outskirts. The AC-type has a much greater range when we look at the entire Middle Area Mandarin, with Xining in Qinghai and Qingshui and Tianshui in Gansu all being AC-type.[4] The circumference of such a concentric distribution around Xi'an is substantially larger. The tone sandhi of the entire Middle Area Mandarin disappearing. It can be explained as follows: first, only a small number of tonal combinations experience tone sandhi, whereas the majority do not. Second, only certain words that share the same tonal combinations are subject to tone sandhi, while other words do not (Zhang, 1990). As a result, the disyllabic third-tone sandhi is not an exception. However, because of the gradual diffusion of the C-type in the lexicon, some dialectal points continued to use the AC-type. The AC-type is an innovative form of the A-type and the C-type represents the end point of its evolution. It is a process in the disappearance of tone sandhi.

3.2. The A-type

The disparity in the rates of evolution of citation tone and sandhi tone served as the original intrinsic motivation for the A-type, and the inverse dissimilation was induced by the bass component in T3 (Hirayama, 1999). However, when did it start to spread? What was the transmission path? Its distribution may provide some hints.

The present A-type and AC-type distribution ranges should be included in the early A-type range of distribution as the A-type was the predecessor to the AC-type. If the Middle Area Mandarin in Shanxi is taken into account[5], it is mainly concentrated in the South Xinjiang region, the Qinlong region, the Weihe Plain of Shaanxi, and the Sushui River Basin of Shanxi. It is important to note that the A-type is also present in the Jin, Hakka, and Jilu Mandarin in addition to the aforementioned regions:

1) The Jin dialect

[4] For tone sandhi in Qingshui, Gansu, see Cao (2021), and for tone sandhi in Tianshui, see Zhang & Deng (2008).

[5] The Middle Area Mandarin within Shanxi is mainly centered in Southern part of Shanxi, where it borders Shaanxi, and the cities in this region, Yuncheng, Hejin, Wanrong, Yongji, Pinglu, and Linyi, all have the A-type, which means that the T3 becomes T1 before T3 in all cities mentioned above.

Li (2014) analyzed and compiled tone sandhi material for 37 Jin dialect points. Table 1 displays the Jin dialects' third-tone sandhi. In addition to the several forms of Mandarin spoken in the Shaanxi Middle area, there are five types of Jin, including B, AB, BC, E, and D type. The B-type signifies T3 turning into T2 before T3. The AB-type indicates that T1 and T3 change to T2 before T3. The BC-type refers to the situation in which T3 changes to T2 before T3 in some cases, but not in others. The E-type indicates that T1 (Yīnpíng 阴平) and T2 (Yángpíng 阳平) merged in citation tone and sandhi tone, resulting in the preceding syllable becoming Píngshēng (平声). Additionally, the D-type indicates that the second syllable has been altered.

Table 1: Types of disyllabic third-tone sandhi in the Shanxi Jin dialect.

Types		Point	Count	Proportion
A-type	T3+T3→T1+T3	Yu County, Zhongyang, Fangshan, Daning, Xinzhou, Dingxiang, Wutai, Dai County, Shuozhou, Pinglu, Hunyuan	11	29.73%
B-type	T3+T3→T2+T3	Ying County	1	2.7%
C-type	T3+T3→T3+T3	Xiaoyi, Jingle, Fenxi, Wuxiang, Xiangyuan, Qinxian, Yangcheng, Gaoping	8	21.62%
AB-type	T1/T3+T3→T2+T3	Yonghe, Datong, Tianzhen, Huairen, Youyu	5	13.51%
AC-type	T3+T3→T1/T3+T3	Xing County	1	2.7%
BC-type	T3+T3→T2/T3+T3	Wenshui	1	2.7%
E-type	T3+T3→Píngshēng +T3	Taiyuan Northern Suburbs, Shouyang, Yushe, Pingyao	4	10.81%
D-type		Jiexiu, Linxian, Lingchuan, Jincheng, Tunliu, Changzhi	6	16.22%
Total			37	100%

Table 1 shows that except for the six points of the D-type, 17 of the remaining 31 points (A, AC, and AB types) exhibit signs of the rule that T3 changes to T1 before T3. The A and AC types, do not need to be mentioned. The AB-type, which is mostly found in northern Shanxi, is thought to be the consequence of the merging of T1 and T3 before T3, followed by the merging with T2. Jin's internal evolutionary rule is the merging of T1 and T3, and the merging with T2 is influenced by Beijing Mandarin's rule. T3 changes to T2 before T3 in Beijing Mandarin. Briefly, there are many dialect points in the Jin dialect area where the former syllable has changed to T1.

2) Hakka

There are six citation tones in the Jiangxi Yudu Hakka, including Yīnpíng(阴平), Yángpíng (阳平), Shǎngshēng (上声), Yīnqù (阴去), Yángqù (阳去), and Rùshēng(入声), and Shǎngshēng (T3) becomes Yīnpíng (T1) before Shǎngshēng (T3) (Xie,

1992/1998: 5). Some Quánzhuóshǎng (全浊上) in Meixian Hakka changed to Yīnpíng when read in conjunction with the Shǎngshēng (Huang, 1988). The value of the citation tone of Yīnpíng (T1) in Jiangxi Dingnan Hakka is 35, and the former syllable changes to low Yīnpíng 13 when the latter syllable is Shǎngshēng (Xu, 2019). Furthermore, according to Huang (1988), one significant feature of Hakka that sets it apart from other dialects is that several characters, which were originally Voiced Shǎng(浊上) characters, are now read as Yīnpíng characters. The alternation of T3 and T1 in citation tones also tends to suggest that the early rule T3+T3→T1+T3 may have been more common in Hakka.

3) Jilu Mandarin

Hirayama (1999/2005) reported that Pinggu and Xiongxian's disyllabic third tone is A-type. Additionally, T1(45) and T3(214) merge before T3(214) in Baoding and T1(33) and T3(213) form T2(35) before T3(213) in Dingxing (Chen & Xu, 1997)[6]. The A-type is distributed sporadically in the Jilu Mandarin region. The Jingnan Incident[7] caused a significant influx of immigrants into the Hebei province, however Baoding, in that province, was not the main destination for those immigrants (Cao, 1997). As a result, the Baoding dialect is still preserving today and is T3+T3→T1+T3, whereas most of the nearby areas are T3+T3→T2+T3.

The A-type geographic distribution evidenced that the T3+T3→T1+T3 is a remnant of an ancient form. As early as the Tang Dynasty, the A-type was completed and dispersed. This is a result of four factors as follows:

First, the Hú rén (胡人) dialect of Chinese gained prestige in the north during the Western Jin Dynasty following the upheaval of the Five Barbarians(五胡乱华). Emperor Xiaowen of the Northern Wei Dynasty (467-499) transferred the capital from Pingcheng (present-day Datong, Shanxi) to Luoyang and implemented the orthography. Then, when the Northern Wei Dynasty came to an end, or divided, one of the branches transferred west to Chang'an (present-day Xi'an, Shaanxi). Later, the Sui Dynasty united the north and south of China, and Chang'an gained political clout (Yang & Huang, 2023). The existence of constant sandhi tones in Shanxi and Guanzhong is closely tied to the Northern Wei dynasty's language policy.

Second, the Tang dynasty established Chang'an as the center of politics, economy, and culture, and the new literary readings based on the Lingua franca of the northern

[6] Numbers in parentheses are tonal values.

[7] The Jingnan Incident(靖难之变) was a war waged by Zhu Di朱棣(1360-1424) against Emperor Jianwen建文帝(1377-？) to seize power from 1399-1402.

Guanzhong region, with the Chang'an accent as the standard, eventually replaced Qiēyùn《切韵》 as the Tang dynasty's standard literary readings in the Táng Xuánzōng 唐玄宗 dynasty (Yang, 2023).

Third, both Quanzhuoshang (全浊上) and Cizhuoshang (次浊上) are pronounced Yīnpíng (阴平) in Hakka. This is because the rule that the Shangshang (上声) becomes Yīnpíng (阴平) before the Shǎngshēng occurred before the tone rule Change from Voiced Shang (上) to Voiced Qu (去), which originated at the end of the seventh century and was frequently used during the Song dynasty (Liu, 1997; Ding, 2005; Xia, 2019).

Fourth, evidence of population migration exists. According to Yudu County's extant genealogical records and the county census of geographical names, most of the ancestors of the various surnames in Yudu County traveled from the north during the Tang Dynasty, although most of them eventually settled in Yudu after a few twists and turns (Xie, 1998: 3-4).

Therefore, Shanxi is the place where the A-type was first bred. The Northern Wei's language policy spread the A-type in the northern Jin dialect to another northern Mandarin. However, when the Xi'an dialect became the primary dialect of the official lingua franca, the A-type became centralized in Xi'an and extended to the Guanzhong, Qinlong, and Shanxi regions as well as the peripheral Mandarin areas. One way that the A-type directly influenced the dialects of the various regions was through the migration from Guanzhong, and another way was through the lingua franca. Naturally, as the Heluo region grew, this influence slowly shifted to the west of the Hangu Pass. The Central Plains accent was retained by the Hakka ancestors who were impacted by the A-type as they moved south. As a result, several of the current Hakka dialects still have typical characteristics of the A-type. Because of Shanxi's mountains and cross-shaped geography, the traditional Mandarin dialect has survived.

3.3. The C-type

Baihe's C-type has anything to do with language contact. Baihe is in southern Shaanxi, and its dialect has been influenced by both Middle Area Mandarin and Southwest Mandarin over time (Ke, 2004). It is most likely the result of simplification following contact between the two dialects that the disyllabic third-tone sandhi does not change. Meanwhile, data shows that what was once an A-type eventually changed to a C-type. In Xi'an, for example, Wu & Zhang (1996) and Wang (1996) specifically described the Xi'an dialect, in which the former T3 becomes T1 before T3, but Ren (2012) and the present fieldwork have found that T3 does not change. This single-point chronological record demonstrates that the tone sandhi of Middle Area Mandarin is rapidly

disappearing. Sandhi's absence has become frequent in all dialect points of Middle Area Mandarin.

4. Conclusion

The findings of this study suggest that the A-type of Shaanxi Middle Area Mandarin originated in the early dialects of northern Shanxi. The Northern Wei Dynasty's language policy profoundly influenced the Guanzhong region, which later spread throughout the nation because of literary readings centered in Chang'an. Xi'an is a key center of influence for the Mandarin of the Middle Area. Dialect interaction gives rise to the marginal C-type, but co-innovation gives rise to the C-type that has emerged in the last 30 years. The AC-type is a transitional stage in which the A-type inherent sound becomes C-type.

References

Cao, Shuji [曹树基] (1997) Zhongguo Yiminshi Diwujuan《中国移民史（第5卷）》[History of Chinese Immigration (Volume V)]. Fuzhou: Fujian People's Publishing House.

Cao, Xinglong [曹兴隆](2021) Gansu Qingshui Fangyan Liangzizu Liandubiandiao《甘肃清水方言两字组连读变调》[Disyllabic Tone Sandhi of Qingshui Dialect in Gansu Province]. *Fangyan* 43(02):223-227.

Chen, Shujing [陈淑静], & Xu, Jianzhong [许建中] (1997) *Dingxing Fangyan*《定兴方言》[*Dingxing dialect*]. Beijing: Local Records Publishing.

Ding, Zhiming [丁治民] (2005) Zhuoshangbianqu Jianyu Beisong Kao《浊上变去见于北宋考》[Tone Change from Voiced Shang(上) to Voiced Qu(去) seen in the Northern Song dynasty]. *Studies of the Chinese Language* 2: 172-176.

Endo, Mitsuaki [遠藤光暁] (2021) Santou hougen keiseizen henchou no chiri bunpu『山東方言軽声前変調の地理分布』[Geographical Distribution of Tone Sandhi before Neutral Tone in the Shandong Dialects]. *The economic review* 13: 131-143. doi: https://doi.org/10.34321/21864

Guan, Xiao [关肖] (2019) *Yongshou(Diantouzhen) Fangyan Yinxi Yanjiu*《永寿（店头）方言音系研究》[*A Study on the Phonological system of Yongshou (Diantou) dialect*]. Literature College of Northwest Normal University. doi: https://doi.org/10.27410/d.cnki.gxbfu.2019.000785.

Huang, Xuezhen [黄雪贞] (1988) *Kejia Fangyan Shengdiao De Tedian*《客家方言声调的特点》[*Characteristics of Hakka Tones*]. *Fangyan* 4: 241-246.

Hirayama, Hisao [平山久雄] (1999 /2005) Guanhua He JinyuFangyan Zhong Shangshangbiandiao De Leixing Jiqi Chengyin《官话和晋语方言中“上上变调”的类型及其成因》. in Hirayama, Hisao[平山久雄] *Pingshanjiuxiong Yuyanxue Lunwenji*《平山久雄语言学论文集》[*Hirayama Hisao Linguistics Proceedings*]. Beijing: The Commercial Press.

Ke, Xigang [柯西钢] (2004) Baihe Fangyan Shengdiao Lueshuo《白河方言声调略说》[A Brief Introduction to Baihe Accent]. *Journal of Ankang Teachers College* 5:53-56+62. doi: https://doi.org/10.16858/j.issn.1674-0092.2004.05.015.

Li, Xi [李晰] (2014) *Shanxi Fangyan Shengdiao Yanjiu*《山西方言声调研究》[*Research on the Tones of Shanxi Dialect*]. Shaanxi Normal University.

Liu, Guanxin [刘纶鑫] (1997) Zhuoshangbianqu Jianyu Nansong Kao《浊上变去见于南宋考》[Tone Change from Voiced Shang(上) to Voiced Qu(去) seen in the Southern Song dynasty]. *Studies of the Chinese Language* 1: 63-66.

Pan, Jiayi [潘家懿] (1990) *Linfen Fangyanzhi* 《临汾方言志》*[Local gazetteers of Linfen*]. Beijing: Yuwen Press.

Qi, Haifeng [亓海峰], & Shi, Feng [石锋] (1998) Guanyu Beifanghua Liandubiandiao De Leixing《关于北方话连读变调的类型》. *Studies in Language and Linguistics* Suppl.

Qi, Haifeng [亓海峰] (2008/2010) *Laiwu Fangyan Yuyin Yanjiu*《莱芜方言语音研究》. Nankai University; Beijing: Chinese Literature and History Press.

Qiao, Guangming [乔光明], & Chao, Baotong[晁保通] (2002) Binxian Fangyan Tongyin zihui 《彬县方言同音字汇》[Homophony Syllabary of Bin County]. *Fangyan* 3:260-271.

Ren, Jie [任洁] (2012) *Jiyu EGG de Guangzhou Jixi Xi'an Fangyan Shengdiao Shiyanyanjiu* 《基于EGG的广州、绩溪、西安方言声调实验研究》[*An experimental study of Guangzhou, Jixi, and Xi'an dialects' tones based on EGG*]. Nanjing Normal University.

Sun, Lixin [孙立新] (2007) *Xi'an Fangyan Yanjiu*《西安方言研究》[*Studies on Xi'an Dialects*]. Xi'an: Xi'an Press.

Sun, Lixin [孙立新] (2017) Meixian Fangyan Liangzizu De Liandubiandiao《眉县方言两字组的连读变调》[Disyllabic Tone Sandhi of Meixian dialect]. *Journal of Xianyang Normal University* 1:40-42+79.

Wang, Junhu [王军虎] (1996) *Xi'an Fangyan Cidian*《西安方言词典》[*Xi'an Dialect Dictionary*]. Nanjin: Jiangsu Education Publishing House.

Wang, Junhu [王军虎] (2010) Shaanxi Fengxiang Fangyan Liangzizu De Liandiao Moshi《陕西凤翔方言两字组的连调模式》[Disyllable Tone Sandhi Patterns of Fengxiang Dialect in Shaanxi Province]. *Fangyan* 3:250-254.

Wu, Tianhui [吴天慧], & Zhang, Ansheng [张安生] (1996) Xi'an Yinxi《西安音系》[The phonology of Xi'an dialect], in Chen, Zhangtai [陈章太], & Li, Xingjian [李行健] (eds). *Putonghua Jichu Fangyan Jiben Cihuiji*《普通话基础方言基本词汇集》[*The Basic Vocabularies of Mandarin Basic Dialects*] : 1079-1081. Beijing: Language & Culture Press.

Wu, Yonghuan [吴永焕] (2020) On the Types and Causes of Disyllabic Third Tone Sandhi in Mandarin Dialects《官话方言双音节连上变调的类型及成因》. *Linguistic Sciences* 2: 207-218.

Xia, Liping [夏俐萍] (2019) Zhuoshangbianqu Yu Zhuoyinqinghua《"浊上变去"与"浊音清化"》[Tone Change from Voiced Shang(上) to Voiced Qu(去) and Initial Change of Devoicing Process]. *Studies in Language and Linguistics* 1:76-83.

Xie, Liuwen [谢留文] (1998) *Yudu Fangyan Cidian*《于都方言词典》[*Yudu Dialect Dictionary*]. Nanjing: Jiangsu Education Publishing House.

Xu, Cuizhen [徐萃珍] (2019) *Dingnan Fangyan Yuyin Yanjiu*《定南方言语音研究》[*Phonological study of the Dingnan dialect*]. Liaoning Normal University.

Yang, Jun [杨军], & Huang, Xiaoshan [黄笑山] (2023) *Jingdianshiwen* De Chuanshan Jiuchan Ji Zhonggu Liangmu De Liubian《<经典释文>的船禅纠缠及中古两母的流变》. Di'erjie Zhongguo Yinyunxue Shuqi Jiangxiban Jiangyi《第二届中国音韵学暑期讲习班讲义》[Handout for the Second Summer Workshop on Chinese Phonetics]. 2023.7.15—20, Guangxi University.

Yang, Xiaoan [杨晓安] (1996) Baihe Yinxi《白河音系》[The phonology of Baihe dialect], in Chen, Zhangtai [陈章太], & Li, Xingjian [李行健] (eds). *Putonghua Jichu Fangyan Jiben Cihuiji*《普通话基础方言基本词汇集》[*The Basic Vocabularies of Mandarin Basic Dialects*] : 1039-1040. Beijing: Language & Culture Press.

Yu, Yongmin [于永敏] (2008) *Hancheng Fangyan Yuyin Yanjiu*《韩城方言语音研究》[*Phonological Study of the Hancheng Dialect*]. Shaanxi Normal University.

Zhang, Chengcai [张成材] (1990) *Shangxian Fangyanzhi*《商县方言志》[*Local gazetteers of Shangxian*]. Beijing: Yuwen Press.

Zhang, Wenxuan [张文轩], & Deng, Wenjing [邓文靖] (2008) Sanshengdiao Fangyan Tianshuihua De Liangzizu Liandubiandiao《三声调方言天水话的两字组连读变调》[The Tone Sandhi of disyllable of Three- tone Dialect in Tianshui]. *Journal of School of Chinese Language and Culture Nanjing Normal University* 4: 178-183+81.

Zhao, Pingjun [赵萍君] (2022) Shanyangxian Zhongyuanguanhua Guanzhongpian Fangyan De Liandubiandiao He Qingsheng《山阳县中原官话关中片方言的两字组连读变调和轻声》[On Sandhi and Neutral Tone of Disyllable in Guanzhong Dialect of Northern Mandarin in Shanyang County]. *Journal of Shangluo University* 3: 25-28+68. doi: https://doi.org/10.13440/j.slxy.1674-0033.2022.03.010

Appendix

List of Used Data

Dialect points	Data sources
Xi’an	Wu & Zhang（1996）
Shangzhou	Zhang（1990）
Meixian	Sun（2017）
Binxian	Qiao & Chao（2002）
Shanyang	Zhao（2022）
Xianyang	Guan（2019）
Baihe	Yang（1996）
Baoji	Nie（1996）
Fengxiang	Wang（2010）
Hancheng	Yu（2008）
Linfen	Pan（1990）
Yuncheng, Xinjiang, Yuanqu, Jixian, Hejin, Wanrong, Yongji, Pinglu, Linyi, Hongdong, Huozhou, Fushan and Yicheng	Li（2014）

Lateral Consonants in Upper and Downstream Dialects of Perak: A Geolinguistic Analysis

Khairul Ashraaf Saari (Universiti Kebangsaan Malaysia)
Nor Hashimah Jalaluddin (Universiti Kebangsaan Malaysia)
Fazal Mohamed Mohamed Sultan (Universiti Kebangsaan Malaysia)

Abstract: The development of language in Malayonesia started with river waterways. This waterway has become the main source of communication for a community in carrying out social, economic, political and other activities. The dynamic interaction between the river's upstream and downstream regions has given an impact towards the phonetic alterations in the studied area. This paper evaluates how the liquid segment has changed in the coda position in Perak's upstream and downstream regions. This paper will use the Generative Phonology Theory to discuss the distribution of dialects in the research region using Geospatial Information System (GIS) technology. In total, 1585 informants from 103 villages were selected to participate in this research. The findings revealed that the liquid segment of Perak's downstream and upstream dialects had undergone alterations due to consonant deletion, segment mergers, and changes to features in the coda position. This article also successfully developed two distribution isoglosses in the state of Perak by classifying the separation of dialect in the downstream and upstream dialects of Perak.

Key words: Downstream and Upstream dialect, Perak River, Lateral, Geolinguistics, Isogloss map, Dialect

1. Introduction

A dialect is a variety of a dynamic language that frequently changes in response to time and circumstances. According to Syahreni Siregar (2017), language variety can be referred to as various forms of language triggered by social factors that may change from region to another region, from one community to another community, from an individual to another individual, and from a situation to another situation. Language variation can be seen through the spread of dialect and the absorption of new dialects in certain area. In addition, language variation can be seen through the level of retention and change of the dialect itself among native speakers. According to Kazuko Matsumoto (2019), language variation has been structured by speakers' social

SAARI, Khairul Ashraaf, Nor Hashimah JALALUDDIN & Fazal Mohamed Mohamed SULTAN. 2023. Lateral Consonants in Upper and Downstream Dialects of Perak: A Geolinguistic Analysis. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 154–171. doi: https://doi.org/10.5281/zenodo.8374663

characteristics in their complex real communities, by the characteristics of their addressees and by the structural characteristics of the language.

Rokhman (2000) has outlined four factors that can influence language variation which are (1) Social factor, (2) Cultural factor, and (3) Situational factor while Paul Kerswill (2012) believed 2 social factors that can affect language variation are (1) Social class factor and (2) Ethnicity factor. Next, Dwi Atmawati (2018) suggests 2 factors of language variation which include (1) Structural factor which is associated to linguistic aspect and (2) Social factor which refers to the relationship of the role between speaker and listener, education, age, sex, place, purpose of speaking and social status. The classification regarding the language variation factor is important in unraveling the cause of language variation in a particular place. However, the information related to topographical information was not emphasized by previous researchers in discussing this factor of language variation. However, in the study of geolinguistics, the discussion regarding geographical information is very important in describing the level of dialect diversity based on topographical information, land contours, landforms and many other factors. Therefore, researchers have suggested the factors of language variation through two branches: (1) geographic factors and (2) social interaction factors.

Geographical factors can be referred as changes from the point of view of the residential area, which involves the shape of the terrain of an area. This topographical information can be a starting point for the classification of dialect distribution. This is referring to the existence of rolling hills, river waterways, height of land contours and others. This view is supported by a study by Marlyna Maros (2010) in which she stated that dialects in the upstream area of the Pahang river are uttered faster when compared to the downstream area due to the wideness and rapidity of the waterways in the Pahang river; Nor Hashimah Jalaluddin et. al. (2017) proved that the spread of the Pattani Malay dialect and Langkawi Malay dialect can be divided through the row of Gunung Raya and Gunung Mat Chincang in Langkawi; and Matthias Urban (2020) who argued that highlands and lowlands will exhibit different phonological forms. Through several proofs, this can then be used as a basis for the argument that the language variation factor can be based on geographical factors.

Unlike the geographical factor which refers to the acquisition of a dialect that develops naturally, this social interaction factor is the result of exposure during social activities. This process can occur in a forced or consensual scenario that can affect the level of dialect distribution in an area. This social interaction factor can be seen through several external domains such as employment domain, migration domain, education

domain, war domain, inter-marriage domain, commerce domain, colonization domain, individual domain (gender, age and identity) and others. Those domains will be a factor in the spread of dialects from one area to another characterized by social interaction with certain communities. According to Shalizaini Uyob & Che Ibrahim Salleh (2017) stated that, human nature can be embodied through the language of speech that is the nature of sharing which becomes unity in social life. For example, Malaysia is the outcome of a collaboration that has made the Malay language the primary means of communication, further reflecting the identity and unity of a multicultural country.

A range of dialect variations will be employed in the family language, which will then be inherited by the next generation of the family as a native language that is widely used in a community. In discussing geolinguistic studies, understanding change factors of language is important in unraveling the cause of this language variation revolution. This dialect is spreading rapidly and has the capacity to traverse national borders. So, the discussion involving the geographical elements of the area is not based on political demarcation but what is important is to ensure that the discussion of this study is able to observe every angle of the study, especially involving native dialects in the upstream and downstream areas of Malay-songai. According to Ahmad Dahlan (2015); Hamidy UU. (2011) waterways in Malaysia (connected between the upstream and downstream areas of the river) are not only capable of being used as a source of commerce routes or fishing activities but also refer to the Malay people's way of life. The connection between the life of the Malay community is very close to the river waterways in the upstream and downstream areas and will be thoroughly discussed through this study. Therefore, this study will discuss the spread of upstream and downstream dialects along the Perak River through a lateral study of the coda position of the final syllable.

2. Upper and Downstream of the Perak River

Perak is the second largest state in Peninsular Malaysia that is popularized by the diversity of ethnics, culture , and linguistics. Based on the record by the Department of Statistic of Malaysia, there was a 2.5 million population in Perak in 2020. Based on the numbers, it has proven that Perak has a high population and is indirectly able to demonstrate the spread of various dialects. According to Asmah Haji Omar (2008), there were 7 dialects that are being used in Perak right now. Three of them were the origin dialects that were separated in Perak such as (1) Parit dialect, (2) Gopeng dialect and (3) Kuala Kangsar dialect. The other four dialects are dialects that had been

influenced by other regions such as (4) Pattani dialect from South Thailand, (5) Rawa dialect from Indonesia, (6) Kedah dialect and (7) Selangor dialect. These seven dialects have been widely used in the state of Perak including in the upstream and downstream areas of the Perak River. This indirectly showcases the diversity of dialects in Perak.

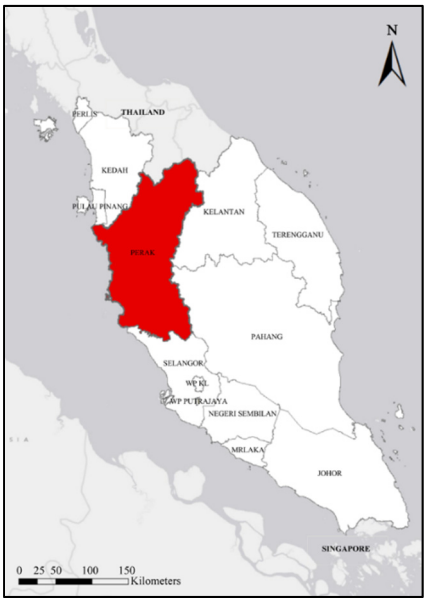


Figure 1 : Location of Perak State
Source : Norzawati Yoep et. al. (2015) Spatio-Temporal Distribution of Malaria in Perak, Malaysia

The state of Perak is one of the states that has a wider range of interaction than other states in Malaysia due to the bordering of the northeastern part with Yala and Narathiwat, Thailand while the western part is bordered with the Strait of Malacca which is famous for its trading activities in the early 15th century. This indirectly demonstrates the friendly community interaction, especially along the Sg Perak estuary. According to Mokhtar Saidin and Jeffrey Abdullah (2007) the Perak River has been the main choice of migration area for the early prehistoric societies of Malaysia (along the downstream and upstream of the Perak River) which has been inhabited since 100,000-200,000 years ago until now. The relationship between the upstream and downstream areas has been built for a long time in the river waterways of Malayonesia. According to Mohd Tarmizi Haran (2018a), this dynamic relationship between the upstream and downstream areas has shaped the civilization politically, economically, and socially in Malayonesia through the flow of raw materials (for example forest products,

agricultural products and mining products) from the upstream and downstream of the river.

Through the development of early civilizations, especially around the upstream and downstream areas of Perak River, the process of communication between local residents has been created. Mohd Tarmizi Hasran (2018b) believed that the diversity and special characteristics of these two areas (upstream and downstream) can be distinguished through the study of dialectal phonemes. The diversity of upstream and downstream dialects of this river is able to exhibit different linguistic forms in describing the geolinguistic analysis in the studied area. So this study will focus on the discussion involving the use of liquids /l/ and /r/ coda position segments along the downstream and upstream areas of Sungai Perak.

3. Research Methodology

Sungai Perak is a river that flows through several traditional villages, houses of worships, historic buildings and locations, sultans' residence palaces, and royal tombs that have stored a variety of nostalgia and historical stories of heroism, loyalty, defiance, institutional glory reign, as well as economic growth (Faizah Mohd Fakhruddin 2022). The Perak River is the second longest river in Peninsular Malaysia which is about 400 kilometers long from the northern part of upstream Perak to the south which is the Bangan Datoh district (the Perak River estuary area). In addition, the discussion regarding the methodology of this study is divided into two parts which are, (1) Data collection process and (2) Data analysis process.

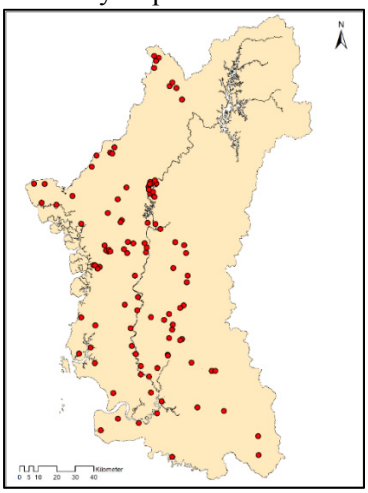


Figure 2: 103 Villages in the study area

This data collection process is done through fieldwork method that covers 103 villages along the Perak River starting from the upstream to the downstream of the river. A total of 1,585 informants were selected as informants representing three (3) age groups; teenagers, adults and the elderly. On average, 15 informants will be selected from each village in the study area. This fieldwork will apply two (2) data collection techniques, namely a questionnaire and a semi-structured interview in discussing the distribution of dialects in the study area. A total of four (4) lexical items have been selected to describe the consonant liquids /l/ and /r/ in the coda position of the syllable, 'cakar', 'sand', 'bantal' and 'kəcil'. The following is a representation of the study data:

Table 1: Display of study data

Standard Malay Language	**Downstream Perak Dialect**	**Upstream Perak Dialect**
/ʧa.kar/ *claw*	[ʧa.kɑ]	[ʧa.kɔ]
/pa.sir/ *sand*	[pɑ.sɛ]	[pɑ.se]
/ban.tal/ *pillow*	[bɑ.tɑ]	[ban.tɑ]
/kə.ʧil/ *small*	[kə.ʧiʔ]	[kə.ʧi]

From the data in Table 1, this data will be discussed in detail through the second methodological process which is the Data Analysis Process. The data analysis process is divided into two analyses, namely (1) Linguistics Analysis and (2) Non-Linguistics Analysis. The Linguistics Analysis will focus on the discussion involving changes in phonological data by using the Generative Phonology Theory which has been developed by Noam Chomsky and Morris Halle (1968) through the publication of his book "The Sound Pattern of English (SPE)". According to Chapman and Raoulege (2009) in Fareed Hameed Al-Hindawi (2018) stated that "Generative Phonology Theory is an approach of generative linguistics whose aim is to establish a set of rules, principles or constraints efficient to produce the surface phonetic forms of a language and to model the internalized linguistic knowledge of native speakers". Through these successfully obtained rules, principles or constraints, it is possible to classify the types of dialects and subdialects spread along the Perak River.

This Linguistics Analysis will be supported with findings from the Non-Linguistics Analysis, which is to discuss the factors that cause changes in phonological

data. The discussion regarding the phonological change factor will be discussed in depth through the two branches that were mentioned at the beginning of the study; the geographical factor and the social interaction factor. The discussion of the study can be focused on socioeconomic, historical, topographical and other factors in unraveling the causes of the change in the distribution of dialects along the upstream and downstream areas of the Perak River. This discussion of non-linguistics analysis will also be aided by the display of isogloss maps generated using Geographical Information Systems (GIS) technology. The display of this isogloss is indirectly able to classify areas by dialect based on the acquisition of study data from Table 1 above.

4. Research Analysis

Research studies involving the presence of /l/ and /r/ segments in the coda position are very varied in the Malay dialect. This has been supported by Zaharani Ahmad (2006) who stated that the liquid consonants /l/ and /r/ at the end of words in the Malay dialect have shown a phonological change including dialect speakers in Perak. Therefore, the discussion of this dialect will be explored in more depth, especially in dividing the distribution area of phonological characteristics of dialects in the upstream and downstream areas of the Perak River.

4. 1. Upper Perak Dialect

According to Nur Habibah C.H. and Rahim Aman (2020), The liquid consonants /r/ and /l/ can appear in the onset position of the first and second syllables in the Upstream Perak dialect, but not in the coda position. Therefore, the liquids segment in the Upstream Perak dialect will undergo the deletion rules in the coda position [consonant /r/ or /l/ → ø / __#]. This phenomenon regarding deletion rules in the Upstream Perak dialect has a close correlation with the Patani dialect. According to Pareeda Hayeeteh (2015) all consonant are not allowed to appear in the coda position of the final syllable including the liquids segment except /ʔ/, /h/, /ŋ/ (see Table 2-5) in the Patani dialect.

Table 2: Representation of Hulu Perak Dialect - lexical /cakar/ *claw*

Input :	/ʧakar/
Deletion Rules : [liquid /r/ → ø / __#]	/ʧaka/
Output :	[ʧaka]

Table 3: Representation of Hulu Perak Dialect - lexical /bantal/ *pillow*

Input :	/bantal/
Deletion Rules : [liquid /l/ → ø / __#]	/banta/
Deletion Rules : [nasal /n/ → ø / V__C]	/bata/
Output :	[bata]

Apart from the deletion rules, nasal consonants are also not allowed to appear in the coda position of the first syllable. According to Asmah Haji Omar (2015), all of the nasal consonants /m/, /n/, /ŋ/ and /ɲ/ in the Upstream Perak dialect can only be in the range of positions before and between vowels. This has caused the occurrence of second deletion rules on nasal consonants (see Table 3). The same thing was also found by Pareeda Hayeetah (2015) which is that the Patani Malay dialect will not allow the presence of a nasal consonant when it is followed by a voiceless plosive consonant /t/ which is homogenous with it.

Table 4: Representation of Hulu Perak Dialect - lexical /pasir/ *sand*

Input :	/pasir/
Deletion Rules : [liquid /r/ → ø / __#]	/pasi/
Vowel Lowering [vowel /i/ → /ɛ/ __#]	/pasɛ/
Output :	[pasɛ]

Table 5: Representation of Hulu Perak Dialect - lexical /kecil/ *small*

Input :	/kəʧil/
Change Rules [liquid /l/ → ʔ / __#]	/kəʧiʔ/
Output :	[kəʧiʔ]

The deletion rules in the liquid segments /r/ and /l/ also occurs in Tables 4 and 5. The lexical /pasir/ will undergo additional rules, namely the vowel lowering /i/ changes to /ɛ/ at the end of the syllable. The same thing was also found by Nur Habibah C.H. & Rahim Aman (2020) who stated that the vowel /i/ in the final syllable of the Upstream

Perak dialect will be realized into four allophones, namely, /i/, /iʸ/, /e/ and /ɛ/. Meanwhile, a different thing is exhibited by the lexical /kecil/ (see Table 5), which is the occurrence a Change Rules involving the segment /l/ which changes to glottal / ʔ / in the coda position of the final syllable [liquid /l/ → ʔ / __#]. This is because, there is no consonant that can occupy the final position of the word in the Upstream Perak dialect except the two consonants /ʔ/ and /h/ (Asmah Haji Omar 2015).

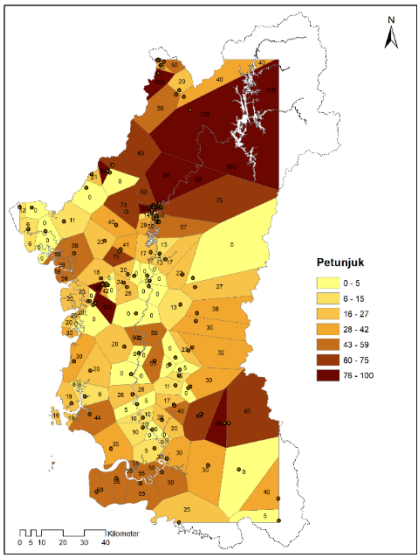


Figure 3 : Isoglos map of Upstream Perak dialect

Figure 3 shows that the distribution of Upstream Perak dialects. The prominent color pattern on the map above serves as a reference for the distribution of Upstream Perak dialects. The colour represents the high level of dialect spread in Upstream Perak, and it fades until the colour represents the lack of dialect spread in the area. Based on the map above, it shows that the spread of the Upstream Perak dialect originates from the Upper Perak River area and then spreads to the middle of the river. The spread of dialects is fading from the Kuala Kangsar district to the estuary of the Perak River. Through the Linguistics Analysis that has been carried out before, it has been shown the form of diffusion of the phonological characteristics of the Patani dialect in the area along Upstream Perak dialect. The spread of the Patani dialect is closely related to geographical factors, which the location of Upstream Perak next to the Southern region of Thailand. According to Yuzlin Yaccob et. al. (2018) the connection between the Upstream Perak dialect and the Patani dialect has been built

for a long time, especially in the Selama and Kerian districts. The connection between these two areas has led to the sharing of phonological features in the Upper Perak River area.

Apart from sharing the same distribution of phonological features, there is some evidence from a historical point of view that there is an entry of the Patani community into the Upstream Perak region. Raja Muktaruddin (1987) stated that most of the Perak natives who live in the Upstream Perak district are originally from the Patani district which is now under the rule of Thailand. The Pattani community in Southern Thailand has migrated to Perak because of the conflict between the Perak Old Kingdom and the Reman Old Kingdom in Thailand (Abdul Halim Nasir 1977). This has shown that there is a very close relationship involving the people in both areas. Through the perspective of geographical factors, the Upstream Perak district and the neighboring country of Thailand are only separated by several mountains and lakes but is connected to the Perak River (Abdul Halim Nasir 1997) and thus makes the interaction process between the two residents in the area easily connected.

4. 2. Downstream Perak Dialect

Similar to the data in the Upper Perak dialect, the Downstream Perak dialect drops the liquid segment in the coda position. Although the Hilir Perak dialect is located faraway from the Hulu Perak district and has a different topography, both dialects have showed similar phonological characteristics through their application of deletion rules [liquid /r/ or /l/ → ø /__#]. According to Nor Hashimah Jalaluddin et. al. (2016) the river waterways in the Kerian River (Upper Perak area) and other tributaries such as the Manik River which has flowed up to the Downstream Perak district has succeeded in becoming a communication route for the previous community and subsequently became an agent of the spread of dialects in Downstream Perak district. However, not all dialect features from the Upstream Perak dialect have successfully spread in the Downstream Perak area, and it only involves the sharing of certain phonological features.

Table 6: Representation of Downstream Perak Dialect - lexical /cakar/ *claw*

Input :	/ʧakar/
Deletion Rules [liquid /r/ → ø / __#]	/ʧaka/
Vowel Lowering [vowel /a/ → ɔ / __#]	/ʧakɔ/
Output :	[ʧakɔ]

Table 7: Representation of Downstream Perak Dialect - lexical /bantal/ *pillow*

Input :	/bantal/
Deletion Rules [liquid /l/ → ø / __#]	/banta/
Output :	[banta]

Table 6 shows the changes involving Deletion Rules in which consonant /r/ is dropped at the coda position and then undergoes vowel lowering rules, /a/ changes to vowel /ɔ/. This variant [ʧakɔ] is widespread especially in the Downstream Perak region. The same thing was also stated by Siti Noraini Hamzah and Nor Hashimah Jalaluddin (2017), the distribution of the phonological form of the final syllable position /-ar/ will be replaced by /-ɔ/ such as the lexical /besar/ becomes [bəsɔ] along the Downstream Perak River villages. The same thing also happens in Table 6 which has shown the phonological change from /ʧakar/ to [ʧakɔ] in the Downstream Perak dialect.

While for the lexical /bantal/ in Table 7, it only involves one rule change, which is the deletion of the segment /l/ in the coda position of the final syllable which becomes [banta]. According to Nor Hashimah Jalaluddin (2015), this consonant [banta] is a new variant that has been created as a result of the clash between the Patani dialect (located in the Upper Perak River region) and the dialect located in the Downstream Perak River through the Perak River communication line. This variation has expanded widely and is spoken as far downstream as the Perak River, particularly in the Lenggong district and north of Kuala Kangsar.

Table 8: Representation of Downstream Perak Dialect - lexical /pasir/ *sand*

Input :	/pasir/
Deletion Rules [liquid /r/ → ø / __#]	/pasi/
Vowel Lowering [vowel /i/ → e / __#]	/pase/
Output :	[pase]

Table 9: Representation of Downstream Perak Dialect - lexical

Input :	/kəʧil/
Deletion Rules [liquid /l/ → ø / __#]	/kəʧi/
Output :	[kəʧi]

Tables 8 and 9 show the phonological representation involving the segment liquids followed by the vowel /i/. The lexical /pasir/ and /kecil/ have changed to [pase] and [keci] in the Downstream Perak River dialect. This has some correlation with the Selangor dialect which is close to the Downstream Perak area. According to Asmah Haji Omar (2015), consonants /l/ and /r/ are only found before and between vowels in the Selangor dialect. Asmah added, the sequence of final syllables /-ir/ in the Selangor dialect will change to /-e/. Vowel /i/ will undergo vowel lowering rules which cause it to change to vowel /e/ after being followed by consonant liquids /r/ in the final syllable. The influence of the spread of the Selangor dialect in the Downstream Perak region is widely used starting from the middle of the Perak River to the estuary of the Perak River.

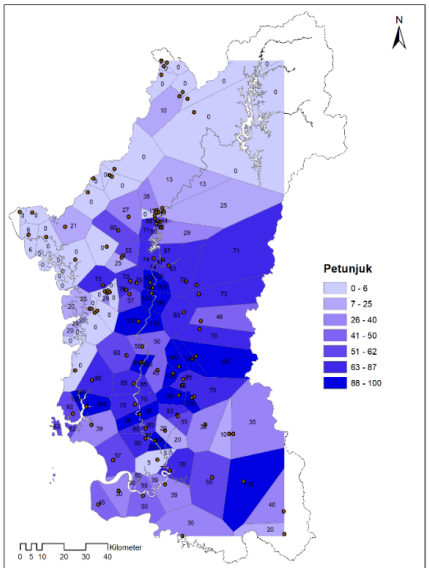


Figure 3 : Isoglos map of Downstream Perak dialect

Figure 3 shows the distribution of Downstream Perak dialects in the state of Perak. The prominent color pattern on the above map serves as a reference for this dialect distribution display. The colour represents the high level of dialect spread in Upstream Perak, and it fades until the colour represents the lack of dialect spread in the area. The linguistics analysis has succeeded in proving that the existence of the Selangor dialect has successfully spread to the Downstream Perak region. The Selangor dialect has spread due to social interaction, which includes usage in the administrative and cultural centers as well as indirect use as a source of Malaysia's national language. According to Siti Noraini Hamzah and Nor Hashimah Jalaluddin (2017), Selangor dialect has spread in the Downstream Perak area. Furthermore, the Selangor dialect was discovered to have spread significantly due to geographical causes, as the bordering territory between the states of Perak and Selangor is not obstructed by mountains or hills. This has indirectly facilitated the clash of community interaction between the two states. The influence that Selangor has as Malaysia's administrative and technological hub has also contributed to the occurrence of movement activities between the communities of Selangor and Perak. According to Hatice Tören (2014), the influence of standard dialects is dependent on the level of migration that occurs from village to city.

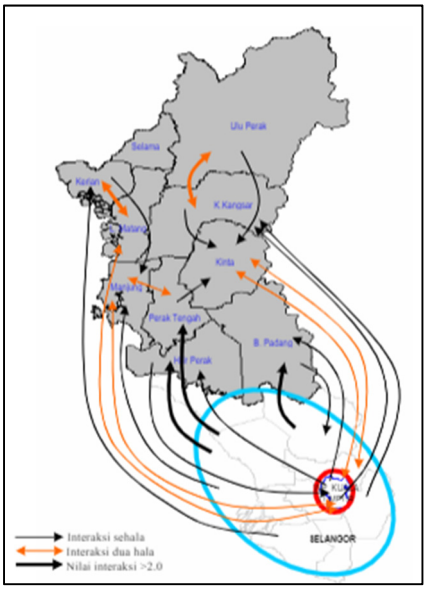


Figure 4 : Inter-migration map of Perak and Selangor
(Mohd Fadzil Abdul Rashid et al. 2012)

According to Mohd Fadzil Abdul Rashid et al. (2012) the migration interaction between Selangor and the Federal Territory of Kuala Lumpur with the districts in Perak has occurred consistently from 1980 to 2000. The districts of Batang Padang, Central Perak and Downstream Perak have received inflows of migration from Selangor continuously from 1980 (see Figure 4). The black arrow indicates one-way migration while the thick arrow refers to a migration value exceeding >2.0 while the orange arrow refers to two-way migration between the two states. The main factor in the occurrence of migration activities between these two states is due to social, economic and physical factors to meet the needs of human life. Therefore, it is not surprising that some phonological features in the Selangor dialect have been successfully found in the Downstream Perak region. Apart from the social interaction factor based on migration, the modernization factor also played a role in the spread of this dialect to Downstream Perak. Siti Noraini Hamzah (2021) stated that most of the speakers in the younger generation in Perak are more inclined to choose the standard pronunciation variant of the Malay language or known as the Selangor dialect as their conversational language.

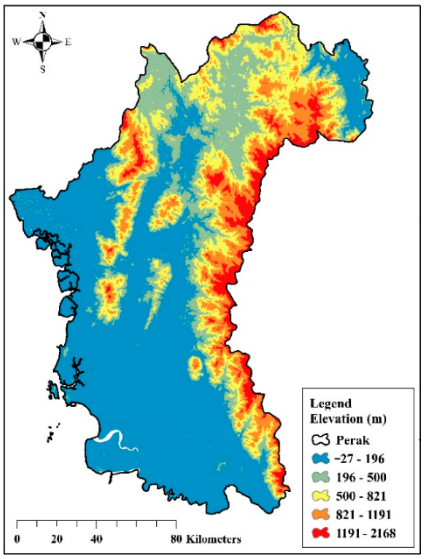


Figure 4 : The elevation map in Perak (Muhammad Talha Zeshan et al. (2021)

According to Muhammad Talha Zeshan et al. (2021) in his study "Monitoring Land Use Changes and Their Future Prospects Using GIS and ANN-CA for Perak Basin, Malaysia" has successfully obtained the Digital Elevation Model (DEM) which refers to elevation data in the Perak River (see Figure 5). Based on the display that was successfully generated, it shows that the contour height of the land surface is able to represent the distribution of the Upstream and Downstream dialects of Sungai Perak (see Figure 2 & 3). The altitude of 500-2168 meters above sea level in Perak has shown the distribution of Upper Perak dialects while the altitude of 500 meters below sea level shows the distribution of Downstream Perak dialects. The pattern of distribution of isoglosses for the Upper and Downstream Perak dialects has shown the same distribution as Figure 4. According to Muhammad Talha Zeshan et al. (2021) again, the Perak River basin area consists primarily of gentle, elongated highlands with mild to steep slopes. Due to its connection to the Perak River network, the highlands in the Upper Perak River region were unable to stop the dialect from spreading to the Downstream Perak River. According to Matthias Urban (2020) states that "There is a phonological difference in the dialect between the people who are in high areas and lower areas." So, this has indirectly strengthened this geographical factor capable of having a great impact on the diversity of dialects and subsequently capable of being one of the main contributors to the area of distribution of a dialect's prominent color pattern.

5. Conclusion

The distribution of dialects upstream and downstream of Perak has been successfully classified through the study of liquid segments in the coda position at the end of the syllable. The dialect distribution area has been successfully identified through the display of the isgolos that has been produced based on the color concentration distribution. Through the display of Figures 3 & 4 it has been shown that the spread of a dialect happens in stages. This has been supported by Asmah Haji Omar (2015) who stated that the spread of this dialect can be connoted as a wave of water. The water wave will form a strong wave after a splash and then the wave will spread and fade away.

This research analysis has proven that the elevation of 500 meters above sea level will exhibit a different phonological distribution. Limitations on the height of this contour can prevent the spread of dialects from happening. However, the dialect's restrictions can be connected by river waterways, which serve as communication links for the surrounding communities. This has indirectly proven that geographical factors are one of the main factors in the occurrence of language variation in an area. This is supported by several arguments from a geographical perspective and topographical data such as rivers, hills, land contours, and others that can examine the factors that led to the spread of a dialect in more depth.

References

Abdul Halim Nasir (1977) Sejarah Perak Siri Pertama. Percetakan Tanda Jaya: Kuala Lumpur.

Abdul Halim Nasir (1977) Lembangan Sungai dalam Peradaban Melayu. Universiti Kebangsaan Malaysia

Ahamd Dahlan (2015) Sejarah Melayu. Jakarta: Kepustakaan Populer Gramedia.

Asmah Haji Omar (2015) Susur Galur Bahasa Melayu. Ed 2. Dewan Bahasa dan Pustaka: Kuala Lumpur

Dwi Atmawati (2018) Influence of Societal Factor on the Emergence of Language Variants: The Study of Indonesia Language Use in Javanese Society. Prosiding Fourth Prasasti International Seminar on Linguistics (Prasasti 2018). Advances in Social Science, Education and Humanities Research.

Faizah Mohd Fakhruddin (2022) Sungai Perak: Sejarah dan Peraddaban. Journal of Islamic. Vol 7 (49).

Fareed Hameed Al-Hindawi (2018) The Basic Tenets of Generative Phonology. Journal of Education and Practice. Vol. 9 (1).

Hamidy UU (2011) Jagad Melayu dalam Lintasan Budaya di Riau. Pekanbaru: Bilik Kreatif.

Hatice Tören (2014) Urbanization, Standard Language & Dialect. Türk Dili ve Edebiyatı Dergisi. Vol. 50 (50)

Kazuko Matsumoto (2019) Language Variation and Change. The Routledge Handbook of Japanese Sociolinguistics. Routledge

Marlyna Maros (2010) Dialek Melayu Pekan: Penggunaan dan Pengekalan. Jurnal Bahasa Vol.10.

Matthias Urban (2020) Mountain Linguistics. Language and Linguistics Compass. Vol 14 (9).

Mohd Fadzil Abdul Rashid, Mohd Roswodi Mat Zin & Norhazlan Haron (2012) Pola dan Penyerakan Migrasi di Negeri Perak dari 1980-2000 dan Penelitian Awal Implikasinya. Prosiding International Conference on Innovation and Technology for Sustainable Built Environment (ICITSBE), April 2012, UiTM (Perak).

Mohd Tarmizi Haran. (2018a) Dialek Melayu Hulu dan Hilir di Timur Semenanjung Malaysia. Jurnal Bahasa. Vol 18 (1).

Mohd Tarmizi Hasran (2018b) Analisis Sejarah Fonologi dan Pengelompokan Dialek Melayu Hulu di Bahagian Timur Semenanjung Malaysia. PhD thesis. Universiti Sains Malaysia

Mokhtar Saidin & Jeffrey Abdullah (2007) Sungai Perak Kuno: Sumbangannya kepada Zaman Paleolitik Malaysia. Jurnal Arkeologi Malaysia. Vol 20.

Muhammad Talha Zeshan, Muhammad Raza Ul Mustafa & Mohammed Feras Baig (2021) Monitoring Land Use Changes and Their Future Prospects using GIS and ANN-CA for Perak River Basin, Malaysia. Water 2021. Vol 13.

Noam Chomsky & Morris Halle (1968) The Sound Pattern of English. Harper & Row: New York.

Nor Hashimah Jalaluddin (2015) Penyebaran dialek Patani di Perak : Analisis Geolinguistik. Jurnal Antarabangsa Dunia Melayu. Vol 8 (1).

Nor Hashimah Jalaluddin, Norlisafina Sanit, Zaharani Ahmad & Harishon Radzi (2016) Lexical Variations of Dialectal Pronouns along the Perak Riverfront: The Geographical Information System Analysis. Jounal of Language Studies (GEMA Online). Vol 16 (1).

Nor Hashimah Jalaluddin (2017) Penyebaran dialek Melayu di Langkawi: Analisis Geolinguistik. Jounal of Language Studies (GEMA Online). Vol 17 (4).

Norzawati Yoep, Hazrin Hasim, Ummi Nadiah Yusoff, Mahani Yusoff & Noor Rizawati Mahpot (2015) Spatio-Temporal Distribution of Malaria in Perak, Malaysia. Advances in Infectious Diseases, 5, page: 154-161.

Nur Habibah C.H. & Rahim Aman (2020) Varian Hulu Perak Utara: Fenomena di Gerik dan Pengkalan Hulu. Jurnal Melayu. Vol 19 (2).

Oktavian Mantiri (2010) Factors Affecting Language Change. SSRN Electronic Journal.

Pareeda Hayeeteh (2015) Penghilangan Konsonan dalam Dialek Melayu Patani. Sarjana Sastera. Yala Rajabhat University.

Raja Muktaruddin (1987) Dialek Perak. Yayasan Perak.

Shalizaini Uyob & Che Ibrahim Salleh (2017) Jati Diri Bahasa Melayu sebagai Bahasa ASEAN. Pertanika MAHAWANGSA. Vol 4 (1).

Siti Noraini Hamzah (2021) Kepelbagaian Varian Leksikal Dialek di Perak: Pendekatan Geographical Information System. Akademika. Vol 88 (1).

Siti Noraini Hamzah & Nor Hashimah Jalaluddin (2017) Migrasi Masyarakat Luar dan Pengaruh Dialek Geolinguistik. Jurnal Bahasa. Vol 17 (1).

Syahreni Siregar (2017) The Influence of Dialect on The Student's Pronunciation in Speaking Ability. Pedagogy Journal of English Language Teaching. Vol 5 (1).

Yuzlin Yaccob, Norlizawati & Noraimanah Saadun (2018) Dialek Patani: Keunikan Bahasa di Hulu Perak. Prosiding Seminar Antarabangsa Arkeologi, Sejarah, dan Budaya di Alam Melayu. Integrasi Nusa Maritim dan Penguatan Jalinan Kebhinekaan Alam Melayu di Asia Tenggara.

Zaharani Ahmad (2006) Kepelbagaian Dialek dalam Bahasa Melayu: Analisis Tatatingkat Kekangan. Jurnal Sains Sosial dan Kemanusiaan. Vol 1 (1).

Geolinguistic Analysis of the Yi Script written in *Hua-Yi Yiyu*

Kazue Iwasa
(Nagoya University of Foreign Studies)

Abstract: In this study, the geolinguistic approach is applied to an analysis of the Yi script in *Hua-Yi Yiyu* (華夷譯語, literally *'Chinese-Foreign Glossaries'*). The Yi script was used to express the Yi language (Tibeto-Burman), spoken by the Yi ethnic group living in southwestern China and northern parts of Vietnam and Laos. The Yi characters show great diversity in shape and phonetic values across regions, and researchers have turned their attention to elucidating the possible routes of their spread and the patterns of change in their forms. To do this, it is necessary to analyse the Yi characters in much narrower areas than before; doing so enables us to trace minute differences in more detail between the research points. To this end, the Lolo version of *Hua-Yi Yiyu* is an ideal source of data, as it contains five distinctive sites that seem to be located in a relatively narrow area. However, two of the sites are unknown, although several studies have been conducted on them. In this study, the author attempts to locate these two sites by focusing on the shapes in the Yi characters rather than their phonetic information.
Keywords: geolinguistic application, the Yi script, Yi-character maps, *Hua-Yi Yiyu*

1. Introduction

This study innovatively applies geolinguistic methods to an analysis of the Yi characters written in *Hua-Yi Yiyu* (華夷譯語, literally *'Chinese-Foreign Glossaries'*).

The Yi script has been used to express the language called Yiyu (彝語: Tibeto-Burman), which is spoken by the Yi ethnic group living in southwestern China and northern parts of Vietnam and Laos. The Yi characters represent great diversity in their shape and phonetic values across regions, and elucidation of the possible routes of their spread and their change process has attracted researchers' attention. Along this line, in an earlier work I drew Yi-character maps based on Swadesh Wordlist and analysed

IWASA, Kazue. 2023. Geolinguistic Analysis of the Yi Script written in *Hua-Yi Yiyu*. In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 172–202. doi: https://doi.org/10.5281/zenodo.8374673

* I deeply thank Professor Mitsuaki ENDO at Aoyama Gakuin University, who always supports and encourages my research. Also, I express my gratitude to Professor Takumi IKEDA at Kyoto University for his support in my research. Finally, Mr Hiroyuki SUZUKI kindly gave me helpful pieces of advice as always, I very much thank him here.

them geolinguistically to explore potential routes of the characters' spread[1]. In the results of this work, 42 common characters were found among the Yi dialects.

However, clarifying the possible routes of the Yi character transmission and determining the change patterns of the characters are still open to discussion. To solve these questions, it is necessary to examine Yi characters in much narrower areas than before, because this approach allows us to capture minute differences in the characters in more detail across certain research points.

To this end, the Lolo version of *Hua-Yi Yiyu* is an ideal source of data, since it contains five distinctive locales that are thought to be in a relatively narrow area—specifically, where Sichuan, Yunnan, and Guizhou intersect. However, two of these five places are still unclear, although some research has been conducted on this issue.

In this study, I first show actual Yi-character maps and their analysis, as well as some remarkable results brought about by the application of geolinguistics to analyses of the Yi characters found in *Hua-Yi Yiyu*. Then, I tentatively identify the two unknown locations.

1.1. Yi language and Yi characters

The language spoken by Yi people, known as Yiyu in China, belongs to the Lolo-Burmese language group, in the Tibeto-Burman language family. It is spoken around the southwestern area of China—in the provinces of Sichuan, Yunnan, Guizhou, and Guangxi—as well as in the northern parts of Vietnam and Laos. There are six dialects in China, and four of them use the Yi script. See Map 1 below: green squares indicate the dialects using the Yi script, whereas orange ones indicate the dialects without scripts.

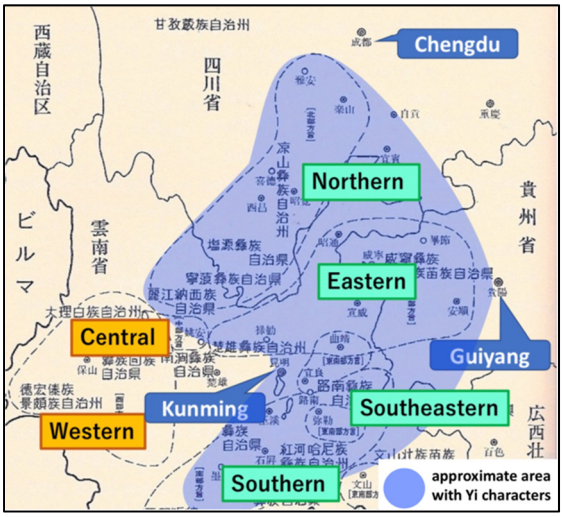
A significant number of manuscripts written in Yi characters are preserved in China, along with a surprising number in Europe, a few in Japan, and possibly three *Hua-Yi Yiyu* in Hanoi[2].

Yi characters are considered to have been completed around the Yuan or Ming dynasty. The Yi script has been used by *Bimo*, religious leaders among Yi people. Throughout history, the script and manuscripts have been handed down exclusively within their clans. One feature which should be noted is that unlike most other characters, Yi characters are not a tool for communicating but one for keeping secrets within every *Bimo*'s clan, from generation to generation. It is incredibly personal writing, so it serves to protect the clan's secrecy.

[1] See Iwasa (2019) for more detail.

[2] I attempted to find them in Hanoi; however, I failed to. Several scholars have also tried to find them but, as far as I know, have never been successful.

At present, the characters are all syllabic, but they probably used to be logographic; for example, some of them are distinguished in writing according to their meaning. This study profits from the nature of Yi characters.



Map 1: Yi dialects in China.

1.2. *Hua-Yi Yiyu* and the Lolo series

Hua-Yi Yiyu[3] is a collective term for various collections, including glossaries (*yiyu*, 譯語), words in common use (*zazi*, 雜字) and received documents (*laiwen*, 來文). It was compiled by official organisations to meet the needs of foreign exchanges around the Ming and Qing dynasties.

Hua-Yi Yiyu can be divided into four categories:

1. Jia books (甲種本). These were compiled by the order of the Hongwu Emperor of the Ming Dynasty and were published in 1389.

2. Yi books (乙種本). In 1407, the Institute for Foreign Services (*Siyi guan*, 四夷館) was established in the Ming Dynasty. There, the books were compiled for the translation of documents written in foreign languages. Later, these books were taken over by a similar organisation (*Siyi guan*, 四譯館) in the Qing Dynasty and kept in order to be revised and modified.

[3] It is translated as *Chinese-Foreign/Ethnic Language Glossaries* in Chunhua et al. (2018).

3. Bing books (丙種本). These books are understood to have been compiled around the end of the Ming Dynasty to train officials who received foreign envoys at an organisation called *Huitong guan* (會同館).

4. Ding books (丁種本). These were compiled in 1748, in the Qianlong era of the Qing Dynasty, at the Institute called *Huitong Siyi guan* (會同四譯館).

As *Hua-Yi Yiyu* comprise valuable linguistic data of languages in and outside China from the 14th century, they provide important pieces of information about the older forms of such languages.

The Lolo series is from the Yi books; it contains five distinctive written copies, the data of which were collected in five different locales. In this paper, I label them each with a number from one to five: *Book 1, Book 2, Book 3, Book 4, and Book 5*.

The Palace Museum in Beijing has five complete copies of the Lolo series. Their photo-offset copies were finally opened to the public in 2018. In Japan, other transcriptions of two of them are preserved, one owned by Professor Imanishi at Kyoto University, and another at Otani University; in this paper, I call them (as necessary) the *Imanishi Book* and the *Otani Book*, respectively. It has been claimed that three *Hua-Yi Yiyu books* could be found at École Française d'Extrême-Orient in Hanoi; however, as yet, none has been found.

In the Lolo series, the original locations of *Book 4* and *Book 5* are unknown. This paper aims to detect them through geolinguistic analysis.

The glossaries consist of 20 sections, such as Astronomy, Geography, Time, and so on, and *Book 1* is edited differently from the other four books. In this study, 138 vocabularies from Astronomy, Geography, and Time were mapped and analysed geolinguistically, some of which are presented in the next section.

	Book 1 (東川)	Book 2 (建昌)	Book 3 (永寧)	Book 4 (X)	Book 5 (Y)
Beijing	1	2	3	4	5
Kyoto	-	Otani book	Imanishi book	-	-
Hanoi	(1)	(3)	(2)	-	-

Chart 1: The Lolo series and the corresponding *Books*

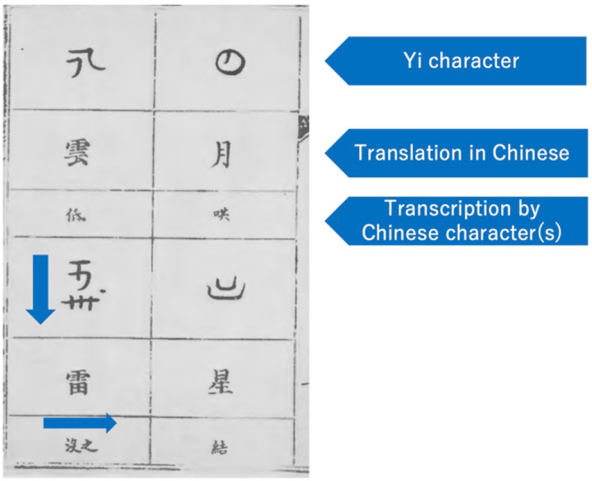


Figure 1: A page of *Book 3*, from Nishida (1979: 3)[4]

After my thorough examination of all seven *Books*, I found quite a few cases where numbers between Yi character(s) and Chinese characters used for transcribing Yi words did not agree: the numbers of Chinese characters for transcription are sometimes more than that of Yi characters, and sometimes less. Addressing this issue, Nishida (1979: 299-301) suggests that there are three phases in the development of Yi characters and that the characters used to be logographic[5]. Then, he concludes that certain Yi characters written in *Book 1* (Dongchuan, 東川) are logographic, since in some cases one Yi character is transcribed by more than two Chinese characters and signifies one single word. However, this is not the case, nor is it only true for *Book 1*.

My investigation into the characters and the phonetic information of the seven *Books* strongly indicates that both locations X and Y should be somewhere in the Eastern dialect area.

[4] The blue arrows indicate the directions of the characters; the explanations in the blue arrows are added by the author.

[5] Nishida (1979) reconstructs the phonological system of Shuiliao Yi based on the phonetic transcription by Chinese characters in the *Imanishi Book*. Additionally, he regards this 'Chinese' as Mandarin, although he still implies the possibility that the phonetic transcription may have been written in Southwestern Mandarin. In fact, according to my investigation, this Chinese transcription is based on Southwestern Mandarin, *Imanishi Book* (i.e., *Book 1*).

It should be noted that the consultants of Yi characters[6] and the pronunciations of the glossary for all the *Books* cannot be identical in each location[7]. Hence, I propose that the Yi characters written in the five *Books* should represent the written and spoken Yi language used between the 15th and the 17th century, as it is highly probable that the Yi characters and pronunciations for the glossary were provided by different consultants. Therefore, this paper pays much more attention to the Yi characters themselves in the *Books* than their phonetic values. This study's focus on the characters themselves is also supported by the logographic nature of the Yi script.

1.3. Goals and Methodology

1.3.1. Goals

In this study, I have set four goals: one has been achieved, one is current, and the other two are long-term goals.

Goal 1: Are certain Yi characters polysyllabic?

Goal 2: Detect the locations of X and Y.

Goal 3: Clarify the possible change pattern(s) in Yi characters.

Goal 4: Trace the feasible route(s) through which Yi characters would have spread.

As for Goal 1, this question has already been answered (in the negative) above: Yi characters are monosyllabic, that is, one Yi character bears only one syllable. Although Nishida (1979) suggests that some Yi characters may signify more than two syllables, this is not true. If one Yi character seems to bear (more than) two syllables in *Hua-Yi Yiyu*, it is because the data for the script and the pronunciation must have been separately collected, possibly by different consultants. Consultants for writing must have given information on the written language, and those for pronunciation must have provided their colloquial vocabulary. This could produce such a discrepancy.

Goal 2 is the current aim of this paper. (Goals 3 and 4 are issues for future work.)

1.3.2. Methodology

Three steps were taken to detect the locations of X and Y.

Step 1: Make a list of all the vocabulary and Yi characters based on the seven *Books*—the five *Books* in the Palace Museum and the two other versions in Japan.

[6] It is very plausible that the Yi characters may have been copied by someone who did not understand the Yi script, since it seems that some copiers did not even recognise where two Yi characters should be divided.

[7] Nie (2020:228) also mentions this possibility, but only in the case of *Book 1*.

Step 2: Draw Yi-character maps based on the list.

Step 3: Analyse the maps and the related data by comparing them with the Yi-character maps generated by the author.

Judging the similarity among Yi characters based on their shapes, and categorising and distinguishing certain Yi characters from each other, I found the following change patterns in the Yi script (shown below): the direction and tendency of shape changes are considered to change from relatively complex to simpler shapes and/or fewer strokes[8].

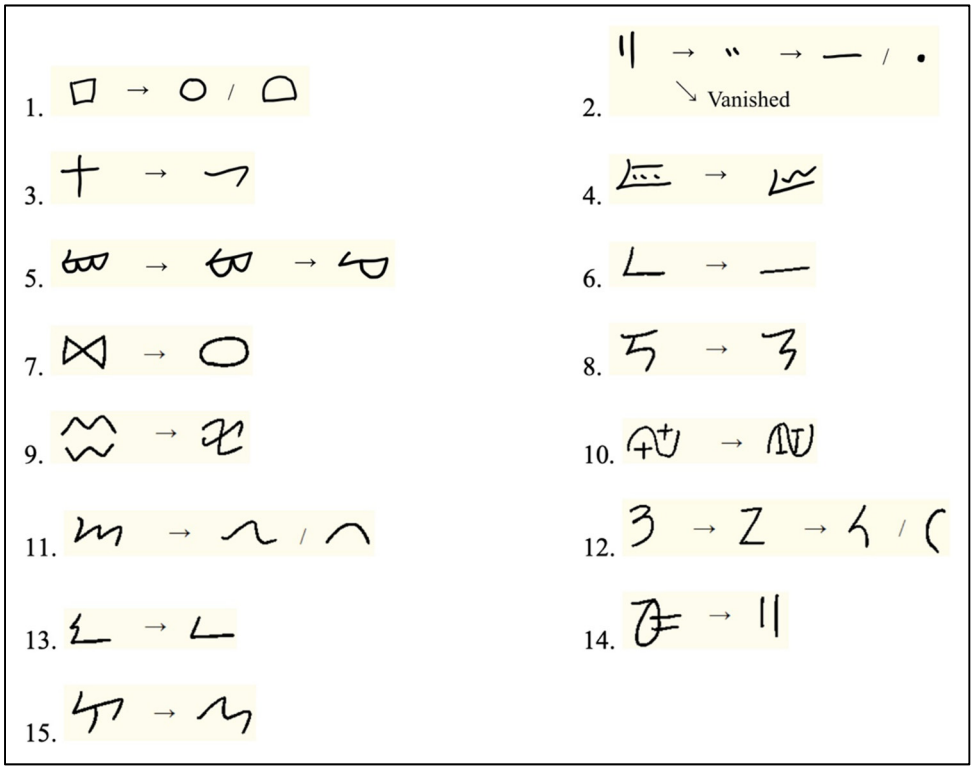


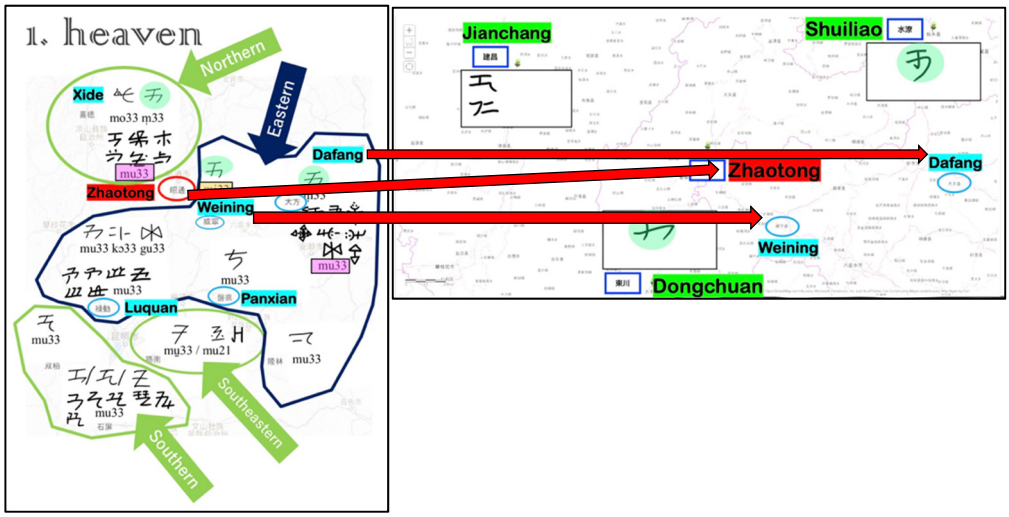
Figure 2: Typical change patterns in Yi characters, from Iwasa (2018: 8)

2. Comparison between the Yi characters written in *Hua-Yi Yiyu* and those on Yi-character maps

In this section, I show types of Yi-character maps. One type is primarily from Iwasa (2018) and appears on the left, and the other was produced for this study and appears

[8] See Iwasa (2018) for more details.

on the right. The maps on the left side show the Yi characters of four dialects[9], while the maps on the right side present the Yi characters written in *Books 1*, *2*, and *3*.



Maps 2 and 3: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

Both maps share three common locations: Zhaotong (昭通), Dafang (大方), and Weining (威寧), indicated by the red arrows on the maps above. The left maps demonstrate the Yi characters of the four dialects: the Northern, Eastern, Southern, and Southeastern, indicated by green arrows (except for the Eastern dialect, which is marked by a dark blue arrow). The main locations are signified by light blue highlights. The right map shows the Yi characters written in *Hua-Yi Yiyu*—in Dongchuan (東川) from *Book 1*, Jianchang (建昌) from *Book 2* and Shuiliao (水潦) from *Book 3*.

[9] The data sources for the Yi characters are *DianChuanQianGui Yihan Jiben Cihui Duizhao Cidian* (『會同滇川黔桂彝漢基本詞彙對照詞典』 (1984)), *Hanyi Jianming Cidian* (『漢彝簡明詞典』(2011)), and *Diannan Yiwen Zidian* (『滇南彝文字典』(2005)). Other sources are consulted as necessary.

	Book 4	Book 5
Yi character		
Transcription in Ch.	米	木
Pinyin	mi	mu

Chart 2: The Yi characters and their transcriptions into Chinese characters in *Books 4* and *5*, with pinyin

In Chart 2 above, the Yi characters written in *Books 4* and *5* are shown at the top of the chart. In the middle row, their transcriptions into Chinese characters are signified. At the bottom, the pinyin of the Chinese characters is provided.

2.1. Geolinguistic analysis of the Yi characters written in *Hua-Yi Yiyu*

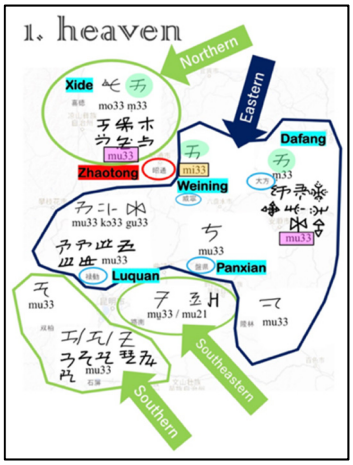
Here, I compare two types of Yi-character maps and analyse them geolinguistically. The colours used for shading characters and highlighting are linked to those in the maps and charts, making it easy to find similar characters on, e.g., maps 2 and 3, because they are shaded in light green on both maps. Likewise, we can also check the phonetic values on the same maps by following pink or orange highlights.

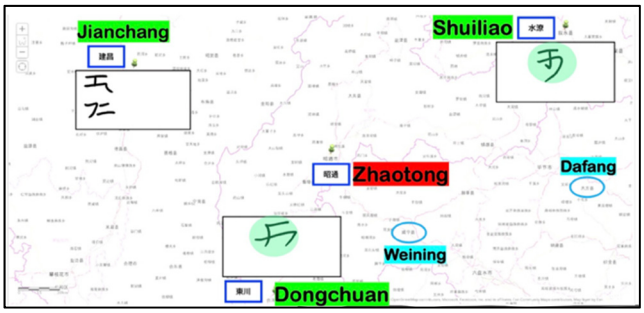
2.1.1. Part I. Astronomy

Here, I explore the maps concerning the seven vocabularies from Astronomy. The numbers on the maps are provided for convenience and correspond to the order in the *Imanishi Book* of *Hua-Yi Yiyu*.

1. Heaven

Both characters signifying ‘Heaven’ in *Books 4* and *5* show great similarity in their shapes to the characters of Dafang, Weining, and Xide (Liangshan, 凉山) on the left map. Given the phonetic values, the characters in *Book 4* seem to come from Weining, and those in *Book 5* from Dafang or Xide.





Maps 4 and 5: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

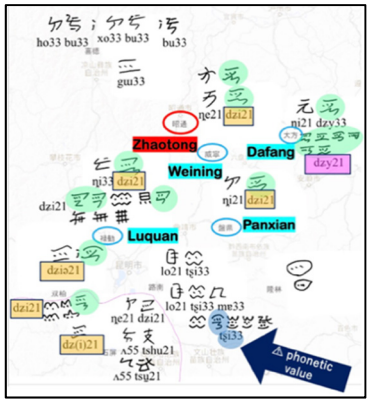
	Book 4	Book 5
Yi character		
Transcription in Ch.	米	木
Pinyin	mi	mu

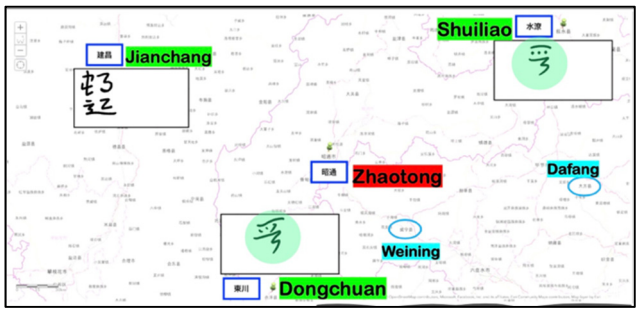
Chart 3: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

2. Sun

This is one case of the 42 common characters; normally, it should be disregarded, because we can hardly find any regional features from them. However, in light of the phonetic information of the Yi characters from *Books 4* and *5*, they seem to belong to the Eastern or Southern dialect, but not likely to the Southeastern dialect.

The transcription in *Book 5* is 'ju' in pinyin, so it seems to have a [y] vowel. It is very plausible that the character in *Book 5* comes from the Dafang area.





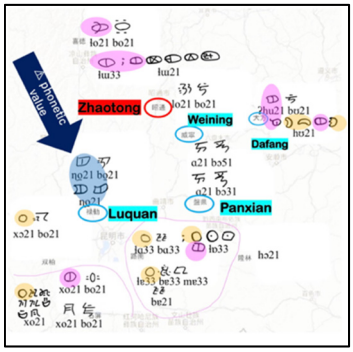
Maps 6 and 7: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

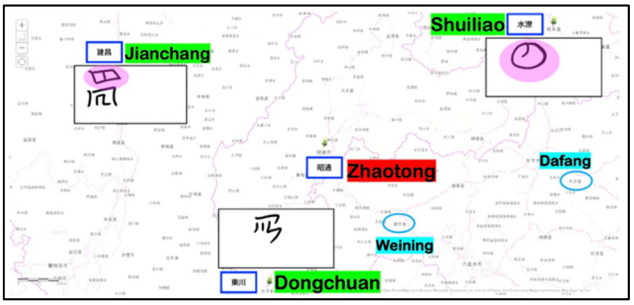
	Book 4	Book 5
Yi character		
Transcription in Ch.	己	舉
Pinyin	ji	ju

Chart 4: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

3. Moon

This is one of the common Yi characters. The characters in Luquan (祿勸) should be excluded from the candidates for determining the locations of Y (*Book 5*), because their pronunciations are different from that in *Book 5*.





Maps 8 and 9: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

	Book 4	Book 5
Yi character		
Transcription in Ch.	合	哄
Pinyin	he	hong

Chart 5: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5* with pinyin

5. Cloud

The shapes of the characters in *Books 4* and *5* suggest that they belong to the Dafang dialect area. However, when the Chinese character transcriptions[10] and phonetic values of the modern dialect are considered, there is a medial [i] in the area. Therefore, the character in *Book 5* must be from the Dafang area.

Maps 10 and 11: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

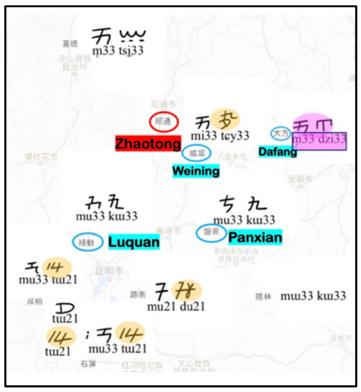
[10] The pinyin for the Chinese character '柴' is 'ye'. However, when all the phonetic data of the Yi dialects on the map are considered, this transcription is very likely an error of confounding with similar Chinese characters as '碟' or '牒', whose pronunciations are 'die' in pinyin.

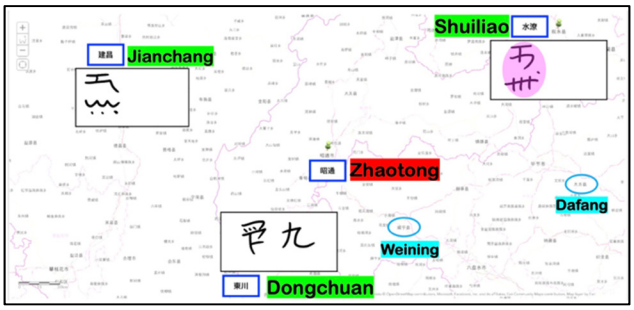
	Book 4	Book 5
Yi character		
Transcription in Ch.	德	某
Pinyin	de	die (※ye)

Chart 6: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with Pinyin

6. Thunder

The characters in *Book 5* are identical to those in Dafang.





Maps 12 and 13: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

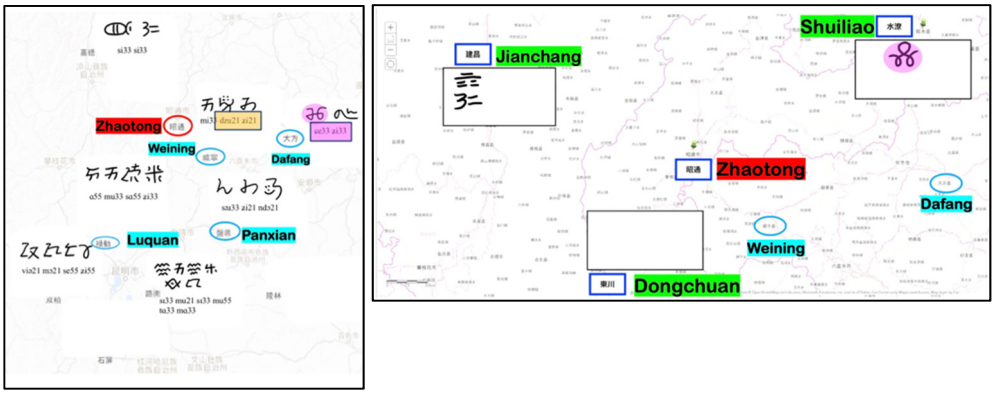
	Book 4	Book 5
Yi character		direction of Yi characters → 
Transcription in Ch.	歇?	遮木
Pinyin	xie?	zhe mu

← direction of Chinese characters

Chart 7: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

15. Rainbow

The first character in *Book 5* is similar to that in Shuiliao and Dafang, and the transcription also seems to correspond well with that in Dafang, whereas the transcription in *Book 4* is close to the one in Weining.



Maps 14 and 15: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

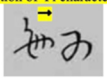
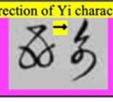
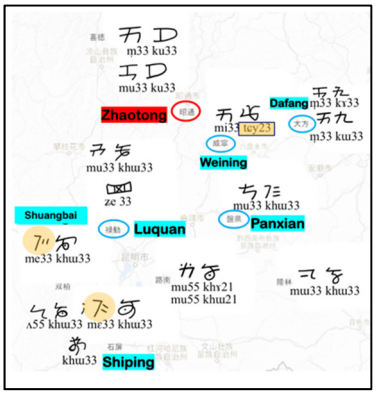
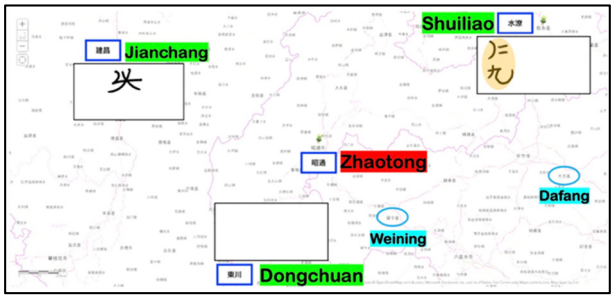
	Book 4	Book 5
	direction of Yi characters →	direction of Yi characters →
Yi character		
Transcription in Ch.	以腳	衣些
Pinyin	yi jiao	yi xie
	← direction of Chinese characters	← direction of Chinese characters

Chart 8: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

16. Smoke

The first character in *Book 4* is like those in Shuangbai, Shiping, and Shuiliao, while the transcription of the second character shows high similarity to the one in Weining.





Maps 16 and 17: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

	Book 4	Book 5
Yi character		
Transcription in Ch.	居麥	煙
Pinyin	ju mai	yan

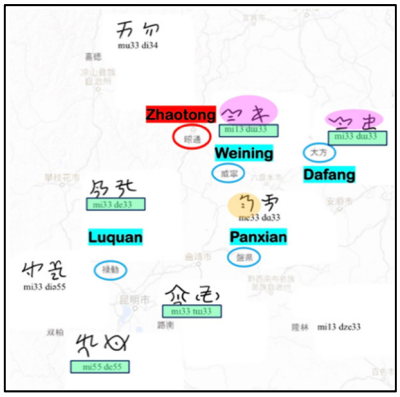
Chart 9: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

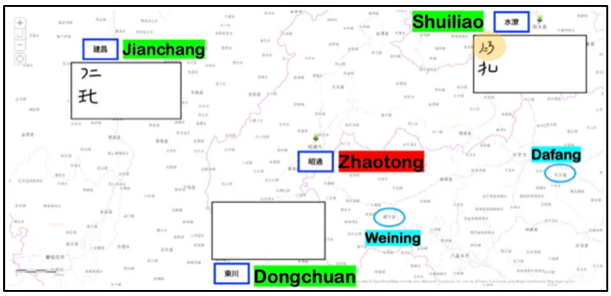
2.1.2. Part II. Geography

Here, I explore the maps concerning the ten vocabularies from Geography. The numbers on the maps are provided for convenience and correspond to the order in the *Imanishi Book* of *Hua-Yi Yiyu*.

6. Local place

The character circled in pink in *Book 5* is similar to those in Dafang and Weining. However, the character circled in orange shows certain similarities to those in Panxian, Lunan and Shuiliao.





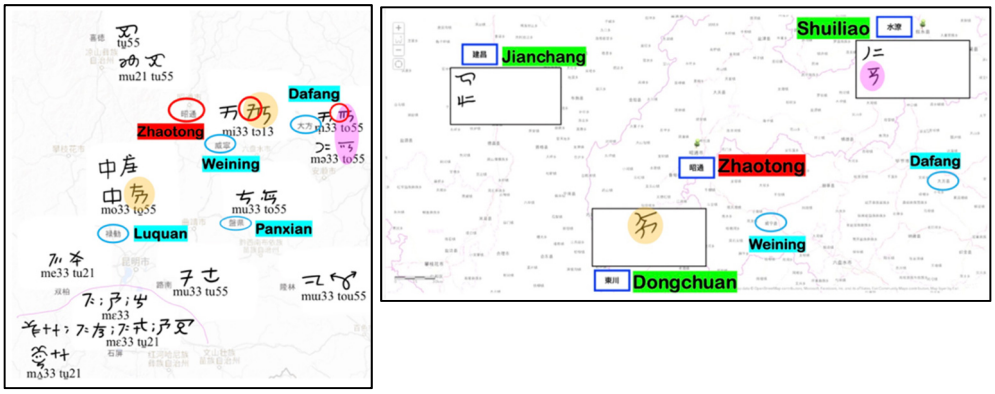
Maps 18 and 19: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

	Book 4	Book 5
Yi character		
Transcription in Ch.	得密	得密
Pinyin	de mi	de mi

Chart 10: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

8. Fire

The character in *Book 5* bears a strong similarity to that in Dafang, whereas that in *Book 4* bears a strong similarity to a character in Weining and Luquan. It seems that the phonetic transcriptions in both *Books 4* and *5* correspond highly with the pronunciations of modern dialects in Weining, Dafang, and Luquan, which all belong to the Eastern dialect.



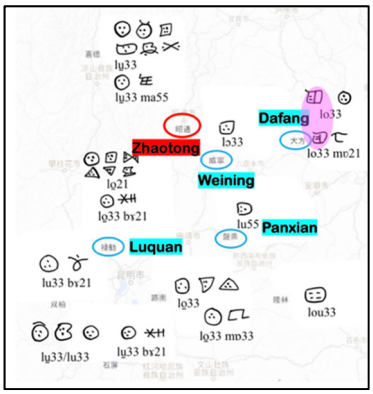
Maps 20 and 21: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

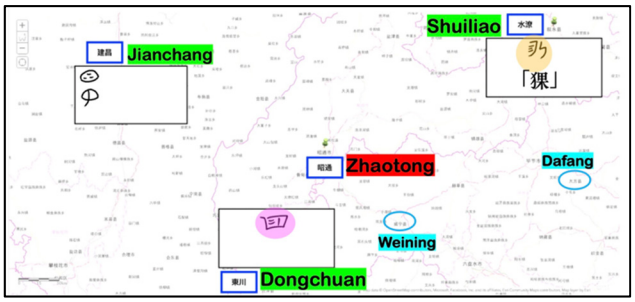
	Book 4	Book 5
Yi character		
Transcription in Ch.	奪	隋（墮？）
Pinyin	duo	（duo）

Chart 11: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

9. Stone

The character in *Book 4* is similar to that in Shuiliao; their transcriptions are also as close, as 「各 (ge)」and 「猓 (guo)」. The language described in *Book 4* may very likely belong to the Eastern dialect. The character in *Book 5* is very similar to that in Dafang, and its transcription also appears to be close to that in Dafang.





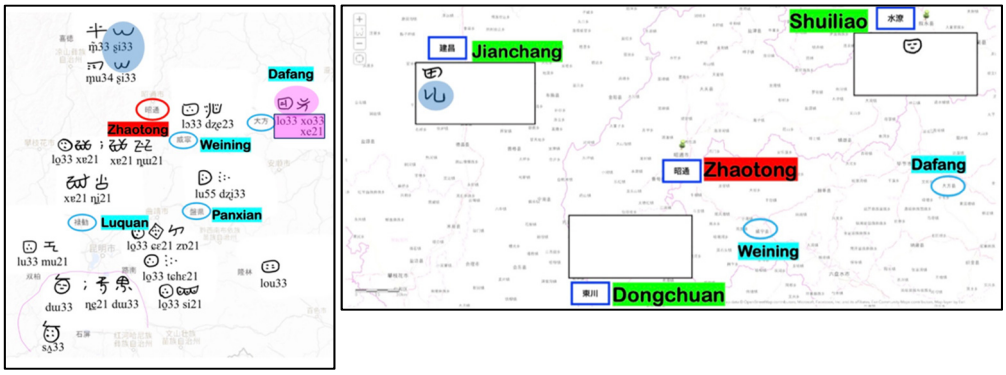
Maps 22 and 23: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

	Book 4	Book 5
Yi character		
Transcription in Ch.	各	洛
Pinyin	ge（?）	luo

Chart 12: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

11. Sand

The characters and transcription in *Book 5* bear a great similarity to those in Dafang. This lexical item probably reflects a word from around the Dafang area used before the 17th century.



Maps 24 and 25: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

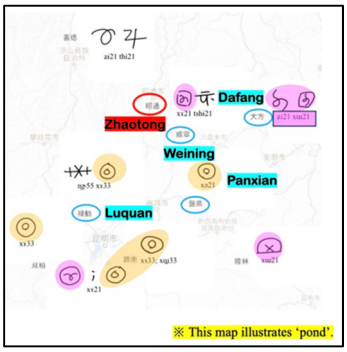
	Book 4	Book 5
Yi character		
Transcription in Ch.	葉撒	黑洛
Pinyin	ye sa	hei luo

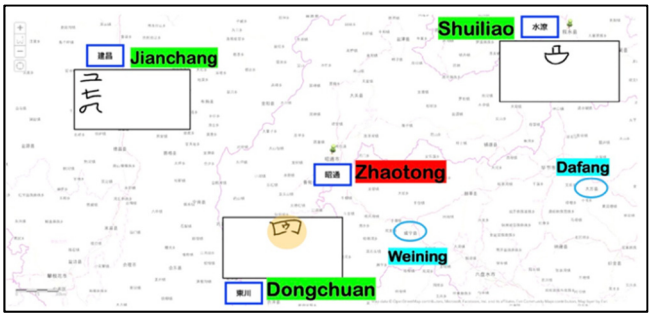
Chart 13: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

12. Sea[11]

Both the characters and transcriptions in *Book 5* indicate this lexical item seems to be from the Dafang area. However, we cannot determine the origin of the character in *Book 4*, because it shows a similarity to those in many places.

[11] The Yi-character map of the four dialects indicates the distribution of 'pond', since there is not enough data for the word 'sea'. This may be because most Yi people live in mountainous areas, so in their vocabularies 'sea' tends to be expressed like 'big pond'.





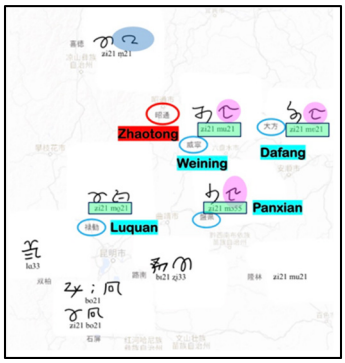
Maps 26 and 27: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

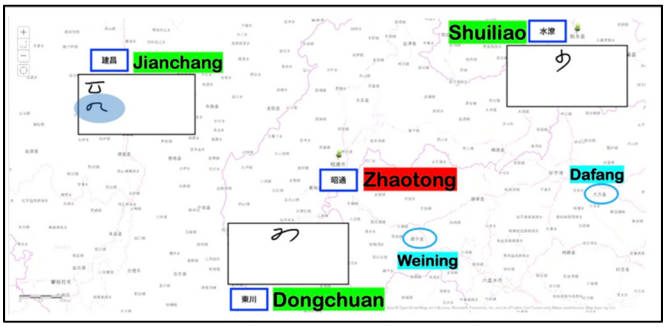
	Book 4	Book 5
Yi character		
Transcription in Ch.	黑	狠以
Pinyin	hei	hen yi

Chart 14: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

14. River

The transcriptions of both *Books 4* and *5* seem to correspond to the modern word forms in the Eastern dialectal area. Therefore, we can determine that the locations of both books should belong to the Eastern dialect.





Maps 28 and 29: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

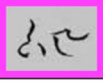
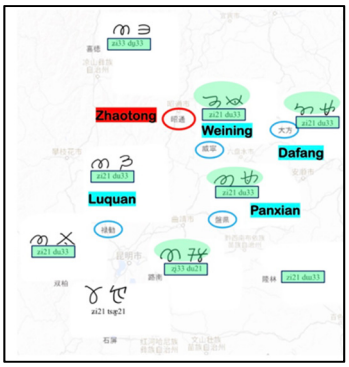
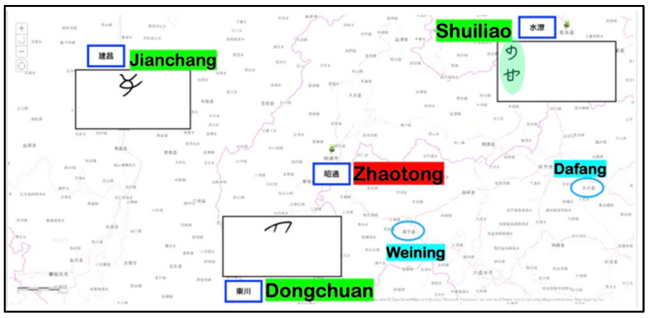
	Book 4	Book 5
Yi character		
Transcription in Ch.	模以	莫以
Pinyin	mo yi	mo yi

Chart 15: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

16. Well

The word forms in *Books 4* and *5* are found in a vast area beyond the specific dialectal areas. Thus, it is difficult to determine the origins of the two books from the data of this lexical item.





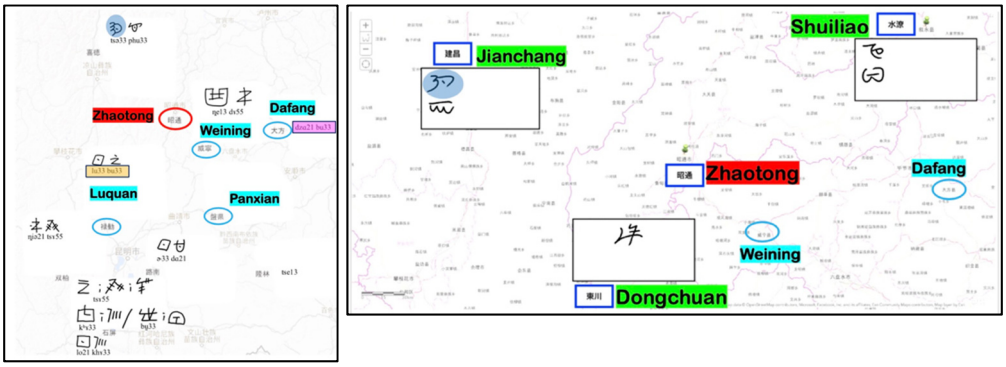
Maps 30 and 31: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

	Book 4	Book 5
Yi character		
Transcription in Ch.	都以	都以
Pinyin	dou yi	dou yi

Chart 16: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

17. Wall

There are no correspondent word forms in the maps with those in *Books 4* and *5*. However, the modern pronunciation in Dafang is similar to that in *Book 5*, and the phonetic value in Luquan is similar to that in *Book 4*. More careful investigation is necessary, because Chinese characters cannot signify voiced consonants (which appear in modern dialects like Dafang and Luquan).



Maps 32 and 33: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

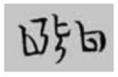
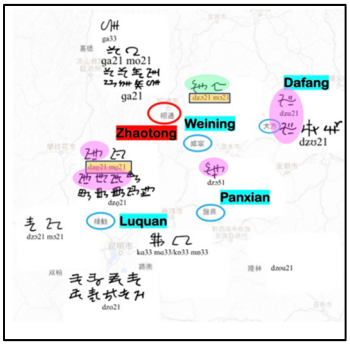
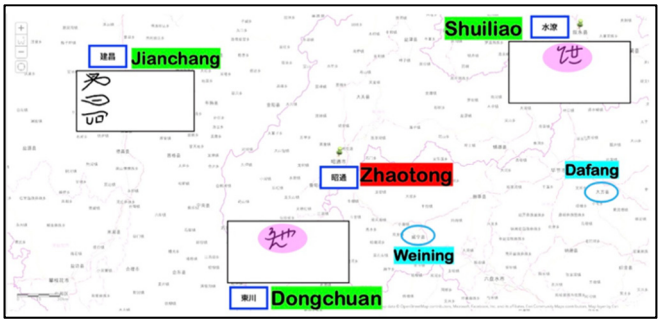
	Book 4	Book 5
Yi character		
Transcription in Ch.	魯補桌	布着
Pinyin	lu bu zhuo	bu zhe

Chart 17: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

19. Road

The characters and transcriptions seem to be close to those in Weining; the modern pronunciation in Luquan is also similar to that in *Book 4*. However, the characters do not correspond. Therefore, it is plausible to determine the origin of *Book 4* as potentially around the Weining area.





Maps 34 and 35: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

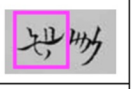
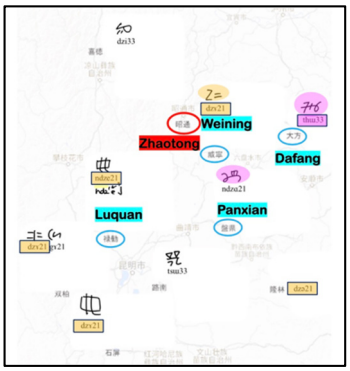
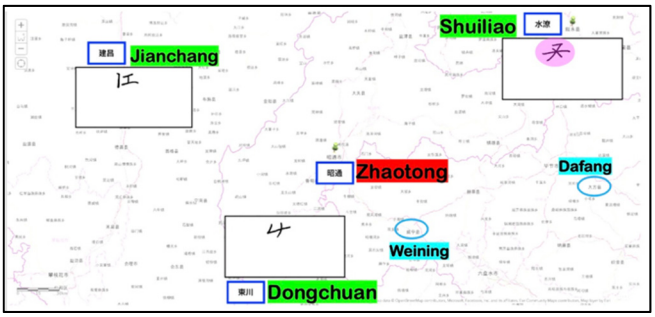
	Book 4	Book 5
Yi character		
Transcription in Ch.	模腳	額腳
Pinyin	mo jiao	e jiao

Chart 18: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

20. Bridge

Both the character and the transcription in *Book 4* bear a strong resemblance to those in Weining, as do those in *Book 5* to those in Dafang. In particular, the transcriptions of 'ze' in *Book 4* and 'tun' in *Book 5* seem to correspond closely to the modern pronunciations of [dzɤ^{21}] in Weining and [thɯ^{33}] in Dafang, respectively.





Maps 36 and 37: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

	Book 4	Book 5
Yi character		
Transcription in Ch.	则	吞
Pinyin	ze	tun

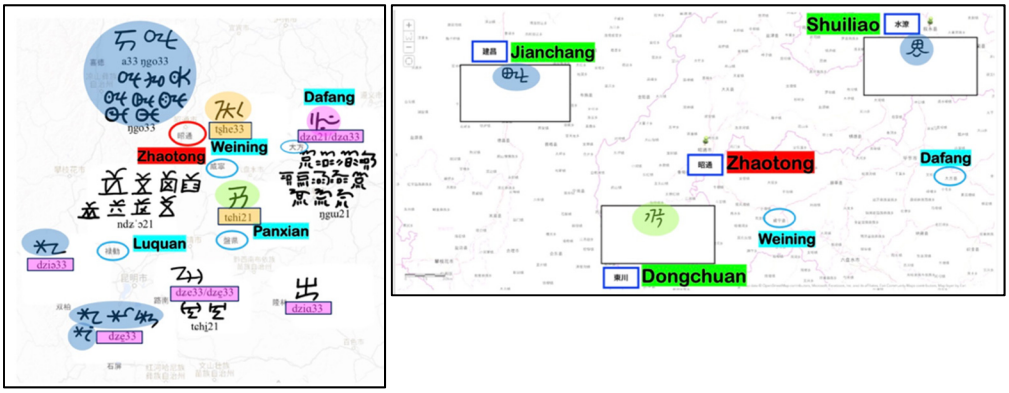
Chart 19: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

2.1.3. Part III. Time

Here, I explore the maps concerning the four vocabularies from Time. The numbers on the maps are provided for convenience and correspond to the order in the *Imanishi Book* of *Hua-Yi Yiyu*.

10. To be cold

The characters in *Books 4* and *5* are very similar to those in Weining and in Dafang, respectively. Additionally, the phonetic values of Weining and Panxian seem to correspond to those in *Book 4*. Therefore, the character in *Book 4* must be from the area around Weining or Panxian.



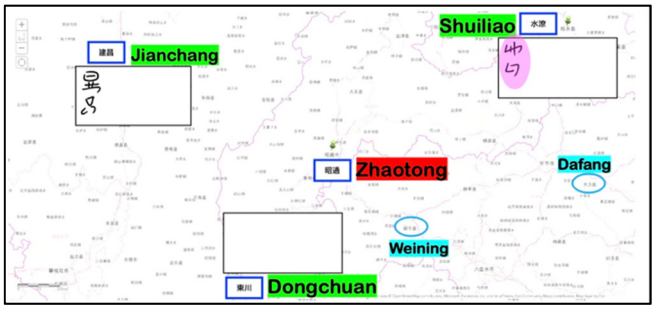
Maps 38 and 39: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

	Book 4	Book 5
Yi character		
Transcription in Ch.	尺	家
Pinyin	chi	jia

Chart 20: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

20. This year

The characters in *Books 4* and *5* correspond to those in Weining and Dafang, respectively. However, while the data in *Book 4* correspond to the data in Weining, and the data in Dafang correspond to those in *Book 5*, their transcriptions also seem to be found in other locations.



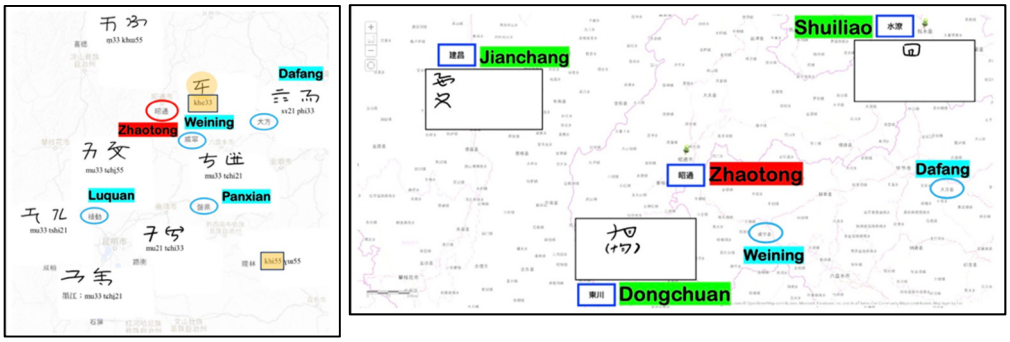
Maps 40 and 41: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

	Book 4	Book 5
Yi character		
Transcription in Ch.	渴赤	課齒
Pinyin	ke chi	ke chi

Chart 21: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

28. Night

Given the transcription in *Book 4*, we can identify two candidate locations. However, in terms of the character, Weining is most likely to be the origin of *Book 4*.



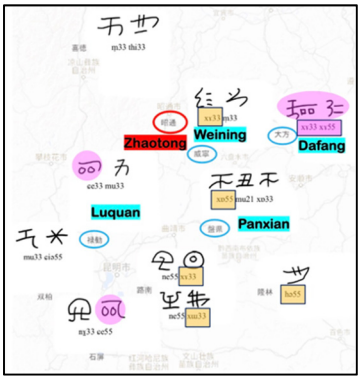
Maps 42 and 43: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

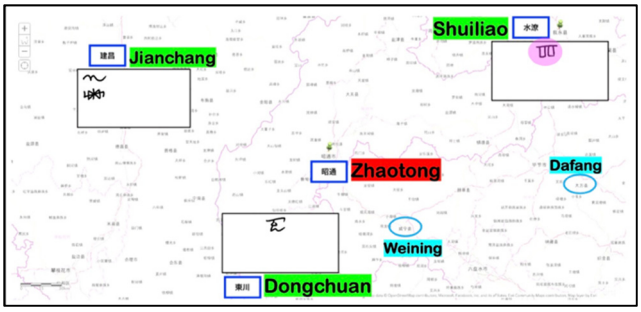
	Book 4	Book 5
Yi character		
Transcription in Ch.	窺	快
Pinyin	kui	kuai

Chart 22: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

30. Morning

Despite slight differences, in characters in *Book 5* and those in Dafang match very closely, as do their phonetic values. Therefore, it is highly probable that the lexical item in *Book 5* comes from the area around Dafang.





Maps 44 and 45: Yi-character map of the whole area, and map based on the data of *Hua-Yi Yiyu*

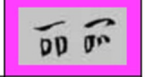
	book 4	book 5
Yi character		
Transcription in Ch.	黑	黑黑
Pinyin	hei	hei hei

Chart 23: The Yi characters and their transcriptions into Chinese characters of *Books 4* and *5*, with pinyin

3. Final Remarks: The origins of *Books 4* and *5*

This study drew on 138 lexical items from seven books of *Hua-Yi Yiyu* to create Yi-character maps, of which 21 were shown. On the presented maps, I have counted the number of identical or similar characters and transcriptions which seem to bear certain resemblances to the phonetic data of the modern dialects. Chart 24 below indicates the results.

As mentioned above, the data of the Yi characters and transcriptions in *Hua-Yi Yiyu* are considered to have been collected separately. This means that we should focus less heavily on the phonetic information and more intensely on the Yi characters themselves (although the phonetic transcriptions are still our important guidance). Through this investigation, the Yi characters' logographic nature also pushed the geolinguistic approach forward.

Chart 24 indicates that the origin of *Book 5* is highly probably around the Dafang area, as its characters (10 of 21 lexical items) and phonetic values (10 of 21 lexical items) display strong resemblances to those of the Dafang region.

At the same time, but with less certainty, the origin of *Book 4* might be around the Weining area, although further research is still essential. According to Nie et al. (2019), the language written in *Book 4* is plausibly a Yi dialect spoken in Weining; my preliminary but thorough investigation concurs. I will continue to explore this issue by focusing more intensively on the Yi characters in *Book 4*.

	similar/ correspondence	(a) found only in Weining	(b) found only in Dafang	(c) found in other Eastern dialects	total number (a)+(b)+(c)	found among other Yi dialects
Book 4	character	5/21	0/21	3/21	8/21	5/21
	phonetic value	4/21	0/21	7/21	11/21	9/21
Book 5	character	0/21	10/21	6/21	16/21	5/21
	phonetic value	0/21	10/21	3/21	13/21	6/21

Chart 24: The number of similarities from the data in *Books 4* and *5*

This study succeeded in indicating the possible origins of *Books 4* and *5*; however, conclusively detecting their origins is still ongoing research. Since both books contain 740 lexical items, I will make the most possible use of them to draw a firm and reliable conclusion in the near future.

References

Chunhua [春花], Li, Ying [李英], and Guo, Jinfang [郭金芳] (2018)「《清乾隆年編《華夷譯語》述論」 [A Study of Hua-Yi Yiyu Compiled in the Qianlong Dynasty]. *Journal of Gugong Studies* 2018.19: 380-396.

Gugong Bowuyuan [故宮博物院] (ed.) (2018) *Gugong Bowuyuan Cang Qianlongnian Bian Hua-Yi Yiyu* 《故宮博物院藏乾隆年編華夷譯語》[*Hua-Yi Yiyu compiled during the Qianlong reign years preserved in the Palace Museum*]. Vol. 1, 5-6. Beijing: The Palace Museum Press.

Ishida, Mikinosuke [石田幹之助] (1973 [1931]) Joshingo Kenkyū no Shin-shiryou 「女真語研究の新資料」[New materials on Jurchen language]. *Tōa Bunkashi Sōkō, Tōyōbunko Ronsō* 54: 3-69.

Iwasa, Kazue (2017) Mapping the Yi characters – an attempt to explore a route of their historical change and origin –. *Studies in Asian Geolinguistics, Monograph series no.2, Proceedings of the Workshop "Geolinguistic Method and Southeast Asian Linguistics"*, 59-67. Fuchu: ILCAA, Tokyo University of Foreign Studies. URL: https://publication.aa-ken.jp/proceedings_geolinguistic_SouthEastAsian_2017.pdf

Iwasa, Kazue (2018) *Remarks on Maps of the Yi Script Based on the Swadesh 100 Wordlist. Studies in Asian Geolinguistics, Monograph Series No. 5*. ILCAA 2015–2017 Joint Research Project Studies in Asian Geolinguistics Research Institute for Languages and Cultures of Asia and Africa Tokyo University of Foreign Studies. URL: https://publication.aa-ken.jp/sag_mono5_yi_2018.pdf

Iwasa, Kazue [岩佐一枝] (2019) Dili Yuyanxue Yuanli de Yingyong –Yi Yiwen Weili-「地理語言學原理的應用 －以彝文為例－」[An application of geolinguistic principles - The Yi script as a case study -], *Collected Papers on Eastern Asian Geolinguistics. Studies in Asian Geolinguistics, Monograph Series*《亞洲地理語言學研究》單刊 第六 東部亞洲地理語言學論文集 *No. 6*, 131-144. ILCAA 2015–2017 Joint Research Project Studies in Asian Geolinguistics Research Institute for Languages and Cultures of Asia and Africa Tokyo University of Foreign Studies. URL: https://publication.aa-ken.jp/sag_mono6_eastern_asian_2019.pdf

Kong, Xiangqing [孔祥卿] (2016) Luoluo Yiyu de Banben ji Jiaokan「猓玀譯語的版本及校勘」[The Editions and Text Criticism of *Luoluo Yiyu*]. *Minzuyuwen*《民族語文》[*Minority Languages of China*] 2016.2: 71-82.

Kuribayashi, Hitoshi [栗林均] (2003) *Kaiyakugo (Kōshubon) Mongorugo Zentango Gobi Sakuin*『華夷譯語（甲種本）モンゴル語全単語・語尾索引』[*Hua-Yi Yiyu (Jiazhongben) Index of All Mongolian Words and Suffixes*], *Northeast Asian Studies Series 10*. Sendai: Center for Northeast Asian Studies, Tohoku University. URL: http://hdl.handle.net/10097/54697

Matsukawa, Takashi [松川節] and Nie, Hongyin [聶鴻音] (2013) Ōtani Daigaku Shozō *Roro Yakugo* Jutsuryaku「大谷大学所蔵『倮儸譯語』述略」[The Summary of Lolo Vocabulary preserved at Otani University]. *Kaiyakugo (Seibankan Yakugo Yonshu Roro Yakugo Isshu Eiin to Kenkyu*『『華夷譯語（西番館譯語四種　倮儸譯語一種）』影印と研究』[The

Photographic Reproduction and study on *Hua-Yi Yiyu* (Four Tibetan Vocabularies and One Lolo Vocabulary)], *Ōtani Daigaku Bunken Kenkyū Sōsho 1*, 23-29. Kyoto: Shokado.

Nie, Peng [聶鵬] (2018) Yongningshu Shuiliao Luoluo Yiyu Jiaoshi「《永寧屬水潦譯語》校釋」 [The collation and annotation on Shuiliao Lolo Vocabulary in Yongning]. *Minzu Yuwen*《民族語文》 [*Minority languages of China*] 3: 81-97.

Nie, Peng [聶鵬] (2020) Gugong cang Dongchuan Fushu *Luoluo Yiyu* Kaolun'「故宮藏東川府屬《猓玀譯語》考論」[A discussion about *Lolo Vocabulary* of Fongchuanfu preserved at the Palace Museum]. *Zhongguo Wenzi Yanjiu* 《中國文字研究》[*Journal of Chinese Writing Systems*] 32: 223-231.

Nie, Peng [聶鵬], Kong, Xiangqing [孔祥卿], and Wen, Zhi [文智] (2019) Gugong cang Disiben《Luoluo Yiyu》Jichu Fangyan Kaozheng 「故宮藏第四本《猓玀譯語》基礎方言考證」[A Textual Criticism on a basic dialect of the fourth book of *Lolo vocabulary* preserved at the Palace Museum]. *Minzufanyi* 《民族翻譯》2: 22-34.

Nie, Peng [聶鵬] and Kong, Xiangqing [孔祥卿] (2020) Dagu Daxue cang《Luoluo Yiyu》Yizhe Shenfen Kao ―Jian Yu Songchuan Jie, Nie Hongyin Xiansheng Shangque 「大谷大學藏《猓玀譯語》譯者身分考－兼與松川節、聶鴻音先生商榷」[A Discussion on the identification of a translator of *Lolo Vocabulary* preserved at Ōtani University]. *Wenxian*《文獻》5: 86-99.

Nishida, Tatsuo (1979) *Roro Yakugo no Kenkyu: Rorogo no Kōzō to Keitō*『倮儸譯語の研究：倮儸語の構造と系統』 [A study of the Lolo-Chinese vocabulary Lolo i-yu: The structure and linage of Shui-liao Lolo]. *Ka-i Yakugo kenkyu sosho 6*. Kyoto: Shokado.

Wen, You [聞宥] (1940) Luoluo Yiyu Kao 「倮儸譯語攷」[On Lo Lo I Yu―Lolo-Chinese Vocabularies]. *Studia Serica*, *Journal of The Chinese Cultural Studies Research Institute*, 77-97. Chengdu: The Chinese Cultural Studies Research Institute, West China Union University.

Yamazaki, Tadashi [山崎忠] (1953) Waga Kuni niokeru *Kaiyakugo* Kenkyūshi 「我が国における華夷譯語研究史」[The Study History of *Hua-Yi Yiyu* in Our Country]. *Journal of the Academic Association of Koreanology in Japan* 5: 45-58.

Yamazaki, Tadashi [山崎忠] (1954) Waga Kuni niokeru *Kaiyakugo* Kenkyūshi Hoi 「我が国における華夷譯語研究史補遺」[Appendix of the Study History of *Hua-Yi Yiyu* in Our Country]. *Journal of the Academic Association of Koreanology in Japan* 6: 163-165.

[Dictionaries]

Bai, Xianyun (ed) [白顯雲主編] (1995)*Yi-Han Zidian*《彝漢字典》[*Yi-Chinese Character Dictionary*]. Kunming: Yunnan Minzu Chubanshe.

Dianchuanqiangui Yiwen Xiezuozu (ed.) [滇川黔桂彝文協作組編] (2004) *Dianchuanqiangui Yiwenziji* 《滇川黔桂彝文字集》 [*Collection of Yi characters of Yunnan, Sichuan, Guizhou, and Guangxi*]. Guizhou/Chengdu/Kunming: Guizhou Minzu Chubanshe. Sichuan Chuban Jituan. Sichuan Minzu Chubanshe. Yunnan Minzu Chubanshe.

Li, Zenghua (ed.) [李增華主編] (2014) *Guyiwen Changyong Zidian*《古彝文常用字典》[*Dictionary of Commonly Used Old Yi Characters*]. Kunming: Yunnan Minzu Chubanshe.

Pu, Zhangkai (ed.) [普璋開主編] (2005) *Diannan Yiwen Zidian*《滇南彝文字典》[*Dictionary of Diannan Yi Characters*]. Kunming: Yunnan Minzu Chubanshe.

Wu, Zili [武自立], Ji Jiafa [紀嘉發] (2011) *Han Yi Jianming Cidian*《漢彝簡明詞典》[*Chinese-Yi Concise Dictionary*]. Chengdu: Sichuan Minzu Chubanshe.

Yunnansheng Lunan Yizu Zizhixian Wenshi Yanjiushi (ed.) [雲南省路南彝族自治縣文史研究室編] (1984) *Yi Han Jianming Cidian*《彝漢簡明詞典》[*Yi-Chinese Concise Dictionary*]. Kunming: Yunnan Minzu Chubanshe.

Zhang, Jinzhi [張晉智] (2014) *Jianming Yi Han Zidian*《簡明彝漢字典（祿勸版）》[*Concise Yi-Chinese Dictionary (Luquan version)*]. Kunming: Yunnan Minzu Chubanshe.

Zhongyang Minyuan Yizu Lishi Wenxian Ban (ed.) [中央民院彝族歷史文獻班主編] (1984) *Dianchuanqiangui Yi Han Jiben Cihui Duizhao Cidian*《滇川黔桂彝漢基本詞彙對照辭典》[*Yi-Chinese Comparative Dictionary of Basic Vocabulary in Yunnan, Sichuan, Guizhou, and Guangxi*]. Beijing: Zhongyang Minyuan Yizu Lishi Wenxian Bianyishi.

Characteristics and Distribution of the Variants of Maternal Kinship Terms in Son Tay Dialect (Hanoi)

Trịnh Cẩm Lan
(Faculty of Linguistics, USSH, VNU Hanoi)

Abstract: The paper describes the features and geographical distribution of local variants of 4 words in Son Tay dialect, Hanoi: mother, grandmother, mother's brother and mother's sister. Approaching local variants from both the directions of Dialectology and Geolinguistics, the study achieves 3 results: (1) Describes the features of the local variants of 4 words: mother (mẹ, mệ); grandmother (bà ngoại, bà vãi); mother's brother (bác, cậu); and mother's sister (già, bá, bác, cô); (2) Explains their existence with chronological data by reconstructing the history of words, with the data on the languages having origin and contact relationships. The results show that variants reflect the linguistic contact between Vietnamese and the Viet-Muong languages, as well as Mon-Khmer and Austronesian, Tai-Kadai and, especially, Chinese; (3) Maps the geographic distribution of variants, explains their distribution in space, and shows their relationship to their changes in time.

Key words: dialect, Son Tay dialect, variant, reconstruction, distribution map.

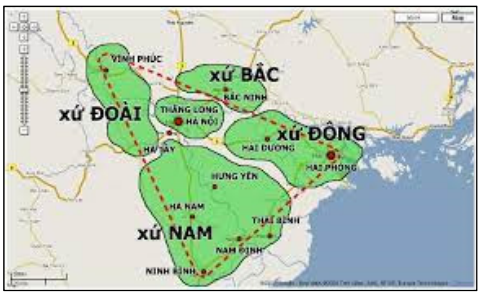
1. Introduction

Son Tay – xu Doai (the Land of Doai) – is the cradle of the ancient Vietnamese (Tran Quoc Vuong, 1997; Lam Ba Nam, 1997; Nguyen Luong Bich, 1997), one of the four areas protecting Thang Long (Hai Dong, Son Tay, Son Nam, Kinh Bac). Although the administrative boundaries have changed over time, this is thus still a special land in terms of history, culture and language, to the west of the Capital. According to many researchers, xu Doai is the place where "the structure of wet rice culture of Thang Long capital is preserved" (Pham Duc Duong, 1997). The two major centers of xứ Đoài during the Tran, Le and Nguyen dynasties were the two towns of Quang Oai and Quoc Oai, which belong to the districts of Quoc Oai, Thach That and Ba Vi today (Dang Van Tu, Nguyen Ta Nhi, 2011: 15-17). In addition, Ba Vi is also said to be the place of existence and development of the ancestral inhabitants of the Vietnamese and Muong people (Lam Ba Nam, 1997), considered the 'original area' of the Viet – Muong

TRỊNH, Cẩm Lan. 2023. Characteristics and distribution of the variants of maternal kinship terms in Son Tay dialect (Hanoi). In Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 203–216. doi: https://doi.org/10.5281/zenodo.8374677

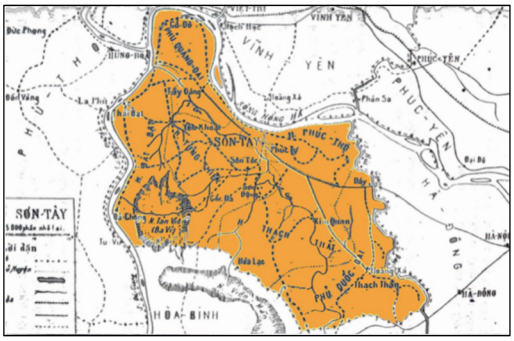
Common people (Tran Quoc Vuong, 1997). These are the two centers for receiving and spreading the culture – language of xu Doai (the Land of Doai).

The two maps below provide a basic visualization of xu Doai. Map 1 shows the location of xu Doai with Thang Long - Hanoi, with the remaining three area of xu Dong, xu Son Nam and xu Kinh Bac, forming four towns. Map 2 is Son Tay province during the French colonial period. This map shows that the boundaries of xu Doai have some historical changes, but the basic shape remains.



Map 1. Location of Sơn Tây – xứ Đoài in the cultural space of the Northern Delta (one of the four towns of Thang Long)

Source: Trần Anh Tuấn, 2019.



Map 2. Map of Son Tay province during the French colonial period (1924)

Source: Đỗ Đình Nghiêm, Ngô Vi Liễn, Phạm Văn Thư, 1930: 101.

The land belonging to the above districts is also an area with a special dialect, called Son Tay dialect. A special feature of this language is a special timbre in the pronunciation of certain phonemes of the system including tones, some vowels and consonants (Nguyen Tai Thai, 2015; Trinh Cam Lan, 2023). In addition, Son Tay dialect also has many different lexical variations, as the result of' the convergence of many dialects. In particular, there are many distinctive features showing that this is a dialect area that still preserves many ancient features of Vietnamese.

The aim of the present study of Son Tay dialect is to: (1) describe the lexical-semantic characteristics of local lexical variations of some kinship words from the maternal side (mother's side), using contemporary materials; (2) compare, and find the etymological origin of, these variations, based on chronological data, to explain the causes of their existence; and (3) produce maps showing the geographical distribution of variations in this space.

The land belonging to the above districts is also an area with a special dialect, called Son Tay dialect. A special feature of this language is a special timbre in the

pronunciation of certain phonemes of the system including tones, some vowels and consonants (Nguyen Tai Thai, 2015; Trinh Cam Lan, 2023). In addition, Son Tay dialect also has many different lexical variations, as the result of' the convergence of many dialects. In particular, there are many distinctive features showing that this is a dialect area that still preserves many ancient features of Vietnamese.

The aim of the present study of Son Tay dialect is to: (1) describe the lexical-semantic characteristics of local lexical variations of some kinship words from the maternal side (mother's side), using contemporary materials; (2) compare, and find the etymological origin of, these variations, based on chronological data, to explain the causes of their existence; and (3) produce maps showing the geographical distribution of variations in this space.

2. Methods and materials

The study was conducted via a survey of 82 places in three districts, Quoc Oai, Thach That and Ba Vi, in the old Land of Doai which is said to be the cradle of the Son Tay dialect.

The data set or this paper is the lexical variants of mother's kinship words in the Son Tay dialect.

In order to explain the presence of these variations, the paper describes synchronously the semantics of variants, while also using the method of comparison and reconstruction of the etymology and finding the historical relationships, to explain the presence of these variants. The comparative results help to further explain the characteristics of the variants and their origins.

3. Results and discussions

The surveyed kinship words of the maternal side consist of 4 words with different variants, that is: (1) indicating "a woman with a child, speaking in relation to her child" (Hoang Phe, 2015: 984) with two words, *mẹ* /mɛ^{21}/ and *mệ* /me^{21}/; (2) referring to "mother's sister, can be used to address" (2015: 38) with words such as *bác* /bak^{35}/, *già* /za^{32}/, *cô* /ko^{33}/, *bá* /ba^{35}/; (3) referring to "mother's brother, can be used to address" (2015: 212) with two words, *bác* /bak^{35}/, *cậu* /kɤ̆w^{21}/; and (4) referring to “grandmother, which can be used to address” (2015: 36) with two words, *bà ngoại* /ba^{32}ŋwaj^{21}/, *bà vãi* /ba^{32}vaj^{325}/.

When studying the lexical differences between dialects, according to Hoang Thi Chau, it is necessary to distinguish two different types: the first, they are partially different words due to the historical development of Vietnamese itself (usually due to phonetic changes); and the second, they are completely different words because they have different origins (Hoang Thi Chau, 2004: 100-101). In trying to explain the origin of the lexical variants, we will examine the origins of the lexical variants from these two perspectives.

3.1. Characteristics and distribution of variants indicating "mother"

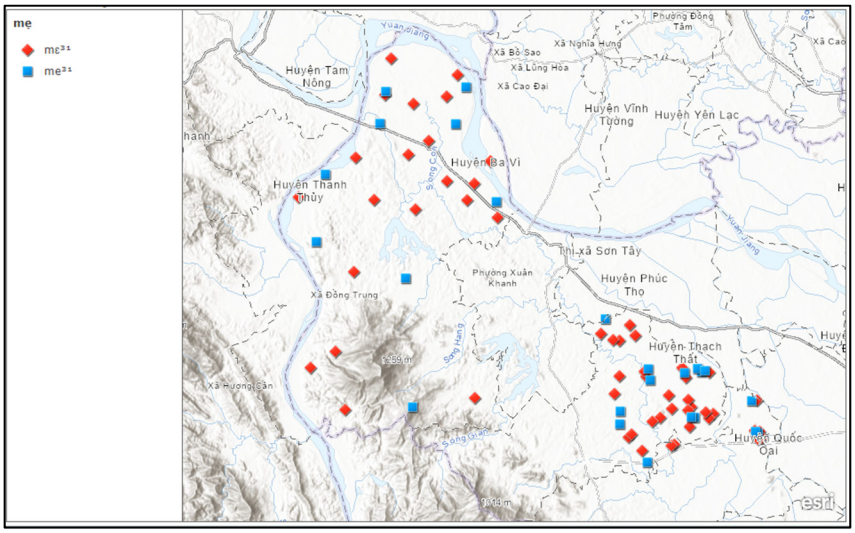
Mẹ (mother) is a word in modern Vietnamese with many local variants such as *bầm* /bɤ̌m^{32}/, *bu* /bu^{33}/, *má* /ma^{35}/, *mạ* /ma^{21}/, *me* /mɛ^{33}/, *mế* /me^{35}/, *mệ* /me^{21}/, *nạ* /na^{21}/ and *u* /ʔu^{33}/ (Hoang Phe, 2015: 985). Besides *mẹ* (common variant), in the study area there is also a dense presence of the word *mệ* in all 3 districts, especially in Ba Vi. These are two different words of the first type, which due to phonetic changes have two different vowels, but with both are the front vowels.

In terms of chronology, in the 17th century, the word *mẹ* was noted to be common in Vietnamese (A. de. Rhodes, 2021: 199). Variants of this word in the Muong dialects, as noted by Nguyen Van Tai, are quite rich, with *me, mệ, mạng, bầm* (Nguyen Van Tai, 2005: 240). Other Austroasiatic languages include /*mɛɛʔ/, /*mɛʔ/, /*maeʔ/, /*mee/, /*mèe/ (F.E. Huffman, 1977). In Proto Mon-Khmer, the reconstructed form of *mẹ* is /*meeʔ/ (Vu Duc Nghieu, 2011: 74). In some other branches of Mon-Khmer, there are also similar forms. For example, there are /*meeʔ/ in Proto West-Bahnaric, /*me:/ in Proto South-Bahnaric, /*me:ʔ/ in Proto Central-Bahnaric, /*maa, mɛʔ/ in Bana, /*me/ in Old Khmer, and /*mae/ in Khmer (A Database of Basic Austronesian Vocabulary – Basic South Asian Vocabulary Database). In terms of etymology, *mẹ* and *má/mạ* are the two forms found in both Austroasiatic and Tai-Kadai languages (Alves 2017). Thus, these reconstructions show the origins of *mẹ* quite clearly.

The relationship between the Mon-Khmer origin, Austroasiatic family with *mẹ* and *mệ* is also supported by synchronous data which shows that these variants are still used quite commonly in the Central dialect, a dialect that preserves many of the original elements of the Vietnamese language. In particular, the *mệ* variant is also found in the Muong Bi, Ba Trai and Muong Khoi regions (the Muongs in or near the former Ha Tay area) and other Muongs in Thanh Hoa, Nghe An and Ha Tinh (Nguyen Van Tai, 2005: 240). This shows that the *mệ* variant in Vietnamese, that is quite common in the study area, is most likely due to the preservation of the ancient pronunciation variant of the Vietnamese language, which is thus still retained by the Muong language. Variations

of the word *mẹ* in other Vietnamese dialects are also very rich (*bầm, bu, má, mạ, me, mệ, mế, mệ, nạ, u* - /bɤ̆m^{32}, bu^{33}, ma^{35}, ma^{21}, mɛ^{33}, me^{35}, me^{21}, na^{21}, ʔu^{33}/) (Hoang Phe, 2015: 985), but the words *mệ* and *mạ* are very common in North Central Vietnam (Nguyen Nha Ban, 1999). They, Ha Tay and North Central Vietnam, are spaces with Viet - Muong cohabitation and/or contact .

The spatial distribution of these variations can be observed in the following map, in which both variants are shown to be widely distributed; however, mẹ is the later, more modern, more densely distributed variant.



Map 3. Distribution of variations of the words *mẹ*

3.2. Characteristics and distribution of words for mother's sister

In the group of kinship words, the word for mother's sister has the most variants: the word *già* appears at 61/82 places of the investigation, *bác* appears at 10/82 places, *bá* appears at 8/82 places, and ***cô*** appears at 3 places. These four lexical variants have quite fundamental differences.

Firstly, *già* is the word most used in the area. This word is annotated as mother's sister and is a local word, synonymous with *bá* (Hoang Phe, 2015: 605). In the Vietnamese Dialect Dictionary, *già* is also annotated as mother's sister, used in the Northern dialect (Pham Van Hao, 2009: 194). In Muong language, data from Nguyen Van Tai show that, among the words for father (or mother) older sister, there is *già/giạ* (2005: 174). In Chut language, the father (or mother) older sister is /ja:^{32}/, while in Proto Mon-Khmer there is /*jaʔ/ (A Database of Basic Austronesian Vocabulary). With

the current data, 'as no further source can be found to help in further in-depth discussion about the origin of this word, we temporarily accept this as a native word, an ancient variant of Vietnamese (at least from the Proto Mon-Khmer period) which is still preserved and used in some Viet and Muong dialects.

The second most commonly occurring word is *bác*. According to the Vietnamese dictionary, *bác* is "father's brother or father's sister-in-law (can be used to address)" (Hoang Phe, 2015: 38). This is a word that is widely used in Vietnamese dialects. In terms of origin, *bác* is a kinship word of Chinese origin. According to Alves, in Chinese loanwords in Vietnamese, there are a series of kinship words such as *bác* (伯 bo), *cô* (姑 gū), *cậu* (舅 jiǔ, Sino-Vietnamese: *cữu*), chú (叔 shū, Sino-Vietnamese: *thúc*) and *chị* (姊 jiě, Sino-Vietnamese: *tỉ*). They are said to be Chinese loanwords at a very early stage (Early Sino-Vietnamese loanwords). These words might have been borrowed from ancient Chinese in the early period of Sino-Vietnamese contact (early AD – Qin dynasties, Han to early Tang dynasties), and were borrowed not only into the Viet-Muong languages but into many other Southeast Asian languages (Alves, 2017). However, in many areas of the Northern Dialect (including Son Tay) in the last few decades, *bác* has tended to move from the paternal side to the maternal side, to refer to mother's sister (and thus competes in terms of use with *già* – an native and older variant).

The third is the word *bá* /ba^{35}/. This word is sometimes thought to be a Sino-Vietnamese word, related to *bác* /bak^{35}/ (Tran Thi Hong Hanh, 2021), and now, *bá* still appears in limited combinations such as *anh em thúc bá* and *thúc bá huynh đệ*. However, this is a word with a very complex etymological origin. Nguyen Van Tai's data also show that the word for father's sister-in-law or father's sister in Vietnamese corresponds to *bả, pả, bá, pá* and *zà/ zạ (/ba^{233}, pa^{323}, ba^{35}, pa^{35}, za^{32}, za^{21}/)* in the Muong dialects (Nguyen Van Tai, 2005: 174). This leads the present researcher to immediately think of the relationship with Muong language, which is evident in the vast majority of the above phonetic variations. However, in turn, *bá* in both Vietnamese and Muong languages is said to be a word of Thai origin (Tai loanword) from the Proto-Vietic period, meaning that its history of borrowing is quite deep. Based on historical phonological evidence, Alves suggests that /*bəːʔ/ in Proto-Vietic means daughter-in-law and /*paːʔ/ in Proto-Vietic means wife of father's brother (i.e. uncle-in-law) or father's (mother's) sister (i.e. aunt) who is now called *bá* (in many places of the North to the North of Central), all of which are of Thai origin (Alves, 2020). The Tay – Nung – Viet dictionary and the Thai – Viet dictionary both have *pả* entries that are annotated as *bá* (means father's sister); while the Thanh Hoa Thai – Viet dictionary also has *pá*

and is also annotated as father's sister (Tran Thi Hong Hanh, 2021). All this evidence supports a Tay – Thai origin for *bá* in the Viet – Muong languages.

From the above it can be seen that these languages use *bá* to refer to the father's sister, but in the Son Tay dialect, *bá* is used to call the mother's sister. Thus, meaning of *bá* changes from the paternal side to the maternal side.

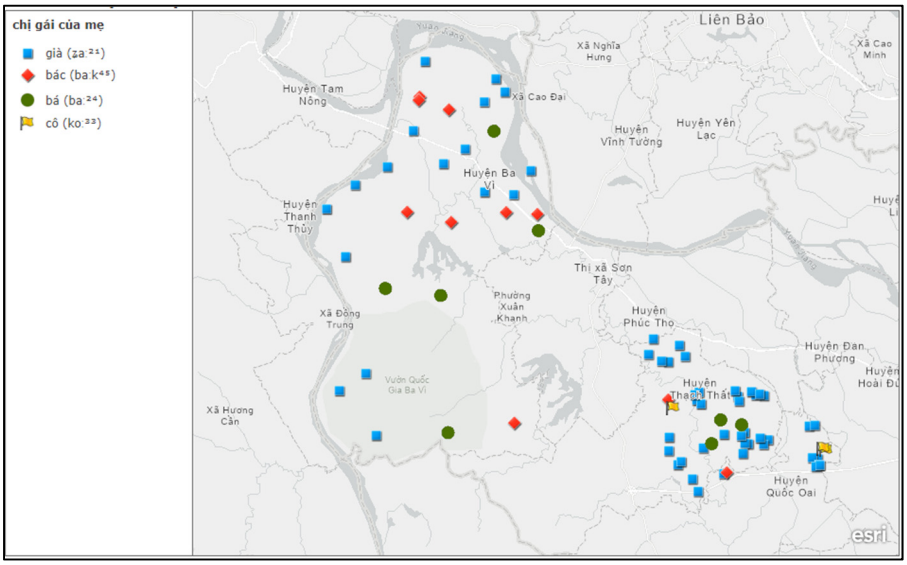
The word *cô* /ko^{33}/ is a word with a phonetic form different from the rest. In the Hoang Phe Dictionary, *cô* is referred to as father's younger/older sister [can be used to address] (Hoang Phe, 2015: 331). In the Vietnamese - Portuguese - Latin Dictionary, *cô* is interpreted as father's sister (2021: 101). In Muong language, the words, /va^{6}, va^{5}, ko^{2}, ba^{3}, ɣwa^{5}, ɔ^{1} and ɔ^{2}/^{1}, are used to call the *father's sister* (Nguyen Van Tai, 2005: 200). In terms of origin, *cô* is a Sino-Vietnamese loanword, borrowed from around Tang dynasty into Proto Viet-Muong, to refer to father's sister (Alves 2017). Thus, in both the original and borrowed form, *cô* was originally a word to call a blood-related woman in the same generation as the father (paternal side) but without distinction between the older and the younger. However, in the Son Tay dialect, there is a phenomenon of using *cô* to refer to the mother's sister. This means that *cô* has changed in meaning. Accordingly, the distinction between the older and the younger is inherently absent, and the distinction between the paternal side and the maternal side is blurred, or even lost, and only the distinction of gender remains. In the Muong data of Nguyen Van Tai, father's older sister is used as *bả, pả, bá, pá* and *zà/zạ*, with no form related to *cô*. In terms of languages that can be contacted, the word for mother's sister in the Tay Thai languages is *pả*, and in the Cham language, *kabăk* is an unrelated form.

Considering all the sources, so far, we have not found a contact that can lead to the transfer of meaning from the paternal side to maternal side. However, it is a fact that, in the Hanoi and surrounding areas, in the last few decades, there has been a quite common trend: calling a mother's younger sister as *cô* (instead of *dì*, the word commonly used before to call mother's younger sister) but mother's older sister as *bác*. Thus, the distinction between the paternal side and maternal side of both these words has tended to fade but the upper-lower distinction still exists. We think of an endogenous cause in Vietnamese, which is a self-semantic change. The first cause is to transfer *cô* from the paternal side to the maternal side, but only to call mother's younger sister. Then, due to the influence of the non-discrimination of older – younger in many kinship words in Vietnamese such as *dì* (mother's older/younger sister), cậu (mother's older/younger brother) and *cô* (father's older/younger sister), *cô* (in the maternal side) gradually has a new semantic feature (older – younger non-discrimination), to call both

[1] We use the author's phoneme recording.

mother's older/younger sister. Of course, this is just a hypothesis and can be re-discussed or even disproved when more available and convincing data is found.

The spatial distribution of these variations can be observed in the following map.



Map 4. Distribution of variants for mother's older sister

It can be seen that *già*, an Austroasiatic variant, is distributed densely and widely in this space; but *bá* and *bác*, variants borrowed by contact, are more sparsely distributed; while *cô*, the new variant (showing the trend of modernization in meaning), only appears in the suburbs of Hanoi. This is a gravity model (a dialect feature spreads, as a 'jumping' type, between urban centers and then to surrounding rural areas).

3.3. Characteristics and distribution of variants for mother's older brother

To refer to the mother's older brother, Sơn Tây dialect uses two words, *bác* /bak^{35}/ and *cậu* /kɤ̆w^{21}/. The word *bác* is found in 62/82 places, while the word *cậu* is found in 20/82 places. These are also two words that have a fundamentally different form. Theoretically, dissimilar origins can be considered.

Bác, as mentioned, is a Sino-Vietnamese word (Alves, 2017). The Vietnamese dictionary considers *bác* as a common word, interpreted to mean father's brother or father's sister-in-law (can be used to address) (Hoang Phe, 2015: 38). Thus, although it is a common word with a wide range of use in all regions, although at first including both the original meaning in Chinese (Dao Duy Anh, 2013:15) and the meaning in

Vietnamese, ***bác*** is now only used to refer to the paternal side. The maternal side meaning (mother's brother) is probably the meaning formed later.

The Vietnamese - Portuguese - Latin dictionary uses "cạu" to refer to the mother's brother (A. de. Rhodes, 2021: 91). In Muong language, to call the mother's older/younger brother, there are also the words *cụ* /ku^{5}/, *cậu* /kɤ̆w^{5}/ and *gụ* /ɣu^{5}/ in different dialects (Nguyen Van Tai, 2005: 188). In the Vietnamese Dialect Dictionary, *bác* is not annotated as a local word. Perhaps, the word *bác* is used to refer not only to the father's brother but also to the mother's brother, in many regions of the Northern dialect, because the distinction between paternal and maternal side gradually disappears, and accordingly, the meaning of the word is expanded. This is a trend of modernization in meaning seen in some other kinship words such as *cô* analyzed above.

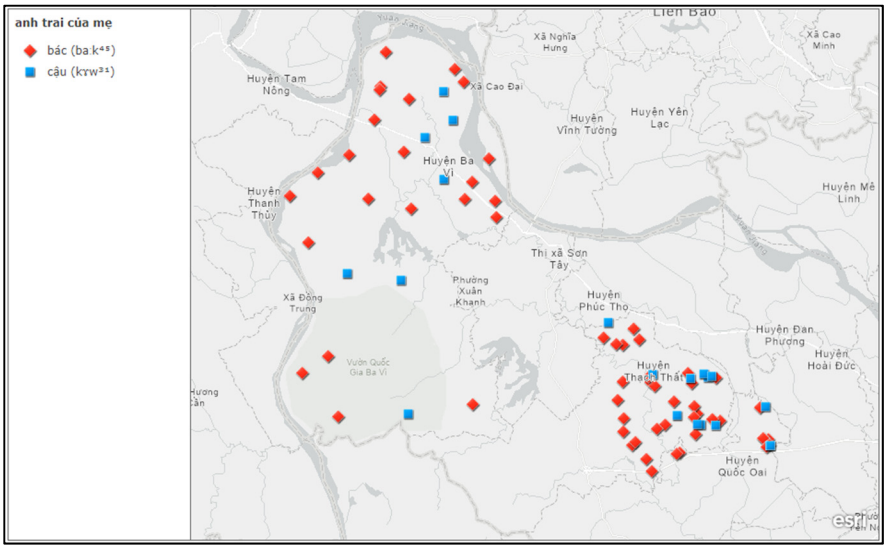
Along with *bác*, Sơn Tây dialect also uses *cậu* /kɤ̆w^{21}/ to call the mother's older brother, a word with a different phonetic form. According to Hoang Thi Chau, completely different words such as these are often due to different origins (2004: 101). However, the materials we found were not like this. Going back in history, we see that the Vietnamese - Portuguese - Latin dictionary has *cạu /kăw^{21}/* (A.de.Rhodes, 2021: 91) to call the mother's older brother. In the 30 Muong dialects, there are 3 words (actually 3 phonetic variants) *cụ* /ku^{5}/, *cậu* /kɤ̆w^{5}/ *and* gụ /ɣu^{5}/ (Nguyen Van Tai 2005: 188), that are used to call the mother's older brother (Nguyen Van Tai, 2005: 188). Two dialects of Cuoi, at Thai Hoa and Tan Hop (Nghe An), also use /ɣu^{54}/ and /ku^{31ʔ}/ (Nguyen Huu Hoanh 2022) with the meaning of mother's older brother. The word *cậu* in Proto Viet – Muong was restored by Alves as /*guB/ (M.Alves 2020). Although located in the center of the Northern dialect, Son Tay's use of cậu is not the same as in the Northern dialect, but more like the Central dialect and other ethnic minority languages of the same family in the Central region. As such, the *cậu* variant in Sơn Tây is much older than in the regions around. According to historians and ethnographers, the Viet Muong residents migrated from North Central Vietnam along the valleys of the western mountains to the North, and stopped in Hoa Binh, Son La, Phu Tho and Son Tay provinces. In addition, Son Tay is one of the stops of these groups of residents (Tran Quoc Vuong 1997, Lam Ba Nam 1997).

Synchronic data show that, in Vietnamese dialects, Central and Southern dialects also use *cậu* to call the mother's older/younger brother (Pham Van Hao 2009: 88). As mentioned, the results of Alves' research on Chinese loanwords in Vietnamese show that the word *cậu* (舅 jiǔ, Sino-Vietnamese: cữu) belongs to the Chinese loanwords from early contacts. According to Alves, *cậu* is one of a group of words that Benedict (1947) calls 'colloquial' words (basic, native words), as opposed to the official Sino-

Vietnamese reading of Chinese characters (Alves 2017). Vu Duc Nghieu also points out that *cậu* is a Vietnameseized Sino-Vietnamese word (2011: 143). Thus, *cậu* is a Chinese loanword but has a very deep history, from very early contacts; and because of this, it is also present in Proto Viet – Muong, Muong, Cuoi languages as it was just mentioned.

However, it is possible to discuss Alves' result further by showing that *bác* (伯 bo) and *cậu* (舅 jiǔ, Sino-Vietnamese: cữu) are both Chinese loanwords from the early contact period. This is the period where Viet and Muong had not been divided. However, the question arises, why is the form of *bác* absent in Muong and Vietic languages? Meanwhile, *cậu* is widely present in many languages related to Vietnamese, with many similar or very similar forms as the result of a historical phonetic evolution.

The spatial distribution of these variants can be observed in the map 5.



Map 5. Distribution of variants of the word for mother's older brother

3.4. Characteristics and distribution of variants indicating the woman who gave birth to the mother (grandmother)

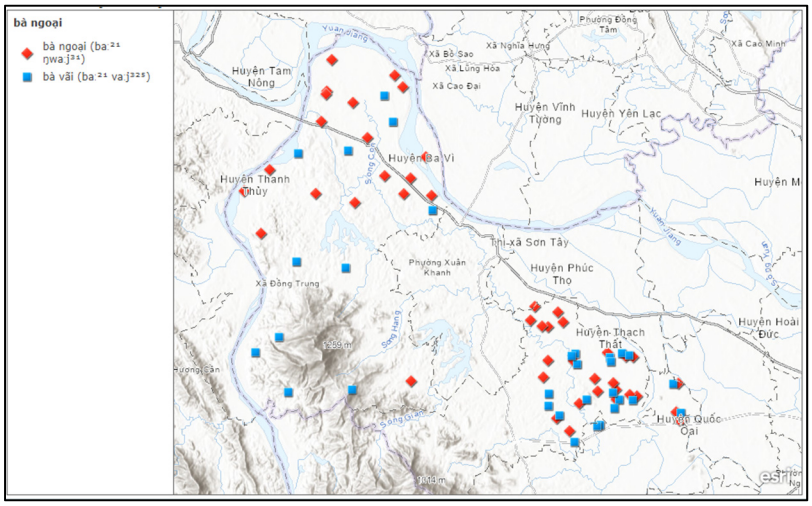
To refer to the woman who gave birth to the mother (grandmother), the districts of Quoc Oai, Thach That and Ba Vi use two words, *bà ngoại* /ba^{32}ŋwaj^{21}/ and *bà vãi* /ba^{32}vaj^{325}/.

The word *bà ngoại* was found in 49/82 survey places, and the word *bà vãi* was found in 33/82 survey places. In Vietnamese, *bà ngoại* is considered a word from the

common vocabulary and is a Chinese loanword (Alves 2017). The Sino-Vietnamese dictionary also shows that both *bà* (婆) and *ngoại* (外) have Chinese etymological origins (Dao Duy Anh 2013: 416).

In the Northern dialect of Vietnamese, the word for grandmother has another variant which is *bà vãi* (Pham Van Hao 2009: 29). The Vietnamese - Portuguese - Latin dictionary has the word *uãi* with the meaning of ancestor, and has the combination *oũ bà oũ uãi* (*ông bà ông vải*) to refer to ancestors in general (A. de. Rhodes, 2021: 316). The word *uãi* in this dictionary is possibly *uãi/vãi* in the combination *bà vãi*. The Muong language has *mệmộng, bả ngwại, pả ngwại* (Nguyen Van Khang 2002), while the Muong dialects have a series of variants such as *bà ngwại, mẹ mộng, mẹ nòn, mẹ ông* and *mộng mụ* (Nguyen Van Tai, 2005: 173). The results of pre-linguistic reconstruction show that, in Proto Mon-Khmer languages, there are the forms, /*yaay/, /*jaʔ/ and /*jaaj/ (Tran Thi Hong Hanh, 2022). With these data, it is possible to infer that *uãi/vãi* with the meaning of maternal in *bà ngoại* is from Austroasiatic origin. The word, *bà vãi*, is a combination of a Chinese element, 婆 (pó), and a Austroasiatic element, *uãi/vãi*. The Austroasiatic word (*vãi*) is competed with by the Chinese loanword and has gradually receded in use in certain regions. Sơn Tây is one of the regions where many ancient elements of Vietnamese are retained, with *bà vãi* as evidence of this.

The spatial distribution of these variants can be observed in the map 6.



Map 6. Distribution of words of grandmother

What is present on the map shows that, although the Chinese variant is very common, the Austroasiatic variant still has a very strong vitality. The dense distribution of *bà vãi* shows that this area can be characterised as quite an ancient linguistic space.

Thus, among the words for maternal kinship, the word *già* (referring to the mother's older sister) and *vãi* (referring to the grandmother) can be considered as of Mon Khmer origin. The distribution of these two elements evenly throughout the map shows that the ancient elements (Proto-Mon Khmer vocabulary) on the migration path of the Proto-Vietic inhabitants were still preserved before being transformed by exposure to the Tai languages (to borrow *bá*) and contact with Chinese (to borrow *bác, bà ngoại* and other Chinese kinship words such as *bác, cô*). What is shown on the map (synchronous) is quite consistent with most of our reconstructions along the chronological scale.

4. Conclusion

The results of this research allow us to confirm the following points:

1. Son Tay dialect preserves many ancient elements of Vietnamese: from Austroasiatic, Proto Mon-Khmer, Proto Viet Muong and Viet Muong.
2. Son Tay dialect preserves many consequences of contact between Vietnamese and geographically close languages from very ancient times: Tai and Chinese languages.
3. The geographical distribution of variants shows that native variants and Early Chinese loanwords from the Vietic period tend to predominate. This situation is quite similar to the ancient dialect island like the Central dialect we all know (Trinh Cam Lan 2021, Trinh Cam Lan 2023a, b).
4. Both Son Tay and Central dialect are two geographic areas witnessing deep Viet – Muong contact.

References

Alves J, Mark (1995) Tonal Features and the Development of Vietnamese Tones. *Working papers in linguistics, Department of Linguistics, University of Hawai'i* 27: 1-14.

Đào Duy Anh (2013) *Hán Việt Từ điển giản yếu* [Concise Sino-Vietnamese Dictionary]. Nxb Văn hoá – Thông tin, Hà Nội.

Nguyễn Nhã Bản (chủ biên) (1999) *Từ điển tiếng địa phương Nghệ Tĩnh* [Nghe Tinh Dialect Dictionary], Nxb Văn hoá thông tin, Hà Nội.

Nguyễn Tài Cẩn (1998) *Giáo trình lịch sử ngữ âm tiếng Việt* [Textbook of Vietnamese Phonetics], Nxb Đại học Quốc gia, Hà Nội.

Hoàng Thị Châu (2004) *Phương ngữ học tiếng Việt* [Vietnamese Dialectology], NXB Đại học quốc gia Hà Nội, Hà Nội.

Phạm Đức Dương (1983), *Nguồn gốc tiếng Việt từ tiền Việt – Mường đến Việt – Mường chung* [Vietnamese origin: from Proto Viet – Muong to Viet – Muong common], trong *Tiếp xúc ngôn ngữ ở Đông Nam Á*, Viện Đông Nam Á xuất bản, Hà Nội.

Phạm Đức Dương (1997), Kinh Bắc – nơi diễn ra sự tiếp xúc Việt Hán, Xứ Đoài - nơi bảo lưu cơ tầng văn hoá lúa nước của thủ đô Thăng Long [Kinh Bac - where the contact between the Viet - China took place, Xu Doai - where the wet rice culture of the Thang Long capital is preserved]. *Tạp chí Nghiên cứu Đông Nam Á*, số 2.

Trần Thị Hồng Hạnh (2021) *Về hai từ bủ, bầm trong thổ ngữ Phú Thọ* [About the two words bủ and bầm in Phu Tho dialect], Kỷ yếu Hội thảo khoa học quốc tế "Nguyễn Tài Cẩn - Tư tưởng, tác phẩm và kỷ niệm", Nhà xuất bản Đại học Quốc gia Hà Nội, Hà Nội, p.133-144.

Phạm Văn Hảo (chủ biên) (2009) *Từ điển phương ngữ tiếng Việt* [Vietnamese Dialect Dictionary], Nxb Khoa học Xã hội, Hà Nội.

Nguyễn Hữu Hoành, Bùi Thị Ngọc Anh, Hoàng Ngọc Tuệ (2022) Cuối language description and extensive glossary, *JSEALS Special Publication* No. 10, University of Hawai'i Press.

Huffman F.E. (1977) An examination of lexicon correspondences between Vietnamese and some other Austroasiatic languages, *Lingua*, Vol., 43, No, 2-3.

Nguyễn Văn Khang (chủ biên) (2002) *Từ điển Mường - Việt* [Muong - Viet Dictionary], Nxb Văn hoá dân tộc, Hà Nội.

Trinh Cam Lan (2022) *Ancient dialect of Area IV in two Nghe Tinh dialect dictionaries (Phonological characteristics of initial consonant system)*, 31st Annual Meeting, The Southeast Asian Linguistics Society, University of Hawai'i, May 18-20. 2022.

Trịnh Cẩm Lan (2023a) Nghiên cứu nguyên âm tiếng Sơn Tây (trường hợp các huyện Quốc Oai, Thạch Thất, Ba Vì) [Study of Son Tay vowels (The case of Quoc Oai, Thach That and Ba Vi districts)], *Tạp chí Ngôn ngữ và Đời sống*, số 2.

Trịnh Cẩm Lan (2023b) Tonal features of Hà Tây dialect (The case of Quốc Oai variant), *The Russian Journal of Vietnamese Studies*, DOI: 10.54631/VS.2023.71-321725, No. 1.2023, p. 95-106.

Lâm Bá Nam (1997) *Người Mường ở Ba Vì (Những ghi chép dân tộc học)* [Muong people in Ba Vi (Ethnographic records)]. Kỷ yếu Hội thảo Sơn Tinh và vùng văn hoá cổ Ba Vì, Sở Văn hoá thông tin tỉnh Hà Tây, Hà Tây.

Đỗ Đình Nghiêm, Ngô Vi Liễn, Phạm Văn Thư (1930) Địa dư các tỉnh Bắc kì, "Tỉnh Sơn Tây" [Geography of the Northern provinces, "Son Tay province"], Nhà in Lê Văn Tân.

Vũ Đức Nghiệu (2011) *Lược khảo lịch sử từ vựng tiếng Việt* [A brief history of Vietnamese vocabulary]. Nxb Giáo dục Việt Nam, Hà Nội.

Hoàng Phê (2015) *Từ điển tiếng Việt* [Vietnamese Dictionary]. Nxb Đà Nẵng, Trung tâm Từ điển học, Đà Nẵng.

Rhodes, A. de (1651, phiên bản 2021), *Từ điển Việt - Bồ - La và các cứ liệu liên quan* [Vietnamese - Portuguese - Latin dictionary and related documents]. Bản dịch Việt ngữ: Đỗ Quang Chính, Hoàng Xuân Việt, Thanh Lãng, Bùi Thị Minh Thuỷ, Nxb Tôn giáo, TP. HCM.

Nguyễn Văn Tài (1972) *So sánh hệ thống ngữ âm tiếng Mường một số vùng quanh Hòa Bình* [Compare the phonetic system of Muong language in some areas around Hoa Binh]. In trong *Tìm hiểu ngôn ngữ dân tộc thiểu số ở Việt Nam*, Hà Nội.

Nguyễn Văn Tài (2005) *Ngữ âm tiếng Mường qua các phương ngôn* [Phonetics of Muong through dialects]. Nxb Từ điển Bách khoa, Hà Nội.

Nguyễn Tài Thái (2015) *Ngữ âm tiếng Sơn Tây* [Son Tay Phonetics]. PhD thesis in linguistics, Vietnam Academy of Social Sciences.

Đặng Văn Tu, Nguyễn Tá Nhí (2011) *Địa chí Hà Tây* [Monography of Ha Tay]. Nxb Hà Nội, Hà Nội.

Trần Anh Tuấn (2019) *Giá trị văn hóa cộng đồng nông thôn vùng đồng bằng Bắc Bộ – Từ văn hóa phi vật thể đến không gian kiến trúc sinh hoạt* [Cultural values of rural communities in the Northern Delta - From intangible culture to architectural living space], TC Kiến trúc Việt Nam.

Trần Quốc Vượng (1997) *Tổng kết hội thảo về Sơn Tinh và vùng văn hoá cổ Ba Vì* [Summary of the workshop on Son Tinh and Ba Vi ancient culture area]*,* Kỷ yếu Hội thảo *Sơn Tinh và vùng văn hoá cổ Ba Vì*, Sở Văn hoá thông tin tỉnh Hà Tây, Hà Tây.