Studies in Asian Geolinguistics, Monograph Series No. 1

Papers from the Third International Conference on Asian Geolinguistics

ILCAA Joint Research Project 2015 - 2017 "Studies in Asian Geolinguistics"
Research Institute for Languages and Cultures of Asia and Africa
Tokyo University of Foreign Studies

Studies in Asian Geolinguistics, Monograph Series No. 1

Papers from the Third International Conference on Asian Geolinguistics

23–24 May 2016,
Royal University of Phnom Penh

(Organized by ILCAA Joint Research Project 2015–2017
"Studies in Asian Geolinguistics")

First published 2016

Edited by Mitsuaki Endo

ISBN 978-4-86337-238-2
(This is an electronic publication in PDF format)
Published by
Research Institute for
Languages and Cultures of
Asia and Africa (ILCAA)
Tokyo University of Foreign Studies
3-11-1, Asahi-cho, Fuchu-shi,
183-8534, Tokyo, JAPAN

Program

Date: 23 - 24 May 2016

Venue: Royal University of Phnom Penh

Russian Federation Boulevard, Toul Kork, Phnom Penh, Cambodia.

23rd May, 2016, Monday

9:00 - 9:10 Opening Ceremony

9:10 - 10:20 Session 1, Chair: Jean-Michel Filippi

10:35 - 12:20 Session 2, Chair: Multamia R.M.T. Lauder

Lunch

14:00 - 15:40 Session 3, Chair: Trần Trí Dõi and Trần Thị Hồng Hạnh

16:00 - 17:45 Session 4, Chair: Chitsuko Fukushima

19:00 - Dinner

24th May, 2016 Tuesday

9:00 - 10:45 Session 5, Chair: Masaaki Shimizu

11:00 - 12:10 Session 6, Chair: Rei Fukui

A longitudinal Geolinguistic Study of Language Change in West Java, Indonesia

Multamia R.M.T. Lauder, Allan F. Lauder

Department of Linguistics, Humanities Faculty, *Universitas Indonesia*

Abstract

This paper questions a number of fundamental hypotheses in historical linguistics and geolinguistics, namely that language change is constant, that it is gradual, occurring over very long time periods, and that the basic vocabulary of languages is especially impervious to change.

The position taken in traditional lexicostatistics and glottochronology is that the rate of change of meanings over time is relatively constant and can be represented with a simple formula where the variables are time, percentage of shared cognates and the retention rate per thousand years. A roughly constant retention rate of around 81% per thousand years for the Swadesh 200 basic word list has been assumed (Atkinson and Gray, 2005). However, work using computer simulation has suggested that language change may be faster in small communities (Nettle, 1999). Indonesia's many regional languages provide an opportunity to examine existing assumptions about the rate of change.

This paper reports on a longitudinal geolinguistic research study of the Bekasi region in West Java where three main languages are found: Javanese, Sundanese and Betawi Malay.

Data was gathered on two occasions separated by two decades. The first study was done in 1978, and the second in 1998. Both studies used exactly the same locations, informants, questionnaire, data gathering and analysis method. A questionnaire was used to elicit 200 basic words from the Swadesh list plus 50 phonological items. Data was gathered in 34 villages spread evenly across the area. Isogloss and dialectometry maps were produced. Comparison of the maps for the same lexical item, sampled across time, showed marked differences in distribution of language variation. There are obvious examples of disappearing lexical items and the appearance of new ones.

Taken together, the results provide new evidence on language change. The results show that language change occurs in periods as short as twenty years, and that the rate of change is not constant, with even basic words undergoing change. The implications of the findings are discussed from a number of perspectives.

Keywords

geolinguistics, comparative linguistics, rate of language change, language mapping, isoglosses, dialectometry, longitudinal research, Swadesh basic word list

1. Introduction

Languages are in a constant state of change (Aitchison, 2012). Various theories and concepts about language change have developed including the idea that contemporary languages have evolved from ancestral languages and that languages may be related into families (Hock and Joseph, 1996) (Campbell, 2013). The earliest work into this was the comparative-historical work into the Indo-European languages (Beekes and Vaan, 2011). This paper seeks to explain changes in language from the standpoint of dialectology based on data from regional languages in the province of Bekasi in Java, a linguistically heterogeneous region with three languages, Betawi, Javanese, and Sundanese.

2. Linguistic Situation

Throughout Indonesia, the presently best estimate for the number of languages is 719. This makes Indonesia the second most linguistically diverse nation in the world, after Papua New Guinea which has 851 languages (Lewis et al., 2015, Martí et al., 2005: 48).

The linguistic situation in Indonesia is by any standard complicated. However, at a policy level, three main categories of language are recognized, namely (1) Indonesian the national language, (2) local languages (*bahasa daerah*), and (3) foreign languages (Alwi and Sugono, 2000, Renandya, 2000: 115).

Indonesian is the national language, but census data reveals that it is not spoken as a first language by all Indonesian citizens. Although its use is spreading, there are still many people who communicate only in their local language. The number of people whose first language is a local language and for whom Indonesian is a second language is 140 million people, about 55.51% of the population.

However, Indonesian has been gaining ground over the last three decades and the number of first language speakers growing. This trend is being driven by three possible conditions:

1. The children of families whose parents are speakers of different regional languages tend to choose Indonesian as the language used in the home seeing it as a practical solution to not having to choose between one or both of the parents' languages;
2. The children in families living in large cities will choose Indonesian to communicate with neighbours and at school even when their parents speak the same regional language; and
3. The children in families living in large cities whose parents speak different regional languages will see the daily use of Indonesian at home as a rational choice.

Presently, the number of native speakers of Indonesian has reached 22.8 million people, or 9.04% of the entire population. The total number of speakers of Indonesian, both first and second language speakers, is therefore 162.8 million people (Lewis et al., 2015).

The regional languages can be described first in terms of their numbers of speakers. Of Indonesia's 719 languages, only fourteen have more than one million speakers. However, taken together, the total number of speakers of these large languages is 174,579,000 people which accounts for 69.22% of Indonesia's 252.2 million total population (Badan Pusat Statistik, 2015). The names and speaker numbers for these large languages is:

Javanese (84.3 million speakers), Sundanese (34 million speakers), Malay (13,040,000 speakers), Batak (7.045 million speakers), Madura (6.77 million speakers), Minangkabau (5.53 million speakers), Betawi (5 million speakers), Bugis (5 million speakers), Aceh (3.5 million speakers), Bali (3.33 million speakers), Makassar (2.13 million speakers), Sasak (2.1 million speakers), Lampung (1.834 million speakers), and Gorontalo (1 million speakers) (Lewis et al., 2015).

These fourteen large languages are all found either in the western part of Indonesia, in Java, Sumatra, and Madura, or in central Indonesia, in Bali, West Nusa Tenggara and Sulawesi. In the eastern part of Indonesia, there are no languages with speaker numbers above a million but the number of languages there is large.

There are also people who are monolingual and only speak their native regional language, numbering recorded 89.4 million people which is 35.44% of the population.

Of the 719 languages recorded in Indonesia, 13 have become extinct. The number of living languages is 706.

If we exclude the fourteen largest languages, this means that there are 692 other living languages with populations of sizes that vary from several hundred thousand down to just a handful of speakers. There are numerous very small languages: 386 languages with up to 5,000 speakers; 233 languages with at least 1,000 speakers; 169 languages with at least 500 speakers; and 52 languages with 100 speakers or less (Gordon, 2005). These very small languages are found primarily in the eastern part of Indonesia. Languages with such small populations are known to be very vulnerable to factors that threaten the existence of those languages.

The languages which have been classified as threatened with extinction has begun to attract the attention of linguists and others (Lauder, 2007a, Lauder, 2011a, Lauder, 2007b, 2011b, Lauder and

Sugono, 2011, Lauder and Lauder, 2012b, c) but we are still some way from widespread, coordinated efforts to revitalize these languages to the point of sustainability. Efforts are under way to provide a policy framework that is geared to provide legal certainty in attempts to ensure that the function and role of these languages is not marginalized (Lauder and Lauder, 2012a). Apart from the importance of these regional languages to the communities themselves, Indonesia is of great interest to scholars because it is home to hundreds of languages. These languages belong to two large language families and the processes of language change and language loss are problems that are now seen everywhere in the world.

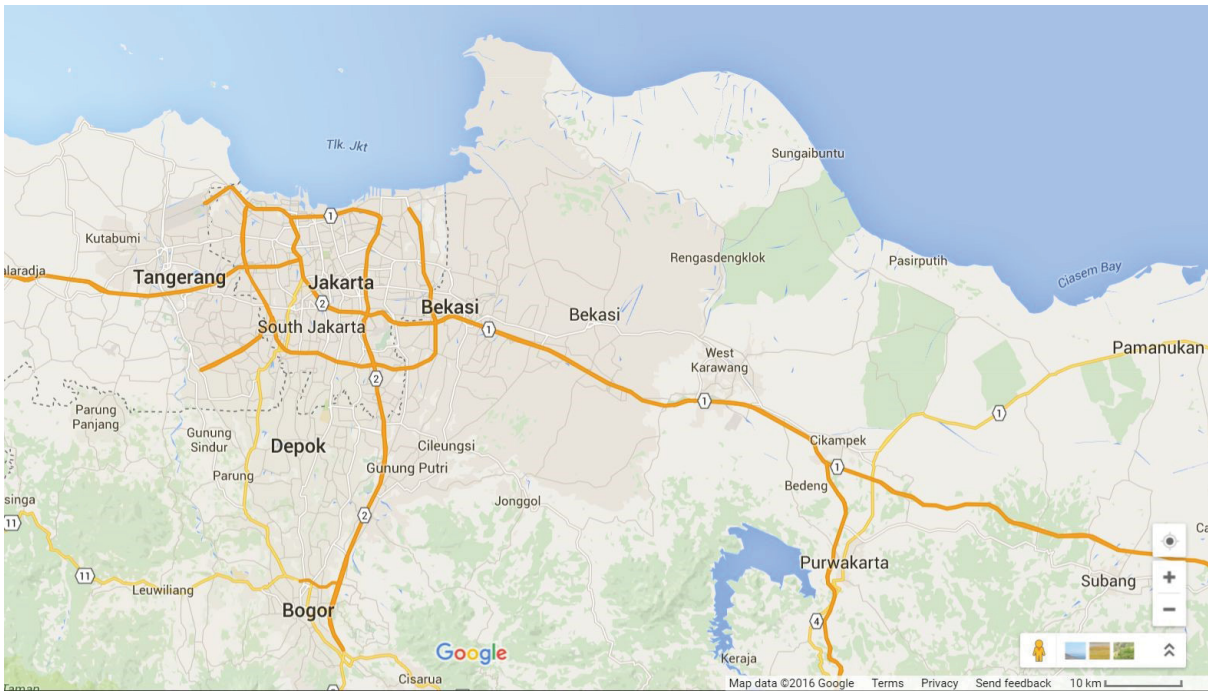
There are currently around 6,000 languages in the world which belong to seventeen major language families (Comrie et al., 2003, Crystal, 1997). These language families are: Indo-European, Uralic, Altaic, Chukotko-Kamchatkan, Caucasian, Afro-Asiatic, Nilo-Saharan, Niger-Congo, Khoisan, Eskimo-Aleut, Na-Dene, Amerind, Dravidian, Sino-Tibetan, Austric (Austronesian), Papuan (Non-Austronesian), and Australian Aboriginal. The indigenous regional languages in Indonesia belong to either the non-Austronesian or the Austronesian language families (Lewis et al., 2015).

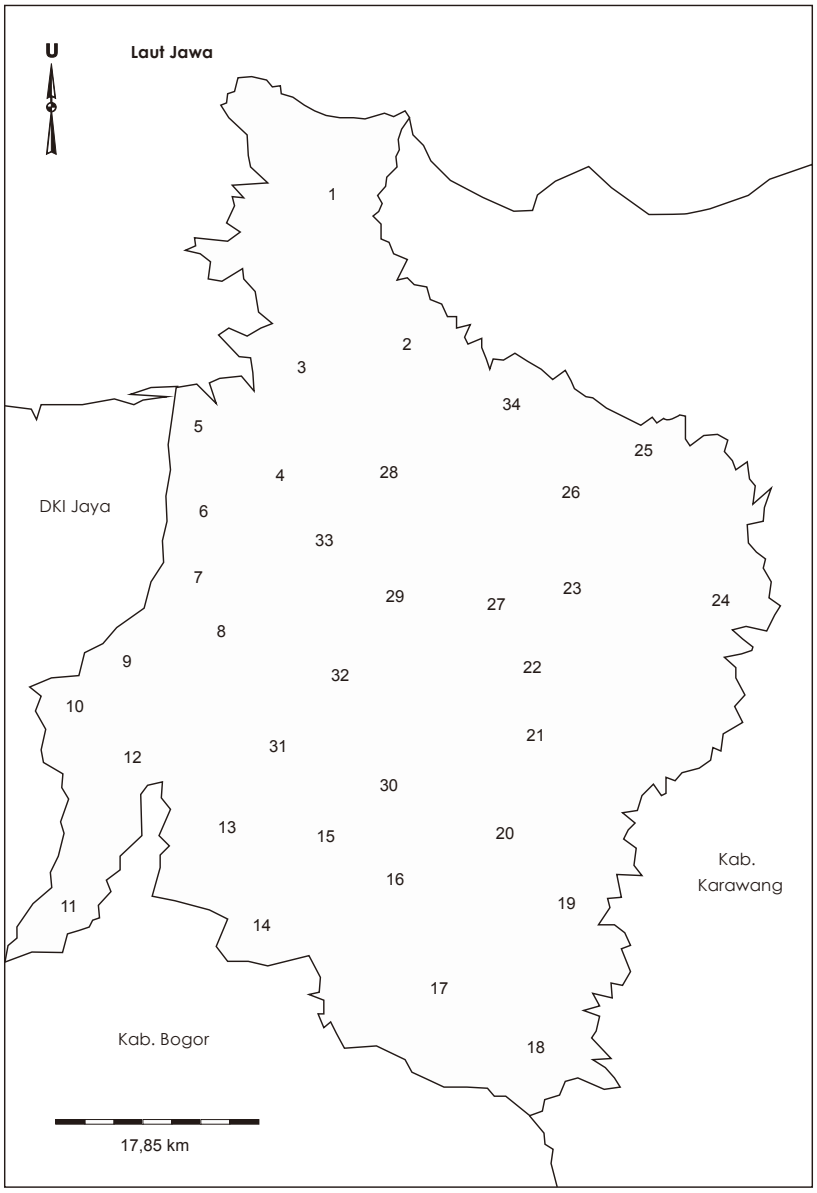
Of the 719 listed languages in Indonesia, 255 can be classified as non-Austronesian languages (Lewis et al., 2015). This is approximately one third of all the languages in Indonesia. They are found spread in the eastern part of Indonesia, in Maluku, East Nusa Tenggara, and Papua.

Comrie, et al. (2003) note that based on the latest research, the languages in Papua are not necessarily fully derived from Proto non-Austronesian. There is still the possibility that the languages actually are Austronesian, but come from different stock or phyla. The research methods for classification of languages in language families is still developing in detail and scope achievable. The placement of Papuan languages as belonging to the non-Austronesian family is still being discussed and debated. There is even the possibility that future research will show that the languages in Papua might belong to not one, but several language families (Comrie et al., 2003).

This brief sketch has provided an introduction to the overall language situation in Indonesia. We focus now on a study of language change in done in a linguistic area in the Bekasi Regency, in West Java province on Java island.

Bekasi





3. Language Change

Language tends to be in a state of constant change or fluctuation (Aitchison, 2012). Changes can be observed at all levels in a language. Many existing studies show such change occurring at the level of phonology or the lexicon (Chambers and Schilling-Estes, 2013). The present study focuses on changes in the lexicon as sampled in a geospatial context.

The total set of words in any language, its lexicon, can be seen as consisting of a small core of 'basic' words which are relatively resistant to language change, and the remainder of the lexicon which may shows words changing, while others may go out of use and new words appear. The field of comparative historical linguistics, based on over a century of scholarship on Indo-European and other languages, has asserted with some certainty that the rate of change in the general lexicon is relatively slow, with a significant change to the lexicon being seen in about a hundred years. But the rate of change for the basic lexicon is thought to be much slower, with such a change taking a millennium. The present study presents data that shows the pace of change may be much quicker than previously assumed.

Other than linguistic processes, extra-linguistic factors also determine the rate at which language changes. Geographical and topographical features such as mountain chains, rivers or stable flat ground, can enhance or impede the ease with which people can travel to get in contact with others in surrounding villages. The topography of Bekasi, the subject of this study, is a relatively small area, relatively flat and open, and as as such there is no impediment to speakers from the three language communities, Betawi, Sundanese, Javanese, being in contact with one another.

4. Data

Data was gathered in the Bekasi Regency which is located to the east of Jakarta. The first study was conducted by Tawangsih[1] (1978) and reported in Tawangsih (1987). Twenty years later the study was repeated by Adriani (1998). Adriani used exactly the same methodology as the earlier research, visiting the same villages, interviewing the same informants if still alive. If the informant had died, then a member of the family living in the same house was found as a replacement. Further, the research used the same questionnaire, the same interview techniques, and the same techniques for making the maps.

5. Discussion

Because every aspect of the research method was the same for both studies except for the time, any difference found by comparing maps shows a change in the linguistic feature shown. The use of the methods of dialectology provide convincing evidence of the details of language change. First, it can show changes in individual lexical items, both the basic vocabulary and cultural words, portrayed spatially in the map. Second, it can show the results for the entire set of words studied, amalgamated either as a whole or grouped into sets of words by semantic field. Third, it can express the extent or degree of change for pairs of words from the two time points quantitatively through the calculation of dialectometry and this can be used as empirical evidence for whether the distance between languages in contact is at the level of a dialect or of a language.

The presentation of the maps produced using dialectology consistently has the earlier study (Tawangsih, 1978) on the left and the later study (Andriani, 1998) on the right. The maps for lexical items shows the villages where data was collected, and a system of symbols to represent different phonological realizations of a lexical item. The pronunciations for lexical items are given in a table in each map. The number and types of pronunciations can be compared for each pair of maps in order to determine the extent of the change as well as its spatial distribution. In the case of the isogloss maps, the thicker the bundle of isoglosses, the greater the difference between one variety of language and its neighbour. For the dialectometry maps, the thickness of the lines separating areas indicates the degree of difference between the varieties spoken in two neighbouring areas.

5.1 Vocabulary Data

The idea of the basic vocabulary of a language is commonly based on high frequency lexical words. In the field of comparative historical linguistics, a basic vocabulary is that set of words, both grammatical and lexical, which is found in all the languages of the world. The basic vocabulary of any language is thought to have the following properties in relation to the question of language change: (1) the basic vocabulary is more resistant to change than the rest of the lexicon, (2) the retention of the basic vocabulary is constant over time, and (3) changes in the basic vocabulary in all languages will be the same.

The basic vocabulary list used by Tawangsih and Andriani in Bekasi consisted of 37 words, namely: child, dog, cloud, father, walk, star, moon, bird, blood, two, sit, four. teeth, nose, rain, mother, fish, this, wife, foot/leg, we, you, chief, we, sea, five, sun, they, mouth, hair, one, me, laugh, sleep, three, and bone.

[1] Tawangsih is the maiden name of the first author of this paper, Multamia RMT Lauder.

A comparison of the 37 basic words in the two studies shows that change occurred in all but three in the brief space of twenty years. This is striking result that contradicts a century of the assumption that the basic words have a high retention rate measured in millennia.

A number of words which are part of the basic vocabulary were found to have continued in use, but they showed a change in their geographic distribution. These words were dog, start, moon, bird, blood, two, four, rain, fish, foot/leg, one, three and bone. However, as many as 21 of the basic words did show a change of some kind, whether phonetic, the appearance of a new word, replacement by a new word, or even the disappearance from use of the word without a new one coined to replace it.

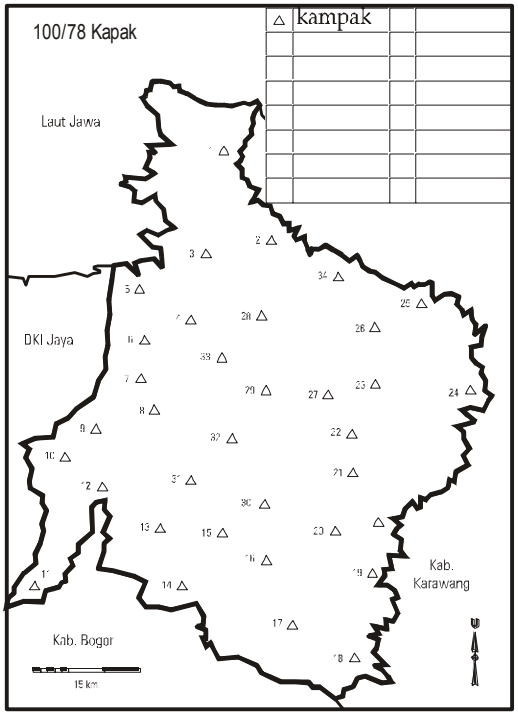
A total of 218 language maps were produced, one for each lexical item, with data collected first in 1978 and then in 1998. A comparison of the maps for any one item, between the 1978 and the 1998 study shows if any variation occurred.

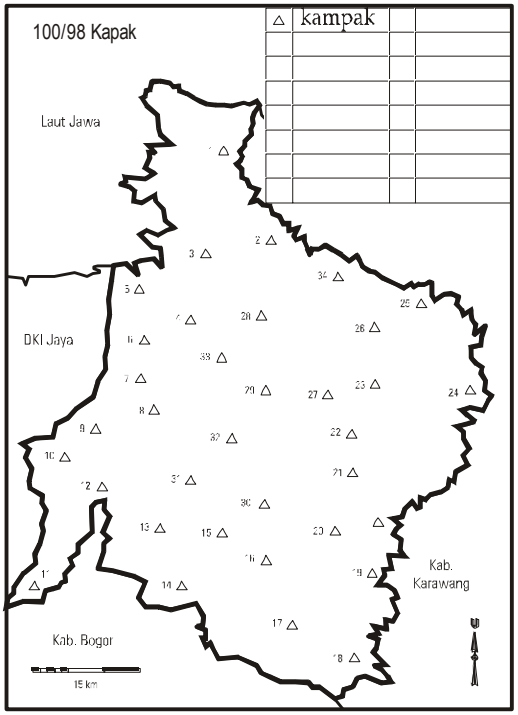
All the lexical variations, shown on 218 language maps, both from the 1978 and the 1998 studies, can be grouped as demonstrating six kinds of change. The type of change for each group, labelled A to F, is described below along with the percentage of lexical items for each group.

- Group A (4.59%): no change occurred;
- Group B (27.98%): no change was seen but some items were found to have moved to a new location (observation point);
- Group C (10.55%): a change in the sound of the word was found and some words were lost without the coining (addition) of a new word;
- Group D (17.89%): the variant of the word used remains the same, but there is the introduction of a new word;
- Group E (0.46%): all variants of the lexical item in use are changed with new word;
- Group F (38,53%): some of the variants of a word demonstrate no change, but some of the variants disappear while there is also the addition of new variants.

5.1.1 Vocabulary Group A Data

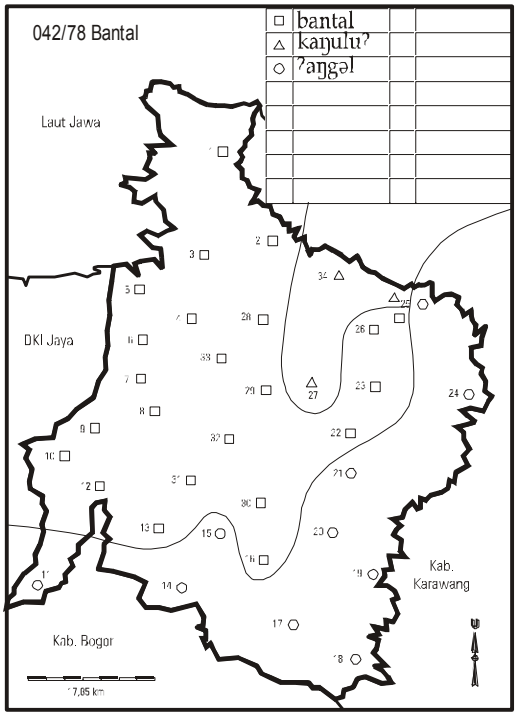
The vocabulary in group A represents 4.59% of all vocabulary and is the fifth largest group out of six. Group A consists of vocabulary where there was no change in the twenty years between 1978 and 1998. This is shown in the maps for KAPAK (axe) which are shown below.

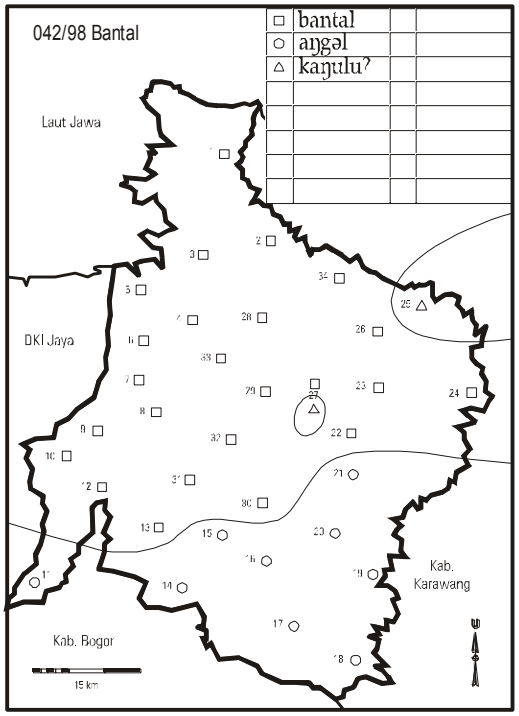




As can be seen in the maps above, there is only one variant for the word KAPAK (axe). The word is part of the lexicons in both the Betawi and Sundanese languages and it is also used by the community of Javanese speakers in Bekasi. No change, either in pronunciation or geographical distribution has occurred in the twenty years between the two studies.

5.1.2 Vocabulary Group B Data



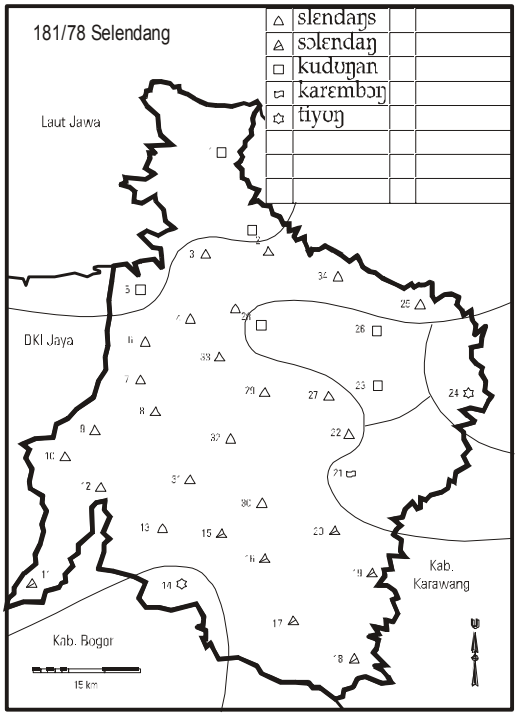


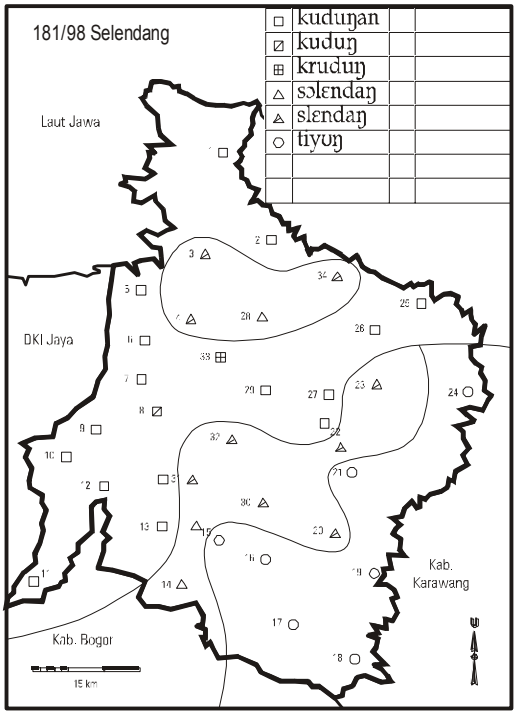
The words in group B make up 27.98% of all words and it is the second largest group. Group B includes words which have not changed but whose geographical distribution has changed. An example of this is shown in the two maps above for the word (042) BANTAL (pillow).

The maps above show that for the word *bantal* (pillow), there are three variants: *bantal* (1-10, 12-13, 22-23, 25-33), *kangulu* (25, 27), and *anggel* (11, 24-25). The numbers refer to the village numbers. The maps show that all three variants have persisted and there has been no change, either phonetic or lexical. However, there have been changes in the geographical distribution for all of them: *bantal* (1-10, 12-13, 22-34), *kangulu* (25, 27), and *anggel* (11, 14-21).

5.1.3 Vocabulary Group C Data

The words in group C represent 10.55% of all words making it the fourth largest group. Group B contains words where there have been changes in pronunciation and some variants have been lost without being replaced by new words. The maps below for word (181) SELENDANG (shawl/scarf) exemplify this.





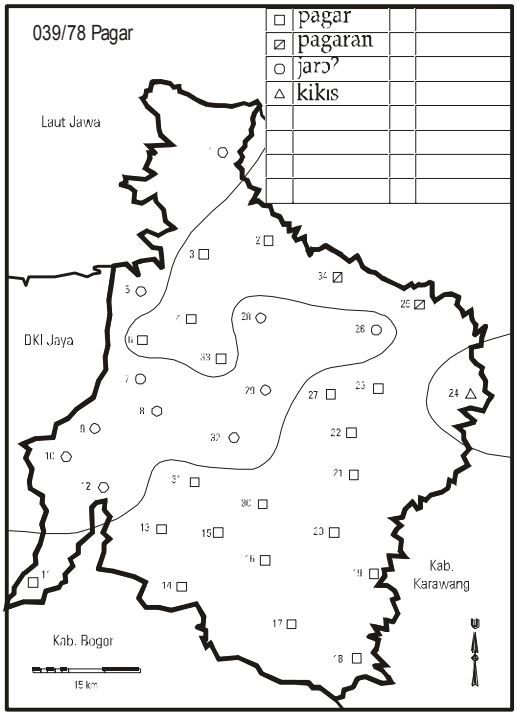
In the 1978 map, it can be seen that there were four etymons for *selendang* (shawl/scarf): *slendang* (2-4, 6-10, 12-13, 22, 25, 27-34), *solendang* (11, 15-20), *kudungan* (1, 2, 5, 23, 26, 28), *karembong* (21), and *tiyung* (14, 24). In the 1998 map, we see a number of changes. The etymon *karembong* has disappeared or is no longer used by the local population and has been replaced by the word *tiyung*. In addition there have been a number of changes in the distribution of words. *Slendang* is now found in (3-4, 20, 22-23, 30-32, 34), *solendang* (14-15), and *tiyung* (15-19, 21, 24). Meanwhile, the use of *kudungan* has become more widespread (1-2, 5-7, 9-13, 22, 25-27, 29, 31) while also experiencing a sound change becoming *kudung* (8) and *krudung* (33).

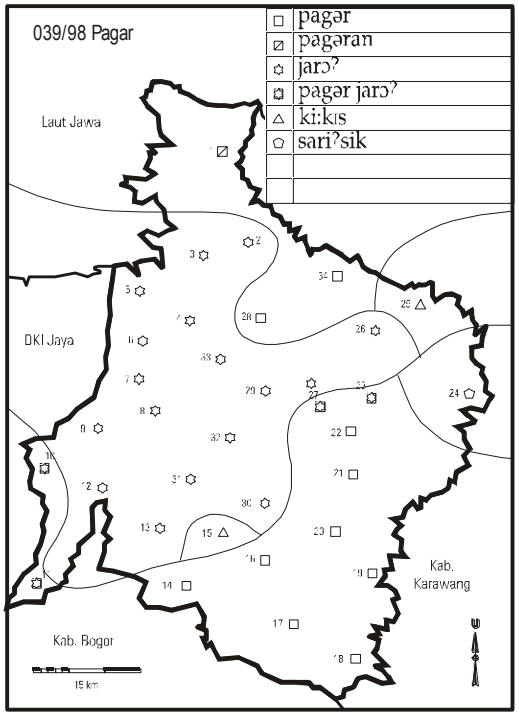
5.1.4 Vocabulary Group D Data

Group D contains 17.89% of all words, making it the third largest group. Group D consists of lexical items which have retained their earlier phonetic forms, but exhibit changes in geographical distribution and also the emergence of new word variants. Map (039) PAGAR (fence) provides an example of this.

On the 1978 map, the variants of *pagar* (fence) were: *pager* (2-4, 6, 11, 13-23, 27, 30, 31, 33); *pageran* (25, 34); *jaro* (1, 5, 7-10, 12, 26, 28-29, 32); and *kikis* (24). On the 1998 map, all four of these variants are still in use, but their distribution has changed: *pager* (14, 16-22, 28, 34); *pageran* (1); *jaro* (2-9, 12, 13, 26-27, 29-33); and *kikis* (15, 25). Also, two new word variants have appeared, namely *pager jaro* (10-11, 23, 27) and *sari'sik* (24).

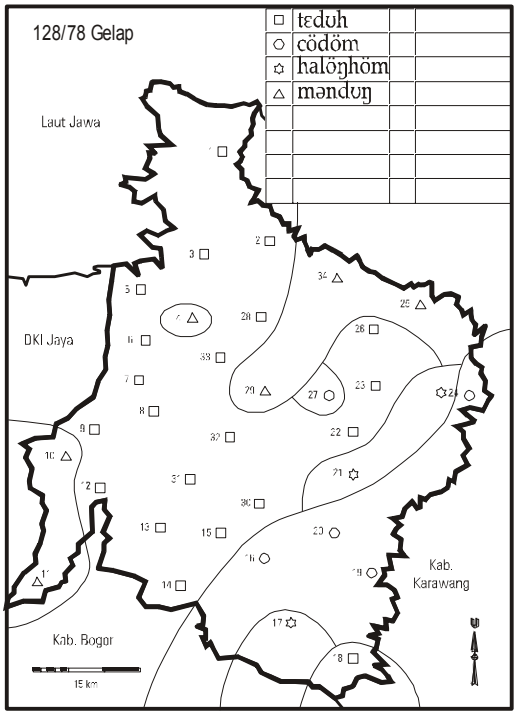
The data clearly show that there has been contact between the three language areas – Betawi, Sundanese and Javanese. Language contact has most likely been a cause leading to changes in the geographical distribution of language variants. The existence of *pager* and *pageran* in 1998 has been put under pressure by the introduction of *jaro* which extends towards the east. In addition, the appearance of the new variant *sari'sik* which in Sundanese appears as *sarigsig* (barrier, divider, bars).

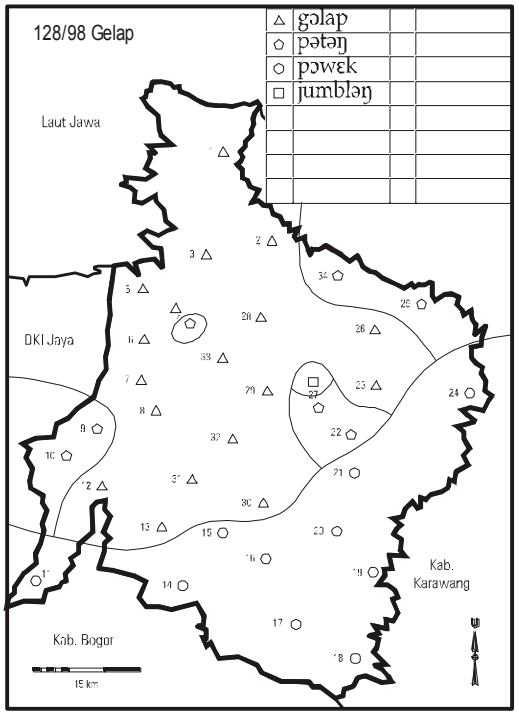




5.1.5 Vocabulary Group E Data

The words in group E make up only (0.46%) of all words making the group the smallest. This group consists of word variants which have all been replaced with new word variants. An example of this is shown in map (128) GELAP (dark).



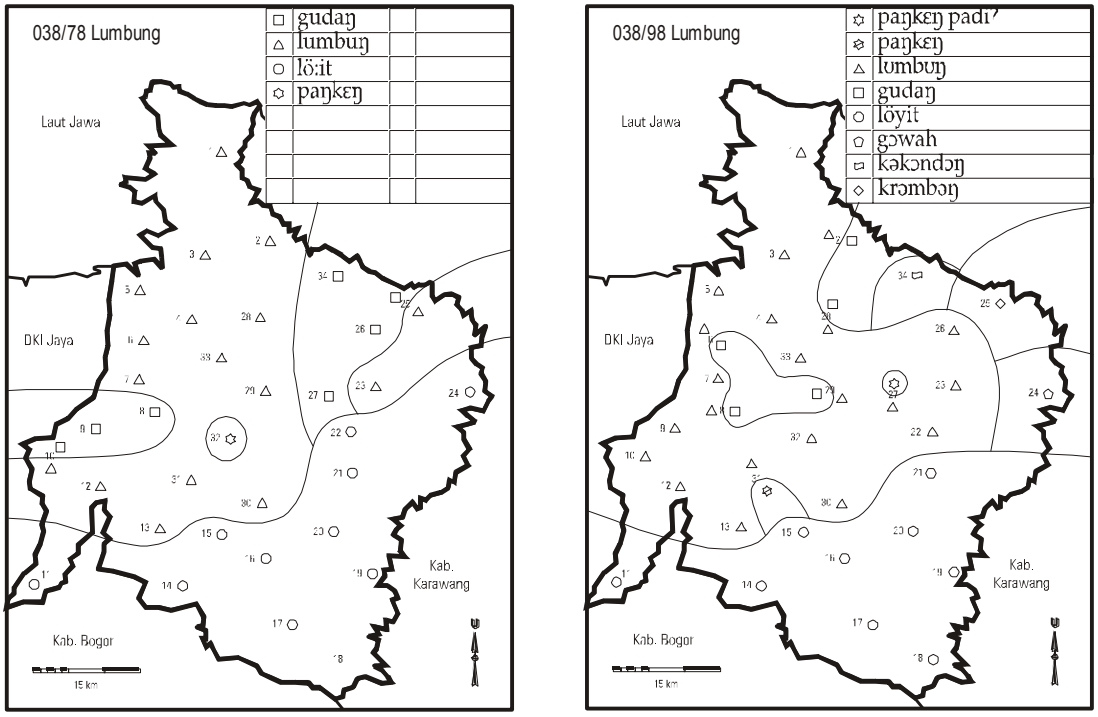


Both the 1978 and the 1998 map contain four etymons. The 1978 map has the word variants *teduh* (1-3, 5-9, 12-15, 22-23- 26, 28, 30-33), *ceudeum* (16, 19-20, 24, 27), *haleungheum* (17, 21, 24), and *mendung* (4, 10-11, 25, 29, 34). None of these are found in the 1998 map having all been lost or fallen

out of use by the local population. Four completely new word variants appear to replace them *gelap* (1-8, 12-13, 23, 26, 28-33), *peteng* (4, 9-10, 22, 25, 27, 34), *powek* (11, 14-21, 24), and *jumbleng* (27).

5.1.6 Vocabulary Group F Data

The vocabulary in group F make up 38.53% of all words, making this the largest of the groups. This makes it the dominant type of change occurring in Bekasi. The vocabulary variants in group F survive or persist with sound changes, but a number of variants also disappear, and there are also cases of new words coming into use. We also see a significant change in geographical variation. These processes are exemplified in the map (1038) LUMBUNG (barn, shed).



Four word forms are found in the 1978 map, namely *gudang* (8-10, 25-27, 34), *lumbung* (1-7, 10, 12-13, 22-23, 25, 28-31, 33), *leuit* (11, 14-21, 24), and *pangkeng* (32). Twenty years later, we see quite sweeping changes. Three of these forms have survived but they have undergone both sound changes and also changes in their distribution, as follows *gudang* (2, 6, 8, 28-29), *lumbung* (1-10, 12-13, 22-23, 26-33), and *pangkeng* (31). The variant *leuit* meanwhile is still in use by the local population but has undergone sound change and distributional change *leuyit* (11, 14-21). Apart from this, four new word variants have emerged, namely *pangkeng padi* (27), *gowah* (24), *kekondong* (34), and *krombong* (25).

5.2 Isogloss Data

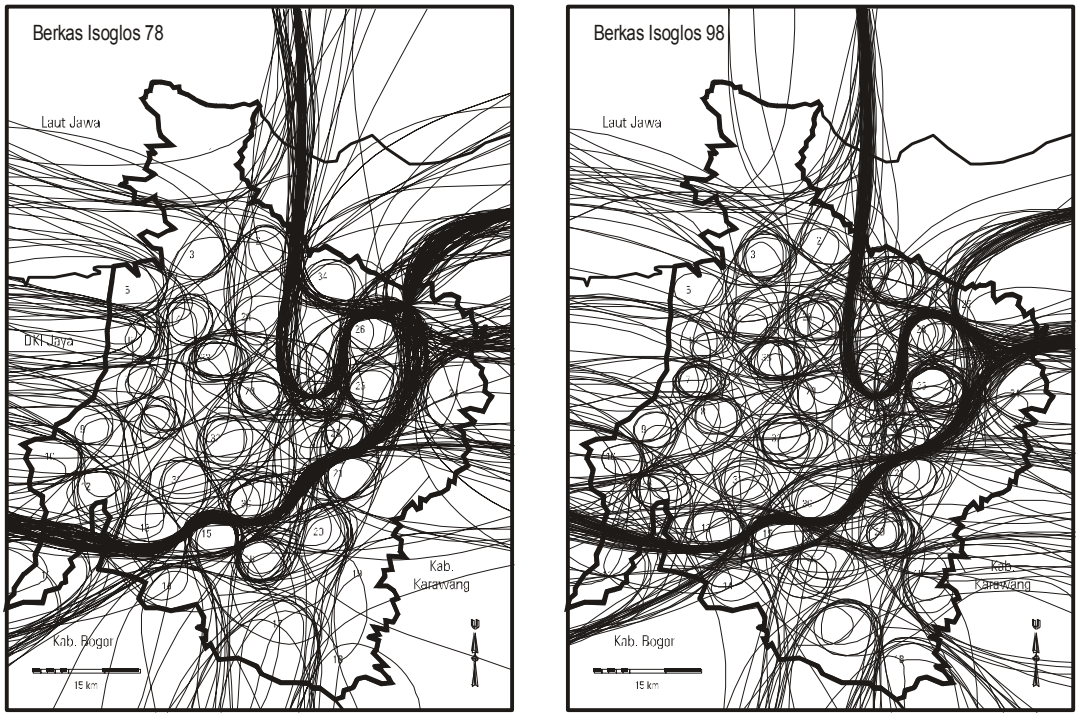
5.2.1 Combined Isogloss Data

An isogloss map is made for each of the 218 lexical items. These 218 individual maps can be combined by layering them on a single map. The maps below show this combined isogloss data, where the thicker the bundling of isoglosses, the more marked the difference between two language areas.

The isoglosses on the 218 language maps, taken together, show that the boundaries separating the Betawi, Sundanese and Javanese language areas still pass between the same data observation points. However, in the space of twenty years, the number of isoglosses demarking the limits of the Javanese language area, has reduced, showing on the maps as thinning of the bundle of isoglosses. This thinning has occurred because, during those twenty years, a process of linguistic accommodation has occurred

in the Javanese 'enclave' toward the Betawi and Sundanese speaking communities that surround them. An opposing trend has also occurred with the thickening of the bundle of isoglosses delimiting the border between the Sundanese and Javanese language areas.

When interpreting this data, it should be borne in mind that any change shown in an isogloss map is the change in the language of individuals based on the lexical, phonological and geographical features. It does not convey or imply anything about the attitudes of individuals or communities about these languages or their boundaries, and in particular should not be seen as a socio-political phenomenon wherein communities appeal to this data to insist on formally recognized shifts in the boundaries of their languages.

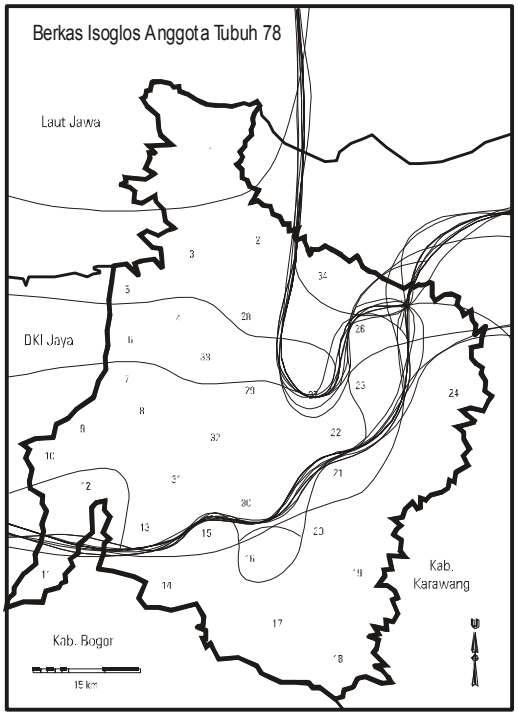


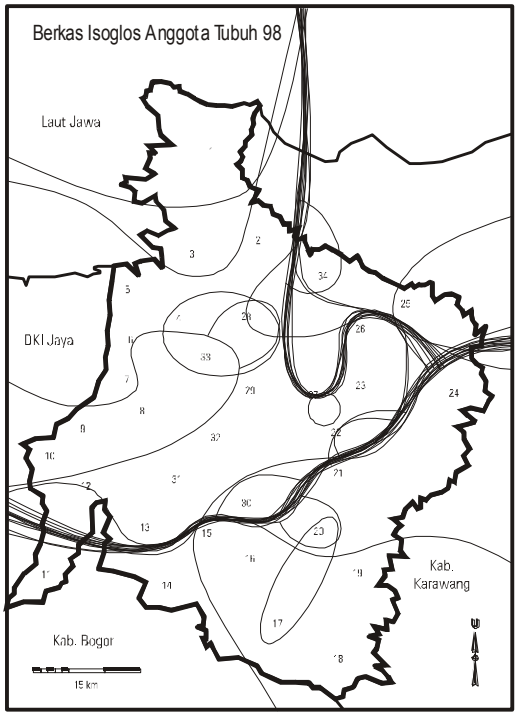
5.2.2 Isogloss Data by Semantic Field

Apart from broad generalizations about the isogloss data as a whole, it is also possible to distinguish more detailed patterns of distribution through looking at the sets of isoglosses by semantic field. An example of a specific pattern of isoglosses can be seen for example in the isogloss map for the semantic fields for parts of the body, kinship and nature. The isoglosses for the semantic field of parts of the body shows a situation where the patterns in the earlier and later studies are similar; in that for kinship, we see the patterns have somewhat changed; while for the semantic field of natural features we see the isoglosses have changed. These three semantic fields, by their nature, also contain a lot of basic words, so any change in these also has implications for the theory of slow and gradual change in the basic lexicon of languages.

The semantic field for PARTS OF THE BODY

The semantic field of parts of the body consists of the words: head, forehead, nose, mouth, lips, teeth, tongue, ear, neck, finger, thumb, index finger, little finger, leg, thigh, bone, hair, eyebrows, and blood.

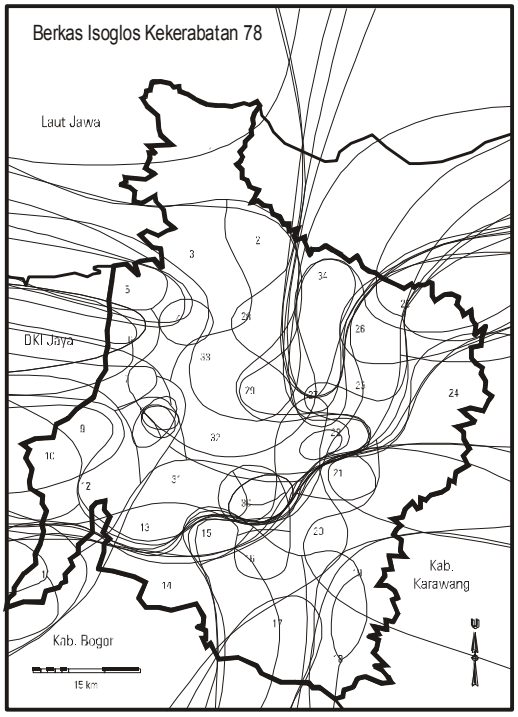


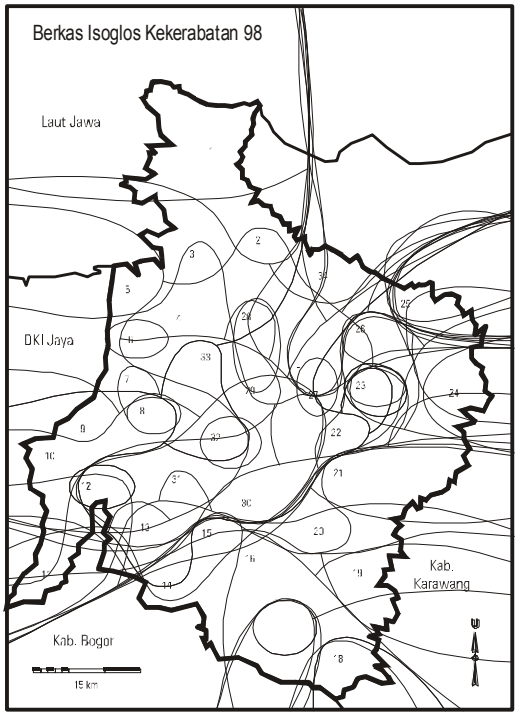


The bundles of isoglosses in the 1978 and 1998 maps resemble each other to a high degree with the isoglosses following the same path. A closer inspection also reveals that there are single isolated isoglosses that have shifted to a different path in the 1998 map. The groups of isoglosses (semantic fields) which exhibit very little change in this way make up 40% of all the groups.

The semantic field of KINSHIP

The vocabulary in the semantic field of kinship includes the following words: grandfather, grandmother, father, mother, older uncle, younger uncle, older aunt, younger aunt, older brother, older sister, younger brother, younger sister, child, older niece/nephew, younger niece/nephew, grandchild, husband, wife, in-laws, daughter/son in-law, and brother/sister-in-law.

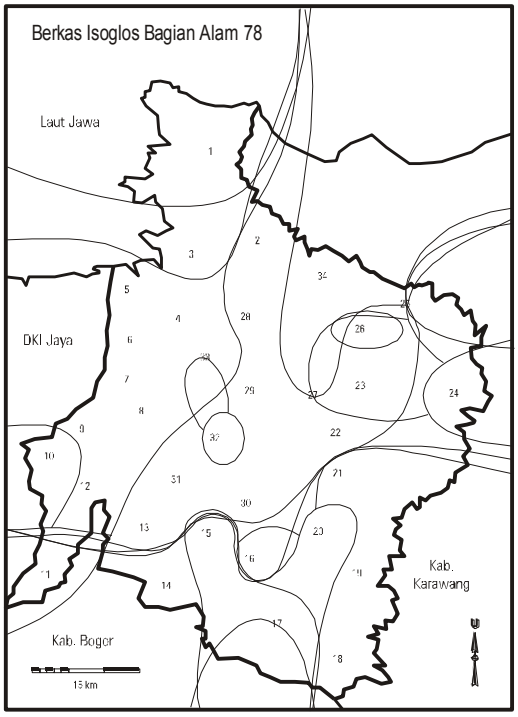


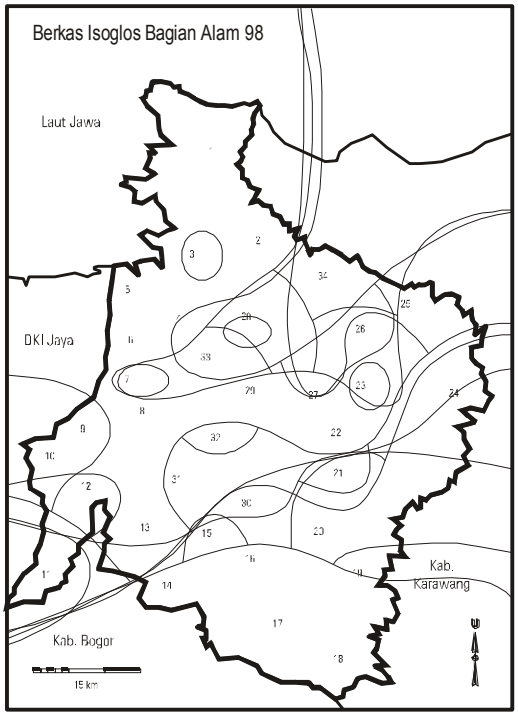


The isogloss maps for the semantic field of kinship bear little resemblance to each other, although some faint similarities still can be seen. For example, traces of the bundle of isoglosses in the south-center and the vertical bundle in the north of the 1978 map can still be seen in the 1998 map. But the relationship is not distinct. Thus, we find evidence that even though this semantic field contains a large number of basic words, the distribution of vocabulary variants is clearly in the process of change.

The semantic field of NATURAL FEATURES

The semantic field of natural features contains the following words: sun, solar eclipse, moon, star, cloud, rain, lightning, lake, river, and sea.





Basic words in the semantic field of natural features are: cloud, star, rain, sea and sun. A comparison of the pattern of isoglosses in the 1978 and 1998 maps immediately shows sweeping change. Change has not led to variations in words at the same observation point, but also led to significant changes in the geographical distribution of vocabulary.

On the basis of the bundles of isoglosses in the aggregated set of 218 isogloss maps and also the individual maps for different semantic fields, we can see that the "boundaries" between the Betawi, Sundanese and Javanese languages are largely located in the same places, although some of the bundles are thinner in the 1998 maps. However, if we focus on the isogloss maps for some of the semantic fields, we can see a tendency for notable degrees of change is also found.

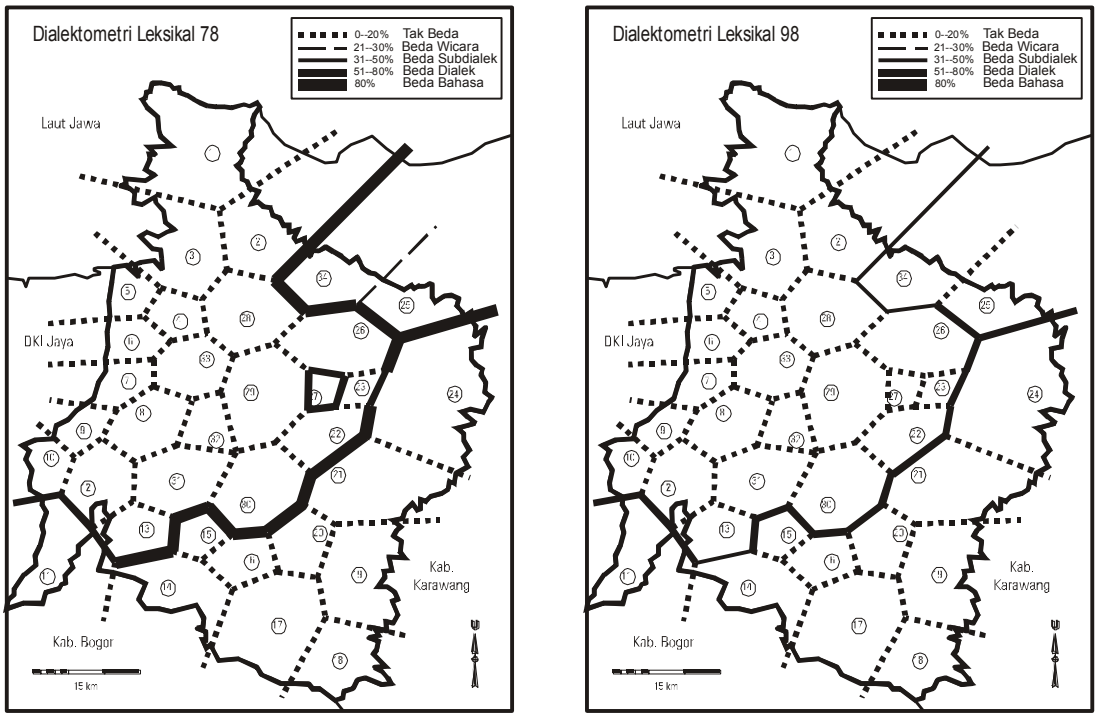
Most importantly, we see that the space of 20 years, although the basic vocabulary has had quite a high retention rate, a significant number are undergoing sound changes, and not a few are being replaced with new words. These findings on what is happening in Bekasi with regard to language change suggest that all three assumptions in comparative historical linguistics about the rate of language change, in particular to changes in the basic vocabulary, are not being supported here.

5.3 Dialectometry Data

While isogloss maps show us the geographical distribution of language variants, dialectometry gives us quantitative information about the degree of difference that has occurred between neighbouring observation points. The results of the dialectometry analysis are a way of quantifying the extent of the impact that the process of linguistic accommodation between speakers of Javanese in isolated areas and the speakers of Betawi and Sundanese in the areas surrounding them.

The degree of difference between contiguous language areas which is derived from dialectometry is expresses as a percentage. Séguy has related the degree of difference by dividing the scale into bands (Séguy, 1973). Séguy's band scale goes from no difference to different language: ≤ 20% no difference, different speech 21-30%, 31-50% different sub-dialect, 51-80% different dialect, and ≥ 81% different languages. The dialectometry analysis for the data from 1978 show that observations points 25, 27, and 34 showed differences in linguistic variation which was above 81% - a difference of language. This data, cross-checked with on the ground observation, showed clearly that Javanese was in use at these points. However, by 1998, between observation point 34 and the points around it where

Betawi was spoken, the degree of difference had dropped substantially to 31-50% a difference only of sub-dialect. Meanwhile, the difference at observation point 27, had dropped by 1998 to below 20% - no difference.



Based on the calculation of dialectometry, it can be proposed that the Javanese vocabulary used twenty years ago at observation points 27 and 34 have tended to change. These changes have occurred primarily as a result of adapting words from the Betawi vocabulary into the Bekasi variant of Javanese. In addition, the area along the border between the observation points for Betawi and Sundanese (10/11; 12/11; 13/14; 31/15; 30/16; 30/20; 22/21; 23/24; and 26/25) has showed a decrease in the percentage difference.

This change is facilitated in Bekasi by the absence of physical or social impediments to cross-cultural communication. The process of borrowing and assimilation in language contact which is known to occur in situations such as this is described by Dixon (1997: 15):

> If two languages are in contact - some of the speakers of each having a degree of competence in the other - they are likely to borrow lexemes, grammatical categories and techniques, and some grammatical forms (in at least one direction, often both directions) and gradually become more similar.

6. Historical Background

Language contact in Bekasi is not a new phenomenon but has been underway for a long time. Several centuries after the Tarumanagara Empire collapsed before the 7th century AD, the famous Pajajaran Kingdom was established under the leadership of King Siliwangi. There is archaeological evidence that prove that present day Bekasi was part of the Pajajaran kingdom. Importantly, there was a main road from the capital city to Karangsambung passing through Cileungsi and Cibarusa. Secondly, the remains of a ship have been found in the Rawa Tembaga area and the remains of a ships chains in the Kobak Rante area which prove that in the past Bekasi was one of the important port towns in the Pajajaran kingdom.

After the collapse of the Pajajaran kingdom due primarily to the spread of Islam, order began to break down in Bekasi. By the end of the 17th century, Bekasi had become part of the kingdom of

Mataram. Banten was strategic for Mataram because of its location. It had immediate borders to the south of the centres of power and also was next to the V.O.C. in the west. ??? At that time, Bekasi was a location where Mataram soldiers were trained for combat and was also important as a rice granary during the war against the V.O.C. in 1626.

Based on reports in Raffles' *History of Java* (Raffles, 1817, 1830), and the *Encyclopaedia van Nederlandsch Indie* (1895) (Lith et al., 1917), it is possible to trace and reconstruct the growth of the population in Bekasi.

When Mataram forces attacked the town of Batavia, (but suffered a crushing loss?) many Mataram soldiers did not return to the capital of Mataram. They chose to remain in the vicinity of Batavia, and some of them found refuge in Bekasi. In Bekasi, they lived in groups around the districts of Sutani, Cabangbungin, and Pabayuran. This was the origin of the enclave of a small Javanese community in Bekasi. They chose to settle on a small island in a lake, as it afforded a natural defensive position against any V.O.C. soldiers who might be sent after them. Taking these things into account, it is not surprising that the speech community was able to maintain its cohesiveness and identity over a long time.

The Betawi community in Bekasi forged a strong identity, and it has been supposed that their vocabulary would also tend toward homogeneity. One study investigated the composition of 200 basic words from the Betawi vocabulary. The study showed that 93% of these words were similar to words in Riau Malay and the remaining 7% were cognates of Javanese, Sundanese, and Balinese words. However, another broader study by Muhadjir contradicted this and suggested that the Betawi lexicon was much more heterogeneous. Muhadjir studied both the basic vocabulary and the cultural vocabulary. He found that words that originated in Malay accounted for only 35.21%. The rest were cognates of various other languages. 18.37% were cognates originating in borrowings from Javanese and 22.05% were cognates originating from Javanese and Sundanese.

7. Conclusion

The findings from the two Bekasi studies can throw new light on our understanding of language change most notably by challenging the long held assumption of slow and constant rate of change for the basic words of a language. Other researchers have questioned these too. Based on studies of aboriginal languages in Australia, Dixon (1997: 9) concluded that "The rate at the which a language changes is not constant and is not predictable." The present work provides evidence that that best describes what is happening in the recent past and the present.

The phenomenon of more rapid and irregularly paced change in language needs an explanation. In Bekasi, a number of extra-linguistic factors are likely to be involved. Bekasi is a socio-culturally heterogeneous area. It is located adjacent to the capital city Jakarta and it has experienced rapid economic growth. A number of government programs designed to stimulate development have been implemented there. These include drives to provide a number of services to all villages including electricity, news media, radio, television, telephones, and internet services. It is feasible to think that any of these could have contributed to the processes of language contact that underlie the rapid changes seen in the languages used there. Bekasi is multilingual and this forces the different speaker communities to become bilingual or multilingual. In the process, we see the features of accommodation as speakers in this multilingual situation are forced to make choices about what language or vocabulary to use. This is highly likely to continue with further diffusion due to geography and demographics.

References

Aitchison, Jean. 2012. *Language Change: Progress or Decay?* 4th edition. Cambridge Approaches to Linguistics. Cambridge: Cambridge University Press.

Alwi, Hasan, and Sugono, Dendy. 2000. Dari Politik Bahasa Nasional ke Kebijakan Bahasa Nasional (From National Language Politics to National Language Policy). In *Politik Bahasa : Risalah*

Seminar Politik Bahasa (Language Politics : Procedings of the Seminar on Language Politics). eds. Hasan Alwi and Dendy Sugono, v-xiv. Jakarta: Pusat Bahasa dan Departemen Pendidikan Nasional.

Andriani, Linda. 1998. Evaluasi Perubahan Leksikal: Studi Kasus Kabupaten Bekasi [An Evaluation of Lexical Change: A Case Study in the Bekasi Regency, Java], Thesis for Bachelors Degree. Department of Language and Literature, Universitas Indonesia.

Atkinson, Quentin D., and Gray, Russell D. 2005. Curious Parallels and Curious Connections-Phylogenetic Thinking in Biology and Historical Linguistics. *Systematic Biology*, 54(4), 513-526.

Badan Pusat Statistik. 2015. *Perkiraan Penduduk Beberapa Negara, 2000-2014*. Jakarta: Badan Pusat Statistik. Online Address: http://www.bps.go.id/linkTableDinamis/view/id/960. Accessed: 3 November 2015.

Beekes, R. S. P., and Vaan, Michiel Arnoud Cor de. 2011. *Comparative Indo-European Linguistics: An Introduction*. Second edition. Amsterdam & Philadelphia: John Benjamins.

Campbell, Lyle. 2013. *Historical Linguistics: An Introduction*. Third edition. Edinburgh: Edinburgh University Press.

Chambers, J. K., and Schilling-Estes, Natalie eds. 2013. *The Handbook of Language Variation and Change*. *Blackwell Handbooks in Linguistics*. Second Edition. Oxford: Wiley-Blackwell.

Comrie, Bernard, Matthews, Stephen, and Polinsky, Maria eds. 2003. *The Atlas of Languages: The Origin and Development of Languages throughout the World*. *Facts on File Library of Language and Literature*. 2nd edition. New York: Facts On File.

Crystal, David. 1997. *The Cambridge Encyclopedia of Language, 2nd Edition*. Cambridge: Cambridge University Press.

Dixon, Robert M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

Gordon, Raymond G., Jr. ed. 2005. *Ethnologue: Languages of the World*. Fifteenth edition. Dallas, Tex.: SIL International.

Hock, Hans H., and Joseph, Brian D. 1996. *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Vol. 93. Trends in linguistics: Studies and monographs. Berlin: Mouton de Gruyter.

Lauder, Allan F. 2007a. Indigenous Languages in Indonesia: Diversity and Endangerment. In *Procedings of Kongres Linguistik Nasional XII*. Surakarta, 3-6 September. Masyarakat Linguistik Indonesia & Universitas Sebelas Maret.

Lauder, Allan F., and Lauder, Multamia R.M.T. 2012a. A Review of Indonesian National Language Legislation. Paper read at *Ninth International Conference: Language, Development and Identity*. Bandung. 16-18 June, 2012. Institut Teknologi Bandung (ITB), British Council and University of Leeds.

Lauder, Multamia R.M.T. 2011a. Kepedulian terhadap Keberagaman melalui Revitalisasi Bahasa Minoritas. In *Ilmu Pengetahuan Budaya dan Tanggung Jawabnya: Analekta Pemikiran Guru Besar FIB UI*. ed. Riris K. Toha-Sarumpaet, 315-330. Jakarta: Penerbit Universitas Indonesia (UI Press).

Lauder, Multamia R.M.T. 2007b. Pemenuhan Hak Akses Informasi sebagai Salah Satu Upaya Memperlambat Laju Kepunahan Bahasa. In *Procedings of Kongres Linguistik Nasional XII*. Surakarta, 3-6 September. Masyarakat Linguistik Indonesia & Universitas Sebelas Maret.

Lauder, Multamia R.M.T. 2011b. Pengelolaan dan Pemberdayaan Bahasa-Bahasa yang Berpotensi Terancam Punah. Paper read at *Pengembangan dan Perlindungan Bahasa-Kebudayaan Etnik Minoritas untuk Penguatan Bangsa*. Jakarta. 15 Desember 2011. Puslit. Kemasyarakatan dan Kebudayaan (PMB), Lembaga Ilmu Pengetahuan Indonesia (LIPI).

Lauder, Multamia R.M.T., and Sugono, Dendy. 2011. Strategies for Language Policy and Planning in Asia and Europe. Paper read at *Steering Committee Meeting of the ASEM Language Diversity Forum*. Hotel Nusa Dua, Bali. 28-29 November 2011. Direktorat Kerjasama Intra Kawasan Amerika dan Eropa, Kementrian Luar Negeri (Indonesia).

Lauder, Multamia R.M.T., and Lauder, Allan F. 2012b. Penelisikan Variasi Bahasa dan Pengelolaan Bahasa di Indonesia [Investigating Language Variation and Language Management in Indonesia]. Paper read at *Seminar Sehari bersama Prof. Dr. Bernd Nothofer Mantan Guru Besar Tamu Program Studi Linguistik Program Pascasarjana UGM dan Kenangan Purnabakti Dr. Inyo Yos Fernandez*. Yogyakarta, 10 September 2012. Universitas Gajah Mada.

Lauder, Multamia R.M.T., and Lauder, Allan F. 2012c. The Importance of Maintaining Linguistic and Cultural Diversity in SEA. Paper read at *National Workshop on Traditional Knowledge and Cultural Diversity*. Dili, Timor Leste, 10th May 2012. UNESCO.

Lewis, M. Paul, Simons, Gary F., and Fennig, Charles D. eds. 2015. *Ethnologue: Languages of the World*. Eighteenth edition. Dallas, Texas: SIL International.

Lith, Pieter Antonie van der, Spaan, A. J., Fokkens, F., and Snelleman, Johannes François eds. 1917. *Encyclopædie van Nederlandsch-Indië*. Second edition. The Hague & Leiden: Martinus Nijhoff & E.J. Brill.

Martí, Fèlix, Ortega, Paul, Idiazabal, Itziar, Barreña, Andoni, Juaristi, Patxi, Junyent, Carme, Uranga, Belen, and Amorrortu, Estibaliz eds. 2005. *Words and Worlds : World Languages Review*. vol. 52. *Bilingual Education and Bilingualism*. Clevedon [England] ; Buffalo [N.Y.]: Multilingual Matters.

Nettle, Daniel. 1999. Using social impact theory to simulate language change. *Lingua*, 108(2-3), 95-117.

Raffles, Thomas Stamford. 1817. *The History of Java*. London: Black, Parbury, and Allen.

Raffles, Thomas Stamford. 1830. *The History of Java*. Second edition. London,: J. Murray.

Renandya, Willy A. 2000. Indonesia. In *Language Policies and Language Education: The Impact in East Asian Countries in the Next Decade*. eds. Wah Kam Ho and Ruth Y. L. Wong, 113-137. Singapore: Times Academic Press.

Séguy, Jean. 1973. La Dialectométrie dans l'Atlas Linguistique de la Gascogne. In *Revue de Linguistique Romane*, 37, 1-24. Paris: Société Linguistique Romane.

Tawangsih, Multamia R.M. 1987. *Bahasa-bahasa di Bekasi*. Jakarta: Panca Mitra.

Tawangsih, Multamia R.M. 1978. Bahasa di Daerah Bekasi: Sebuah Kajian Lokabasa [Languages in Bekasi: A Geolinguistic Study], Thesis for Bachelors Degree. Department of Language and Literature, Universitas Indonesia.

Correspondence between initial sounds in ancient Sino-Vietnamese words in Viet and Muong languages

Trần Trí Dõi

(Prof.Dr., Faculty of Linguistics, USSH, Vietnam National University).

Abstract: In their development history, the Viet and the Muong evolved into individual but relative languages by the 13th – 14th centuries after the period of *Viet-Muong Common*. Both had the same borrowings from Sino when they were still one as *Ancient/Archaic Viet-Muong*. For this reason, their phonetics as well as geographical differences among Muong dialects and Vietnamese provide good evidence for identifying *ancient Sino-Vietnamese* and *Vietnamized Sino-Vietnamese* word classes borrowed from Sino. This will be demonstrated through the correspondence between initial sounds in ancient Sino-Vietnamese words in Viet and Muong languages in this paper.

Key words: *ancient Sino-Vietnamese, Vietnamized Sino-Vietnamese, Vietnamese language, Muong language, geographical dialect*

1.**Ancient Sino-Vietnamese Words in Vietnamese and Muong Languages**

1.1. Separation of Vietnamese and Muong Languages in the Viet-Muong Group

Vietnamese (or Viet) and Muong are monosyllabic languages in the Viet-Muong group (Vietic) which belongs to the Mon-Khmer branch in the Austroasiatic family which is considered native to the cultural area of Southeast Asia. Throughout their development history, Viet and Muong had been identified as different parts (or dialects) of proto Viet-Muong/proto Vietic, or PVM for short), ancient/archaic Viet-Muong (AVM) and Viet-Muong Common (VMC). Only after the period of VMC did Viet and Muong start to separate and evolve in their own manner into independent individuals in the group (Trần Trí Dõi, 2011:98-174). This separation is believed to have occurred around the 13th-14th centuries.

The difference in the lexicons of Viet and Muong is most probably due to the differences in Sino loans. Phonetically, unlike Muong, all pre-nasal or glottalized initials in PVM became nasal sounds in Viet under the nasalization rule (Trần Trí Dõi, 2011: 186-197). Also, as the two initials in VMC voiceless plosive series became implosives (which are referred to as voiced sounds represented by *b* and *đ* at present), the alveolar-dental fricative **s* turned to *t* as a result of stopization (Trần Trí Dõi, 2015). On the other hand, in some cases, open syllables with two narrow vowels [*i] and [*u/*o] changed to half-open ones with a short middle, more open vowel (**i* > *ây, ay;* **u* > *âu, au/ao*) followed by a semi-vowel as the ending sound (Trần Trí Dõi, 2015a:85-90).

Thus, in the Viet-Muong group, Viet and Muong shared the same history of development from PVM (around 1000 B.C.E) until the 13th-14th centuries C.E., and historical changes in the two languages in this period should, in principle, be treated in the same way. Only when they had evolved into individuals (around the 13th-14th centuries) did they experience different historical changes. Therefore, we posit that

historical phonetics of Muong can provide good evidence for investigating Sino loans in the Viet-Muong group.

1.2. Ancient Sino-Vietnamese words in Viet and Muong

Throughout history, Vietnamese in particular and the whole Viet-Muong group in general did borrow words from Sino. However, researchers are short of convincing evidence to prove borrowings between the two languages in the PVM period, so borrowings that have been addressed are those of the AVM period. In his 1948 analysis of Sino loans in Vietnamese, Vương Lực divided them into three word classes of Sino origin, namely *Hán Việt ngữ*, *Hán Việt ngữ cổ* and *Hán Ngữ Việt hóa* (lit. *Sino-Vietnamese, Ancient Sino-Vietnamese* and *Vietnamized Sino,* respectively) (Vương Lực, 1982). Vương Lực's view to such identification was subsequently shared by various scholars in Vietnam and beyond, with slight changes in naming, e.g. *Hán Việt ngữ cổ (Ancient Sino-Vietnamese)* was called *Hán-Việt cổ (Archaic/Ancient Sino-Vietnamese)* (ASV), *Hán Việt ngữ* was called *Hán-Việt* (Sino-Vietnamese, SV) and *Hán ngữ Việt hóa* was called *Hán-Việt Việt hóa* (Vietnamized Sino-Vietnamese VSV).

Vietnamese words of Sino origin identified as ASV are those borrowed in the AVM period, e.g. *đầu* (頭 - head)*, buồng* (房 - room)*, muôn* (萬 – ten thousand/many), *mùa* (務 - season), *mùi* (味 - smell), etc., and they are currently used as pure Vietnamese words. In this historical period, as Viet and Muong were still one language, those borrowed into Viet probably entered Muong as well. However, the distinction between ASV and VSV words remains controversial. At present, the lists of these loans (first developed by Vương Lực) are cited by both Chinese and Vietnamese scholars, with great differences in numbers. For instance, Nguyễn Văn Khang presented a list of 108 ASV words, but only 96 of them were noted as cited from Vương Lực (Nguyễn Văn Khang, 2007:245-256). Meanwhile, Vũ Đức Nghiệu referred to Nguyễn Văn Khang's list of 108 words. He also made a new list of 288 instances[1] based on Lý Lạc Nghị, and mentioned Hà Tuệ Thành's 196 instances (Vũ Đức Nghiệu, 2011:154-159).

For the sake of this paper, we only limit ourselves to investigating the correspondence of initials in ASV using Nguyễn Văn Khang's list cited from Vương Lực, which is currently most widely accepted. Other lists are still under much debate, as a number of words are variably regarded as ASV by some but as VSV by others, or vice versa. That again explains why we rely on Nguyễn Văn Khang's list of ASV cited from Vương Lực.

2.Data and Geography concerning Correspondence of Initials

2.1.Geography of Viet and Muong

When providing explanation for the development history of the Viet-Muong group in general, and of Viet and Muong in particular, we pointed out that the difference in their geographical residence is one of the causes to their historical difference. We stated, "the reason why Viet entered a new period of development was that speakers of

[1] When enumerating Ancient Sino-Vietnamese words from Lý Lạc Nghị's list, Vũ Đức Nghiệu stated that Lý Lạc Nghị "did not make a separate list of ancient Sino-Vietnamese words". The list presented in Vũ Đức Nghiệu's book was based on "his notation and explanations" in the Vietnamese translation of "Tracing the origin of Sino" (Vũ Đức Nghiệu, 2011: 156).

VMC in the delta and in the mountains experienced various degrees of social impacts" (Trần Trí Dõi, 2011: 155). Such impacts were linguistically attributed to the "Sinicization" process at differing extents. Observing current settlements of Muong and Viet speakers, we can see the difference in the geographical distribution of the two communities (see Annex: Map 1), which resulted in this historical impact. While Viet speakers mostly dwelled along the coastal delta, Muong speakers took the geographical intersections between the plain and the high mountains in the west as their home. Thanks to more convenient transport links in the coastal delta, the indigenous Viet people enjoyed more contact with the Chinese. Meanwhile, the Muong definitely suffered from transport difficulties, and consequently much less contact with the Chinese than the Viet. We posit that such geographical differences concerning their residence served as key drives to the divergence of VMC into independent individuals around 700 years ago, resulting in linguistic differences described earlier in section 1.1.

Geographical differences not only existed between Viet and Muong, but among various dialects of Muong as well. In a recent study, J.D. Phan once again[2] pointed out that even dialects of Muong did differ greatly in a variety of ways across geographical areas (John D. Phan, 2012:1-18). In his account, the southern dialects of Muong extended northward as far as the Muot village in Cam Thuy, Thanh Hoa while the central dialects of Muong were represented by Na Bai (Cao Phong, Hoa Binh), and the northern dialects of Muong by Choi hamlet (Tan Son, Phu Tho). Nguyễn Văn Tài's phonetic analysis of Muong dialects also showed the same differences among those three regions of Muong dialects (Nguyễn Văn Tài, 2005). Given the current research situation, when dealing with geographical differences among Muong dialects, we can make temporary use of the same southern, central and northern distinction across the three provinces of Thanh Hoa, Hoa Binh and Phu Tho as those used by John D. Phan.

2.2.Correspondence of initials in Viet and Muong among Ancient Sino-Vietnamese words

According to Nguyễn Văn Khang, the list of Ancient Sino-Vietnamese words in Vietnamese enumerated by Vương Lực contains 96 instances. Among these, considering the place and manner of articulation of initials in syllables, we find rich correspondence between Viet[3] and Muong. In terms of the place of articulation, initials in ASV include bilabial, dental, alveolar, palatal, velar and glottal consonants; in terms of the manner of articulation, they include plosive, fricative and nasal consonants.

However, for the sake of investigation in this paper, we only pay attention to the correspondence between *nasal* and *lateral* [m, n, ɲ, ŋ, l] among ASV in Viet and Muong as some examples. We choose nasals and laterals because in Vietnamese phonetic history, these initials are proved to have been the best preserved since PVM[4]. This means that in principle, in their development history, these initials remain almost unchanged. Therefore, given geographical differences between Viet and Muong speakers, correspondence of

[2] In his classification of Muong dialects to find data for his study of the Vietnamese language, H. Maspero used geographical distribution as one of the criteria (H. Maspero, 1912: 5).

[3] We hereby make temporary use of *Sino-Vietnamese pronunciation* and *Ancient Sino-Vietnamese* as transcribed by Nguyễn Văn Khang. In fact, some pronunciation of these words of Sino origin may need reconsidering.

[4] See Nguyễn Tài Cẩn (1995), Trần Trí Dõi (2009, 2011, 2011a)

initials among ASV and Muong dialects may reveal interesting facts concerning ASV in language history. At present, the way the Muong people use these words of Sino origin is hardly different from that of the Viet. Although not all ASV words in current use in Vietnamese may find their corresponding forms in Muong, the words identified may generate some observations of possible borrowings.

More specifically, of the 1,000 words in 30 dialects classified by Nguyễn Văn Tài[5] (NVT), among the *nasal* and *lateral* initials, only 3 out of 23 are bilabial and velar, namely *mả, múa, ngẩng (tomb, dance, raise* (one's head), respectively*)*. Muong data analyzed by Nguyễn Văn Tài are given in the first column of the following table. Meanwhile, in Muong "Dictionary" by Nguyễn Văn Khang[6] (NVK, chief editor), the number of correspondence, though larger, stands at 15 instances, namely 4 instances of [l] out of 6 *(lầu, lìa, liềm, lồng – floor, separate, sickle, cage)*, 7/7 instances of [m] *(mèo*, *mạng*, lông *mày, mả, mùi, múa,* sương *mù – cat, web, eyebrow, tomb, smell, dance, fog*), one instance of [n] out of 2 *(nộp – pay)*, 3/7 instances of [ŋ] *(ngửa*, *ngà* voi, *ngan – supine, ivory, white duck*). In "Muong Corpus"[7] published in 1987 with 1909 entries, we find only 6 out of 23 instances of correspondence, including 1 out of 6 of [l] *(lừa - donkey*), 3 out of 7 of [m] *(mèo*, lông *mày, múa – cat, eyebrow, dance)*, and 2 out of 7 of [ŋ] *(ngửa*, *ngăn – supine, prevent/resist)*. As both of the "Dictionary" and "Muong Corpus" are based on the Muong dialect in Tan Lac (Muong Bi) and Lac Son (Muong Vang) in Hoa Binh province, they represent the central Muong. Putting them all together, we get the following correspondence.

Material of Muong language (IPA)

Nº	SV		ASV	**NVT**	**NVK**	**VX**
1	樓	lâu	lầu (floor)	-	[lɤ̆w^{2}]	-
2	離	li	lìa (depart)	-	[lie^{2}]	-
3	鐮	liêm	liềm (sickle)	-	[liem^{2}]	-
4	連	liên	liền* (immediately, at once)	-	-	-
5	籠	lung	lồng (cage)	-	[loŋ^{2}]	-
6	驢	lư	lừa (donkey)	-	-	[lɯə^{1}]
7	猫	miêu	mèo (cat)	-	[mɛ̀w^{4}]	[mɛw^{1}]
8	命	mệnh	mạng (lives/fate)	-	[maŋ^{4}]	-
9	眉	mi	(lông) mày (eye brow)	-	[măj^{2}]	-
10	味	vị	mùi* (smell)	-	[muj^{4}]	[muj^{1}]
11	舞	vũ	múa (dance)	[muə^{3}]	[muə^{5}]	[muə^{3}]
12	霧	vụ	(sương) mù (fog/mist)	-	[mu^{2}]	-
13	墓	mộ	mả (grave)	[ma^{4}]	[ma^{3}]	-
14	南	nam	nôm	-	-	-
15	納	nạp	nộp (to hand in)	-	[nop^{1}]	-
16	辱	nhục	nhuốc (sullying)	-	-	-
17	疑	nghi	ngờ/ngỡ* (doubt/think)	-	-	-
18	逆	nghịch	ngược (against, opposite)		-	-
19	瓦	ngõa	Ngói (tile roof)	-	-	-

[5] See Nguyễn Văn Tài (2005: 233-257)
[6] See Nguyễn Văn Khang (2002). We use IPA to transcribe these initials.
[7] Published by the Soviet Union's Academy of Science (AN.SSSP, 1987).

20	御	ngự	ngựa (ngăn) horse (to prevent)	-	[ŋăn^{1}]	[ŋan^{2}]
21	仰	ngưỡng	ngửa/ngẩng (turn upward)	[ŋɔk^{3}]	[ŋa^{4}]	[ŋak^{3}]
22	牙	nha	ngà (ivory)	-	[ŋa^{2}]	-
23	雁	nhạn	ngan (musk duck)	-	-	-

Note:

- NVT is material of Muong language collected in Huy Thuong – a commune in Phu Yen district, Son La province, Vietnam. (21^{o}16'41''B, 104^{o}41'34''Đ)

- NVK is material of Muong language collected in Phong Phu commune which is in Muong Bi area of Tan Lac district, Hoa Binh Province, Vietnam. (20^{o}37'51''B, 105^{o}12'43''Đ)

- VX is material of Muong language collected in An Nghia commune which is in Muong Vang area of Lac Son district, Hoa Binh province, Vietnam (20^{o}24'4''B, 105^{o}30'18''Đ)

1. Discussion

With the data at hand, we can see *initials* in ASV in Viet and Muong are treated almost the same. Such similarity is found not only between Viet and Muong, but among various Muong dialects as well. This means Sino loans which are referred to as ASV have been preserved in Muong in the same way as in Viet. This also means those words of Sino origin were borrowed into Viet in AVM period.

Another issue may arise through the investigation of ASV in Viet and Muong. That is, most probably, thanks to the identification of correspondence of Sino words classified as ASV and VSV in Viet and Muong, we may acquire more clues to re-assert the two word classes in Vietnamese, i.e. thanks to Muong data, we can deal with the controversy over some words of Sino origin which are either believed to be ASV or VSV in different views.

REFERENCES

- Nguyễn Tài Cẩn (1995), *Giáo trình lịch sử ngữ âm tiếng Việt (Sơ thảo*), Nxb Giáo dục.

- Trần Trí Dõi (2009), *Về nguồn gốc lịch sử của dãy âm đầu mũi [m, n, ɲ, ŋ] hiện nay trong tiếng Việt,* Ngôn ngữ, No 11(246)/11-2009, tr 1 - 11.

- Trần Trí Dõi (2011), *Giáo trình lịch sử tiếng Việt*, Nxb Giáo dục Việt Nam.

- Trần Trí Dõi (2011a), *Một vài vấn đề nghiên cứu so sánh lịch sử nhóm ngôn ngữ Việt - Mường,* Nxb Đại học Quốc gia Hà Nội.

- Trần Trí Dõi (2015), *Hiện tượng tắc hóa (stopization) âm đầu trong lịch sử tiếng Việt*, Hội thảo Quốc tế"Ngôn ngữ học Việt Nam 30 năm đổi mới và phát triển", *Viện Ngôn ngữ học,* Hà Nội ngày 23 tháng 08 năm 2015, 09 tr A4

- Trần Trí Dõi (2015a), *Góp thêm ý kiến về sự tương ứng giữa vần *[U], *[O] với [ÂW/ĂW] và [AW] trong ngữ âm lịch sử tiếng Việt*, Tạp chí *Khoa học Xã hội và Nhân văn* (Vietnam Social Sciences and Humanities), số 1 (10/2015), tr 85-90.

- H. Maspéro (1912), *Etude sur la phonétique historique de la langue annamite: Les initiales*, BEFEO, XII, n^{o}1, pp 1-127.

- Nguyễn Văn Khang *chủ biên* (2002), *Từ điển Mường - Việt*, Nxb Văn hoá dân tộc.

- Nguyễn Văn Khang (2007), *Từ ngoại lai trong tiếng Việt*, Nxb Giáo dục.

- Vương Lực (1982), *Long trùng tịnh điêu trai văn tập*. Trung Hoa thư cục, 1982 (tiếng Trung)

- Vũ Đức Nghiệu (2011), *Lược khảo lịch sử từ vựng tiếng Việt*, Nxb Giáo dục Việt Nam.

- John D. Phan (2012), *Mường is not a subgroup: Phonological evidence for aparaphyletic taxon in the Viet-Muong sub-family*, Mon – Khmer Studies 40: 1-18

-A.N.SSSP, (1987), *Jazuk Mương*, "Nauka", Moskva.

- Nguyễn Văn Tài (2005), *Ngữ âm tiếng Mường qua các phương ngôn,* Nxb Từ điển bách khoa.

Annex I* (Map I): *Geographical differences between Viet and Muong at present

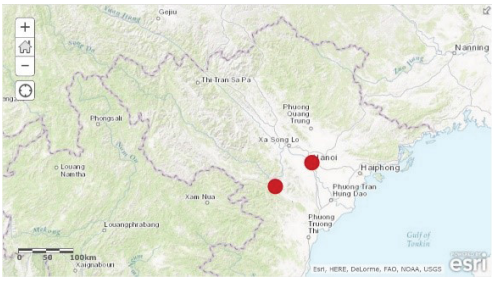
Annex II* (Map II): *Geographical distribution of southern, central and northern dialects of Muong

***Annex III*: List of Ancient Sino-Vietnamese ASV** in Muong language

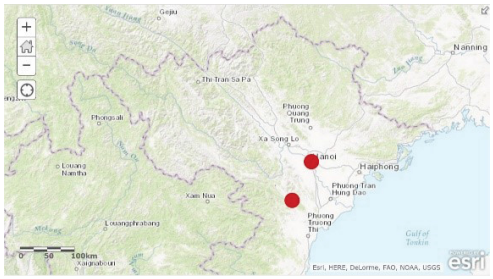
- NVT is material of Muong language collected in Huy Thuong – a commune in Phu Yen district, Son La province, Vietnam. (21°16’41’’B, 104°41’34’’Đ)

- NVK is material of Muong language collected in Phong Phu commune which is in Muong Bi area of Tan Lac district, Hoa Binh Province, Vietnam. (20°37’51’’B, 105°12’43’’Đ)

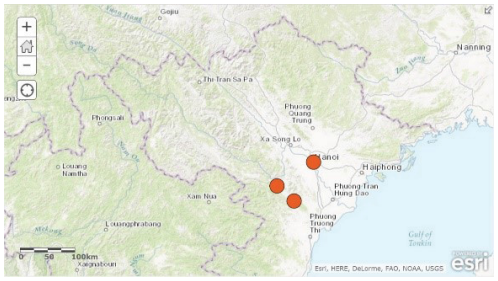
- VX is material of Muong language collected in An Nghia commune which is in Muong Vang area of Lac Son district, Hoa Binh province, Vietnam (20°24’4’’B, 105°30’18’’Đ)



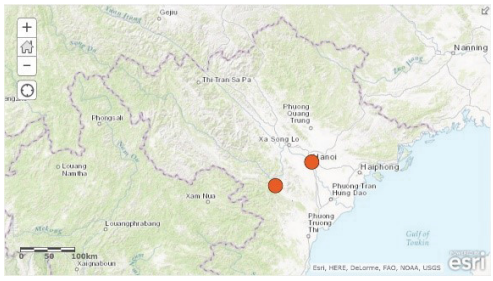
“樓, 離, 鐮, 籠” (●: l)



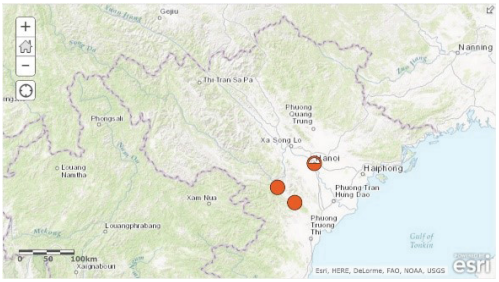
“驢” (●: l)



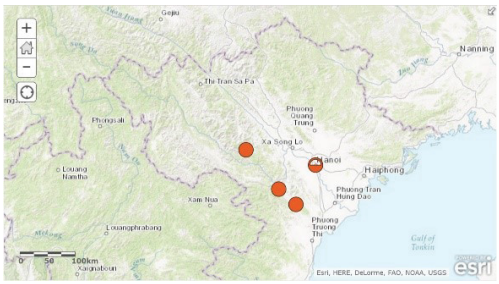
“猫” (●: m)



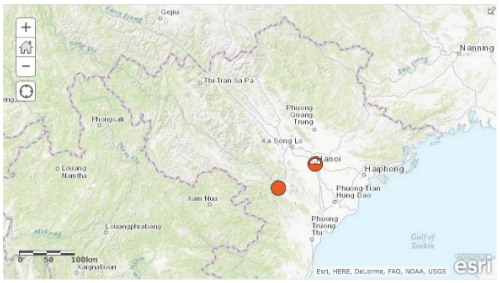
“命, 眉” (●: m)



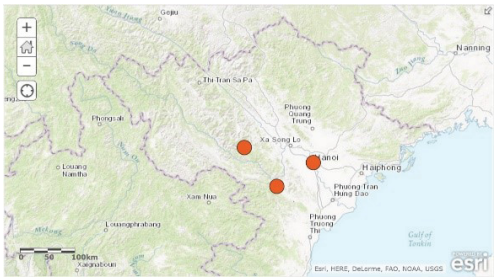
“味” (●: m, ◒: m>v)



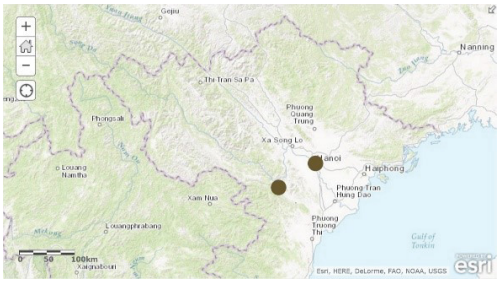
“舞” (●: m, ◒: m>v)



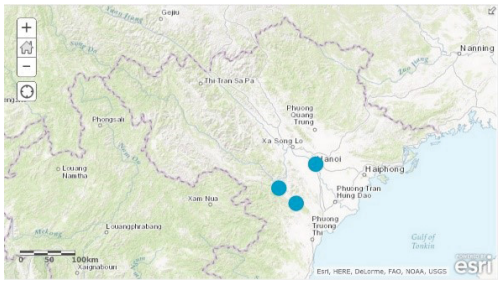
“霧” (●: m, ◒: m>v)



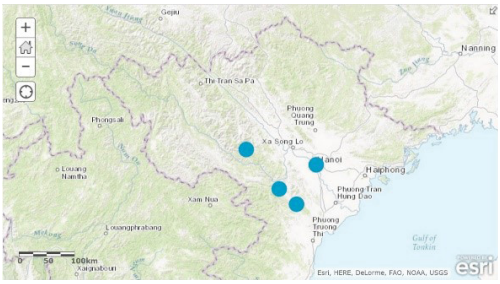
“墓” (●: m)



“納” (●: n)



“御” (●: ŋ)



“仰” (●: ŋ)

“牙” (●: ŋ, ◒: ŋ>ɲ)

Trần Trí Dõi
Faculty of Linguistics, USSH, Vietnam National University
336, Nguyễn Trãi, Thanh Xuân, Hà Nội Việt Nam
Email: ttdoihanh@gmail.com.

A Phonological Basis for Rethinking Vietnamese Isoglosses

Shimizu Masaaki

Graduate School of Language and Culture, Osaka University

Abstract

This study aims to rethink Vietnamese isoglosses through analysis of the dialectal data collected in September 2011 and September 2012. The only comprehensive work on Vietnamese dialectology so far is Hoàng (1989). However, in the nearly 20 years since its publication, many obvious changes occurred in the language's phonological and lexical distribution. We will clarify which features were subject to change and which were not. The main objective of this study is the distribution of the initial consonants that show dialectal variations.

1 Introduction

Vietnamese dialectal variation has been an object of many scholars' interests (Cadier 1902, Maspero 1912, Jones&Huỳnh 1957, Thompson 1959, Hoàng 1989, etc.), and it reflects the geographical spread of the Việt (Kinh) tribe through its long history of migration (Đặng 1994). From the first description of each dialect—especially those of the Christian missionaries—until now, each dialect had undergone several kinds of changes, among which the phonological changes are quite important when considering isoglosses.

2 Previous literature

The only comprehensive work on Vietnamese dialectology is *Tiếng Việt trên các miền đất nước* (Vietnamese of each region) by Hoàng Thị Châu, published in 1989. An important contribution of this work is the fact that it showed clear phonological evidence of dialect division. The isoglosses shown there are still meaningful and remain the basis for dialectal division today. Using the same data as ours, Kondo (2012) proposed new evidence for Vietnamese isoglosses based on the lexical distribution to ensure that the Hoàng's (1989) dialect-division proposal was still significant.

3 Data

The data used in this paper were collected in September 2011 and September 2012 in Vietnam. Most of the contributors were the university students. The data consist of two portions: reading sounds from the syllable list and the lexical questionnaire. This study is based on the sound data. Sixty-one consultants from 60 different birth places were required to read the syllable list aloud as written in the modern Vietnamese orthography, *Quốc Ngữ*, three times for each syllable. The syllable list consists of three sets: initial consonants with simple rhymes (e.g., /-a^{1}/ and a few /-ɔ^{1}/ cases), rhymes with simple initials (e.g., /t-/, /k-/, etc.), and the 6 tones in an open syllable, /ta/, and in a sentence (*Anh Hà có thể đã hẹn*. [ʔaŋ^{1} ha^{2} kɔ^{5} tʰe^{3} ɖa^{4} hɛn^{6}] "Mr. Ha could have made an appointment").

4 Methods and Results

All sound data compiled for reading the syllable set were transcribed with IPA and compared among the dialects. In this study, we concentrate on the initial consonants, among which the letters that show dialectal variation are: *ch*, *d*, *g*, *gi*, *kh*, *r*, *s*, *tr*, *v*, and *x* (see Table.1). The Northern phonemes for all of these letters are fricatives or affricates. The phonetic realization of each letter is as follows:

(1)

Orth.	Sound variation among dialects				Maps
r	[z],	[ʐ],	[r~ɽ],	[ɣ]	(Fig.1)
tr	[tʃ],	[tʂ],	[ʈ]		(Fig.2)
s	[s],	[ʂ],	[ʃ]		(Fig.3)
ch	[tʃ],	[c]			(Fig.4)
d	[z],	[ɗ],	[j]		(Fig.5)
gi	[z],	[ʐ],	[j]		(Fig.6)
v	[v],	[w],	[j]		(Fig.7)
x	[s],	[ʃ]			(Fig.8)
g	[ɣ],	[ɡ]			(Fig.9)
kh	[x],	[kʰ]			(Fig.10)

The distribution of each realized form is plotted in the maps in Figs. 1-10. For some letters, such as *d* and *ch*, the distribution is quite clear-cut, while for the others, it is not. When we compare our results with Hoàng's (1989), shown by the red dashed lines in the maps, the distribution of some letters, such as *r*, *d* and *gi*, coincide with Hoàng's (1989), while others do not. Hereafter, we will attempt to interpret them to consider the isoglosses of present Vietnamese.

5 Discussion

According to Hoàng (1989), the Vietnamese dialects are divided into three groups: Northern (NOR), Central (CEN) and Southern (SOU) dialects. Concerning the initial consonants, the isogloss that divides the NOR and CEN dialects is shown in Maps 1, 2, and 3, while the one that divides the CEN and the SOU dialects is shown in Map 7. The former is located at the border of the Thanh Hoá and Nghệ An provinces, and the latter is at the border of the Thừa Thiên-Huế and Quảng Nam-Đà Nẵng provinces, which coincides with the Hải Vân Pass. First, based on these borders, we will allot the sound variants to three dialects. Their distribution with the Proto Vietmuong (PVM) and 17C phonemes corresponding to each of them (Shimizu 2015) are as follows:

(2)

PVM		17C	SOU	CEN	NOR	Orth.
*p *b	>	*β~βj	j/w/v	v	v	(*v*)
*t *d	>	*ð~ðj	j	j/ɗ/z	z	(*d*)
*c *ɟ	>	*c	c	c	tʃ	(*ch*)
*c *ɟ	>	*ɟ	j	j/ʐ/z	z	(*gi*)
*s	>	*ɽ	ɽ~r/ʐ/ɣ	ʐ	z	(*r*)
*tʃ	>	*ʃ	s	s/ʃ	s	(*x*)
*k *g	>	*ɣ	ɡ~ɣ		ɣ	(*g*)
*kʰ	>	*kʰ	kʰ~x		x	(*kh*)
*Cl-	>	*tl~bl	ʈ~ʈʂ		tʃ	(*tr*)
*Cr-	>	*ʂ	ʂ~ʃ		s	(*s*)

5.1 Geographical Interpretation

Four isoglosses can be settled through Hoàng's (1989) study and our own, and their phonological evidence is summarized as follows:

(3)

Isoglosses	Hoàng (1989)	Data in 2011, 2012
① Thanh Hoa-Nghệ An	*r* [z-ʐ], *tr* [ʈʃ-ʈ], *s* [s-ʂ]	*r* [~z], *tr* [~ʈʃ], *ch* [c~] *
② Ngang Pass	*d* [z-j], *gi* [z-j]	*ch* [~ʈʃ], *d* [z-j], *gi* [j~]
③ Hải Vân Pass	*v* [v-j]	
④ Bình Thuận-Ninh Thuận		*v* [w/j~]

* [~A] denotes that the sound A distributes from the origin northward, while [A~] from the origin southward.

Two isoglosses ① and ② still are meaningful for the present situation, while as far as the initial consonants are concerned, a very important line (③), which is regarded as the border between the CEN and the SOU dialects, does not suit the present situation. The point here is the reading of the letter *v*. According to Hoàng (1989), in the SOU dialect, the letter *v* is usually read as [v] only in the cultural lexicon (i.e., Sino-Vietnamese), while it is read in others as [j]. However, at present, the NOR and CEN [v] have spead to the SOU region.

5.2 Historical Interpretation

The chronological order of each phoneme in (2) and its dialectal distribution can be estimated as follows:

(4)

Orth.	17C						Present	Dialectal distribution
v	*βj	⟍	j	>			j	SOU
			β	>			v	CEN/NOR
d	*ðj	⟍	j	>			j	SOU/CEN
			ð	>			z	CEN/NOR
ch	*c		c	>			ʈʃ	SOU/CEN > NOR
gi	*ɟ		j	>	ʐ̩	>	z	SOU > CEN > NOR
r	*ɽ	⟍	ɽ	>	ʐ̩	>	z	SOU > CEN > NOR
							ɣ	SOU (Mekong Delta region)
x	*ʃ		ʃ	>			s	CEN > SOU/NOR
g	*ɣ		g	>			ɣ	SOU/CEN > NOR
kh	*kʰ		kʰ	>			x	SOU/CEN > NOR
tr	*tl~bl		ʈ	>	ʈʂ	>	ʈʃ	SOU/CEN > NOR
s	*ʂ		ʂ	>	ʃ	>	s	SOU/CEN > NOR

As for the initial consonants, it is quite obvious that the NOR phonemes always show the most innovative forms. Meanwhile, in some cases, such as *gi* and *r*, the SOU dialect preserves the oldest form, and the CEN dialect shows the intermediate forms. In other cases, the CEN dialect preserves the older form (e.g., for *x*) or the SOU and CEN dialects preserve the older ones (e.g., for *ch*, *g*, *kh*, *tr*, and *s*).

6 Conclusion

Through the discussion above, as for the initial consonants, the border between the CEN and SOU dialects are becoming unclear, compared to the situation 20 years ago. However, the border between the NOR and CEN dialects is still quite obvious, which makes the status of NOR speech outstanding, and can give way to the spread of NOR speech to other regions.

Acknowledgements

Our thanks go to Prof. Trần Trí Dõi with the University of Social Sciences and Humanities, Hanoi, for his great help in our research.

References

Cadière, Léopold E. 1902. *Phonétique annamite (dialecte du Haut-Annam)*, E. Leroux, Paris, France.

Đặng Thu. 1994. *Di dân của người Việt từ thế kỷ X đến giữa thế kỷ XIX* [Migration of the Viet from the Xth century to the middle of the XIXth century], Trung tâm Nghiên cứu Dân số và Phát triển, Hà Nội, Việt Nam.

Hoàng Thị Châu. 1989. *Tiếng Việt trên các miền đất nước: Phương ngữ học* [Vietnamese of each region: Dialectology]. Nhà xuất bản Khoa học Xã hội, Hà Nội, Việt Nam.

Jones, Robert B., Huỳnh Sanh Thông. 1957. *Introduction to Spoken Vietnamese*, American Council of Learned Societies, Washington, USA.

Kondo Mika. 2012. A preliminary survey of the isoglosses in Vietnam, *Papers from the First International Conference on Asian Geolinguistics*, December 14-15, 2012, Aoyama Gakuin University, Tokyo, Japan: 45-64.

Maspero, Henri. 1912. Études sur la phonétique historique de la langue annamite: les initiales, *BEFEO*, 12: 1-126.

Shimizu Masaaki. 2015. A Reconstruction of Ancient Vietnamese Initials Using Chữ Nôm Materials, *NINJAL Research Papers*, 9: 135-158.

Thompson, Laurence C. 1959. Saigon Phonemics, *Language*, 35(3): 454-476.

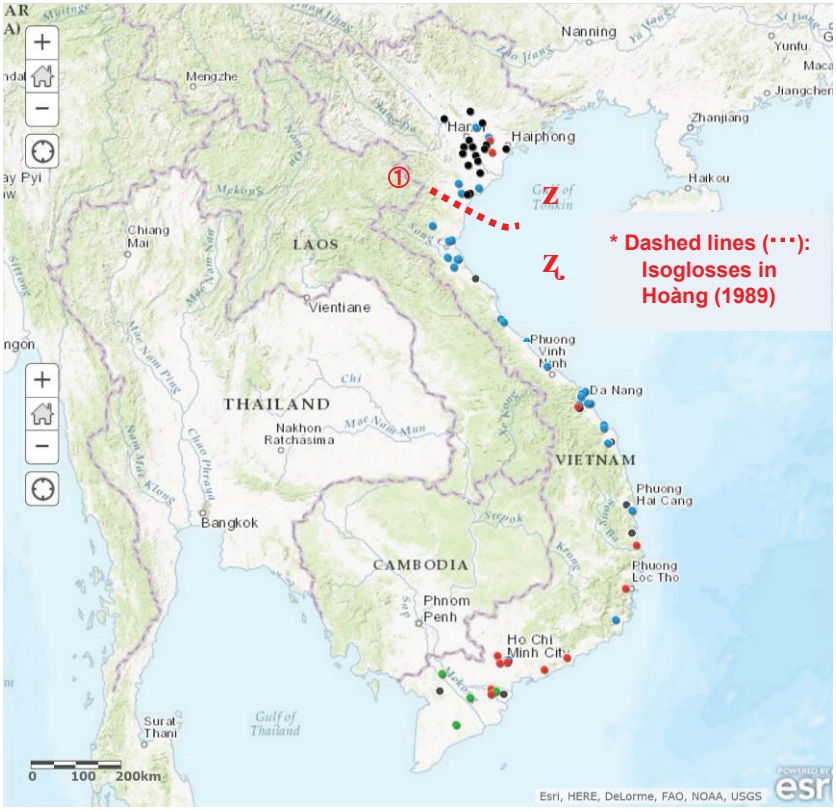


Fig.1 *r* (z: ●, ʐ: ●, r~ɽ: ●, ɣ: ●)

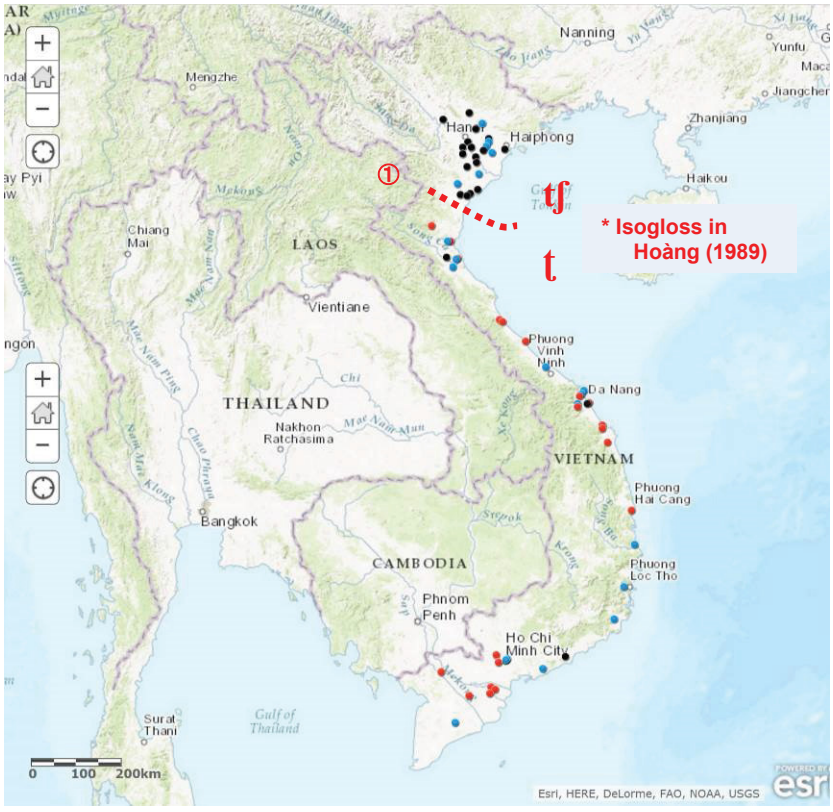


Fig.2 *tr* (tʃ: ●, ʈʂ: ●, ʈ: ●)

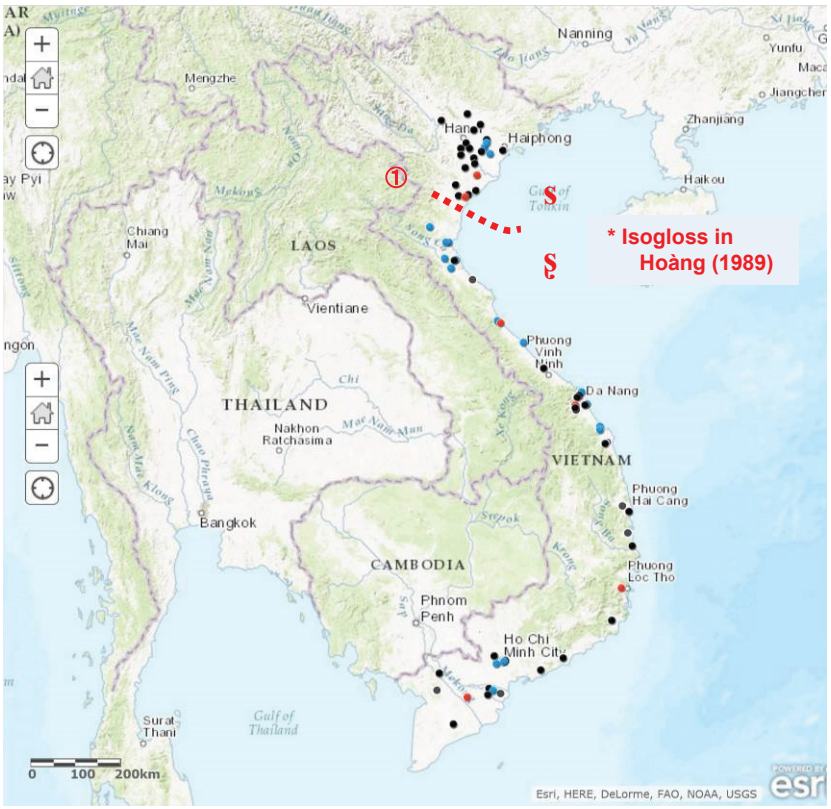


Fig.3 *s* (s: ●, ʂ: ●, ʃ: ●)

Fig.4 *ch* (tʃ: ●, c: ●)

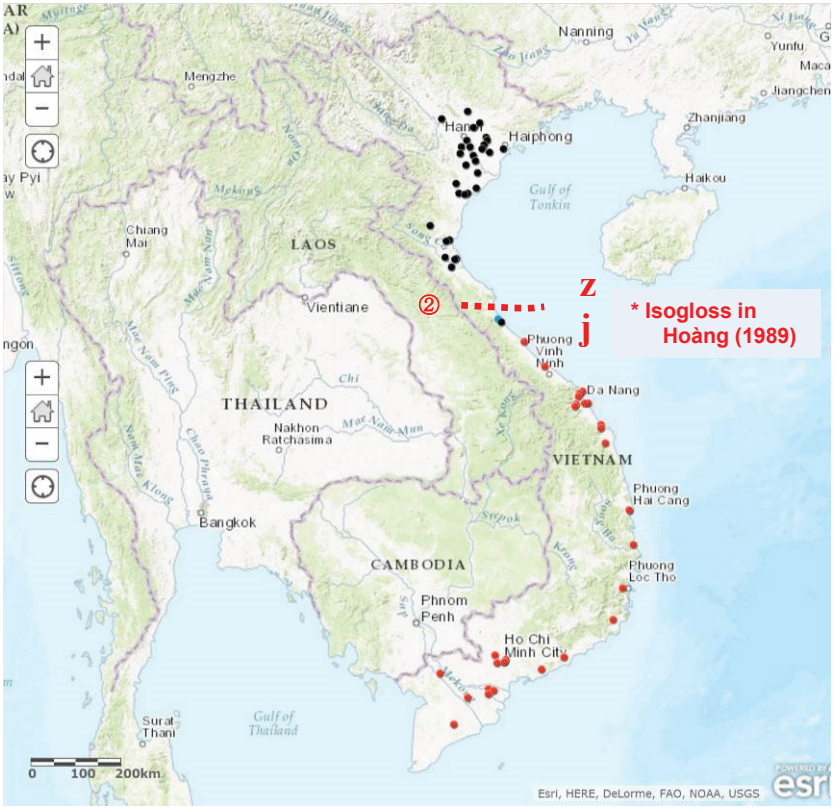


Fig.5 *d* (z: ●, ɗ: ●, j: ●)

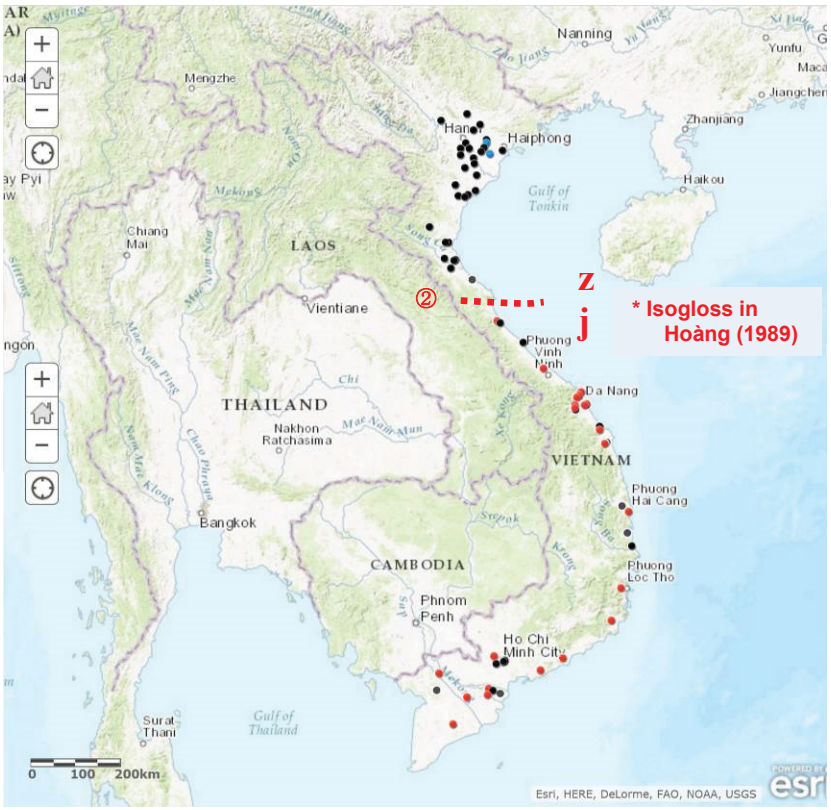


Fig.6 *gi* (z: ●, ʐ: ●, j: ●)

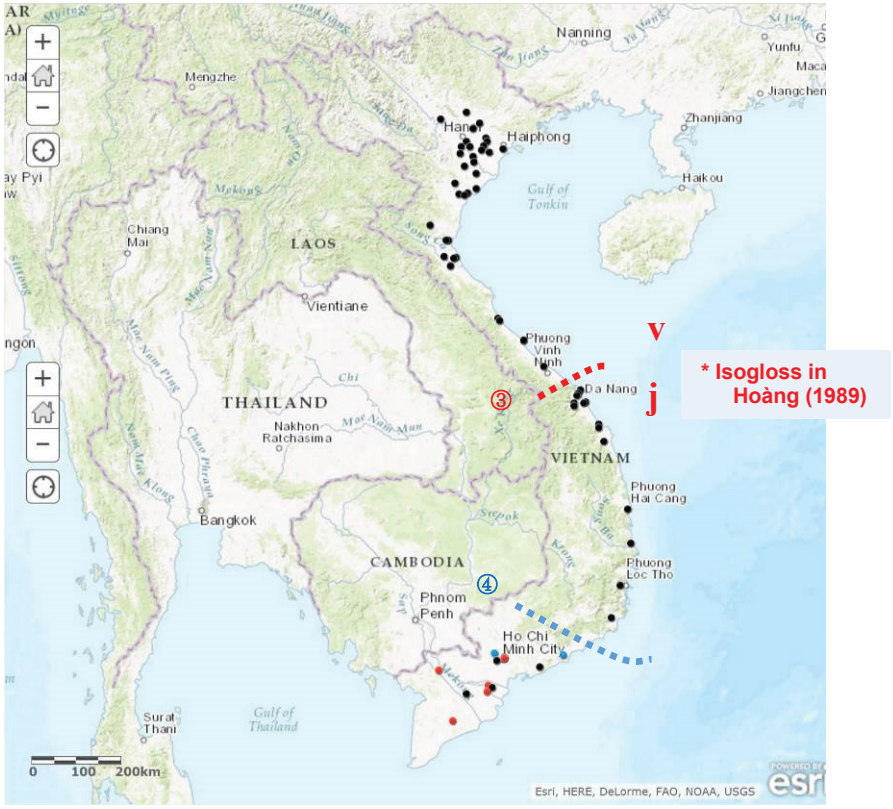


Fig.7 *v* (v: ●, w: ●, j: ●)

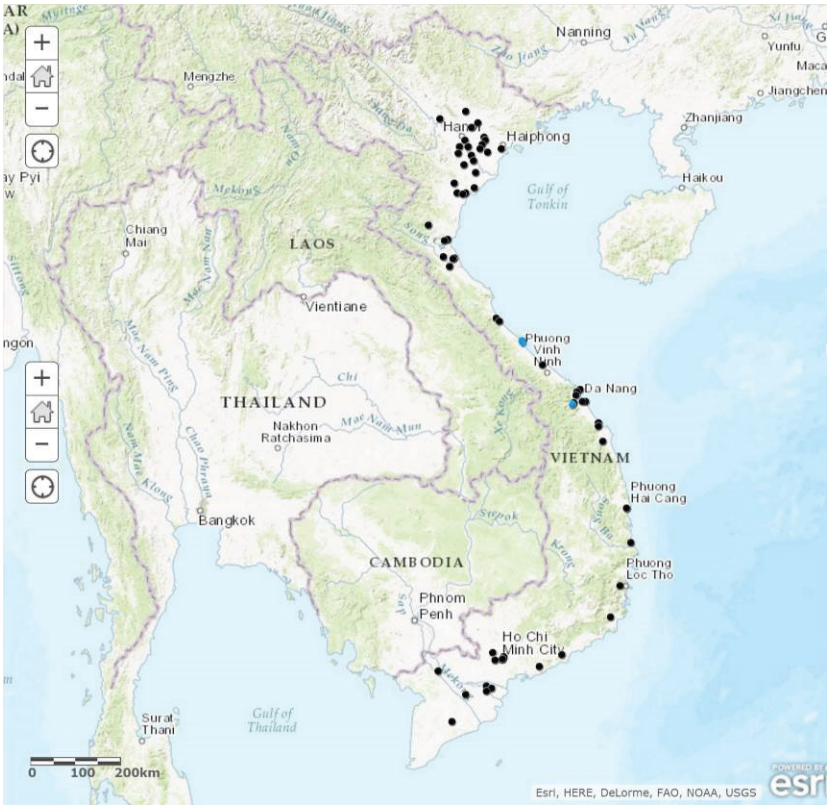


Fig.8 *x* (s: ●, ʃ: ●)

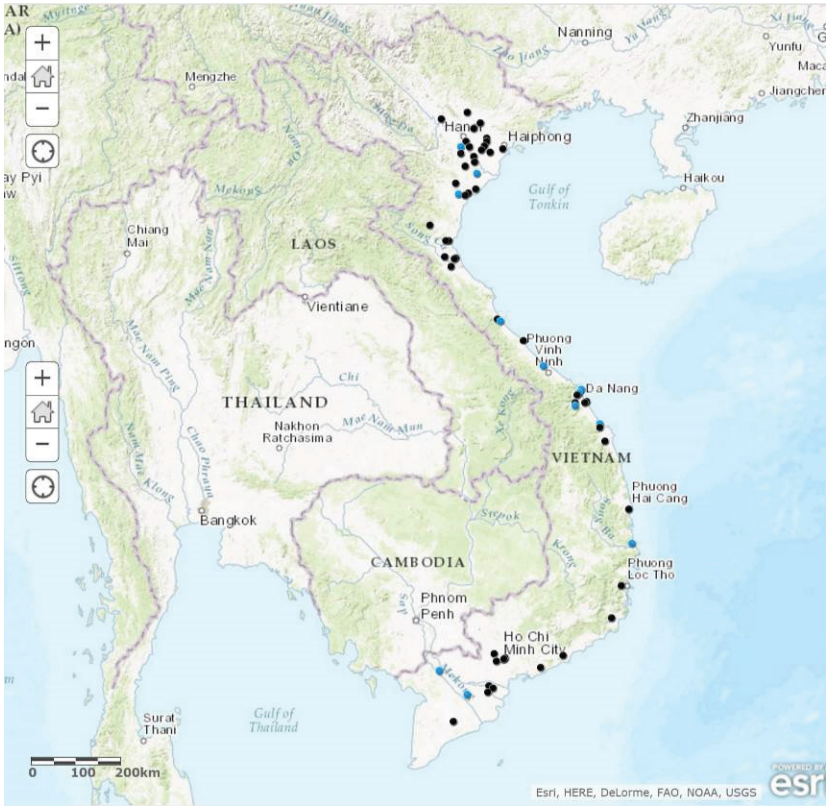


Fig.9 *g* (ɣ: ●, g: ●)

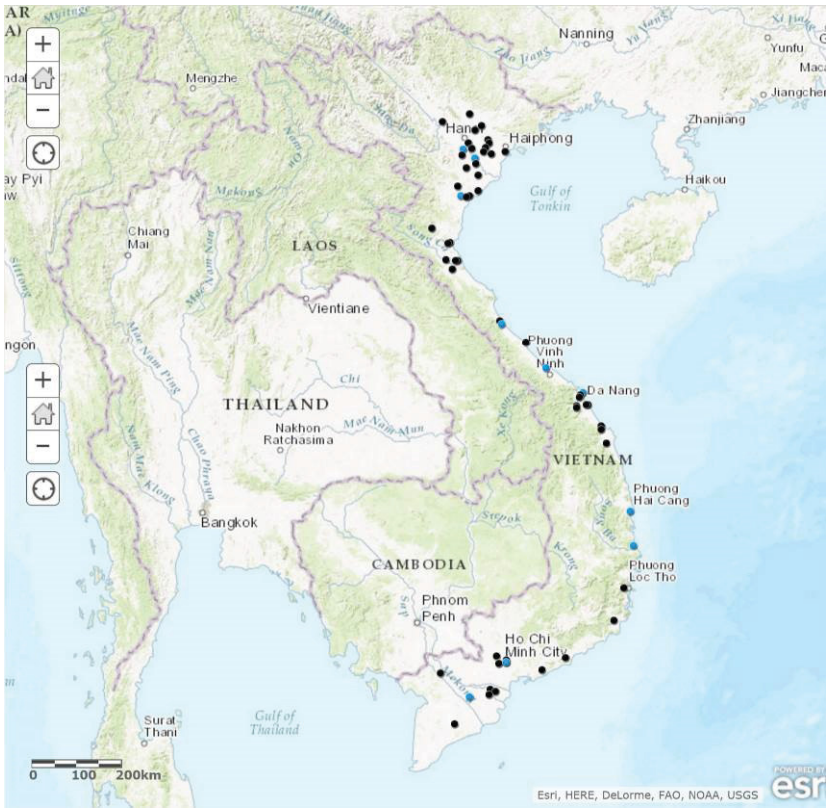


Fig.10 *kh* (x: ●, kʰ: ●)

Table.1

No.	province, etc.	Birth place	*ra*	*trà*	*so*	*cho*	*da*	*giá*	*và*	*xa*	*gà*	*khó*
1	Thái Nguyên	ucky	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
9	Vĩnh Phúc	ducbac	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
15	Bắc Giang	quenham	zaː1	ʈʂaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
6	Bắc Ninh	bacninh	ɽaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
8	Hải Dương	namsach	ɽaː1∽ʐaː1	tʃaː2	sɔː1	tʃɔːɑ1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔːɑ5
19	Hải Dương	ngocchau	raː1∽zaː1	ʈʂaː2	sɔː1∽ʂɔː1	tʃɔː1	zaː1	ʐaː5	vaː2	saː1	ɣaː2	xɔː5
5	Hưng Yên	vinhkhuc	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
20	Hải Dương	trungkhanh	raː1	ʈʂaː2	ʂɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
13	Hưng Yên	thuanhung	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
3	Hà Tây	hatay	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2∽gaː2	xɔː5∽kʰɔː5
11	Hải Phòng	dongphuong	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
18	Hải Dương	tucuong	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
17	Hải Dương	hungthai	raː1>zaː1	ʈʂaː2>tʃaː2	ʂɔː1>sɔː1	tʃɔːɑ1	zaː1	ʐaː5>zaː5	vaː2	saː1	ɣaː2	xɔːɑ5
10	Hà Tây	unghoa	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
2	Hưng Yên	antao	raː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5∽kʰɔː5
7	Hà Nam	lynhan	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
4	Hà Nam	thanhliem	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
16	Nam Định	daithang	raː1∽zaː1∽ʐaː1	ʈʂaː2	ʃɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2∽gaː2	xɔː5
70	Thanh Hoá	thanhtho	ʐaː1	ʈʂaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
14	Nam Định	nghiahung	ʐaː1∽raː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
68	Thanh Hoá	hunglocth	zaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
67	Thanh Hoá	hoangkhanh	ɽaː1∽ʐaː1	tʃaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	gaː2	kʰɔː5∽xɔː5
64	Thanh Hoá	hauloc	zaː1	tʃaː2	ʃɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
73	Nghệ An	quyhop	ɽaː1	ʈaː2	ʂɔː1	cɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
69	Nghệ An	dienthap	ʐaː1	ʈaː2	ʂɔː1	cɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
72	Nghệ An	phuthanh	ʐaː1	ʈʂaː2	ʂɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
62	Nghệ An	vandien	ɽaː1	tʃaː2	ʂɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
65	Nghệ An	hungloc	ɽaː1	ʈaː2	ʂɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
71	Nghệ An	vinh	ʐaː1	ʈʂaː2	sɔː1	tʃɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
12	Hà Tĩnh	buixa	ʐaː1	ʈʂaː2	ʂɔː1	cɔː1	zaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
40	Quảng Bình	lytrach	raː1	ʈaː2	ʂɔː1	cɔː1	ɗaː1	jaː5	vaː2	saː1	ɣaː2	xɔː5
61	Quảng Bình	haidinh	ɽaː1	ʈaː2	ʃɔː1	cɔː1	zaː1	zaː5	vaː2	saː1	gaː2∽ɣaː2	kʰɔː5
38	Quảng Trị	vinhkim	ʐaː1	ʈaː2	ʂɔː1	cɔː1	jaː1	zaː5	vaː2	ʃaː1>saː1	ɣaː2	xɔː5
39	Huế	quangdien	ʐaː1	ʈʂaː2	sɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	gaː2	kʰɔː5
44	Đà Nẵng	thoquang	ʐaː1	ʈʂaː2	ʂɔː1	tʃɔː1	jaː1	jaː5	vaː2	saː1	gaː2	kʰɔː5
54	Đà Nẵng	danang1	ʐaː1	ʈʂaː2	sɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	gaː2	xɔː5
50	Đà Nẵng	camle1	ʐaː1	ʈaː2	sɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	ɣaː2	xɔː5
49	Hội An	cudai	ɽaː1	ʈaː2	ʂɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	ɣaː2	xɔː5
53	Hội An	hoian	ʐaː1	tʃaː2	sɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	ɣaː2	xɔː5
46	Quảng Nam	daihoa	ʐɔ̃ɑː1	ʈʂɑːˑ2	ʃɔː1	cɔː1	jaː1	jaː5	vwɑː2	sɔ̃ɑː1	ɣɑː2	xɔː5
43	Quảng Nam	duyhoa	rɔ̃aː1	ʈɔ̃aː2	sɔː1	cɔː1	jɔ̃aː1	jɔ̃aː5	vɔ̃aː2	ʃɔ̃aː1	gɔ̃aː2	xɔː5
52	Quảng Nam	tamgiang	ʐaː1	ʈaː2	ʂɔː1	cɔː1	zaː1∽jaː1	zaː5	vaː2	saː1	gaː2	xɔː5
48	Quảng Nam	tammydong	ʐɔ̃aː1	ʈɔ̃aː2	ʂɔː1	cɔː1	jɔ̃aː1	jɔ̃aː5	vɔ̃aː2	sɔ̃aː1	ɣɔ̃aː2	xɔː5
42	Quảng Ngãi	tunghia	ʐaː1	ʈaː2	sɔː1	cɑɔː1	jaː1	jaː5	vaː2	saː1	ɣaː2	xɑɔː5
56	Bình Định	quynhon	ʐaː1	ʈaː2	sɔː1	cɔː1	ɖaː1∽jaː1	jaː5	vaː2	saː1	ɣaː2	kʰɔː5>xɔː5
57	Phú Yên	tuyhoa	raː1	ʈʂaː2	sɔː1	cɔː1	jaː1	zaː5	vaː2	saː1	gaː2	kʰɔː5
41	Khánh Hoà	dienkhanh	raː1	ʈʂaː2	ʃɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	ɣaː2	xɔː5
58	Ninh Thuận	phuocson	ʐaː1	ʈʂaː2	sɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	ɣaː2	xɔː5
31	Long An	myhanhbac	raː1∽ɽaː1	ʈaː2	sɔː1	cɔː1	jaː1	jaː5	ʊaː2	saː1	ɣaː2	xɔː5
24	Bình Thuận	thuanquy	raː1	tʃaː2	sɔː1	cɔː1	jaː1	jaː5	waː2	saː1	ɣaː2	xɔː5
26	tp HCM	binhthanh1	ɽaː1	ʈʂaː2	sɔː1	cɔː1	jaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
27	Long An	tphcm2	ʐaː1	ʈʂaː2	ʂɔː1	cɔː1	ɟaː1	dʒaː5>ɟaː5	vaː2	saː1	ɣaː2	xɔː5
25	Tiền Giang	tphcm1	raː1	ʈʂaː2	ʂɔː1	cɔː1	jaː1	zaː5	jaː2	saː1	ɣaː2	xɔː5>kʰɔː5
23	tp HCM	binhchanh	raː1	ʈaː2	ʂɔː1	cɔː1	jaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
30	BR-VT	xuyenmoc	raː1	ʈʂaː2	sɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	ɣaː2	xɔː5
32	An Giang	kienan	ɣaː1	ʈaː2	sɔː1	cɔː1	jaː1	jaː5	jaː2	saː1	ɣaː2∽gaː2	xɔː5
21	Bến Tre	p3bentre	raː1	ʈaː2	sɔː1	cɔː1	jaː1	jaː5	jaː2	saː1	ɣaː2	xɔː5
35	Bến Tre	luongquoi	ɣaː1	ʈaː2	ʂɔː1	cɔː1	jaː1	zaː5	vaː2	saː1	ɣaː2	xɔː5
22	Bến Tre	phuochiep	raː1	ʈaː2	sɔː1	cɔː1	jaː1	jaː5	jaː2	saː1	ɣaː2	xɔː5
36	Vĩnh Phúc	tambinh	ɣaː1∽gaː1	ʈaː2	ʃɔː1	cɔː1	jaː1	jaː5	vaː2	saː1	gaː2	kʰɔː5
28	Sóc Trăng	tanlong	ɣaː1	ʈʂaː2	sɔː1	cɔː1	jaː1	jaː5	jaː2	saː1	ɣaː2	xɔː5

Domaaki as a language of northern Pakistan: from a geolinguistic point of view

YOSHIOKA Noboru

National Museum of Ethnology, Japan

Abstract

This is a pilot study to draw the linguistic similarity and dissimilarity among languages graphically as treating Domaaki, spoken in northern Pakistan, being a geneologically once or still conrtoversial language as the main subject of discussion. By means of geolinguistics I plot maps of languages/dialects with differently coloured symbols which reflect the ratio of idiosyncretic items and the deviation of similarity to any ideal language(s). As a result Domaaki appears to be a geneologically Central Indic language being more affected by Dardic languages than Burushaski, indeed.

1 Introduction

The Domaaki language, an Indo-Aryan language spoken in northern Pakistan, has been considered either as a Dardic language or as a Central Aryan language. Even though there were actually (but few) description on the language, its geneological position was, or is still, controversial among scholars. I think it is caused by that the characteristics of Domaaki include the ones of both language groups. While the strict position is unknown within the Indo-Aryan group, Domaaki shows much similarity to another surrounding language, Burushaski, too.

Here, I try to illustrate how the Domaaki language is close to Dardic languages, Central Aryan languages, and Burushaski graphically by taking a hint from geolinguistic ways to show data.

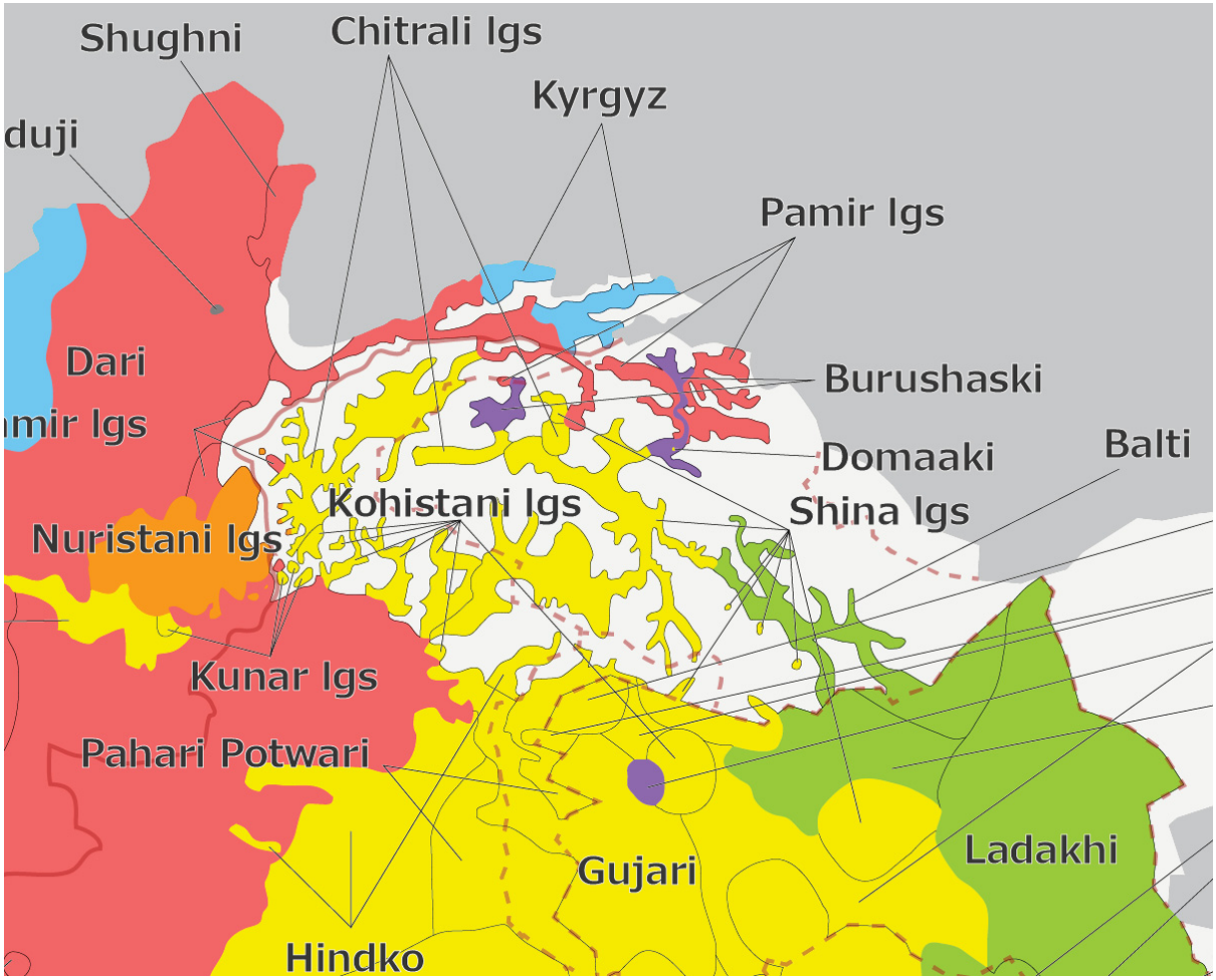
2 Domaaki and the surrounding languages

Domaaki is the ethnic language of Doma /ɖóma/ people. But actually many Doma in northern Pakistan do not speak Domaaki, and they have selected to use other languages most of which are the languages of other ethnic groups, such as Burushaski, Shina, Khowar, Wakhi, Balti.

In northern Pakistan, there are many languages from different language families / branches (see Map 1 on the next page). Domaaki itself is a Central Indic language as an outland far from the area of the Central languages including Hindi-Urdu. This language is completely enclaved by Burushaski which is a language isolate (purple in the map). And around Burushaski there are some Dardic languages, including Shina, Khowar, Kalasha, and Kashmiri, which are classified as Northwestern Indic languages (Indic is yellow in the map). While, north to Burushaski, an Eastern Iranian language Wakhi is spoken (Iranian is red). In the west side of northern Pakistan, there is also a Nuristani language (orange coloured) Kati spoken by a few people. The Nuristani branch belongs to Indo-Iranian group, but it is on the same level as Indic (Aryan) and Iranian. At the southeast to Shina area, they speak Balti being a Tibetan language (light green in the map).

Now, based on my research, the Domaaki language is spoken only in two communities. The first is Mominabad village of Hunza valley, in Hunza-Nager district in Gilgit-Baltistan administrative territory, and the second is Bedishal hamlet at Uyum Nager village of Nager valley, in the same district. Doma people living in the other places, they can no longer speak the language. The speakers of Domaaki are in total less than 500 people at most. They use Burushaski for daily life in their community nowadays, too.

Historically Domaaki has sometimes been considered as a Dardic language like to the other Indo-European languages in this area. Or some scholars may be still thinking so, while many evidences



Map 1. Languages and Language families / branches in northern Pakistan

say "no". Because Domaaki shows much similarity with Dardic languages I think. And that is why I want to show the characteristics of Domaaki now.

In this pilot study, I treat Urdu and the following languages and dialects which I have surveyed on by myself: Mominabad Domaaki (2 consultants' data), Bedishal Domaaki, Karimabad Burushaski, Ganish Burushaski, Nager Burushaski, Yasin Burushaski, Gilgit Shina, Ishkoman Khowar, Yasin Khowar, and Rumbur Kalasha.

3 How to draw the similarity and dissimilarity among languages

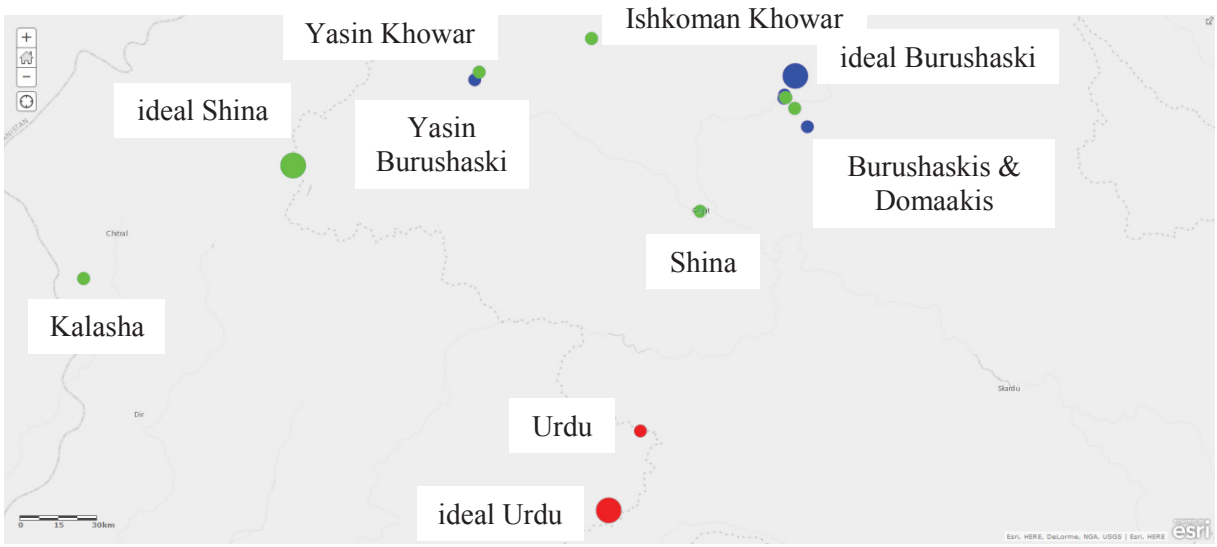
Taking a hint from geolinguistic way to visualise, I illustrate the distance among the characteristics of languages in northern Pakistan here. On Map 1 above the language groups are distinguished by colours, but the distinction is based on the geneological classes and the colouring does not reflect actual similarity and dissimilarity among languages, indeed.

Meanwhile in geolinguistics, I felt that the results of study usually show dis/similarity among languages despite of language families. Some of them show the cases that geographic closeness reflects linguistic similarity, and something else show that geographic closeness sometimes cannot surpass the geneological bonds. Anyway it seems the geolinguistic ways are valid to know linguistic dis/similarity among neighbouring languages. The ways can be useful to illustrate the characteristics of a once controversial language, Domaaki.

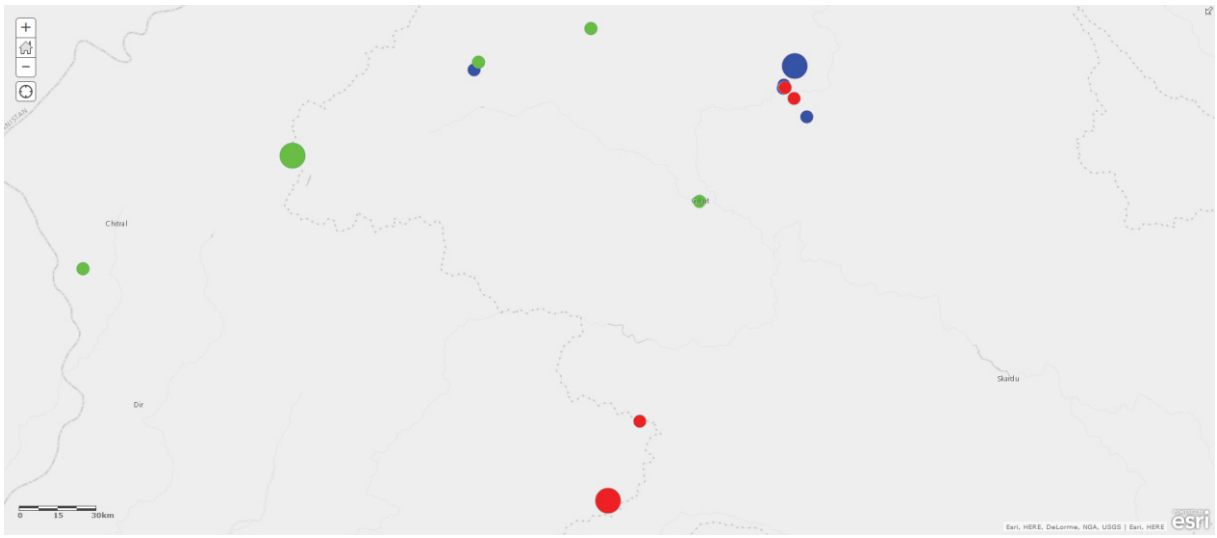
In the following procedure I try to draw out the character of Domaaki:

1. Fixing three primary colours, red, green, and blue, on the three ideal (typical) languages concerning to Domaaki, that is, Urdu as a Central Indic language, Shina as a Dardic language, and Burushaski as a circumjacent language, respectively.
2. Comparing the linguistic items of 11 sample languages (see §2 above) to the ones of three ideal languages.
   i. When an item is (apparently) cognate with any ideal language, then the said language will be coloured with the colour of the ideal language.
   ii. When an item is not cognate with any ideal language, then it will not get any colour.
   iii. When an item is cognate with two or three ideal languages, then it gets all colours of the ideal langauges concerned. (For example, an item cognate both with Urdu and Shina, then the item will be cloloured yellow, which is made with red and green.)

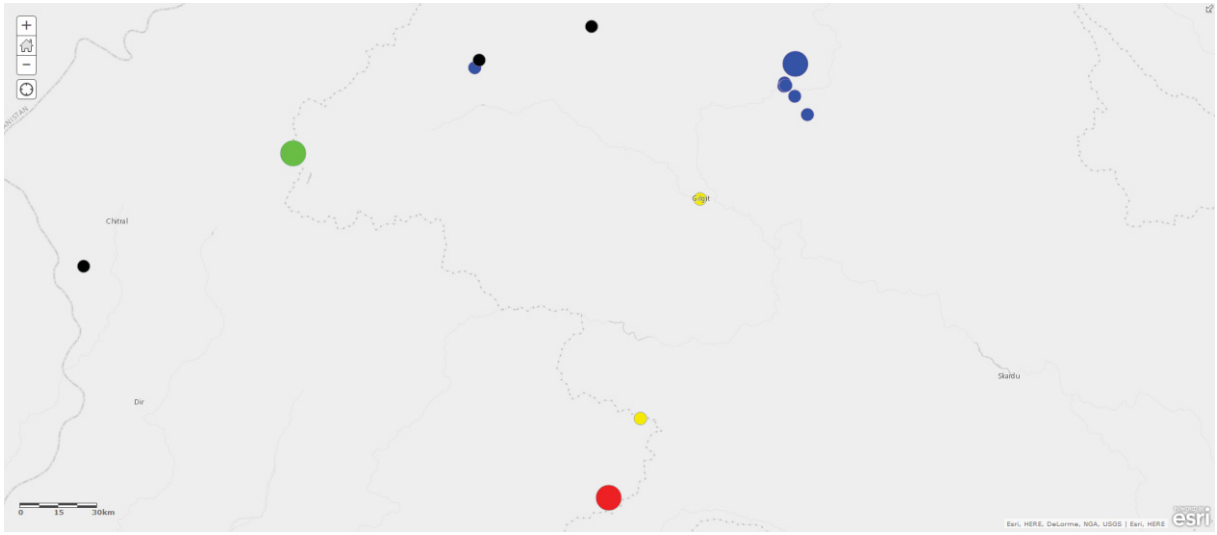
At this stage, a map is drawn for an item. The following three maps are examples of basic vocaburaly items "stone", "water", and "write". In this scale, the dialects of Burushaski and Domaaki in Hunza and Nager valleys are overlapping, but to explain the making method of maps and to inform how languages are similar or dissimilar in this wider ranged area. Notice that Urdu position in these maps are not strict, and the actual position of Urdu is more southward. Ideal languages are plotted with larger circles.



Map 2. The map of "stone"



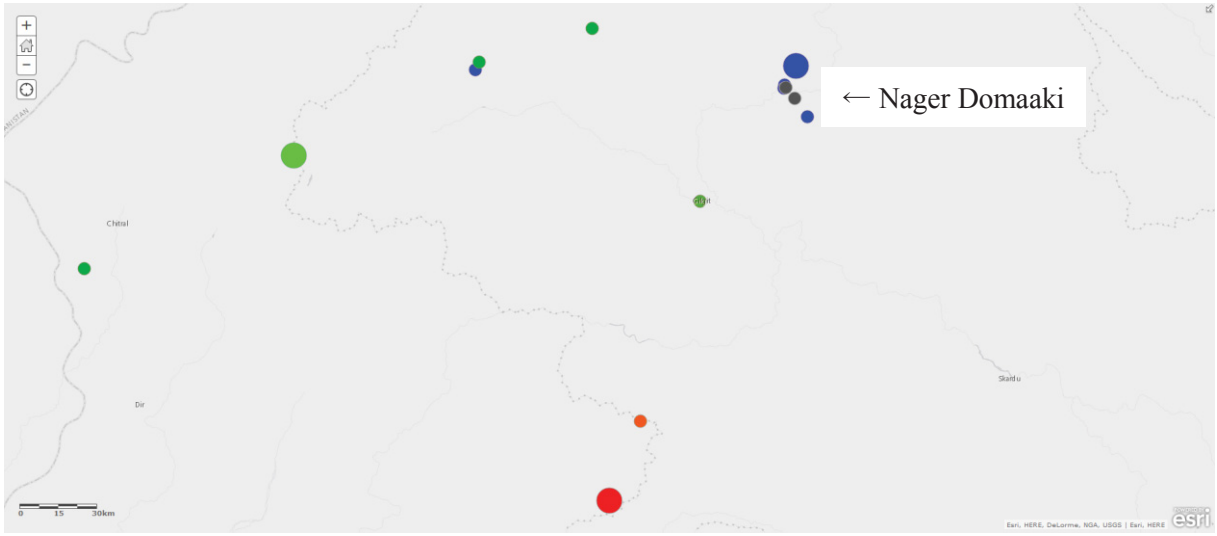
Map 3. The map of "water"



Map 4. The map of "write"

In Map 4, Shina and Urdu show the same origin and then they are shown in yellow colour, while other Dardic languages, i.e. Khowar and Kalasha, have other type of forms and then they are black (no colour) here.

3. Piling up given maps for each item. At the time, RGB colour scores (8-bit) of each point are calculated into the average marks.



Map 5. The piled map of "stone", "water", and "write"

For example, Nager Domaaki in Map 5 is coloured with gray [85.85.85: RGB], as a result of three scores, green [0.255.0] in Map 2, red [255.0.0] in Map 3, and blue [0.0.255] in Map 4.

In this way, we can see the likelihood of each language comparing to ideal languages.

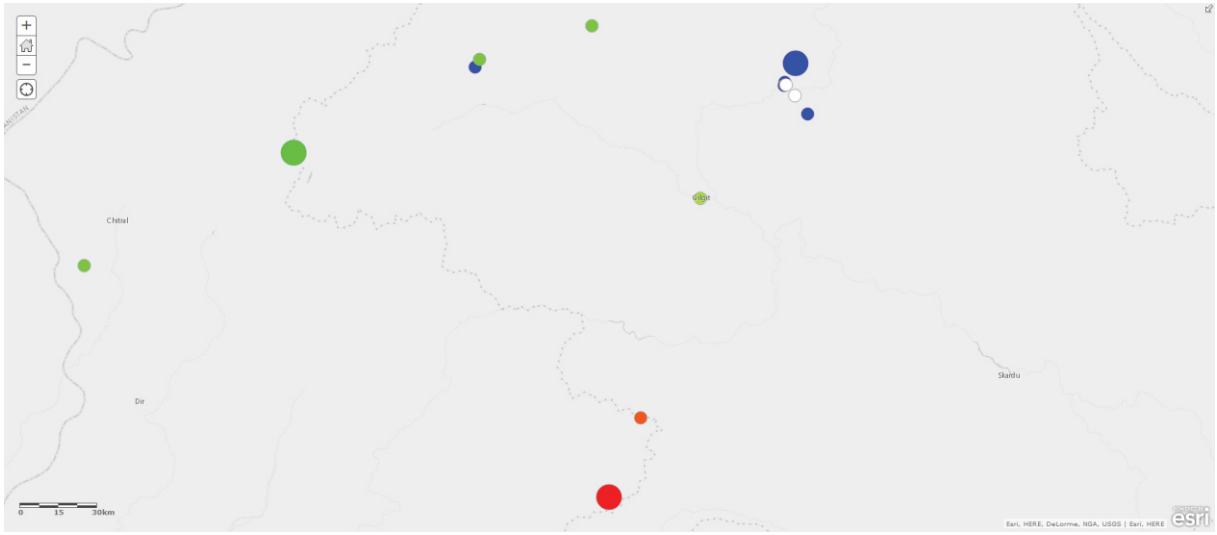
Luminousity reflects the degree how many items which are (not) cognate with the ones of some ideal language a said language has. The darker a language symbol is, the more idiosyncratic items the language has.

Saturation indicates the degree how much a language inclines toward some ideal languages. If a language is symbolised more colourfully, it means the language is more specifically similar to less than three ideal languages.

Of course, hue shows to which ideal language a given language is similar on asked items.

But less luminous colours are surely difficult to distinguish by human eyes, so that I make one more kinds of maps which is inflated luminousity.

4. Inflating luminousity of maps made in the #3 process for the sake of perceptibility, as adding same scores on each of RGB to make the largest score up to 255.



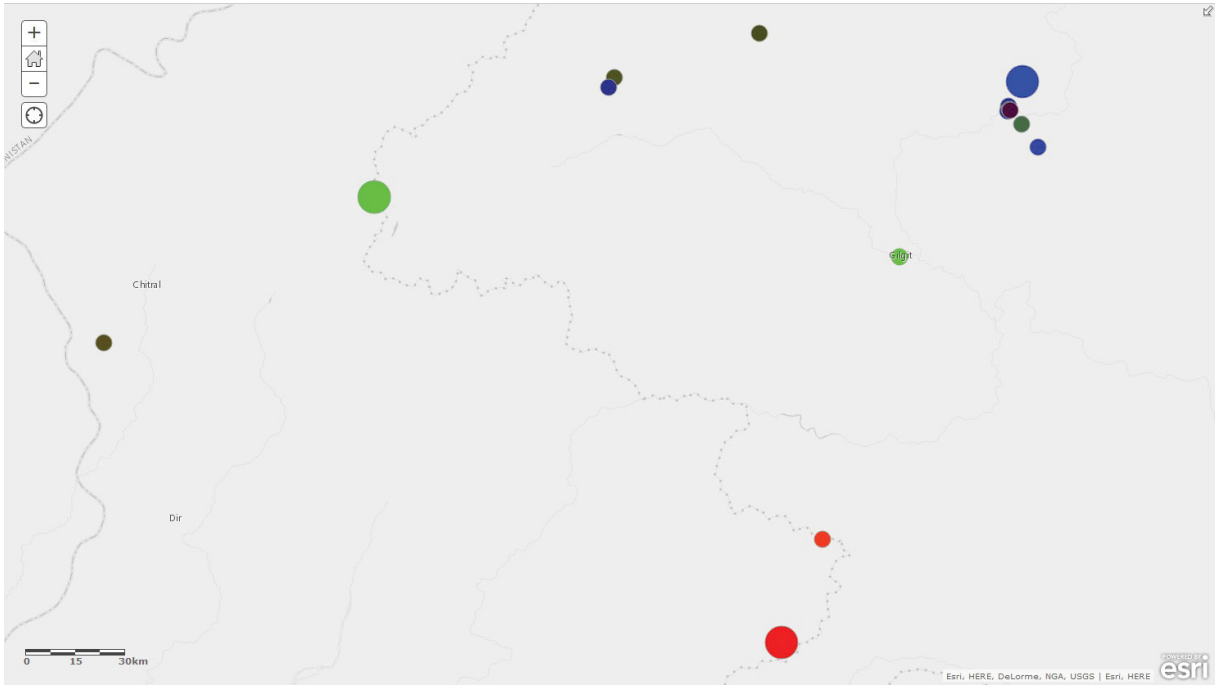
Map 6. The inflated map of "stone", "water", and "write"

Now the luminousity is amplified and symbols remain only saturation and hue values. For example, Nager Domaaki in Map 6 is symbolised with completely white colour, because the data of the idiolect is trichotomic and its [85.85.85] score has become [255.255.255]. It is not similar to any ideal language specifically here.

4 Vocaburaly

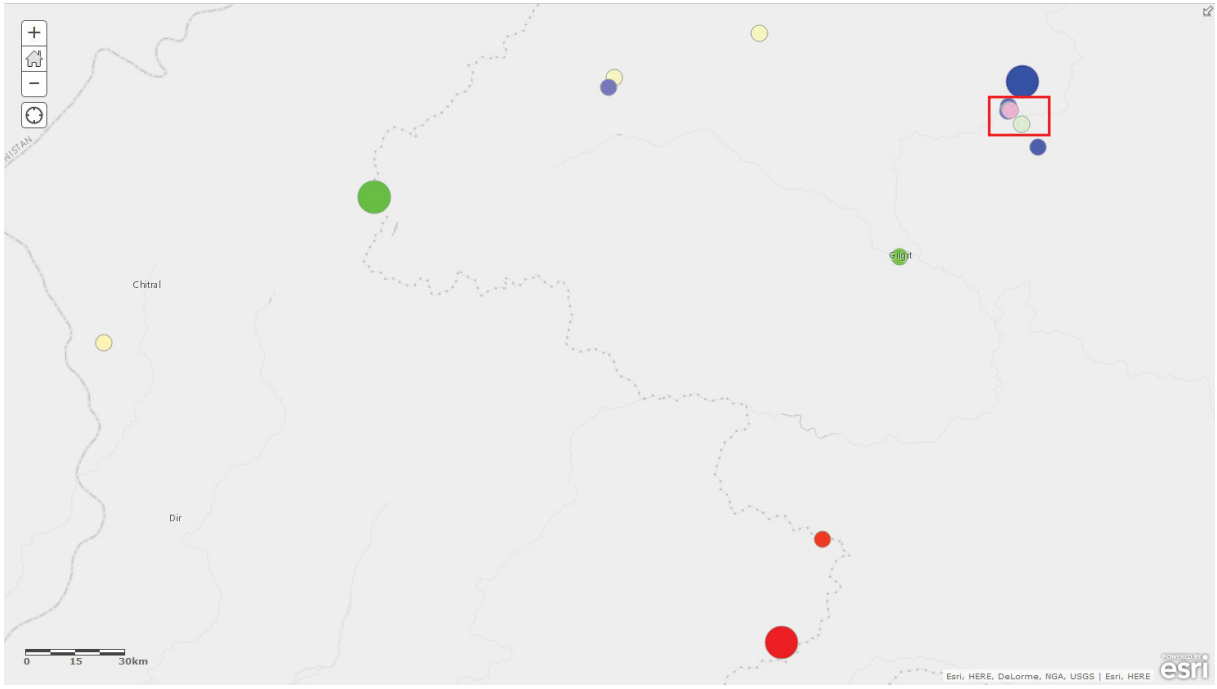
At first I illustrated the languages of northern Pakistan on 480 basic vocabulary. In my fieldwork I almost always start survey with 500 basic vocabulary based on Tokunaga et al. (1967). Some words in the questionnaire are not suite for northern Pakistan, and then I treat only 480 words' data for this study.

The result is as follows:



Map 7. The piled map of 480 vocabulary

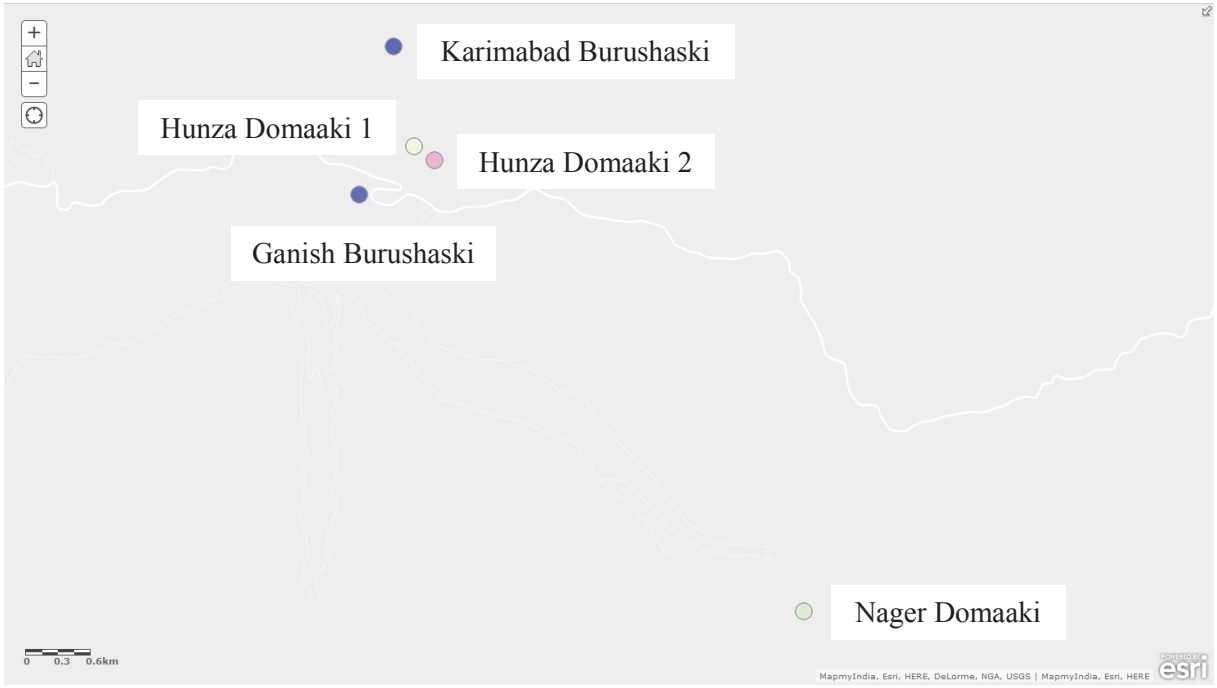
Except Burushaski, Shina, and Urdu, all languages have got very dark. They are not quite similar to any of ideal languages.



Map 8. The luminousity inflated map of 480 vocabulary

The luminousity inflated map shows the points: Khowar and Kalasha being yellow coloured have both characteristics of Dardic (Shinaish) and Central (Urduish). This fact can be said that Shina may not be a typical Dardic language.

To see the Domaaki situation, I enlarge the enclosed part in Map 8 below.

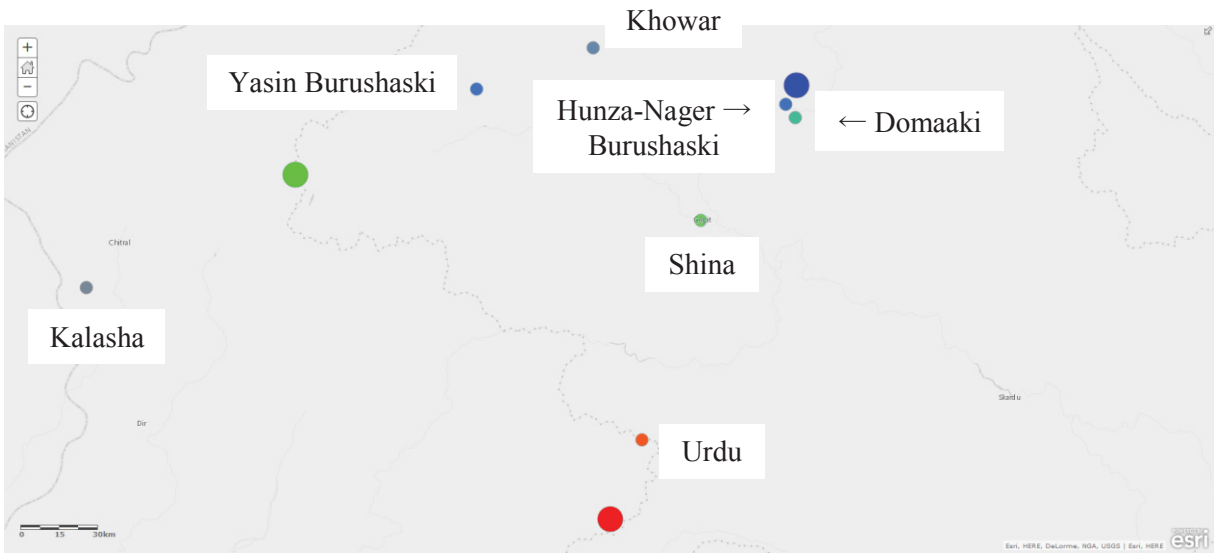


Map 9. The luminousity inflated map of 480 vocabulary in Hunza-Nager

We can see that three Domaaki symbols get different colours. Nager Domaaki, of a 50's man, shows the most Shinaish characteristics and Hunza Domaaki 2, of a 20's man, is sufficiently Urduish. Hunza Domaaki 1, of a 30's woman, takes a middle position between the two idiolects.

5 Grammatical topics

I illustrate similarity maps not only on 480 vocabulary, but also on 15 grammatical items in the same way.



Map 10. The piled map of 15 grammatical items

The 15 items are as follows:

Numeral base (decimal / vigesimal)	Indefinite singular affix (y/n)	Indefinite plural affix (y/n)
Fixed-segment of echo formation (/w/, /ɕ/, /m/)	Causative derivational affix (y/n)	Anticausative derivational affix (y/n)
Retroflex affricates (y/n)	Voiced aspirate consonants (y/n)	Nasal vowels (y/n)
Complex predicates (y/n)	Passive construction (y/n)	Impersonal causal expressions (y/n)
Use of prefixes (y/n)	Gender system (2-way / 4-way)	Neutralisation of negative copulae (none / all / partial)

Map 10 is not inflated on the luminousity. There are several areal features on the grammars in this area, northern Pakistan, that is a reason we would misunderstand what are inherent characteristics and what are a posteriori auquired characters under language contacts for a long period. Actually, Domaaki shows bluish green colour here, which is more bluish than other Dardic languages but rather greenish than bluish at the score [R:44 G:121 B:99].

Of course, it is very problematic, I am feeling, about that here I treat every grammatical features as if they have the same value each other.

6 Conclusions and further issues

This is a pilot study to illustrate linguistic similarity and dissimilarity by means of geolinguistic graphical display.

There still remain many points to be improved, actually. However I could aware some characters anew with this study, such as that the personal vocabularies of three Domaaki consultants vie with each other whether they have more Dardic or Central Indic cognate words than the other branch's and Burushaski words. I was always feeling in fieldworks that Domaaki has so many Burushaski characteristics that it might be a creole between Indic and Burushaski, but now my such misunderstandings are completely wiped out by the results. Domaaki is never a kind of Burushaski creole. Illustrated maps in this study tell us that Domaaki is possibly a language of Central Indic group geneologically, whereas it is well affected by Dardic, or specially Shina, as well as Burushaski. The reason why Dardic influence is stronger than Burushaski effect may be at the primitive commonality of the genes of Domaaki, Shina, and Burushaski or at the original similarity of lingustic parts among the languages concerned.

It is interesting that western Dardic languages, i.e. Khowar and Kalasha in this study are considerably different from Shina as an eastern Dardic language both on the vocabulary and the grammar. There are more western and eastern Dardic languages in a wide area from northwestern Pakistan to northwestern India. As a farther issue, it should be discussed on where, or which language, is the centre of Dardic languages. Shina does not seem so.

Acknowledgements

I am deeply grateful to my field consultants: Qhushan (Mominabad Domaaki), Ali Ahmad Jan (Mominabad Domaaki), Gohar Ali (Bedishal Domaaki), Amir Qhan (Karimabad Burushaski), Musa Baig (Ganish Burushaski), Ainur Qhayat (Nager Burushaski), Arshad Ali (Yasin Burushaski), Masood Rehman (Gilgit Shina), Said Ghulam Nabi Wafa (Ishkoman Khowar), Sharif Ullah (Yasin Khowar), and Yasir Arafat (Rumbur Kalasha). For the remaining flaws I can blame only myself. This work was supported by JSPS KAKENHI Grant Number 15H05380.

Reference

Tokunaga, Yasumoto et al. (eds.) 1967. *Linguistic Questionnaire for Asia and Africa 2*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Geographical Distribution of the /r/ type sounds in Zhuang

Mitsuaki Endo

Aoyama Gakuin University

Abstract

A chronological order concerning / r / type sounds in Zhuang is given based on processes of merger as well as geographical distribution. Complicated types are found in the Southern group; however, it is easier to reconstruct the changing process step by step in such an area. Directionalities of sound changes were extracted from these concrete changes.

1 Introduction

This paper discusses the geographical distribution of the / r / type sounds in Zhuang, a Tai-Kadai language located mainly in the Guangxi Zhuang Autonomous Region, China. Yuan (1963) [1] first gave 11 corresponding rules with full data for 98 words/morphemes among the 51 Zhuang dialects, as follows:

Reconstructed Forms	Sound Correspondences among Dialects	Lexical Items	Frequency
*ʔr	hj : ɣ : r : ð : z : h : j : l : hl	1-9	9
*ʔrᵤ	hj : ɣ : r : ð : z : j : l : hl : jw : zw : lw : w(v)	10-15	6
*thr (*tr)	hj : h: ɣ : r : ð : z : j : l : th (t)	16-26	11
*thrᵤ (*trᵤ)	hjw : hw : hj : h : ɣ : r : ð : z : j : jw : l : lw : th (t)	27-28	2
*khr (*kr)	hj : ɣ : r : ð : z : j : l : hl : khj : h	29-36	7
*khrᵤ	hjw : hw : hj : h : ɣ : rjw : rw : r : ð : zw : z : jw : j : lw : l : w : hl : khj : kh	37-41	4
*r (*rᵤ)	hj : ɣ : r : ð : z : θ : j : l	42-67	26
*nr (*nrᵤ)	hj : ɣ : r : ð : z : θ : j : l : n	68-75	8
*ɣ	hj : ɣ : r : ð : z : θ : j : l : hl : ɬ	76-91	16
*ɣᵤ	hjw : hj : hw : rw : ɣ : r : ð : z : θw : jw : lw : hl : ɬ : h	92-95	4
*ŋr	hj : ɣ : r : ð : z : θ : j : l : n : ŋ : h	96-98	3

Table 1. Sound Correspondences concerning / r / in Zhuang

These rules were extracted with an orthodox comparative linguistic method, and they serve as sound change chains among cognate words/morphemes. Here, we can ask further questions, such as what are their geographical distibutions, and how did these differences come about. As a matter of fact, Yuan (1963: 191) declared that this detailed dialectal information was given for drawing dialect maps. Now, we are going to put his plan to practice.

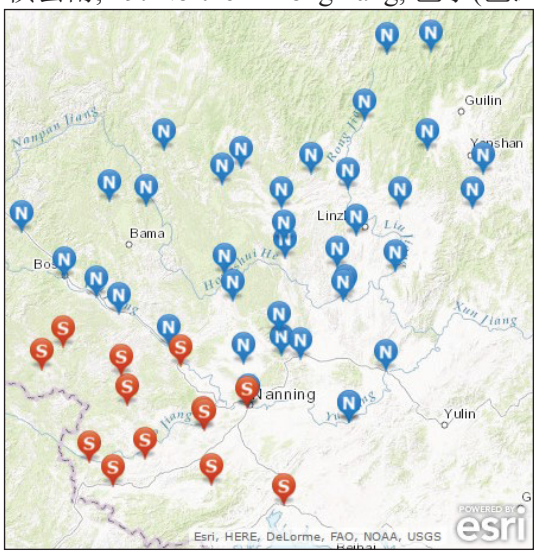
[1] This paper was included in Yuan (2010) as well, but 4 lines were lacking from line 15 of page 173, due to misprinting.

2 Places and subdivisions of the Zhuang dialects

The places which Yuan (1963) reported are as follows: 1. Liujiang, 柳江; 2. Northern Laibin, 来宾(红水河北); 3. Yishan, 宜山; 4. Liucheng, 柳城; 5. Xincheng, 忻城; 6. Luocheng, 罗城; 7. Huanjiang, 环江; 8. Damiaoshan, 大苗山（大庙山）; 9. Rong'an, 融安; 10. Sanjiang, 三江; 11. Yongfu, 永福; 12. Longsheng, 龙胜; 13. Hechi, 河池; 14. Nandan, 南丹; 15. Donglan, 东兰; 16. Du'an, 都安; 17. Mashan, 马山; 18. Shanglin, 上林; 19. Southern Laibin, 来宾(红水河南); 20. Wuyi, 武宣; 21. Xiangxian, 象县; 22. Luzhai, 鹿寨; 23. Guixian, 贵县（贵港）; 24. Lipu, 荔浦; 25. Yangshuo, 阳朔; 26. Binyang, 宾阳; 27. Northern Hengxian, 横县北; 28. Southern Hengxian, 横县南; 29. Northern Yongniang, 邕宁(邕江北); 30. Wuming, 武鸣; 31. Pingguo, 平果; 32. Tiandong, 田东; 33. Tianyang, 田阳; 34. Baise, 百色; 35. Tianlin, 田林; 36. Longlin, 龙林; 37. Lingle, 凌乐; 38. Fengshan, 凤山; 39. Qinxian, 钦县; 40. Southern Yongning, 邕宁(邕江南); 41. Long'an 隆安; 42. Northern Fusui, 扶绥(北); 43. Southern Fusui, 扶绥(南); 44. Shangsi, 上思; 45. Chongzuo, 崇左; 46. Ningming, 宁明; 47. Longzhou, 龙州; 48. Daxin, 大新; 49. Tiandeng, 天等; 50. Debao, 德保; 51. Jingxi, 靖西.

The Zhuang language is divided into two dialect groups: The Northern dialects consist of places No. 1 to No. 38, and Southern dialects consist of No. 39 to No. 51 as seen in map 1. The Northern and Southern groups, belong to the Northern Tai and Central Tai, respectively according to the standard classification of Tai languages as given in Li (1977).

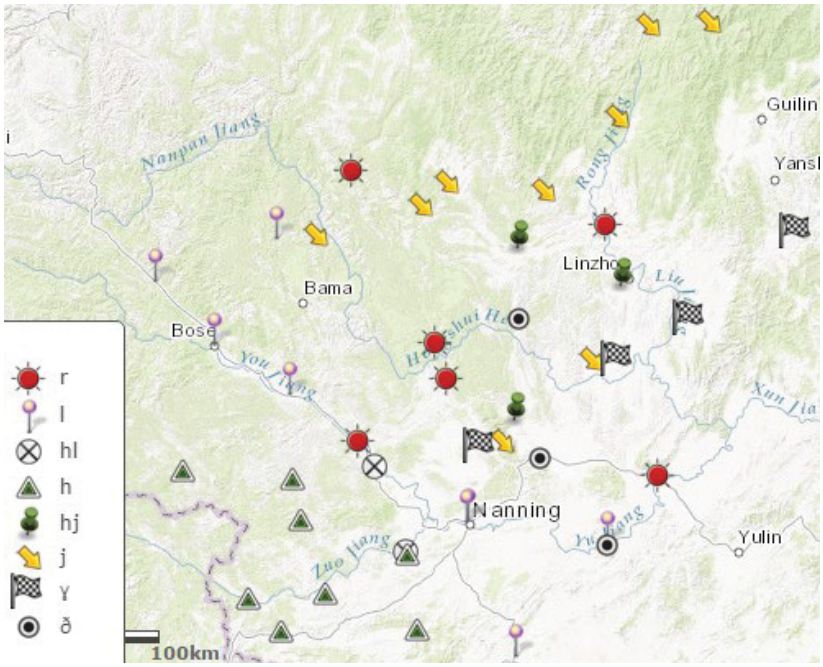
Hereafter, all maps are drawn with Arc GIS Online.



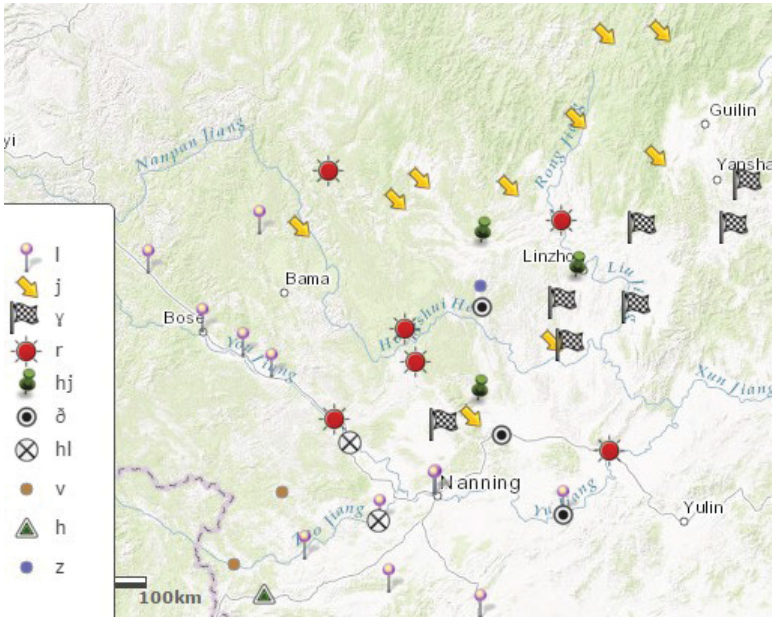
Map 1.
Northern and Southern groups of Zhuang

3 Specimens of individual maps: lexical items No. 1 and No. 12

The 98 lexical items are drawn in linguistic maps as seen in the Appendix. Here are two specimens for lexical items No. 1 and No. 12.



Map 2. Lexical item No. 1 Stem morpheme of *Amaranthus tricolor*



Map 3. Lexical item No. 12 *mushroom*

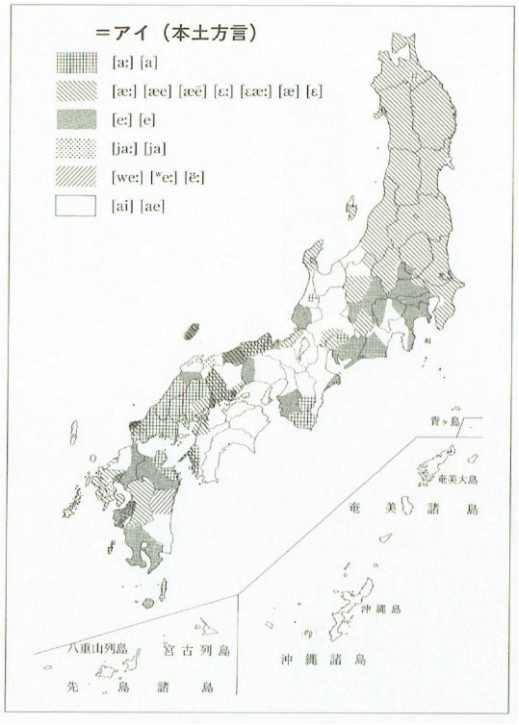
For each lexical item, only data for a cognate form with / r / type sound is found in Yuan (1963). Therefore, places having such a form differ from word to word; in some items, almost all dialects occupy the same form, while in others, only few places occupy such a form.

It is observed that rivers and roads play a role in the dispersion of a linguistic feature. For example, form [l] is found along the Youjiang River and Yujiang River in map 2, as well as along the Zuojiang River in map 3, and so on. Wei (2015) pointed out a similar observation drawing a linguistic map of "house" with / r / type initial consonant in Zhuang. Furthermore, he postulated the following changing process using a diagram as seen in Pan (2010):

$$*r \rightarrow r \rightarrow l / hl \rightarrow j / hj \rightarrow z \rightarrow ð$$
$$\nearrow\ ɣ \qquad\qquad \searrow\ ɬ \qquad\qquad \searrow\ θ$$

This is an interesting and meaningful attempt, but since this directionality was given merely by theory of sound change, it is necessary to find evidence to prove the hypothesis,

Unfortunately, geographical distributions give little help to decide the chronological order of a phonological phenomenon. Since the norm of ABA distribution or peripheral distribution depends on the First Principle of Linguistics by Saussure, e.g. arbitrariness of language sign, it is applicable to a lexical item which is a combination of sound and meaning. However, cognate forms are used to examine a phonological phenomenon. They have a common origin by definition. And a sound can change into another sound in separated areas independently. A well-known example among Japanese dialectologists is that the proto *ai rests intact in the central area of the main island of Japan, while it changed into [e:, ɛ:, æ:] etc. in peripheral areas as seen in map 4 cited from Sato (2002: 314).



Map 4. Reflections of *ai in Japanese dialects

4 Merging Types of Proto-Phonemes

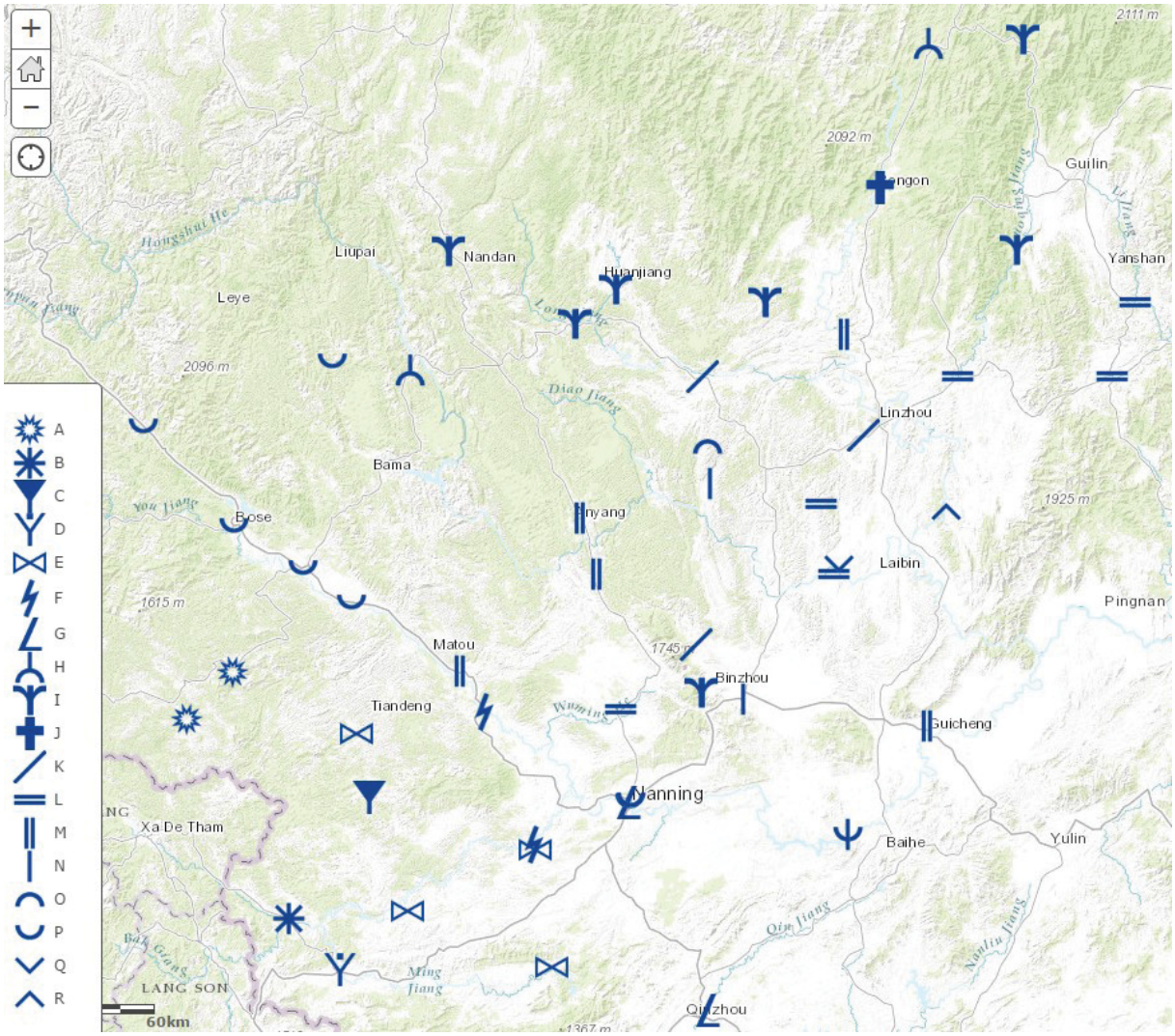
However, according to comparative linguistics, there is a principle concerning sound categories: given that two (or more) sound categories in a dialect correspond only to one caterory in another dialect, if there is no conditioning factor (e.g. complementary distribution) between these categories, then a merger occurred in the later dialect. Endo (2015) displayed geographical distributions and relative chronology of tone category types in Tai-Kadai using this method, and extracted directionalities of tone value change. Here, we apply this point of view to the / r / type initial consonants in Zhuang. Table 2 shows the various reflection types regarding merging patterns of proto categories. Types A to G belong to the Southern Group, while types H to R belong to the Northern Group. Proto categories *hr, *thr, *khr occur in upper tone categories which derived from voiceless initials, the others occur in lower tone categories which derived from voiced initials.

Types	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**
***hr**	h	h	h	h	h	hl	l	j	j	j	hj	ɣ	r	ð	z	l	j	ɣ, h, hj
***thr**	th				th	th	th											
***khr**	khj	khj		khj	l	hl	l											
***r**	l	l	l	l		l		l	r	θ								
***nr**	n	n	n	n	n	n	n											
***ɣ**	r	ɬ	ɬ	l	l	hl	l											
***ŋr**	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	n											
Places	50 51	47	48	46	43 44 45 49	41 42	39 40	10 15	6-8 11- 14	9	1 3 18	19- 22 24 25 30	4 16 17 23 31	5 26 27	36	28 29 32- 35 37 38	2	21

Table 2. Reflecting types of proto-phonemes among Zhuang dialects

Essentially the same table is already given in Yuan (1963: 190), in which the proto categories with rounded medial u̯ are omitted as well. There are some amendments: 1) *ʔr by Yuan (1963) is replaced by *hr, following the reconstruction by Li (1977: 148-151) and Liang & Zhang (1996: 101). The reason for the reconstruction of *ʔr opposing to *r is the fact that these morphemes are read in upper tone classes, denoting the voiceless origin of the initial consonants. Although a three-fold distinction consisting of glottalized, voiceless and voiced sonorants exists in the Kadai branch, voiceless sonorants commonly exist in the Tai branch as indicated in traditional orthographies in Siamese, Laos, and so on. Moreover, *hr or *r̥ is far more easier to explain the changing process to the modern descendants like [h], [hl], and [hj]. 2) Types A to R are distinguished and given names by this paper. 3) Correction of misprinting and errors in writing: Type H was written as seen in place 5, but actually it should be 15, and 10 should be also added; Type I was written as seen in places 8 and 13, but actually places 6, 7, 11, 12, 14 should belong to this type instead of the original Type M. 4) Type R is separated from original Type L, since there are many cases of other reflections as [h] and [hj] in this place.

The geographical distributions of these types are displayed in map 5 below.



Map 5. Geographical distribution of merging types concerning / r / type sounds in Zhuang

5 Chronological Order of Each Type based on Merging Process

Now we trace the merging process from the proto system to each type step by step. This process is displayed on the map comparing geographical distribution. Finally, directionalities of sound value changes will be extracted from this process.

5.1 Changing Process in the Southern Group

Type A distinguishes all 7 categories in the proto system; therefore it should be the oldest. As for the changes of sound values, *hr > h and *thr > th are a simple drop of medial *r. In *khr > khj, medial *r changed into j. This is a frequently observed pattern of sound change. *r > l is also a well-known type of change, for example, Bangkok or Southern district Siamese r- initial is pronounced as l in Isan Siamese or Lao. As a result of this change, the distinction between *r and *l in Proto-Zhuang coalesced into single l-, while this distinction was kept in the Northern Group. By and large, the Southern Group is conservative in terms of protection of proto distinctions regarding / r / type sounds, while in this regard the Northern Group kept the older distinction of categories. *nr > n and *ŋr > ŋ are again a simple drop of medial *r. Regarding *ɣ > r, the exact phonetic nature of *ɣ is also a problem, and no further comment is given here.

Types	A	B	C	D	E	F	G
*hr	h	h	h	h	h	hl	l
*thr	th				th	th	th
*khr	khj	khj		khj	l	hl	l
*r	l	l	l	l		l	
*nr	n	n	n	n	n	n	n
*ɣ	r	ɬ	ɬ	l	l	hl	l
*ŋr	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	n

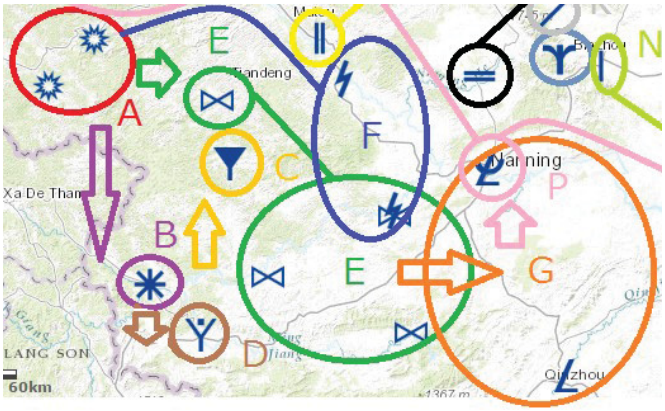


Table 3. and Map 6. Changing process in the Southern Group

The change from type A to type B is accomplished only by one step: th > h, and in type C, the further change khj > h also occurred. These changes are the loss of closure between tongue and palate. The nature of change between [r] in type A and [ɬ] in types B & C for *ɣ is uncertain.

Type D still keeps [khj], so it is relatively older than type C. At the same time, it has a shared innovation from proto *hr and *thr to [h] with type B. Therefore, type D came from type B, and the distinction between *r and *ɣ merged into [l].

Type E has kept *th, so it is older than types B, C, and D, and it should be derived directly from type A in this regard. However, type E has a common innovation with type D regarding the merger of *r and *ɣ. Besides, type E further experienced the merger of *khr into *r and *ɣ.

Type F has a shared innovation with type E regarding the merger of *khr and *ɣ. However, it keeps the distinction between *khr and *r; hence it couldn't be derived from type E. Also, type F keeps proto *thr category; therefore it couldn't be derived from type B, C, or D, but from type A or the proto type. An innovation from *hr to [hl] occurred and coalesced with *khr and *ɣ. The sound value [hl] for *hr and *khr is also noteworthy. The first element [h] may be regarded as the retention of the first element of proto initials *hr and *khr.

Type G is similar to type F regarding the merger of *hr, *khr, and *ɣ, but the actual sound is [l], not [hl] of type F. In this regard, type G is the same as type E, as well as the merger of *r to [l]. Moreover, proto *ŋr merged to the descendant of *nr, and became [n].

The whole changing process can be diagrammed as in table 4 below:

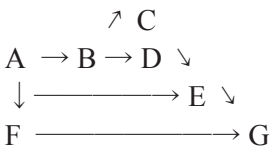
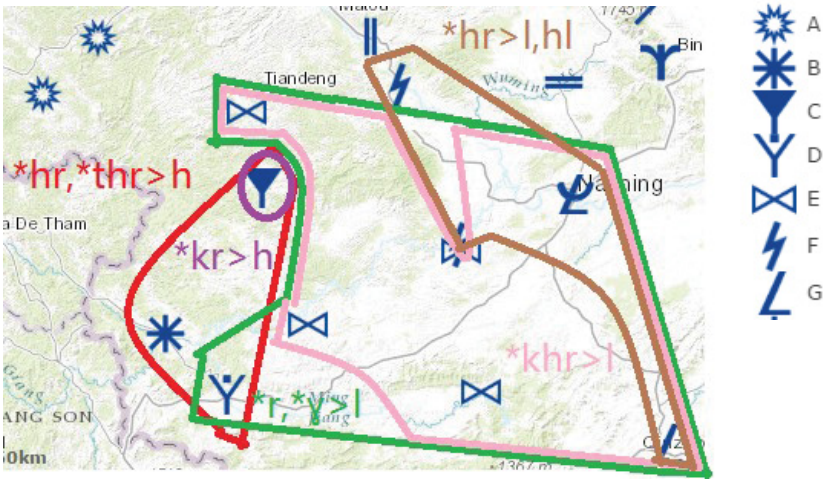


Table 4.
Chronological order of merger types in the Southern group



Map 7.
Shared innovations
in the Southern group

Shared innovations in the Southern group are shown in map 7. Here, the district where *kr > h is included with that of *hr, *thr > h; while *khr > l in *r, *ɣ > l, respectively. This relationship is comparable to a subset in a Venn diagram. In such a situation, the change with a bigger area occurred earlier than that with a smaller area, given that these changes have close relationships, especially in cases of a push chain or drag chain.

5.2 Changing Process in the Northern Group

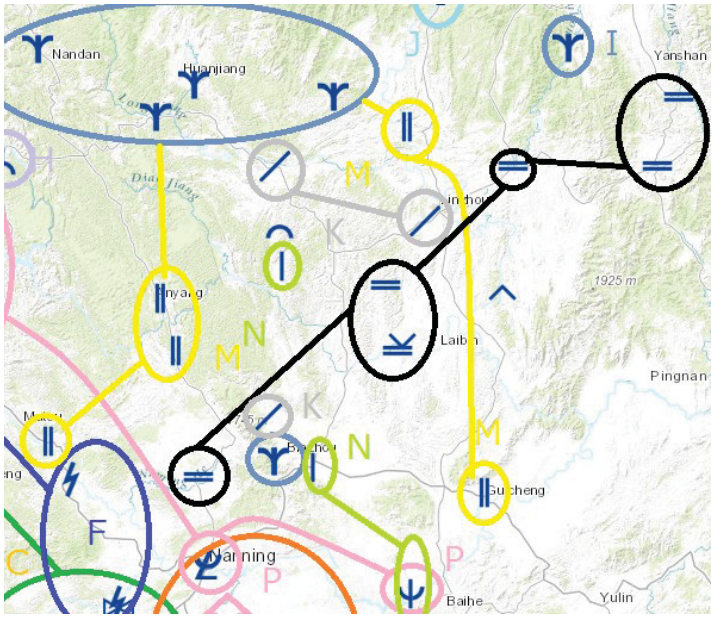
In the Northern group, there are two subgroups: types H, I, and J show bipartition in which the upper series consonants merged to [j] and the lower series consonants merged together to become [l, r, or θ] respectively; types K to R display a consistent reflection regarding phonetic diversity [hj, ɣ, r, ð, z, l, and j] as seen in table 5.

Types	C	G	H	I	J	K	L	M	N	O	P	Q	R
***hr**	h	[l]	j	j	j	hj	ɣ	r	ð	z	l	j	ɣ, h, hj
***thr**		th											
***khr**		[l]											
***r**	l		l	r	θ								
***nr**	n	n											
***ɣ**	ɬ	[l]											
***ŋr**	ŋ	n											

Table 5. Types in the Northern group and type C & G in the Southern group

The former subgroup is mainly located in the Northern Guangxi area. This subgroup has a similarity with type C of the Southern group, with respect to the merger of upper consonants. Based on the extent of the merger, the former subgroup may be relatively earlier than the later subgroup.

Type P has a similarity with type G of Southern group, since the former type can emerge if [th] and [n] of the later changed to [l]. Geographically speaking, Type P is located next to type G, especially in Nanning. Type P possibly dispersed from Nanning along the Youjiang river.



Map 8. Types in the Main Part of Northern group

Type M and type I are similar in that they have [r] in common. They are geographically adjacent and type M seems to be distributed in the peripheral area in the Northern group. Maybe type M is the oldest among the later subgroup.

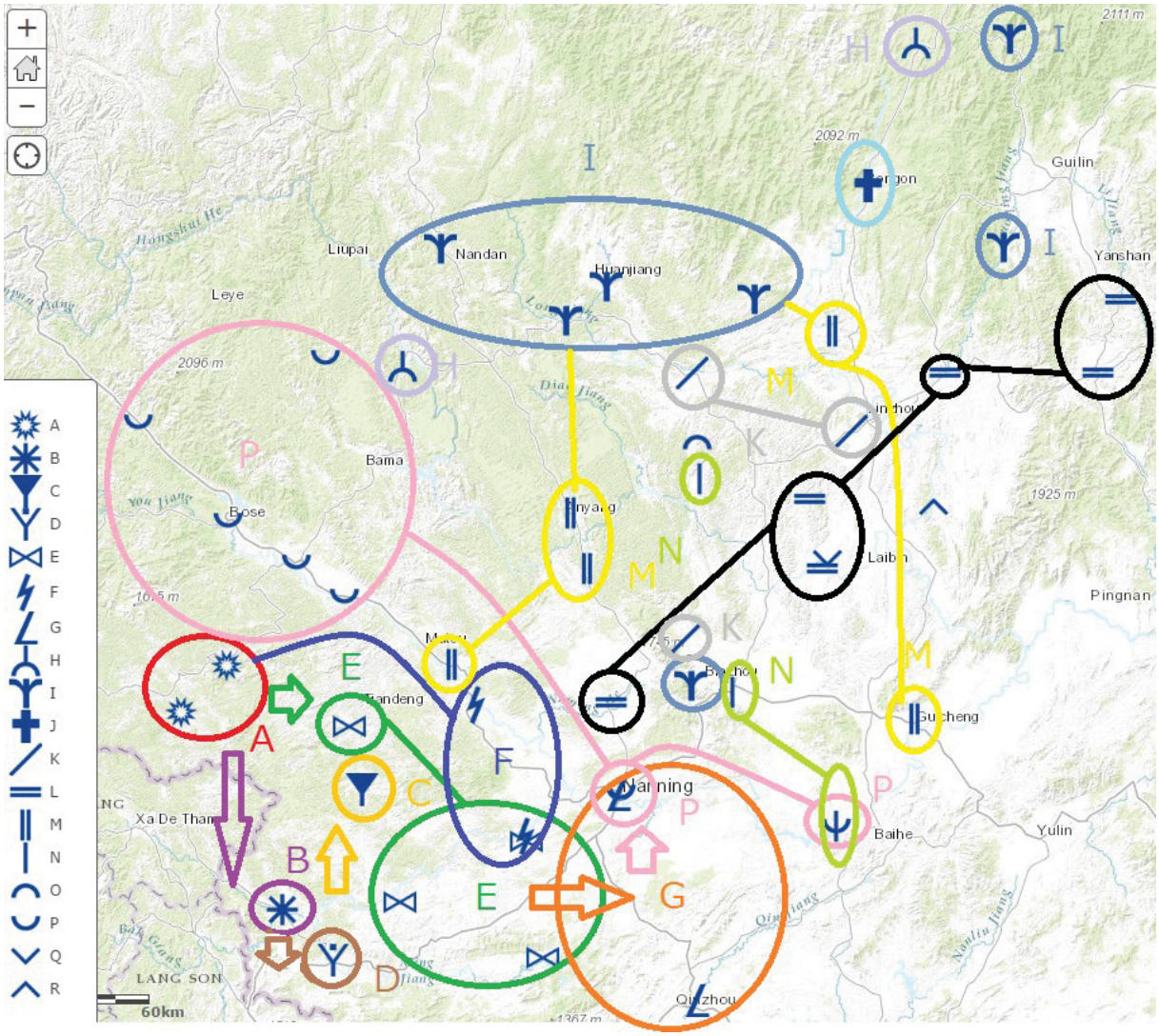
Type K with [hj] is located at the inner side of type M, so it derived from type M, that is to say, a change [r] > [hj] occurred.

Type L with [ɣ] seems to be the newest form in the Zhuang language, since it occupies the central area including Wuming, which is the standard variety of Zhuang.

Other than these, types N, O, Q are scattered, and experienced minor change. Type R is a mixed type.

6 Conclusion

The entire picture of merging types in Zhuang is seen in Map 9.



Map 9. The Whole Process of Change of / r / type sounds in Zhuang

Thanks to the preservation of older systems in the Southern group, a relatively solid changing process was traced in this region. Also, we can extract the directionalities of sound change from these concrete changes as follows:

1) *hr > h in types A, B, C, D, E
2) *hr > hl in type F
3) *hr > l in type G
4) *thr > th in types A, E, F, G
5) th > h in types B, C, D
6) *khr > khj in types A, B, D
7) khj > h in type C
8) khj > hl in type F
9) hl > l in types E, G
10) *r > l in types A, B, C, D, E, F, G
11) *nr > n in types A, B, C, D, E, F, G

12) *ɣ > r in type A
13) *ɣ > l in types B, C, D, E, G
14) *ɣ > hl in type F
15) *ŋr > ŋ in types A, B, C, D, E, F
16) ŋ > n in type G
17) *hr, *thr, *khr > j in type H, I, J
18) th > l in type P
19) n > l in type P
20) r > hj in type K
21) r > ɣ in type L
22) r > z in type O
23) r > ð in type N
24) r > j in type Q

The changes from proto form to the modern forms are provisional, especially for *ɣ. Directionalities 18 to 24 in the Northern groups have less reliability.

In sum, a relatively safer interpretation of geographical distribution based on merging types of several categories can be given, although complicated situations are observed in individual maps.

Acknowledgements

This study was supported by the Grants-in-Aid for Scientific Research KAKENHI (C) from JSPS, 15K02525, 2015-2017. I appreciate Dr. Donna Erickson for correcting grammatical mistakes of my English.

References

Endo, M. 2015. Geographical Distribution of Tone in Tai-Kadai. *Keizai Kenkyu*, Aoyama Gakuin University, 7:1-35.

Li, F. 1977. *A Handbook of Comparative Tai*. Honolulu: The University Press of Hawaii.

Liang, M. & Zhang, J. 1996. *Dongtai Yuzu Gailun* [Introduction to Tai-Kadai]. Beijing: Zhongguo Shehui Kexue Press. [in Chinese]

Pan, W. 2010. Cong Dili Shishi Huanyuan Lishi Zhenshi [Historical Real Time Sound Shifts as Reconstructed from Geographical Apparent Time]. *Minzu Yuwen*, 2010(1):3-12. [in Chinese]

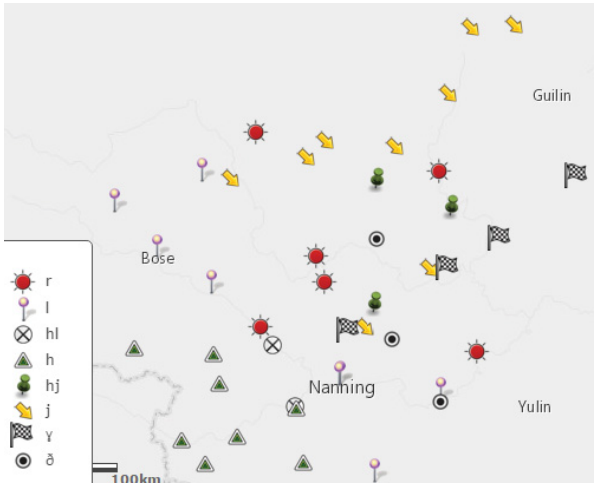
Sato, R. 2002. *Hogen no Chizucho* [Dialect Atlas]. Tokyo: Shogakukan. [in Japanese]

Wei, J. 2015. Yuyan Jiechu Shiyu xia Zhuangyu Liuyin *r de Tezheng Fenbu [Feature Distribution of Glide [r] in Zhuang Language in Perspective of Language Contact]. *Zhongyang Minzu Daxue Xuebao (Zhexue Shehui Kexue Ban)*, 42(4):136-141. [in Chinese]

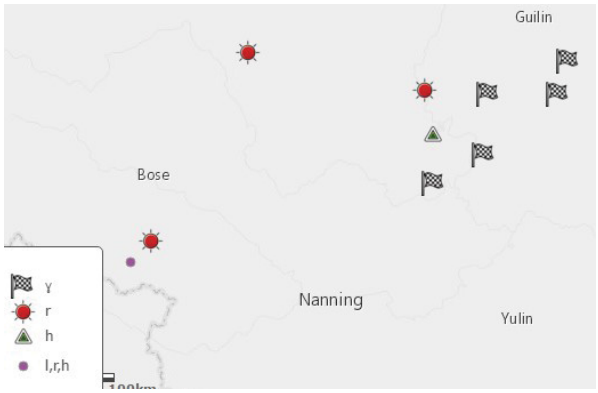
Yuan, J. 1963. Zhuangyu / r / de Fangyin Duiying [Sound Correspondences of / r / among the Zhuang Dialects]. *Yuyanxue Luncong*, 5:187-218; also In: Wang, F. & Sun H. eds.2010. *Yuan Jiahua Wenxuan* [Selected Works of Yuan Jiahua]. 155-178. Beijing: Beijing University Press. [in Chinese]

Zhang, J. et al. 1999. *Zhuangyu Fangyan Yanjiu* [Studies in Zhuang Dialectology], Chengdu: Sichuan Minzu Press. [in Chinese]

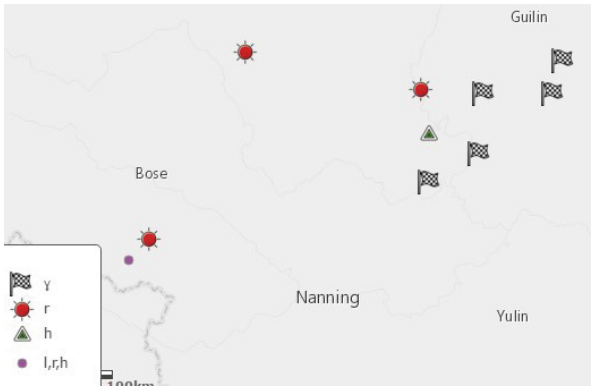
Appendix



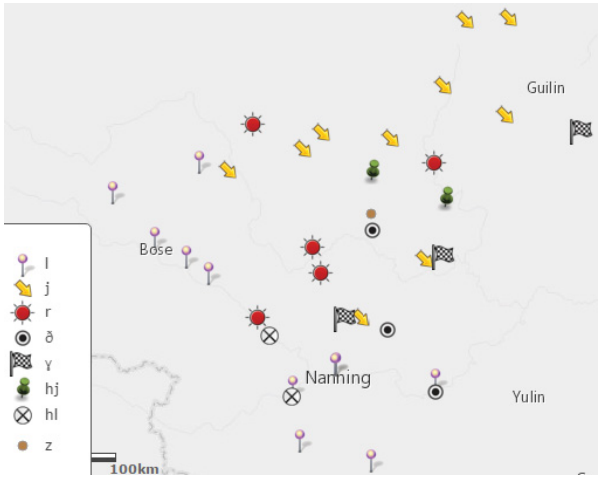
No. 1 “苋”菜



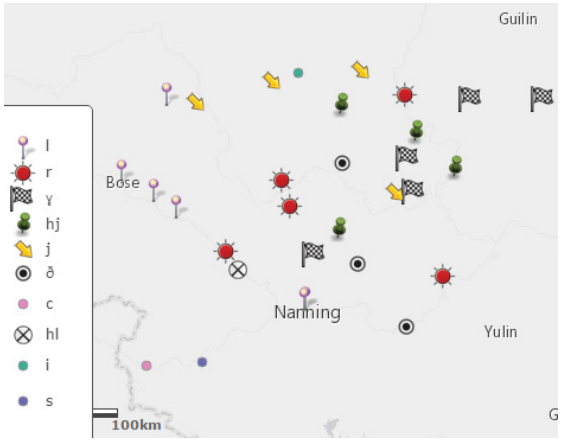
No. 2 丝瓜



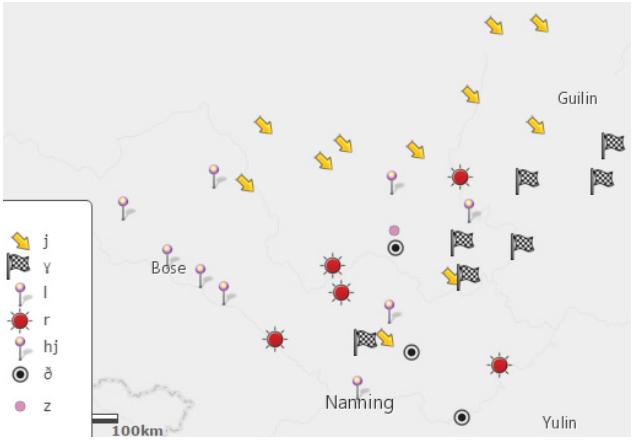
No. 3 牛“瘟”



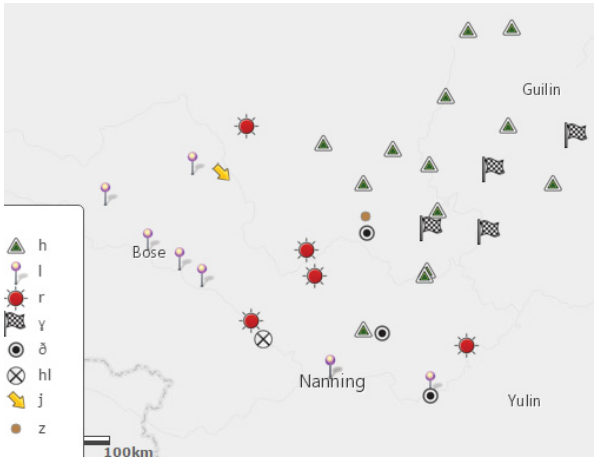
No. 4 臼



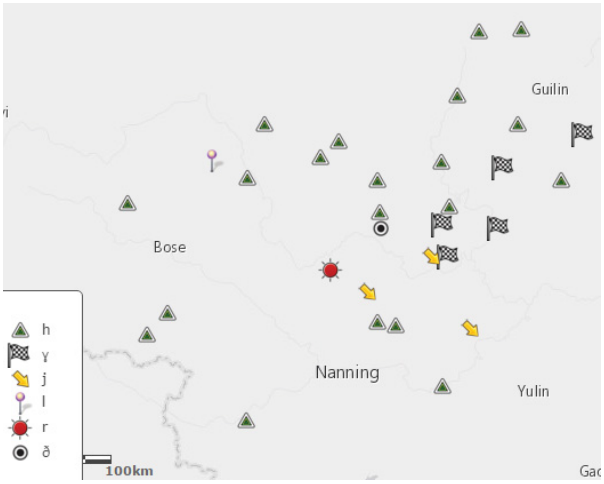
No. 5 “切”肉



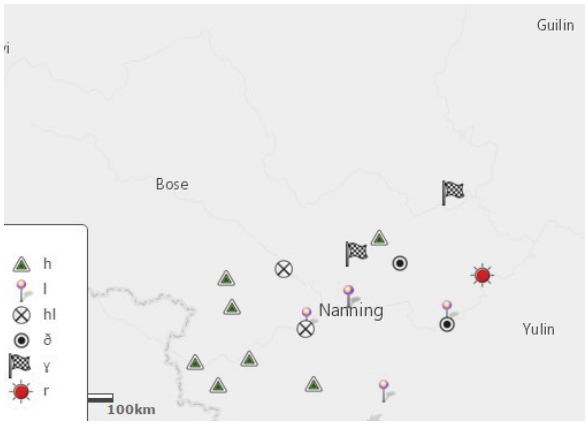
No. 6 蚊帐



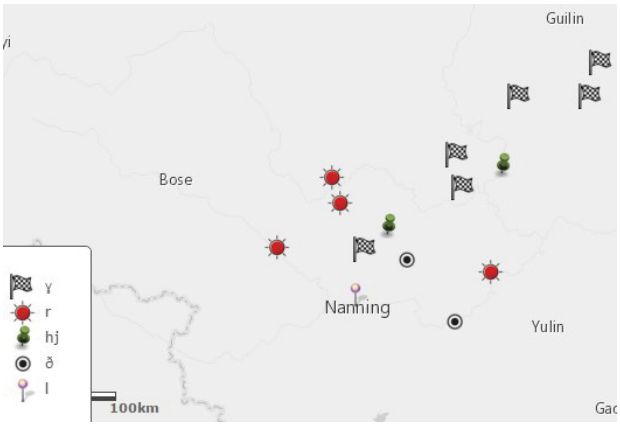
No. 7　鸡“冠”



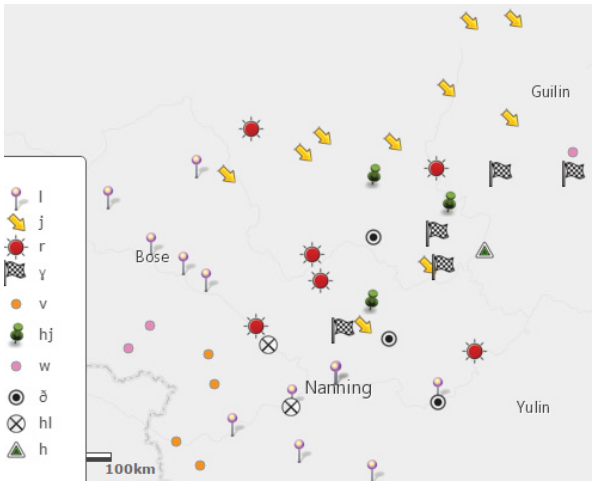
No. 8　慈菇



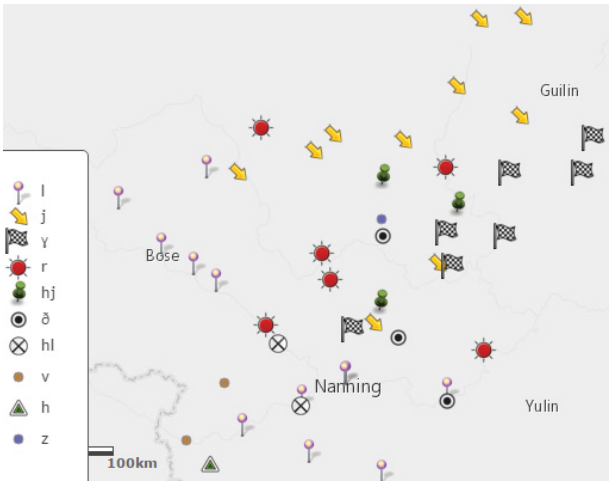
No. 9　炒菜“锅”



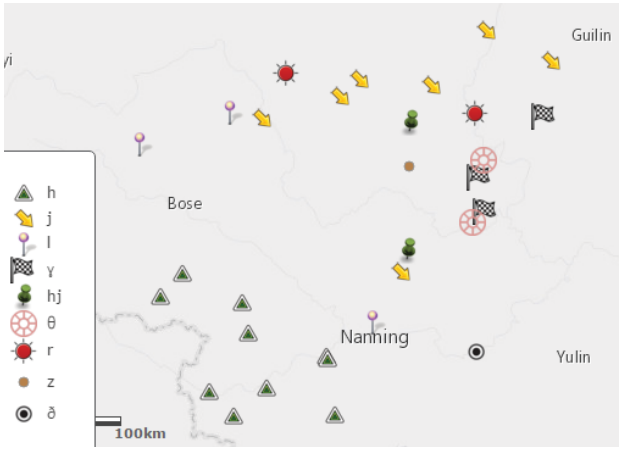
No. 10　痣



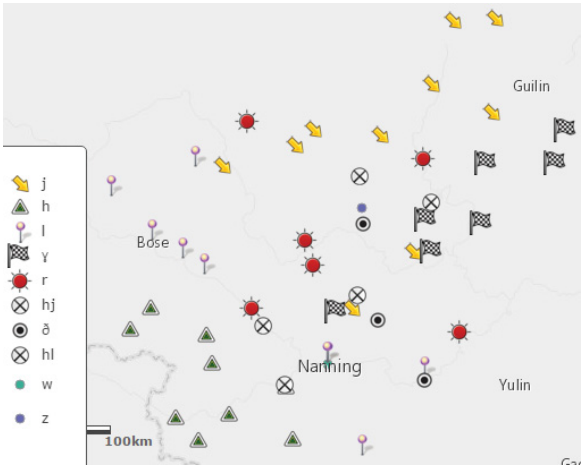
No.11　“梳”头



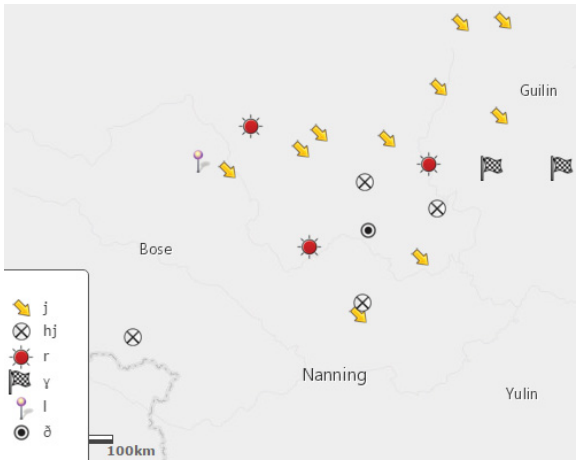
No. 12　菌子



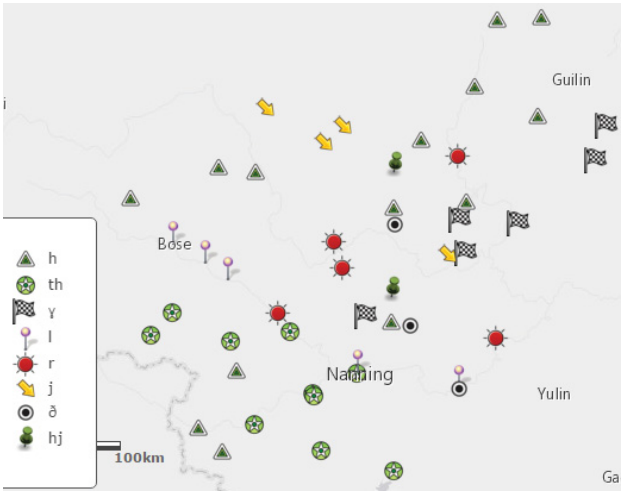
No. 13　打呵欠



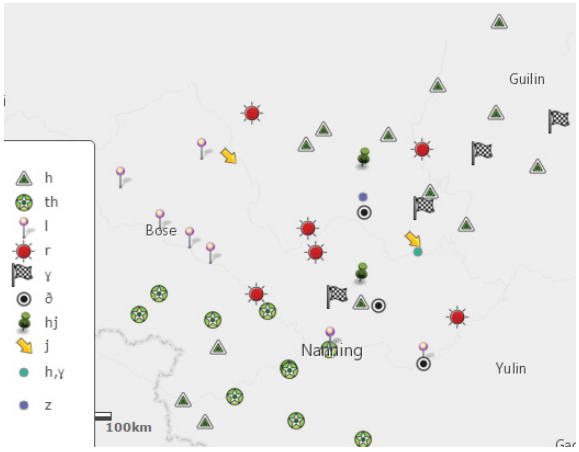
No.14　狗“叫”



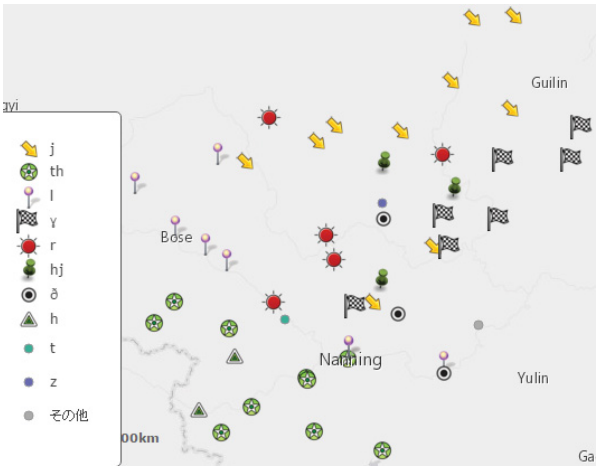
No. 15　“耙”田



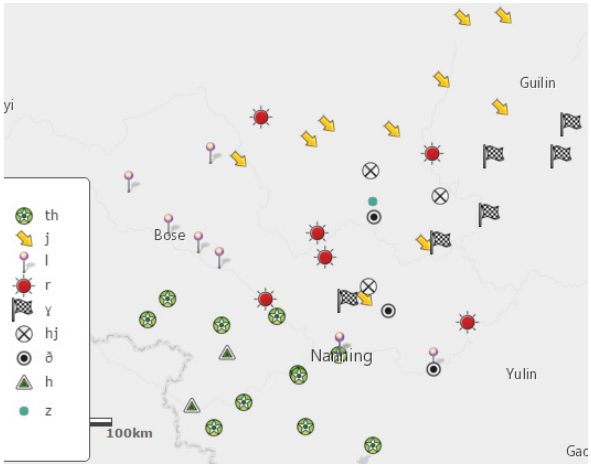
No. 16 石头



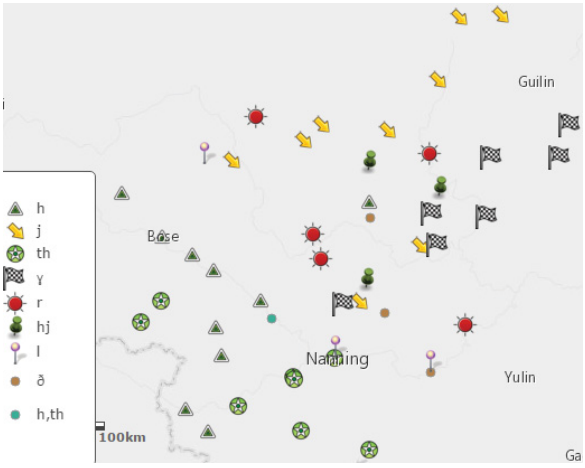
No. 17　尾巴



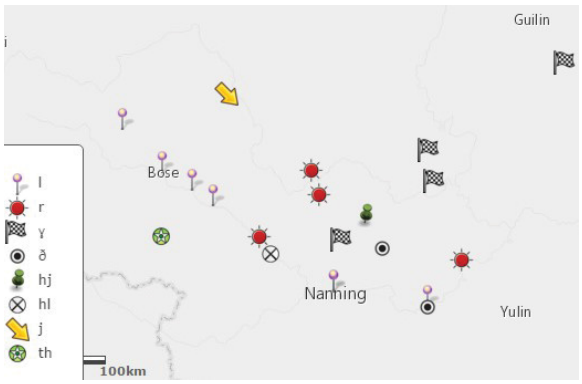
No. 18 头虱



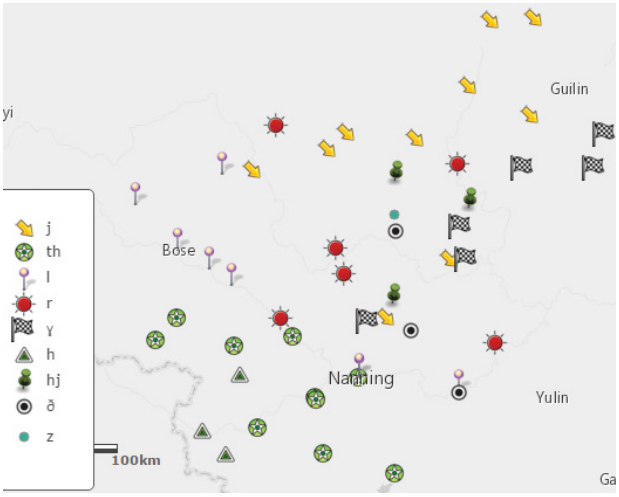
No. 19　暖和、温暖



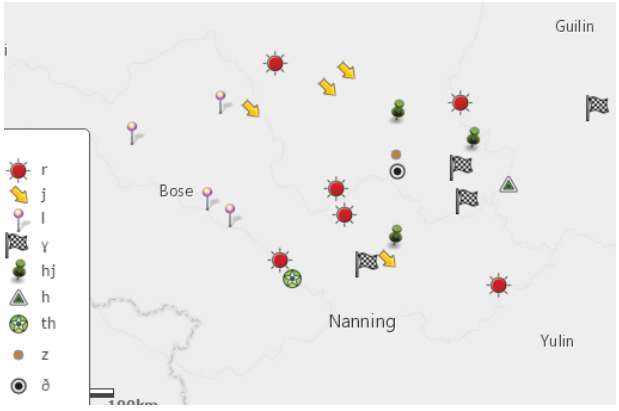
No. 20　看见



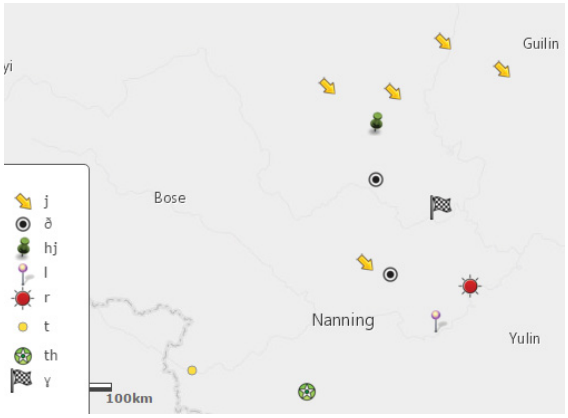
No.21　蜜蜂



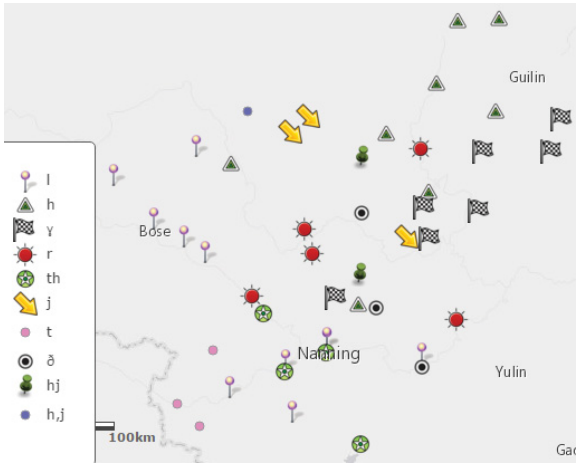
No. 22　“挑”水



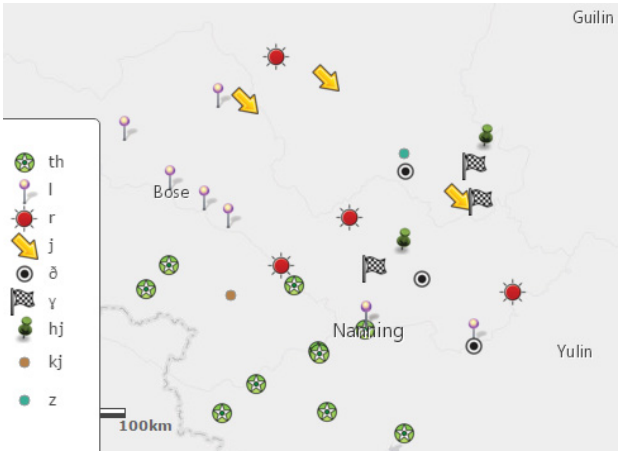
No.23　“敲”门



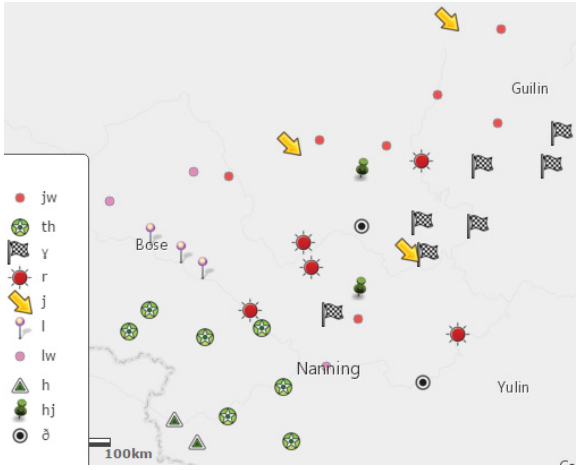
No. 24　斗笠



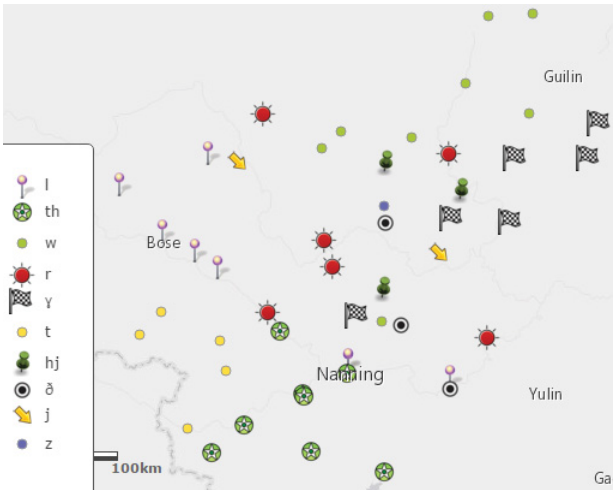
No. 25　水“满”了



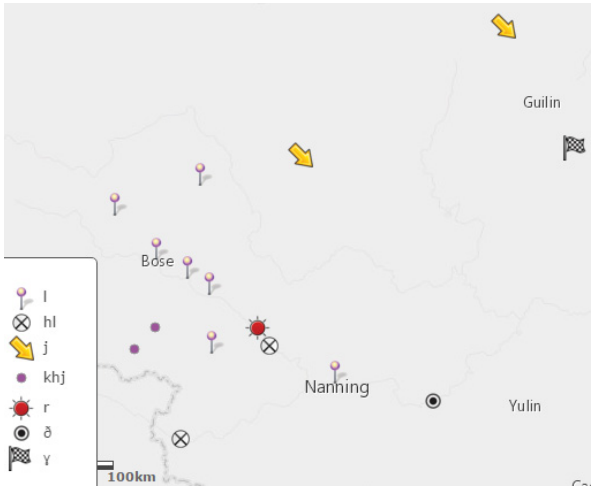
No. 26　“砍”树



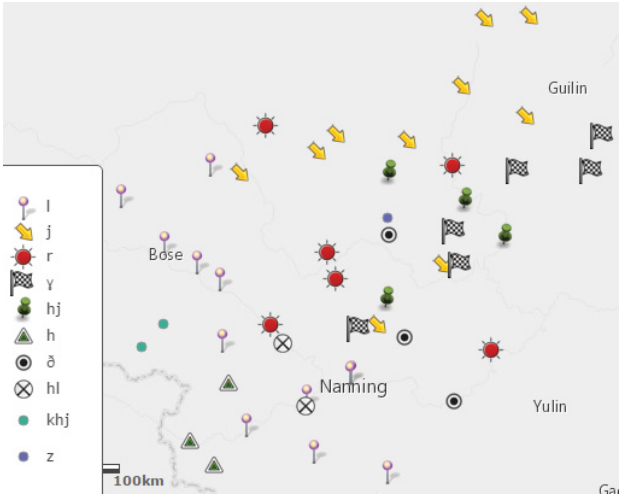
No. 27　抬



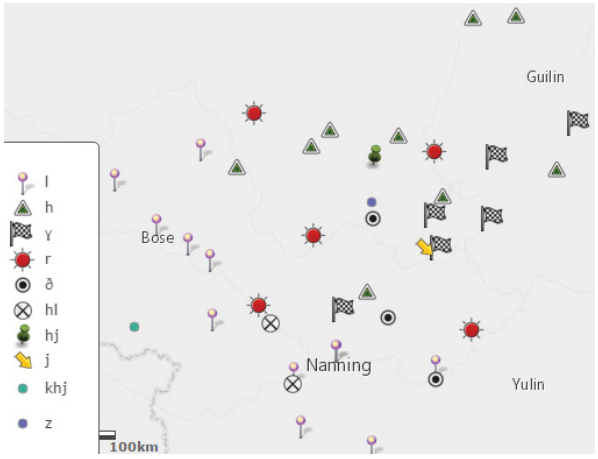
No. 28　屁



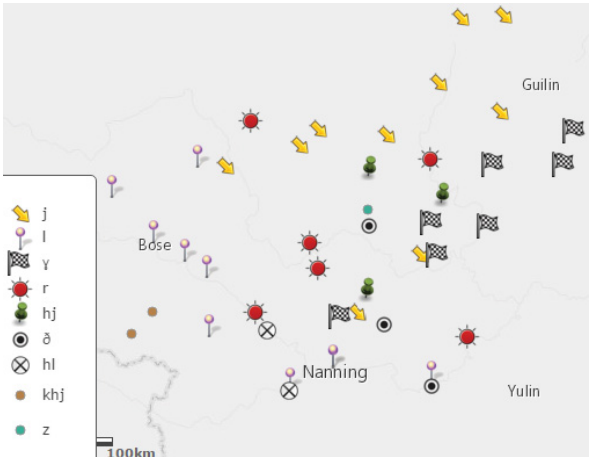
No. 29　蓝靛（植物）



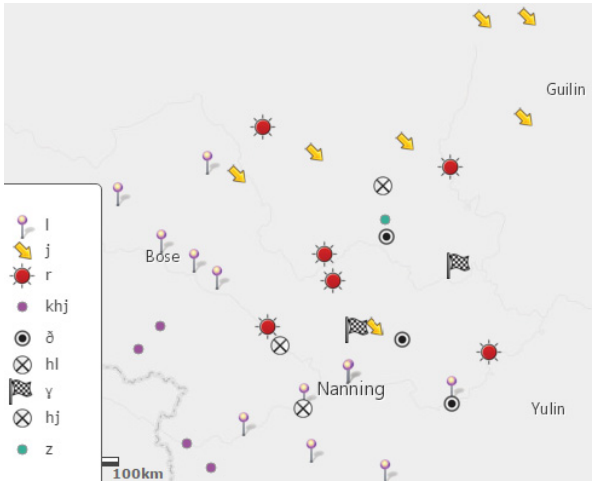
No. 30　六



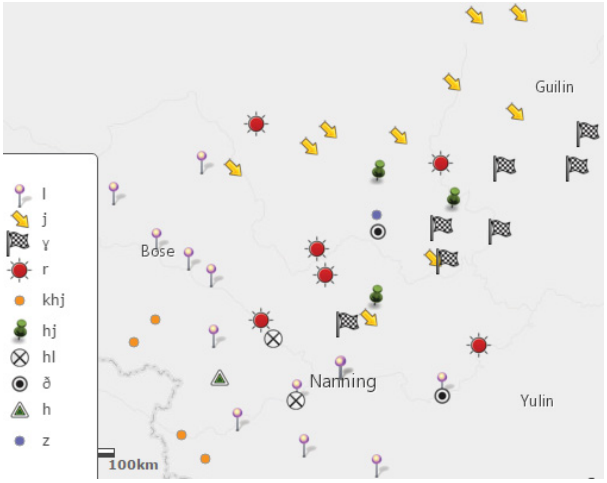
No. 31　碗“柜”



No. 32　寻找

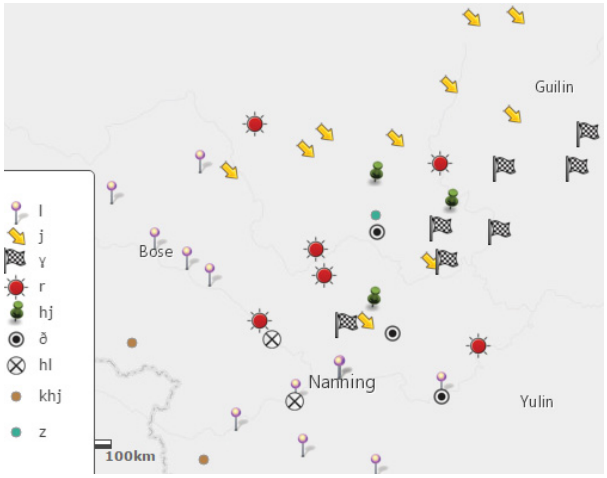


No.33　筛子

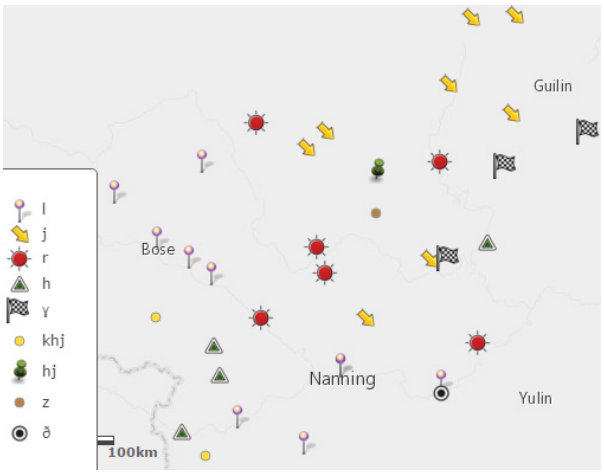


No. 34　“枫”树

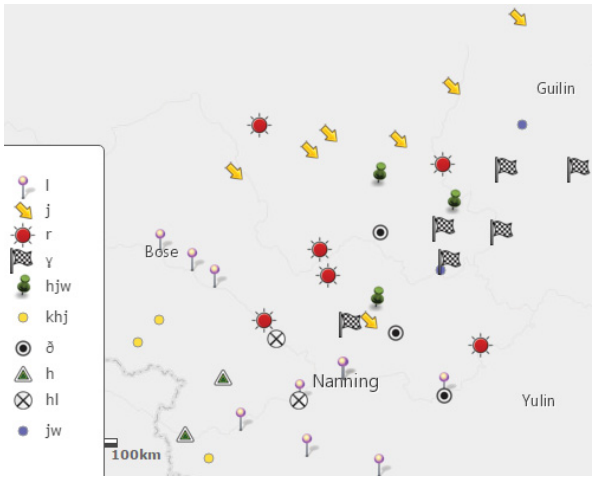
、



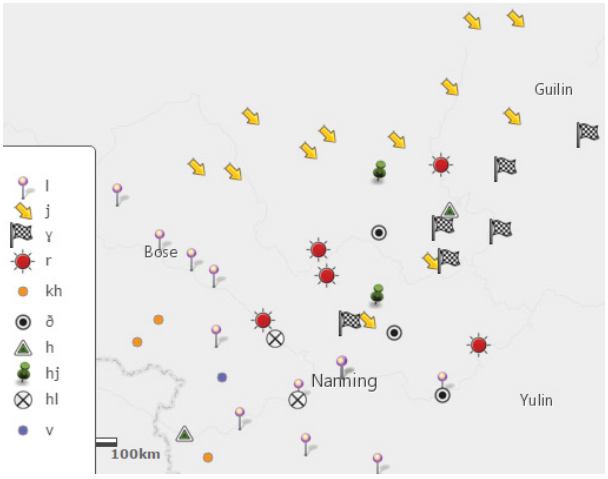
No. 35　斑鸠



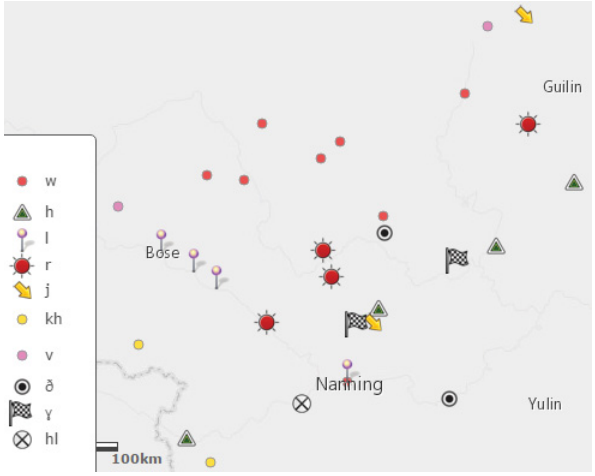
No. 36　可以携带的鸡“笼”



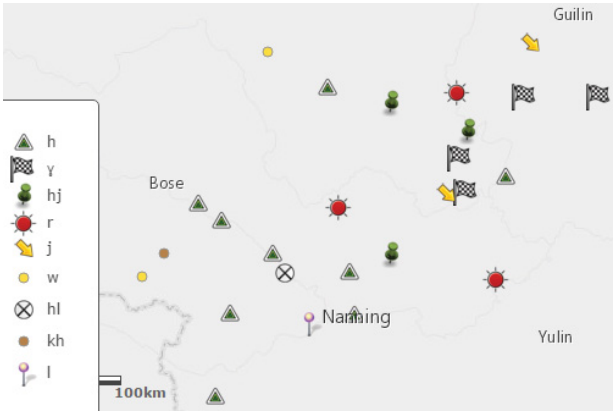
No. 37 “量”布



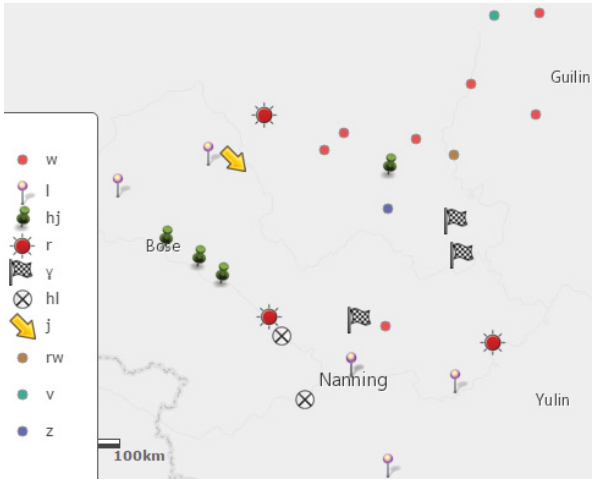
No. 38 笑



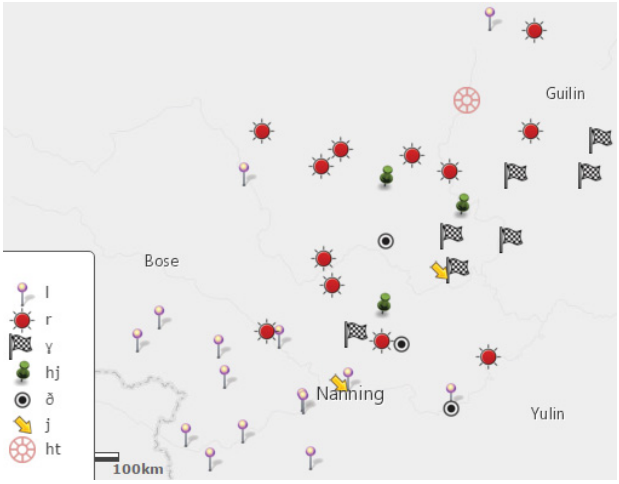
No. 39 溪（山涧）



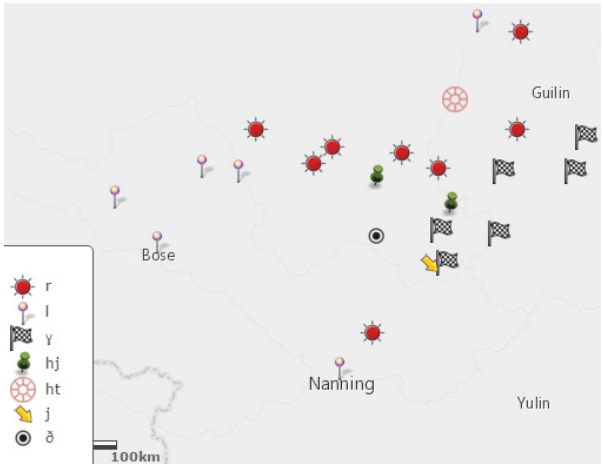
No. 40 一“围”一“抱”



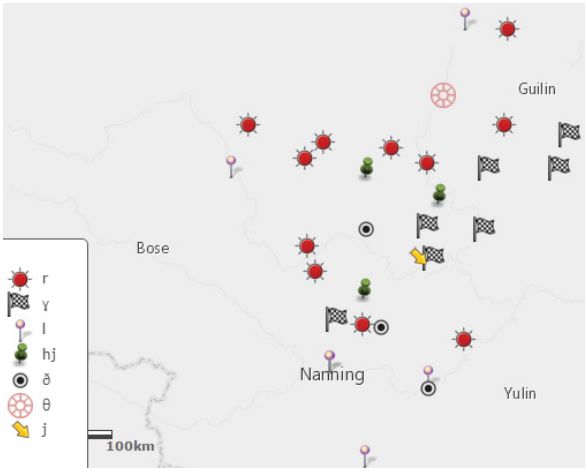
No.41 道路



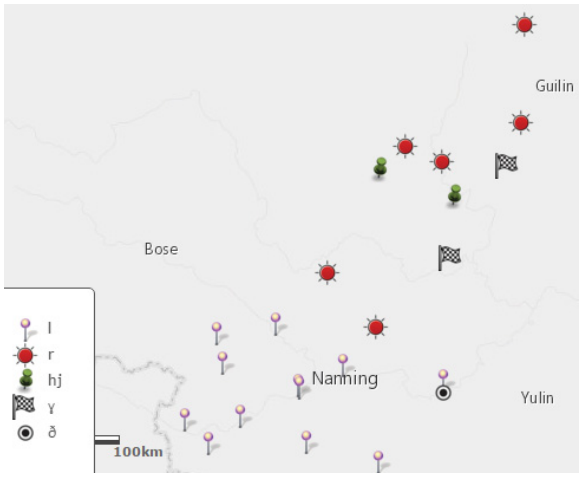
No. 42 “麻”脸



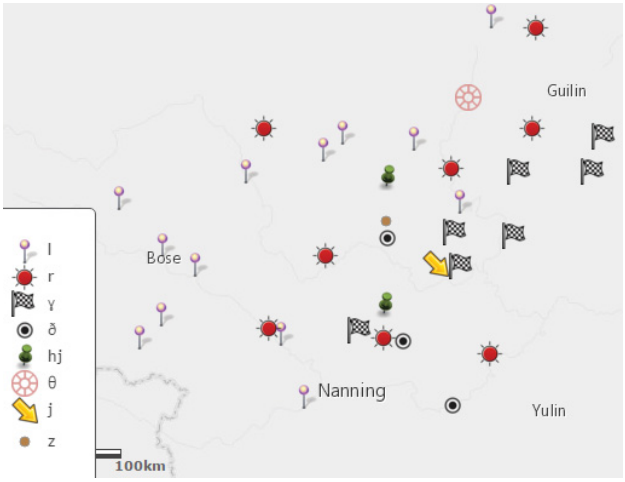
No. 43　野猪



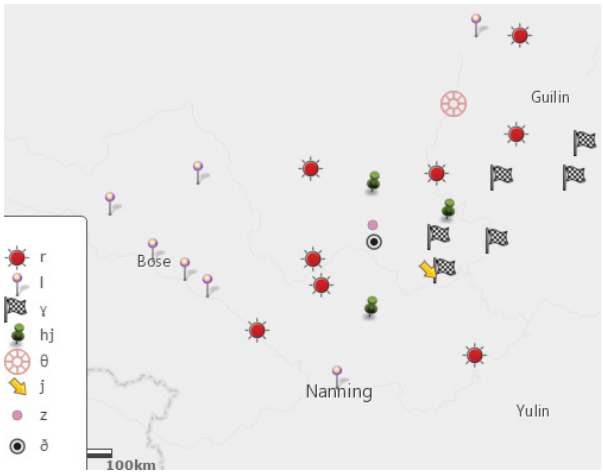
No. 44　刀“利”



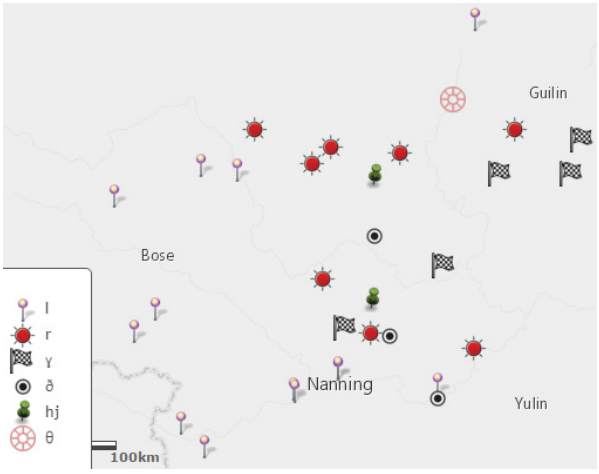
No. 45　大叶榕树



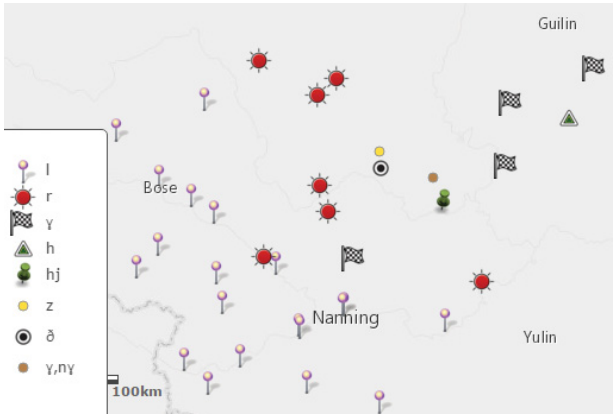
No. 46　路“滑”



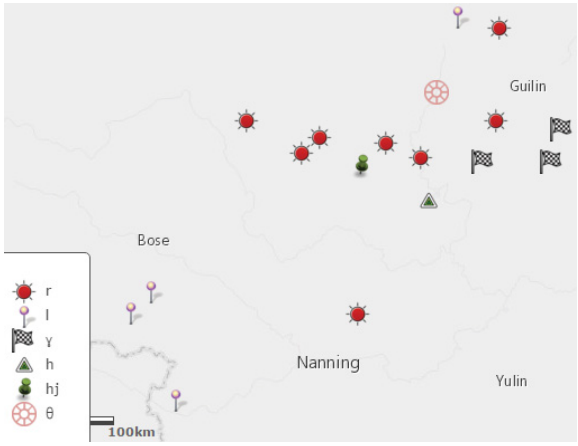
No. 47　“剪”布



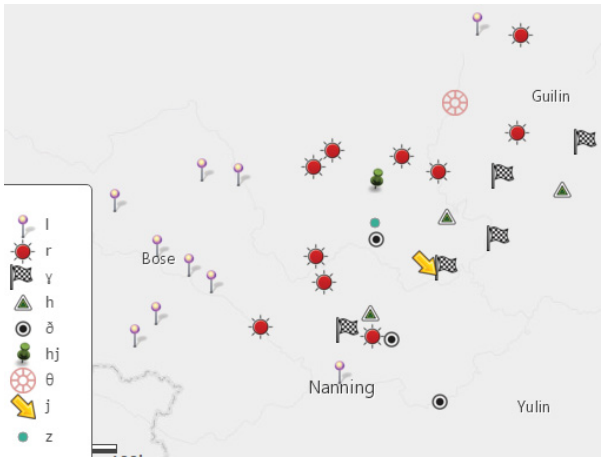
No. 48　“拉”车



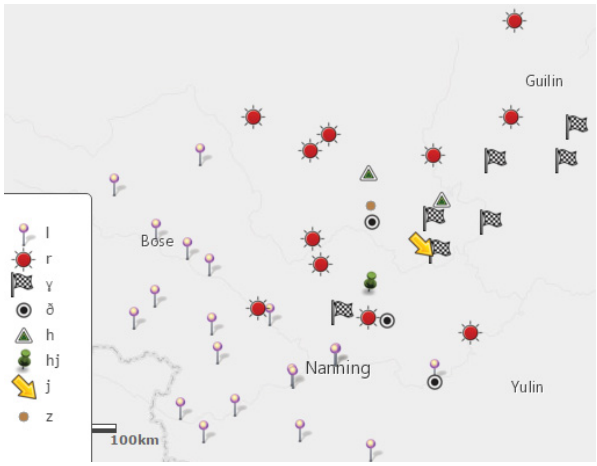
No. 49 “午”饭



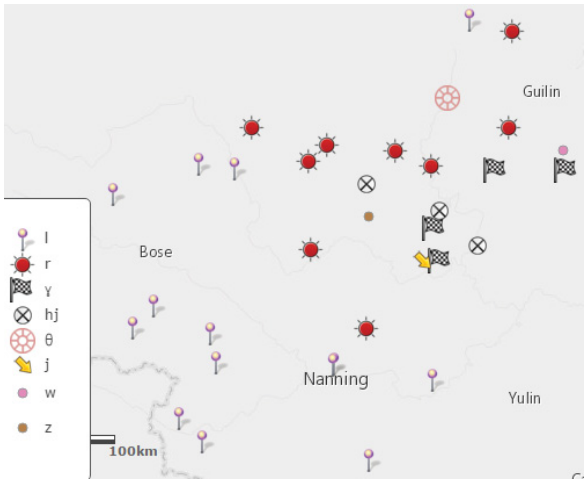
No. 50 爪子



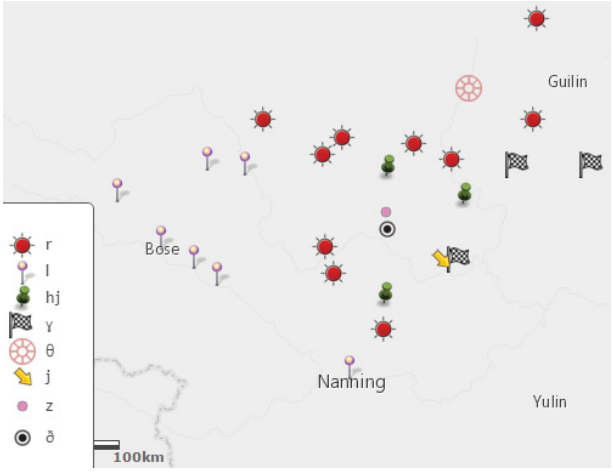
No. 51 萤火虫



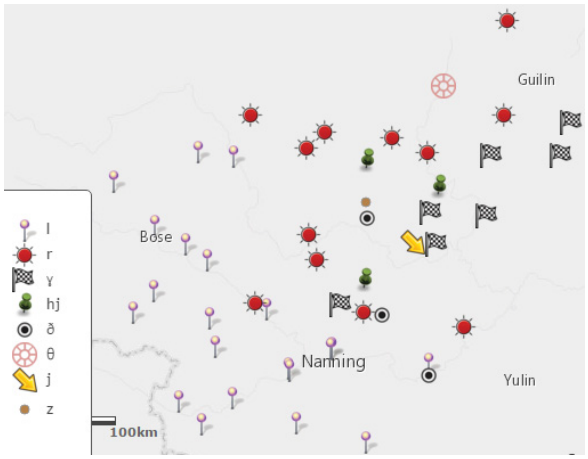
No. 52 天“旱”



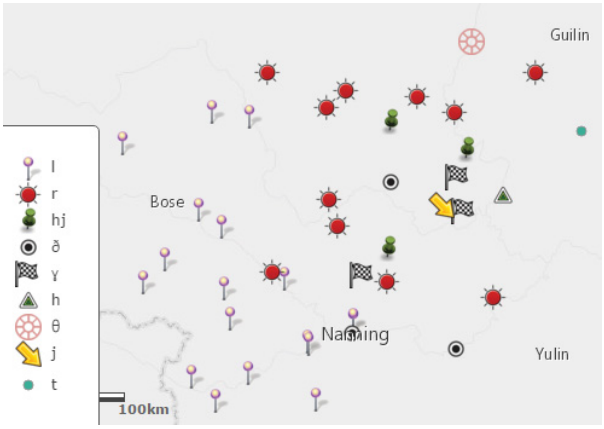
No. 53 鹰



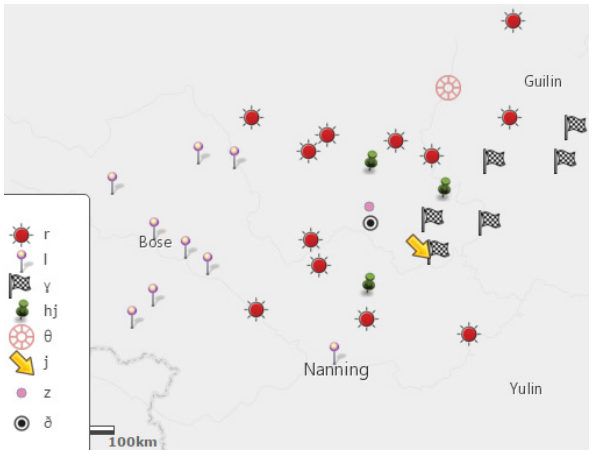
No. 54 早



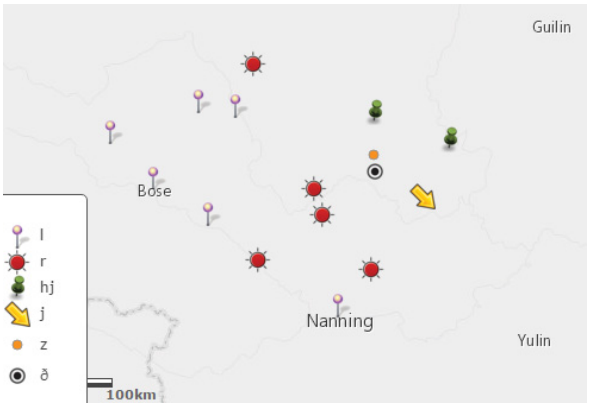
No.55　风



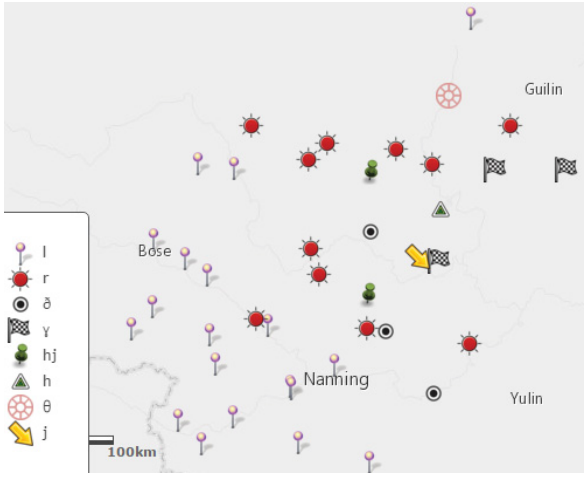
No. 56　㖞（洞场）



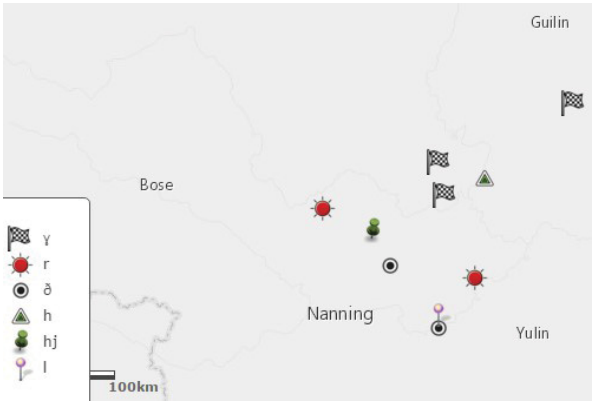
No. 57　“赎”回来



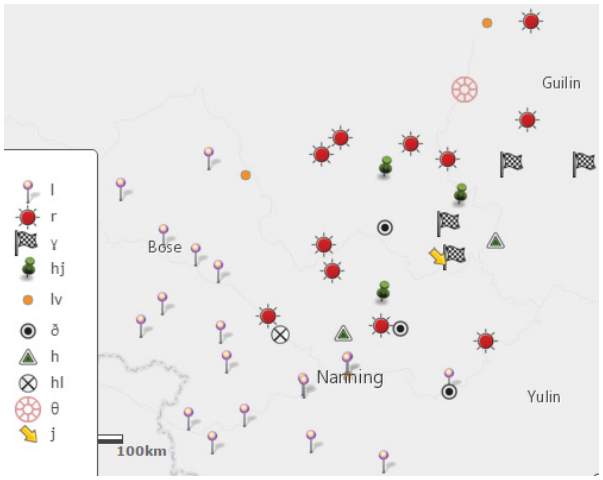
No. 58　“跟”我来



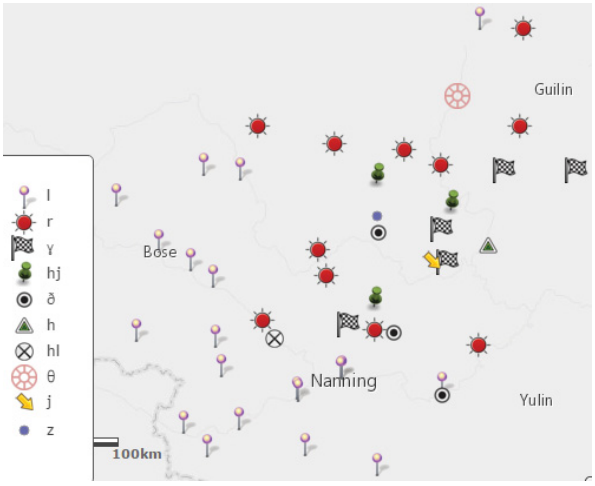
No. 59　牛“栏”



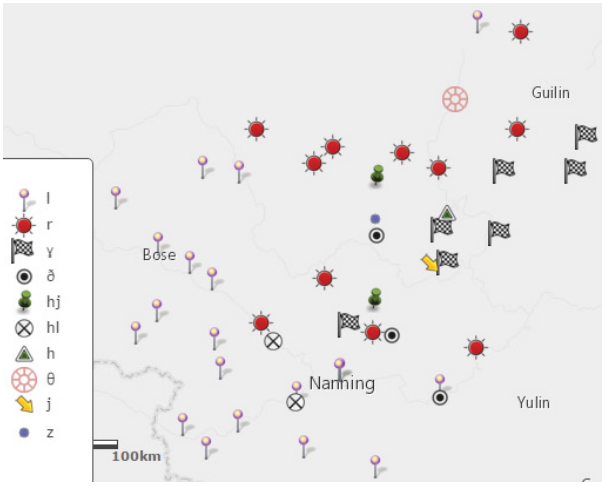
No. 60　“浇”水



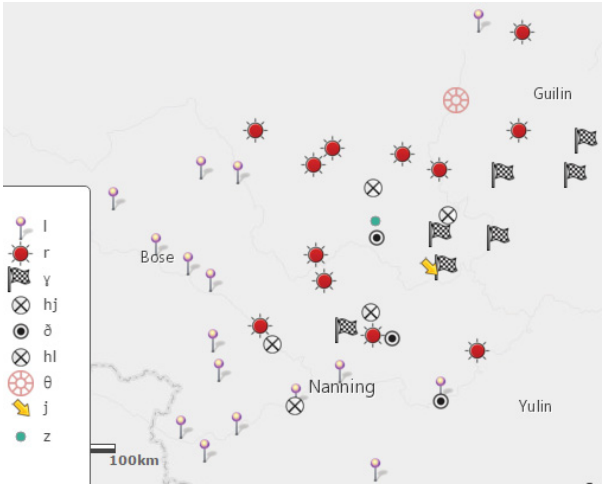
No. 61　船



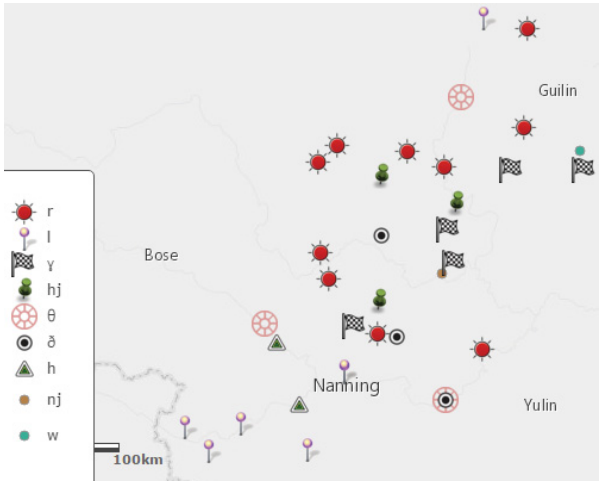
No. 62　“后”日



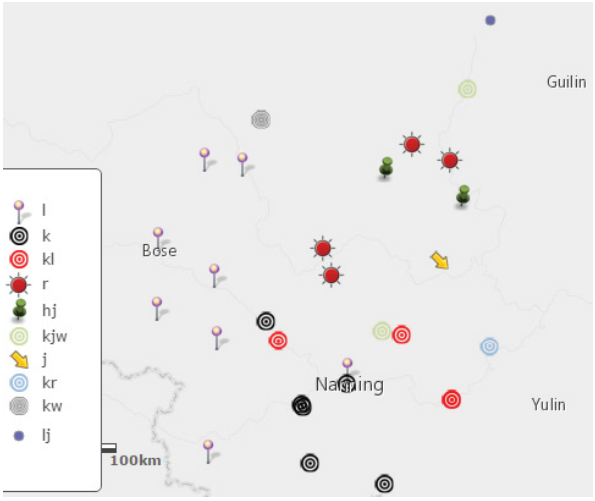
No. 63　臭虫



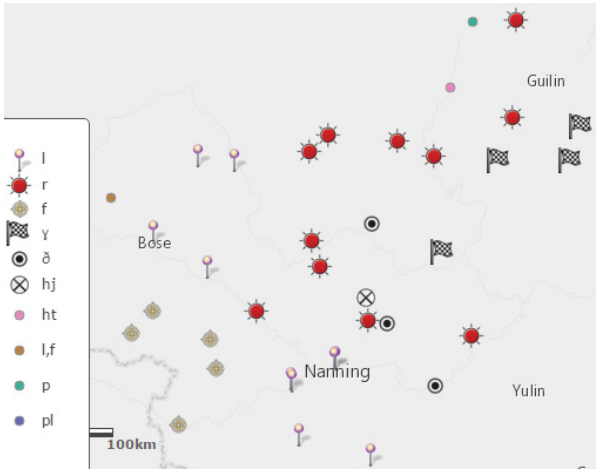
No. 64　树“根”



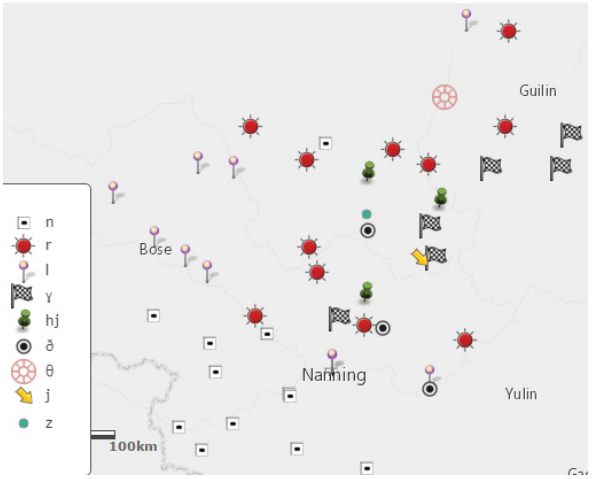
No. 65　蝌蚪



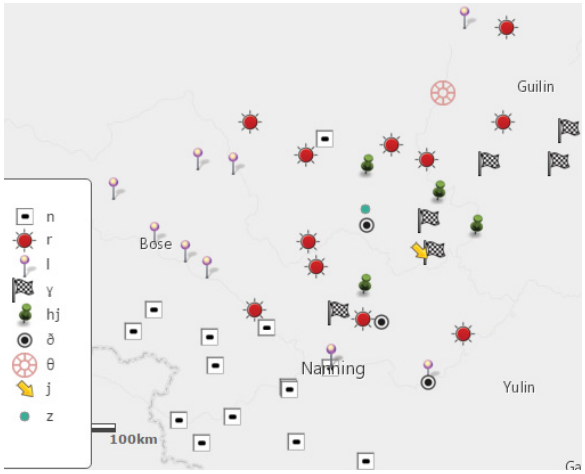
No. 66　石头从山上“滚”下来



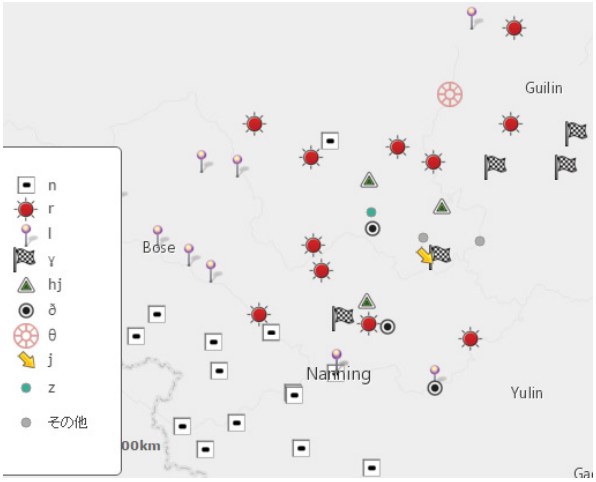
No. 67　水“溢”



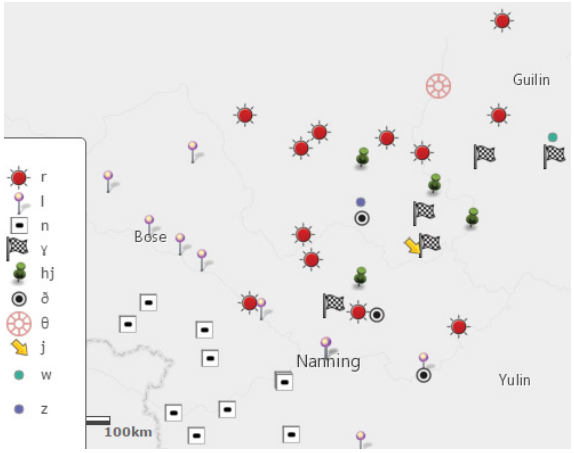
No. 68　水



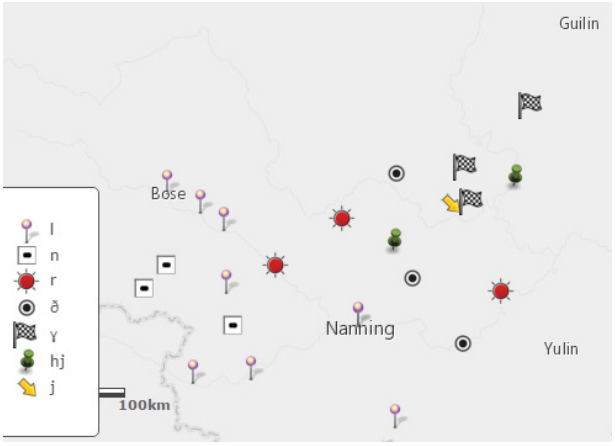
No. 69　鸟



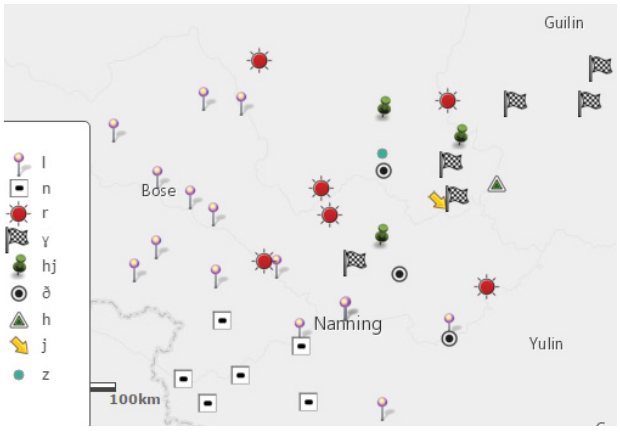
No. 70　“露”水



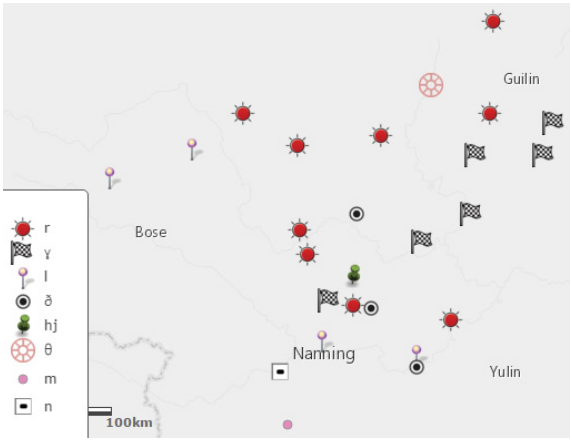
No. 71　“外”边



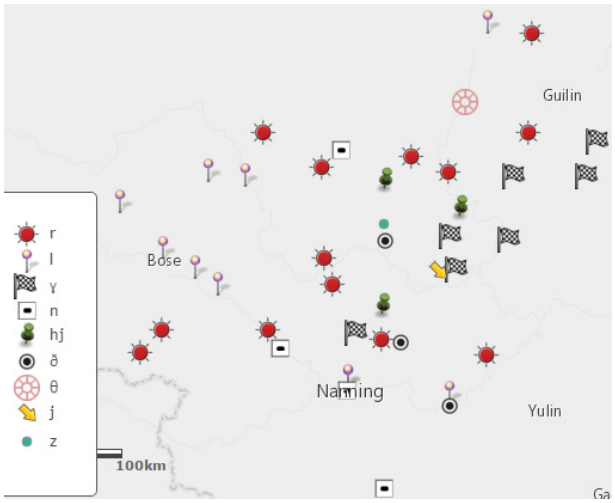
No. 72　“洪”水



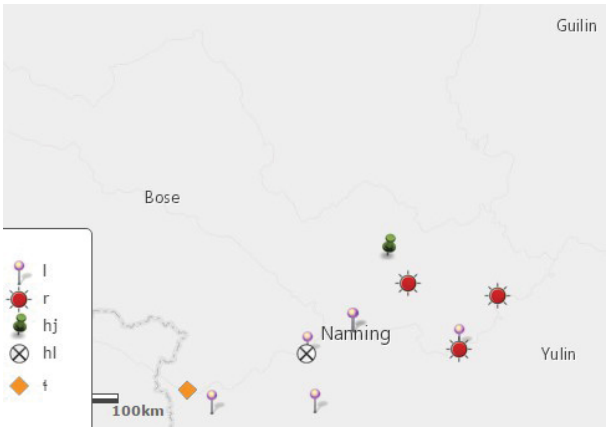
No. 73 “下”小牛，“下”山



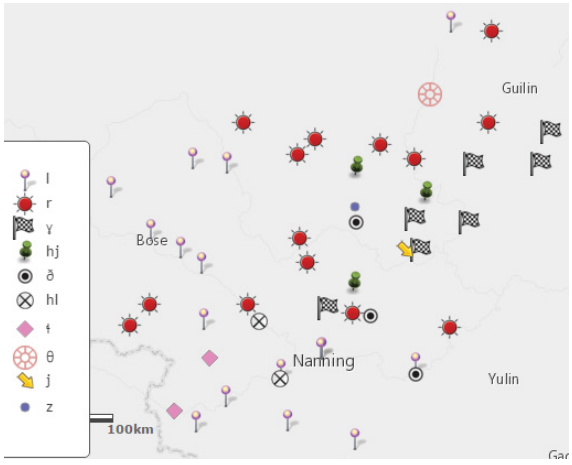
No. 74 “打”雷、雷“响”



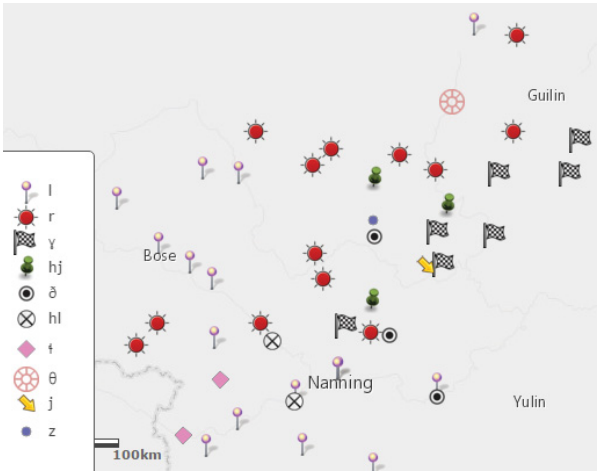
No. 75 笋



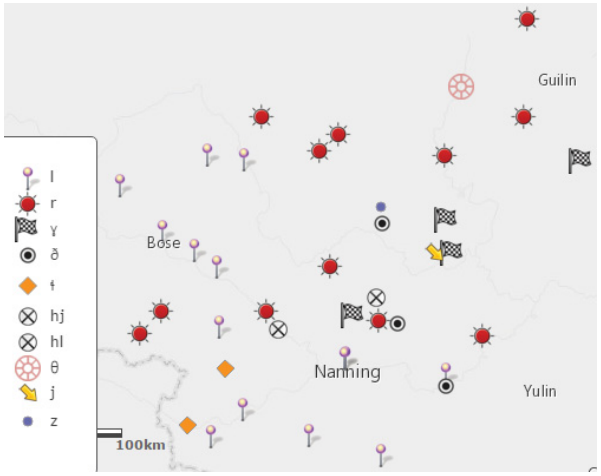
No. 76 小叶榕树



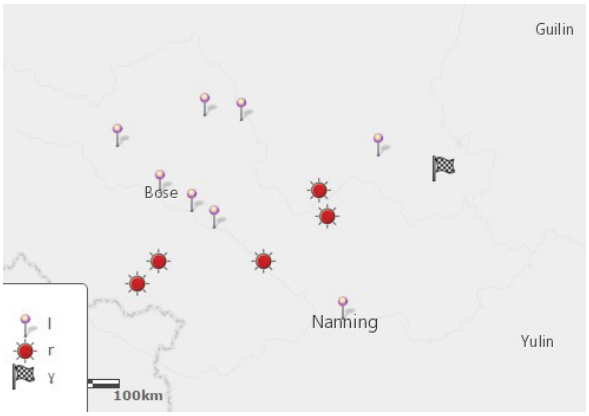
No. 77 细糠



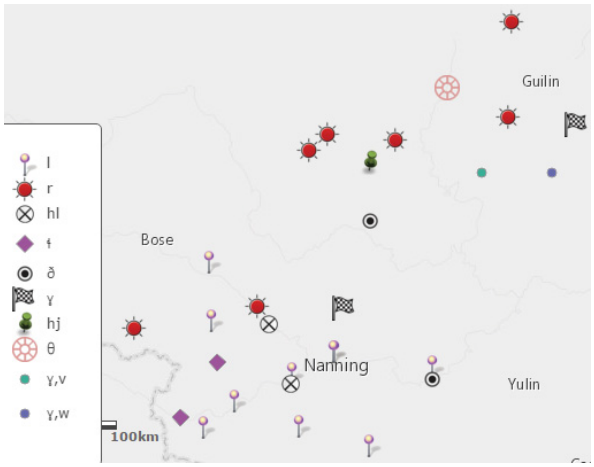
No. 78 “长”短



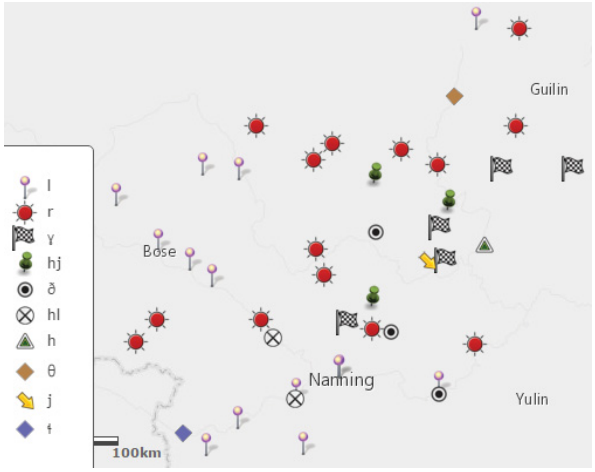
No. 79　气力，出“力”



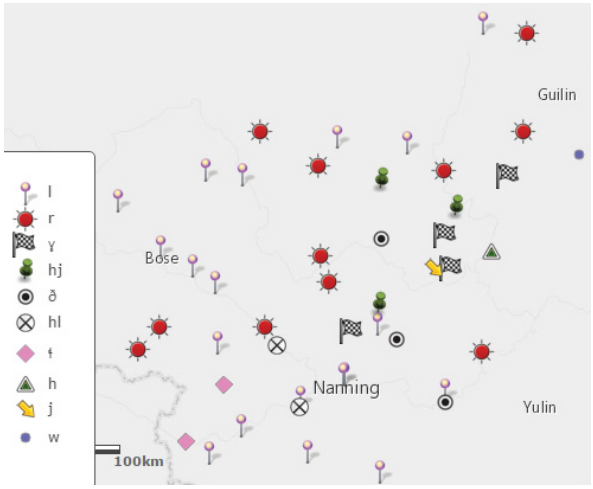
No. 80　木棉



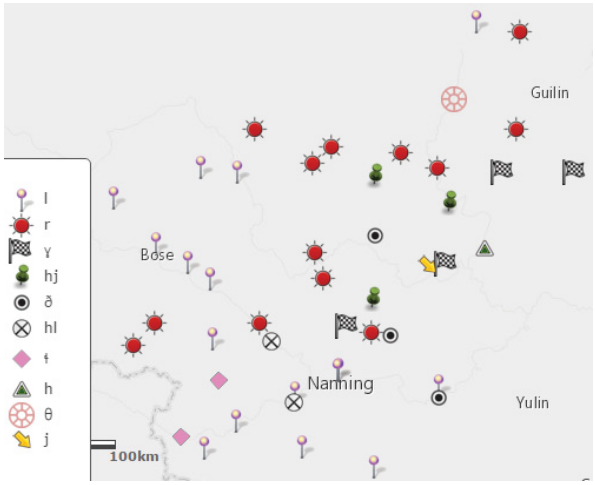
No. 81　苦楝树



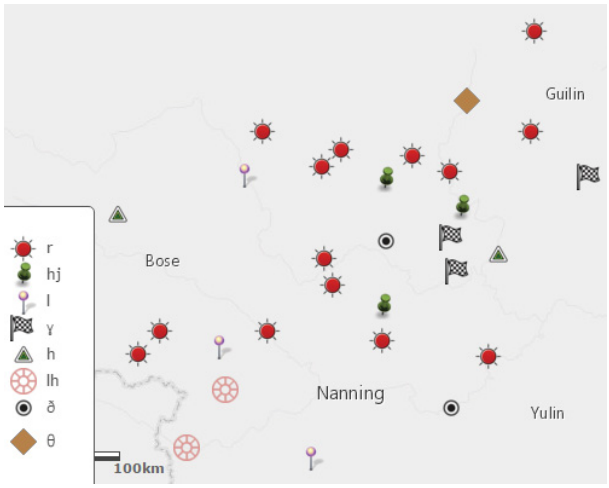
No. 82　知道，晓得



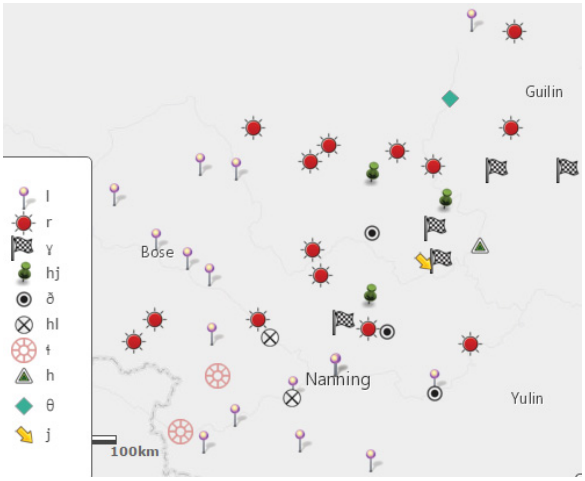
No. 83　“漏”雨



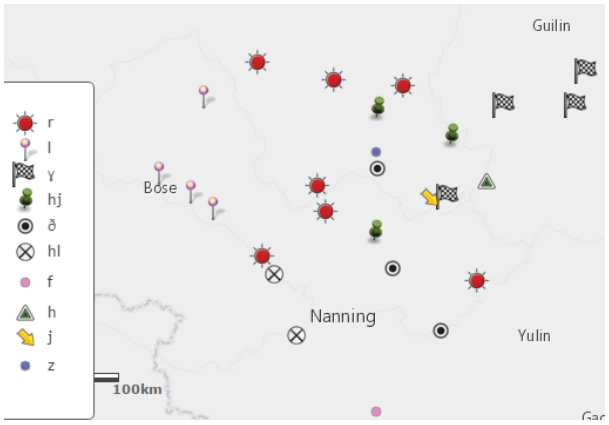
No. 84　鸟“窝”



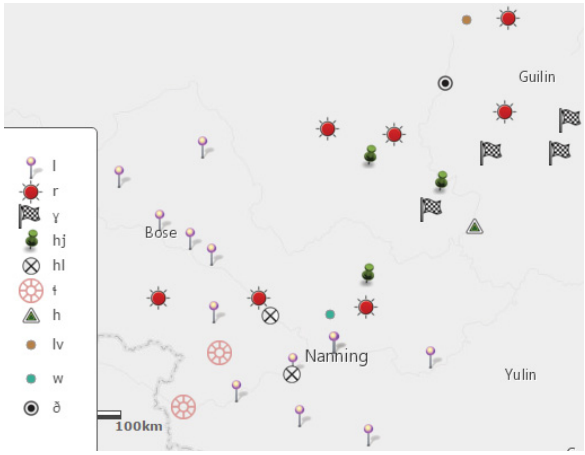
No. 85　虎“叫”，“吼”



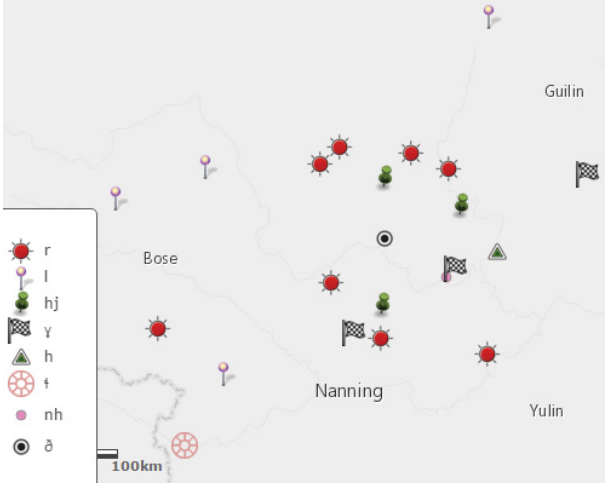
No. 86　天“亮”了



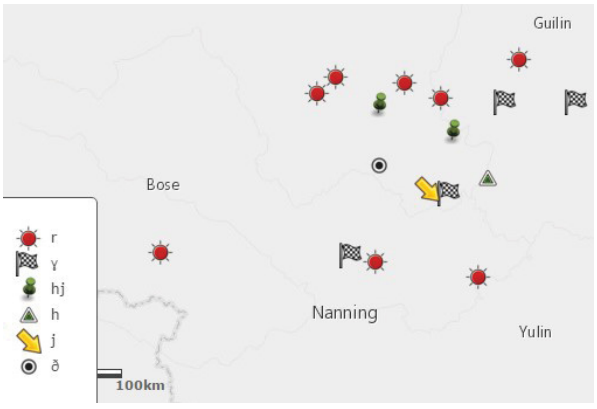
No. 87　牛“浸”水



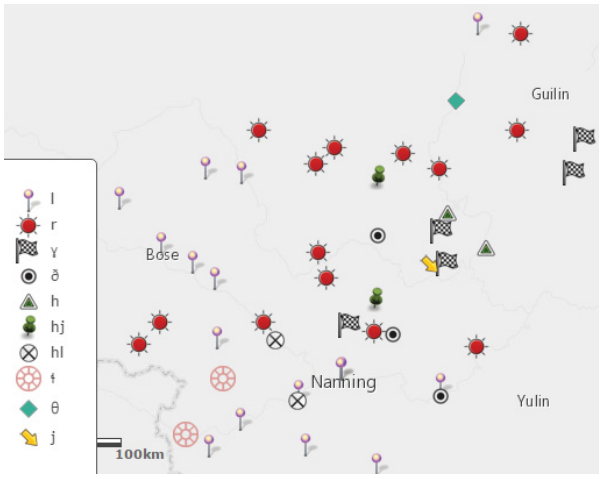
No. 88　呕吐



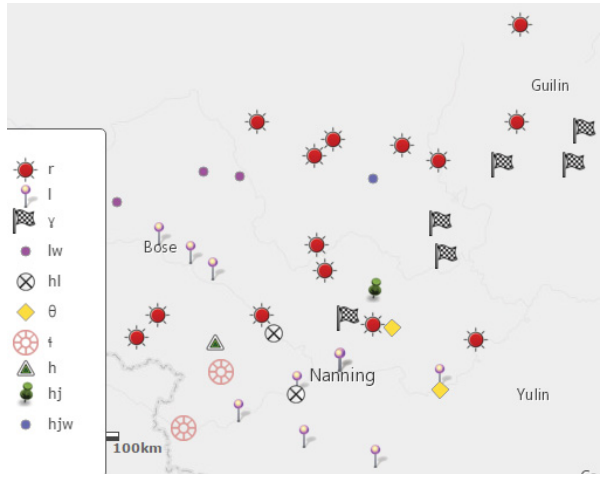
No. 89　衣胞，胎衣



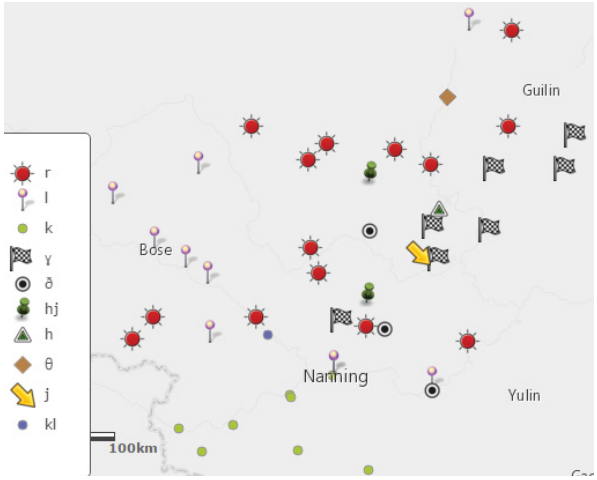
No. 90　丑



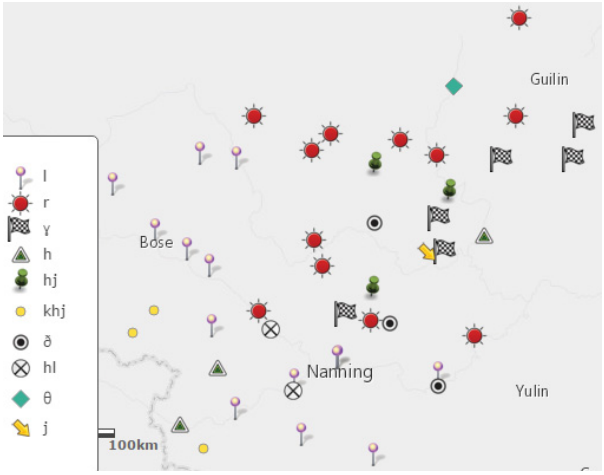
No. 91　畲地



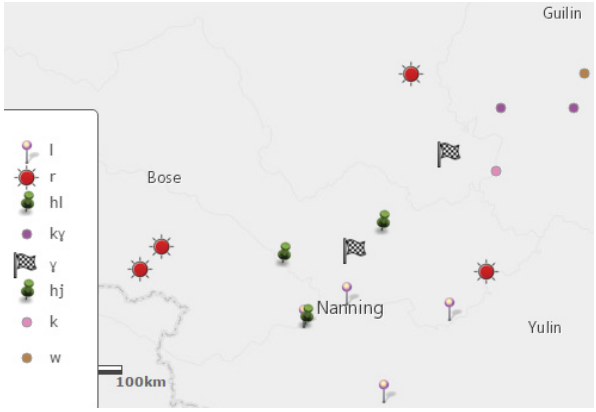
No. 92　鸡“虱”



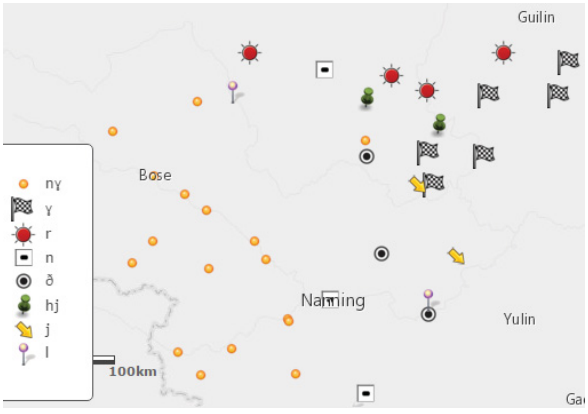
No. 93　糠，稻壳



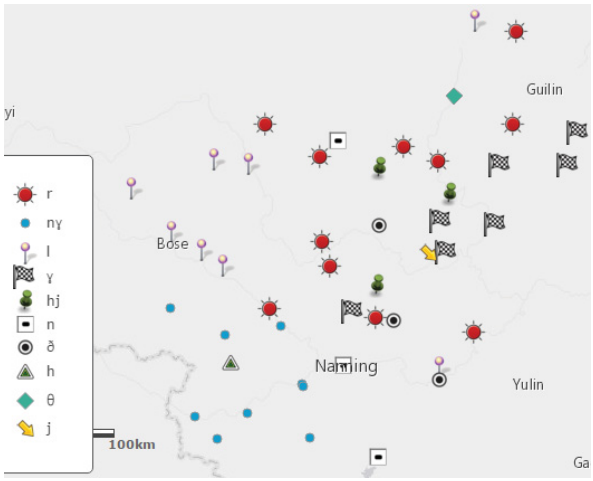
No. 94　“耳”朵



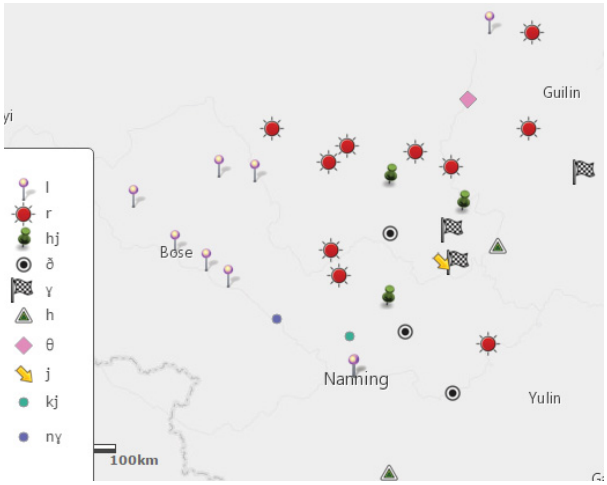
No. 95　不便于搬移的鸡“笼”



No. 96　影子



No. 97　芝麻



No. 98　干枯

Cultural Contexts of the Expansion of a Tibetan Word *'bras* 'Rice' in the Easternmost Tibetosphere

Hiroyuki Suzuki[a] and Sonam Wangmo[a]

[a]IKOS, University of Oslo

Abstract

This article attemps to provide linguistic maps regarding the word form 'rice' used in the languages spoken in the easternmost Tibetosphere, focusing on Tibetic and Qiangic languages spoken in Sichuan and Yunnan, China, and to examine how the Tibetan word form *'bras* has widely expanded, not only in Tibetic languages but also in non-Tibetic languages and varieties. Our principal hypothesis is: Tibetan *'bras* has been used as a religious material and it is thus regarded as a cultural word, therefore many languages and varieties spoken in the Tibetosphere employ the same cognate.

1 Introduction

Suzuki et al. (2016ab) have drawn linguistic maps of 'rice plant' and 'rice' in general in Tibeto-Burman languages, with which we can see that word forms for 'rice' employed in most Tibetic languages as well as languages spoken in the eastern Tibetosphere are common to each other, i.e., a form corresponding to Written Tibetan (henceforth WrT) *'bras*.[1] However, as we can imagine, most parts of the Tibetosphere are not suitable to cultivate rice and this word must not be a basic word in these languages. Indeed, it is known that more than 70 per cent of word forms are shared with varieties of the Tibetic languages (Jin ed. 1983:144), and more than 90 per cent of word forms correspond to a WrT form. Hence, it is not quite peculiar that the word for 'rice' is also shared with many varieties within the Tibetic languages. However, if the word 'rice' is acquired through a cultural contact, how can this word form be widely shared within the Tibetic languages distributed in the widest area among the Tibeto-Burman languages? For this question, the present authors will raise a hypothesis that the word 'rice' spreaded as a religious word all over the Tibetosphere, and because of this reason, this word can be borrowed by other non-Tibetic languages spoken in the eastern Tibetosphere, such as rGyalrongic and Qiangic languages, some of which have originally had their own word forms of 'rice' (and a semantic division of 'rice' if applicable).

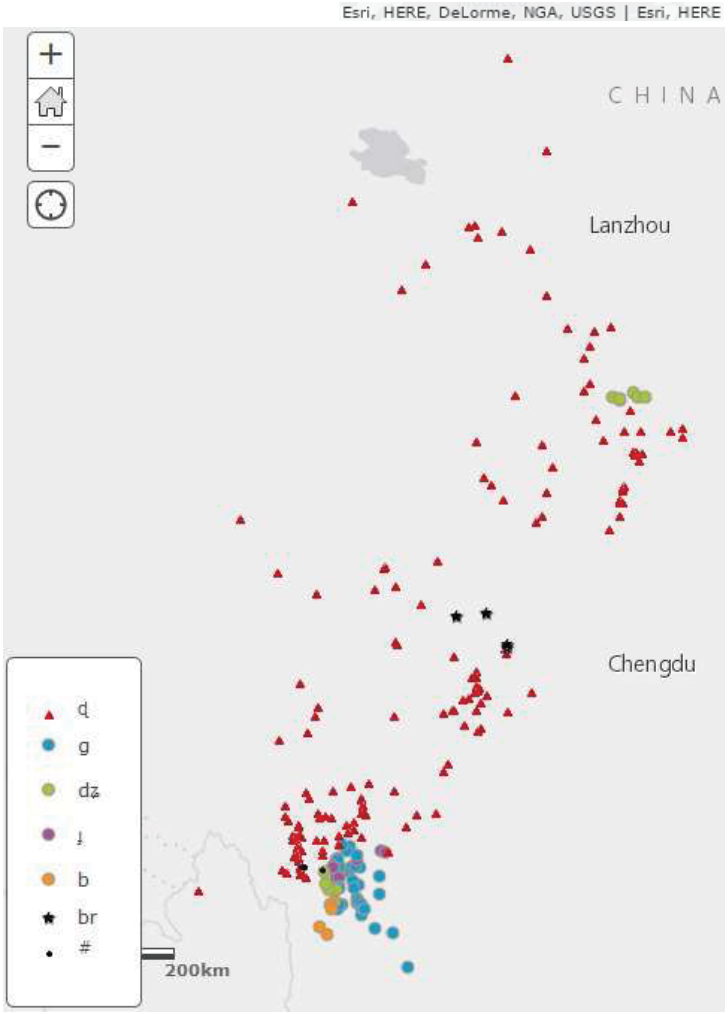
This article provides a detailed description of the geolinguistic analysis of the word forms for 'rice' derived from WrT *'bras* in the languages spoken in the easternmost Tibetosphere. The geographical scope of the eastern Tibetosphere follows the definition of Suzuki (2015), and the easternmost Tibetosphere corresponds the places confronting to other cultural areas. The article focuses on Tibetic and Qiangic languages spoken in Sichuan and Yunnan. The linguistic maps reflect so-called 'regiolects', i.e., dialects with regional differences. Sociolects, which certainly exist in the given area, are not dealt with in this article.

2 WrT *'bras* and its phonetic variation

There are many phonetic realisations of the word form derived from WrT *'bras*, some of which are: [ⁿɖɛ:], [ᵐɖi:], [ᶮʝi:], [ⁿdʑe:], [ᵐbɛ˗:], [ᵐbɛˢ:], [ᵐbrɛ:], [ⁿɖɯ ɦɯ], [ᵑgɯ:], and [ŋgi:]. Paying attention to the

[1] Note that WrT *'bras* corresponds to Proto-Tibeto-Burman (PTB) **b-ras* 'RICE / FRUIT / BEAR FRUIT / ROUND OBJECT'. Tibetic languages principally employ this PTB etymon for 'rice' among the Tibeto-Burman languages. See STEDT: http://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl/etymon/2071, accessed 16th March 2016.

initial sound of this word form, the distribution of various phonetic realisations attested in the Tibetic languages are displayed as Map 1 (from Suzuki 2016):



Map 1: Distribution of the main initial (with a glide) sound corresponding to WrT *'bras*.[2]

Other than them, phonetic forms attested in non-Tibetic languages are following:

Table 1: WrT *'bras* in non-Tibetic languages[3]

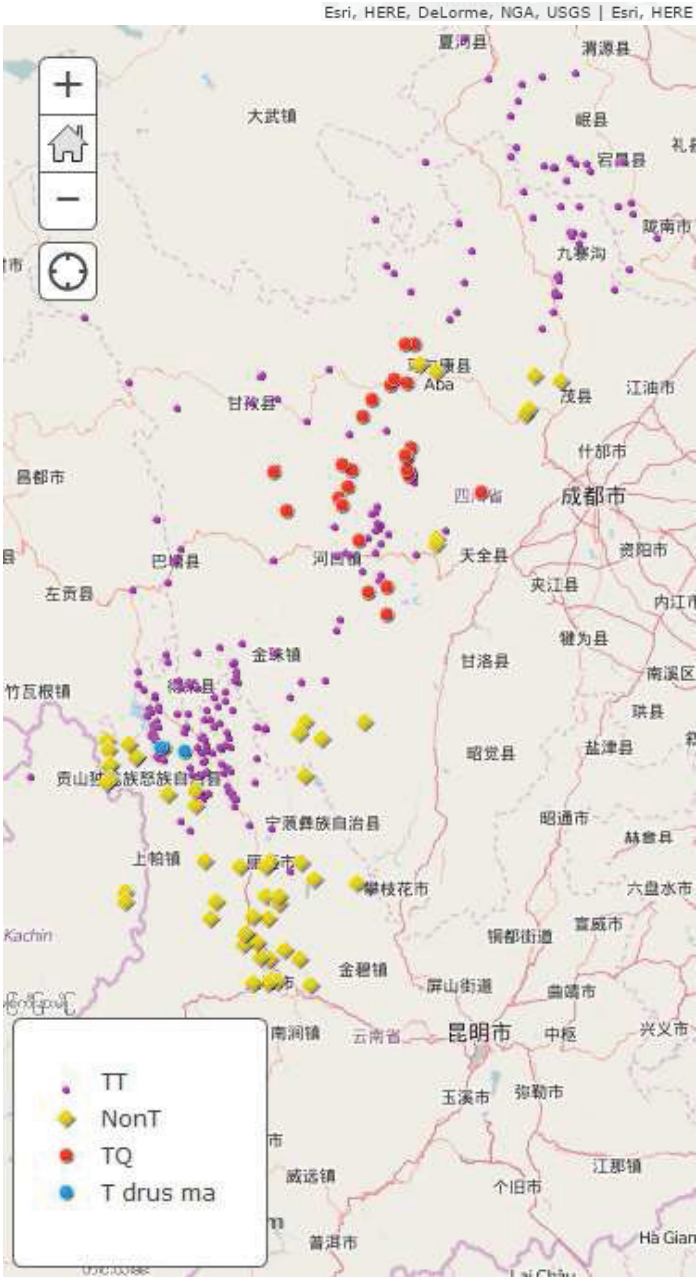
Language	Dialect	Word form for WrT *'bras*
Chuchen rGyalrong	Munashan	^{m}bras
bTsanlha rGyalrong	Sengge	^{m}bras
sTau	Mazur	^{m}brɛ
Geshitsa	brGyargyud	^{m}brɛ
Lhagang Choyu	Thamkhas	^{m}ɖwa
nDrapa	Ngwirdei	^{m}ɖɛ
Darmdo Minyag	Lhatseshis	^{n}ɖʐe
Nyagrong Minyag	Shoring	^{n}ɖri

[2] This map is designed with ArcGIS online. The legend does not reflect the preinitial feature (prenasalisation in most cases); 'ɖ' includes both a plosive /ɖ/ and an affricate /ɖʐ/; '#' means lack of the form corresponding to WrT *'bras*.

[3] The data has been collected by the first author. The suprasegmental description of word forms is uniformly omitted.

Even though the phonetic variation is wide, it is easy to understand that they are derived from the single WrT form *'bras*. Phonetic variation is generally not a criterion to classify word forms. However, an irregular sound correspondence should be noted, because it might show a spreading process of the irregular form.[4]

An overall distribution of the word form 'rice' derived from WrT *'bras* over the languages spoken in the eastern Tibetosphere is displayed in Map 2.



Map 2: Distribution of the word form 'rice' in the languages in the eastern Tibetosphere.[5]

[4] A partial discussion for the irregular phonetic form of WrT *'bras* 'rice' was provided in Suzuki (2012).

[5] Legend: TT: Tibetic languages using WrT *'bras*; TQ: : non-Tibetic (especially Qiangic) languages using WrT *'bras*; NonT: non-Tibetic languages not using WrT *'bras*; T drus ma: Tibetic languages using WrT *drus ma*.

As shown in Map 2, the issue regarding this article is mainly limited to the Minyag-rGyalrong area and the north-western part of Yunnan because these areas display a complicated situation. The authors are going to analyse these two cases in detail. First, we describe the usage of rice in the religious life under the Tibetan cultural area, taking Lhagang, the sacred place worshipped by Tibetans inhabiting the surrounding areas, as an example, and show the importance of the rice in their belief. Second, we analyse the way of spreading word (phonetic) forms for 'rice' by drawing specific linguistic maps. The basic data is common to Suzuki et al. (2016ab), the project of *Studies in Asian Geolinguistics*.

3 Use of 'rice' in everyday life and rituals: example from Lhagang Village

Lhagang Village is located on the Minyag Rabgang region,[6] a part of the easternmost Tibetosphere, where a monastery with a locally well-known Bodhisattva statue lays[7] and attracts many pilgrims not only from Minyag but also from its surrounding areas including rGyalrong. Under this perspective, Lhagang Village functions as a 'crossroad' of various local cultures within the easternmost Tibetosphere.

At present, rice is widely eaten as a part of staple food by Tibetans in Minyag and rGyalrong. Since rice does not grow on the plateau of Minyag Rabgang, it is certainly 'imported', at least in Lhagang Village, from Dartsendo Town (known as Lucheng Town), the administrative centre of this region. However, there are no rice fields in the town. Hence, rice should be transported from other places, perhaps from the Han territory such as Yaan. In the rGyalrong valley, rice can be cultivated, however, we rarely see rice field there. Rice in the rGyalrong area should also be imported from the contacting Sinosphere.

Rice is also employed when people practise rituals. Lhagang Monastery belongs to the Sakyapa sect of Tibetan Buddhism; however, the use of rice in rituals is quite common to any sects. In Lhagang Village, we can principally see two rituals using rice: *bdun mtshon chus skyes* and *'bras bsres ma sku*. *bdun mtshon chus skyes* is to consecrate rice to water by soaking it in water or alcohol; *'bras bsres ma sku* is to prepare boiled rice cooled and hardened with butter in a small bowl, used when a monk comes to a laypeople's house to recite a sutra for eliminating misfortunes and driving bad luck out from the house. In addition, Tibetans put crops including rice in a *maṇḍa*, a circle-shaped religious box symbolising a mandala in order to make offerings. According to the abbot of Lhagang Monastery, there are three monastic rituals utilising rice: *maṇḍa bzhi mchod*, *rab gnas*, and *sbyin bsreg cho ga*. Other than them, rice is also used in any rituals as a replacement of white stones and/or white ritual objects just when they lack. This situation indicates that rice plays an important role in religious ceremonies even though its use is limited.

Ritual use of rice is widespread in the Tibetosphere. Since it is not recent expansion, the supply of rice to the Tibetosphere in the past and present is a question, which has not been well investigated so far. The present description is not enough to figure out a complete way from the provenance to the destination of rice. However, we can understand how rice is treated in the Tibetosphere and functions in the Tibetan culture. Therefore, the languages spoken under the strong influence of the Tibetan culture must have received the word *'bras* 'rice' as one of the cultural objects.

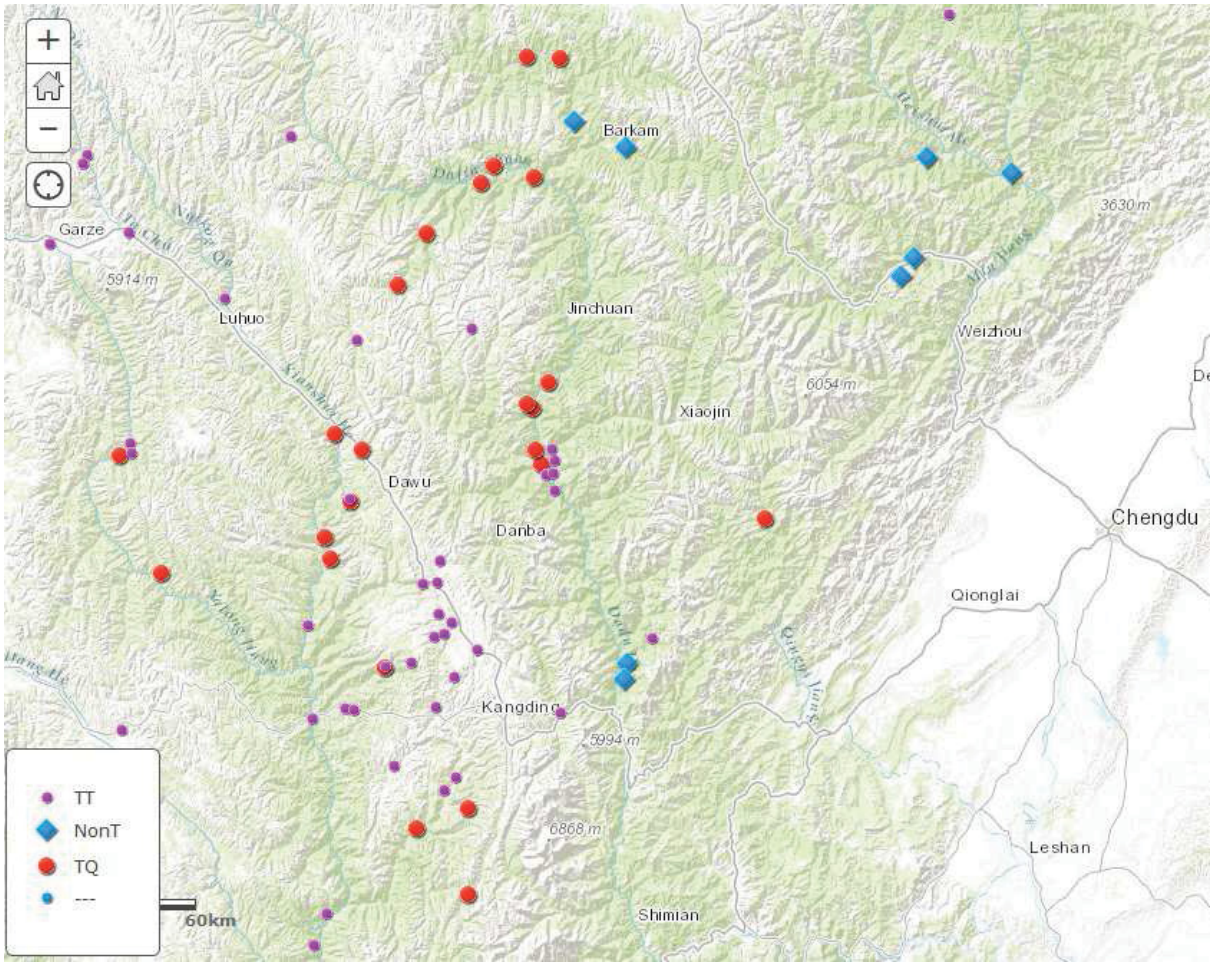
4 Word spreading process of WrT *'bras*: cases of two areas

4.1 Minyag-rGyalrong area

A linguistic map regarding the word form for 'rice' in the languages spoken in the Minyag-rGyalrong area is following:

[6] Administratively, Lhagang Village is in Tagong Town, Kangding Municipality, Ganzi Tibetan Autonomous Prefecture, Sichuan Province.

[7] See Sonam Wangmo (2013) and Suzuki & Sonam Wangmo (2015) for details.



Map 3: Word form for 'rice' in the Minyag-rGyalrong area.

In this region, many non-Tibetic languages and varieties employ the form corresponding to WrT *'bras*, so do all the Tibetic languages reflected on Map 3. The majority of non-Tibetic languages spoken there belong to the Tibetosphere, which means that the influence of the Tibetan culture and custom is extremely strong. Hence, this distribution is not extraordinary.

First of all, we should note that the border area of the varieties using this word form and those using their own word form. There are two places to be described: Situ-rGyalrong varieties in Maerkang County and nGochang (generally known as Guiqiong) in Kangding Municipality. Situ-rGyalrong varieties generally have one form for the whole 'rice' category, whereas nGochang, at least three forms (Suzuki et al. 2016b). This situation implies that nGochang has once been spoken in a rice cultivation area, or been a desceendent of a language spoken in a rice cultivation area. Another view can be pointed out: the contact of nGochang with Sinitic varieties, which have at least four words for the 'rice' category. In fact, the form for 'rice plant' of Qianxi nGochang is a Sinitic loan (*guzi*). In Situ-rGyalrong, there is an inherent word for 'rice', /k^{h}re/, which is, in fact, problematic; it might be an earlier Tibetan loan corresponding to WrT *khre* 'millet'. Situ-rGyalrong is mainly spoken inside steap valleys, where cultivation of crops except for barley is difficult, hence the rice has not occupied an important place in the language landscape. However, varieties spoken in the region which is much closer to the Tibetic-spoken area have accepted a form corresponding to WrT *'bras* 'rice' as shown in Map 3, see the distribution of TQ. Moreover, various phonetic forms of 'rice' related to WrT *'bras* in rGyalrongic languages (see Table 1) reflect the origin and period of borrowing from Tibetic varieties. For example, the pronunciation /^{m}bras/ attested in Situ, Chuchen, and bTsanlha rGyalrong has already disappeared in

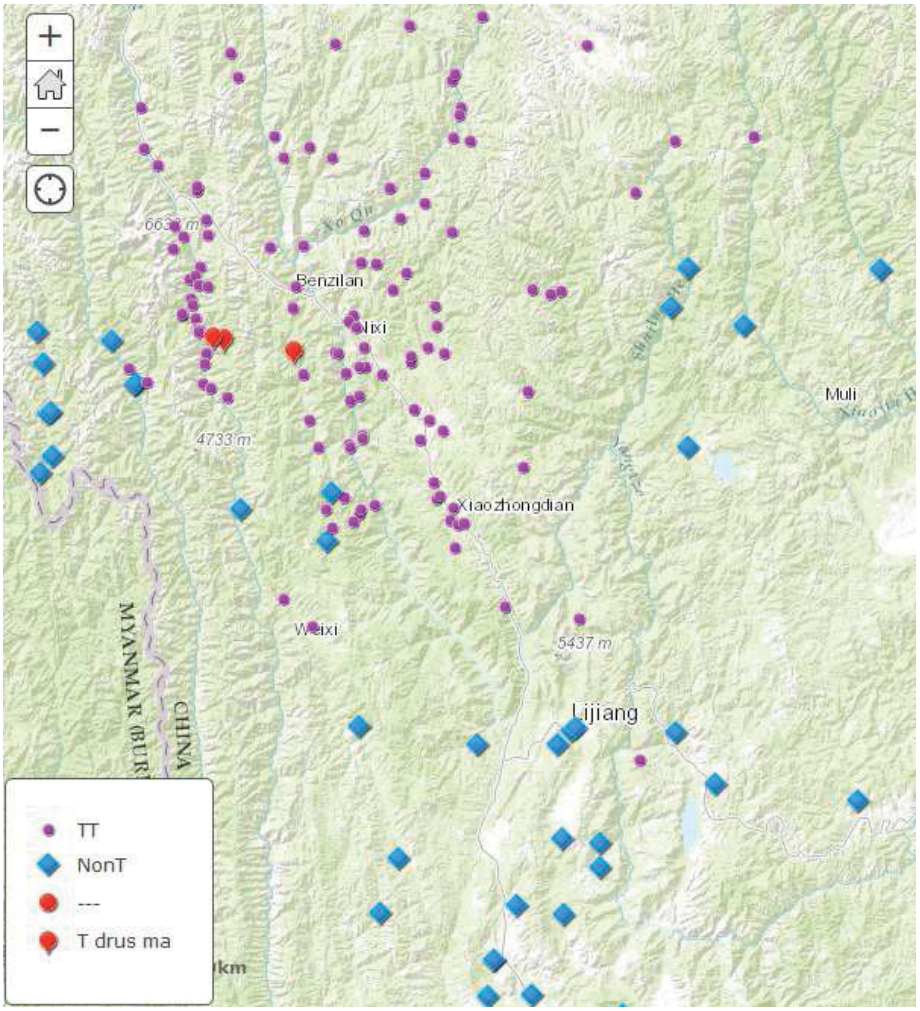
the majority of Tibetic varieties surrounding the rGyalrongic languages. [8] It means that this word form is an archaic loan.

Next, we will consider several factors that non-Tibetic languages accepted the Tibetan loan regarding 'rice'. The primary purpose of use of 'rice', especially 'rice grain', within the Tibetosphere might be religious ceremonies mentioned in Section 3. Since the Tibetic-spoken area is normally located higher than 3,000m altitude, rice cannot grow; however, Tibetan inhabitants employ rice for several special religious ceremonies, whether they practise Buddhism or Bon. If their use of rice is highly associated to the religious purpose, the word for 'rice' itself can be counted as a cultural word.

To sum up, the distribution of the word form is basically related to WrT *'bras* in non-Tibetic varieties in Map 3 is connected to Tibetic varieties. In addition, this word form is not a recent loan but an archaic one judging from the phonetic variation attested in rGyalrongic languages. A detailed process of borrowing should be investigated by referring to the historical sound development of Tibetic varieties.

4.2 Yunnan area

A linguistic map regarding the word form for 'rice' in the languages spoken in the Tibetosphere in Yunnan is following:



Map 4: Word form for 'rice' in the Tibetosphere in Yunnan.

[8] In fact, it is extremely difficult to find a variety of any Tibetic languages which has a phonetic form as /^{m}bras/. For example, an initial /^{m}br/ is attested near the rGyalrongic-spoken area (see Map 1), but its rhyme is not /as/. On the other hand, Ladak (Tibetic language spoken in North India) has an /as/ rhyme but its initial is /br/, lacking prenasalisation.

An interesting point in the Tibetosphere of Yunnan which is different from the case in Minyag-rGyalrong region is that there are no non-Tibetic languages and varieties employing the form corresponding to WrT *'bras*. Moreover, some Tibetic varieties do not use WrT *'bras*-form for 'rice' (T drus ma-type in Map 4). In Yunnan, we should pay more attention to exceptional phonetic realisations, especially a velar sound, attested in Tibetic varieties mainly distributed in Shangri-La Municipality (see Map 1) which are not reflected on Map 4, as well as varieties with a semantic division between 'rice plant' and 'rice grain'.

The idea 'every word has its history' is the most fundamental for geolinguistc research. If a given sound change cannot be explained in a straight way of the regular process, we should seek other factor(s) which caused the exception. The phenomenon observed in many varieties spoken in the central area of Shangri-La Municipality is that the velar sound /ᵑg/ appears on the position where the apparition of /ⁿdʑ/ or /ᶮɟ/ is expected. It is certain that some varieties has a regular sound correspondence between WrT *'br* and /ᵑg/, however, the distribution of such varieties is limited, and it is also complicated that they give some influence to varieties spoken in a wider region. Returning to Map 1, we notice that the distribution of /ᵑg/ forms a 'line' from the central area of Shangri-La Municipality to Lijiang Municipality. What does this shape of distribution mean? Thinking of this issue with other background information of the region and history, we can raise a hypothesis that there has been influence from Naxi, previous prestige language functioned while the Mu-chieftain period from Ming to Qing dynasties. In Lijiangba Naxi, phonetic realisations among prepalatal, palatal, and velar are not well distinct. In Shangri-La, rice is not cultivated but used as religious purposes as well as frequently eaten by inhabitants at present. In addition, the provenance of rice as a commercial item is Dali and Lijiang. In other words, the word for 'rice' was somewhat influenced from the pronunciation of Naxi-speakers and the oral form might have transmitted from south to north.[9]

Since the climate condition is appropriate for rice cultivation, Tibetans practise to plant rice in a part of the Tibetosphere of Yunnan, especially along lower Jinshajiang within Shangri-La and Weixi. In this area, Tibetans' language also have a semantic division within the 'rice' category, i.e., 'rice plant' and 'rice grain' (Suzuki 2016). Of them, the form of 'rice plant' generally corresponds to WrT *'bras*, as reflected on Map 4, while that of 'rice grain', to WrT *drus ma* (see also Suzuki 2012). The class 'T drus ma' of Map 4 could have been generated by replacing WrT *'bras* for WrT *drus ma* over the whole semantic division of 'rice' because 'rice' as a food is more important than that as a plant in the non rice cultivating area.

5 Conclusion

The word form of 'rice' in the Tibetic languages in the eastern Tibetosphere mainly corresponds to WrT *'bras*, and its geographical distribution is nearly pervasive. Most regions do not belong to the rice cultivation area, however, varieties have the same stem for rice. It is probably because the rice is used for religious rituals, whether they are of Bon or Buddhism. The rice is generally a kind of staple food, but in the case of Tibet, it can be used for a religious purpose.

In the Minyag-rGyalrong area, the loan of the word form WrT *'bras* is certainly related to the distribution of non-Tibetic languages. Most varieties spoken in the vicinity of Tibetic-spoken area employ a WrT *'bras* form for 'rice'. Its expansion is highly connected with the strength of Tibetan cultural influence.

In the Tibetosphere in Yunnan, however, a complicated system is attested. Several dialects spoken under the rice cultivation culture distinguish 'rice grain' from 'rice plant' by using different stems. The irregular sound correspondence of WrT *'bras* is also seen in Yunnan, which might have spreaded from the Naxi area to its north following the cultural influence of Lijiang.

[9] See Suzuki (2016) for a detail.

Acknowledgements

We should like to express my gratitude to my Tibetan friends who helped us and taught us their mother tongue. Field research was funded by four Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science [JSPS]: “Linguistic Substratum in Tibet” (headed by Yasuhiko Nagano, No. 16102001), “Dialectological Study of the Tibetan Minority Languages in the Tibetan Cultural Area in West Sichuan” (headed by the present author), “International Field Survey of the rGyalrongic Languages” (headed by Yasuhiko Nagano, No. 21251007), and “Study on the Dialectal Development of Tibetan Spoken in Yunnan, China, through a Description of the Linguistic Diversity” (headed by the present author, No. 25770167), as well as and a private financial support of Tibetan Studies Committee of Yunnan Ethnology Association headed by Xu Jianhua.

References

Jin, P. ed. 1983. *Zangyu Jianzhi*. Minzu Chubanshe.

Sonam Wangmo. 2013. *Lhagang Monastery in Myth, History and Contemporary Society*. Universitetet i Oslo M.Phil. Thesis.

Suzuki, H. 2012. À propos du terme ‘riz’ et de l’hypothèse du groupe dialectal Sems-kyi-nyila en tibétain du Khams. *Revue d’études tibétaines* 23:107-115.

Suzuki, H. 2015. A geolinguistic description of terms for ‘sun’ in Tibetic languages in the eastern Tibetosphere. *Studies in Asian Geolinguistics* 1:79-85.

Suzuki, H. 2016. A geolinguistic description of terms for ‘rice’ in Tibetic languages of the eastern Tibetosphere. *Studies in Asian Geolinguistics* 2 (in press).

Suzuki, H. and Sonam Wangmo. 2015. Quelques remarques linguistiques sur le tibétain de Lhagang, «l’endroit préféré par le Bodhisattva». *Revue d’études tibétaines* 32:153-175.

Suzuki, H., S. Shirai, K. Kurabe, K. Iwasa, S. Ebihara, and I. Matsuse. 2016a. Rice plant: Tibeto-Burman. *Studies in Asian Geolinguistics* 2 (in press).

Suzuki, H., K. Kurabe, K. Iwasa, S. Shirai, S. Ebihara, and I. Matsuse. 2016b. Geolinguistic analysis of ’rice’ in Tibeto-Burman. *Studies in Asian Geolinguistics* 2 (in press).

A preliminary report on the distribution of 'eclipse' in languages of Mainland Southeast Asia

Keita Kurabe

Japan Society for the Promotion of Science / ILCAA, Tokyo University of Foreign Studies

Abstract

The aim of this paper is to provide a preliminary overview of expressions for solar and lunar eclipses in East and mainland Southeast Asian languages from the geolinguistic point of view. Our findings, based on an investigation of about 100 languages and dialects, can be summarized as follows. (a) At least 80 languages, irrespective of genetic affiliation of them, express solar and lunar eclipses employing ingestive verbs with morphemes for the sun or the moon and predator nouns which consume it. (b) Ingestive verbs involved in the eclipse expression are almost always 'to eat' or 'to swallow'. (c) Predator nouns involved, in most cases, are some kinds of animals such as a dog, frog, tiger, flying squirrel, spirit, dragon, snake, centipede, and fish. (d) Languages sometimes exhibit some kinds of asymmetries between solar and lunar eclipses.

1 Introduction

Many languages of East and Southeast Asia express the eclipse of the sun and the moon employing ingestive verbs such as 'to eat' and 'to swallow' in addition to morphemes referring to the sun and moon. The aim of this paper is to explore expressions for solar and lunar eclipses in East and mainland Southeast Asian languages and to provide a preliminary overview of them. Our primary finding, based on an investigation of about 100 languages and dialects, is that at least 80 languages, irrespective of genetic affiliation of them, express solar and lunar eclipses employing ingestive verbs with morphemes for the sun or the moon and predator nouns which consume it.

2 Solar and Lunar eclipses: A quick exemplification

Many languages of East and Southeast Asian languages, regardless of genetic affiliation of them, express solar and lunar eclipses employing ingestive verbs, especially 'to eat' and 'to swallow'. First observe this in the following examples, which appear to be in a borrowing relationship, where Japanese and Chinese express solar and lunar eclipses employing a verb 'to eat' in addition to morphemes for the sun and moon.

(1) Mandarin
a. rì-shí
sun-eat
'solar eclipse'
b. yuè-shí
moon-eat
'lunar eclipse'

(2) Japanese
a. nit-syoku
sun-eat
'solar eclipse'
b. get-syoku
moon-eat
'lunar eclipse'

Although the expressions in Mandarin and Japanese given above do not mention about what/who eats the sun and moon, many languages in East and mainland Southeast Asia, including Mandarin, have expressions for the solar and lunar eclipse that explicitly express the agent of consumption. Consider:

(3) Mandarin [Sinitic]
tiān gǒu chī yuèliang
sky dog eat moon
'lunar eclipse' (lit. 'celestial dog eats the moon')

(4) Lahu [Tibeto-Burman] (Matisoff 2006:77)
mû-ni lâ câ ve
sun tiger eat PTCL
'have an eclipse of the sun' (lit. 'tiger eats the sun')

(5) Kadu [Tibeto-Burman] (Dr. Keisuke Huziwara, p.c., 2015)
sə^{1}miʔ^{2} pə^{1}touŋ^{2} nyon^{2}
sun python swallow
'have an eclipse of the sun' (lit. 'python swallows the sun')

(6) Shan [Tai-Kadai] (Sao Tern Moeng 1995)
kop^{4} kin^{1} lɤn^{1}
frog eat moon
'lunar eclipse' (lit. 'frog eats the moon')

(7) Yay [Tai-Kadai] (Hudak ed. 1991:28)
baaŋ^{2} kɯn^{1} dɯan^{1}
flying.squirrel eat moon
'lunar eclipse' (lit. 'flying squirrel eats the moon')

(8) Kaco' [Austroasiatic] (Dr. Nathan Badenoch, p.c., 2016)
ka:ŋgp̚w lə:n khɒj
catfish swallow moon
'lunar eclipse' (lit. 'catfish swallows the moon')

(9) Sedang [Austroasiatic] (Smith 2012:424)
kɔxɛ́ ka hài
centipede eat sun
'solar eclipse' (lit. 'centipede eats the sun')

(10) Mong Njua [Hmong-Mien] (Lyman 1974:199)
kláŋ nâo hli
spirit eat moon
'lunar eclipse' (lit. 'spirit eats the moon')

3 Previous studies

There are seemingly no previous linguistic studies that investigate the expressions for solar and lunar eclipses in East and mainland Southeast Asian languages. In folklore and mythological studies, on the other hand, it is known that Chinese and some indigenous peoples in East and Southeast Asia have folklore or mythology that attributes eclipses to devouring of the sun and moon by animals or mythological beings. Wallis (1918:369), for example, mentions a Chinese belief about eclipses, remarking that "the sun or moon is being swallowed by a dog or other beast. They accordingly beat gongs to rescue it by frightening away the devourer." This belief can be traced back to ancient China, as the Huainanzi (before 139 BC) records a belief that attributes lunar eclipses to devouring of the moon by a frog in the moon. Also, Hinduism, as found in a narrative of Mahābhārata, attributes eclipses to Rāhu, the severed head of an Asura, Svarbhānu, which swallows the sun.

4 Eclipses with ingestive verbs

Solar and lunar eclipses in East and Southeast Asian languages, as exemplified in Section 2, tend to be expressed employing ingestive verbs, such as 'to eat' and 'to swallow'. This section provides a close look at these expressions, classifying them from three different perspectives. (a) What consumes the sun and moon? (b) In what manner? (c) What consumes the sun and moon in what manner? In what follows, these three perspectives will be discussed in order.

4.1 What consumes the sun or moon?

Eclipse expressions with ingestive verbs can be divided into two types in terms of the agent of ingestion: one which does not overtly express the agent and one which does. Japanese belongs to the former type, as illustrated by (2) in Section 2. The same holds for Sgaw Karen (TB), Rawang (TB), Garo (TB), Vietnamese (AA), and Buginese (AN), among others. Mandarin and most Sinitic languages have both expressions as illustrated by the Mandarin examples given in (1) and (3). Eclipse expressions with overt agents prevail in languages of East and mainland Southeast Asia. (11) lists languages that have this type of expressions, being classified based on language families and branches.

(11) *a. Sino-Tibetan*
[Sinitic] Gan, Hakka, Hui, Jin, Mandarin, Min Bei, Min Dong, Min Nan, Min Zhong, Wu, Xiang, Yue; [Angami-Pochuri] Sema; [Jinghpaw-Luish] (Jinghpaw) Jinghpaw, Gauri, Duleng, (Luish) Kadu, Ganan; [Lepcha] Lepcha; [Naic] Naxi; [Ngwi-Burmese] (Burmish) Zaiwa, Lhaovo, Lacid, Ngochang, (Ngwi) Azhe Yi, Lahu, Lisu, Nisu Yi, Nuosu Yi, Sani Yi; [Northern Naga] Chang; [Qiangic] Northern Qiang, Southern Qiang; [Tani] Apatani; [Tangkhul] Tangkhul

b. Tai-Kadai

[Kra] Gelao; [Kam-Sui] Dong, Maonan, Mulam, Sui, [Hlai] Hlai?; [Ong Be] Be; [Tai] (Central) Western Nung; (Northern) Bouyei, Saek, Yongbei Zhuang; (Southwestern) Ahom, Black Tai, Dehong, Lao, Lue, Shan, Thai; (others) Lungming, Yay

c. Austroasiatic

[Bahnaric] (Central) Bahnar, (North) Kaco', Rengao, Sedang, (South) Mnong; [Katuic] Kui; [Khasian] Khasi; [Khmeric] Khmer; [Khmic] Bit, Khmu; [Palaungic] (Angkuic) Man Met, (Waic) Wa; [Aslian] Senoi

d. Hmong-Mien

[Hmongic] (Chuanqiandian) Chuanqiandian, Diandongbei, Luobohe, Mong Njua, White Meo, (Xiangxi) Xiangxi, (Qiandong) Qiandong; [Mienic] (Biao-Jiao) Biao Min Yao, (Bunu) Bunu, (Mian-Jin) Mian

The map given below shows the geographical distribution of languages with the overt agent eclipse expression. Observe in the map that this type of expressions is widely distributed in the region, being observable in China (incl. Hainan), Vietnam, Cambodia, Laos, Thai, Burma, NE India (incl. Nepal, Bhutan), Malaysia, and Indonesia.

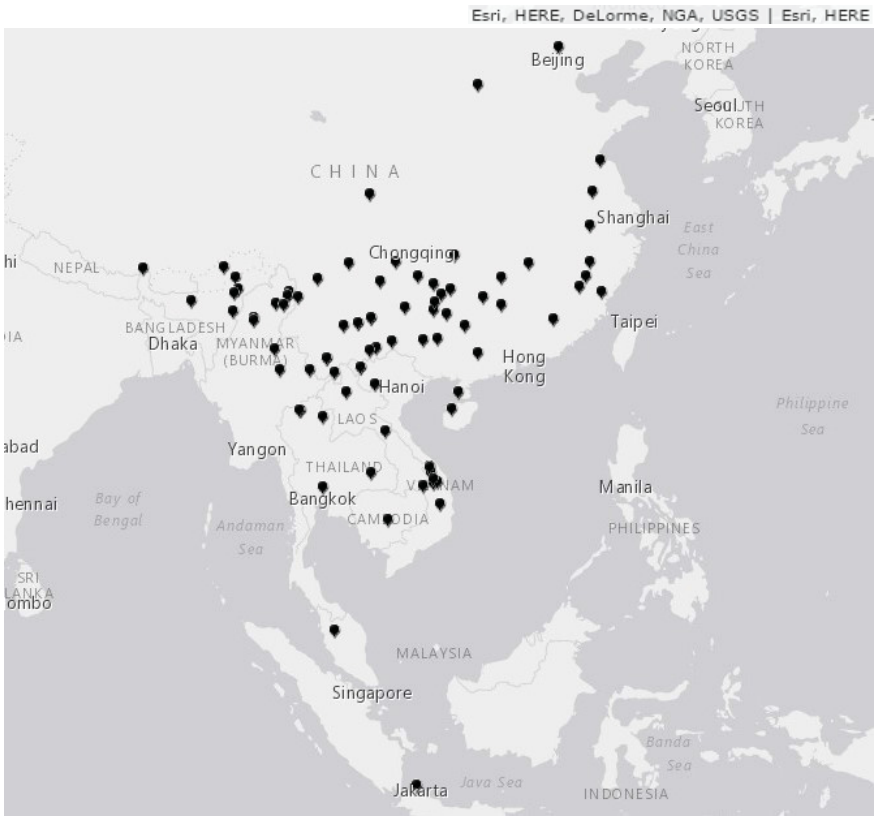


Figure 1: Geographical distribution of eclipse expressions with overt agents (produced with ArcGis)

The following list provides languages with the overt agent eclipse expression grouped in terms of genetic affiliation and of semantic types of the agent, i.e. what consumes the sun and moon. Observe in the list that, although some languages make use of inanimate agents such as wood, the sun, and the moon, most languages employ some kinds of animals in order to express the solar and lunar eclipse. Recurrent animal motifs distributing cross genetic affiliation are a dog and a frog.

(12) *a. Sino-Tibetan*

DOG: Azhe Yi, Lacid, Lhaovo, Lisu, Naxi, Nisu Yi, Ngochang, Nuosu Yi, Qiang, Sani Yi, Zaiwa; FROG: Apatani, Jinghpaw, Lahu; PYTHON: Kadu; CELESTIAL DOG: Gan, Hakka, Hui, Jin, Mandarin, Min Bei, Min Dong, Min Nan, Min Zhong, Wu, Xiang, Yue; SNAKE: Ganan; TIGER: Azhe Yi, Chang, Lahu, Lepcha, Sani Yi, Sema, Tangkhul;

b. Tai-Kadai

FROG: Ahom, Black Tai, Bouyei, Dehong, Lao, Lue, Maonan, Saek, Shan, Sui, Thai; DOG: Gelao, Mulam; DOG [PSN]: Black Tai; FLYING SQUIRREL: Yay; WOOD: Be; MYTHICAL CREATURE: West Nung; CELESTIAL DOG: Lungming; SUN: Be (i.e. 'sun eats the moon)

c. Austroasiatic

CATFISH: Kaco'; CENTIPEDE: Sedang; FISH (kind of): Bahnar, Sedang; FROG: Bit, Khasi, Khmu, Man Met; DOG: Paraok Wa; MOON: Mnong (i.e. 'moon swallows the sun'); RAHU: Khmer, Kui, Senoi

d. Hmong-Mien

CELESTIAL DOG: Luobohe Miao, Xiangxi Miao; SPIRIT: Mong, Chuanqiandian Miao, Diandongbei Miao, Mong Njua, White Meo

The map given below shows the geographical distribution of animal motifs employed for the eclipse expressions. Observe in the map that a dog and a frog prevail in East and mainland Southeast Asia.

Figure 2: Geographical distribution of animal motifs utilized for eclipses (produced with ArcGis)

4.2 In what manner?

Eclipse expressions with ingestive verbs can also be classified in terms of what verb is involved. The following list provides languages classified in terms of genetic affiliations and of ingestive manners. Observe that prevailing verbs across genetic affiliation are 'to eat' and 'to swallow'. Note that languages in the parentheses are of the agentless type that expresses the eclipse with no devourer, and languages without parentheses are of the type that explicitly expresses the devourer.

(13) a. *Sino-Tibetan*

EAT: Azhe Yi, Gan, Hakka, Hui, Mandarin, Jiangwei Mandarin, Jin, Min Dong, Wu, Xiang, Yue, Lahu, Qiang, Zaiwa, Lacid, Lisu, Lhaovo, Nisu Yi, Ngochang, Nuosu Yi, Sani, Yi, Sema, Tangkhul, Chang; SWALLOW: Apatani, Jinghpaw, Jiangwei Mandarin, Kadu, Ganan, (Abor-Miri, Lushai, Sgaw Karen, Garo Rawang); BITE: Linwu Hunan, Lepcha, Naxi; CRACK (as seeds between one's teeth): Min Bei, Min Zhong

b. Tai-Kadai

EAT: Thai, Shan, Lao, Ahom, Black Tai, Dehong, Gelao, Lue, Bouyei, Yay, Saek, Sui, Lungming, Maonan, Mulam, Be, West Nung, (Kam); SWALLOW: Yongbei Zhuang, Hlai

c. Austroasiatic

EAT: Bit, Khasi, Kui, Man Met, Paraok Wa, Sedang, (Jeh); SWALLOW: Khmu, Senoi, Bahnar, Mnong, Kaco', Khmer, Plei Bong Bahnar, Rengao, (Vietnamese, Car)

d. Hmong-Mien

–EAT: Biao Min Yao, Chuanqiandian Miao, Diandongbei Miao, Luobohe Miao, Mong Njua, White Meo, Xiangxi Miao, (Qiandong Miao); SWALLOW: Mian Yao

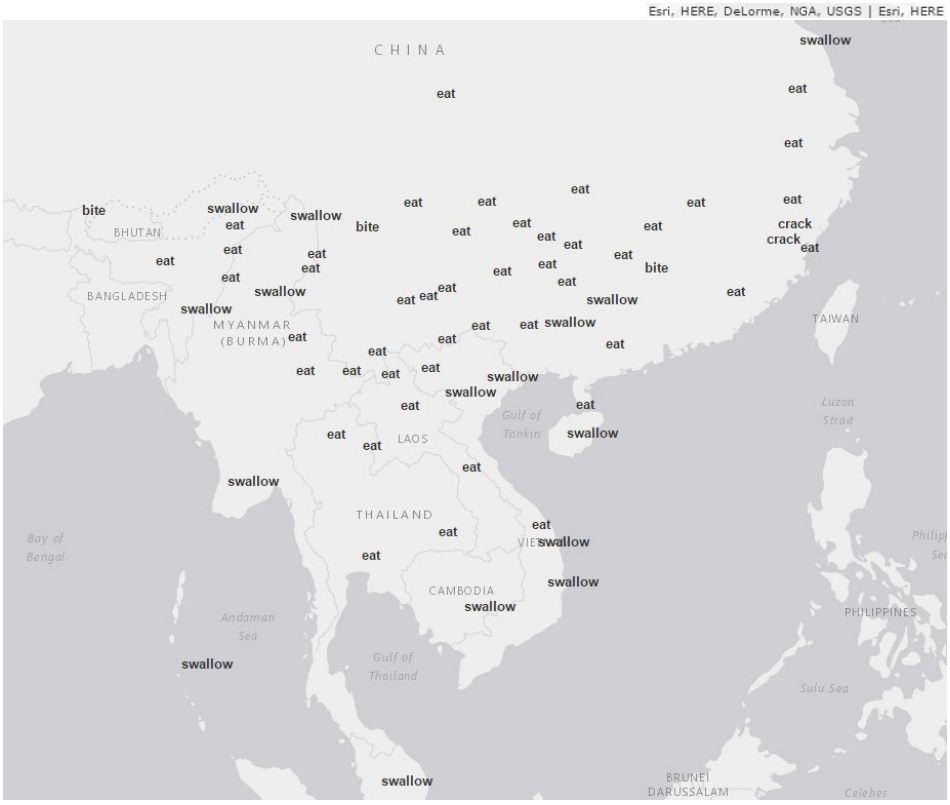


Figure 3: Geographical distribution of ingestive verbs employed for eclipses (produced with ArcGis)

Figure 3 shows the geographical distribution of ingestive verbs employed for the eclipse expression. Observe that the verb 'to eat' is predominant in the region followed by the verb 'to swallow', which appears to be distributed in more peripheral areas.

4.3 What consumes the sun and moon in what manner?

It is of interest to ask what eater is combined with what ingestive verb. The following is a summary of combinations of animals and ingestive verbs involved in eclipse expressions. It should be noted that this list is based on a limited data set and thus can be falsified when more data are taken into account.

(14) Combination of animal motifs plus verbs
–CENTIPEDE: eat
–FISH (kind of): eat
–FLYING SQUIRREL: eat
–MYTHICAL CREATURE: eat
–SPIRIT: eat
–SUN: eat
–WOOD: eat
–CATFISH: swallow
–MOON: swallow
–RAHU: swallow
–SNAKE/PYTHON: swallow
–FROG: eat, swallow
–CELESTIAL DOG: eat, bite, crack
–DOG: eat, bite
–TIGER: eat, bite

5 Symmetry and asymmetry of solar and lunar eclipses

It is not always the case that both solar and lunar eclipses are encoded by a symmetrical expression. Observe first in the following examples from Jinghpaw that eclipses of the sun and moon are expressed by symmetrical expressions that employ the same morphemes except the sun and moon.

(15) Jinghpaw [Tibeto-Burman] (my field notes)
a. jan ɕùʔ məyùʔ
sun frog swallow
'solar eclipse'
b. ɕəta ɕùʔ məyùʔ
moon frog swallow
'lunar eclipse'

There are also languages that encode the solar and lunar eclipse using asymmetrical expressions. Lahu, Sani Yi, Sedang and Black Tai employ asymmetrical expressions for solar and lunar eclipses in that two different animal motifs are utilized.

(16) Lahu [Tibeto-Burman] (Matisoff 2006:77)

a. mû-ni lâ câ ve
sun tiger eat PTCL
'have an eclipse of the sun'

b. ha-pa pā câ ve
moon frog eat PTCL
'have an eclipse of the moon'

(17) Sani Yi [Tibeto-Burman] (Wu and Ji eds. 2011:488, 687)

a. $lo^{21}tʂi^{33}mɒ^{33}$ $lɒ55$ $dzɒ^{21}$
sun tiger eat
'solar eclipse'

b. $łɒ^{33}bɒ^{33}mɒ^{33}$ $tʂ‘i^{21}$ $lɪ^{33}$ $dzɒ^{21}$
moon dog ? eat
'lunar eclipse'

(18) Sedang [Austroasiatic] (Smith 2012:423–4)

a. ko'xế ka hài
centipede eat sun
'solar eclipse'

b. kau ka khế
fish(type.of) eat moon
'lunar eclipse'

(19) Black Tai [Tai-Kadai] (Baccam Don et al. 1989)

a. ma cuəŋ ma ŋɛɛŋ kin taavan
dog Chuang dog Ngaeng eat sun
'solar eclipse'

b. kop kin bɯən
frog eat moon
'lunar eclipse'

Liangshan Yi also encodes solar and lunar eclipses with asymmetrical expressions, but this time with different verbs. Compare:

(20) Liangshan Yi [Tibeto-Burman] (Wu and Ji eds. 2011:488, 687)

a. $ho^{33}bu^{33}$ $sɿ^{33}$
sun die
'solar eclipse'

b. $ło^{21}bo^{21}$ $k‘ɯ$ $dzɯ$
moon dog eat
'lunar eclipse'

6 Summary and prospects

This paper provided a first look at expressions for solar and lunar eclipses in East and mainland Southeast Asian languages. Our findings, based on an investigation of about 100 languages and dialects, can be summarized as follows. (a) At least 80 languages, regardless of genetic affiliation of them, express solar and lunar eclipses employing ingestive verbs with morphemes for the sun or the moon and predator nouns which consume it. (b) Ingestive verbs involved in the eclipse expression are almost always 'to eat' or 'to swallow'. (c) Predator nouns involved, in most cases, are some kinds of animals such as a dog, frog, tiger, squirrel, spirit, dragon, snake, centipede, and fish. (d) Languages sometimes exhibit some kinds of asymmetries between solar and lunar eclipses.

The future directions for this study may be summarized as follows. (1) More data are required. This study is far from conclusive as it is based on limited data (about 100 languages). (2) The genesis of the widespread eclipse expression is also an important issue. There are three possibilities: diffusion, chance, or polygenesis (universal). It is highly likely that the resembling expressions for eclipses in East and mainland Southeast Asian languages stem from a common world view that is caused by cultural and linguistic diffusion. It should be noted, however, that similar expressions can be found in other parts of the world as well. For example, Campbell, Kaufman and Smith-Stark (1986:553), which provide semantic calques characteristic of the Meso-American linguistic area, list "eclipse: eat the sun/moon; the sun/moon dies; sun/moon to rot" as an areal calque. It is also pointed out by folklore and mythological studies that folk beliefs that attribute eclipses to consumption of the sun and moon by animals and mythological beings are found around the world, whether they are in relationships of cultural diffusion, accidental similarity, or universal tendency. Thus, Norse mythology tells the story of wolves pursuing the sun and moon to devour them, and a popular belief about eclipses among North American Indian "was that some mythological being, such as Coyote, a bird, a dog, Frog, Lizard, Rattlesnake, etc., was eating the sun or moon. In many tribes noise was made to frighten the aggressor and bring back the sun or moon to life" (Leach and Fried eds. 1972: 337-8). The Sandwich Islander says that "the moon is bitten, pinched, or swallowed" during an eclipse (Wallis 1918: 369).

References

Baccam, Don, Faluang, Baccam, Hung Baccam, Dorothy Fippinger. 1989. *Tai Dam-English, English-Tai Dam vocabulary.* Eastlake: Summer Institute of Linguistics.

Campbell, Lyle, Terrence Kaufman and Thomas C. Smith-Stark. (1986) Meso-America as a linguistic area. *Language* 62.3: 530-70.

Hudak, Thomas J. 1991. *William J. Gedney's the Yay language: glossary, texts, and translations*. Ann Arbor, Michigan: Center for South and Southeast Asian Studies, University of Michigan.

Izushi, Yoshihiko. 1927. Jōdaishina no hi to tsuki to no setsuwa ni tuite [On the sun and moon myths in ancient China]. *The Toyo Gakuho* 16:3. pp.411-26.

Leach, Maria and Jerome Fried (eds.), 1972. *Funk and Wagnalls standard dictionary of folklore, mythology and legend.* New York: Funk and Wagnalls.

Lyman, Thomas A. 1974. *Dictionary of Mong Njua: a Miao (Meo) language of Southeast Asia.* The Hague: Mouton.

Matisoff, James A. 2006. *English-Lahu lexicon*. University of California Publications in Linguistics, Vol. 139. Berkeley, Los Angeles, London: University of California Press.

Means, Nathalie and Paul B. Means. 1987. *Senoi-English, English-Senoi dictionary*. Petaling Jaya: Pelanduk Publications.

Premsrirat, Suwilai. 2002. *Thesaurus of Khmu dialects in Southeast Asia*. Bangkok: Institute of Language and Culture for Rural Development, Mahidol University.

Sao Tern Moeng. 1995. *Shan-English dictionary*. Kensington, Maryland: Dunwoody Press.

Smith, Kenneth D. 2012. *Sedang Dictionary with English, Vietnamese, and French glossaries.* Salaya, Thailand: Mahidol University and Summer Institute of Linguistics.

Wallis, Wilson D. 1918. Prodigies and portents. In James Hastings (ed.), *Encyclopaedia of religion and ethics*. Vol. 10. 362-76. New York: Charles Scribner's Sons.

Wang, Xin. 2015. Tiangou shi ri (yue) kou [A consideration on the dog eats the sun (moon)]. In Kazuhiko Komatsu (ed.), *The Tradition and Creation of Yōkai Culture: From the Viewpoint of Inside and Outside*, 67-81. Kyoto: International Research Center for Japanese Studies.

Wu, Zili and Jiafa Ji (eds.), 2011. *Han-Yi jiangming cidian [Concise Chinese-Yi dictionary]*. Sichuan minzu chubanshe.

Some Notes on *Sun in Sinitic*

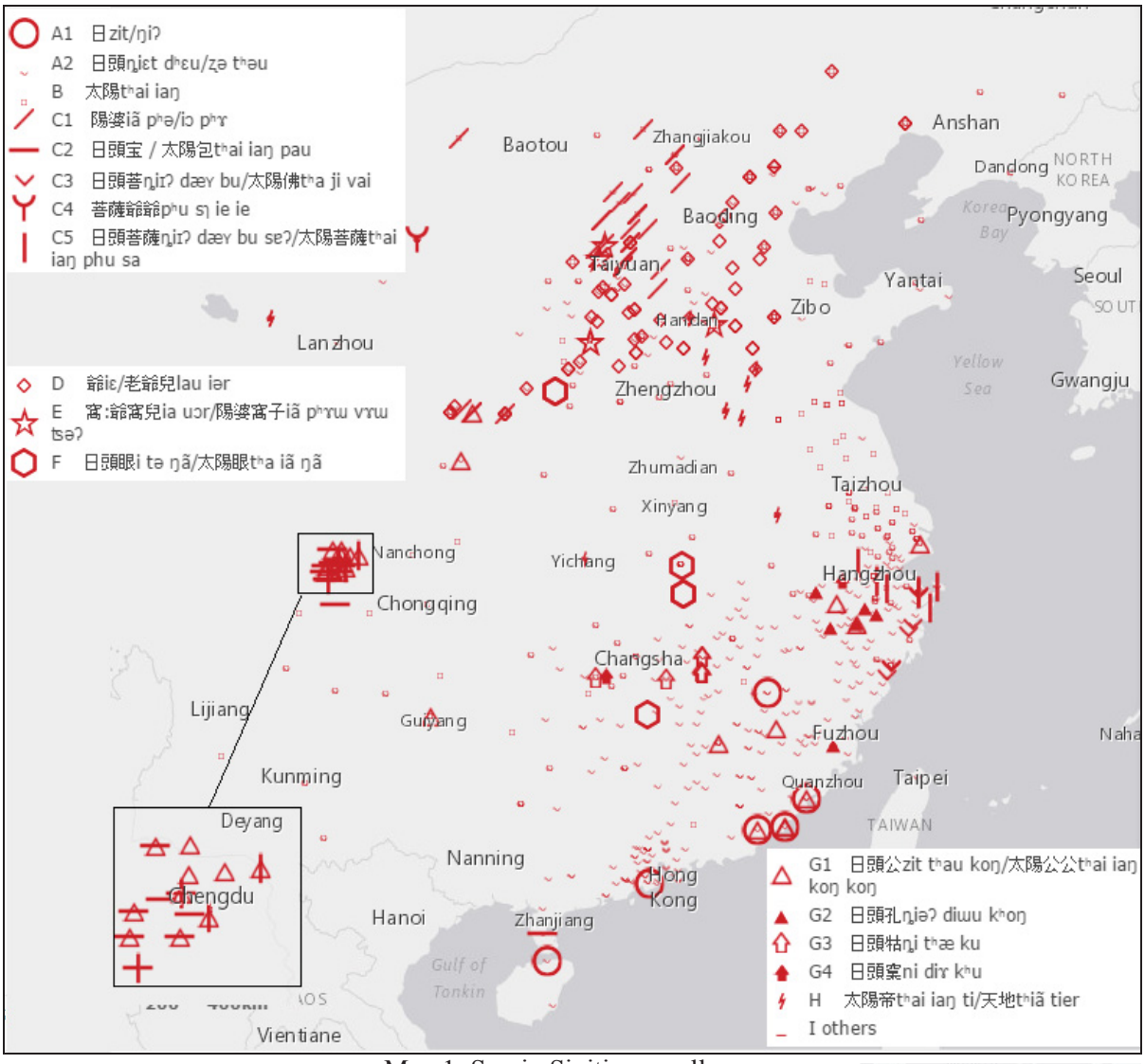
Kenji Yagi

Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies

Abstract

This article discusses three topics related to maps of "sun" in Sinitic and other neighbouring languages. One is the possibility that the form *yang*陽 existed before the appearance of the form *taiyang*太陽, the second is that forms including "eye" are also found in Sinitic, and the last is that the forms including labial onset show continuous distribution with some southern languages; some of these forms may have a common origin.

1 Introduction



Map 1: Sun in Sinitic overall map

The map of "sun" in Sinitic has already been drawn by Cao (2008) and Matsue (2009), and the main distributions of the map are interpreted in Matsue (2009) and Ueya and Yagi (2015). For simplicity, first I provide an outline of the distribution, and then discuss three topics related to other neighbouring languages, especially in the southern region.

Legends and brief explanation
A: *ri*日type　A1: Monosyllabic type of *ri*日　A2: *rtou*日頭, *retou*熱頭, etc.
B: *taiyang*太陽 type
C: *b-*/ *p-* type: Having a labial onset. C1:陽婆,C2:太陽包,C3:日頭菩, etc., see Section 4.
D: *ye*爺 type
E: *wo*窩 type
F: *yan*眼 type
G: *k-* type: Having a onset k- or k^h-

The main categories of words representing "sun" in Sinitic are divided into *ri*日 type and *taiyang*太陽 type (Map 1), which are distributed all over China (although *taiyang* are relatively seldom seen in southern dialects), and many dialects contain both types. The monosyllabic forms of *ri* (Middle Chinese (MC) form reconstructed by Guo (2010) is *nʑiĕt), already found in oracle bone script, are also distributed in some southern dialects; however, many other dialects contain forms like *ritou*日頭 or *retou*熱頭. The form *taiyang* is attested in the literature from the Han period onward.

2　About the origin of *ye*爺 and *yangpo*陽婆

2.1　Continuous distribution of C1 (yangpo陽婆), D (ye爺), E (wo窩), and F (yan眼)

C1(yangpo陽婆) and D(ye爺) types, which have rather concentrated distributions, are distributed around Shanxi or Hebei, as are E (wo窩) and F (yan眼) types, both quite minor. The distribution of C1 and D is approximately complementary, and E and F seem to be distributed in the peripheral area of the distribution of C1 and D (Map 2).

2.2　The sound similarity of *ye*爺 and *yang*陽

	Sound weakening	Paronymic attraction	Folk etymology
陽 MC:jĭaŋ ＞	陽iɐ̃（Daixian代县[1]）	iɐ̃ > ŋiɐr眼	日頭眼ʐʮ tʰəu ŋiɐr（Yongji永济[1]）
	陽yo（Lanxian岚县[1]）	yo > uə 窝	爺窩iɛ uər（Guantao馆陶[2]）
	陽iɛ（Xinxian忻县[1]）	iɛ > ə王（və?）	王子ə zə?（Loufan娄烦[3]）
	陽iã（Jining集宁[4]）	iã > ia爺	爺爺ia ia（Wanrong万荣[1]）
			陽婆
			太陽

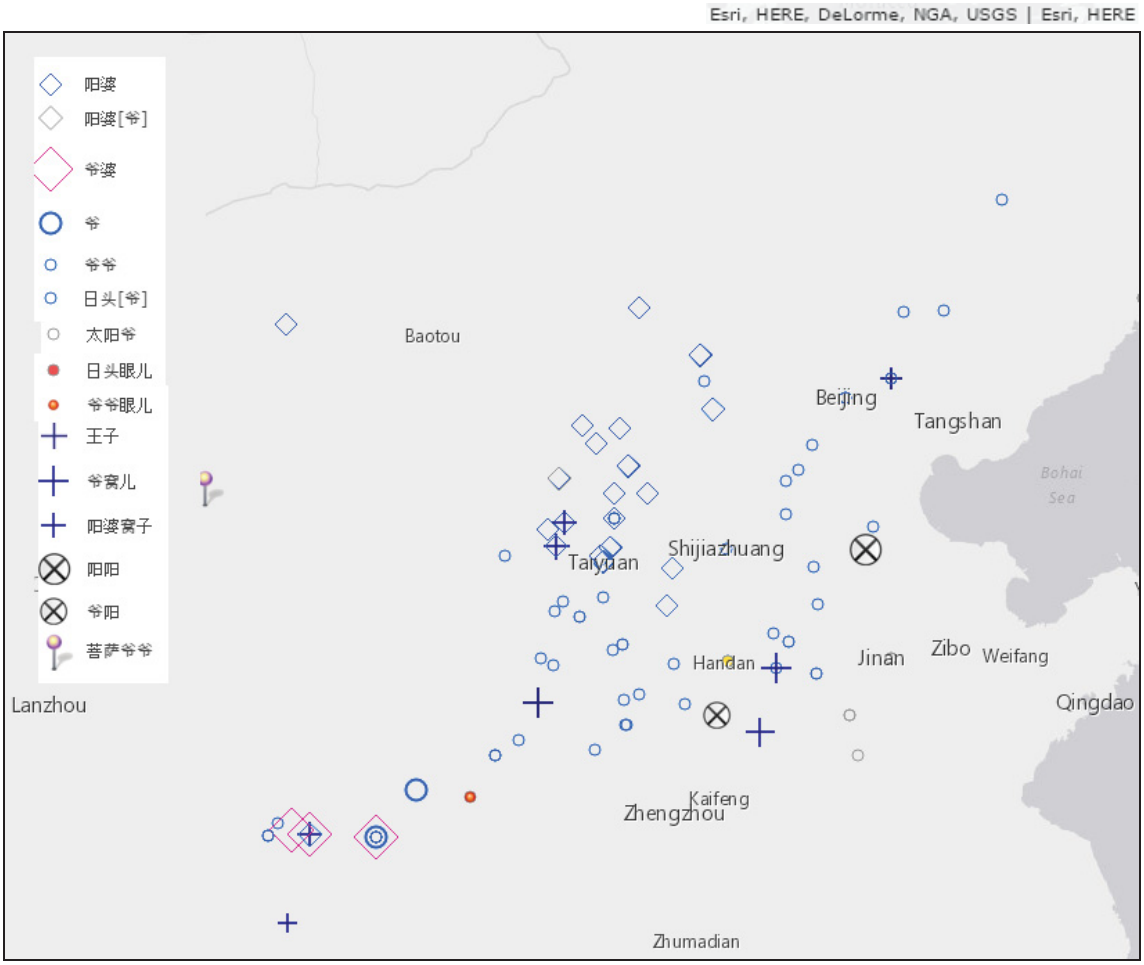
Table 1: Summary of the process of weakening of *yang*陽and appearance of new forms

The sound difference between ***ye***爺(MC: *jĭa[1]) and *yang*陽(MC: *jĭaŋ[1]) in Middle Chinese is the presence of the nasal final "-ŋ". Especially around Shanxi province, the nasal final is frequently weak or

[1]Hou 1993
[2]Hebeisheng 1999
[3]Guo 2005
[4]Chen 1996

dropped; therefore, the 爺 and 陽 sounds are often close to or corresponding to each other. For example, in Xin忻 prefecture, the sound of the "sun" is "iɛ pʰɛ (陽婆)"; and "grandfather" is "iɛ iɛ (爺爺）"(Hou 1993). Here, the 陽 and 爺 sounds completely correspond. Further, in some dialects like Fufeng扶風 (in Shaanxi), 陽婆 is written as 爺婆.

From the continuous distribution and sound similarity of *ye*爺 and *yang*陽, we can presume the possibility that *ye*爺 is derived from *yang*陽. First, the sound of *yang*陽 weakened, and then through paronymic attraction or folk etymology, *ye*爺 appeared. The appearance of E (*wo*窩) or F (*yan*眼) also can be similarly interpreted. A summary of this process is shown in Table 1.



Map 2: The distribution of *ye*爺 and *yangpo*陽婆

3 Distribution of yan眼, kong孔 etc., and cross-linguistic similarity of word formation

Map3 shows the distribution of words related to "eye" or "hole" in Sinitic, Tai-Kadai, and Austroasiatic. The Tai-Kadai map is made by Endo Mitsuaki, and the Austroasiatic map by Kondo Mika.

Legends of Map 3 (in Sinitic)

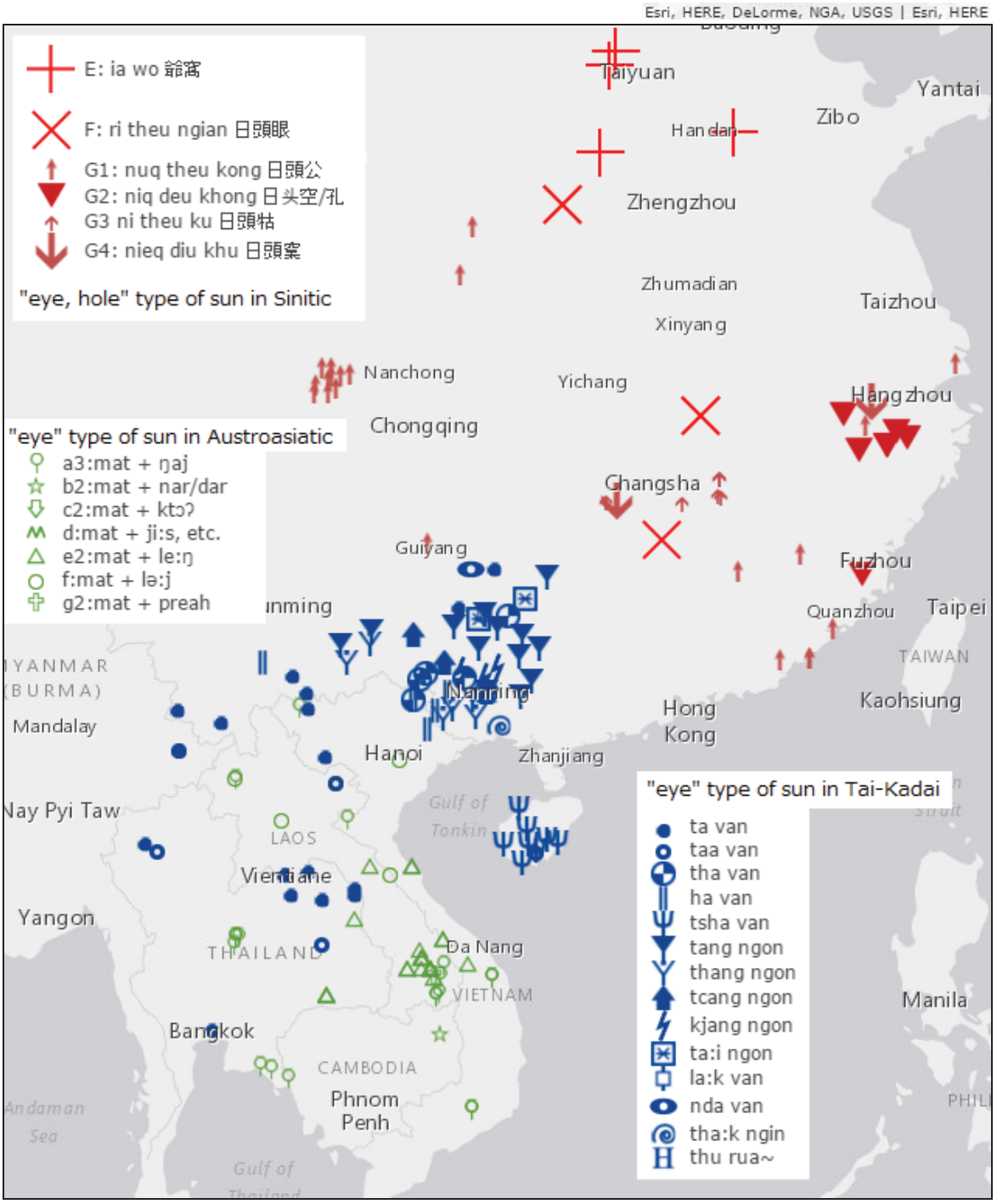
E: *wo*窩type：爺窩児[iɛ uər] 陽婆窩子[iɑN pʰɤɯ vɤɯ tsəʔ]

F: *yan*眼type：日頭眼[i tə ŋaN] 太陽眼[tha ioN ŋɑN]

G：*k*-type：

G1: *gong*公（koŋ）type：日頭公[zik t'au koŋ] 太陽公公[tʰɑ ɦiAN koŋ koŋ]

G2: *kong*空/孔 (kʰoŋ)type：日頭孔[ŋiəʔ tiu kʰoŋ] 日頭空
G3: *gu*牯(ku)type：日頭牯[ŋi tʰæ ku]
G4: *ku*窠(kʰu)type：日頭窠[ni diɤ kʰu]



Map 3: The distribution of "eye" types in Sinitic, Tai-Kadai, and Austroasiatic

3.1 About the "eye" type

In some Chinese dialects, the forms contain *yan*眼"eye" (F type). As mentioned in Endo (2015), forms like "eye" + "day or sky" are frequently observed in Tai-Kadai or Austroasiatic, etc.; therefore, there is

cross-linguistic similarity of word formation. However, in Sinitic, the forms already include a word representing "sun," like *ri*日 or *taiyang*太陽. In addition, the Chinese *yan*眼 can also represent "hole," and forms like *kong*孔 (G2) or *wo*窩 (E), which represent "hole," are found in other dialects, therefore the Chinese *yan*眼 and Tai-Kadai or Austroasiatic "eye" might be etymologically different. The possibility that *yang*陽 changes into *yan*眼 was discussed in section 2.

3.2 About the changing process in Sinitic

G (*k*-) includes elements with several unrelated meanings, such as *gong*公 "gentleman, grandfather," kong孔/空 "hole," *ku*[5]窠"nest," *gu*牯"bull," etc. These forms are taken as one type because of their sound similarity. They all have a syllable with onset *k*- and a rounded back vowel like *-u-* or *-o-*, and are divided into four subgroups based on two criteria: (1) whether aspirated (G2, G4) or not (G1, G3), (2) containing a nasal final (G1, G2) or not (G3, G4).

G1 (*gong*公) has most the widespread distribution. *gong*公 is a kind of reverential form, and the background taking "sun" as an object of worship might relate to its formation.

G2 (kong孔/空) is concentrated in Zhejiang浙江, as is G1, therefore, G2 (considering the "sun" as a "hole in the sky") may be derived from G1 type.

G3 and G4 are also scarcely distributed, and concentrated around Jiangxi江西 or Hunan湖南. In these areas, *gu*牯 is used as a suffix denoting males, therefore, it is possible to think that *gong*公 existed there first, and when the sound and the meaning weakened, *gong*公 became *gu*牯. People later considered *ku* 窠 (also a kind of "hole") more suitable, and G4 was formed.

4 About the *b-/ p-* type

4.1 The distribution of *b-/ p-* type

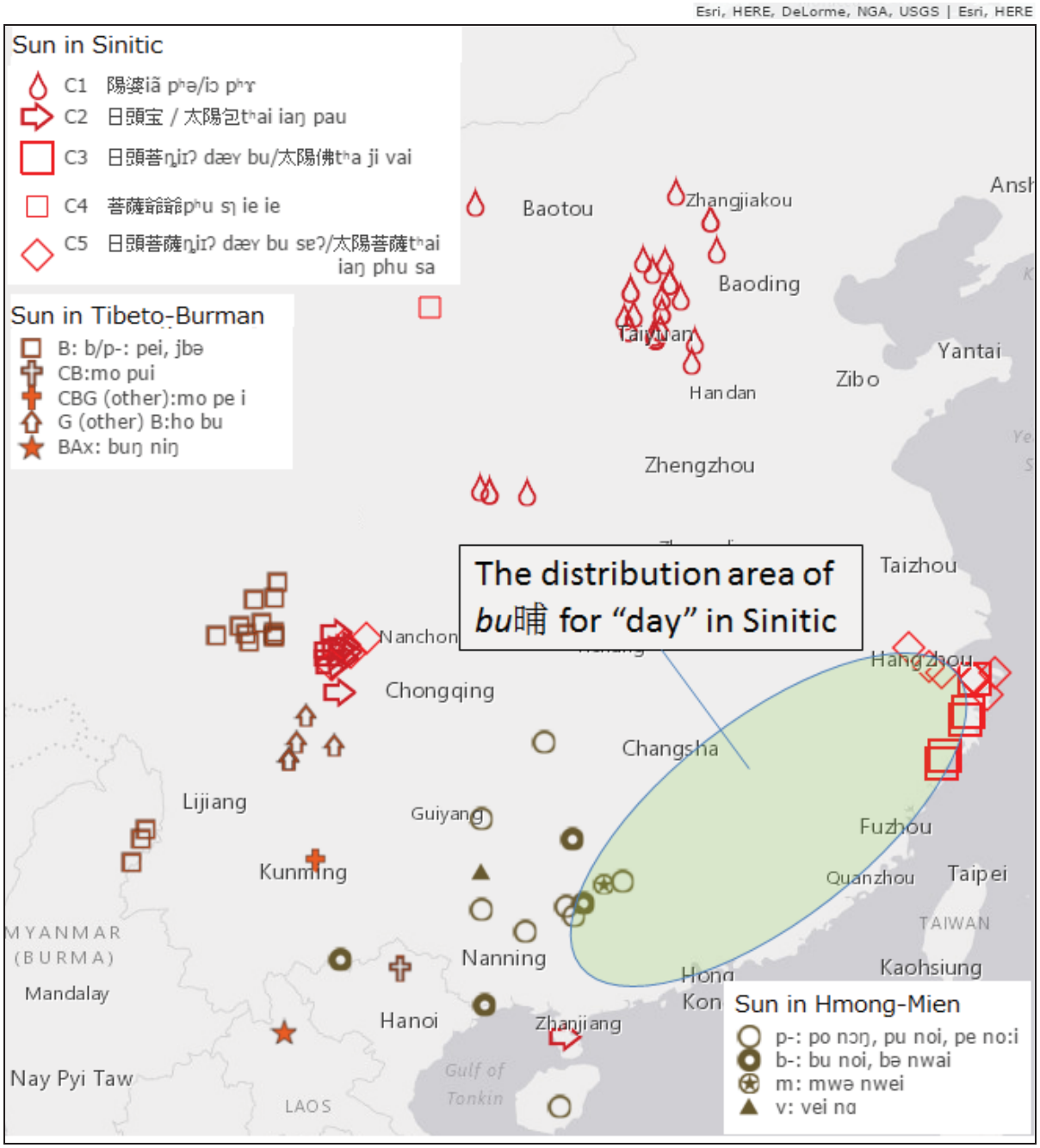
The *b-/ p-* type contains a labial onset syllable like *bao*包 or *pu*菩 in their forms, classified into five subtypes.

C1: *yangpo*陽婆[iɑN pʰə] [iɛ pʰɛ],etc.
C2: *taiyangbao* 太陽包 / *ritoubao*日頭宝[ziek tʰau pɔ] ,etc.
C3: *retoufo*熱頭佛[ŋ̍ieʔ də və] 日蒲頭 [ŋ̍ɪɪʔ bu dœʏ] □日[bi nʌʔ][bɪŋ nʌʔ],etc.
C4: *pusayeye*菩薩爺爺[pʰu sɿ ie ie] ,etc.
C5: *ritoupusa*日頭菩薩[ŋ̍ɪɪʔ dæʏ bu sɐʔ] 太陽菩薩[tʰa ɦiaN bu saʔ],etc.

C1 is mainly distributed in the northern Shanxi area; *po*婆 means "old woman." C2 is seen in Chengdu 成都 and Haikang海康; *bao*包 means a "bundle," however, the cognate labial syllable can also be written as *bao*宝 "treasure." C3 is mainly distributed in the coastal area of Zhejiang province. Many C3 labial onset syllables are written as 佛 (Buddha). However, there are also forms like 蒲[bu] or [bi], which have unclear etymologies. C4 and C5 contain *pusa*菩薩, "a Buddhist saint," in the front (C4) or back (C5) of the forms.

*fo*佛(Buddha) or *pusa*菩薩 is the Buddhists' object of worship, so it is clear that these forms originate from a social back-ground considering "sun" as an object of worship. However, it is not clear why they contained the Buddhist term. Now, I wish to propose the following possibility: a form like *bu* existed at first to represent the sun, and *ri* or *taiyang* followed and stayed as a stem of the form.
People forgot the original meaning of *bu* and newly interpreted it as *fo*佛 or *pusa*菩薩 depending on the sound similarity.

[5] The standard *pinyin* form of 窠 is *ke,* but I refer to the dialectal form here.



Map 4: The distribution of *b-/ p-* type in Sinitic, Tibeto-Burman, and Hmong-Mien

4.2 Continuous distribution of *b-/ p-* types between Sinitic, Tibeto-Burman, and Hmong-Mien

The most important fact supporting this possibility is that forms with *b-/ p-* onset syllables are also seen in Tibeto-Burman or Hmong-Mien, and their distribution are adjacent, especially in the Sichuan area.

In Tibet-Burman, *b-/ p-* onset syllables are found in forms like "pêi, paF, jbə, ɣbə, ʁmbɣi etc". (B(*b/p-*) type in Shirai, et al.2015). Forms with *b-/ p-* onset syllables found in Sinitic or Hmong-Mien are all polysyllabic, and monosyllabic forms are only found in Tibeto-Burman, so these Tibeto-Burman monosyllabic forms may maintain the oldest form.

In Hmong-Mien, *b-/ p-* onset syllables are found in the previous positions of disyllabic forms like bu noi, pu n̥oi, bə nwai, and po n̥ɔŋ. The corresponding forms of the latter syllable, like nuai, n̥u, and noŋ are used solely to represent "sun" in other Hmong-Mien dialects (Taguchi (2015)).

4.3 Distribution of *bu*晡 representing "today, the day before yesterday, etc."

In many languages, the word representing “sun” is also used to represent “day” (Fukushima, forthcoming); *ri*日 in Sinitic, and *van* in Tai-Kadai represent both “day” and “sun” as well. Therefore, it is worthwhile to survey words representing “day.”

In southern China, some dialects contain *bu*晡 in the forms representing “today,” “tomorrow,” “yesterday,” “the day before yesterday,” “the day after tomorrow,” and so on (see Iwata 2009; Cao 2008). For example, in Meixian梅県, “today” is represented as [kim pu ŋit]今晡日, “yesterday” as [tshiu pu ŋit] 秋晡日, and “the last day of the year” as [ŋian sam səp pu]年三十晡 (Huang 1995).

I show an outline of the distribution of *bu*晡in map 4. For a precise distribution, see Iwata (2009). This bilabial plosive onset syllable is similar to the labial syllable in the *b-/ p-* types in Sinitic or Hmong-Mien. Besides, the distribution of *bu*晡 adjoins the distribution of Hmong-Mien *b-/ p-* types around Hunan and Guangdong as well as the distribution of Sinitic *b-/ p-* types (C3, C4) around Zhejiang.

This situation can support the idea that southern China had wider distributions of *b-/ p-* types of “sun” in the past.

5 Conclusion

In this article I proposed some new interpretations of “sun” in Sinitic, especially with respect to neighbouring languages.

In Section 2, I pointed out the continuous distribution of D (*ye*爺) and C1 (*yangpo*陽婆), and the sound similarity between them, and proposed the possibility that they have the same etymology “*yang* 陽,” and may stay in an older form than B (*taiyang*太陽) because they still do not have the prepositional element “*tai*太.”

In Section 3, I discussed the formation of G (k-), and pointed out the similarity with Tai-Kadai or Austroasiatic, that they possess the element “eye” in their forms.

In Section 4, I pointed out the continuous distribution of C (*b-/ p-*) type between Sinitic and Tibeto-Burman or Hmong-Mien. At least some of them might have the same origin.

Sinitic languages are distributed over a vast area, and even though they have similar forms, it is uncertain whether they have a common origin. Therefore, we must continue discussing individual dialects.

References

Cao, Zhiyun 曹志耘 2008. 汉语方言地图集·词汇卷，北京：商务印书馆.

Chen, Zhangtai 陈章太and Li, Xingjian 李行健 1996. 普通话基础方言基本词汇集·词汇卷上，北京：语文出版社.

Endo, Mitsuaki 2015, “Sun” in Tai-Kadai : *Studies in Asian Geolinguistics*,1: 18-19

Guo,Xiaozhen and Zhang, Xianping 2005, 楼烦方言研究.山西人民出版社

Guo, Xiliang 郭錫良 2010 漢字古音手冊.商務印書館

Ota, Itsuku太田斎 2005. “媳妇”が関わる民間語源三題―中国語西北方言の例を中心に,神戸外大論叢 56-7:15-41.

Hebeisheng, Guantaodifangzhibianzuanweiyuanhui 河北省馆陶地方志编纂委员会1999. 馆陶县志. 中华书局

Hou, Jingyi 侯精一 1993. 山西方言调查研究报告. 太原 : 山西高校联合出版社.

Huang, Xuezhen黄雪貞1995. 梅県方言詞典. 江蘇教育出版社

Iwata, Ray 岩田礼 2009. 汉语方言解释地图. 東京: 白帝社.

Kondo, Mika 2015 “Sun” in Austroasiatic : *Studies in Asian Geolinguistics*,1:20-21

Matsue, Takashi 松江崇 2009. 太阳 : 汉语方言解释地图. 東京:白帝社.

Shirai, Satoko, Kurabe, Keita, Iwasa, Kazue, Suzuki, Hiroyuki and Ebihara, Shiho 2015. Sun: Tibeto-Burman: *Studies in Asian Geolinguistics*,1: 14-17

Taguchi, Yoshihisa 2015, “Sun” in Hmong-Mien: *Studies in Asian Geolinguistics*,1:12-13

Ueya, Takashi and Yagi, Kenji 2015. “Sun” in Sinitic languages : *Studies in Asian Geolinguistics*,1: 10-11

Wang, Weihui 汪维辉 2014. 说“日”“月”, 中国语言学报,16：73-96.

Kinship Terms in China Guangxi Liujia Dialect

Takashi Ueya [a]

[a] Kyoto University of Foreign Studies

Abstract

This paper discusses the Kinship terms for the Older Generation in Chinese Liujia dialect (六甲话). In this paper, we address two topics. 1) geographical distribution of the words for "Father's Younger Brother (FyB)", and 2) kinship terms for "Father-in-law" and "Mother-in-law".

1 Introduction

The Liujia dialect is a sub-dialect of the Chinese Pinghua dialect(平话), spoken natively by the Liujia ethnic group (六甲人) in the Sanjiang-Dong Autonomous County (三江侗族自治县) of Guangxi Zhuang-nationality Autonomous District(广西壮族自治区). Sangjiang is located in northern Guangxi province. (Map1)

2 Kinship terms for the older generation

Kinship terms for the Older Generation in the Liujia dialect are listed below.

FF	公	[koŋ^{53}]
FM	婆	[po^{343}]
MF	外公	[mui^{22} koŋ^{53}]
MM	外婆	[mui^{22} po^{343}]
F	爷	[ie^{53}]
M	□[1]	[ȵie^{22}]
FeB	大爷	[taai^{22} ie^{53}]
FyB	□	[niaŋ^{33}]
	□	[maan^{35}]
FZ	□	[pa^{33}]
FeBW	□□	[taai^{22} ȵie^{22}]
FyBW	□□	[niaŋ^{33} ȵie^{22}/ maan^{35} ȵie^{22}]
FZH	□□	[ku^{22} ie^{53}]
MB	舅	[kiau^{35}]
MeZ	大□□	[taai^{22} in^{343} ȵie^{22}]
MyZ	姨	[i^{343}]

[1] □ represents dialectical elements which has no appropriate Han characters.

MBW	舅母□	[kiau^{22} mu^{35} ŋ̩ie^{22}]
MZH	姨夫爷	[i^{343} fu^{53} ie^{53}]

Key to abbreviations:

P=parents, F=father, M=mother, B=brother, Z=sister, S=son, D=daughter, G=sibling, C=child, H=husband, W=wife, e=elder, y=younger

3 Kinship terms for FyB -[maan^{35}] and [niaŋ^{33}]-

As noted in the above list, two terms FyB exist in the Liujia dialect. In this chapter, we analyze the process by which the term for FyB came to be in the Liujia dialect, using a linguistic map. Map 2 indicates the geographical distribution of FyB in the Guangxi Chinese dialect. The data used in this map is based on Xie (2007).

Kinship terms for FyB in the Guangxi Chinese dialect can be classified into five types.

Type A： 叔 [ʃuk] [su] [sɔo] etc.
Type B ： □ [man] [mɐi][ma ma] etc.
Type C ： □□ [ie ie]
Type D ： □□ [ne ia]
Type E ： □□ [t͡sie]

In northern Guangxi province, where Sanjiang is located, Type B, C and D are distributed. There seems to be no doubt from geographical distribution that [maan^{35}] and [niaŋ^{33}] have strong relation to Type B and Type D. The detailed data and locational point are as follows.

【Type B】
[man^{42}] (*Lingui Wutong Pinghua* 临桂五通平话) ,
[miɛ^{23} miɛ^{23}] (*Guilin Chaoyang PInghua* 桂林朝阳平话) ,
[ma$^{\sim 53}$] (*Yongfu Taocheng Pinghua* 永福桃城平话),
[mɐi^{21}] (*Rongshui Xiancheng Tuguaihua* 融水县城土拐话),
[ma^{55} ma^{55}] (*Guanyang Wenshi Xiangyu* 灌阳文市湘语)

【Type D】
[ne^{52} ia^{21}] (*Longsheng Hongyao Pinghua* 龙胜红瑶平话)

According to informants, [maan^{35}] has a slightly different meaning from [niaŋ^{33}]. The informants stated that if one's father has younger brothers, one would refer to the older younger brothers as [niaŋ^{33}], while one would refer to the youngest younger brother of one's father [maan^{35}]. In short, [maan^{35}] means "the youngest FyB".

From this information, we can derive the following hypothesis.
(1) In Sanjiang, there is only one term for "FeB". However, it is unclear whether this term is [niaŋ^{33}] or [maan^{35}].
(2) Liujia speakers have taken to using the new term "FeB" in certain occasions. It may be that the term was brought into Sanjiang from nearby areas.
(3) While the terms [niaŋ^{33}] and [maan^{35}] were in "competition" for a while, [maan^{35}] subsequently acquired the new meaning of "the youngest FyB", and since then the terms have coexisted. This situation continues up to the present day.

4 Kinship terms for F/M by marriage

In this chapter, we discuss the kinship terms for F/M through marriage. Married couples in China continue to have different surnames, thus in Liujia dialect, when we refer to F/M by marriage, we must take their surname into consideration.

For example, If EGO's WM has the same surname, MM is regarded as Ego's FZ. At the same time, WF becomes Ego's FZH.

[case1]

EGO's surname = WM's surname

WM → FS (pa^{33})

WF → FSH (ku^{22} ie^{53})

ex.

荣*Rong*（WF）——————侯*Hou*（WM）

|

侯*Hou*(EGO)——————荣*Rong*（W）

On the other hand, if EGO's WM has a different surname, WM is regarded as Ego's MZ. At the same time, WF becomes Ego's MZH.

[case2]

EGO's surname ≠ WM's surname

WM → MS (taai^{22} in^{343} ȵie^{22})

WF → MSH (i^{343} fu^{53} ie^{53})

ex.

荣*Rong*（WF）——————吴*Wu*（WM）

|

侯*Hou*(EGO)——————荣*Rong*（W）

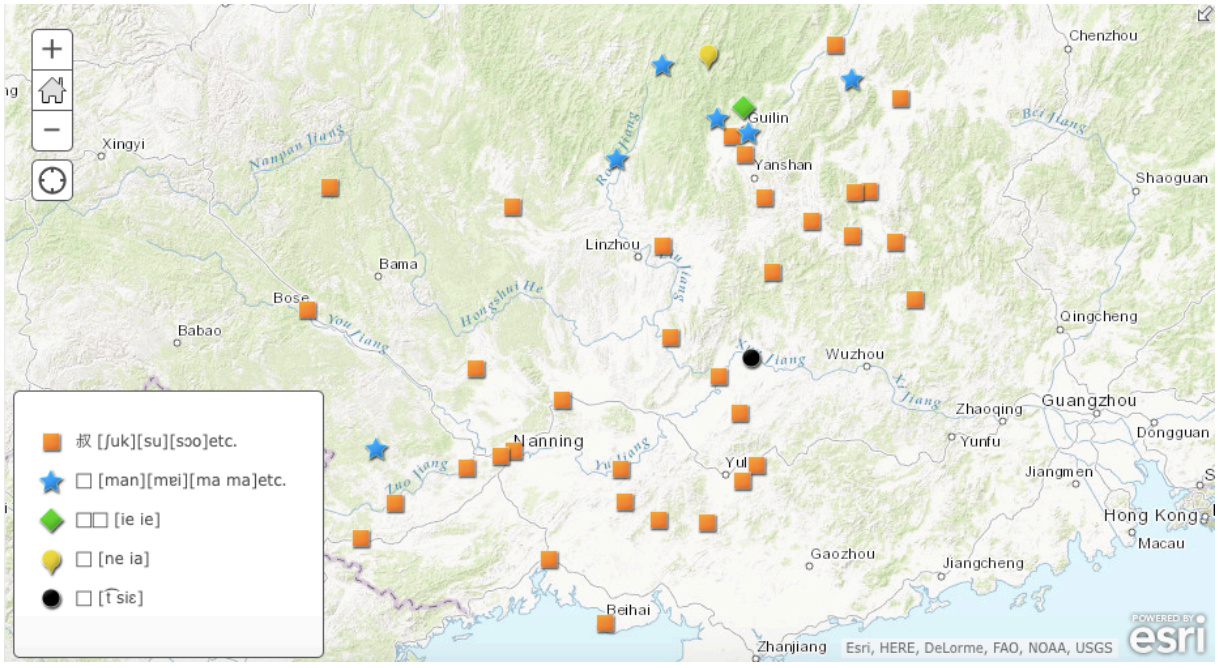
5 Conclusion

This paper addressed the process of formation of FyB in the Liujia dialect. Most notably, there exist two terms for FyB, namely [maan^{35}] and [nianʃ^{33}]. It is highly possible that these two forms originate in Chinese minority languages. I would like to address this question in in more detail in future research.

References

XIE, Jianyou 谢建猷. 2007. Studies on the Han Chinese dialects of Guangxi 广西汉语方言研究 ,Guangxi renmin chubanshe.

Map 1 Guangxi dialect of Chinese



Map 2 The geographical distribution of FyB in the Guangxi

Basic color terms in Ainu

Mika Fukazawa

Graduate School of Humanities and Social Sciences, Chiba University
/JSPS Research Fellow

Abstract

This is the preliminary study of the basic color terms (BCTs) in the Ainu language. We will review what colors are "basic" in the language and culture and test the validity of the current evolutionary sequence proposed by Kay, Berlin, Maffi and Merrifield (1997). Ainu belongs to the minor line of the sequence, which is related to the generic puzzle regarding the origin of yellow/green terms. In this study, we drew the geographical distribution of the color foci of yellow/green/blue (Y/G/Bu) terms and verified the "Composite Category Rule" and the evolution of the BCTs. The result for the developing process of Y, G and Bu in Ainu aligns with Kay et al.'s (1997) sequence: Stage I and II (unknown) > Stage III (Y/G/Bu) > Stage IV (Y and G/Bu) or (Y/G and Bu) > Stage V (Y, G, and Bu). In the Ainu culture, besides black, white, red and Y/G/Bu, the colors of gold and silver have their own words, which were borrowed from Japanese before the other intersections of primaries, e.g., pink, orange, brown and so on.

1 Introduction

Since Berlin and Kay (1969) proposed the influential hypothesis of "basic color terms" (BCTs), empirical studies and the World Color Survey (WCS)[1] have developed the categorization and the evolutional sequence of BCTs. Berlin and Kay's original sequence of BCTs is as follows: I – white and black, II – red, III – green or yellow, IV – green and yellow, V – blue, VI – brown and VII – purple, pink, orange and gray. Against stage I (white and black), Heider insisted the Dugum Dani has two term systems, *mili* 'dark; cool colors' and *mola* 'light; warm colors' (Heider 1972; Heider and Oliver 1972). Regarding to stages III to V, Kay (1975) introduced the new category, "grue," which should not divide into green and blue or be labeled with two basic color terms.[2] Kay and McDaniel (1978) suggested four "composite categories": "warm" (red/yellow), "light-warm" (white/red/yellow), "dark-cool" (black/green/blue) and "grue" (blue/green). They also suggested "derived categories," i.e. brown, orange, pink, purple and gray, which encode the intersection of the six primary categories: white, black, red, yellow, green and blue. For example, the derived category of pink encodes the intersection of red and white.

Figure 1 shows the evolutionary typology of BCT systems based on the WCS data (Kay, Berlin, Maffi and Merrifield 1997). Ninety-one of the 110 WCS languages (83%) belong to the middle row of Figure 1, which is the "main line" of evolutionary development. Ainu was known as a $III_{Y/G/Bu}$ language that has one term for the yellow/green/blue (Y/G/Bu) category (c.f. Hattori 1964). Only two of the $III_{Y/G/Bu}$ languages, Karajá and Lele, are also found in the WCS data. There is not enough evidence to suggest a scenario for the origins of Y/G composites, and we still need to discover the solution of the Y/G mystery,[3] since Y and G are separate composites in stage I of the main line. I will focus on the BCTs and the historical evolution of Ainu as a stage $III_{Y/G/Bu}$ language. Although it is difficult to do

[1] "This project, a collaboration of researchers at the University of California, Berkeley, and the Summer Institute of Linguistics (SIL), collected colour-naming data from 110 languages being studied by SIL linguist-translators in the late 1970s." (Kay and Maffi 2013)

[2] One of the evidences was that Japanese *ao* represents both 'green' and 'blue.'

[3] Maclaury (1987) introduced a "yellow-with-green" category in Shuswap and suggested the idea that it diffused among Salishan languages as a concept that yellow and green belong together.

fieldwork on this language at present, I will try to draw the historical outline of the BCTs in Ainu through philological investigation and the geographical distribution of the terms.

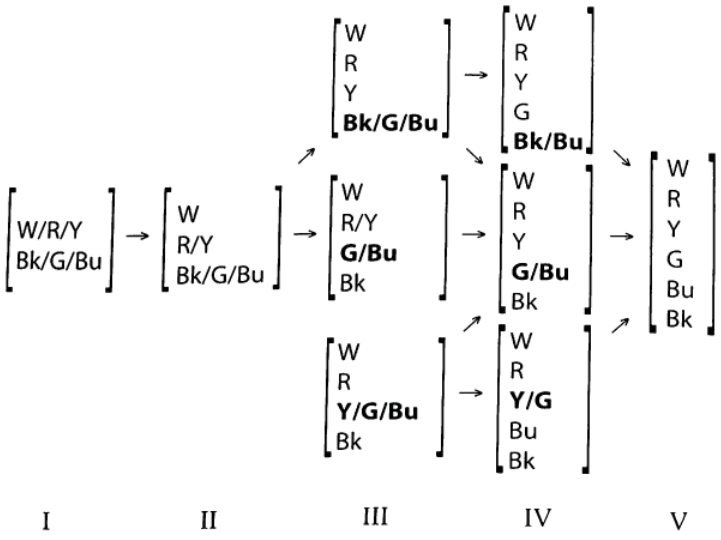


Figure 1: Types and Evolutional Stages of Basic Color Term Systems;
Kay, Berlin, Maffi and Merrifield (1997) [4]

2 What are "the basic color terms" in Ainu

The Ainu language[5] is one of the indigenous languages in Japan, spoken throughout Hokkaido, Sakhalin, and the Kuril Islands. It has five vowels /i, e, a, o, u/ and eleven (or twelve including a glottal stop) consonants /p, t, k, c, s, m, n, r, w, y, h, (')/. There are few studies of BCTs in Ainu, but Hattori and Chiri (1960) and Hattori (1964) documented the rich words for color in the each dialect. In addition to this, Tamura (1974: 26) mentioned the BCTs and different color names in Ainu. Table 1 shows Tamura's description of the BCTs from her field data.

Term	Gloss	Reference
kúnne	'black, dark'	black~dark gray; dark colors: dark brown, dark blue, dark purple etc.
retár	'white'	white ~ light gray; light colors: light yellow, light pink, light blue etc.
húre	'red'	red ~ reddish brown/maroon; reddish purple ~ orange
síwnin	'green'	light green ~ deep green; yellow ~ bluish purple

Table 2: Basic Color Terms in Ainu

According to Tamura (1974: 27), the terms for BCTs with the suffix *-no* stand for their color foci, "ideal colors," and those terms using the prefix *ru-* represent the marginal range of the colors as "non-ideal colors": i.e. blackish, whitish, reddish and greenish. Some derived categories may be expressed with the prefix *ru-* '-ish': *ru-hure* means 'pink' and 'orange' and *ru-siwnin* means 'light blue' in the Obihiro dialect (Hattori 1964; Sawai and Tamura 2005). McNeil (1972) also referred to the BCTs of Ainu and

[4] W = white; Bk = black; R = red; Y = yellow; G = green; Bu = blue.

[5] The language is a polysynthetic SV/AOV, basically head-marking type and allows 'pro-drop,' which is compensated with verbal agreement (singular/plural forms and personal affixes/clitics).

has been cited many times (c.f. Moss 1989), however, his description includes some errors. In the next section, I will correct them and draw the geographical distributions of the BCTs in Ainu.

3 Geographical Distributions of the BCTs in Ainu

Here, we will discuss (i) the etymology of the BCTs in Table 1 above, (ii) the foci of the terms for yellow/green/blue and (iii) the evolutionary sequence of BCTs in Ainu.

3.1. Black

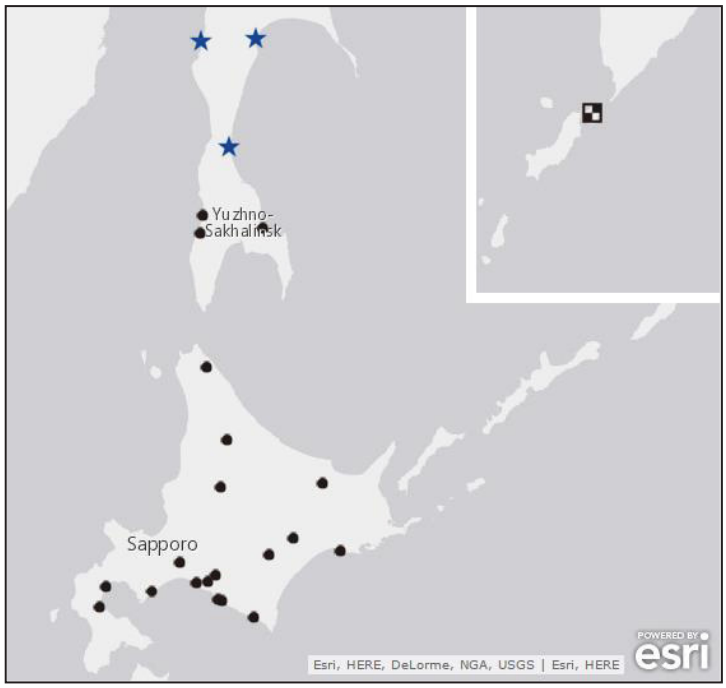


Figure 2: Black

- *kúnne*:
  Hokkaido (without unaccented dialects)
- *kunne*:
  Sakhalin (Ochiho, Tarantomari, Maoka)
  Hokkaido (Samani, Bihoro, Kushiro and Shizuai)

★ *kurasno*:
Sakhalin (Shiraura, Raichishka, Nairo)

ekuroku/ekuroko:
Northern Kurils (Shumshu)

The majority of Hokkaido dialects have a pitch accent,[6] while in the Sakhalin dialects, vowel length is distinctive (Chiri 1942 and Tamura 2000). The word *kúnne* for 'black' in Hokkaido corresponds to *kunne* in Sakhalin[7] (Hattori and Chiri 1960). Etymologically, Hattori (1999 [1959]: 152) suggested that *kunne* originated from √*kur* + the copula *ne*. The root *kur* is originated from the noun *kur* for 'shadow.'

The words in Sakhalin and Northern Kurils, *kurasno* and *ekurok*(*u/o*), seem to have the root √*kur* 'shadow.' The suffix *-asno* is parallel with *-asnu* in Hokkaido, which attaches to roots or bases, and creates adjective-like intransitive verbs, as in (1).

(1)	a. *kamásnu*	*kam* ('meat') *-asnu*	'to have a clear/beautiful skin'
	b. *tumásnu*	*tum* ('power') *-asnu*	'to be healthy; strong.'
	c. *monásnu*	*mon* ('hand') *-asnu*	'to be a quick worker'
	d. *kurásnu*	*kur* ('shadow') *-asnu*	'to be iridescence'

(Tamura 1996; the Saru dialect)

Tamura (1974: 26) said that in Hokkaido, the term *ekúrok* is "an emotional expression" and means 'to be deep black; to be completely dark.' As Hattori (1999 [1959]) and Tamura (1974) mentioned, if these three terms—*kúnne*, *kurasno* and *ekuroku/ekuroko*—include the root √*kur* 'shadow,' the term for 'black' would be encoded by the low brightness. In Raichishka (Sakhalin), there is also the term *kunne* in addition to *kurasno*, but it is specially used by old people. The term *kúnne* is older than the terms

[6] The dialects of Samani, Bihoro, Kushiro and Shizunai in Hokkaido have no accent (Hattori and Chiri 1960).

[7] In addition to this, Hattori and Chiri (1960) recorded the word for 'black' as *kuɴne* in Ochiho, Sakhalin. Sometimes, maybe idiolectally, [ɴ] is pronounced as /n/, but not [n] in Ainu.

kurasno and *ekuroku/ekuroko*, which may be newly selected as forms for 'black' in Sakhalin and Northern Kurils, respectively.

3.2. White

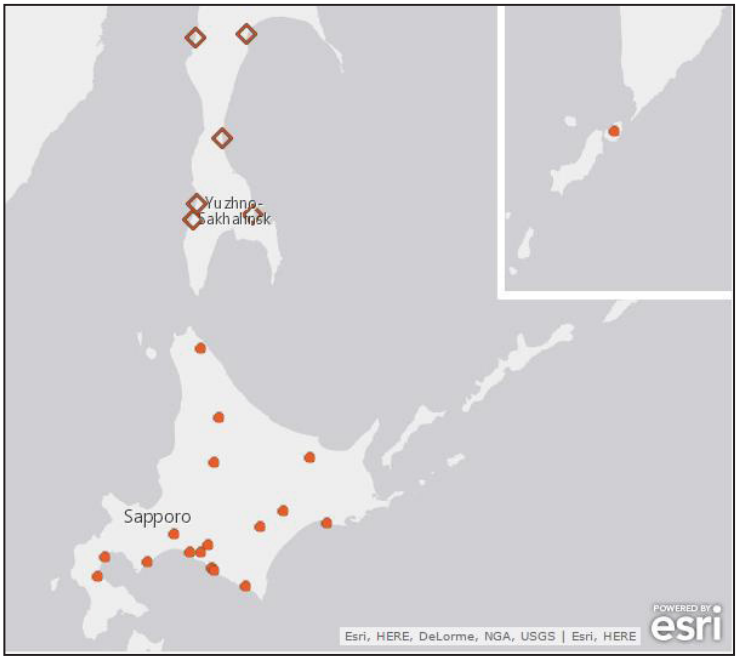


Figure 3: White

retár:
Hokkaido (without unaccented dialects)
retar:
Hokkaido (Samani, Bihoro Kushiro and Shizunai)
retara:
Northern Kurils (Shumshu)

tetara:
Sakhalin

According to Chiri (1942) and Tamura (2000), the coda /r/ in Hokkaido has changed to /h/, or /rV/ in Sakhalin. Following this rule, the term *retara* is expected in Sakhalin dialects, but it does not exist. This dialects have the term *tetara*, even though there is no generic rule of the onset /t/ and /r/ as a phonetic correspondence between Sakhalin and Hokkaido. In Hattori and Chiri (1960), only the Nairo dialect (Sakhalin) tends to have the onset /t/, which corresponds to /r/ in the other dialects. Moreover, I will mention the previous studies of *tetara*. Kindaichi (1931: 10-11) described that investigators often mistake *retara* for *detara* because this is an intermediate sound between /t/ and /d/ and /r/. Chiri (1942: 462) explained that in Sakhalin, the onset /r/ is confused with /t/ or /d/: e.g. *retara* > *tetara* 'white.' We can see both forms in the writings record of Kurils, *retara* (Torii 1903) and *tetar* (Dybowski 1982).

3.3. Red

A uniform distribution of 'red' is shown on the map because the terms *huure* and *húre* comply with the accentual pattern in Sakhalin and Hokkaido that we introduced in section 3.1 above. In Tamura (1974), *hure* encodes the ranges of red ~ reddish brown/maroon and reddish purple ~ orange, and the focal color is just red like the color of blood. According to McNeil (1972), the term *hu* represents 'red-green,' because Chiri (1953) shows the term *hu* for 'row,' which is translated in composite words as Japanese *kanji* 青 (blue): *hu-ham* for green leaf (Japanese:青葉) and *hu-kina*/*hu-mun* for green grass (Japanese: 青草). However, this includes three matters at least. First, Japanese *ao* (青) reminds us of 'fresh' and the color of grue (green and blue). It cannot be equal to the terms blue and/or green in English. Second, Chiri (1953) does not refer to *hu* as the color of 'blue/green.' He just explains *hu* means 'row,' and the opposite is the term *ci* for 'mature/die.' Third, it is a serious mistake to refer to *hu* as the term for 'red.' The term *hure* cannot be separated into *hu* and *re*. If the term *hu* were 'red,' *re* would be the causative suffix and *hure* would mean 'to make it red.' The term *hure* is in fact an intransitive verb.

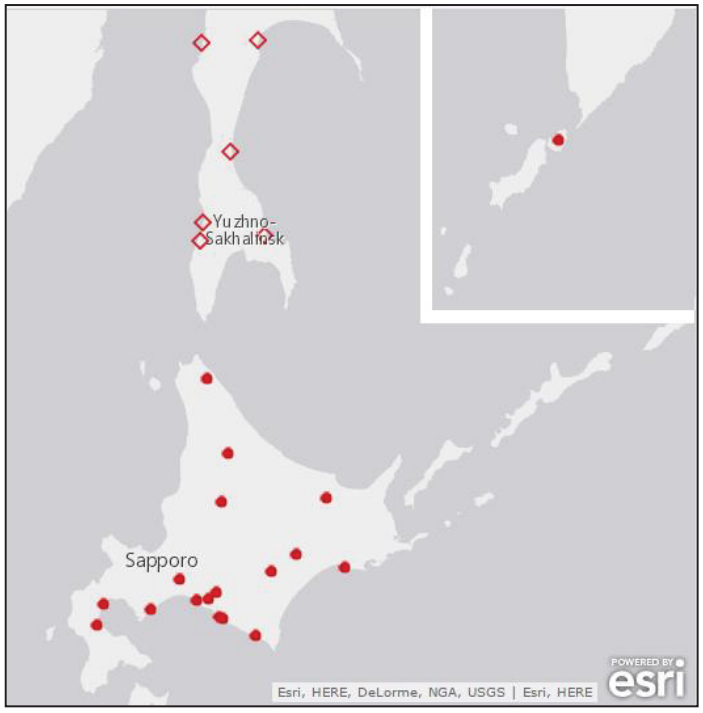


Figure 4: Red

- húre:
  Hokkaido (without unaccented dialects)
- hure:
  Hokkaido (Samani, Bihoro Kushiro and Shizunai)
  Northern Kurils (Shumshu)

◇ huure:
Sakhalin

3.4. Yellow/Green/Blue

In Ainu, besides the terms *kunne* 'black', *retar* 'white' and *hure* 'red' above, the term *siwnin* (*suynin* in Bihoro) is assumed to be one of the BCTs. McNeil (1972) presented *siwnin* 'blue-yellow' as a BCT. Hattori (1964) recorded that in some dialects, 'reddish yellow' is expressed by *hure*, while 'bluish yellow' is expressed by *siwnin*/*suynin*. Moreover, Tamura (1974) notes the dialectal differences in the foci of *siwnin*/*suynin*: 'light blue' and 'yellowish green.' For example, in Biratori (Hokkaido), the prototypical color is a grass green, but in Bihoro (the northeastern Hokkaido) it is a blue such as the color of sky and the sea (Tamura 1996; Hattori 1964). Figure 5 shows the distribution of the terms for 'green/blue/yellow,' and Figure 6 plots their color foci on the map.

Figure 5: Green/blue/yellow

- síwnin:
  Hokkaido (without unaccented dialects)
- siwnin:
  Hokkaido (Samani and Shizunai) and Sakhalin

◇ suynin:
Hokkaido (Bihoro and Kushiro)

☆ teunin:
Northern Kurils (Shumshu)

(no data): Hokkaido (Nukibetsu)

In Figure 5, the pair of *iw* and *uy* in the terms *siwnin* and *suynin* is a characteristic metathesis in Ainu. The Bihoro and Kushiro dialects tend to have *uy* terms, e.g. *riwka*:*ruyka* 'bridge,' *nociw*:*nocuy* 'star,' *ciw*:*cuy* 'to stab' and so on. In Northern Kurils, the term *teunin* for 'green' might not have originally displayed the focal color of 'green,' since in Hokkaido, the term *téwnin* means 'sparkle' or 'twinkle.' This may be one piece of evidence for "brightness naming" in MacLaury (2005).[8]

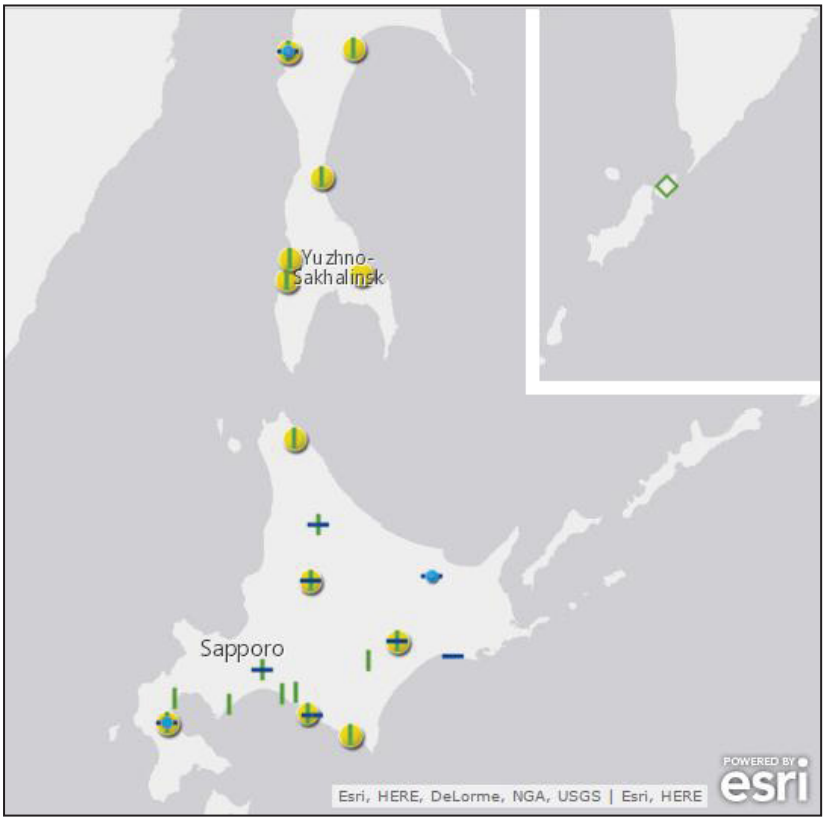


Figure 6:
Yellow/Green/(Light) Blue;
The foci of *siwnin*/*suynin*
and *teunin*

Blue
Light blue
Green
Green (< sparkle; twinkle)
Yellow

Now, let us look at the map of Figure 6, in which we indicated the color foci. There are many dialects that have a yellow-green(-blue) category, while no dialect has a yellow-blue (i.e. without green) category. When the term can indicate a light blue, it is also possible to indicate just a blue. This can be explained by Figure 7 and the "Composite Category Rule," that is, "A possible composite category is any fuzzy union of a subset of [primary categories] which, in Figure [7], forms an unbroken associational chain not crossing the diagonal line" (Kay, Berlin and Merrifield 1991: 16).

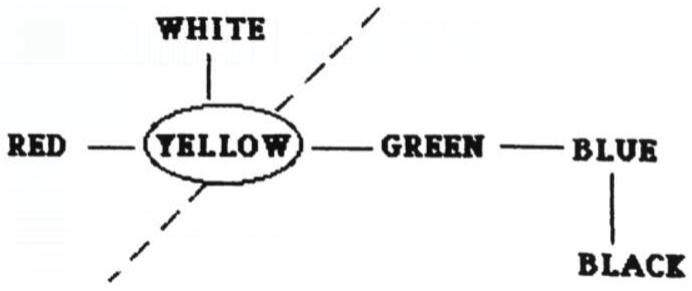


Figure 7

[8] With his empirical data of Lele (Afro-Asiantic), Stage $III_{Y/G/Bu}$ languages are classified into the type of "brightness naming" opposed to "hue naming." MacLaury (2005: 209) reported "The Lele speaker focuses the middle brightness range of *bole* in green [...] and extends it throughout the colors that an English speaker calls yellow, red, and blue."

Roberson (2005: 65) also gives a detailed account of color categorization as follows; "No language has ever been reported to have a category that includes two areas of color space (e.g. yellow and blue) but excludes an area between them (green). There is no associative chain of similarity that could connect yellow to blue without passing through green."

Note that even if a dialect is displayed with a stage V icon (●, | or —), it may in fact be positioned in stage III in Figure 6. We assume that all of the dialects are situated in stage III and then marked only with the focal color on the map. This strategy would be successful for finding the evolutionary development of the Y/G/Bu category in the Ainu language. The marginal dialects such as ABA distribution have the stage III icon, while, as Tamura (1974) mentioned before, the focal color of *siwnin/suynin* is blue in the eastern Hokkaido dialects, and it is green in the western dialects. The stage IV icons are sandwiched between the icons of stages III and V. This is the evidence for the intermediate stage. Figure 8 shows the two possible developmental paths of Y/G/Bu systems, adapted from Figure 1 (Kay, Berlin, Maffi and Merrifield 1997) above.

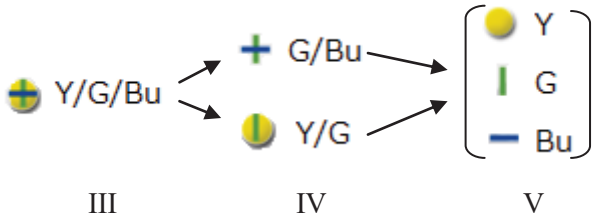


Figure 8

Ainu can be considered as a non-partition language that conforms to the Emergence Hypothesis (EH), according which "not all languages necessarily possess a small set of words or word senses that each signify a color concept and whose significata jointly partition the perceptual color space" (Kay and Maffi 1999: 744; Kay et al. 2009: 9). Moreover, Kay and Maffi (1999: 755) and Kay et al. (2009: 41) suggested "a plausible solution to the apparent mystery of Y/G composites":

> EH languages may develop somewhat along the lines of Yélî Dnye, assigning basic terms, according to principles [Black & White and Red], only to restricted [Black, White, and Red] violating Partition. Subsequently, Partition comes into play and a Y/G/Bu term appears, covering the remaining primary colors. There is some suggestive that no reliable conclusion can be drawn here. In some cases, the Y/G/Bu term may then divide into Bu and Y/G terms.

The term *siwnin*/*suynin* would be etymologically related to the term *siw*/*suy* 'bitter, sour in taste.' In some dialects, *siwnin* itself can express this meaning. If it originated from the color of an 'unripe or immature (plant),' it is no wonder that the Y/G/Bu composites occur in order to cover the remaining primaries.

4 Non "Basic Color Terms" — But Essential Colors

What is the 'color' for Ainu people? The term for 'color' in Ainu is not an original word, but rather *iro* (the possessive form *iroho*), which is borrowed from the Japanese *iro* (色). In the Ainu cluture, bark used to make fabric and textiles is dyed the color of black, red or Y/G/Bu with bark or grass. One common linguistic feature of the BCTs is that they have transitive forms with a causative suffix *-ka*: *kunne-ka*, *retar-ka*, *hure-ka* and *siwnin-ka* 'to make it black, white, red and Y/G/Bu.' This is why we suggest that these colors are the BCTs for the Ainu language and culture.

As for the other essential colors, Refsing (2013: 220) mentioned, "besides black and white and possibly gold and silver, there are only two words for colors: *hure* (red) and *siwnin* (yellow) in the Ainu language." Although the terms for 'gold' and 'silver' may indicate their minerals, certainly these color terms appear frequently in Ainu oral literature and history. In some cases, they might intend to show

metonymically gorgeous, luxury and beauty. Only these colors are clearly expressed by borrowing words from Japanese *kogane* (黄金) and *sikokane* (白金), as compared to *konkane*/*konkani* and *sirokane*/*sirokani*. The beginning sentence of the first story in Yukie Chiri's (1978[1923]: 10–35) "Ainu Shin'yoshu (A collection of the Ainu epics of the gods)" would be famous in Ainu literature: "sirokanipe ranran piskan, konkanipe ranran piskan"—"(Silver raindrops are falling around and gold raindrops are falling around)."

5 Conclusions

This study has investigated the BCTs in the Ainu language and their geographical distribution. Ainu has four BCTs: black, white, red and yellow/green/blue (Y/G/Bu). The terms for 'white' and 'red' show uniform distributions, while the terms for 'black' show a dialectal distribution. The older form for 'black' would be *kúnne*, and then *ekuroku*/*ekuroko* and *kurasno* appeared in Northern Kurils and Sakhalin, respectively. These forms take the root √*kur* 'shadow' and originate from the meaning of 'dark.' The terms for 'Y/G/Bu' are not very different formally, but the foci vary from dialect to dialect. The result would not deny the current evolutionary sequence of WCS (Kay et al. 2009); however, the distribution of the foci is suggestive for the other non-partition (i.e. EH) languages. It seems that the Y/G/Bu terms in Ainu may be related to the word for 'bitter' and they may originally represent the color of an 'unripe or immature (plant).' Interestingly, *konkane*/*konkani* 'gold' and *sirokane*/*sirokani* 'silver' are also essential color terms in the Ainu culture. These colors possess their own terms, which were borrowed from Japanese prior to other terms that name the fuzzy intersections of primaries: brown, pink, purple, orange, and gray.

At present, we cannot investigate the Ainu language in detail in the field, because a few "native speakers" only know limited words and cannot speak this language freely. However, young generations of Ainu people are now learning it as second language; they are writing articles in Ainu and speaking it in their community. They value their own dialects, and to some extent, have positive attitudes toward newly coined words and borrowed words from Japanese. The movement and the study of the Ainu language has entered a new phase. We will observe, in the future, how Y/G/Bu terms will be developed and which terms the Ainu people will select for foci of Y, G and Bu in their current society.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 16J01055.

Appendix

List of Used Data:

Dialects	Materials
Yakumo, Horobetsu, Asahikawa, Nayoro, Bihoro, Sooya and Raichishka	Hattori and Chiri (1960) and Hattori (1964)
Oshamambe, Nukibetsu, Niikappu, Samani, Ochiho, Tarantomari, Maoka, Shiraura and Nairo	Hattori and Chiri (1960)
Biratori and Fukumitsu	Hattori and Chiri (1960), Hattori (1964) and Tamura (1996)
Obihiro	Hattori and Chiri (1960) and Sawai and Tamura (2005)
Kushiro and Shiranuka	Hattori and Chiri (1960) and Fujimura (1986)
Chitose	Nakagawa (1995)
Shizunai	Okuda (1999)
Mukawa	Nakagawa (2014)
Hombetsu	Sawai (2006)
Shumushu (Northern Kuril)	Torii (1903) and Murayama (1971)

The Atlas of Ainu Dialects

Hokkaido:
1. Yakumo, 2. Oshamambe, 3. Horobetsu, 4. Biratori, Fukumitsu, 5. Nukibetsu, 6. Niikappu, 7. Samani, 8. Obihiro, 9. Kushiro/Shiranuka, 10. Bihoro, 11. Asahikawa, 12. Nayoro, 13. Sooya, 14. Chitose, 15. Shizunai, 16. Hombetsu, 17. Mukawa, 18. Nemuro

Sakhalin:
19. Ochiho, 20. Tarantomari, 21. Maoka, 22. Shiraura, 23. Raichishka, 24. Nairo

Kuril Islands:
25. Shumshu

References

Berlin, B. and P. Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Chiri, M. 1942. Ainu gohoo kenkyuu: Karafuto hoogen wo chuusin toshite [The study of Ainu grammar: Specially for Sakhalin dialects]. *Chiri Mashiho chosakushuu* [*Collected works of Chiri Mashiho*], 3, 1973[1993]. Tokyo: Heibonsha.

Chiri, M. 1953. Bunrui Ainugo Jiten: Shokubutsu-hen [Classified Ainu Dictionary: The Plant volume]. *Nihon joomin bunka kenkyuu-jo Ihoo* [*Institute for Japanese Folk Culture Bulletin*], 64. Kanagawa: Institute for Japanese Folk Culture.

Chiri, Y. 1978[1923]. *Ainu Shin'yoshuu*. Tokyo: Iwanami Shoten.

Dybowski, B. S. 1982. Słownik narzecza Ainów, zamieszkują wyspę Szumszu w łańcuchu Kurylskim przy Kamczatce. *Kitachishima Ainugo* [*Northern Kuril Ainu*], 134–244, Murayama, S., 1971. Tokyo: Yoshikawa Kobunkan.

Fujimura, H. 1986. Urawakai Sinutapka-bito no hanashi, *Katari no naka no seikatsushi* [*Ainu Life in Oral Tradition*], 187–249. Sapporo: Ainu Mukeibunka Denshoo Hozonkai.

Hattori, S. 1999[1959]. *Nihongo no Keitoo* [*The genealogy of the Japanese language*]. Tokyo: Iwanami Shoten.

Hattori, S. (ed.) 1964. *Ainugo hoogen jiten* [*Ainu dialect dictionary*]. Tokyo: Iwanami Shoten.

Hattori, S. and M. Chiri. 1960. Ainugo shohoogen no kisogoi tookeigakuteki kenkyuu [A lexicostatistic study on the Ainu dialects]. *Minzokugaku kenkyuu* [*The Japanese journal of ethnology*], 24(4): 31–66.

Heider, E. R. 1972. Probabilities, sampling, and the ethnographic method: The case of Dani Colour Names. *Man*, 7(3): 448–466.

Heider E. R. and D. C. Oliver. 1972. The structure of the color space for naming and memory in two languages. *Cognitive psychology*, 3: 337–354.

Kay, P. 1975. Synchronic variability and diachronic change in basic color terms. *Language in society*, 4: 257-270.

Kay, K. and L. Maffi. 2013. Green and blue. *The world atlas of language structures online*. Dryer, M. S. and M. Haspelmath (eds.). Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/134, Accessed on 2016-05-04.)

Kay, P. and C.K. McDaniel 1978. The linguistic significance of the meanings of basic color terms. *Language*, 54(3): 610–646.

Kay, P and L. Maffi. 1999. Color appearance and the emergence and evolution of basic color lexicons. *American anthropologist*, 101: 743–760.

Kay, P., Berlin, B. Maffi, L., Merrifield, R. and C. Richard. 2009. *The world color survey.* Stanford: CSLI Publications.

Kindaichi, K. 1931. Ainu yuukara gohoo tekiyoo. [An outline of the grammar of Ainu epic poetry]. *Ainu jojishi: Yuukara no kenkyuu* [*Ainu epic poetry: research on yukar*], 1. Tokyo: Tooyoo Bunko.

MacLaury, R. E. 1987. Color-Category Evolution and Shuswap Yellow-with-Green. *American Anthropologist*, 89: 107–124.

MacLaury R. E. 2005. So-called Brightness in color ethnography: Potentials for LCD technology in Fieldwork and categorization research. *Cross-cultural research*, 39(2): 205–227.

Moss, A.E. 1989. Basic Colour Terms: Problems and Hypotheses. *Lingua*, 78: 313–320.

Murayama, S. 1971. *Kitachishima Ainugo* [*Northern Kuril Ainu*]. Tokyo: Yoshikawa Kobunkan.

Nakagawa, H. 1995. *Ainugo Chitose hoogen jiten* [*The Ainu-Japanese dictionary: Chitose dialect*]. Tokyo: Sofukan.

Nakagawa, H. (ed.). 2014. *Ainugo Mukawa hoogen Nihongo Ainugo jiten* [*The Japanese-Ainu dictionary: Mukawa Dialect*]. Chiba: Chiba University. (Available online at http://cas-chiba.net/Ainu-archives/)

Okuda, O. 1999. *Ainugo Shizunai hoogen bunmyaku-tsuki goishuu* (*CD-ROM tsuki*) [*Ainu Shizunai dialect lexicon in context (with CD-ROM)*]. Ebetsu: Sapporo Gakuin University.

Refsing, K. 2013. Ainu Spririts singing: The living world of Chiri Yukie's Ainu Shin'yōshū [Review]. *Journal of Japanese studies*, 39(1): 216–221.

Roberson, D. 2005. Color categories are culturally diverse in cognition as well as in language. *Cross-cultural research*, 39(1): 56–71.

Sawai, H. 2006. *Ms. Tomeno Sawai's basic vocabulary of the Tokachi dialect of Ainu* [in Japanese]. Sapporo: The Hokkaido Ainu Culture Research Center.

Sawai, H. and S. Tamura (eds.) 2005. *Ainu language: Transcribed materials of the Obihiro Dialect—Basic vocabulary of Haru Hirono recorded by Suzuko Tamura* [in Japanese]. Ebetsu: Sapporo Gakuin University.

Tamura, S. 1996. *Ainugo Saru hoogen jiten* [*The Ainu-Japanese dictionary: Saru dialect*]. Tokyo: Sofukan.

Tamura, S. 2000. *The Ainu language*. Tokyo: Sanseido.

Torii, R. 1903. *Chishima Ainu* [*Kuril Ainu*]. Tokyo: Yoshikawa Kobunkan.

Phonetic observations on Korean dialects made by Ogura Shinpei

Rei Fukui

Department of Korean Studies,
Graduate School of Humanities and Sociology,
The University of Tokyo

Abstract

This paper[1] discusses various problems found in phonetic descriptions on Korean dialects made by Ogura Shinpei. Pitch accent, vowel length, the qualities of some Chejudo vowels, and some initial consonants are discussed. As for pitch accent, two kinds of newly found manuscripts are introduced. By comparing these manuscripts and the corresponding descriptions appeared in the published version, we can understand more closely his observation of Korean pitch accent.

1 Introduction

Ogura Shinpei (1882–1944) was one of the early scholars who founded Korean linguistics in the first half of the 20th century. He conducted a series of dialect surveys on Korean dialects throughout the Korean peninsula in 1910s through 1930s, and the results had been published in numerous articles and monographs. In 1944 he tried to publish a comprehensive work on Korean dialectology by bringing together all the materials he had gathered through his entire life and major articles he had written on Korean dialectology. The result was a two volume book entitled as *Chōsengo hōgen no kenkyū* (Studies in Korean dialects, 1944). However, his health did not allow him to see its publication as he fell ill and passed away early that year so that its publication was completed by his two disciples[2].

His observation of Korean dialects is priceless in two respects: first, it encompasses all the areas in the Korean peninsula, which is now impossible because of the political separation between North and South. Secondly, his observation was the earliest among any modern systematic dialect surveys. Despite all the advances, however, it is sometimes difficult to understand what he had meant by his phonetic transcriptions. For example, we sometimes find inconsistent descriptions across different parts of the book. In some cases, his observations can simply be wrong: vowels that we normally expect long are often described as short, and in the case of a vowel found in the Chejudo dialect, all other researchers after him make a different observation. As for pitch distinctions, which are prominent in some parts of northern and southern dialects, he gives us very little information.

The purpose of this paper is to examine various problems found in his description of accent, vowel length, vowel system (especially that of the Chejudo dialect), some initial consonants (especially those beginning with /n/ and /r/ before /i/ and /j/) in order to evaluate correctly his description and make the best of his data[3]. As for the problems concerning the accent, although he made only fragmentary descriptions in published work, we would like to introduce two kinds of hitherto unknown manuscripts containing his account on Korean accent.

---

[1] This paper is a revised version of my presentation at the 3rd International Conference on Asian Geolinguistics held at the Royal University of Phnom Penh on 24 May 2016. Part of this study is supported by Grant-in-Aid for Scientific Research ((C) 15K02504) from the Ministry of Education, Science and Culture.

[2] A detailed account on how this book was published was written by Shibata Takeshi (1918–2007) in the postscript of this book.

[3] A more general acount on his dialect surveys is given in Fukui (2016). This paper discusses his phonetic descriptions in more detail.

In a sense, examining his description of Korean dialects is a kind of text criticism that is needed for historical linguistics, and we are dealing with what may be termed as the *historical geolinguistics* since a little more than 100 years have passed when he first began to investigate Korean dialects.

2 Ogura Shinpei's system of phonetic transcription

His system of phonetic transcription for describing Korean dialects differs slightly according to articles but the one used in Ogura (1944) is like shown in Table 1. There are two kinds of transcriptions: for each item, a Hangul letter is shown first, followed by a letter-to-letter transliteration for it, and then a phonetic transcription enclosed in square brackets. The former is used for citing examples from past literature written in the Hangul script, and the latter for the purpose of indicating phonetic transcription of modern dialect data.

Table 1. Phonetic transcription used in Ogura (1944).

1. ㅏ = a [a]	2. ㅑ = ia [ja]	3. ㅓ = ɔ [ɔ]	4. ㅕ = iɔ [jɔ]
5. ㅗ = o [o]	6. ㅛ = io [jo]	7. ㅜ = u [u]	8. ㅠ = iu [ju]
9. ㅡ = ụ [ụ]	10. ㅣ = i [i]	11. ㆍ = ɐ	12. ㅐ, ㆍㅣ = ai, ɐi [ɛ]
13. ㅒ = iai [jɛ]	14. ㅔ = ɔi [e]	15. ㅖ = iɔi [je]	16. ㅚ = oi [ø]
17. ㅟ = ui [wi]	18. ㅢ = ụi [ụi]	19. ㅘ = oa [wa]	20. ㅝ = uɔ [wɔ]
21. ㅙ = oai [wɛ]	22. ㅞ = uɔi [we]	23. ㄱ = k [k, g]	24. ㄴ = n [n]
25. ㄷ = t [t, d]	26. ㄹ = r [l, r]	27. ㅁ = m [m]	28. ㅂ = p [p, b]
29. ㅅ = s [s, t]	30. ㅇ = ŋ [ŋ]	31. ㅈ = č [tʃ, dʒ]	32. ㅊ = čʻ [tʃʻ]
33. ㅋ = kʻ [kʻ]	34. ㅌ = tʻ [tʻ]	35. ㅍ = pʻ [pʻ]	36. ㅎ = h [h]
37. ㅺ = sk [ʔk, ʔg]	38. ㅼ = st [ʔt, ʔd]	39. ㅽ = sp [ʔp, ʔb]	40. ㅆ = ss [ʔs]
41. ㅉ = sč [ʔtʃ, ʔdʒ]	42. ㅿ = ż		

This system is basically in accordance with the IPA but there are several differences. First, the symbol [ụ] (Overdot U) is used for the Hangul letter '으' which is now usually transcribed by [ɨ] or [ɯ], and for vowels found in the Cheju island corresponding to Middle Korean 'ㆍ' and 'ㆍㅣ', he uses [*o*] (Italic O) and [*ö*] (Umlaut Italic O) respectively, apparently being used as ad hoc symbols.

As for the symbol [ụ] (No. 9 in Table 1), he used [ɯ̇] (Overdot ɯ) in his previous works such as Ogura (1931a, 1931b). He explains reasons for changing it as follows:

> In the past I used [ɯ̇] for '으', but in order to simplify the symbol shape and to keep harmony with the second class of the vowel '어', I changed it to [ụ]. 「従来私は 으 に対しては [ɯ̇] を用ひ来ったが，字画を簡単ならしむると，어 の第二類に関係を保たしむるとの理由により，之を [ụ] と改めた」(Ogura 1934: 94, translation mine, the same hereafter).

The second reason has a relationship with vowels that correspond to the Hangul letter '어'. In some dialects including conservative Seoul, there are two kinds of pronunciation for this single letter: one is a vowel usually transcribed as [ɔ], and the other, a central (and usually long) vowel which is usually transcribed as [ə:] or [ɨ:]. He was well aware of this and had called the former the first class, and the latter the second class, transcribing by the symbols [ɔ] and [ʉ] respectively.

As for the vowels found in Chejudo dialects, we will discuss later in section 5.

3 Pitch accent

Some dialects of the Korean language are known for having a distinctive pitch accent system. The two most typical cases are found in Hamgyŏng dialects, located in the north-eastern area of the peninsula, and in Kyŏngsang dialects, located in the south-east area.

Despite this fact, Ogura Shinpei left us very few descriptions on Korean accent compared to the total amount of his works on Korean dialects[4]. There are no articles or monographs that exclusively deal with accent. But fragmentary descriptions are found sporadically in various works such as the following:

Kokugo oyobi chōsengo hatsuon gaisetsu (An introduction to the pronunciation of Japanese and Korean)『国語及朝鮮語発音概説』(1923)
Nanbu chōsen no hōgen (Southern Korean dialects)『南部朝鮮の方言』(1924)
Chōsengo hōgen no kenkyū (Studies in Korean dialects)『朝鮮語方言の研究』(1944)

Among these works the most detailed description is found in the last one, in which two kinds of descriptions on the accent system of several Kyŏngsang dialects and a tentative accentual classification of the Chŏlla dialects are given (2nd vol., 431–439).

Recently I have found two kinds of manuscripts of his that correspond to the two kinds of Kyŏngsang accent description just mentioned. In the following, I will first cite all the passages in which he touched upon Korean accent as exhaustively as possible, and then try to compare the most detailed part of his published description and corresponding newly found manuscripts which contain a more interesting account on the Kyŏngsang accent.

3.1 *An introduction to the pronunciation of Japanese and Korean* (1923: 119)

In this book he briefly mentions about the accent in Korean. He distinguishes 'word accent' and 'sentence accent', and states that both Kyŏngsang and Chŏlla dialects have a kind of prominent accent. As for the phonetic realization of the accent, both pitch and intensity seem relevant in this description.

(1) In Korean, word accent and sentence accent differ according to regions. In my own investigation, people in south and north Kyŏngsang and Chŏlla dialects have in general a prominently lowering accent. For example, ... (example sentences omitted) … putting accents in this way, they constantly go on speaking with an alternating wave of intensity. …「朝鮮語においては語アクセント及び文アクセントは地方によって夫々相違がある。余の実査した所では慶尚南北道・全羅南道の人々のアクセント，殊に文アクセントは，一般に著しく降下的即ち頭あがりである。例へば此等の地方の人々の談話は「사람이만타」・「山이높다」の사・만・山・높にアクセントを置き，絶えず強弱の波動を繰返して進んで行くのである。…」

3.2 *Southern Korean dialects* (1924: 47-48)

In this monograph the following description on 'sentence accent' is given for southern Korean dialects. Also, a map indicating the area having a distinctive accent system is attached in the appendix (See Figure 1). In this description, a more detailed account on the geographical distribution of the dialects having this kind of accent is given.

(2) It is easy to observe that South and North Kyŏngsang dialects have sentence accent, i.e., prominent highs and lows of pitch. The investigation of this system of accent is the most difficult part in dialect surveys and I myself have been unable to fully attain the goal, but to sum up, it places prominence on the first syllable and the discourse goes on with a head-rising tone in these areas.

[4] The author (and perhaps many others ?) has been half unconsciously thinking that this has something to do with his linguistic background. He was born and raised in Sendai, a city known for the lack of distinctive pitch like Tokyo, Kyoto and many other dialects.

And the geographical distribution is as follows: most prominent areas are all parts of South and North Kyŏngsang, (coastal) Kangwŏn area south of Chumunjin. Its influence goes on to all South Chŏlla dialects but it is not so prominent as in the Kyŏngsang areas, except for southern coastal places like Kohŭng, Pŏlgyo, Posŏng and Changsŏng, in which it is as prominent as in Kyŏngsang areas. It is noteworthy that it is not found in South and North Chungch'ŏng, North Chŏlla, and Kangwŏn area north of Yangyang, less prominent as in the case of Seoul areas.「慶尚南北道地方の朝鮮語に文アクセント (Sentence-accent) 即ち語調の高低の顕著なることは何人も容易に認め得る所である。此のアクセントの研究は方言調査上最も困難を感ずる所で，余も未だ十分の目的を達することを得ぬのであるが，要するに同地方に於ける此の種のアクセントは毎音節の初の部分に力を置き，即ち頭上がりの調子を以て談話を継続して行くのである。而して其の分布の状態を見るに，慶尚南北両道の全部及び江原道注文津以南を以て最も顕著なる地方とする。而して其の影響の及ぶところは全羅南道一円であるが，該地方に於けるアクセントは慶尚南北道に於けるもの程顕著ならず，唯南方海岸に面し慶南地方と比較的交通の頻繁なる高興・筏橋・宝城・長城地方が慶尚道と同程度に於て著しきを発見するのである。忠清南北道・全羅北道の全部，江原道襄陽以北にありては此のアクセントが認められず，京城地方と大差なきは注意するに足る現象である」

3.3 *Studies in Korean dialects* (1944)

Lastly, we would like to examine descriptions found in this work. First, all passages containing brief descriptions on accent of any kind are listed, and then the above mentioned detailed description found on pp. 431–439 will be discussed.

(3) … a kind of orange is also called [kam-dʒa], but in the case of potato the accent is placed on the second syllable whereas in the case of orange it is on the first syllable.「蜜柑も [kam-dʒa] であるが，馬鈴薯はアクセントが第二音節にあり，蜜柑は第一音節にある。」(1st vol., p. 196. A note for the item 'potato' on Kyŏngsangdo Hapch'eong (陜川))

The word [kam-dʒa] denoting a kind of orange (柑子) is no longer used so that it seems difficult to verify its accent. In the case of the word for 'potato', the accent is placed on the second syllable in many Kyŏngsang dialects and this is in accordance with his observation.

(4) (explaining accusative markers) … In south Hamgyŏngdo north of Munch'ŏn [kɛ-rù ʔtɛ-rin-da] (hit a dog), [sul-ù mɔŋnùn-da] (drink alcoholic beverages). (accent is placed on [rù] and [ù])「咸鏡南道にありては，文川以北では，[kɛ-rù ʔtɛ-rin-da], [sul-ù mɔŋnùn-da]（[rù] 及び [ù] にアクセントを置く）」(2nd vol., p. 327)

Although this is for the purpose of explaining accusative markers, he made a comment in passing, noting that the accent is placed on the accusative markers, implying that in other cases accent is placed on the noun. This seems to be his sole mention on the Hamgyŏngdo accent.

Next, the following passages all concern about dialects in the vicinity of Kyŏngsangdo, either with or without an accent system. Some Chŏlla dialects are reported to have an accent system similar to that of Kyŏngsang dialect ((5) and (6)), whereas no such accent is heard in North Ch'ungch' ŏng dialects (7).

(5)「南原郡 雲峰面の東隣なる東面及び山内面の一部には慶尚南道方言の影響あり，殊に慶尚道方言に存する独特のアクセントが聴取される。雲峰面自身の方言は東隣の東面及び西隣の二白面のそれとも趣を異にし，恰も両者の中間にある如き観を呈して居る。」(2nd vol., p. 525)

(6)「全羅南北道順天・求礼・南原・長水・茂朱諸郡内の慶尚南北道に接する地方には，一般に慶尚道特有のアクセントが行はれて居る。」(2nd vol., p. 608)

(7) 「慶尚北道各地に同道特有の文アクセントが顕著であるが，忠清北道方言ではそれが聞かれない。」(2nd vol., p. 616)

Next, the following passage is the most detailed description he had made on the accent system of Kyŏngsang dialects (2nd vol., pp. 431–439). As I said above, this passage consists of two kinds of descriptions. First, a comparison of 'word accent' for some eight places in Kyŏngsang dialects are given. The number of words investigated is about eighty. The places are mostly located in North Kyŏngsang area but a few South Kyŏngsang places are included as well.

(8-1) Everyone can notice that a kind of special accent exists in the Yŏngnam (=Kyŏngsang) dialects. I have been unable to fully investigate it until now, but based on a few materials at hand I will try to deal with one aspect of this topic, although insufficient. First, as for word accent, I will cite a few words from the dialects such as Taegu (大邱), Kimch'ŏn (金泉), Ŭisŏng (義城), Yŏngch'ŏn (永川), Andong (安東), Yech'ŏn (醴泉), Yŏngdŏk (盈德), Yangsan (梁山), and compare among them. 「嶺南地方の方言に一種特別のアクセントの存することは，何人にも直ちに気がつくことである。私は今日までそれに関し十分な調査を遂げたことが無いが，手許にある資料を基礎として，不完全ながら少しくこの問題の一端に触れて見よう。先づ，単語のアクセントであるが，それについては，大邱・金泉・義城・永川・安東・醴泉・盈德・梁山諸地点の方言中より若干語を抽出し，比較を試みることとする。」

In this explanation he only mentions 8 places but in fact he sometimes adds data from Kimhae (金海) , Ulsan (蔚山) and some other places.

He then presents accent data for disyllabic and trisyllabic words according to the place of accent as are shown in Table 2. The labels such as 2-1, 2-2, 3-1 etc. are added by the author to indicate the number of syllable and the location of accent. (In this table only one example word for each accent group is shown but a total of 80 words are presented in his description. Also, Square brackets and hyphens indicating syllable boundary are omitted due to limited space in the table.)

Table 2. Accent group and example words.

	gloss	Taegu	Kimch'ŏn	Ŭisŏng	Yŏngch'ŏn	Andong	Yech'ŏn	Yŏngdŏk	Yangsan	Kimhae
2-1	woman	tʃi´dʒip		tʃi´dʒip	tʃi´dʒip	tʃi´dʒip	tʃi´dʒip		tʃidʒip	tʃidʒip´
2-2	potato	kamdʒa´	kamdʒa´	kamdʒa´	kamdʒa´	kamdʒa´	kamdʒa´	kamdʒa´	kamdʒa´	
3-1	lizard	to´mɛpɛm	to´mɛpɛm	to´mɛpɛm		to´mɛpɛm	to´mɛpɛm		to´mɛpɛm	
3-2	wild goose	kirɔ´gi		kirɔ´gi	kirɔ´gi	kirɔ´gi	kire´gi	kirɔ´gi	kiri´gi	kiri´gi
3-3	boy	mɔsùma´	mɔsùma´	mɔsùma´		mɔsùma´	mɔsùma´			mɔsù´ma

(´ : accent, words without an accent mark: level tone (平板), shaded cell: exceptional accent.)

Examples of words for each accent group are summarized in Table 3.

Table 3. Words belonging to each accent group.

2-1	[tʃi´-dʒip] (woman), [tʃim´-tʃi] (kimchi), [ka´-mɛ] (iron pot), [ku´-si] (manger), [kɔ´-buk] (turtle), [kɔ´-pʻum] (bubble), [nùi´-bi] (silkworm), [na´-bu] (butterfly), [tal´-bi] (hairpiece), [to´-mɛ] (chopping board), [to´-mi] (sea bream), [mol´-gɛ] (sand), [mal´-bam] (water chestnut), etc.
2-2	[kam-dʒa´] (potato), [kun-dùi´] (swing), [ku-mu´] (hole), [ka-ri´] (powder), [ku-du´] (shoes), [no-ri´] (deer), [tu-bu´] (bean curd), [to-tʃʻi´] (axe), [tɔm-bul´] (bush), [tɛ-tʃʻu´] (jujube), [toŋ-tʻɛ´] (ring), [tol-gɛ´] (broad bellflower), [mun-dʒi´] (dust), etc.
3-1	[to´-mɛ-pɛm] (lizard), [mi´-nu-ri] (wife), [ʔpɛ´-da-dʒi] (drawer), [sa´-ma-gu] (blotch), [ɔ´-gùm-ni] (molar),

	[tʃɔn´-d ùn-da] (endure)
3-2	[ki-rɔ´-gi] (wild goose), [ka-si´-gɛ] (scissors), [ʔkɛ-go´-ri] (frog), [na-sɛŋ´-i] (shepherd's-purse), [na-mak´-sin] (wooden shoes), [mu-dʒi´-gɛ] (rainbow), [pɔ-bɔ´-ri] (dumb)
3-3	[mɔ-sù-ma´] (boy)

He then concludes this section in the following way:

(8-2) As can be seen from these accent groups, Taegu and neighbouring places have much in common but places such as Yangsan (梁山), Ulsan (蔚山) and Kimhae (金海) show slightly different tendencies.「以上（イ）より（ホ）に至る現象を通覧するに，大邱付近の語アクセントには自ら共通点を発見することができるが，梁山・蔚山・金海地方になると，少しくその趣が異なって来ることが看取される。」

The above is his description on 'word accent' of Kyŏngsang dialects. In this description, he made only two kinds of accentual distinction for disyllabic words and three for trisyllabic words. However, we now know that many of these dialects can distinguish four kinds of pitch patterns for dissyllabic words, and four or five kinds for trisyllabic words as can be seen in the following table.

Table 4. Typical Kyŏngsang accent systems.

A typical south Kyŏngsang system (Pusan)			A typical north Kyŏngsang system (Taegu)		
H(H)	HH	HHL	H(H)	HH	HHL
H(L)	HL	HLL	H(L)	HL	HLL
L(H)	LH(L)	LHL	H:(H)	LH(L)	LHL
	LH(H)	LHH		H:H	LHH(L)
					H:HL

(H = high, L = low, ':' = long. The tones of case markers are shown in parentheses.)

In view of the accentual system made by recent researchers, Ogura Shinpei did not seem to be able to observe patterns like HH, HHL, and the ones having a long vowel in the initial syllable, and hence the distinction between HHL and LHL, etc. However, he sometime mentions a pitch pattern like 'level (or flat)' (平板) for a dialect like Yangsan seen above so that he seemed to have been aware of the existence of pitch patterns that cannot be described simply by the location of accent.

Recently, two kinds of handwritten manuscripts are found in one of his documents recently donated to the Research institute for Oriental cultures, Gakushuin University (Tsuji et al. (eds.) (2016)). In the following, these two manuscripts will be referred to as M1 and M2. M1 is entitled as 'Akusento shirabe (Accent survey)', consisting of five sheets of manuscript paper (see Figure 2, at the end of this paper). By comparing M1 with the description we have just seen above, it is easy to see that he had written the published version based on the data included in this manuscript. However, there are several differences: in the manuscript, three more places (Yŏngju 栄州, P'ohang 浦項 and Hamyang 咸陽) are added, and it has a more complicate system of symbols indicating accents than the published version. Basically a small circle (○) seems to indicate the accent but two other symbols, namely —○ and ◎ are also used. Meanings of these two symbols are not obvious. The symbol —○ is sometimes placed between two syllables, possibly meaning level pitch. The symbol ◎ is used, for example, for the word [tʃidʒip´] (woman) for the Kimhae dialect. It may be the case that this symbol was used to emphasize that this place had a different accent than all other places.

Next, we would like to turn to the second part of the descriptions on the accent of Kyŏngsang dialects. It deals with the 'sentence accent', making contrast with the previous section that deals with the 'word accent'. However, the dialects discussed here is not the same as the previous section in that it only deals with south Kyŏngsang dialects such as Tongnae, Pusan and Kimhae.

(9-1) The above is a description of phenomena based on independent word accent. I did not have yet chance to investigate accentual alternation when a case marker or particle is attached to a word, making a phrase or a sentence. However, I would like to present some results obtained from my investigation conducted in Tongnae, Pusan and Kimhae (results of the three places coincided) in order to be served for reference.「以上は独立せる単語のアクセント現象である。これに助詞が附き句をなし，文をなす場合に起るアクセント変化現象などは，未だこれを調査する機会を得なかったが，次に嘗て私が東萊・釜山・金海において実査した結果（三箇所調査の結果が完全に一致した）を掲げ，参考に供する（上部の語は単独に発音される場合のアクセント，下部は文中において発音される場合のアクセントである)。」

He then presents the following data. The leftmost word in square brackets shows the word accent, and the following sentence the 'sentence accent'.

(9-2)
[ka-si´-gɛ]（scissors） tʃɔ´-sa-ram- ùi ka-si´-gɛ-rùl ka´-dʒɔ-o-nù-ra (Bring his scissors.)
[ka-ri´]（powder） ko-tʃʻi´-ka-ri-nùn mɛ-u´ mɛp-ta (Red pepper powder is very hot.)
[jɔ´-rum]（summer） jɔ´-rum-i[5] tʃi-na´-go ka-ù´-ri tø´-jɔt-ta (Summer is past and autumn has come.)
[ha´-nal]（sky, heaven） pjɔ-ri´ ha´-na-re ʔpɔn-tʃɔk´ ʔpɔn-tʃɔk´-han-da (Stars twinkle in the sky.)
[sa-ram´]（person） tʃo-hùn sa-ram´-ùl sa-raŋ´-han-da (To love a good person.)
[ka´-ma-gi]（crow） ka´-ma-gi han-ma´-ri-ga na-mu´ u´-e an´-dʒɔ-it-ta (A crow is sitting on the branch.)

In this description, only the location of accent is indicated by the symbol ´ as before. However, we have another manuscript (M2) for this part of description which is much more interesting than M1. M2 is written on a sheet of paper and entitled as the 'Example of Tongnae dialect accent (東萊方言Accent例)'. The date and place are also recorded: 'Recorded on Nov. 14, 1932, at the Tongnae elementary school (「昭和七年十一月十四日調　東萊普通学校にて」)'.

It is easy to see that this manuscript corresponds to the published version of the passage we have just seen. However, the method of description is quite different (see Figure 3 attached at the end of this paper). Three kinds of symbols are used: acute accent (´), grave accent (`) and bar (－). Sometimes no symbol is written on a syllable especially near the end of a sentence. If we count this case as a kind of zero notation, then we have four kinds of symbols. The meanings of these symbols are not explained in anywhere so that we only have to guess. The following is a tentative reading by the author based on the actual south Kyŏngsang pitch patterns and the way these symbols are combined.

´	High (accented)	H
`	Low	L
－	Mid	M
No symbol	Extra low	(unmarked)

In this manuscript, words and sentences appeared in (9-2) are all included, with the addition of some monosyllabic words like 'star', 'dog', 'cow', 'horse' and 'water'. Also, in the published version, transcription of words and sentences is made in phonetic symbols whereas it is recorded in the Hangul script in M2.

[5] [jɔ´-rum] must be a mistake for [jɔ´-rùm].

	MML	MMML HL MM
雁	기리기	기리기가 날아 간다 (wild goose—A wild goose flies.)
	MHM	HMML MHL HM
鋏	가시개	져사람의 가시개를 가져오느라 (scissors—Bring his scissors.)
	MH	MHMM MH MM
粉	가리	고치가리는 심히/매우 밉다 (powder—Red pepper powder is very hot.)
	HM	HLM MH MHL HM
夏	여름	여름이 지나고 가을이 되엿다 (summer—Summer is past and autumn has come.)
	HL M	MH HLM MH MH H
天	하늘 星 별	별이 하날에 뻔썩뻔썩 한다 (stars—Stars twinkle in the sky.)
	M M M	MH HL HL
犬	개 牛 소 馬 말	개와 소와 말이 온다 (dog, cow, horse—A dog, a cow and a horse come.)
	MH	MH MH MH MH
人	사람	조흔 사람을 조와(사랑)한다 (person—To love a good person.)
		HM HM HL HL
水	물	맑은 물이 흘너 간다 (water—Clean water flows.)
	MML	MML MHLM MH HL HL MM
烏	가마기 (ka-ma:-gi)	~~한마리~~ 가마기 한마리가 나무 우에 안저 잇다 (crow—A crow is sitting on the branch.)

If we compare the description of accent in M2 with that of the published version, the high pitch (H) found in M2 generally coincide with the location of accent in the published version. All other pitch variations like mid, low, extra-low are omitted in it. Also, pitch patterns such as HH, HHL are not recorded as such and they tend to be transcribed as HL, HM, MML. Such a tendency is the same as in the case of M1. It is also interesting to note that pitch patterns like LH, and HL tend to be recorded as MH and HM and so on, which reminds us of the way tones are recorded in the works by Hŏ Ung (1963) and Kim Ch'agyun (1980).

The pitch patterns for the word 'wild goose' and 'crow' in M2 is almost the same (MML) but the former is deleted in the published version and the latter does not coincide with the data found in the published version. Let us compare these data with those of some modern Kyŏngsang dialects, based on the data found in Kim Ch'agyun (1980) and Son Chaehyŏn (2007).[6]

	Ch'ang'wŏn (昌原)	Chinju (晋州)	Kyŏngju (慶州)	Mup'ung (茂豊)	Taegu (大邱)
Wild goose	M3	M3	M3	M3	3D (=H2M)
crow	H2M	H2M	H2M	H2M	3D (=H2M)

If Ogura Shinpei's data for 'wild goose' (MML) is correct, it coincides with all the data shown in Kim (1980). But the data for 'crow' does not coincide with any of these data.

3.4 Accentual Classification of Chŏlla dialects

After the detailed discussion on Kyŏngsang dialects, he briefly mentions about Chŏllado dialects, thereby classifying the types of accents found in these areas.

(10-1) The accent system of Chŏlla dialects is different from that of Kyŏngsang dialects in many respects. I did not have had a chance to closely examine the difference until now, but it seems like there are three types of accent systems in Chŏllado. 「全羅道方言のアクセントは慶尚方言のそれ

[6] Data for the Taegu dialect are taken from Son (2007) and all others are from Kim (1980). In the notation by Kim (1980), M3 is phonetically eqauvalent to MML, and H2M is equivalent to MHM. The notaion 3D in Son (2007) is equivalent to Kim's H2M.

と趣を異にせる点が多い。私は今日までその異同を精査する機会を得ないが，全羅道方言には大体次の如き三種の型が存するやうに思はれる。」

And then he describes the three accent types in the following way:

(10-2)
A. Kwangju type: the accent is always located on the second syllable for disyllabic or longer words. 「光州型　二音節また三音節以上の語において，アクセントが常に第二音節に来る。」
B. Yŏsu type: the accent is located on the first syllable in the case of disyllabic words, and for words longer than that it is located on the second syllable. 「麗水型　二音節の語にありては第一音節に，三音節以上の語においては第二音節にアクセントが来る。」
C. Mokp'o type: the accent is always located on the first syllable for disyllabic or longer words. 「木浦型　二音節また三音節以上の語において，アクセントが常に第一音節に来る。」

4 Long vowels

In this section we would like to examine his description of long vowels. In order to do so it is convenient to classify long vowels according to their sources:

(1) those that correspond to Middle Korean rising tone
(2) those that emerged as the result of the contraction of adjacent vowels
(3) those with unknown origin, especially in the case of borrowings

The following words, taken from the first volume of Ogura (1944), show examples of the above three types of long vowels.

(1) [pɛ:l], [pi:l] (star), [sɛ:m][7] (fountain, spring), [sa:-ram] (person), [to:-kʻi], [to:-tʃi] (axe), [pu:-ri] (beak), [ko:m] (bear), [pɔ:l], [pɔ:-ri] (bee), ……
(2) [no:l], [no:-ri] (sunset glow), [ka:l] (autumn), [kjɔ:l], [tʃʉ:k], [tʃù:l], [tʃu:l-gi] (winter), [ma:l], [mɔ:l], [mo:l] (village), [jɔ:l] (shallows), [kɛ:m], [ʔkɛ:m], [kɛ:mi] (hazel nut), [mɛ:m], [mɛ:-mi] (cicada), ……
(3) [kol-lo:-si], [kol-lo:-sin] (rubber shoes (Hg.[8]) < Rus. галоши), [ka-rùm-da:-si] (pencil (Hg.) < Rus. карандаш), [hɛ:m] (a side dish (Hg.)), [mu-du:-kɛ] (beer bottle (Hg.) < Rus. бутылка), ……

Among these three types, the first group of words with a long vowel corresponding to the MK rising tone pose a question. That is, there are plenty of words that had originally a rising tone and pronounced with a long vowel in many modern dialects and yet are transcribed without the vowel length in Ogura (1944). For example, words such as [øn][9] (left), [pjɔŋ] (illness), [kjɛ-dʒip], [kɛ-dʒip], [ke-dʒip] (woman), [hjo-dʒa] (dutiful son), [mu-daŋ] (shaman), [sa-ma-gui] (blotch), [tol] (stone), [pɔm] (tiger), etc. are always written without vowel length for any dialects in his data.

Compared to this, the second type of long vowels seems to be transcribed fairly correctly.

As for the third type of long vowels, it is usually difficult to know the reason why they have a long vowel. However, it is interesting to note that in the cases of borrowing from Russian the location of long vowel seems to coincide with that of the accent in Russian.

[7] For this item, it is interesting to note that there are dissyllabic varieties such as [sɛ-am] and [si-am] besides the ones with a long vowel.

[8] Hg. is the abreviation for the Hamgyŏng dialect.

[9] It seems like the vowel [ø] always appears without vowel length.

To say a few more words on the problem of type (1), he was well aware that MK rising tone (indicated by two dots) corresponded to the long vowel as can be seen from the following explanation.

> [pu:] in the word [pu:-ri] (beak) is pronounced long. In *Hunmongjahoe*[10], this word was recorded as pu-ri with two dots at its left side, in order to indicate that this syllable should be pronounced long. 「[pu:-ri] (嘴) の [pu:] にある母音は長音に発音される。「訓蒙字会」に「嘴」をpu-riとし、puの左側に二点を附してあるのはpuを長母音に発音すべきことを示したものである。」(Ogura 1944, 2nd vol., p. 87)

5 On Chejudo vowel system

In this section we would like to discuss problems found in his descriptions on the Chejudo vowel system. This dialect is famous for preserving the phonological contrast between the vowel corresponding to the Middle Korean vowel ᄋᆞ /ʌ/ and all other vowels. In all other dialects this phonological distinction is lost and in many cases it merged with the vowel /a/ in the first syllable.

Ogura (1944) transcribes this vowel by the symbol [*o*] (Italic O), but this is obviously not a good selection because it will be easily confused with a normal [o], and having a phonetic distinction between italic and upright shapes is not a typographically good solution. The symbol that comes most closely to this Chejudo vowel is [ɔ] in the current IPA framework. Unfortunately, however, this symbol was used to represent the vowel corresponding to the Hangul letter 어, in spite of the fact that this vowel is phonetically a central vowel [ə] in this dialect, so that he seemed to have no other choices.

Next, we would like to discuss a yet unresolved problem concerning the vowel system of this dialect, namely the phonetic quality of the vowel corresponding a Middle Korean diphthong ᄋᆡ. Ogura (1944) represents this vowel by the symbol [*ö*] (Italic Umlaut O). Although he does not give any explanation at the introductory part of this book, a detailed description is given in the chapter describing various characteristics of the Cheju dialect (2nd vol., pp. 459–461).

> As I told above, there is a clear distinction between [*o*] (ᄋᆞ) and [a] (아). However, this kind of feature manifests itself not only in the cases of single vowels but in the cases of vowels combined with other vowels. Therefore, there is a clear distinction between the vowels 애 (represented by [ɛ]) and ᄋᆞ이 (represented by [*ö*]). For example:
>
> | 애 [ɛ] | ᄋᆞ이 [*ö*] |
> |---|---|
> | [kɛ-mat] (bay)・[kɛ] (dog)・ | [pɔn-g*ö*] (thunder)・[ʔsil-g*ö*] (liver)・ |
> | [pe-gɛ] (pillow)・[tʃo-gɛŋ-i] (clam). | [k*ö*-u-ri] (earthworm)・[k*ö*k] (guest). |
> | …… | |
> | [hɛ] 해 (harm). | [h*ö*] ᄒᆞ이 (sun, year). |
>
> These examples clearly preserve the distinction in pronunciation in every places (in Cheju).

> 「[*o*] (ᄋᆞ)と [a] (아)との間に明かに発音上の区別の存することは，上述の通りであるが，その特質は独り単母音の場合に於てのみならず，これが他の母音と結合した場合に於ても，発音上に区別が現はれて来る。即ち애（[ɛ]を以てあらはす）とᄋᆞ이（[*ö*]を以てあらはす）との間にも明白な区別が行はれるのである。例えば，
>
> | 애 [ɛ] | ᄋᆞ이 [*ö*] |
> |---|---|
> | [kɛ-mat] (浦)・[kɛ] (犬)・ | [pɔn-g*ö*] (電)・[ʔsil-g*ö*] (肝)・ |
> | [pe-gɛ] (枕)・[tʃo-gɛŋ-i] (蛤)。 | [k*ö*-u-ri] (蚯蚓)・[k*ö*k] (客)。 |
> | ……（中略）…… | |
> | [hɛ] 해 (害)。 | [h*ö*] ᄒᆞ이 (日)。 |
>
> の如きは各地とも最も明らかに発音上の区別を保存する。」

[10] A Sino-korean dictionary compiled in 1527.

Similar explanation is also found in Ogura (1931a: 32–33, also 1944: 2nd vol.).

> 「의　一般には 애 と同音である故に，ɛ を以って表はす。但し，済州島方言にありては，오 は両唇の円みを帯びた音である故に，애 の如く全然開口音とはならず，외（其の条参照）種の音となり，唯 외 よりも稍々開口の形を取る。私は之に対してöの記号を与える。済州島人が 애 と 의 とを明らかに区別して発音することは，陸地人の奇異に感ずる程である。例えば陸地方面では 대（竹）・듸（対）とも何れも tɛ: と発音するが，済州島人は前者を tɛ:，後者を t*ö* と発音し，又陸地方面では 해（害）・희（日）とも何れも hɛ: と発音するが，島人は前者を hɛ:，後者を h*ö* と発音して区別する。」

In short, the vowel represented by the symbol [*ö*] by him is a front rounded vowel that is a little more open than the vowel corresponding to the Hangul letter 외 ([ø]). In the framework of the current IPA system the vowel symbol that comes most closely to such a vowel would be [œ], front, half-open, rounded vowel. (See Figure 4, attached at the end of this paper, for a vowel diagram shown in Ogura (1931a))

By the way, this observation by Ogura Shinpei has not been followed by other researchers. Umeda Hiroyuki (1960: 34), Lee Kimun (1977), Hyŏn P'yŏnghyo (1985: 372–375, 461), Chŏng Sŭngch'ŏl (1995) almost unanimously state that this vowel is pronounced as [e], so that his observation may be simply wrong. However, it may be equally possible that this vowel had lost lip rounding and changed ([ö] > [e]) sometime after his observation, because he states the quality of this vowel so clearly in the above mentioned two works.

6 Problems in initial consonants /n, r/ before /i, j/

The initial consonants /n/ and /r/ are not generally pronounced when followed the vowel /i/ or the glide /j/ in almost all central and southern dialects so that, for example, the surname Lee (李) is pronounced as [i:] in these areas. However, in northern dialects such as P'yŏng'an and Hamgyŏng, these combinations can be pronounced. It is also noteworthy that the combination /ni/ can also appear in the Cheju dialect. Ogura (1944) gives minute descriptions for these cases, as shown in the following examples (all taken from the 2nd volume of Ogura (1944)):

> P'yŏng'an (平安道)「ni は語頭にある場合には京城地方では [i] に発音されるが，本道【＝平安道】にありては明かに原音を存する。例へば ni（歯）・nip-sa-kui（葉）・ni-čɔ-pɔ-rita（忘れる）・nik-ta（熟す）・ni-rɔ-na-ta（起つ）は京城の発音では … (中略) … であるが，本道では [ni], [nip-sa-gui], [ni-dʒɔ-pe-ri-da] , [nik-ta], [ni-rɔ-na-da] と発音される。黄州邑内では [ni] 或は [i] の何れにも発音されるが，郡内天柱面にありては，平安南道の如く [ni] に発音されるというふことである。」(p. 258)

> P'yŏng'an (平安道)「ri は語頭にある場合には，京城地方では [i] となるが，本道では多くの場合原音のままに発音される。例へば「李書房」・「臨津江」・「隣家」・「里洞」…（中略）…但しこの場合の発音を精細に観察すると [ni] の如くにも [ri] の如くにも [di] の如くにも聞かれる。」(p. 260)

> Hamgyŏng (咸鏡道)「ni は平安南北道にありては，語頭にありても明かに [ni]（或は [di] に近く）に発音される特徴を有するが（例，「歯」[ni],「葉」[nip-sa-gui] など），咸鏡南道にありては，豊山・長津郡内に僅かに「葉」を [nip-sa-gui] といふこと以外，すべて京城地方と同じく ni は [i] となり，黄海道にありては，黄州に於て著しく平安道の影響を受けて「歯」[ni]・「葉」[nip-sa-gui]・「熟す」[nik-ta] など語頭に [n] を存するが，南下するに従って漸くその影を潜め載寧・瑞興に於て僅かに「歯」を [i] 及び [ni] と発音する外，他の地方にありては，すべて京城地方と同じく [i] の音に変化する。」(p. 305)

Hamgyŏng (咸鏡道)「ri は語頭にありては，咸鏡北道の国境地方及び平安南北道のほとんど全部に亘って原音を存し，[ri] に発音されるが，咸鏡南道及び黄海道に於てこの現象の存するのは豊山地方のみであり，……」(p. 308)

Cheju (済州島)「n を頭音とする語のうち，… i と結合する ni（虱）・ni（歯）の如きは，次の如き形に於て n の原音を存し，しばしば屢々 [di] 音の如くにする聞かれる。」(p. 464)

7 Concluding remarks

Ogura Shinpei was born and raised in Sendai. One of the most famous features of the Sendai dialect is the lack of distinctive pitch. Therefore, it may be possible to speculate that this linguistic background had something to do with the fact that he left us so little information on the accent system of Korean dialects. However, as we have seen above, he was in fact very conscious of the accent system of Kyŏngsang dialects. Newly found manuscripts on word and sentence accent of Kyŏngsang dialects also reveals his concern on this topic. He was at least able to identify the location of accent but he seemed to have difficulty in treating words with a level pitch pattern, or patterns that cannot be stated simply in terms of the location of accent. On the other hand, the reason why the accent system of Hamgyŏng dialects are ignored, except for a brief mention of it, is unclear.

His descriptions of long vowels are not satisfactory, especially in the cases of the ones that originate from Middle Korean rising tone. It may also be possible that it had something to do with his linguistic background, because Sendai is located at the south border of the area so-called ‘syllabeme’ dialect where the distinction of syllable length is not so punctual or at least not the same as in the case of the so-called moraic dialects including Tokyo and many other dialects.

His description on the Chejudo vowel [*ö*] has not been verified by any other researches so that this still remains as an open question.

Acknowledgements

I would like to thank Professor Mitsuaki Endo for inviting me to this conference and to all the participants who gave me useful comments.

References

Chŏng, Sŭngch’ŏl (1995) *Cheju pang’ŏn ŭi t’ongsa ŭmunron* (Diachronic phonology of the Cheju dialect). Seoul: T’aehaksa.

Choi, Hakkŭn (1978) *Han’guk pangŏn sajŏn*. (A Dictionary of Korean dialects) Seoul: Hyŏnmunsa.

Fukui, Rei (2016) Dialect Surveys Conducted by Ogura Shinpei: In Order to Make Use of the Data Contained in Chōsengo Hōgen no Kenkyū. *Tokyo University Linguistic Papers* 37. Departmet of Linguistics, the University of Tokyo.

Han’guk Chŏngsin Munhwa Yŏn’guwŏn (ed.) (1987–1995) *Han’guk Pangŏn Charyochip* (Materials in Korean dialects). 9 vols. Han’guk Chŏngsin Munhwa Yŏn’guwŏn.

Hŏ Ung (1963) *Chungse kugŏ Yŏngu* (A study of Middle Korean). Seoul: Chŏng’ŭmsa.

Hyŏn, P’yŏnghyo (1985) *Chejudo pang’ŏn yŏngu* (A Study of the Cheju dialect). Seoul: Yiyu Ch’ulp’ansa.

Kim Ch’agyun (1980) *Kyŏngsando Pang’ŏn ŭi sŏngjo Yŏngu* (A study of Korean tones). Seoul: Kwahaksa.

Lee, Iksop et al. (2008) *Han’guk pangŏn sajŏn* (Linguistic atlas of Korea). Seoul: T’aehaksa.

Lee, Kimun (1977) Chejudo pang'ŏn ŭi 'ʌ'e kwallyŏn toen myŏt munje (Some problems concerning the vowel 'ʌ' of the Cheju dialect). *Kuge kungmunhak nonch'ong* (Festschrift for the 70th birthday of Professr Lee Sung'nyŏn). Seoul: Tower Press.

Ogura, Shinpei (1923) *Kokugo oyobi chōsengo hatsuon gaisetsu* (An introduction to the pronunciation of Japanese and Korean). Keijō (Seoul): Chikazawa insatsujo shuppanbu.

Ogura, Shinpei (1924) *Nanbu chōsen no hōgen* (Southern Korean dialects). Keijō (Seoul): Chōsenshi gakkai.

Ogura, Shinpei (1931a) Saishūtō hogen. (The Chejudo dialect) *Seikyū gakusō* (青丘学叢) 5: 26–70. Keijō (Seoul): Seikyū gakkai. (reprinted with modification in Ogura (1944, 2nd vol.: 453–500)

Ogura, Shinpei (1931b) On the phonetic notation of Korean vowel. *Study of Sounds*. 4: 139–148. Phonetic society of Japan.

Ogura, Shinpei (1934) Genbun no rōmaji hyōkihō (Method of transcription of the Hangul script). *Oda sensei shōju kinen chōsen ronsyū*. Keijō (Seoul): Ōsakayagō shoten. 諺文のローマ字表記法.『小田先生頌壽記念朝鮮論集』京城：大阪屋號書店

Ogura, Shinpei (1944) *Chōsengo hogen no kenkyū* (Studies in the Korean dialects). 2 vols. Tokyo: Iwanami Shoten.

Son, Chaehyŏn (2007) *Kankokugo shohōgen no kijututeki kenkyū* (Accent in Korean dialects). Ph D dissertation. University of Tokyo.

Tsuji, Yamato et al. (eds.) (2016) Ogura Shinpei kankei bunsho mokuroku: Gakushūin Daigaku Tōyōbunka kenkyuujo shozō (A catalog of documents related to Ogura Shinpei: Research institute for Oriental cultures, Gakushuin University). *Ocasional Papers* 60.

Umeda, Hiroyuki (1960) On the phonemes of Cheju dialect of Korean. *Nagoya Daigaku Bungakubu Kenkyuu Ronsyuu* 22, Bungaku 8: 17–46. Nagoya University.

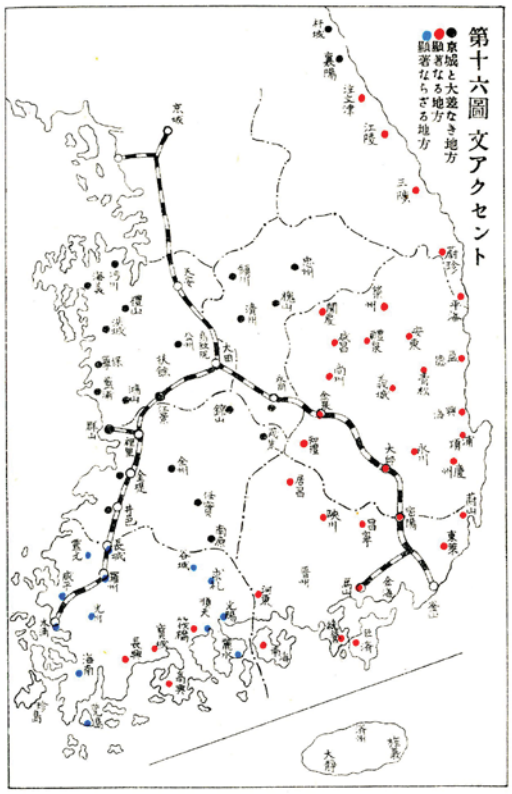


Figure 1. Geographical distribution of southern dialects with 'sentence accent' (Ogura 1924, appendix)

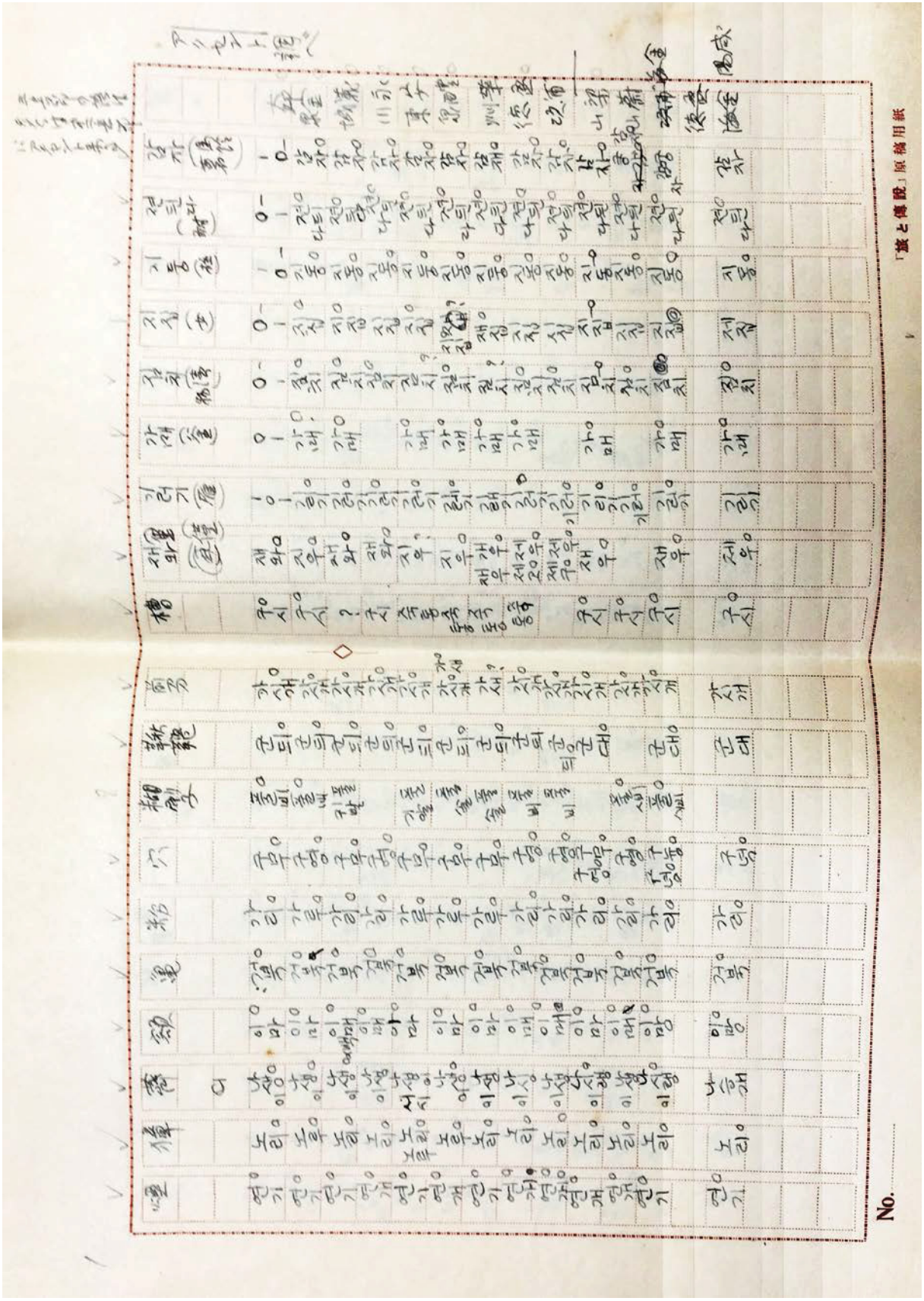


Figure 2. The first sheet of the manuscript entitled '*akusento shirabe* (accent survey)'.

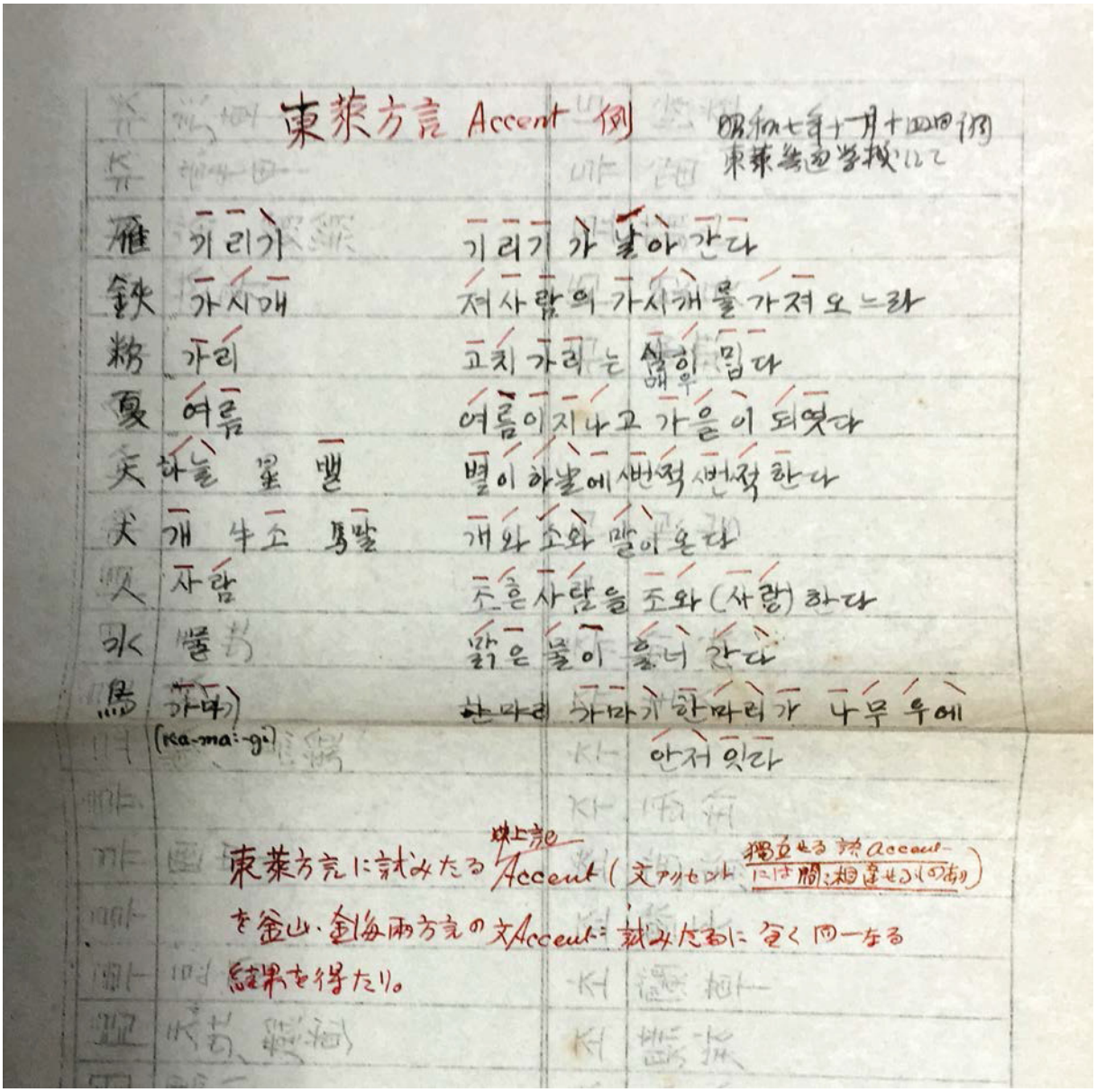


Figure 3. Manuscript entitled '*Tōrai hōgen no akusento rei* (example of accent in the Tongnae dialect)' written in November 14, 1932.

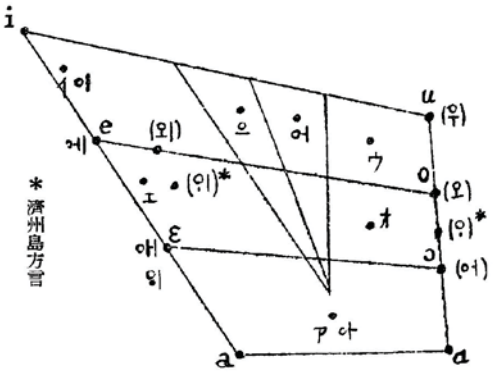


Figure 4. Vowel diagram drawn by Ogura Shinpei (Ogura (1931a)).

Geographical Distribution of Quasi-Nominal Particles in Japanese Dialects

Chitsuko Fukushima

Faculty of International Studies and Regional Development
University of Niigata Prefecture

Abstract

The author has compared the results from different geolinguistic surveys made decades apart to examine dialectal variation and changes. The data of the third nation-wide geographical survey of Japan (FPJD) are now comparable with those of the older surveys (LAJ and GAJ). With a focus on quasi-nominal particles in Niigata dialects, linguistic maps from FPJD and GAJ were integrated to show the differences in variation. In this study, the local variants of quasi-nominal particles in previous dialectal studies were examined and mapped. It was shown that a complicated system of quasi-nominal particles got simpler. The evidence of regional changes and generational changes confirmed the existence of a retreating isogloss of the peculiar form *(C)an*, ex. *aran*.

Keywords

mapping integrated geolinguistic data, generational change, retreating isogloss

1 Introduction

The author has been involved in the study of dialectal variation and changes by comparing the results from different geolinguistic surveys made decades apart (Fukushima 2013 and 2016). The data of the third nation-wide geographical survey of Japan, the Field-Research Project for Analyzing the Formation Process of Japanese Dialects (FPJD), are now available and comparable with those of the older surveys such as *Language Atlas of Japan* (LAJ) and *Grammar Atlas of Japan* (GAJ). The author was in charge of the FPJD survey in Niigata Prefecture. Linguistic maps of Niigata Prefecture were produced from the data of FPJD and GAJ with a focus on quasi-nominal particles (Fukushima 2014a and 2014b)[1]. In addition to linguistic maps of individual questionnaire items, geolinguistic data of relevant forms were integrated and plotted in the maps to show the differences in variation. In this study, previous dialectal researches were studied to find the local variants of quasi-nominal particles in order to explain what really happened in this area.

[1] The GAJ data and the FPJD data used for this study can be downloaded at the websites cited below.

2 Quasi-nominal particles in Japanese

Quasi-nominal particles in Standard Japanese are *no* or *n* as shown below. There are two typical uses of quasi-nominal particles ("Q-N p." in the gloss). The first type is the pronominal use.

(1) sore wa watasi ***no*** *da* "That is mine."
pronoun particle pronoun **Q-N p.** aux.
"it" topic "I" pronominal copula

(2) *soko ni iru* ***no*** *wa sensei da* "the one who is there is a teacher"
pronoun particle verb **Q-N p.** particle noun aux.
"there" locative "exist" pronominal topic "teacher" copula

The second type is quasi-nominal particles that are used to make functional words such as *noda*, *noni*, *node* etc.; thus called the functional use. This type often appears as *n* in conversation.

(3) *kare wa jakuba e iku* ***no*** *daroo* "I think he will probably go to the city hall"
pronoun particle noun particle verb **Q-N p.** aux.
3rd p. sing. topic "city hall" allative "go" inference

Since *no* or *n* is used as pronouns or as part of functional words which conclude nominal phrases/clauses, they are called quasi-nominal particles.

3 Change in quasi-nominal particles in Niigata dialects

It has been found out that there is much variety of quasi-nominal particles in Niigata dialects. Based on the comparison of GAJ and FPJD data, the following linguistic maps have been drawn[2] (Fukushima, in press). The maps are those of three examples shown above from two surveys (Figure 1A to 3B): 1A & 1B, WATSHI/ORE NO "mine"; 2A & 2B, ARU/IRU NO "one that exists"; and 3A & 3B, IKU NO DAROO "will probably go". The variants found in the maps include:

no, n same as standard Japanese
non a local variant of *no*
ga an older variant found in classical Japanese
gan and its variants

The form *gan* which probably originated from *ga* + *no* and its variants are found in many localities in GAJ. The variants include (cf. *aru* "to exist") :

gan *aru gan*
(V)an *aru an* (< *aru gan*)
wan *aru wan* (< *aru gan*)

[2] The software Illustrator CS2 is used to draw original maps of Figure 1A to 3B.

jan *soi jan* (< *soi gan* "it is so.")

(C)an *aran* (< *aru gan*)

gat(ta) *aru gatta* (< *aru gan da*) [Irihirose in Central Niigata]

(ak)kan *akkan* (<*aru gan*) [Awashima, a small island in the north]

Here *(V)* denotes a vowel preceding the form involved, and *(C)* denotes a consonant preceding the form involved. The form expressed as *(C)an* (denoted by a black triangle) is peculiar because it is the same form as a negative form in Western dialects when a particle follows a verb (ex. *aru gan* > *aran*). The probable direction of the change is as follows:

gan > *(V)an* > *wan* > *(C)an* (after a vowel /u/ or /o/)

> *jan* (after a vowel /i/ or /e/)

The form *(C)an* is found at three separate localities in FPJD, and its distribution in GAJ maps is similar but different. If the form *(C)an* is new, then why does it show such a distribution?

4 Integrating the distribution of quasi-nominal particles in Niigata dialects

Quasi-nominal particles were also found in other linguistic maps; thus they were integrated at each locality by adding frequencies of each form (Fukushima 2014b). When tallied, examples were divided into the two types of quasi-nominal particles: pronominal and functional. Based on the table of frequencies, a colorful pie chart showing the percentage of each form was drawn at each locality[3]. The distribution of these maps can be interpreted as follows. As for Figure 4 and 5 mapped from the GAJ data:

1. For the pronominal use, the typical dialectal form *gan* (red) and the standard form *no* (blue) are found especially in the central area of Niigata. The variants of *gan* (white, green, yellow, and pink) are found surrounding the *gan* area. The classical form *ga* (gray) is found in the western area of Niigata. The newer word *non* is used on Sado Island.
2. For the functional use, the variants of *gan* are found more than *gan*. The form *n* (black) is used more than *no*, which also applies to Standard Japanese.
3. Among the variants of *gan*, *(V)an* (white) is used at more localities than *(C)an* (yellow).

As for Figure 6 and 7 mapped from FPJD:

1. The points 1) and 2) shown by GAJ figures also apply to the FPJD figures.
2. The form *(V)an* (white) is not found in the pronominal use but is used at more localities in the functional use. The form *(C)an* (yellow) is used frequently at three localities in both of the two uses and sporadically at other localities in the functional use.

[3] The GIS software SIS 7.0 is used to draw these maps. The GAJ and FPJD maps were redrawn because the colors should be reassigned to compare with new maps and the table of frequency was revised to correct some misclassification, ex. *igannesse* [*(C)an* >> not counted] (in Murakami).

5 Variants of quasi-nominal particles shown by previous dialectal studies

Considering the change of distribution, it seems that the form *(V)an* changed into the form *(C)an* in the peripheral area in Niigata. However, why the distribution of the form *(C)an* is limited to the separate localities is still unknown. Ohno (1983:66) stated that the form *(C)an* was used after the adverbial form of a verb in the dialect of Nagaoka located in the central area of Niigata in the 1980s, ex., *asobu an* > *asoban*. Therefore, older dialectal researches were studied to search for the examples of quasi-nominal particles in local dialects.

Ohashi (2002) which included reports of dialectal surveys conducted from 1970 to 1999 was examined. The frequency of examples of quasi-nominal particles found in each report was also added up to make maps similar to those mapped from GAJ and FPJD (See Figure 8 and 9)[4]. The distribution pattern is not so different from that of Figure 4 to 7, but we can see more localities where more variants such as *gan*, *(V)an*, *wan*, *jan*, and additionally *(C)an* are used at the same time. These variants are used for different functions or with different connotations. For example, in Furutsu, Niitsu, *gan* is used for pronominal use or for special collocations such as *ganni, (V)an* is an old-fashioned form but is used most frequently, *jan* and *wan* appear as variants of *(V)an,* and the form *n* is also used as a standard form but mostly as part of the form *nda* (Ohashi 2002:208).

We can find three linguistic areas in Niigata: (Area A) most of mountainous Niigata where *gan* and its variants are used, (Area B) the western part of Niigata where *ga* is used, and (Area C) Sado and part of Niigata where *non* is used. In the peripheral parts of Area A, the ratio of *gan* decreases and that of *(C)an* increases. Actually these are also shown by Figure 4-7.

The pie chart does not show the existence of low frequency forms clearly, so the bar graph is used to show the system of quasi-nominal particles at each locality (see Figure 10). Here even one example is treated as a member of the system. Thus you can see what kind of variants of quasi-nominal particles are used in each locality. This map denotes the distribution of different sets of variants of quasi-nominal particles. The map showing different sets of variants was made for the FPJD data (see Figure 11). If you compare Figure 11 with Figure 10, you can find changes to opposite directions in Area A. At some localities close to the sea, *gan* is still used but the variants of *gan* are not. At some peripheral localities, *gan* is not used and the variants of *gan* prevail, which is the case of three separate localities where the form *(C)an* prevails. In both cases, traditional variants declined while standard variants are added.

The two figures also tell us that the form *(C)an* was used in larger area in Niigata before but now its use is limited at separate localities. This means that the isogloss of *(C)an* is not

[4] The point of Furutsu, Niitsu is moved a little westwards in the maps to show the distribution more clearly. This also applies to the following maps.

"advancing" but "retreating" according to the term Tokugawa (1964) used. However, we should not forget that the form developed its use in the local speech in such localities.

6 Regional changes and generational changes

This interpretation can be confirmed by other evidences. First, the regional changes of *(C)an* forms are shown in a map based on Ohashi (2002), GAJ, and FPJD (see Figure 12). The localities where *(C)an* is used sporadically surround or connect those where it is used frequently. Second, using two databases of dialectal conversation, NHK (1999) and NINJAL (2013), the informants' usage of *(C)an* forms were examined and plotted on a map (see Figure 13). The generational changes were found at all four localities. The older generation used *(C)an* and the younger generation did not use it. At one locality in the west, the form used by the older generation was not *(C)an* but *(C)a:* (*mitora:da* < *mitoru an da*), which could be a variant of *(C)an*; the form used by the younger generation was *ja* after /i/, *soi ja* da.

7 Conclusion

In order to understand the regional variation, the variants of quasi-nominal particles in each locality were considered. The variation is actually a combination of different sets of quasi-nominal particles, and there are transitional areas. A complicated system with more variants rolled to a simpler system with less variants in some localities but in more than one way. Figure 8, 9 and 10 show regional variants such as *(V)a*, *(V)on*, *(V)on* at some localities, but they are not found in the FPJD data. This might be another example of the system getting simpler in newer dialects. Finally, the isogloss of the peculiar form *(C)an* is retreating, not advancing, from the geographical point of view.

Acknowledgements

This paper is part of the outcomes of the collaborative research project "Field-Research Project for Analyzing the Formation Process of Japanese Dialects", carried out at the National Institute for Japanese Language and Linguistics, and was supported by JSPS Grant-in-Aid for Scientific Research (A) 23242024, (C) 25370487, and (C) 16K02688.

References

Fukushima, C. 2013. Revisiting regional variation on an island after thirty years. *Proceedings of Methods XIV: Papers from the Fourteenth International Conference on Methods in Dialectology, 2011*, 305-314. Peter Lang.

Fukushima, C. 2014a. Regional Variation and Change of Quasi-Nominal Particles in Japanese Dialects. A paper presented at ICAG-2 at Pathumwan Princess Hotel, Bangkok, Thailand on May 24, 2014. *Papers from the 2nd International Conference on Asian Geolinguistics,* 32-40.

Fukushima, C. 2014b. Niigata-ken ni okeru juntai-joshi no bumpu to henka: GAJ to FPJD o hikaku site [Distribution and changes of quasi-nominal particles in Niigata Prefecture: Comparing GAJ and FPJD]. A paper presented at gengo chirigaku forum (Forum for Linguistic Geography) at Toyama University on November 30, 2014.

Fukushima, C. 2016. Tracing real and apparent time language changes by comparing linguistic maps. In: Marie-Hélène Côté, Remco Knooihuizen & John Nerbonne (eds.), *The future of dialects*, 363–376. Language Science Press, Berlin, German.

Fukushima, C. in press. Juntai-joshi no bumpu to henka [Distribution and changes of quasi-nominal particles]. In: Takuichiro Onishi (ed.), *Kukan to jikan no naka no gengo [Language in space and time]*. Asakura Shoten, Tokyo, Japan.

National Institute of Japanese Language and Linguistics (NINJAL) (ed.). 2003. *Nihon no fukusato kotoba shusei. [Collection of Japanese dialects] Vol. 7 Gunma & Niigata*. Kokusho Kankokai, Tokyo, Japan.

National Language Research Institute (present NINJAL) (ed.) 1966-1974. *Nihon gengo chizu. [Linguistic atlas of Japan (LAJ)]*. National Printing Bureau, Tokyo, Japan.

National Language Research Institute (present NINJAL) (ed.) 1989-2006. *Hogen bumpo zenkoku chizu. [Grammar atlas of Japan (GAJ)]*. National Printing Bureau, Tokyo, Japan.

Nihon Hoso Kyokai (NHK) (ed.) 1999. *CD-ROM ban zenkoku hogen shiryo. [CD-ROM version National dialect data]*. NHK Shuppan, Tokyo, Japan.

Ohashi, K. 2002. *Niigata-ken hogen no kijutsu-teki kenkyu 2: Hyogen-ho hen [Descriptive studies of dialects in Niigata Prefecture Vol.2: Volume of Expressions]*. Koshi Shoin, Niigata, Japan.

Ohno, S. 1983. Gendai hogen ni okeru rentai kaku-joshi to juntai-joshi. [Adnominal case particles and quasi-nominal particles in contemporary dialects]. *Nihon gakuho* 2, Department of Japanese Studies, Graduate School of Letters, Osaka University.

Tokugawa, M. 1964. Togosen o megutte [On isoglosses]. In: Nihon hogen kenkyukai (ed.) *Nihon no hogen kukaku [Dialect demarcation in Japan]*. Tokyo-do, Tokyo, Japan.

Referred URLs

GAJ Data Download http://www2.ninjal.ac.jp/hogen/dp/gaj_all/gaj_all.html

FPJD Data Download http://www2.ninjal.ac.jp/hogen/dp/fpjd/fpjd_index.html

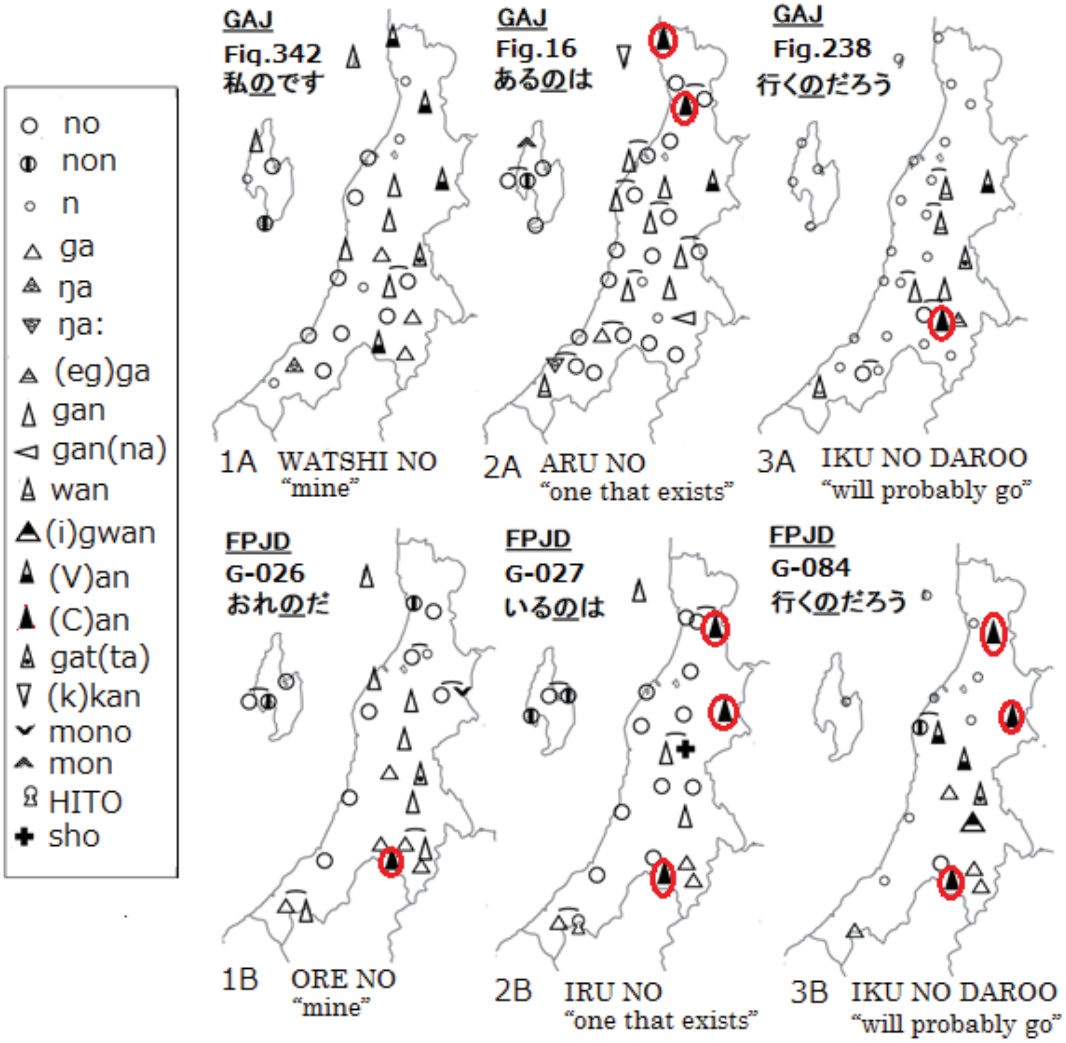


Figure 1A WATASHI NO 'mine' from GAJ

Figure 1B ORE NO "mine" from FPJD

Figure 2A ORU NO "one that exist" from GAJ

Figure 2B IRU NO "one that exist" from FPJD

Figure 3A IKU NO DAROO "will probably go" from GAJ

Figure 3B IKU NO DAROO "will probably go" from FPJD

Figures adapted from Fukushima (in press)

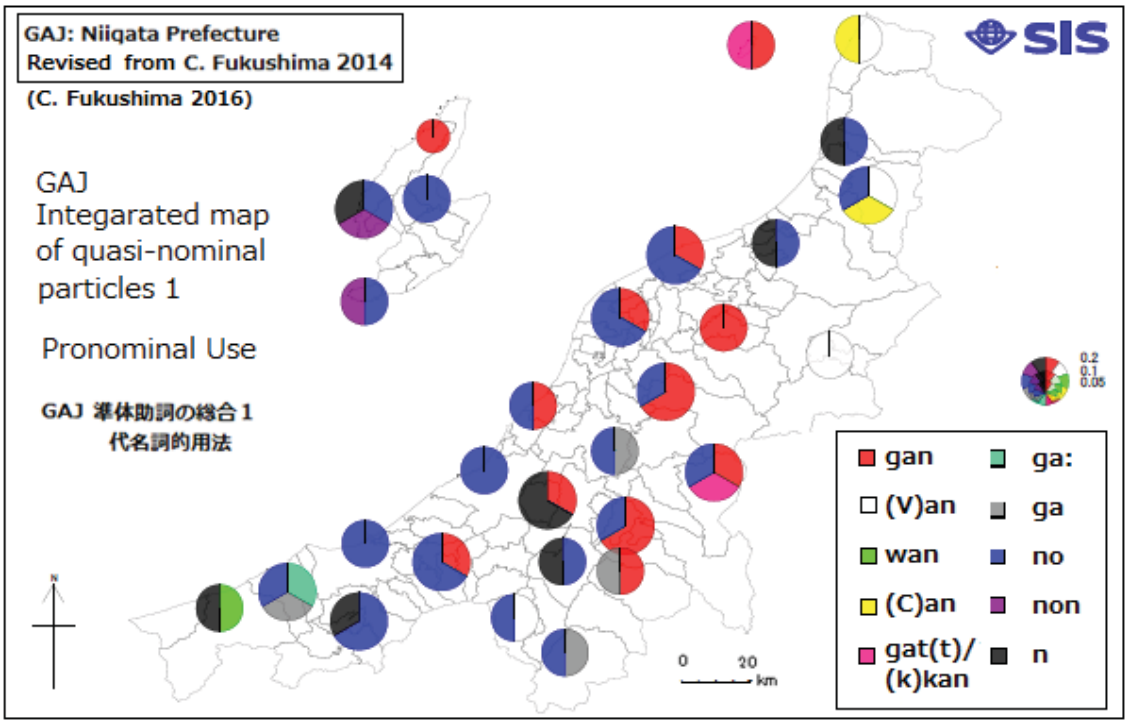


Figure 4 GAJ Integrated map of quasi-nominal particles 1: Pronominal use

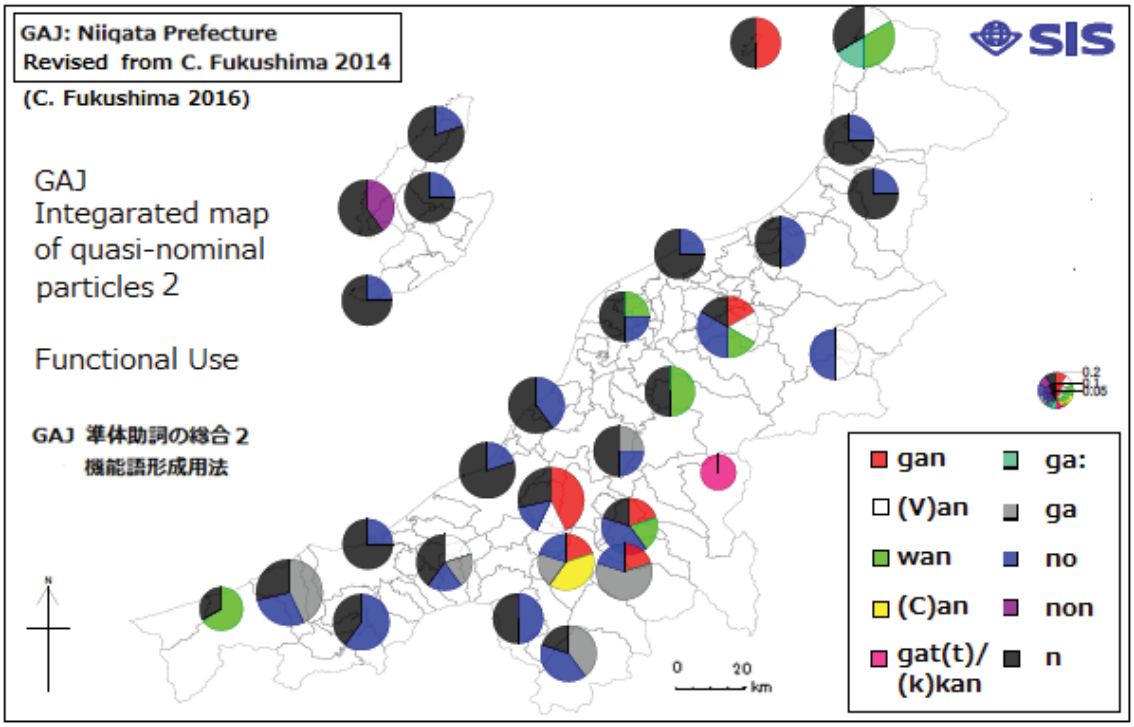


Figure 5 GAJ Integrated map of quasi-nominal particles 2: Functional use

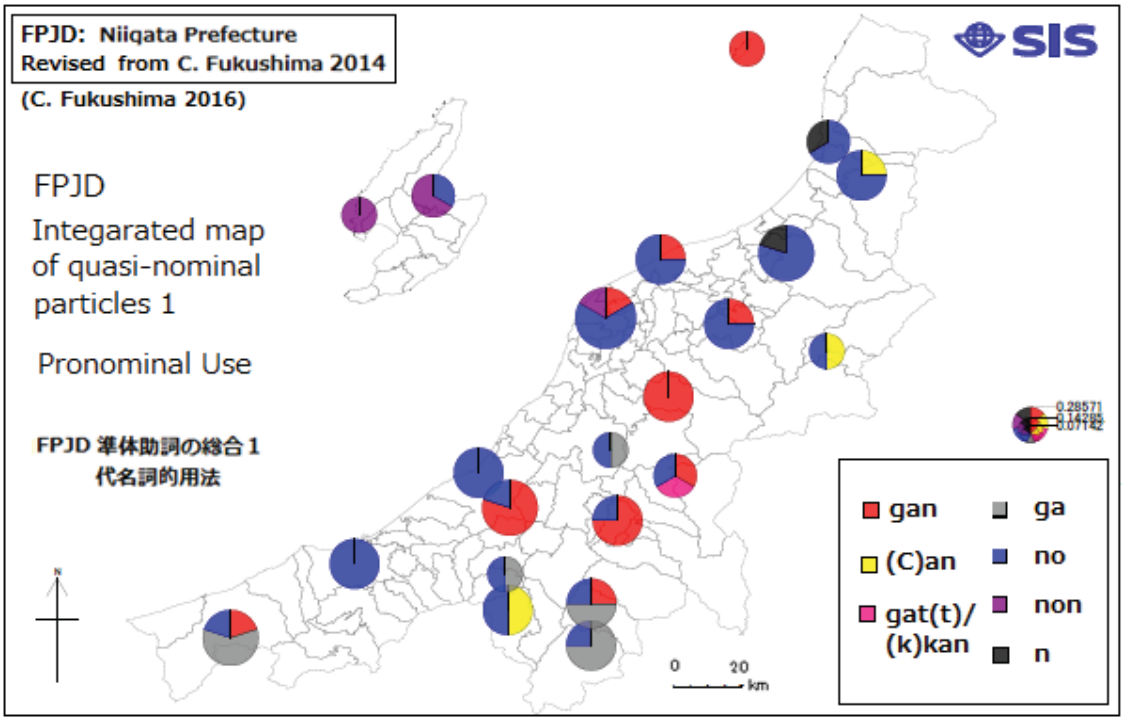


Figure 6　FPJD Integrated map of quasi-nominal particles 1: Pronominal use

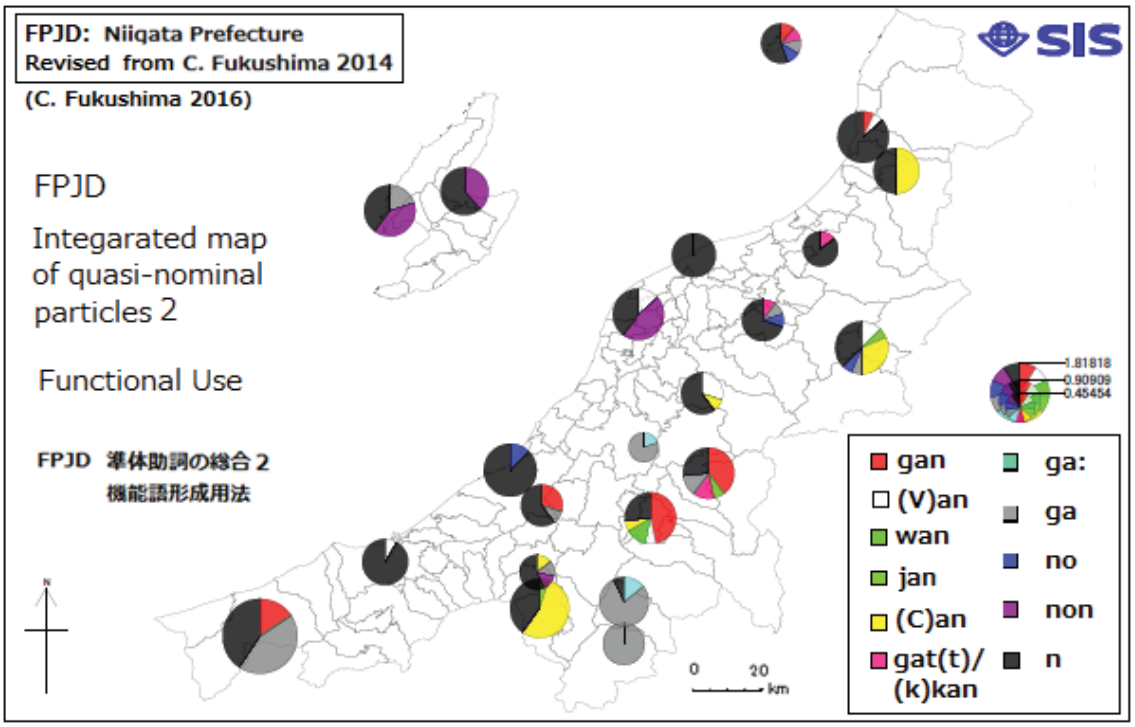


Figure 7　FPJD Integrated map of quasi-nominal particles 2: Functional use

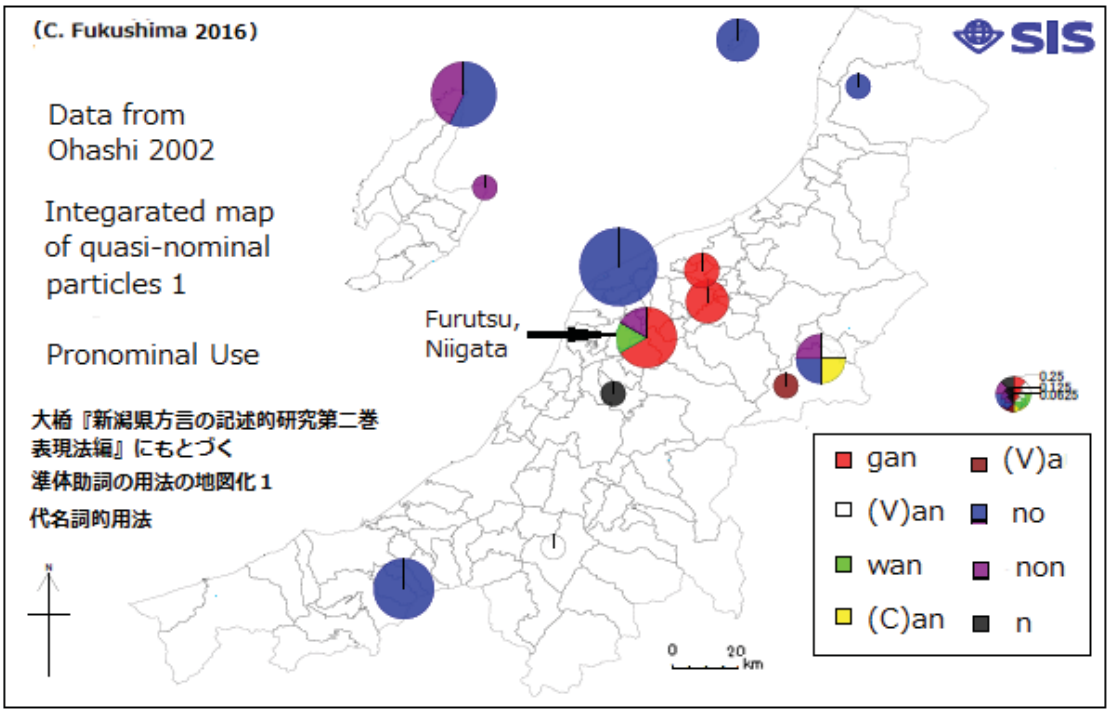


Figure 8 Integrated map of quasi-nominal particles from Ohashi data 1: Pronominal use

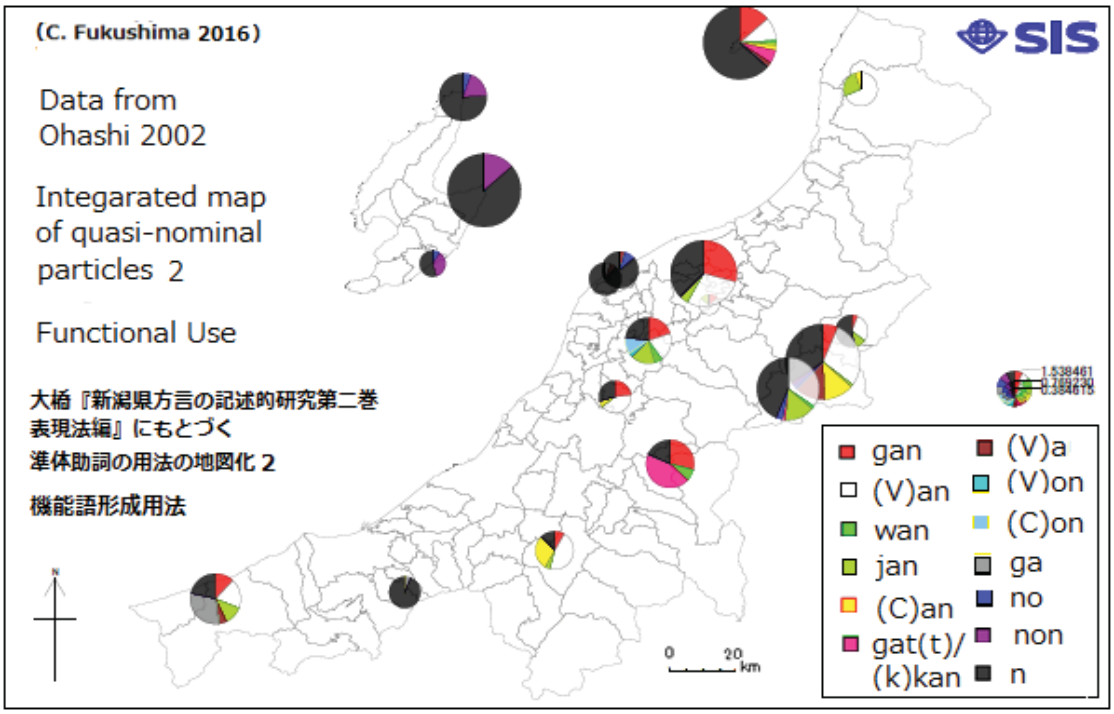


Figure 9 Integrated map of quasi-nominal particles from Ohashi data 2: Functional use

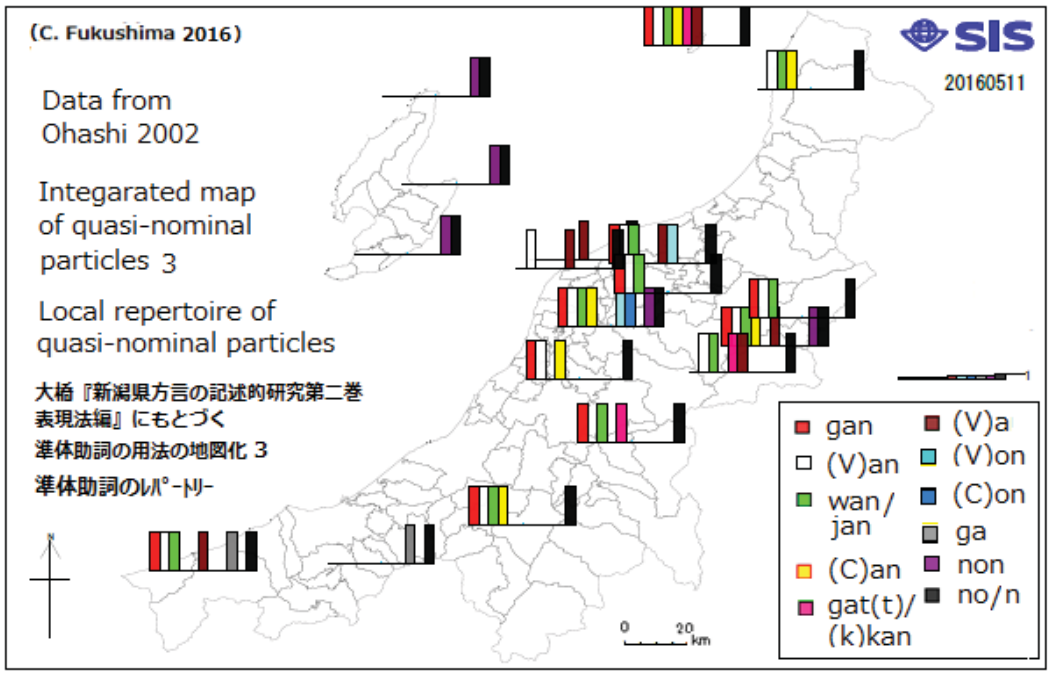


Figure 10 Integrated map of quasi-nominal particles from Ohashi data 3: Sets of variants in each locality

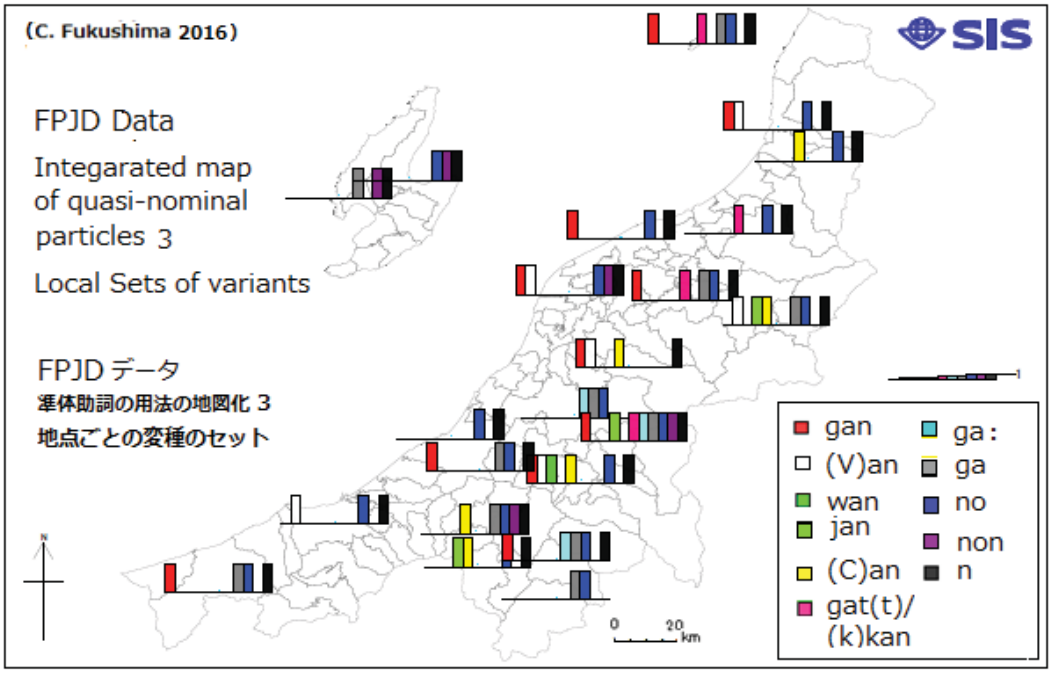


Figure 11 Integrated map of quasi-nominal particles from FPJD data 3: Sets of variants in each locality

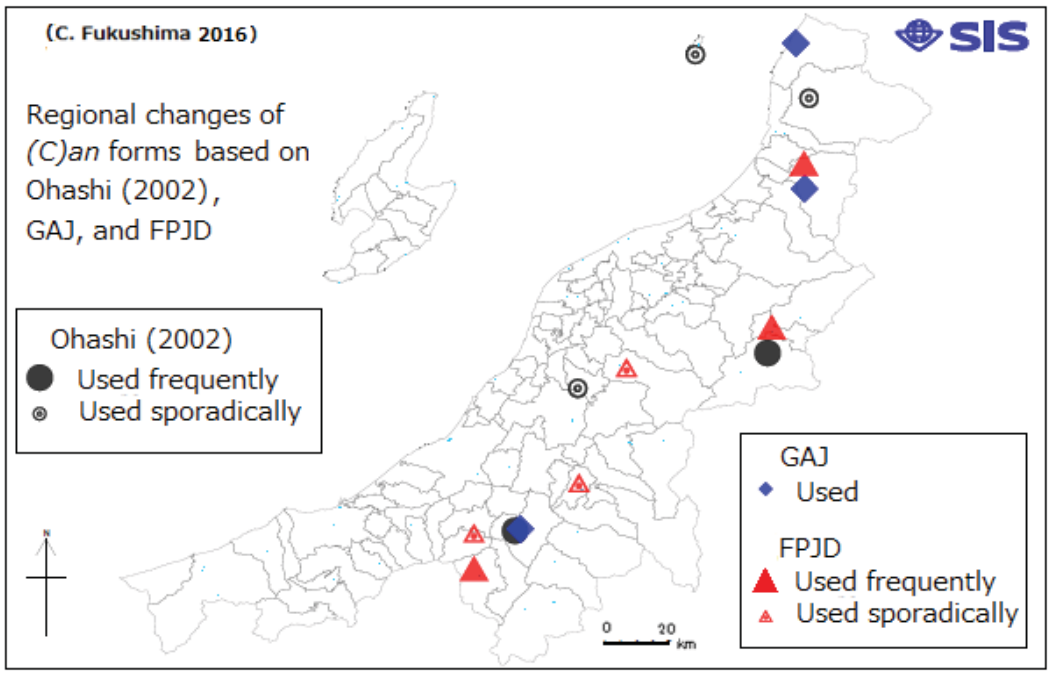


Figure 12　Regional Changes of *(C)an* forms in the data of Ohashi (2002), GAJ, and FPJD

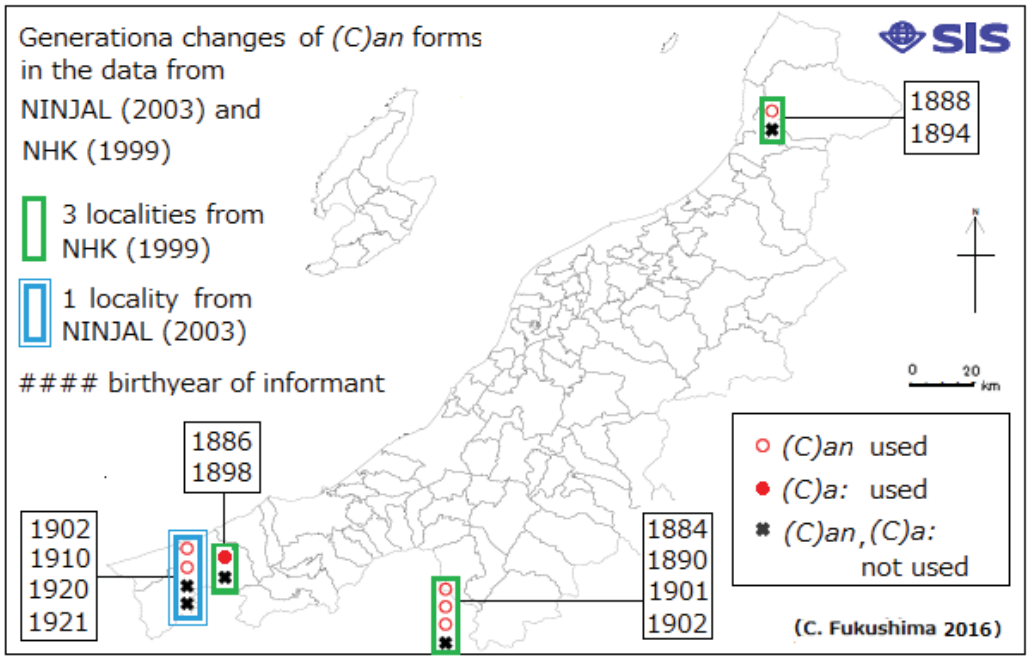


Figure 13 Generational changes of *(C)an* forms in the data from NHK (1999) and NINJAL (2003)

The Linguistic Atlas of Cambodia and Standard Khmer Language

Jean-Michel Filippi and Chem Vatho

Royal University of Phnom Penh

Master Program in Linguistics

xunwu41@gmail.com chemvatho@gmail.com

Abstract

The project of the Linguistic Atlas of Cambodia (LAC) relies on the necessity to give an accurate representation of the spatial linguistic variation on the territory of the Kingdom of Cambodia. We intend to present here our preliminary work that has consisted in defining Standard Khmer Language (SKL) from a phonetic / phonemic point of view. Our approach doesn't consist only in general terms but in order to make it easily available, we created a "Khmer Pronouncing Dictionary". As far as Khmer language is concerned, this work is the first of the kind, and it aims to describe the phonetics of both Standard Khmer and Phnom Penh dialect.

In the future, this Phonetic / phonological variation survey will be based on our Khmer Pronouncing Dictionary (KPD).

As a matter of fact, even if the notion of SKL is often quoted, its content has never been identified in a substantial way. Most of SKL definitions are negative in the way that they use but a limited number of features (the "r" treatment has been almost proverbial) to tell us what SKL is not, but certainly not what it is. Such an intuitive approach of SKL cannot replace a positive knowledge of its content.

Our approach aimed at defining 1. The very notion of SKL and its origin, 2. At describing its content and 3. At examining its relationship to Phnom Penh dialect.

The results we obtained may now allow us to use SKL as a substantial base of comparison. SKL being, at least in theory, used in the media and in teaching, it can exert an influence on the local pronunciations and as SKL can be more or less identified to Phnom Penh dialect, the general improvement of the road and communication network may lead us to contemplate the effects of a diffusion.

1. The current knowledge about Khmer geolinguistics

Khmer language spatial variation has never really been a consistent research field. Prior to 1975, a number of surveys were published about the phonetic and phonemic of Khmer. They can be divided into two types: traditional structural descriptions and a few attempts to define standard Khmer as opposed to nonstandard and / or local dialects.

1.1 Structural descriptions

Most of our current knowledge about Khmer language is related to surveys published in a rather long period that started from the early 20th century to 1975. These surveys greatly differ from each other mainly in the description of Khmer vowels, their inventory and their systemic representation. Above all, what is particularly confusing is the fact that the authors never mentioned the dialectal origin, social or spatial, of the language they described.

Amongst the first real linguistic descriptions[1], Finot (1902) and Maspéro's (1915) work are totally disembodied and it is almost impossible to guess where their informants were from.

With Martini (1946) we are better informed for the first time as he warns us that the local source of his description is Phnom Penh[2]. Unfortunately this is true only in appearance for several reasons. At that time, the biggest part of Phnom Penh population was Chinese Tew Chiu and there wasn't a deeply rooted ethnic Khmer population. I fact a closer examination of Martini's description leads us to suspect that the base of the pronunciation he described consisted of several dialects. The reason is that Martini chose his informants in the Buddhist Institute among monks and laymen with very different origins.

The next descriptions, Henderson (1952), Gorgoniyev (1966) and Khuon Sokhmapu (1970) are not really based on a real spoken language, but on the social dialect of one leading scholar of the time who was trying hard to forget his local dialectal features in order to produce a hyper normative level of language in which the influence of the Khmer script is obviously dominant.

1.2 The first mentions of "dialects"

A number of surveys mentioned in passing the word dialect. Huffman (1967) claimed to have based his description of Khmer language on "the Takeo" dialect. As a matter of fact, Huffman worked with only one of Takeo province.

In the 1960s, the problem of standard Khmer was raised for the first time by Noss (1966) and Sakamoto (1968). It didn't rely on a detailed survey of a social dialect in a given formal context and its potential location in Cambodia but in a simpler way that opposed standard Khmer to the Phnom Penh dialect. We can at least know what standard Khmer is not.

As a conclusion, we can assert that spatial linguistic variation in Cambodia has not been until now a research field and the reason of it pertains to two facts.

[1] There are other non-linguistic descriptions like the Abhitiensap that was found in Kompong Cham and that can be considered as the first text about Khmer language. The Abhitiensap is mentioned in passing by Maspéro in his grammar.

On the one side, there was never in Cambodia any appeal for a survey about language spatial variation. Variation is an essentially comparative notion and its study very often implies the existence of a center (the norm) and a periphery (various forms differing form the norm). In Cambodia each monastery had its own script rules based on the local pronunciation and that resulted in a one monastery-one peculiar script strategy, in other words: a total atomization.

A second reason is that Khmer language was often seen as a language that didn't display any serious variation worthy of study. According to Huffman (1967): "The dialects of Cambodian are remarkably homogeneous, and the degree of dialect variation seems to be roughly comparable to that between the various dialects of American English. Even the dialects spoken in northeastern Thailand and western South Vietnam appear to be mutually comprehensible with the dialects spoken in Cambodia, the differences being perhaps on the order of those between British and American English."

There is of course no evidence at all to back Huffman's assertion as notions such as "remarkably homogeneous" and "mutually comprehensible" are too ambiguous to be used as criteria (Filippi, 2007) and in fact, apart the small number of surveys at our disposal (Filippi, 2002), we have until now no real material to discuss Khmer language geographic variation.

2. The survey of spatial variation and the notion of standard Khmer

Within the Royal University of Phnom Penh Master Program in Linguistics, it has been possible to put in place an introductory course in descriptive phonetics. This course aimed at presenting the framework and principles of phonetic analysis, to train the students in the use of the International Phonetic Alphabet (IPA). Another course about language variation was put in place too, the major part of it being dedicated to the teaching of dialectology, history and methodology.

2.1 Surveying standard Khmer pronunciation

Prior to the description, we needed to elucidate the very notion of standard Khmer, at least to have a starting point to construct a questionnaire and to have a base of comparison[3]. We soon discovered that there was not, at least from an official point of view, such a notion as standard Khmer. As a matter of fact, there didn't exist any book, recommendation or any set of prescribed rules that would define a "good" pronunciation.

We then carried out a survey about testing the "desirable and less desirable ways of speaking". We selected a number of persons working in fields including administration and education. All of them held degrees such as BA, MA, in short, they were educated people from Phnom Penh and other parts of the country. The survey consisted in having the informant reading minimal pairs, words in a formal context and small passages, answering a questionnaire of more than 2000 words, and collecting informants' opinions on recordings of Khmer language in formal and less formal contexts.

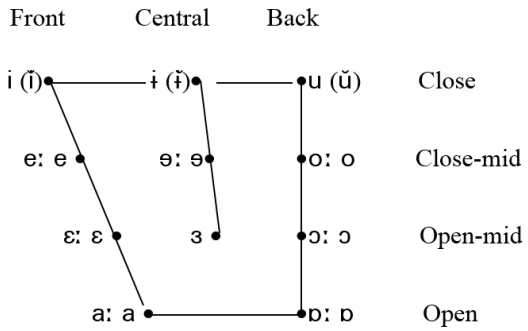
In spite of the fact that the notion of standard Khmer was never mentioned, there was a recurring theme: *"nijiey cbah"*. It means literally "to speak clearly" and it refers to is a desirable way of speaking that could be identified as standard Khmer. It corresponds roughly to the language level that should be used in formal contexts such as administration, radio, television and teaching.

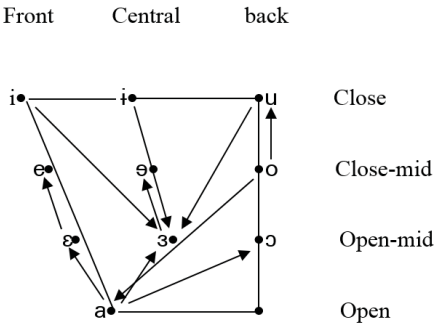
The following is an overview of our results. The data is displayed in a broad transcription. For the present purpose, we don't really need to take smaller details into consideration.

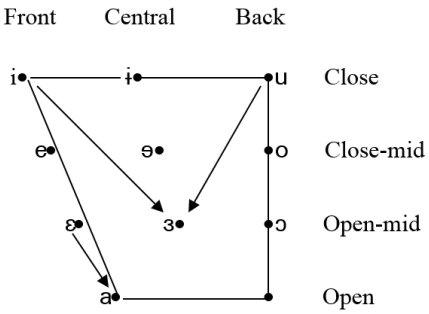
2.1.1 Consonants

	Labial	Dental	Palatal	Velar	Glottal
Plosive	p b	t d	c	k	ʔ
Nasal	m	n	ɲ	ŋ	
Trill		r			
Fricative		s			h
Approximant	β̞		j		
lateral approximant		l			

2.1.2 Vowels and diphthong







The interesting fact is that these 31 distinctive units appeared to be more or less equivalent to the Phnom Penh dialect.

2.2 The Khmer pronouncing dictionary. Standard Khmer and Phnom Penh dialect.

In order to have a clear overview of the situation, we decided to make a dictionary with all the data we got. This meant a dictionary that would display for each word: the standard pronunciation and a less formal local Phnom Penh pronunciation.

2.2.1 The dictionary

We attempted to display as many variants as we could. Below is an example of the way data has been displayed:

ការហារ kaṅhār
[kɑŋha:] [cɑŋhal]
[təŋhal]

ការហែន kaṅhaen
[kɑŋhaɛn] [kəŋhaɛn]

កែងកង keṅ kaṅ
[keŋkɑ:ŋ]

We displayed first the most formal pronunciation and in the end the most casual.

It can be noticed that an important feature that characterizes Phnom Penh casual speech is the total absence of “r” in the initial position or after a consonant. Phonetically, the non-formal variant is a complex realization that involves:

- Breathy voice
- Pitch variation

- Diphthongization
- A higher degree of closure

A few examples are given in the table below:

Khmer script	Standard pronunciation	Non standard	Meaning
ប្រាំ	Pram	p̪a̪am	Five
ត្រង់	trɑŋ	to̪a̪ŋ	Right
រយ	rɔ:j	ɦɔ̪:j	Hundred
ស្រែ	sraɛ	sɛ̪:	Rice field
ស្រោម	sraɔm	sɔ̪:m	Envelope
ត្រីយ	traəj	tə̪:j	Shore

This feature has generally been described as belonging to Phnom Penh. It is not really the case as we can find it all along the Mekong and the Bassac Rivers.

2.2.2 Standard Khmer pronunciation and its diffusion

A question remained to be asked: why would standard Khmer pronunciation be a prerequisite in order to describe Khmer language phonetic variation?

First of all, we need to indicate a pronunciation that reflects a formal use and that could be used as a base of comparison and this pronunciation has never been extensively described with its peculiar formal features as well as in opposition to a more casual one. Then, we have seen that the main phonetic features of standard Khmer are equivalent to the ones of the Phnom Penh dialect.

These two facts will have an important consequence: the diffusion of a number of features from Phnom Penh to the province.

This formal way of speaking (standard Khmer) is actually exerting a non negligible influence in the various urban centers through education and media. This doesn't mean that the existence of the local dialects is under a threat, but that a part of the educated population of the province is or will be able to use at least two variants: the formal and the local one.

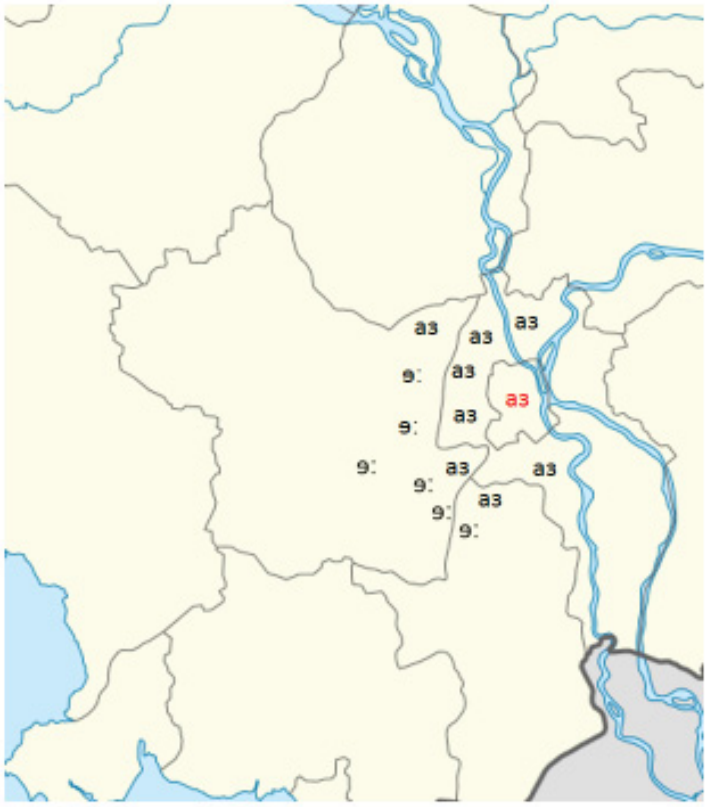
The Phnom dialect has already began spreading along the most important communication ways to the most important urban centers.

An obvious example can be found in the Kompong Spoeu Province, South West of Phnom Penh:

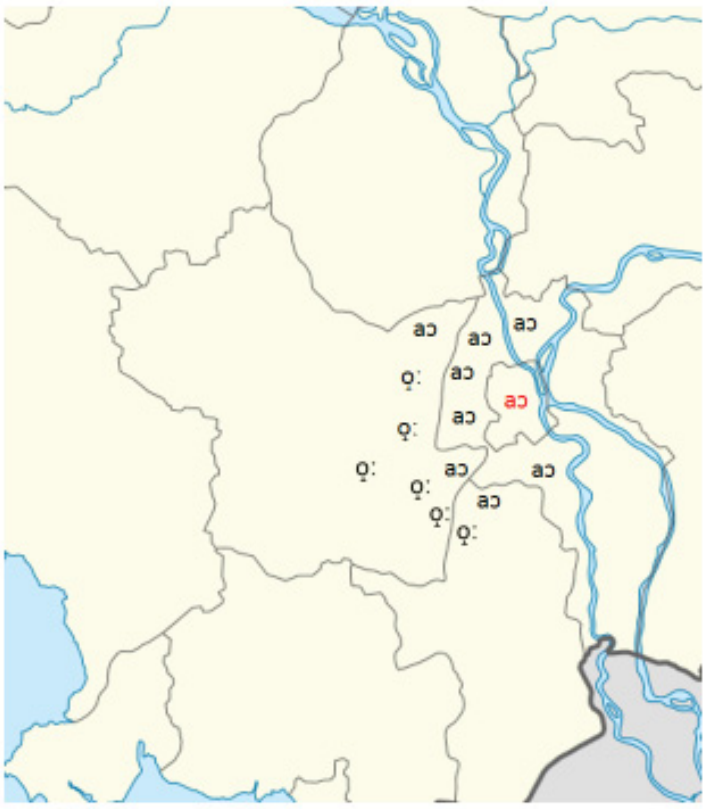
	South West	Phnom Penh
ដែរ	[dẹ:]	[daɛ]
ដើម	[dɘ:m]	[daɜm]
កោរ	[kọ:]	[kaɔ]

The Phnom Penh diphthongs are realized as monophthongues in Kompong Spoeu province. This monophthongization can be further exemplified by the following maps:

ដើម



កោរ



It is nevertheless interesting to notice that diphthongs predominate in the Kompong Spoeu province urban centers all along the national road. These centers are much more subject to Phnom Penh linguistic influence.

3. Conclusion

In its first step, this Linguistic Atlas of Cambodia (LAC) will deal primarily with phonetics and phonology. For practical reasons we have to postpone the remaining lexical, morphological and syntactic components to a further step.

References

Filippi, J-M. 2002. « La description des dialectes du Khmer ». Actes du Congrès de l'Agence Universitaire Francophone.

Filippi, J-M. 2007. *Norme et phonétisme dans le Cambodge moderne*. Udaya. Number 4.

Filippi, J-M et Hiep, C. V. 2006. *Introduction to phonetics*. Phnom Penh, FUNAN. (In Khmer language).

Finot, L. 1902. « Notre transcription du cambodgien ». BEFEO, T. 2, PP. 1-15

Gorgonijev, Ju. 1966. *Grammatika khmerskogo jazika*. Moskva, Nauka.

Henderson, E.J.A. 1952. "The Main Features of Cambodian Pronunciation". BSOAS, Vol. XIV: Part 1: 151-174.

Huffman, F.E. 1967. *An Outline of Cambodian Grammar*. Cornell University, PhD.

Khuon, S. 1970. *Système de la phonétique khmère*. Section de linguistique appliquée de la faculté des lettres et des sciences humaines de l'Université de Phnom Penh. (En langue Khmère).

Martini, F. 1946. « Aperçu phonologique du cambodgien ». BSLP 42, Fascicule 1, No.124, PP. 112-131.

Maspero, G. 1915. *Grammaire de la langue khmère (Cambodgien)*. Paris, Imprimerie nationale.

Noss, R.B. 1966. "The treatment of /r/ in two modern Khmer dialects". Studies in comparative Austro-Asiatic linguistics, Norman Zide (ed).

Sakamoto, Y. 1968. "The standard Khmer Language and its Phnom Penh Dialect". Tônan ajia Kenkyû 6 - 2.