Studies in Asian Geolinguistics, Monograph Series No. 4

Papers from the Fourth 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. 4

Papers from the Fourth International Conference on Asian Geolinguistics

Report of ILCAA JOINT RESEARCH PROJECT 2015–2017
"STUDIES IN ASIAN GEOLINGUISTICS"

First published 2018

Edited by Hiroyuki SUZUKI & Mitsuaki ENDO

ISBN 978-4-86337-291-7

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
https://publication.aa-ken.jp/

Table of Contents

Cover photo: Humanities Faculty of *Universitas Indonesia*.
Retrieved from http://static.panoramio.com/photos/large/114886365.jpg

Program of the Fourth International Conference on Asian Geolinguistics

Date: May 4 - 5, 2018

Venue: Room 4101, 4th Building, Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia

DAY 1 Friday May 4

9:15-10:15	Registration	
10:15-10:30	Opening ceremony	Dean of Faculty of Humanities; Prof. Mitsuaki Endo
10:30-12:00 Chair: Allan Lauder	Multamia R.M.T. Lauder and Allan F. Lauder	A Real Time Geolinguistic Study of Language Change in Bekasi, West Java
	Sri Munawarah	An Overview of Language Variation in Depok, West Java
	Mesiyarti Munir	Language Contact and Language Change in the History of *Betawi Ora* Dialect
12:00-13:30	Lunch & Friday prayer	
13:30-15:00 Chair: Multamia Lauder	Nor Hashimah Jalaluddin, Adriana Santa Tinggom, Siti Noraini Hamzah, Junaini Kasdan, Harishon Radzi	The Similarities and Dissimilarities of Satun Malay Spoken in Malaysia and Thailand: A Geo-linguistic Analysis
	Junaini Kasdan, Nor Hashimah Jalaluddin, Harishon Radzi, & Adriana Santa Tinggom	The Variant of Emotion Lexicons in Langkawi Island: A Geolinguistic Analysis
	Yoichi Nagato	Arabic word sharing in Indian ocean languages
15:00-15:30	Coffee break	

15:30-17:00 Chair: Shinsuke Kishie	Rei Fukui	Palatalization and hypercorrection in the history of Korean
	Mitsuaki Endo	Correlation between onset and vowel, and the principle of “wider distribution” as revealed in the changing process of the forms for “rain” in Tai-Kadai
	Chitsuko Fukushima	Variation and Change of Adjectives in Niigata Dialects
18:00-19:30	Dinner	

DAY 2 Saturday May 5

9:30-10:30 Chair: Chitsuko Fukushima	Mika Fukazawa	Geographical Distribution and Patterns of Basic Ainu Vocabulary in Hattori and Chiri (1960)
	Satoko Shirai	A geolinguistic approach to the basic vocabulary in the Western Sichuan Ethnic Corridor languages
10:30-11:00	Coffee break	
11:00-12:00 Chair: Rei Fukui	Hiroyuki Suzuki & Sonam Wangmo,	Geolinguistic approach to the route of Tibetic loanwords in Lhagang Choyu
	Shinsuke Kishie and Yukako Sakoguchi	Distribution and diffusion of the Dialectal Formatives in the Region of Seto Inland Sea
12:00-13:00	Lunch	
13:00-14:30 Chair: Nor Hashimah Jalaluddin	Agusniar Dian Savitri	Bhâsa Bhâbhien as a Madurese Language Dialect: The Case of Bawean isolect on Bawean Island
	Supriatnoko	Javanese Language in Cirebon: the Language at the Bordrline
	Suprayogi Salim	Language Distribution and Reconstruction of Migration in Northern Pringsewu, Lampung Province.

14:30-15:00	Coffee break	
15:00-16:00 Chair: Mitsuaki Endo	Nadia Almira Sagitta	The Mandar language isolate in Makassar, South Sulawesi
	Atsuko Utsumi	Address Terms in Indonesian
16:00-16:30	Closing ceremony and photo session	

A Real Time Geolinguistic Study of Language Change in Bekasi, West Java

Multamia R.M.T. Lauder and Allan F. Lauder
Department of Linguistics, Humanities Faculty, *Universitas Indonesia*

Abstract

Longitudinal geolinguistic studies can reveal the extent of change in language areas over time. This is particularly useful in situations where there is language contact as in Indonesia with its diverse patchwork of indigenous, regional languages, many of which are under pressure from language change and language loss. Geolinguistics can be of great use in understanding the processes involved in such situations. The present study focuses on the Bekasi region in West Java, an area to the east of the capital city, Jakarta. Three main languages are found there: Javanese, Sundanese and Betawi Malay. We might ask how much can these languages change over a relatively short period of time and can we find evidence of language loss.

This paper presents evidence from geolinguistic research carried out in the area over a thirty-two year period, starting in 1978. We draw on three geolinguistic studies, done over that time period: in 1978, 1998, and 2010. The three studies are comparable because they all used exactly the same locations, informants, questionnaire, data gathering and method of analysis. The use of a consistent methodology provides the empirical basis for a real time study. The comparison of the maps for the same lexical items, sampled across time, allows us to see degrees of difference in the distribution of language variation, in particular whether the degree of changes are such that the language is endangered.

Keywords

Geolinguistics, language change, rate of change, real time study, language contact

Introduction

Over the course of a human life time, anybody may observe that certain aspects of their language have changed (Sankoff, 2017). New words appear, coming into widespread use; older words become outdated and fall into disuse. This is part of the phenomenon of language change (Aitchison, 2001). From a scientific perspective, language change leads us to ask such questions as whether present day languages that resemble one another are related and share a common ancestor language, and what phenomena are involved in language change. A further question is what the rate of language change is (Nettle, 1999).

In historical linguistics, there are methods for establishing the time frames for significant change in the lexicon of language. Swadesh (1952, 1955) introduced the idea that comparison of lexical data could be used to work out the time line of language change. This was influential in the field of lexicostatistics (McMahon and McMahon, 2013). From early on, the assumption was that language change was constant, that the basic vocabulary was most resistant to change and that change occurred over periods of thousands of years (Gudschinsky, 1956).

More recently, genetic and cultural models of change have been used in computational modelling to estimate the rate of language change (Baker, 2008, Zuraw, 2006). Such studies also suggest that significant changes to a language may occur over periods of a thousand or many thousands of years.

The present study presents a review of geolinguistic studies that offer evidence from language use at three different points of time. The research was carried out in Bekasi, an area lying to the east of

Jakarta, the capital city of Indonesia. It involves three indigenous languages, all members of the Austronesian family of languages: Javanese, Sundanese and Betawi Malay.

Empirical evidence of the rate of language change, or more specifically a rapid rate of change, is important firstly, because it has implications for the assumptions and criteria that historical linguistics and computational models are based on. Secondly, identifying extremely rapid language change is important because it relates to the the well-documented phenomenon of language endangerment and language death. It is suggested that development, globalization and modernity are accelerating the pace and scale of language change to an extent that has never been seen before (Crystal, 2000, Nettle and Romaine, 2000). According to UNESCO's World Language Report (Barreña et al., 2000, Martí et al., 2005), intergenerational transmission is occurring smoothly in only 30% of the world's languages. This process is eroding linguistic diversity, leading to loss of diversity, of knowledge and weakening of local identity.

Language Change

As Aitchison (2001: ix) observes, language change is a field of study that involves a particularly wide range of areas. See for example Labov (2010). Therefore, we focus here on the aspects of language change that are relevant to our study, namely language change in relation to geolinguistics, rate of change, and language contact.

The study of language change arose in the field of historical linguistics and the two terms are closely linked (Aitchison, 2001: ix). This can be seen in the following definitions of language change:

> In historical linguistics, a general term referring to change within a language over a period of time, seen as a universal and unstoppable process (Crystal, 2008).
>
> In historical linguistics, the study of the diachronic process(es) of change in language elements and language systems (diachronic linguistics) (Bussmann, 1996).

Historical linguistics focused on tracing earlier forms of a language and showing the relationships between languages in the form of genetic trees representing members of language families. It also gave rise to geolinguistics which studies the geographic or spatial distribution of language areas (Tillery and Bailey, 2003: 351).

All aspects of language are involved in language change, though most attention has been paid to phonology and lexis, where change is most noticeable and frequent (Crystal, 2008). The present study focuses on lexical items and gathers phonological data for all variants found.

Another phenomenon that is related to language change and relevant to the present paper is that of language contact (Hickey, 2010, Thomason, 2006, Trudgill, 1989). Crystal (2008) defines language contact as:

> A term used in sociolinguistics to refer to a situation of geographical continuity or close social proximity (and thus of mutual influence) between languages or dialects.

Contact between languages provide the necessary conditions for such things as an increase in loan words or lexical borrowing, and phonological change (Labov, 2010).

Methods

This article discusses from a language change perspective three geolinguistic studies that were all performed in Bekasi, an administrative area lying to the east of Jakarta, the capital of Indonesia. The first of the the studies was Tawangsih (1978), the second, twenty years later was Andriani (1998), and the third Ramawirawan (2010). The reseach design has two elements: the specifics for each of the individual studies, and the overall method for comparing them.

At the time of the 1978 study, three languages were reported in use in the area: Javanese, Sundanese and Betawi Malay. These languages are among Indonesia's largest languages. Information on speaker populations is given in Simons and Fennig (2018). The number of speakers of Javanese is given as approximately 84 million, Sundanese 34 million, and Betawi Malay 5 million.

The individual research designs were the same. That is, the second and third studies were resurveys that kept all details of their research design the same as the first study in 1978. The details of this are as follows:

- The data was collected in a total of 34 villages which sampled the whole area evenly.

- Informants at the time of the 1978 study were middle aged people who had relatively low mobility and were considered by other members of the speech community to be 'typical' users of the language. In the second study in 1998, and the third in 2010, the researchers returned to the same household where data was gathered in 1978. In 1998, in many cases, the 1978 informant was still alive but now in old age. Where the informant was still alive, the same informant was used again. Where the original informant had passed away, a family member who was of the same or nearly the same generation as the earlier informant was used in their place. The criteria and protocols for choosing informants reduced the likelihood that demography could be a hidden variable that might have influenced the data.
- Data on a set of lexical items was elicited from the informants using a questionnaire based on the Swadesh 200 list. This list was established by Morris Swadesh for work in lexicostatistic dating (Swadesh, 1955). The questionnaire contained 218 words. The words in the list were chosen to possess two properties: universality and resistance to change. By universality, we mean that the words are found in practically all of the world's languages (Youn et al., 2016). These words are often referred to as basic words (Greenhill and Gray, 2012, Tadmor et al., 2010). Basic words are usually high frequency words (Calude and Pagel, 2014). Basic words are known to be more resistant to change than other words. The use of the Swadesh list for research in historical linguistics and language change is widespread and relevant to geolinguistic studies of Austronesian languages (Greenhill et al., 2008, Heggarty, 2010, Holman et al., 2011).
- Researchers recorded the responses immediately using IPA on the questionnaire. The data gathering sessions were done with other members of the family and community present. These provided feedback affirming or otherwise commenting on the informant's responses. These onlookers could strengthen the reliability of the responses.
- Two kinds of maps were produced from the data: isogloss maps and dialectometry maps, sometimes called spider-web maps. Each lexical item has its own map, and in addition, isoglosses are bundled according to semantic field or some other grouping. Another map is produced which aggregate the 218 basic words into one map showing all isoglosses. In the isogloss maps, the thickness of the bundles of isoglosses allow us to distinguish how much of a difference there is between adjacent areas. The other kind of map is the dialectometry maps. These show degree of sameness or difference between areas. The methods of geolinguistics help us to map the variation of lexical items (Chambers and Trudgill, 1980).
- In this paper, we compare the results from three different studies. By comparing the maps from the first study with the maps from the subsequent studies, we can see whether language boundaries have moved or if the degree of difference has increased or decreased. The larger the number of lexemes that have changed, the more extensive the change. This is the macro-context of the research design.
- The present research uses the approach of replicating an earlier study so that an earlier study can be compared with a later study holding all other variables the same. This approach is considered the most desirable but because of practical difficulties in doing it is extremely rare (Cukor-Avila and Bailey, 2013, Tillery and Bailey, 2003). The methodology of redoing a research study at a later time is known as a 'real time' study. Real time and apparent time are terms describing different research designs in sociolinguistics . Both of them concern diachronic or longitudinal phenomena, but they differ in terms of their sampling. Real time studies collect data at intervals over a particular period. They do this by replicating the research design for the first study across all subsequent studies. Apparent time studies on the other hand collect data at only one time. However, they get separate samples from different age groups in the speech community. The distribution of linguistic variation in the different age groups is a surrogate for sampling at chronological intervals (Sankoff, 2006, Turell, 2003). Further reading can be found in Cukor-Avila and Bailey (2013). The issue of real time and apparent time methodologies in geolinguistics can be found in Tillery and Bailey (2003). Discussion on the lifespan as a sampling frame is covered in Sankoff (2017).

Results

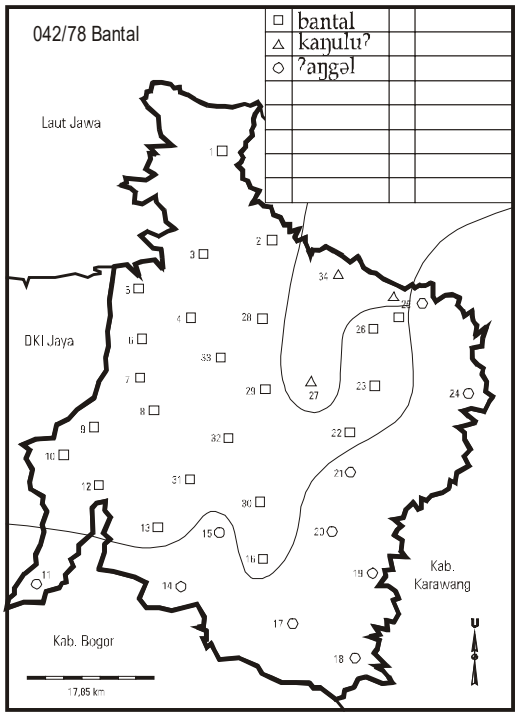
Three linguistic areas were identified in the Tawangsih (1978) study. In the west is a Betawi Malay linguistic area in the districts Cabangbungin, Babelan, Tarumajaya, Bekasi, Setu, and Pondokgede. In the south is a Sundanese linguistic area in the following districts: Pondokgede, Setu, Cibarusa, Lemahabang, and Pebayuran. There is also a Javanese linguistic area in the villages Karangharja, Sukamanah, and Sindangsari. This is a small 'speech island', surrounded by a Betawi Malay speaking area.

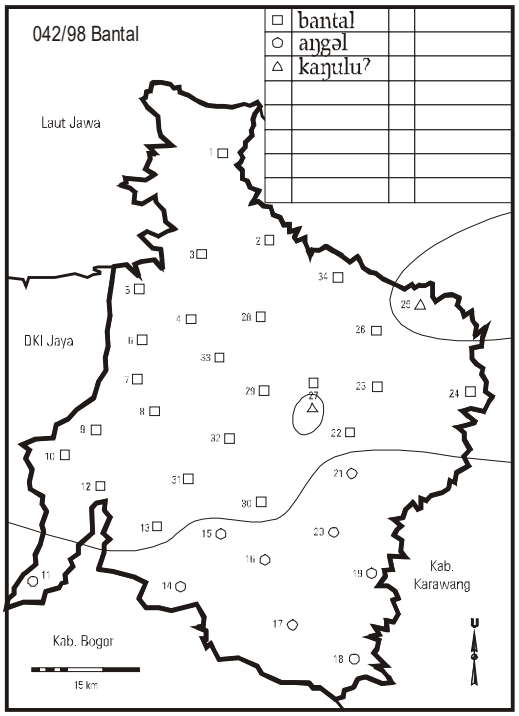
Isogloss data

A single isogloss map was made for each of the 218 lexical items in all three studies.

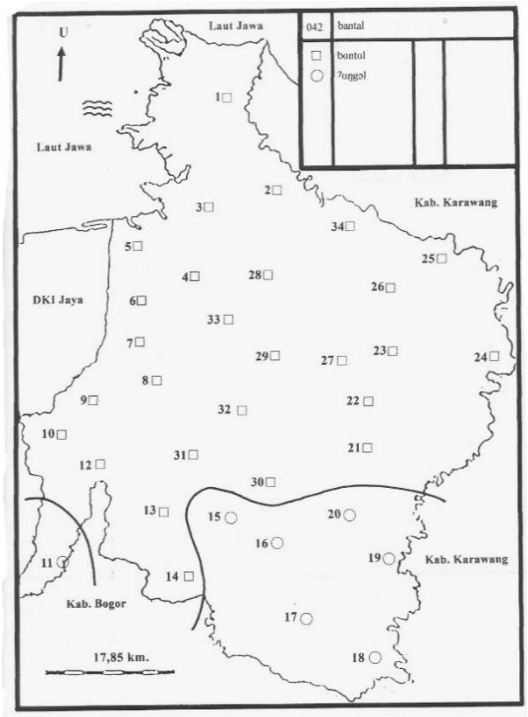
Individual isogloss map

The following maps show the isoglosses for the word BANTAL (pillow) in the 1978, 1998 and 2008 studies.





The 1978 map (left) and 1998 map (right) above show that for the word (042) BANTAL 'pillow', there are three words: *[bantal]* which detected as a Betawi Malay word used in villages 1-10, 12-13, 22-23, 25-33, *[kaɲulu]* which recognized as a Javanese word used in villages 25, 27, 34*,* and *[aŋgəl]* which is a Sundanese word used in villages 11, 24-25. The maps from 1978 and 1998 show that all three words from three different languages 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: the use of *[bantal]* spreads in villages 1-10, 12-13, 22-34, *[kaɲulu]* was only used in two villages 25, 27*,* and *[aŋgəl]* in villages in the south 11, 14-21.



The map on the left is map (042) BANTAL 'pillow' from Ramawirawan (2010). In this map we can see firstly that there are only two lexical items, namely the Betawi Malay word *[bantal]* and the Sundanese word *[aŋgəl]*. The use of the Javanese word *[kaɲulu]* in the 1978 map in villages 25, 27, dan 34 is gone. In the 1998 map, the word *[kaɲulu]* is only found in use in villages 25 and 27. This means the disappearance of the Javanese word *[kaɲulu]* and its replacement with *[bantal]*.

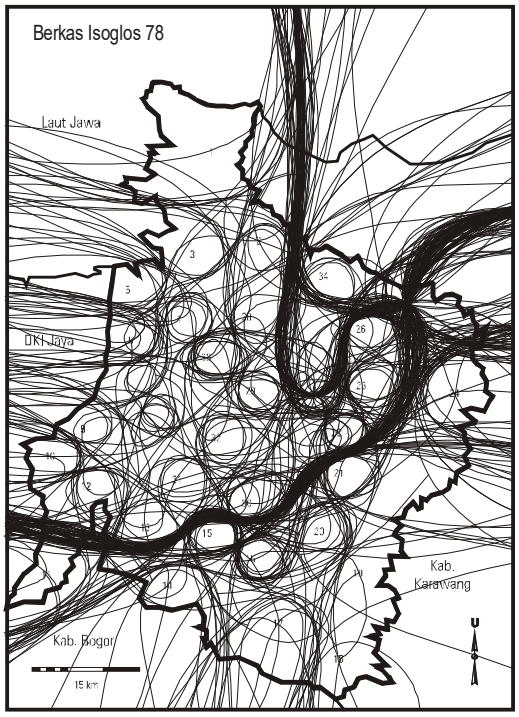
The sequence of maps shows clearly the progressive replacement of the Javanse word with a Betawi Malay one. The former Javanese speech area, villages 25, 27, and 34 formed a speech 'island', surrounded by the Betawi Malay speech area. We see the possibility that smaller speech 'islands' could be more susceptible to language change, especially if it involves intensive language contact.

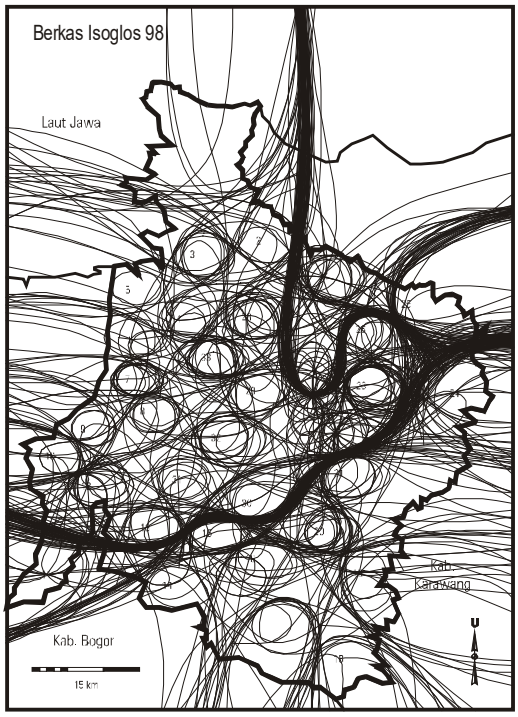
Bundle of isoglosses maps

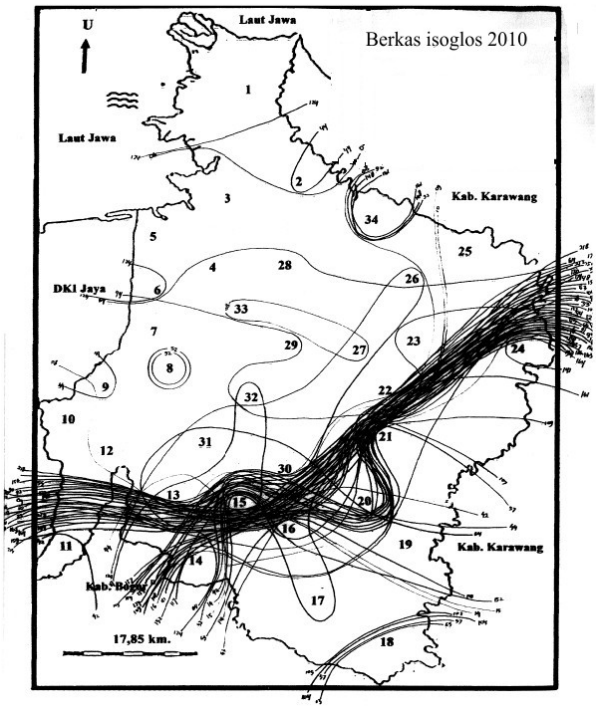
While the individual lexical item isogloss maps provide a detailed picture of lexical variation and the geospatial patterning, the bundling or grouping of all 218 maps into a single map allow us to visualize the language and dialect boundaries.

The isoglosses on the 218 language maps, taken together, from 1978 and 1998 shows no marked change for the position of the language area boundaries. Specifically, we note that the boundaries separating the three different languages areas still pass between the same data observation points. The 1998 map shows basically the same pattern of isogloss bundles as the 1978 map. There are three main language areas: Betawi Malay in the west, Sundanese in the south and Javanese in the northeast partially surrounded by Betawi Malay.

However, an important change can also be seen. 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 language contact and linguistic accommodation has occurred in the Javanese 'enclave' with a shift toward the Betawi Malay and Sundanese languages spoken by the surrounding communities.







The isogloss bundles in the 2010 map show a markedly different picture. There are now only two language areas, separated by the thick bundle of isoglosses separating the Betawi Malay area in the northwest and the Sundanese area in the southeast. The Javanese area, villages 25, 27, and 34, is gone. The Javanese language enclave, formerly surrounded by Betawi Malay is no longer distinguishable. The processes of language contact has led to language loss.

When interpreting this data, it should be borne in mind that any change shown in an isogloss map reflect the change in language use based solely on the lexical and phonological features. It does not tell us if any syntactic changes have taken place. Nor does it convey or imply anything about the attitudes of individuals or communities about these languages or their boundaries.

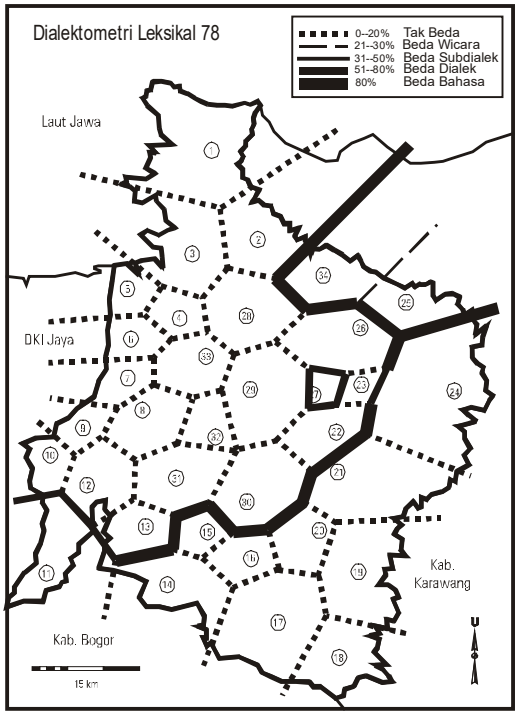
In particular, linguists should be vigilant to prevent such findings being exploited for socio-political ends where communities use the findings to support initiatives that seek formal recognition in shifts in the boundaries of their languages.

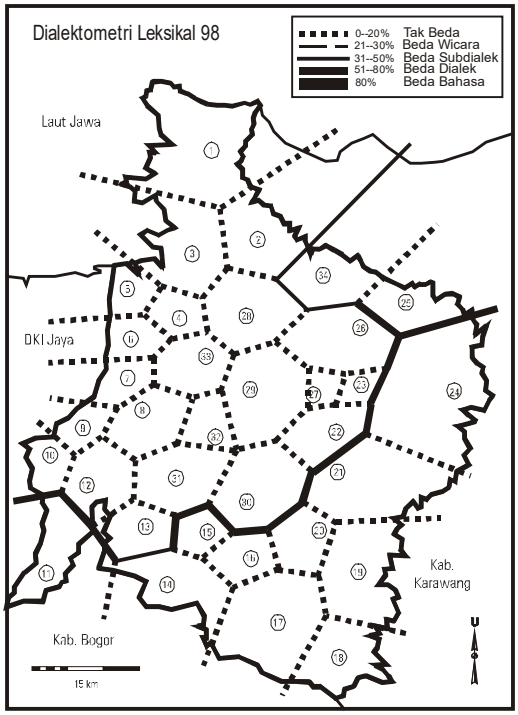
Dialectometry Data

While isogloss maps show us the geographical distribution of language variants, dialectometry gives us quantitative information about the degree of difference 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.

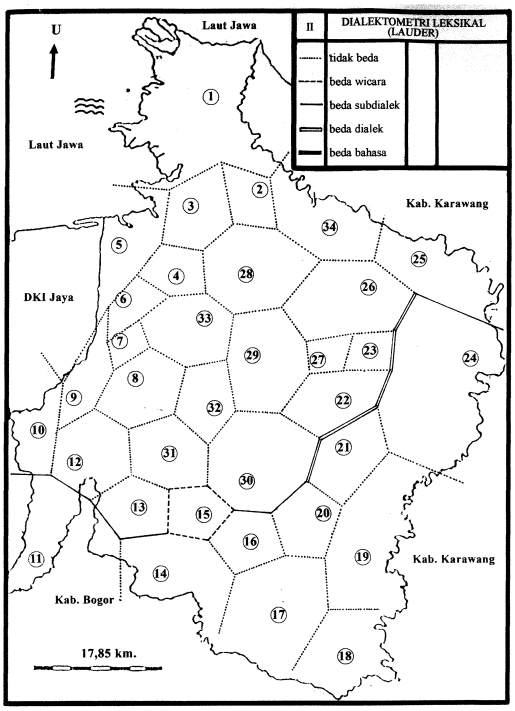
The degree of difference between contiguous language areas is derived from dialectometry analysis which is expressed as a percentage. Séguy (1973) has related the degree of difference to different kinds of linguistic variation by dividing the scale into bands. Séguy's band scale goes from no difference 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 seen 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 degree of difference.



The findings from the dialectometry analysis in Ramawirawan (2010) is shown in the 2010 map (left). It shows clearly that the degree of difference separating the three language areas has significantly lessened. In particular, the line that divided the Betawi Malay and the Sundanese areas has been replaced with a double line that indicates only a difference of dialect, and no longer a difference of language. Whereas the area in the previous studies showed three language areas, we now have two dialects. Further, there is little indication of the former existence of a Javanese language area. The difference at observation point 25, 27, and 34 in 2010 has dropped to below 20%. Such a small difference means that there is no difference in words or pronounciation between people in the former Javanese villages and the Betawi Malay villages.

This is a marked change. It most likely means that there has been intensive contact between villages, made possible by the development of better transport infrastructure, and by the use of Indonesian in schools and the mass media. The changes observed in the 2010 study exceed those described in those found in the 1998 research (Lauder and Lauder, 2016).

Discussion

The results from the real time geolinguistic research in Bekasi show clearly that during the 32 years between the 1978 and 2010 studies, language change has occurred with the Javanese, Sundanese and Betawi Malay languages.

A comparison of the 1978 with the 1998 findings show a change in the distribution of language variation. However, the boundaries between the Javanese, Sundanese and Betawi Malay areas have not moved; while the degree of difference between them has lessened. However, the research done in 2010 gives a different picture of language change. The isogloss bundles that had marked the location of the Javanese language area have been reduced to the point that the language area is seen as having disappeared. Meanwhile, the dialectometry calculations show that the degree of difference between the Javanese area with its neighbors have lessened to the smallest level on the scale. This picture is clear evidence of the process that could lead to language loss.

Throughout the historical linguistics literature, we are given to understand that language change proceeds slowly in relation to a human life-span and takes a thousand years. The studies that measure change make use of the basic words in the Swadesh list which are considered to be highly resistant to change compared with other words in the lexicon (Gudschinsky, 1956).

Language change is a phenomenon that affects all the world's languages. Languages are always subject to change. What needs to be taken into consideration is the rate of language change. Change is certain, but there remain questions about whether change is constant or fluctuates.

The real time, diachronic research done in Bekasi which used the Swadesh basic word list provides evidence that language change which seems to be working at a rate that is faster than a time frame of thousands of years. It seems that the factors that are likely responsible for speeding up the rate of language change in the multilingual Indonesian context are intense language contact between neighboring languages. Such contact is principally in the form of borrowing and assimilation as set out 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.

We also note in particular the loss of the Javanese language area. This is somewhat surprising, given that Javanese, with 84 million speakers, is the largest regional language in Indonesia and is not considered to be endangered at this time. We note that the (former) Javanese language area was an 'island', surrounded by the Betawi Malay area. As a kind of linguistic island, it was isolated from the main Javanese speaking areas. This could have been one of the reasons for its loss. A similar situation can be found in a study of Nonthaburi Malay in Thailand (Tadmor, 2004).

The findings of the geolinguistic research in Bekasi can make a contribution to our understanding of the rate of change in multilingual contexts, but things are likely to be different in monolingual areas.

Conclusion

The methods of geolinguistics, with the production of language maps, are a tried and tested method to get an accurate picture of the distribution of language variation and also, with real time studies, of language change. This makes geolinguistics of value to understanding and documenting the more than 700 regional languages found in the world's largest archipelagic country.

There is now evidence from the geolinguistic study in Bekasi that language change can occur at much more rapid rates than previously assumed, and this can be the starting point for further studies which investigate in more detail what factors are implicated in such a high rate of change. Further, and also important is to do further research that can give us a theoretical model than can account for rapid change and also provide a benchmark for rate of language change. More linguistic evidence from language use would be useful to supplement computational approaches that model language change.

Finally, the finding that this study reveals that language change may occur at a fast rate, has implications for the phenomenon of language endangerment, language loss, and language death. If language change at this rate is widespread, it means that our response to it should be prompt and effective. This is because the loss of languages is closely linked to personal identity, and the los of human knowledge, culture and heritage.

References

Aitchison, Jean. 2001. *Language Change: Progress or Decay?* 3rd edition. Cambridge: Cambridge University Press.

Andriani, Linda. 1998. Evaluasi Perubahan Leksikal: Studi Kasus Kabupaten Bekasi, Skripsi Sarjana Sastra. Program Studi Sastra Indonesia, Fakultas Sastra, Universitas Indonesia.

Baker, Adam. 2008. Computational approaches to the study of language change. *Language and Linguistics Compass*, 2(2), 289-307.

Barreña, Andoni, Etxebarria, Maitena, Idiazabal, Itziar, Juaristi, Patxi, Junyent, Carme et al. 2000. World Languages Review: A Preliminary Approach. In *Linguistic Heritage of India and Asia*. eds. Omkar N. Koul and L. Devaki, 318-336. Mysore: Central Institute of Indian Languages.

Bussmann, Hadumod. 1996. *Routledge Dictionary of Language and Linguistics*. Translated by Trauth Gregory P. and Kazzazi Kerstin. London & New York: Routledge.

Calude, Andreea S, and Pagel, Mark. 2014. Frequency of use and basic vocabulary. *Multilingual Cognition and Language Use: Processing and Typological Perspectives. Amsterdam: John Benjamins Publishing*, 45-73.

Chambers, J. K., and Trudgill, Peter. 1980. *Dialectology*. Cambridge textbooks in linguistics. Cambridge [Eng.] ; New York: Cambridge University Press.

Crystal, David. 2000. *Language Death*. Cambridge: Cambridge University Press.

Crystal, David. 2008. *A Dictionary of Linguistics and Phonetics*. 6th Edition. The Language Llibrary. Malden, MA & Oxford: Blackwell.

Cukor-Avila, Patricia, and Bailey, Guy. 2013. Real Time and Apparent Time. In *The Handbook of Language Variation and Change*, Second Edition. eds. J. K. Chambers and Natalie Schilling-Estes, 239-262. Oxford: Wiley-Blackwell.

Dixon, Robert M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
Greenhill, Simon J, and Gray, Russell D. 2012. Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation. *Diachronica*, 29(4), 523-537.
Greenhill, Simon J., Blust, Robert, and Gray, Russell D. 2008. The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics. *Evolutionary Bioinformatics*, 4, 271.
Gudschinsky, Sarah C. 1956. The ABC's of lexicostatistics (glottochronology). *Word*, 12(2), 175-210.
Heggarty, Paul. 2010. Beyond lexicostatistics: How to get more out of 'word list' comparisons. *Diachronica*, 27(2), 301-324.
Hickey, Raymond ed. 2010. *The Handbook of Language Contact*. Malden, MA and Oxford: Wiley-Blackwell.
Holman, Eric W, Brown, Cecil H, Wichmann, Søren, Müller, André, Velupillai, Viveka et al. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, 52(6), 841-875.
Labov, William. 2010. *Principles of Linguistic Change: Vol. III Cognitive and Cultural Factors*. Malden, MA & Oxford: Wiley-Blackwell.
Lauder, Multamia R.M.T., and Lauder, Allan F. 2016. A Longitudinal Geolinguistic Study of Language Change in West Java, Indonesia. ed. Jean-Michel Filippi. In *Papers from the Third International Conference on Asian Geolinguistics (ICAG 3)*: 1-18. Phnom Penh, Cambodia, 23-24 May 2016. Royal University of Phnom Penh.
Martí, Fèlix, Ortega, Paul, Idiazabal, Itziar, Barreña, Andoni, Juaristi, Patxi et al. eds. 2005. *Words and Worlds : World Languages Review*. vol. 52. *Bilingual Education and Bilingualism*. Clevedon [England] ; Buffalo [N.Y.]: Multilingual Matters.
McMahon, April, and McMahon, Robert. 2013. Lexicostatistics and Glottochronology. In *The Encyclopedia of Applied Linguistics*. ed. Carol A. Chapelle. London: Willey-Blackwell.
Nettle, Daniel. 1999. Is the rate of linguistic change constant? *Lingua*, 108(2-3), 119-136.
Nettle, Daniel, and Romaine, Suzanne. 2000. *Vanishing Voices: The Extinction of the World's Languages*. Oxford: Oxford University Press.
Ramawirawan, Catra Ditya. 2010. Perubahan Leksikal di Bekasi, Skripsi Sarjana Humaniora Program Studi Indonesia, Fakultas Ilmu Pengetahuan Budaya, Universitas Indonesia.
Sankoff, Gillian. 2006. Age: Apparent time and real time. In *Encyclopedia of Language and Linguistics*, Second Edition. Vol. 1. ed. Keith Brown, 110-116. Oxford: Elsevier.
Sankoff, Gillian. 2017. Language change across the lifespan. *Annual Review of Linguistics*.
Séguy, Jean. 1973. La Dialectométrie dans l'Atlas Linguistique de la Gascogne. In *Revue de Linguistique Romane*. 37, 1-24. Paris: Société Linguistique Romane.
Simons, Gary F., and Fennig, Charles D. eds. 2018. *Ethnologue: Languages of the World*. Twenty-first edition. Dallas, Texas: SIL International.
Swadesh, Morris. 1952. Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society*, 96(4), 452-463.
Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, 121-137.
Tadmor, Uri. 2004. Dialect endangerment: The case of Nonthaburi Malay. *Bijdragen tot de Taal-, Land-en Volkenkunde*, 160(4), 511-531.
Tadmor, Uri, Haspelmath, Martin, and Taylor, Bradley. 2010. Borrowability and the notion of basic vocabulary. *Diachronica*, 27(2), 226-246.
Tawangsih, Multamia R.M. 1978. Bahasa di Daerah Bekasi: Sebuah Kajian Lokabasa, Skripsi Sarjana Sastra. Program Studi Sastra Indonesia, Fakultas Sastra, Universitas Indonesia.
Thomason, Sarah G. 2006. Language Change and Language Contact. In *Encyclopedia of Language and Linguistics*, Second Edition. Volume 6. ed. Keith Brown, 339-346. Oxford: Elsevier.
Tillery, Jan, and Bailey, Guy. 2003. Approaches to real time in dialectology and sociolinguistics. *World Englishes*, 22(4), 351-365.
Trudgill, Peter. 1989. Contact and isolation in linguistic change. In *Language Change: Contributions to the Study of its Causes*. eds. Leiv Egil Breivik and Ernst Häkon Jahr, 227-237. Berlin and New York: De Gruyter.

Turell, Maria Teresa. 2003. Apparent and real time in studies of linguistic change and variation. *Noves SL. Revista de Sociolingüística*, 1-10.

Youn, Hyejin, Sutton, Logan, Smith, Eric, Moore, Cristopher, Wilkins, Jon F et al. 2016. On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences*, 113(7), 1766-1771.

Zuraw, Kie. 2006. Language Change: Probabilistic Models. In *Encyclopedia of Language and Linguistics*, Second Edition. Volume 6. ed. Keith Brown, 349-357. Oxford: Elsevier.

An Overview of Language Variation in Depok, West Java

Sri Munawarah[a]

[a] Department of Linguistics, Faculty of Humanities, Universitas Indonesia

Abstract

Depok as an autonomous region was projected by the government to be a residential, education, tourism, and water absorption area. Increasing educational facilities and public facilities such as malls also have a significant impact in the development of Depok City. Depok's demographic structure, along with an increasing number of means of transportation, enables high levels of interaction, even to the outskirts. Based on the history and geographical location, there is a hypothesis stating that in Depok there are speakers that speak Betawi and Sundanese. Depok's geographical location directly adjacent to DKI Jakarta, Bogor, Bekasi, and Tangerang is also expected to affect the linguistic situation. In addition, the city of Depok has transformed into an urban city. Several colleges, apartments, and star hotels have made the city of Depok as an urbanization destination.

The city of Depok is directly adjacent to the city of Jakarta and is a buffer to be directed as residential, educational, trade, tourism, and water absorption areas. Meanwhile, the southern part borders with Bogor, which was once a municipal entity. Depok's population also grew rapidly after joining with several sub-districts from Bogor. Therefore, the authors are interested to review the research that has been done Sandra Dewi (1997) and Yohanes Wahyu (2010). If Dewi focuses only on Betawi language in Depok, Wahyu (2010) examines the languages that exist in Depok along with the joining of several areas of Bogor.

The emergence of variations in language, such as dialect and its intricate dissemination, makes dialectology an appropriate science for obtaining visual data on the spatial distribution of language variations. Finally, the map in the dialectology will provide basic information to the public on the approximate number of regional cultures and their variations throughout Indonesia (Lauder, 2007: 4).

There was found the differences between the research result of Dewi (1997) with research result of Wahyu (2010). The research conducted by Dewi showed that there is one language in Depok, that is Betawi Ora. After thirteen years since that research, Wahyu showed that in Depok there are two dialects that have quite high differences. Different dialects in Depok can be said as different languages. However, the reality on the ground shows that Sundanese in Depok is contaminated with Betawi so that the difference level only reaches dialect differences level (less than 70%).

Keywords: Language variation, dialectology, Betawi language, Sundanese language

1 Introduction

Regional languages are a valuable asset of the nation. In the explanation of Article 36 of the 1945 Constitution it is implied that the state respects and maintains the languages that its supporting societies recognize as a symbol of social identity. Regional languages continue to grow along with the development of the area itself, and can even affect the surrounding area. For example, Depok, which originated from a sub-district in Kewedanaan neighborhood (Sub Regency) of Parung District, Bogor Regency, experienced astonishing development in various sectors: development, economy, education, and the development of language. Depok has continued to evolve from the status of the sub-district into

an administrative city (1981) and more recently changed again to a municipality (1999) and then a subdistrict (2010).

Linguistic issues are an ever-growing issue. Various factors, such as social events, politics, history, and the dynamics of the life of Indonesians when interacting or in language contact situations with regional and foreign languages, encourage the need for a guide in managing language as a national cultural treasure. Similarly, language development in Depok may also experience developments that are influenced by these factors.

Before holding the status of the district, Depok was only 12 km ^{2} of private land bought by Dutch merchants. Thus, it is possible that formerly the Dutch language was used in Depok. However, after becoming part of the district in Kawedanan in the Bogor area, Depok is expected to see many speakers speak Sundanese. Rapid progress has made Depok change its status back into an administrative city. In addition, Depok has also served as an autonomous region as well as a projected as a residential area, while Jakarta has become the center of government.

According to Irsyam (2017), formerly Depok was only a remote village in the middle of nowhere. On May 18, 1696, a Dutch merchant who was a former high-ranking VOC official, Cornelis Cahstelein, bought land in Depok and a few areas of South Jakarta and Ratujaya Bojong Gede. In 1871, the Dutch Government allowed the Depok region to form its own government and president. The government was a *Gementee* ruled by a president as the highest governing body.

Until 1942, *Gemeente* Depok held power over sub-districts under the mandate (9 foremen) and assisted by village police barracks and *cumitir* or minister of barns. Initially, the total area of the *Gemeente* Depok territory was only 1,244 hectares, but was later abolished in 1952 after an agreement of the right of release between the Government of Indonesia and the leadership of *Gemeente* Depok. (Irsyam, 2017)

In naming Depok, according to local elders in Depok, the name *Depok* comes from the word *De Volk* meaning 'settlement to be proud of'. There is also a suggestion made by Chastelein that Depok stands for De Everste Protestante Organisatie van Christianen (Irsyam, 2017). Meanwhile, in Sundanese, the word *depok* means 'hermitage'. Until now the origin of the name Depok is still debated. The government's projection of Depok as an autonomous region for resettlement, education, tourism, and water absorption creates new speculation that can affect language mapping. Increased access to education and public facilities such as malls also provide a significant impact communications between different language groups. Depok's demographic situatin sees an increasing means of transportation which allow high levels of interaction even in peripheral areas. Based on its history and geographical location a hypothesis has emerged that states that in Depok there are speakers who speak Betawi and Sundanese. Depok's geographical location is directly adjacent to DKI Jakarta, Bogor, Bekasi, and Tangerang and this is also expected to affect the language situation. In addition, the city of Depok has transformed into an urban city. Several colleges and hotels have made Depok City have attracted people and thus accelerate urbanization.

The emergence of variations in language, such as dialects and its spread, makes dialectology an appropriate field which can provide visual data on the spatial distribution of language variation. The geolinguistic maps can provide basic information on the approximate number of regional cultures and languages and their variations throughout Indonesia (Lauder, 2007: 4).

As a rule, dialectology research focuses on rural areas as the object of its research. However, the present research is an urban dialectology study. The language situation in rural and urban areas are different. The condition of the language in urban areas is more complex than in rural ones because it can be affected by migration. The growth of the city of Depok as a settlement city, has experienced population movement from the towns around Depok and this has led to the contact of languages that became the object of dialectology research.

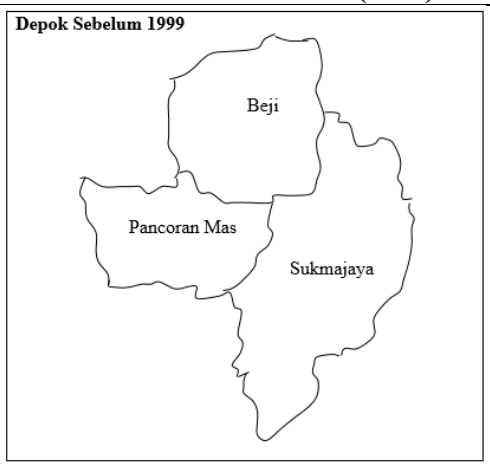
According to Holmes (2013: 131), language is used to characterize people into certain groups and to create different aspects as part of their social identity. The division of community groups also creates language variations of each group. Language variations are classified into two types, namely regional variation and social variation. Communities in a particular region have different characteristics in their language to people in other areas. This is what regional variations mean. This variation is

influenced by geography (Wardhaugh, 2006: 40). Meanwhile, social variations are influenced by various social factors, such as age, sex, occupation, and social class.

The city of Depok is directly adjacent to the city of Jakarta and acts as a buffer zone with functions such as residential, educational, trade, tourism, and water catchment area. Meanwhile, the southern part borders Bogor, which was once a municipality. Depok's population also grew rapidly after joining several sub-districts from Bogor. Therefore, the authors are interested to review the research that has been done Sandra Dewi (1997) and Yohannes Wahyu (2010). While Dewi focuses only on Betawi language in Depok, Wahyu (2010) examines the languages that are used in Depok along with the joining of several areas of Bogor. In addition, Dewi's research was conducted in 1997 when Depok was an administrative city (kotif). Meanwhile, 13 years later Depok has experienced an expansion of territory and seen its status change into a municipality. Dewi's research shows that there is only one language in Depok, Betawi Ora. Wahyu (2010) meanwhile, concludes that in Depok there are two dialects that have a significantly high difference. Different dialects in Depok can be said to be different languages because the level of difference reaches almost 70% (minimum limit to be expressed as a language difference according to Lauder (2007)). The linguistic thing that is interesting to research in Depok City is the change of linguistic situation that followed Depok city's ongoing urbanization. .

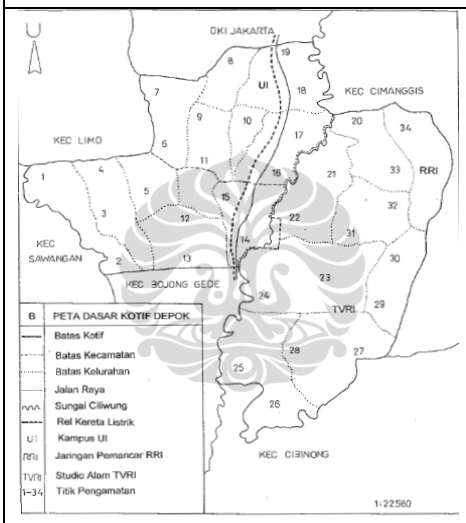
2 Comparison of Research by Dewi (1997) and Wahyu (2010)

Presently, Depok has changed from being an administrative city *(kotif)* to a municipality. The region also experienced an expansion and increase in the number of sub-districts. At the time of research conducted by Dewi (1997), Depok consisted of only three districts, namely Beji Sub-district, Pancoran Mas Sub-district, and District Sukmajaya.

Dewi Observation Area (1997)	Wahyu Observation Area (2010)
	

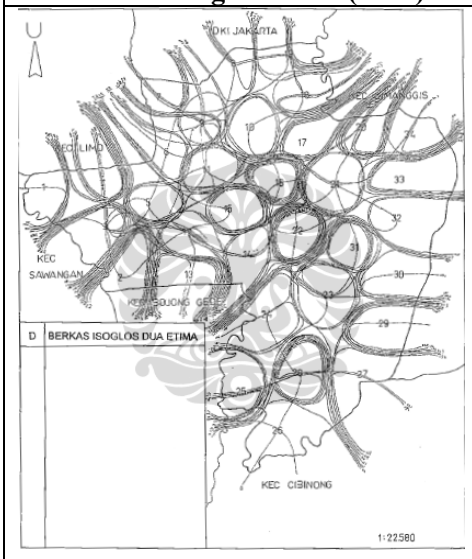
1.1 Areas of Observation

Two years after the research was completed (1999) , the number of sub-districts in Depok increased by three. These were Cimanggis Sub-District, Sawangan Sub-district and Limo Sub-district. It also added five fractional villages from Bojong Gede Sub-district, Bogor. Today, Depok already has 11 districts resulting from the division of working area based on the decision of the Depok local government in 2007.

Dewi's 34 Observation Points (1997)	Wahyu's 11 Observation Points (2010)
	

1.2 Points of Observation

The isoglos file that has been created shows a very thick bundle at point 8, the point of observation Tapos district. At other points, there are very few isoglos bundles. From the results of making isoglos bundles can be concluded that there is a fairly high difference between Tapos and other observation points in Depok. Bundles of isogloss lines that are not so thick in other points of observation illustrate that there is little difference there in terms of vocabulary.

Dewi isoglosse files (1997)	Wahyu isogloss file (2010)
	

1.3 Isogloss Files

Dewi (1997) says that the result of isogloss compilation (isogloss file) and the calculation of dialectometry can lead to the conclusion that the language in Depok is homogeneous. It is a Betawi subdialect suburb. This is unlike the results in Wahyu (2010) who says that in Depok there is only one language with two dialects Betawi Ora and Sundanese. Sundanese in Depok, particularly in Tapos (TP 8), is already mixed with Betawi because of the abundance of indigenous Betawi-speaking people who have come from other sub-districts. However, based on dialectometric calculations, the difference

between the Sukmajaya point with Tapos (TP 8) is 67, 53%. That number is just short of the 70% which would make the varieties in Sukmajaya and Tapos different languages . In Tapos subdistrict, there are only two neighborhoods where Sundanese is spoken, namely Leuwinanggung and Cimpaeun Urban Village.

Wahyu (2010) says that there are three languages in Depok, namely Sundanese in the east; Betawi Ora in the west, and the Betawi dialect dialect not found in 'ora' vocabulary in the south, north and center. Meanwhile, Dewi (1997) states there is one language, the Outer Betawi dialect (Betawi Pinggiran) in almost all areas. Wahyu also says that from the eleven observation points observed , only in the Tapos (TP 8) points is Sundanese used, while Betawi is in use in an area spread across ten points, namely Beji (TP 1 1), Pancoran Mas (TP 2), Cilodong (TP 3), Sukmajaya (TP 4), Limo (TP 5), Sawangan (TP 6), Cipayung, (TP 7), Cimanggis (TP 9), Cinere (TP 10), and Bojongsari (TP 11).

This means there is a difference between the results of Dewi (1997) and Wahyu (2010). Dewi's research shows that there is one language in Depok, Betawi Ora. After thirteen years, Wahyu finds that there are two dialects that have quite a high degree of difference. Variation at the level of dialect in Depok can be considered a difference of language. However, the reality on the ground shows that Sundanese in Depok now uses lexical items from Betawi so that the difference between the two is that of dialect and does not reach the difference of language.

3 Changes of Language Variation and Distribution in Depok City

Dewi (1997) research result shows there is only one language used in Depok, the Betawi language's peripheral dialect. At that time, Depok was still an administrative city. Muhadjir (1984: 5) said in broad outline the dialect of Jakarta is divided into two subdialects called inner Betawi and outer Betawi. Outer Betawi is found in the outskirts of Jakarta, and are often called "Betawi Ora" or "Malay Ora". The name Betawi Ora comes from the characteristic the use of the Javanese word '*ora*'' not. This illustrates that in the subdialects of the outer Betawi, we find words derived from the Javanese language and which are not used in subdialect Inner Betawi. In addition, there are phonological differences that distinguish the two dialects. For example in the Betawi Dalam dialect most of the final vowels, which in Indonesian are pronounced a, are pronounced [*ɛ*], whereas in the Betawi Pinggiran dialect they are pronounced with [*ah*] as follows.

Indonesian	**Inside the City Sub-dialect**		**Border Sub-dialect**	
Saya	→	*sayɛ*	→	*sayah*
Apa	→	*apɛ*	→	*apah*
Maŋga	→	*maŋgɛ*	→	*maŋgah*

In addition, the consonants *b, d, g,* and *h* rarely appear at the end position in the Betawi Dalam subdialect, whereas in the fourth suburb the consonant may appear in that position (Muhadjir, 1972).

In Depok, Betawi Ora language is used. Although the word *ora* as a marker of Betawi Ora was not found in the whole area of Depok, where consonants *b, d, g,* and *h* as a marker of the Pinggiran dialect can appear in the final positions are found, for example *lalab 'salad', tu ŋ ked* 'sticks ' *, gludug* ' lightning '. The word 'no' in Betawi in Depok is realized with the words *kaga?* , Ə *ŋ ga?* , and *ora*.

The study of Wahyu (2010) gives almost the same picture, that the Betawi language used by the people of Depok is the Betawi Pinggiran dialect. However, it is found in only one observation point (observation point 8 (Tapos District)) that tends to use Sundanese. Geographically, the village is located in the border area between Depok City and Bogor City. Apparently, the change of Depok's status from sub-district to administrative city also changed the boundaries of language and dialect areas.

3.1 Findings of Non-Betawi Vocabulary in Depok

Castles says the composition of the Betawi population since the 17th century has been comprised of diverse ethnic groups that have come together as a result of migration from within and from outside Indonesia (Muhadjir, 1984: 1). Residents of Jakarta City and the suburbs, according to Castles cited by

Muhadjir, include among others consist of Europeans, Chinese, Arabs, Javanese, Sundanese, South Sulawesi, Sumbawa, and Ambon (Muhadjir, 1984: 5). The varied composition of the origin of residents of Jakarta and the suburbs, also affect the socio-cultural population, and the language used. Hans Kahler said the Malay used there is a Malay dialect with elements of Balinese, Javanese, Sundanese, Chinese, Arabic, Portuguese, Dutch, and English (Muhadjir, 1984: 5). The findings of language vocabulary in Depok are presumably derived from the vocabulary of Java, Sundanese, and some Balinese and Dutch.

3.1.1 Javanese vocabulary

Depok City including West Java Province is directly adjacent to DKI Jakarta Province in the north. In terms of geography, it is estimated that the Sundanese language is used in most areas of West Java Province as the largest "contributor" to the language Betawi Pinggiran dialect used in Depok. However, in fact, the vocabulary derived from the Javanese language is more prevalent in Betawi in Depok than in the Sundanese vocabulary.

The abundance of Javanese vocabulary as assumed by Muhadjir (1984) in Depok was known by the time of the VOC Governor Jan Pieterszoon Coen was in power, during which the Mataram forces attacked Batavia several times. However, these attacks failed. When the troops of Mataram led by Sultan Agung attacked Batavia, Muaraberes, a small port on the banks of the Ciliwung River (not far from Depok) was used as a concentration center of Mataram troops. It is along this Ciliwung River Sultan Agung came to Batavia. Consequently, it is believed that Depok is populated by the remnants of the Mataram troops who spoke Javanese. Examples of Javanese vocabulary in Depok Betawi is as follows.

VOCABULARY	TRANSLATIONS
Bocah	‘child’
lanaŋ	‘boy’
Rowahan	‘kenduri’
jə̲riji	‘finger’
Gandul	‘papaya’
Wadon	‘woman’
Gludug	‘lightning’

3.1.2 Sundanese Vocabulary

Although not as much as the vocabulary of the Javanese language, Sundanese vocabulary can also be found in Betawi in Depok. Historically, Depok was once ruled by the Sunda Kingdom. This can be proven from the archeological remains of the Sunda Kingdom in Depok. The names of the villages in Depok still use Sundanese names, such as Parung Bingung, Parung Balimbing, Cisalak, Cilodong. Below is an example of Betawi is Depok which is considered to be derived from Sundanese.

VOCABULARY	TRANSLATIONS
uwa?	‘elder brother from parents’
paraji?	‘baby shaman’
sauŋ	‘loom’
baŋor	‘naughty’
paraŋ	‘sickle’
ə̲nɛŋ	‘call for little girls’

The influence of vocabulary from other regional languages is found only one or two words. For example, the word *kumpi* ' *grand dad* ' and 'great-grandson', which is recognized as Betawi language (Malay language Jakarta) allegedly comes from the Balinese language. (Muhadjir, et al., 1979: 1)

4 Conclusion

Based on the discussion that has been done, it can be concluded that the research results of Dewi (1997) showed that there is only one language in Depok, the Betawi Ora. Thirteen years since then, it is shown that there are two dialects that have a high enough differences in Depok. Based on the dialectometric calculation, the dialect differences in Depok almost can be said as two different languages, although has not yet reached 70% of differences. Some changes of research result of Sandra Dewi and research of Wahyu can be seen as follows.

Some Differences in the Research Results of Dewi and Wahyu

	Dewi (1997)	**Wahyu (2010)**
1.	The research area consiststs of three districts: Beji, Pancoran Mas, and Sukmajaya	The research area consiststs of eleven districts: Beji, Pancoran Mas, Cilodong, Sukmajaya, Limo, Sawangan, Cipayung, Tapos, Cimanggis, Cinere, and Bojongsari
2.	The result is one language, the Betawi language in the outskirts of almost all regions	The result is two languages, Betawi and Sundanese. However, the difference between the two is less than that of language.
3.	The use of the word 'ora' in the western part of Pancoran Mas and Beji, and southwest of Sukmajaya	No use of the word 'ora' in the three research districts Sandra Dewi. The word 'ora' is only found in the westernmost part of present-day Depok, the Bojongsari subdistrict (11)
4.	Homogeneous language, Betawi Pinggiran dialect.	Heterogeneous situation with Betawi and Sundanese languages. However, through intense language contact and borrowing, the difference between them is at the level of dialect

References

Ayatrohaedi. 1979. *Pedoman Penelitian Dialektologi*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.

_____. 1997. *Kamus Khusus Penelitian Kekerabatan dan Pemeta Bahasa-Bahasa Daerah di Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

_____. 2002. *Pedoman Penelitian Dialektologi*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.

Balai Penelitaian Bahasa Singaraja. 1975. *Kamus Indonesia-Bali*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Bloomfield, L. 1992. *Language: Bahasa*. Jakarta: Gramedia Pustaka Utama.Chaer, Abdul. 1976. *Kamus Dialek Jakarta.* Jakarta: Ende-Flores: Nusa Indah.

Chambers, J.K. dan Peter Trudgill. 2004. *Dialectology Second Edition*. United Kingdom: Cambridge University Press.

Dewi, Sandra. 1997. "Bahasa Betawi di Kotif Depok Kabupaten Bogor". Skripsi Fakultas Sastra Universitas Indonesia.

Francis, W.N. 1983. *Dialectology: An Introduction*. United States: Longman Group Limited.

Grijns, C.D.. 1991. *Kajian Bahasa Melayu Betawi*. Jakarta: Balai Pustaka.

Holmes, J. 2003. *Introduction to sociolinguistics. 2nd ed.* England: Pearson Education Limited.

Ikranegara, Kay. 1988. *Tata Bahasa Melayu Betawi*. Jakarta: Balai Pustaka.

Irsyam, Tri Wahyuning M. 2017. *Sejarah Depok 1950—1990.* Jakarta: Yayasan Pusaka Obor Indonesia.

Kridalaksana, Harimukti. 2008. Kamus Linguistik (edisi keempat). Jakarta: Gramedia Pustaka Utama.

Lauder, Multamia MT. 1979. "Bahasa di Daerah Bekasi: Sebuah Kajian Lokabasa". Skripsi. Jakarta: Fakultas Sastra Universitas Indonesia.

_____. 1993. *Pemetaan dan Distribusi Bahasa-bahasa di Tangerang.* Jakarta: Departemen Pendidikan dan Kebudayaan.

_____. 2002. "Reevaluasi Konsep Pemilah Bahasa dan Dialek untuk Bahasa Nusantara." Jurnal Makara Sosial Humaniora Vol 6. No. 1 Juni. Depok: Universitas Indonesia.

_____. 2007. *Sekilas Mengenai Pemetaan Bahasa*. Jakarta: Media Eka Sarana.

Malingkas, Yulianti. 1996. "Situasi Kebahasaan di Kelurahan Pancoran Mas, Depok." Skripsi Fakultas Sastra Universitas Indonesia.

Muhadjir. 1984. *Morfologi Dialek Jakarta*. Jakarta: Jambatan.

_____. 2001. Bahasa Betawi: Sejarah dan Perkembangannya. Jakarta: Yayasan Obor Indonesia.

Muhadjir, dkk. 1979. *Fungsi dan Kedudukan Dialek Jakarta.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

_____. 1986. *Peta Seni Budaya Betawi*. Jakarta: Dinas Kebudayaan DKI Jakarta.

Sugono, Dendy. 2008. Pusat Bahasa. *Bahasa dan Peta Bahasa di Indonesia*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.

Wahyu, Yohanes. 2010. "Variasi dan Distribusi Bahasa Betawi di Kota Depok." Skripsi Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia.

Language Contact and Language Change in the History of *Betawi Ora* Dialect

Mesiyarti Munir
Department of Linguistics, University of Indonesia

Abstract

This research analyzes Betawi Ora dialect in Jabodetabek in term of language contact and language history. This aim of this research is to view language contact and language change since the reasearch of Grinjs and today situation and how the history has made it to be. The mapping of Betawi Ora dialect tells that the big migration from the city to the urban of Jakarta made this dialect become shifting and near to extinct. Compared to the map of Grinjins research in 1970, the dialect Ora speakers were experienced a shifting in the location of the speakers. Native speakers who still occupy the Urban of Jakarta such as Kampung Rambutan, Jagakarasa, and Jatinegara have moved to outside Jakarta such as Bogor and Bekasi. The result of this research shows that Betawi Ora dialect area is shifted to the East and Southeast of the Jabodetabek. The language contact with the Jakarta immigrants from various languages in Jakarta become the reason why this dialect. Some of the typical vocabulary that this dialect possesses does not survive the next generation of dialects. Based on the using of dialect, the speakers of Betawi Ora there are three speakers which are the speakers of dialect that is still used Betawi Ora definitively, the speakers who still use the typical dialect vocabularies of Betawi Ora, and the speakers who are no longer using the typical dialect vocabularies. The dialect mapping and dialect using analysis show that the dialect Betawi Ora in Jabodetabek is not extinct but shifted due to factors outside of language

Keywords: *Betawi Ora, Dialect, language mapping, language contact, language using, dialectology, Jabodetabek,*

I. Introduction

1.1 The Background of the study

The Betawi Ora dialect comes from ethnic Javanese descent who settled in the Jakarta area after failing in war against the VOC soldiers (Muhadjir, 2000, p. 71). Soerjapranata's 1927 study in Tawangsih (1990, p.32) found that villagers Sudimara in Ciledug Tangerang uses the language Betawi Ora for using the word ora (derived from Javanese), while the Betawi people using the Malay language is the word no and *êngga*. According to Muhadjir (2000, p. 72), the Betawi dialect Tengahan and Betawi Pinggiran or the so-called with Betawi Ora having differences based on phonological features and vocabulary. The final vowel e in particular which in Indonesian is pronounced as a, in a suburban dialect pronounced with *a* 'or *ah*. So (s)*ayè* pronounced *saya* means me '. Based on use of vocabulary, subdialek Betawi Tengahan not found words derived from the Java language, while subdialek Betawi Pinggiran many found vocabulary derived from the Java language words like *bocah*, *lanang*, *kulon*, and *jeleh* not found in the Betawi dialect dialect.

Grijins research, 1991 found that the use of Betawi Ora dialect vocabulary is still commonly found in Jakarta City, such as Tanah Abang, Jatinegara, Jagakarsa, Grogol, Petamburan, Kampung Rambutan, and several other areas in Jakarta. The Betawi Ora dialect is used in everyday conversation and in art

performances such as Lenong, Wayang Kulit, and Mask Melayu. The art of Wayang Kulit was found in other arts. Wayang Kulit is displayed in Betawi Tengahan and Betawi Ora.

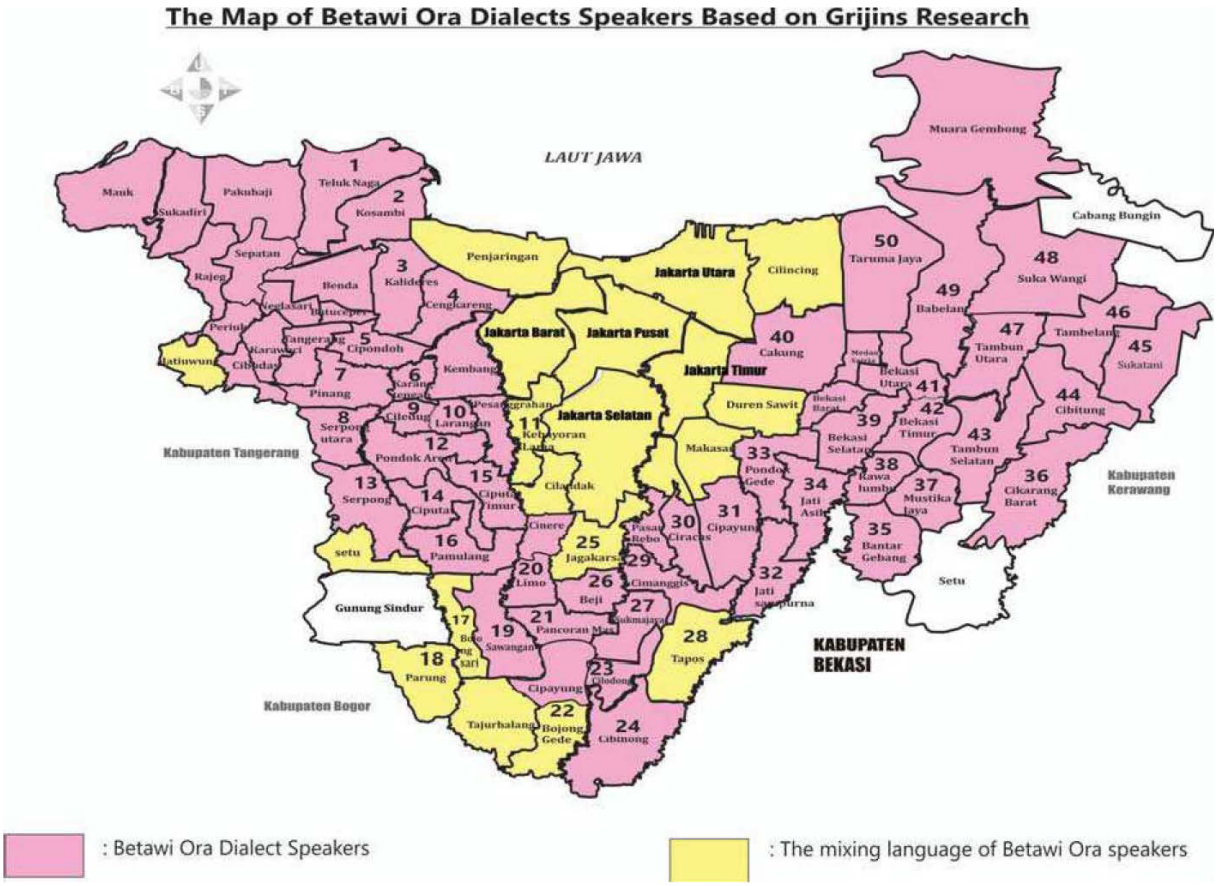


Figure 1
The Map of Betawi Ora Dialects Speakers Based on Grijins Research

Currently, the speakers of Betawi Ora dialect are concentrated in some areas that still have cultural figure of Betawi. According to some Betawi figures, speakers of the Betawi Ora dialect are in the vicinity of the area Depok distric, Tangerang Selatan subdistrics, Municipality of Tangerang, Kota Bekasi, Bekasi Regency, and a small part in Bogor Regency. Betawi dialect Ora on this area is still running because it is still supported by Betawi Ora art activities though very rarely done. Most speakers of the Betawi Ora dialect inhabits the periphery close to the City of Jakarta. However, the consequences the development of settlements with many centers average shopping and residential complex comes from the immigrants, the speakers of dialect Betawi Ora is getting shifted away from the city area Jakarta. This happens in almost at most in area of South Tangerang City, Tangerang City, and Bekasi city. In general, the location of Betawi dialect speakers Ora is in the occupied suburb area of Jakarta as a large area of Jabodetabek (Jakarta, Bogor, Depok, Tangerang, and Bekasi).

Betawi Ora dialect research in Jabodetabek area has never been done specifically and thoroughly to date. Dialectological research ever done only is sporadic and has not provided a comprehensive overview of the location of dialect speakers Betawi Ora at the moment. Thus, then the mapping again the Betawi Ora dialect needs to be done as part of the dialect geography research. Language mapping this is very important because the resulting map of this research can contribute to that real in education as material considerations for designing local content curricula in school. In addition, it can be used as a tool for language development and language preservation.

Efforts to preserve local languages can be done with language mapping so that location can be known and the condition of the dialect speakers. In addition, with language mapping can be obtained in documentation writing as an attempt to keep the language from being destroyed although it is no longer used by generation's successor. According to Tawangsih (1990, p.3), language mapping is necessary in the area multilingual because the language contact is not avoided. Thus, Jabodetabek area is an area with multilingual conditions very complex is necessary mapping a thorough language using the study dialectology. This language preservation effort appears as a result of worries about the extinction of many languages occurs in regional languages in Indonesia. This matter LIPI disclosed that based on SIL data (Lauder, 2004) there are 169 languages in Indonesia that are threatened extinct because it only has 500 speakers.

1. 2. The research problem

Betawi Ora dialect is currently on the verge of extinction compared to the research of Grinjs in . In this regard, the researcher will trace the existence of the Betawi Ora dialect is integrated in Jabodetabek area. Based on the background, the extinction of the Betawi dialect Ora has a poor language transmission process because the younger generation is reluctant to use it because considered not cool, plebeian, and uneducated such as the results of research Rahayu (1988) in the area Ciledug Tangerang, Tawangsih (1990) in the Tangerang area, and Revelation (2011) research only found the vocabulary "ora" is only in the westernmost part of Depok ie Bojongsari subdistrict. In addition, developments population and urban planning led these dialect speakers increasingly marginalized from the city of Jakarta. This research also will show map of usage area the current Betawi Ora dialect in today usage compared to the research of Grijins in 1991. Some areas that use the Betawi Ora dialect in the Girijins research have not been found anymore because their generation of speakers is gone and moved. This happens in some region of Betawi Ora dialect speakers in Tangerang as the Betawi Ora dialect speakers expressed almost extinct.

2. Research methods

2.1 Data Source

The scope of this research is 50 sub-districts spread in the DKI Jakarta, Bogor, Depok, Tangerang, and Bekasi (Jabodetabek) are still suspected has Betawi Ora dialect speakers. Because lack of information about the location of Betawi dialect speakers Ora, then the determination of observation points is based on literature study, information Betawi figures and residents local, and direct tracking in the field. Based on literature study found some research which is related to the Betawi Ora dialect, among others is Tawangsih (1990) research in Tangerang, Pratiwi's (1998) research in Cipayung District, and research Yohannes (2012) in Depok Municipality. Betawi Ora dialect speaker based on literature study found in 12 sub districts. More details can be seen in the following figure.

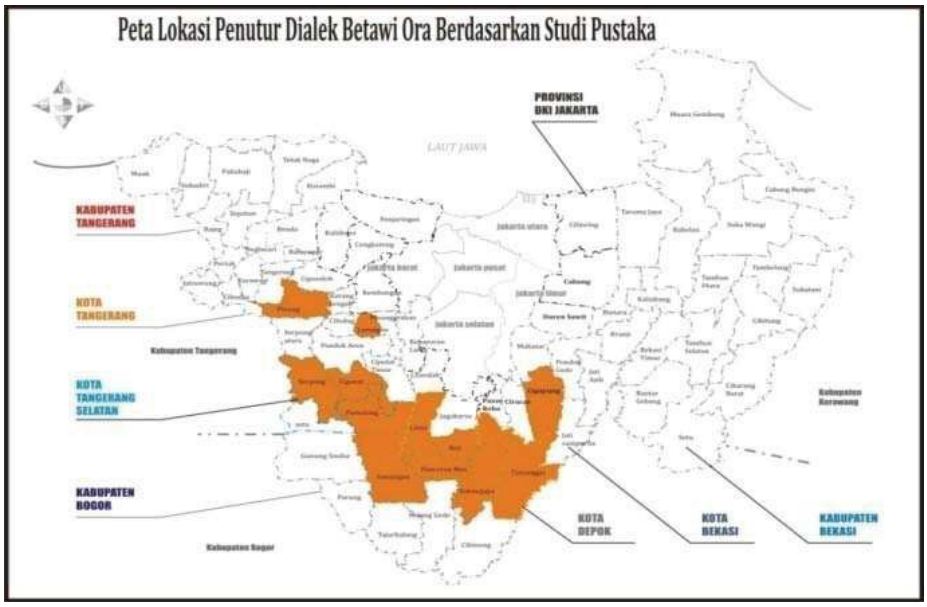


Figure 2
The Map of Betawi Ora Dialect Speakers Based on the Library Studies

The next step is direct tracking in the field. At this stage the information obtained that speakers of the Betawi Ora dialect spread in 40 districts in Jakarta area. The following map the location of Betawi Ora dialect speakers based on information from Betawi public figures and local residents.

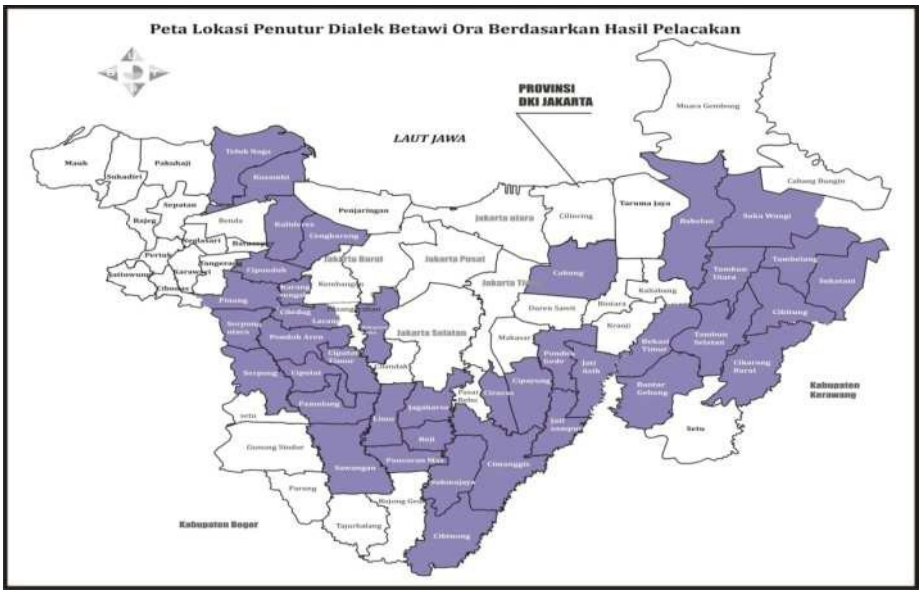


Figure 3
The Map of Betawi Ora Dialect Based on the Information from Betawi People and Locals

Based on literature study, information about Betawi public figures and local people, as well as direct tracking in the field, it can be determined that the observation point in this study amounted to 40 observation points. After the research was done, the observation point increased to 50 observation points from. The addition of 10 observation points are located in Depok, Bogor, and Bekasi are Bojong Sari,

Parung, Bojong Gede, Cilodong, Tapos, Mustika Jaya, Rawalumbu, South Bekasi, and Bekasi Utara. Here's a picture of the observation point map.

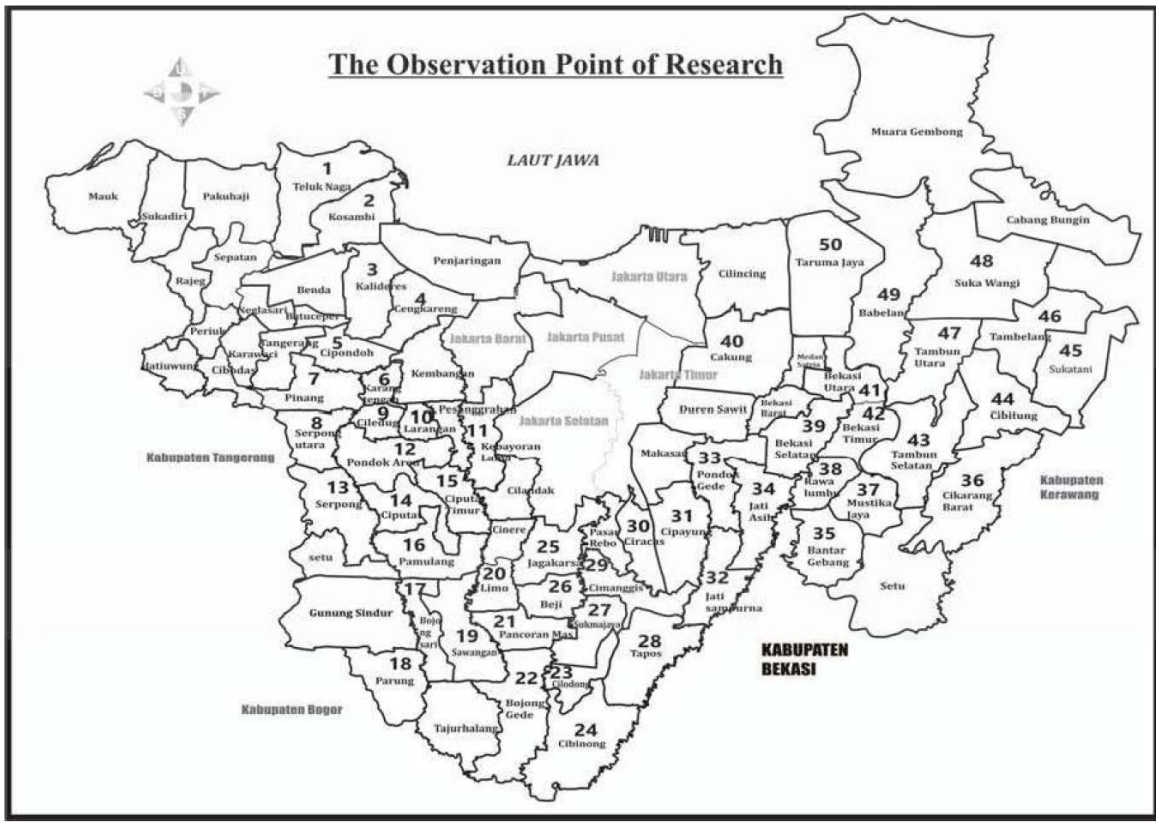


Figure 4
The Observation Point of the Research

The informants of this research are Betawi Ora dialect speakers in Jabodetabek area obtained based on the result of research, references of Ora Betawi culture experts, local people, and tracking. In this study selected informants aged 40-60 years with the consideration that the age of the informant still mastered the Betawi Ora dialect. In addition, the informants are speakers and native Betawi Ora residents of each sub-district with maximum high school education level so that the language used is still pure and that has a good memory and speech. Informants in this study are two informants for the point of observation, so that the number of informants as many as 100 speakers Betawi Ora dialect.

The form of questionnaire used to find the Betawi Ora dialect in this study is 200 basic vocabulary of Swadesh and basic cultural habitat sourced from kinship research standards and language mapping from the Ministry of National Education (2011). The questionnaire used in this study amounted to 375 vocabularies consisting of 200 Swadesh vocabularies and 175 basic cultural vocabularies of 52 vocabulary body parts, 25 vocabulary kinship systems and 98 motion and work vocabularies (attached) and 10 simple sentence questions (attached) to prove whether Betawi Ora dialect vocabularies is still used in everyday life. The vocabulary is considered a relatively more difficult vocabulary to change

3. Literature Review

In general, Chambers and Trudgill (1980, p.4) observed that dialect is a collection of variations languages separated by existing geographic factors because the language touch, but speakers can still mutually intelligible. From the second the statement, it can be concluded that dialect is a variation of interconnected languages and separated by geographical factors but still mutually exclusive understand.

The concept of mutual understanding (mutual intelligibility) has the basic principle that the spatial distance is proportional to the level of understanding. That is to say, a region of isolates has reciprocal understanding according to distance proximity to the deployment center. Closer with the center of the spread, the higher the level mutual understanding; the farther away from the center spread, the lower the level of understanding of lead behind it.

Muhadjir (2000, p. 72) states that dialect Betawi Tengahan and Betawi Ora have characteristics typical of a different kind of speech. First, Middle dialect with the final vowel feature e, whereas Betawi Ora vowel e is pronounced with a 'and ah. Become S) *ayê* pronounced as saya '; what ape pronounced. Second, the consonant h which is contained in the end on the Betawi Ora dialect is spoken the same as the language Indonesia, like blood, hard, split, and so forth. Whereas in the dialahan dialect pronounced as *darê*, *susê*, and *belê.* The third prominent speech feature is a voiced consonant utterance: *b*, d, and g. Consonant contained in words like *bedug*, *urug*, *maulud* spoken in the language Indonesia, which in Betawi Ora dialect is spoken in a voiced consonant. The fourth characteristic is the field prominent vocabulary like ora 'not' in the language Betawi Pinggiran. Other common vocabulary used in the Betawi Ora dialect comes from the vocabulary Javanese, such as *bocah*, *lanang*, *kulon* not found in Betawi dialect. This tracking research uses features typical of Betawi Ora dialect according to Muhadjir (2000).

Language variations by Chambers and Trudgill in his book The Language Variation and Change (2000, p. 9) is a study of observing languages the use of language in the social sphere and grouping languages according to distribution social. According to Chaer (2004, p.62) language variations caused by the existence of social interaction activities done by a very society or group diverse and because of its speakers inhomogeneous. In case of variation of this language there are two view. First, the variation is seen as a result the social diversity of the speakers of that language and the diversity of language functions. So variations of that language occurs as a result of social diversity and diversity of language functions. Second, the variations of that language is there to fulfil its function as a tool interaction in various community activities variety. Weinreich (1979, p.5) explains that touch language is considered by some Anthropologists as one of the aspects of cultural contact and interference language intervention in terms of cultural dissemination and acculturation. Language contact occurs not solely caused by a touch of language between the two region I village, but also determined by the presence the history of social relations between the population of society including economic, political and demographic factors. It occurs due to community interaction in a certain village with the adjacent village with him. Sometimes interactions that occur between the boundary society is an interaction which affect each other and become an interaction the positive (neighbouring territory) or even is a mutually incomprehensible interaction others and cannot influence each other

The dialectological study was conducted by Tawangsih on year 1979 entitled "Language in Bekasi Region a Lokabasa ". According to this study, there are the Javanese language that is in the middle of the village using the Malay language of Jakarta. The village Javanese language is allegedly there are speakers of Betawi dialect Ora. Other dialectological research is also done by Tawangsih in 1990 in the Tangerang area entitled 'Mapping and Distribution of Languages at Tangerang. In this study found that area-use Malay adjacent to Special Capital Region of Jakarta, regional-use language Sunda is directly adjacent to Bogor Regency; and area-use Javanese language is directly adjacent with North Serang District. In addition, too found some villages that use dialect Betawi Ora is only used by the younger generation. Betawi Ora dialect research conducted by Rahayu in 1988 entitled "Betawi Ora Dialect Mapping in Ciledug District ". This research aims to charting the Betawi Ora dialect in Ciledug Subdistrict, Tangerang. The results show that the vocabulary is found in many Ciledug areas that come from Javanese or Javanese vocabulary, Sundanese, and Balinese. The Betawi Ora dialect in this region is rare used by the younger generation because of embarrassment an inferiority.

4. Research Discussion

4. 1 the History of Betawi Ora Dialect

The Raid of Mataram (1628-1629) and Jan Pieterszoon's Policy Coen up to the time Independence Cultural repertoire search and language in Bekasi area Culture and Malay Language Betawi Dialek Bekasi is an exploration effort which is quite large. Excavation, can be drawn from the time the great kingdoms (Pajajaran, Tarumanegara) with elements Hindu-Buddhist culture. Furthermore, during the VOC and occupation The Netherlands, where the cultural process in Bekasi get influence from elements of other cultures. Bekasi original culture and language expected to experience the process cultural marginalization is not just due to the inclusion of cultural elements of the para entrants, but also internal factors, namely the desire of the community to develop and preserve its own culture is lacking got great attention.

During the Dutch East Indies, J.P. policy events The Coen forming a territory special form of Weltevreden level very Praja City area influence the development of culture and language in Jakarta and surrounding (including in Bekasi). However, the impact of formation Weltevreden by forming a zone to maintain security, a zone was created buffer especially in the east and west of Batavia (especially the region Bekasi and Tangerang), namely the area inland known as Ommelanden2 designation2 Ommelanden is the name of the hinterland and the area around Batavia, which owns its own history. For example when the Netherlands first occupied Batavia, the inhabitants were forced to leave the countryside enrich the treasures of culture and language in the Bekasi area has its own uniqueness compared with culture and language in Batavia (Jakarta tempo first) and separated from the region Batavia (Cribb, 1991, 31).

In the Ommelanden area in addition to community dominated natives also live in the community of the Sundanese and Javanese tribes. The Sundanese are predominant population at around the hills of Priangan, while the Javanese dominate along the coast in the direction Banten and Cirebon. Besides that, there are Chinese settlements mutually dispersed and always are in the assimilation process with local people. Mostly Their profession is the landlord and traders, so it's reasonable if then the economy is controlled by ethnic Chinese, and some Arabs.

Population migration and value culture and language in the region Bekasi surrounding, also mark with the appearance of Events Mataram I and II raids to Batavia (1628-1629) who gave contribution to development language and culture in Bekasi. Evidence history is shown in existence Historical Site "Saung Rangon in around Batavia, among them to Bekasi area and Tangerang with the intention of creating a zone a buffer that is not inhabited for the sake of the security of the Dutch East Indies government at the time J.P Coen (The Liang Gie, 1992). The presence of these other ethnicities shows mobility high population. Change is seen from the amount residents of Bekasi between 1927 until 1940. In 1927, the population of Bekasi is 162,000 people, consisting of 30 the population of Europe, 7500 residents Foreign Orient, and 154,470 inhabitants' indigenous people ("Bekasi", ENI, 1927: 237-238). Meanwhile, on in 1940 the number of residents in this area reaches 200,000 people, based on submitted report data accept position (Memoir van Overgave) Resident Batavia L.G.C.A. van der Hoek (Indonesia, 1980: 268).

When viewed from some data, it turns out cultural diversity it is more prevalent in the region Bekasi, (especially the cultural element Betawi Pinggiran or Betawi Ora). However, in its development precisely in line with growth development and migration of the population Bekasi City shifted to the area Bekasi Regency as well as elements the culture it embraces, then a therein lies the problem so cultural or bias occurs cultural identity of the people of Bekasi. In addition, the allocation Betawi cultural area of Jakarta (City) with a cultural area Bekasi (Betawi Ora) so there is sort of cultural arrogance at Jakarta area that the Center Betawi culture is DKI Jakarta are in the assimilation process with local people. Mostly Their profession is the landlord and traders, so it's reasonable if then the economy is controlled by ethnic Chinese, and some Arabs.

4. 2 Lexical Variations

Based on the etyma classification, look more lexeme that comes more than one etyma rather than lexeme that comes from more than one etyma. This matter indicates that lexical variation in Jabodetabek is

not shows a high variation. By percentage lexical variation in Swadesh vocabulary, variation sequence the largest to the smallest is one etyma (71 percent), two etyma (18 percent), three etyma (7percent), four etyma (2 percent) and last order occupied five etyma and six etyma both equal
1 percent. More details, can be seen in diagram as follows

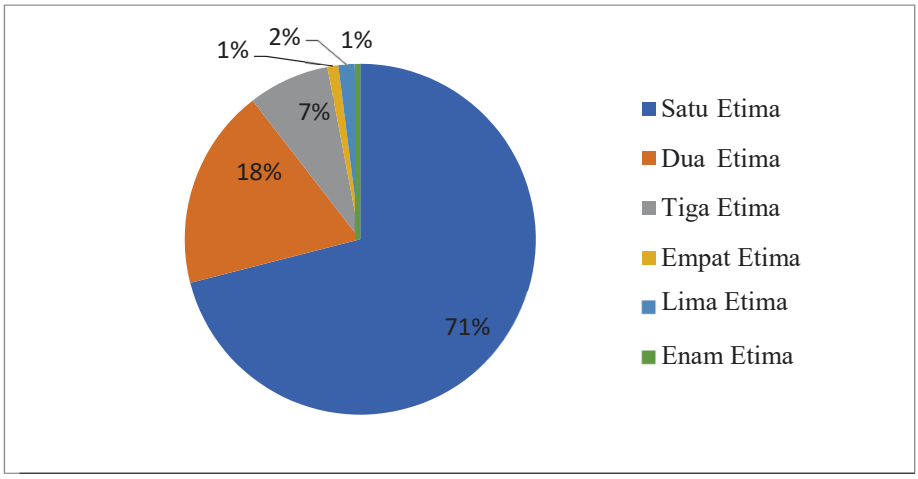


Diagram 1
The Percentage of Lexical Variation in Swadesh Vocabularies

The percentage of lexical variation in the mean field of the kinship system, the order of the most large to the smallest variations is one etyma (56 percent), two etyma (32 percent), three etyma (8 percent), and six etima (4 percent). More details, can be seen in the diagram as follows.

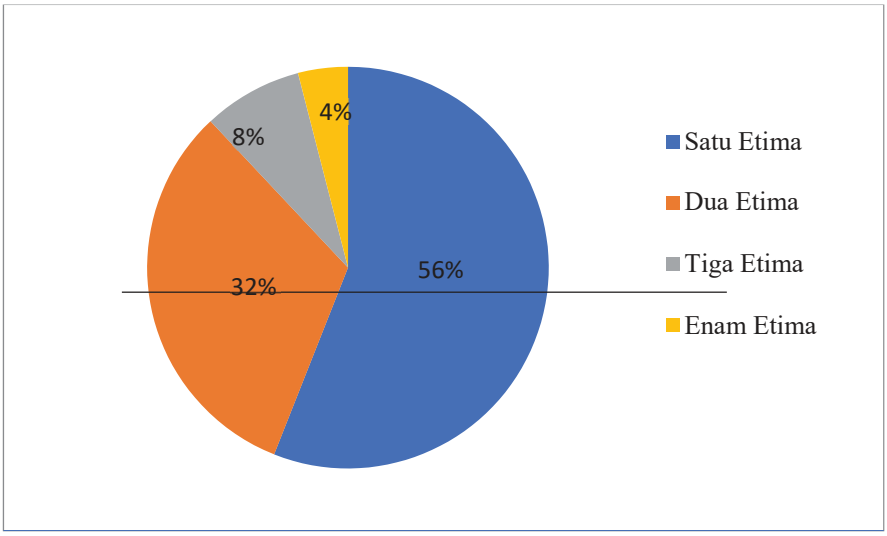


Diagram 2
The Percentage of Lexical Variation In The Meaning Of Kinship System

The merger of isogloss vocabulary files Swadesh, isogloss field of body meaning field, isogloss field of meaning of kinship system, and isoglos field of motion and work mean field shows result of isogloss file which progressively thickened in this isogloss joint file map. Very thick isoglosic bundles are seen in some TP 17 and TP 20; TP 28 and 29, TP 29; TP 35 and 37; TP 38 and 39; TP 41 and TP 42; TP 44 and TP 48. The thick lines in the combined isogloss file indicate a high lexical variation compared to other observation points. This lexical variation signifies a change in the use of Betawi Ora dialect vocabulary

used. The high lexical variation is caused by touching languages from other languages or dialects that are geographically adjacent. The vocabulary of ora and the vocabulary of the Betawi Ora dialect has begun to use off and is not used at the observation point on the line above. The vocabulary found is a vocabulary derived from Betawi and Indonesian dialects such as [kaga?], [ŋga?], [Pərəmpuan], [ana?], and [lakI?].

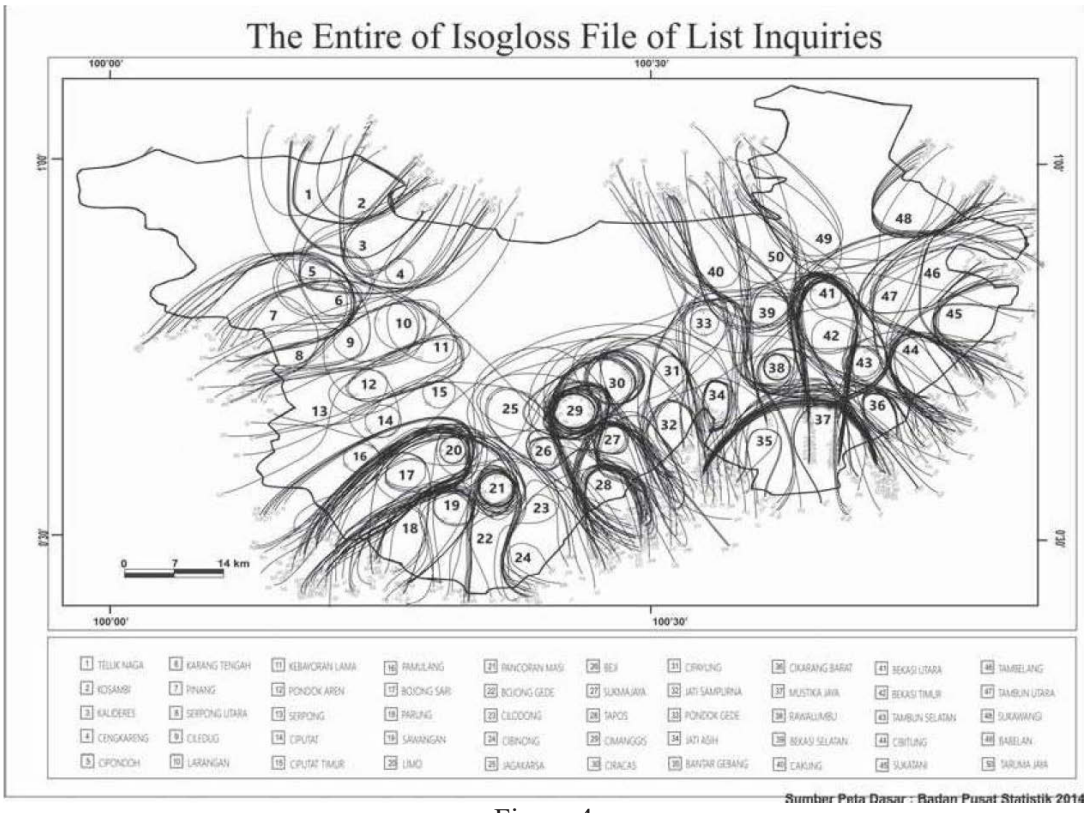
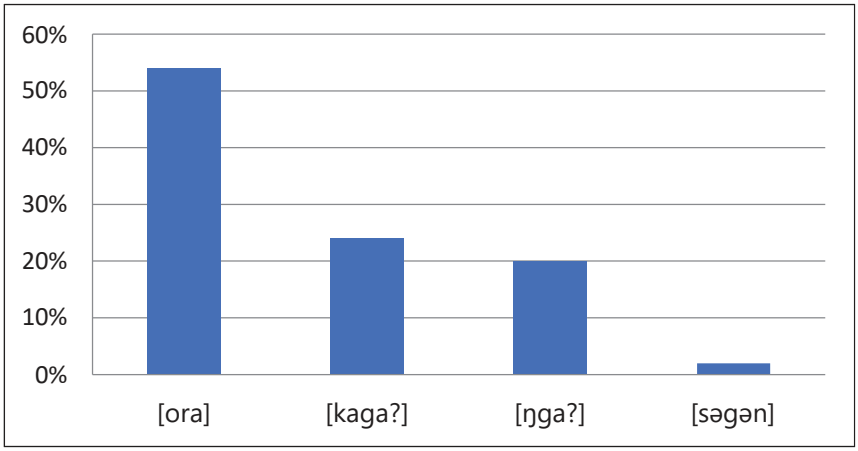


Figure 4
The Entire of Isogloss File of List Inquiries

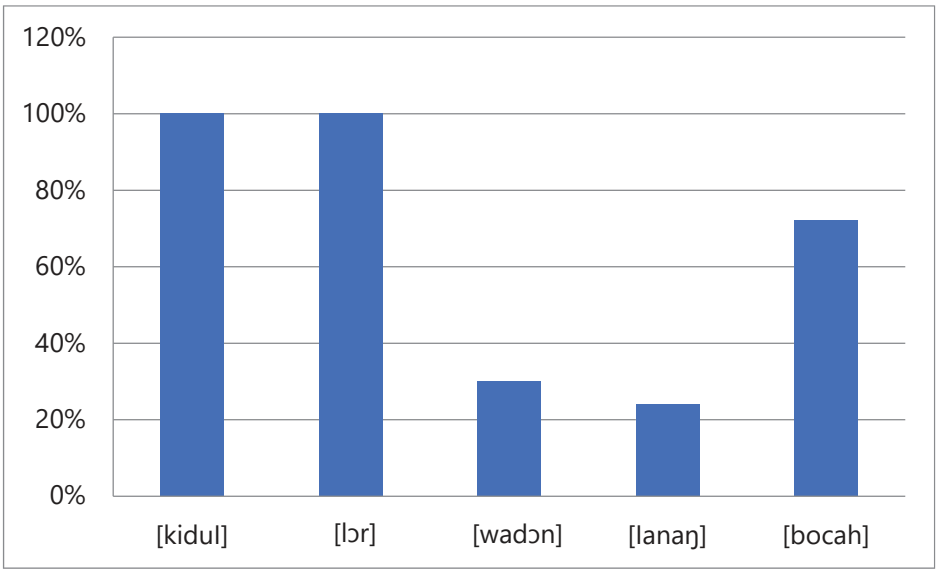
4. 3 **the Use of Betawi Ora Dialectic Vocabulary in Sentence**

According to the percentage of Betawi dialect Ora usage in the sentence, the order of use of the vocabulary the highest to the lowest is the vocabulary KIDUL and LOR as much as 100 percent), vocabulary MORE as much as (72 percent), ORA vocabulary (52 percent), WADON vocabulary (30 percent), vocabulary LANANG (24 percent), NGGAK vocabulary (24 percent), and KAGAK vocabulary (20 percent) out of a total of 50 dots. This indicates that the typical vocabulary dialect is generally still used by speakers Betawi Ora dialect. Point of observation is still a lot using a distinctive dialect vocabulary next to east and southeast of Jabodetabek region. For further details, can be seen in the diagram as follows.



Graph 1.
Percentage of Vocabulary Usage *Tidak* Betawi dialect

The use of vocabulary of TIDAK in Betawi Ora dialect there are four variations of vocabulary, ie [ora] as much as 54 percent, [kaga?] as much as 24 percent, [ŋga?] as much as 20 percent, and [səgən] as much as 2 percent. Most vocabulary users [kaga?] And [ŋga?] Claim to be embarrassed to use [ora] vocabulary because it is considered plebeian and incompatible with the present. The informants acknowledged that beforehand they still used the vocabulary [ora] when meeting with their peers. The younger generation of Betawi Ora dialect speakers generally do not use the [ora] vocabulary because it is not a cool language so they are embarrassed if their parents use this vocabulary in communicating. However, the researchers found that most of the younger generation of Bekasi and Kota Bekasi still use the vocabulary [ora] because they consider Betawi Ora dialect is Bekasi language. These findings indicate that the Betawi dialect has shifted the name, Bekasi is actually the Betawi Ora dialect. Speakers using the vocabulary [səgən] found in this study are speakers who recognize themselves as Betawi Tengahan instea d of Betawi Ora, but in direct interviews some Javanese vocabulary such as [wadon] is still used. When viewed from the percentage of the [ora] dialect Betawi dialect Ora by 54 percent it can be concluded that this vocabulary is still used by the speaker. Here is a graph of the percentage of vocabulary use typical of Betawi Ora dialect.



Graph 2.
Percentage of Typical Vocabularies Usage of Betawi Ora Dialect

Conclusion

Based on the results of the mapping that has been done with attention to the discussion of isogloss file and the use of the current dialect of Betawi Ora, some things that can be concluded are as follows. Compared to the Grijins research map on the use of the Betawi Ora dialect some of the areas currently using Dialect Betawi Ora is not found anymore, such as Kampung Rambutan, Jagakarsa, Pasar Rebo, and some areas urban Jakarta. Betawi Ora no longer same Based on the analysis of lexical variation, it appears that lexical variation in Betawi Ora dialect does not show high variation. These measurements are based onthe etyma classification indicating that the group derived from one etyma group showed the highest percentage. The highest percentage of lexical variation derived from a single etyma of 63 percent indicates that the typical Betawi Ora dialect vocabulary is still used in most of the allegedly spoken areas of Betawi Ora dialect**.** Based on the results of isogloss beam analysis shows the result of isogloss file which is getting thickened in this isogloss joint file map. Very thick isoglosses bundles are seen in some TP 17 and TP 20; TP 28 and 29, TP 29; TP 35 and 37; TP 38 and 39; TP 41 and TP 42; TP 44 and TP 48. The thick line at this point of observation indicates a greater variety of Betawi Ora dialect vocabulary compared to other regions. The high lexical variation is caused by touching languages from other languages or dialects that are geographically adjacent. The vocabulary of ora and the vocabulary of the Betawi Ora dialect has begun to wear off and is not used at the point of observation that lies on the thick line above. Regional use of Sundanese vocabulary which includes Based on the analysis of the use of Betawi Ora dialect vocabulary, it is known the following things.

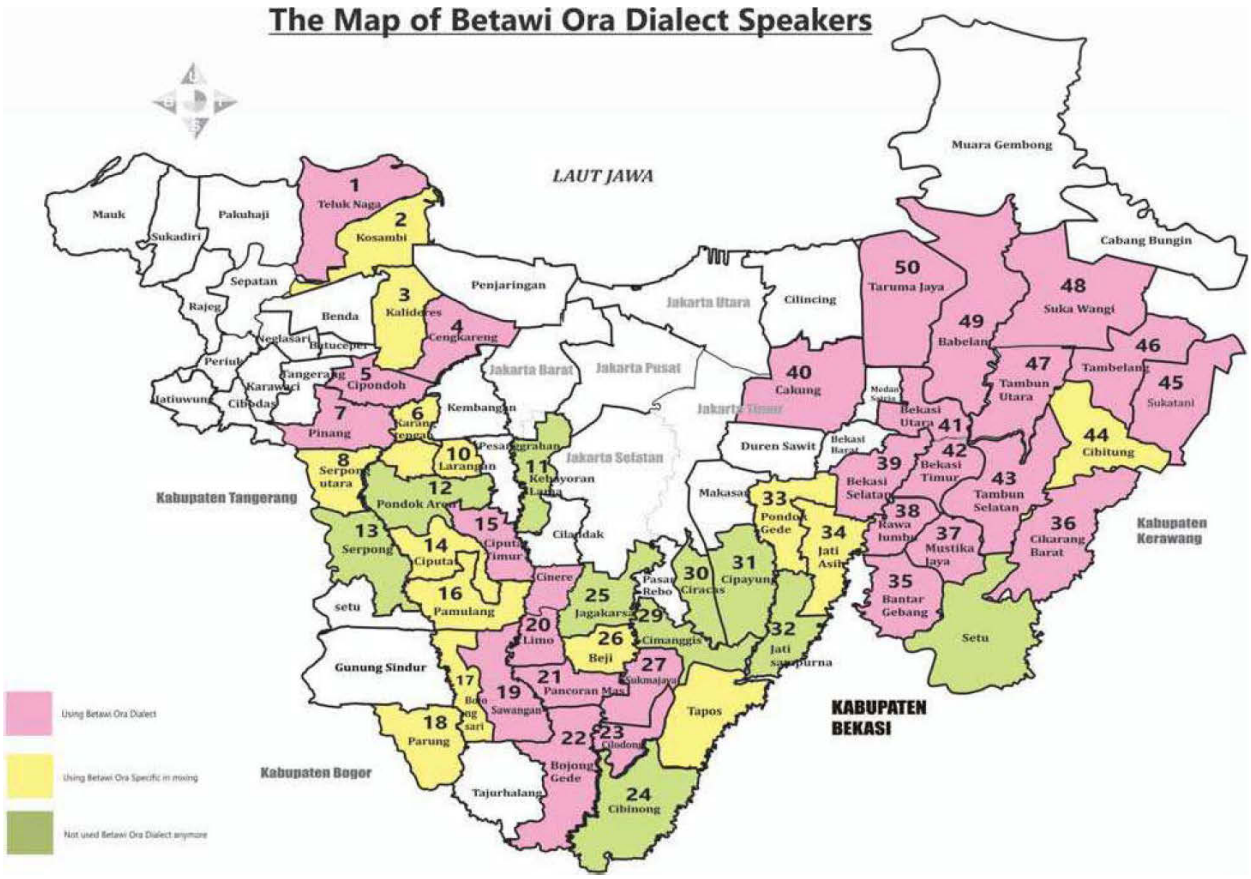


Figure 5
The Result of the Research

The speakers of the Betawi Ora dialect are divided into three categories, namely the dialect of Betawi Ora definitively, the dialect speakers who still use the ora vocabulary and the vocabulary of the Betawi Ora dialect, the dialect speakers who do not use the Ora vocabulary, but still use the Betawi Ora dialect vocabulary, and speakers of Ora dialect who have not used the Ora vocabulary and the vocabulary of the Betawi Ora dialect. a. The speaking region of Betawi ora dialect definitively, the Betawi Ora dialect speakers who still use ORA vocabulary and dialect's distinctive vocabulary. The area is Teluk Naga District (TP 1), Cengkareng Sub-district (TP 4), Cipondoh Sub-district (TP 5), Pinang Subdistrict (TP 7), East Ciputat Subdistrict (TP 15), Sawangan Sub-district (TP 19), Bojong Gede Sub- (TP 22), Cilodong District (TP 23), Sukmajaya District (TP 27), Bantar Gebang Sub-District (TP 35), West Cikarang Sub-District (TP 36), Mustika Jaya Sub-District (TP 37), Rawalumbu Sub- Bekasi Selatan Sub-district (TP 39), Cakung Sub-district (TP 40), North Bekasi Sub-district (TP 41), East Bekasi Sub-District (TP 42), Sukatani Sub-District (TP 45), Tambelang Sub-district (TP 46), North Tambun Sub-), Sukawangi District (TP 48), Babelan Sub-district (TP 49), and Taruma Jaya Sub-district (TP 50). b. The area of Betawi Ora dialect speakers who still use the distinctive dialect of Betawi Ora dialect. The speakers of the Betawi Ora dialect of this category still use a distinctive dialect vocabulary but have not used the word ora. These areas are Kosambi Sub-district (TP 2), Kalideres Sub-district (TP 3), Karang Tengah Sub-district (TP 6), North Serpong Sub-district (TP 8), District of Banangan (TP 10), Ciputat Sub-district (TP 14), Pamulang Sub- TP 16), District Bojongsari (TP 17), Parung Subdistrict (TP 18), Limo Subdistrict (TP 20), Beji Sub-district (26), Cibinong District (24), and Kecamatan Cibitung (44). c. The areas of Betawi Ora dialect speakers who have not used the oral vocabulary and dialect vocabulary are Kebayoran Lama (TP 11), Pancoran Mas (TP 21), Jagakarsa (TP 25), Tapos (TP 28), Cimanggis (TP) 29), Ciracas District (TP 30), Cipayung District (31), Jati Sampurna Subdistrict (32), Pondok Gede Subdistrict (33), and Jati Asih Sub-District (34).

The results of the isogloss file analysis and the analysis of the use of the Betawi Ora dialect vocabulary indicate the existence of parallels, i.e. the thick lines in the isogloss file show the use of oral vocabulary and the dialect vocabulary which is faded and not used anymore because of the influence of variations of vocabulary from dialect or other languages. The results of this study indicate that the Betawi Ora dialect is not extinct, but has shifted the East and Southeast region due to the development of existing infrastructure in the city of Jakarta, changes in urban spatial, the number of industrial areas and regions. In addition, the culture of *Bedol Desa* that existed in the speakers of Betawi Ora dialect also helped them move massively leaving the suburbs of Jakarta. Based on the map of the use of Betawi Ora dialect vocabulary, it is seen that the current dialect of Betawi Ora dialect increasingly heading to the east and southeast of Jabodetabek area. The findings of this study prove that the hypothesis of this study in accordance with the reality in the field. The Betawi Ora dialect is considered plebeian by its speaker and embarrassed to use it in daily conversation.

References

Aitchison, Jean. (2004). *Language Change*: Progress or Decay. Cambridge: Cambridge.

Ayatrohaedi. (2002). *Guidelines for Dialectological Research*. Jakarta: Language Center of the Ministry of National Education. Education and culture.

. (1983). *an Introduction to Dialectology*. Jakarta. Department Development and Language Development Center

Chambers, J. K., Peter Trudgill and Natalie Schilling-Estes (Eds). (2003). *the Handbook of Language Variation and Change*. Blackwell Publishing.

Chambers, J.K. and Peter Trudgill. (1980). *Dialectology. Cambridge*: Cambridge University Press.

Campbell, Lyle, and Mauricio J. Mixco. (2007). *A Glossary of Historical Linguistics*. Edinburgh: Edinburgh University Press.

Chudori, Mudjahid. (2011). *Java Grammar Banten*. Pustaka Sarana Cipta

Crystall, David. (2000). *Language Death. Cambridge*. Cambridge University Press.

Dalby, Andrew (2003). *Language in Danger, the Loss of Linguistics Diversity and The Threat to Our Future*. Columbia University. New York

. (2004). *Dictionary of Languages*. New York. Columbia University Press.

Dewi, Sandra. (1997). *Betawi Language Mapping in Kotif, Depok*. Essay. Faculty of Literature. Depok: University of Indonesia.

Faizah. (1987). *Dialect Jakarta in Kebon Jeruk. Depok*. Essay. Faculty of Literature. Depok: University of Indonesia.

Grijns.c.d (1991*). Study of Malay-Betawi Language*. Jakarta. Graffiti

Harrison, David. (2007). *When Language Die The Extinction of The Worlds Language and The Erosion of Human Knowledge*. Oxford University Press

Institute of Bases & Literature Sundanese. (1975). *General dictionary of Basa Sunda*. Bandung. Tarate Publisher Bandung.

Lauder, Multamia R.M.T. (1993). *Mapping and Distribution of Languages in Tangerang*. Jakarta: P3B.

Lauder, Multamia R.M.T. 2004. "Optimizing Indonesian Based Corpus Linguistics". Paper of Indonesian Language and Literature Scientific Meeting XXVI, Muhammadiyah University, Poerwokerto 4-5 October 2004.

. (2007). *Overview of Language Mapping*. Jakarta. Akbar Media Aksara.

Lewis, M.P (ed). (2009). *Etnologue: Language of The World*, 16th edn. Dallas, Text, SIL International

Ikranegara, Kay. (1998). *the Grammar of Betawi Malay*. Balai Pustaka. Jakarta.

Jones, Mari. C and Sigh, Ishtla. (2005). *Exploring Language Change*. New York. Routledge

Kawi, Djantera. (1991). *Banjar Dialects and Sub dialect*. Dissertation. Faculty of Letters. Depok. University of Indonesia.

Krauss, M. (2007). *Classification and Terminology for Degrees of Language Endangerment*. New York. Routledge.

Kusumawardhani, Queen Arum. (2012). Liyan in Betawi Architecture Case Study at Betawi Ora House in South Tangerang. Thesis. Department of Architecture FT UI. Depok. University of Indonesia.

Ministry of National Education. (2012). *Kinship Research and Language Mapping in Indonesia*.

Malmkjaer, Kirsten. (1995). *the Linguistic Encylopedia*. London and New York: Routledge

Moseley, Christopher (eds). (2010). *Atlas of the Worlds Language in Danger*. Unesco Publishing.

Mesthrie, Rajend, and Bhatt, M. Rakesh. (2012). *World Englishness, the Study of New Linguistic Varieties*. Cambride. Cambridge University Press

Muhadjir. (2000). Betawi. *History and Development*. Jakarta: Yayasan Obor Indonesia.

Muhadjir and Ayatrohaedi. (1979). *Function and Position of Dialect Jakarta*. Language Development and Language Development Center for Education and Culture. Jakarta.

Mukti, US. (1977). District Dialect Citeureup, Gunung Putri, Cileungsi. Thesis. Faculty of Languages and Letters. Jakarta: IKIP Jakarta

Muhadijr, Lauder, Multamia., & Rahmat, Ruciat. (1986). "Tracking the Traces of Java in Betawi". Paper at Periodic Scientific Meeting Research and Study Center of Nusantara Region: Javanology in Yogyakarta.

Mahsun. 1995. *Diacronical Dialectology: An Introduction*. Yogyakarta: Gadjah Mada University Press.

Unesco (2003*). Language Vitality and Endangerme*nt. Document Submitted to the International Expert Meeting on Unesco Program. Safe Guarding of Endangered Language, Paris 10-12 March 2003. http:www.unesco.org/culture/ich/doc/src/00120-EN.pdf (accessed on 28 November 2013).

Utomo, Sutrisno.Sastro (2007). *The Complete Dictionary of Javanese Language*. Kanisius Publisher

Pratiwi, Endang Hesti. (1996). Betawi Language in Cipayung DKI Jakarta: A Language Mapping and Language Analysis. Essay. Faculty of Literature. Depok: University of Indonesia.

Rahayu, Rr. Sri Yuniasih. (1988). *the Mapping of the Betawi-Ora Dialect in Ciledug District*. Essay. Faculty of Literature. Depok: University of Indonesia.

Robins, and Uhlenbeck (1991). *Endangered Language*. Oxford University Press.

The Similarities and Dissimilarities of Satun Malay Spoken in Malaysia and Thailand: A Geo-linguistic Analysis

Nor Hashimah Jalaluddin[a,]
Junaini Kasdan[b,]
Harishon Radzi[a]
[a]Linguistics Program, Universiti Kebangsaan Malaysia
[b]Institute of the Malay World and Civilization

Abstract

Langkawi Island in the northwest of Malaysia and Satun in the southern part of Thailand share many interesting facts and fictions. The two neighborhoods are closely related because they traditionally experienced similar historical, geographical, economic, linguistic and cultural backgrounds. The most important trait is that the communities speak the same subdialect of Malay dubbed as Satun Malay. Accordingly, the aim of this paper is to examine the phonology of Langkawi Satun Malay (LSM) and Thailand Satun Malay (TSM), particularly focusing on their similarity and dissimilarity features. Data are collected in the field from various subjects, namely young, adult, and old informants. The available data are broadly transcribed and the phonological features are systematically deduced based on the given regular patterns. The phonological features are then represented and mapped accordingly by employing the geo-linguistic analysis. The phonology of LSM and TSM can be divided into two categories, namely (i) Malay phonological features, and (ii) Thai phonological features. In the former, the phonological rules involved are commonly found in many Malay dialects, while in the latter, the rules are strictly associated to Thai grammar. For the purposes of GIS analysis, the representations merely focus on Thai phonological features that were imported into TSM and LSM, namely monosyllabic word formation, tones, aspiration and long vowel. It is apparent that the distributions of these features in both localities are dissimilar and not uniformed.

Keywords: similarities, dissimilarities, geo-linguistic, Satun Malay, phonology,

1 Introduction

Langkawi and Satun are neighboring districts which are situated and administered by two different nation states, namely Malaysia and Thailand respectively, but yet they share many things in common because they traditionally experienced similar historical, geographical, economic, linguistic and cultural backgrounds. Geographically, the two districts are separated by Andaman Sea, but they can easily be reached by ferry. Thus, the peoples are moveable and interconnected quite freely. Based on historical evidences, the two regions were once subsisted as one kingdom. However, after Anglo-Siamese treaty was officiated in 1909, they were separated and administered by two different sovereign countries. Langkawi belongs to Malaysia, while Satun goes under Thai administration. As mentioned, although the lands are physically divided, the communities are mutually assimilated as they shared the same cultural and linguistic features which accordingly tied the peoples in harmony until today (Kamaruddin 1999). In their daily communication, the peoples of Langkawi and Satun generally speak the same dialect of Malay commonly dubbed as Satun Malay. It must be noted that Malay is a minority language in Thailand, and it is spoken mostly by Moslem Thai who live in Yala, Pattani, Narathiwat, Songkla and Satun (Luangthoangkham 2008, Phaithoon 2005). In his preliminary study, Worawit (1999) generally

points out that Thailand Satun Malay (TSM) is quite similar to Langkawi Satun Malay (LSM). In attempting to suffice his previous work, the current study aims to examine and describe the phonological aspects of TSM and LSM, primarily focusing on their phonological similarities and dissimilarities respectively.

2 Previous Researches on LSM and TSM

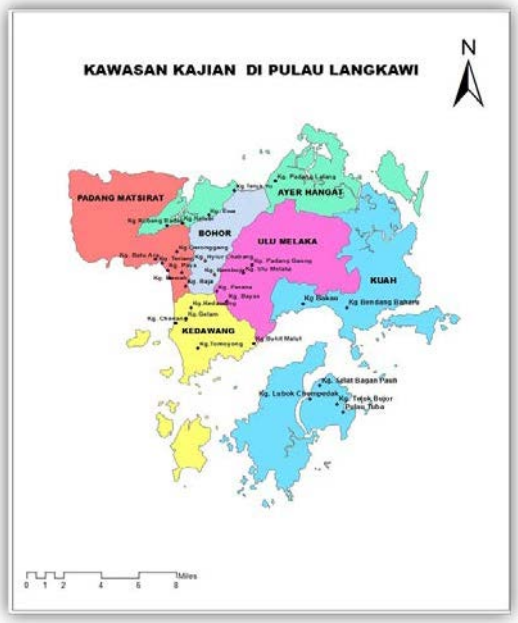
As reported in the litrature, there are a few studies has been done on LSM (Nor Hashimah et. al. 2017, Hayati et. al. 2016) and TSM (Kamaruddin 1999). As noted, it has been observed that phonologically there are similarity and dissimilarity fatures betwen these two subdialects. In addition, there are also some elements of Thai grammar that had been imported into the grammar of TSM and LSM through loanwords. In his preliminary study on Malay dialects spoken in Thailand, Amon (1986) has identified nine different groups of Malay dialects, and in Satun itself there are four subdialects, namely, Ban Khuan, Ce Bilang, Tammalang and Urok Lawoi in Pulau Adang. In another study, Gandour (1979) looks into how tones are assigned to Thai loanwords. Gandour claims that tone assignments are heavily dependent on word final syllables. Malay language on the other hand is a non tonal language. However, tones can still be heard in both Satun subdialects. Court 1975 (in Gandour 1979) reported that the high rising falling tone is assigned to syllables ending in a sonorant segment that occurs in phrase final position, and the high level tone is assigned elsewhere. For TSM the high falling tone is compatible with Malay intonation. Apart from linguistics, Norizah & Nor Hisham (2010) and Hayati et.al (2016) have look into sociolinguistic aspects of TSM. It is observed that TSM has no economic value in Thailand. Thai education policy has marginalised the role of TSM and its usage is confined only to family members, friends and religious domain. TSM is fluently spoken by adults of 50 years old and above, and on the contrary it has become a second language to the younger generation.

Thus far there is no linguistic study that focuses on the distribution of LSM and TSM in both loacalities. Accordingly, this study proposes that a geo-linguistic analysis which incorporates both linguistic and non-linguistic features is able to depict dialect distribution in a systematic, precise and accurate map produced by GIS (Onishi 2010, Teerarojanarat and Tingsabadh 2011, Nor Hashimah 2015, Nor Hashimah et. al. 2016, Nor Hashimah et. al. 2017, Siti Noraini 2018, Zaharani et. al. 2018). The non-linguistic attributes such as topography, migration, history and socio-culture provide extra informations that can offer a better analysis about dialect distribution in one particular area.

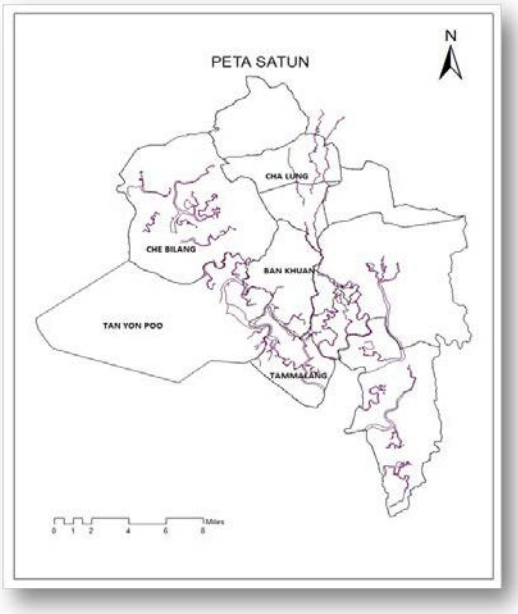
3 Methodology

The present study is conducted in accordance with the standard procedures established in a research metodology. Primary and authentic data are collected by the researchers in the field by various methods, such as observation, questionnaires, interviews and voice recordings. This procedure is essential and appropriate as it is closely related to data analysis and elucidation of final results (Nor Hashimah Jalaluddin et. al. 2017). The fieldworks were carried out in Langkawi Island and Satun. In the former, there are six villages involved, namely Kuah, Padang Mat Sirat, Kedawang, Bohor, Air Hangat and Ulu Melaka (see map 1). The informants were native speakers of Langkawi together with Thai respondents who have been resided in Langkawi for a long time. There are totally 120 informants, and they are divided into three groups, namely (i) young (age between 15-25 years), (ii) adult (age between 26-49 years), and (iii) old generation (50 years and above). The similar procedures and criteria are implemented in Satun district which involved three villages, that is (i) Che Bilang, (ii) Tammalang and (iii) Ban Khuan (see map 2). Che Bilang is located in a coastal area which is separated from Langkawi by Andaman Sea. Tammalang is bordered with Kuala Perlis in the south, and Ban Khuan is situated in the interior part which received a lot of Thai influence. As mentioned, in Satun only the adults and old informants can speak TSM fluently.

The available data are broadly transcibed and their phonological features are systematically deduced and categorized based on the regular patterns surface in the outputs. These phonological informations are subsequently imported to the GIS software. The software then analytically produces precise and scientific maps depicting the distribution of each linguistic item under study in both Langkawi and Satun.



Map1: Langkawi Island



Map 2: Satun

4. Discussion and Findings

Based on the regular phonological patterns, the phonology of LSM and TSM can be divided into two categories, namely (i) Malay phonological features, and (ii) Thai phonological features. In the former, the phonological processes or rules involved are commonly found in many Malay dialects, while in the latter, the rules are strictly associated to Thai grammar. It is apparent that in these two categories, both LSM and TSM display some similarities as well as dissimilarities respectively. In the following section, the discussion is divided into two parts:

i) Malay phonological features
ii) Thai phonological features

4.1 Malay Phonological Features

As mentioned above, Malay phonological features referred to phonological rules that are visibly active in many Malay dialects. There are four regular rules that have been identified, namely (i) lateral deletion rule, (ii) plosive substitution rule, (iii) fricative formation rule, and (iv) vowel rounding rule.

4.1.1 Lateral Deletion Rule

It is common in many Malay dialects that the lateral consonant /l/ cannot occur in word final position, and it is resolved by segmental deletion (i.e. /bɑntɑl/ ‘pillow’ becomes [bɑntɑ]). This rule is visibly active in both LSM and TSM. However, in the former, there is an additional rule that changes the vowel /ɑ/ to /ɛ/, and the output surfaces as [bɑntɛ].

Table 1: Lateral deletion word finally

lexicon	Lateral deletion	LSM/TSM
/bantal/ 'pillow'	[banta] [bantɛ]	LSM , TSM TSM

4.1.2 Plosive Substitution Rule

Here is another instance where sonorant consonants are not permitted in word final position. Unlike the lateral /l/, the prohibition of nasal segments in this environment is resolved by plosive substitution rule where the final nasal is replaced by a homogenous plosive segment (i.e. /ajam/ 'chicken' and /ikan/ 'fish' become [ajap] and [ikat] respectively). This rule is visibly active only in TSM, particularly in Che Bilang and Ban Khuan. The rule never applies in LSM and the nasal consonants are retained in the lexicons (i.e. [ajam] and [ikan]). Surprisingly, the rule also failed to operate in Tammalang, another village in TSM. It must be mentioned that even though there are variants, the lexicons are mutually comprehensible among the speakers in both subdialects. Other examples are shown in table 2 below.

Table 2: Substitution of nasals by homogenous plosives

Lexicon	Nasal to plosive	LSM/TSM
/ajam/ 'chicken'	[ajap] [ajam]	TSM LSM
/tadʒam/ 'sharp'	[tadʒap] [tadʒam]	TSM LSM
/ikan/ 'fish'	[ikat] [ikan]	TSM LSM
/hutan/ 'forest'	[utat] [utan]	TSM LSM
/makan/ 'eat'	[makat] [makan]	TSM LSM

4.1.3 Fricative Formation Rule

Third regular rule involved the derivation of fricative sounds [ɣ], [ʁ] and [ʕ] from the trill consonant /r/, and the process is dubbed as fricative formation rule which is governed by the occurrence of /r/ in the syllable positions. The fricatives [ɣ] and [ʁ] occur in the onset, while [ʕ] occurs in the coda. The distinction between TSM and LSM is that the former utilizes velar fricative [ɣ] and the later employs uvular fricative [ʁ]. They are similar in a situation which involved the occurrence of pharyngeal fricative [ʕ] in the coda, and this is a salient feature of Malay dialects spoken in northern part of West Malaysia. Relevant examples of lexicons are displayed in table 3.

Table 3: Transformation of trill to fricatives

lexicon	/r/ to [ɣ] and [ʁ]	TSM/LSM
/rambutan/ 'a kind of fruit'	[ɣambot] [ʁambot]	TSM LSM
/kari/ 'curry'	[kaɣi] [kaʁi]	TSM LSM

4.1.4 Vowel Rounding Rule

All the rules we discussed thus far only involved consonantal segments. What follows is a rule that modifies a vowel sound. Both subdialects undergo the same phonological rule called vowel rounding which transforms the low vowel /ɑ/ into mid low back vowel [ɔ] word finally. Relevant examples are exemplified in table 4 below.

Table 4: Word final /ɑ/ becomes [ɔ]

Lexicon	[ɔ] word finally	TSM/LSM
/pɑhɑ/ 'thigh'	[pɑhɔ]	TSM LSM
/sukɑ/ 'happy'	[sukɔ]	TSM LSM
/kitɑ/ 'we'	[kitɔ]	TSM LSM

4.2 Thai Phonological Features

As mentioned earlier, TSM is a marginal dialect spoken by minority Muslim communities in Thailand. Hence, the dialect is easily influenced by a predominantly Thai language. Inevitably, some Thai grammatical rules are imported into TSM via loanwords and subsequently disseminated to LSM. Based on the available data, four visibly active rules have been identified, namely (i) monosyllabic word formation, (ii) tone assignment, (iii) aspiration, and (iv) long vowel. Some of these rules are shared features, and others mark the distinction between LSM and TSM.

4.2.1 Monosyllabic Word Formation

Generally Malay words are disyllabic and Thai words on the other hand are monosyllabic. When Malay disyllabic words in TSM and LSM are transformed into monosyllabic words, it is believed that the transformation is triggered by Thai phonology. Monosyllabic word can be derived either by syllable reduction or resyllabification.

Table 5: Monosyllabic word formation

Lexicon	monosyllabic	TSM/LSM
/ɲiur/ 'coconut'	[ɲɔ],	TSM
/nenek/ 'grandma'	[nɛʔ] [toʔ]	LSM, TSM LSM, TSM
/dɑun/ 'leaf'	[do^{w}n], [do^{w}n]	LSM, TSM
/ləpat/ 'sweet delicacy'	[p^{h}at]	LSM, TSM
/ɑbɑŋ/ 'brother'	[bɑŋ] [bɑ]	LSM, TSM TSM

4.2.2 Tone Assignment

Another salient feature that characterizes the distinction between Thai and Malay is that the former is a tonal language and the latter is non-tonal. Gandour (1979) claims that there are five different tones in Thai, namely the mid falling, low falling, high falling, high raising and low rising. These tones are phonemic in Thai. According to Court (in Gandour 1979), TSM assigns the high falling tone / ˆ/ in the presence of sonorant sounds in a word final position. The high rising tone / ˊ / is used elsewhere. Interestingly, based on the available data, two other types of tones have been identified, namely low rising tone / ˇ/ and falling tone / ˋ /. Table 6 demonstrates some examples of Malay lexicons in LSM and TSM that have undergone tone assignment.

Table 6: Tone assignment

Lexicon	Lexicons with tones	TSM/LSM
/aɪr/ ‘water’	[aʲjâ], [aʲjǎʕ]	LSM, TSM
/ular/ snake’	[uljǎʔ]	LSM, TSM
/ajam/ ‘chicken’	[ajǎp]	TSM
/makan/ ‘eat’	[makàt]	TSM
/hutan/ ‘forest’	[utàt]	TSM
/durian/ ‘durian’	[dəjǎn], [dəjàt]	LSM, TSM
/ikan/ ‘fish’	[ikàt]	TSM
/tadʒam/ ‘sharp’	[tadʒǎp]	TSM
/ləmbut/ ‘soft’	[ləmuìt]	LSM, TSM

4.2.3 Aspiration

As established in the literature, aspiration is not distinctive in Malay phonology. Thus, aspirated consonants never surface in native lexicons. In Thai on the other hand, aspirated and unaspirated consonants are contrastive, and therefore aspiration is phonologically significant in the language. In the contexts of LSM and TSM, aspiration only occurs in the latter, and this linguistic feature is imported into the subdialect through Thai loanwords, as exemplified in Table 7 below. In sum, aspiration is a salient feature that distinguishes the two subdialects of Satun Malay.

Table 7: Loanwords with aspirated consonants

Lexicon	Aspirated Consonant	TSM/LSM
/kʰaj/ ‘chicken’	[kʰaj]	TSM
/kʰom/ ‘sharp’	[kʰom]	TSM
/laŋkʰa/ ‘hut’	[laŋkʰa]	TSM
/kʰǒŋ/ ‘objects’	[kʰǒŋ]	TSM
/pʰuk ʁâw/ ‘we’	[pʰuk ʁâw]	TSM
/hǔa kʰaù/ ‘knee’	[hǔa kʰaù]	TSM
/kʰa/ ‘thigh’	[kʰa]	TSM

4.2.4 Long Vowel

Besides aspiration, vowel length is also not contrastive in Malay phonology. Thus, the distinction between short and long vowels never occurs in native lexicons. Once again long vowels are imported into TSM only through Thai loanwords. Similarly to aspiration, long vowel is another phonological feature that differentiates TSM from LSM.

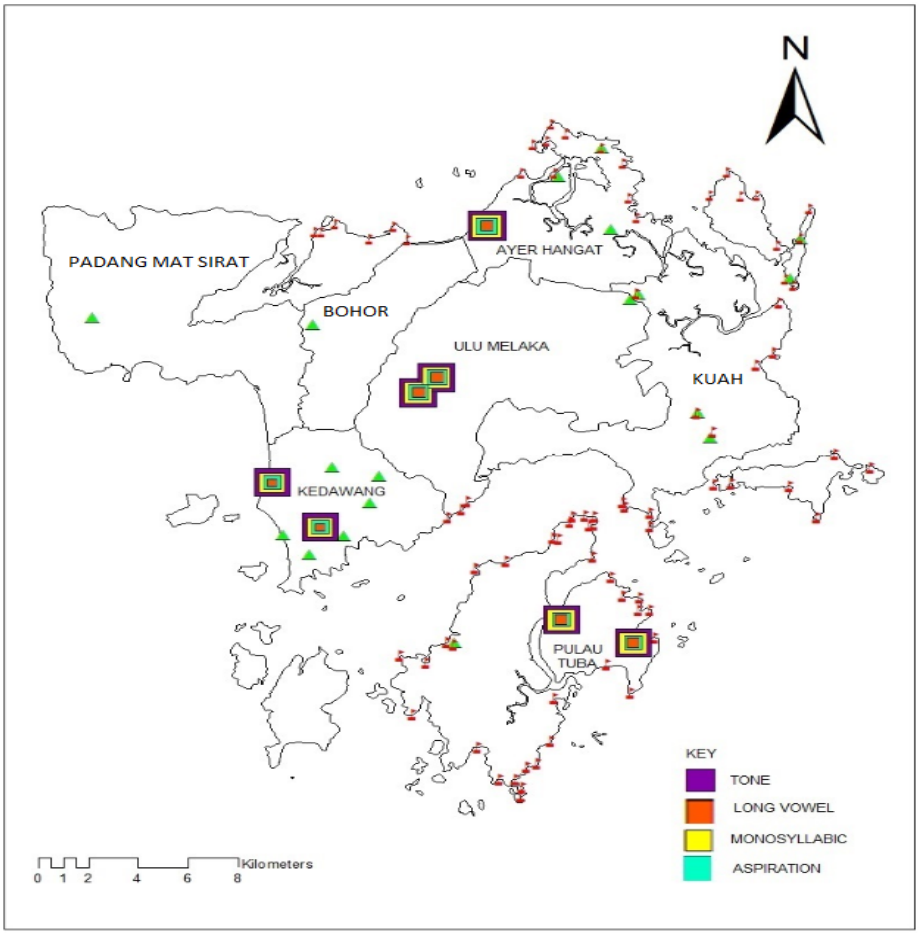
Table 8: Loanword with long vowel

Lexicon	Long vowel	TSM/LSM
/wɔː/ 'son'	[wɔː]	TSM
/pà:/ 'forest'	[pà:]	TSM
/mɑj:/ 'mother'	[mɑj:]	TSM
/luː mɑŋkut/ 'mangoesteen'	[luː mɑŋkut]	TSM
/luːp/ 'ironing'	[luːp], [tuː]	TSM
/luːʔ nun/ 'jackfruit'	[luːʔ nun]	TSM
/tɑː jɑ/ 'grandma'	[tɑː jɑ],	TSM
/tɔː/ 'a kind of vegetable'	[tɔː]	TSM

4.3 A Geo-linguistic Analysis

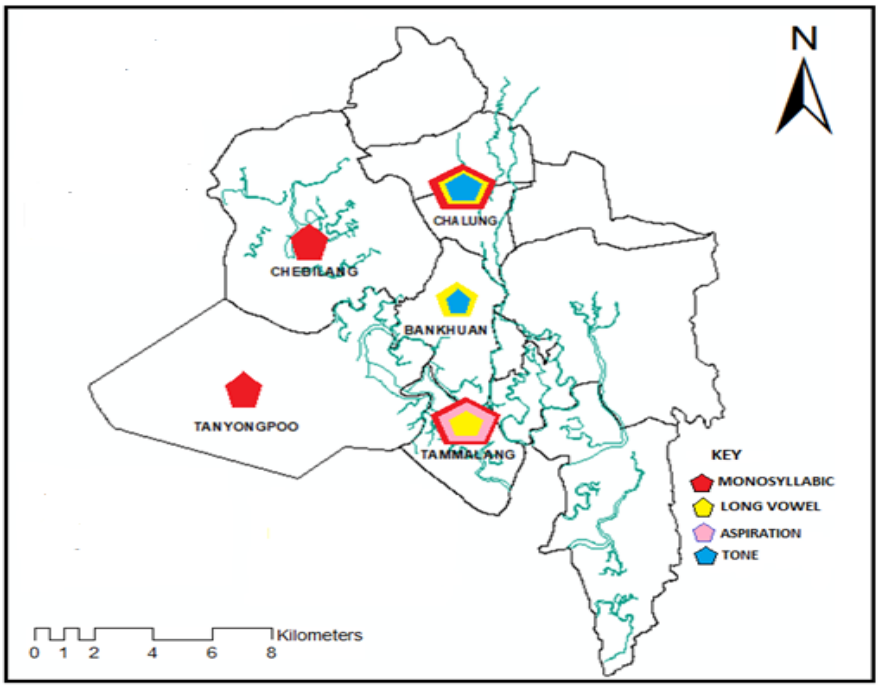
As shown above, there are phonological features that are shared by LSM and TSM, and at the same time there are features that exclusively occur in a particular subdialect. This implies that the dissemination of Satun Malay phonological features is not uniformed in LSM and TSM, and the distribution of these features will be captured and mapped by employing the GIS analysis. For the purposes of this paper, we merely focus on the Thai phonological features that were imported in the two subdialects, namely monosyllabic word formation, tones, aspiration and long vowel. The distribution of these features in LSM and TSM is depicted in map 3 and 4 respectively.

As can be seen in map 3, the feature of phonological behaviors such as monosyllabic aspiration, tones and long vowel are confined to certain areas only that is the central part of Langkawi. Here is the dominant residence of LSM speakers. Nor Hashimah et. al (2017) emphasizes that the topography of Langkawi acts as a hindrance factor in determining the dissemination of these features to Padang Mat Sirat, Bohor and Kuah. All these villages are known as the original residence of the locals.



Map 3: Distribution of Thai phonological features in LSM

Likewise, the distribution of Thai phonological features in TSM is also not uniformed. As shown in map 4, there are areas which only have one feature, while the others have more than one. For instance, in Che Bilang and Tanyongpoo (sub-village of Che Bilang) only the monosyllabic feature is utilized, while the other features such as tones, aspiration and long vowel are totally neglected. In stark contrast, Chalung, Tammalang and Ban Khuan utilize more Thai features which are imported through loanwords. However, it must be mentioned that the imported features retained in each village are still not the same. The distinction between Chalung and Ban Khuan is that apart from sharing the same phonological features of tone and long vowel, the former has an additional monosyllabic feature as compared to the latter. Tammalang which is situated in the southern part of Satun also has three phonological features. The salient characteristic of Tammalang is that it is the only region in Satun that has the phonological feature of aspiration besides monosyllabic and long vowel.



Map 4: Distribution of variants in Satun

5. Conclusion

The dissemination of Satun Malay phonological features in Langkawi and Satun is best understood if it is represented in a map where non-linguistic factor, such as topography can also be taken into account in explaining the distributional pattern. Non-linguistic factors such as migration, history and socio-cultural elements in both localities also play important roles in explaining the phenomena. In sum, linguistic and non-linguistic elements are equally essential in linguistic study, and they are incorporated under GIS analysis where the findings are more explicit, systematic and scientific.

6. Acknowledgements

Our thanks go to Universiti Kebangsaan Malaysia for granting research funds for this project (TD-2015-004 – The Sustainability of Malay Civilization in Langkawi and Satun)

References

Asmah, H. O. 1985. *Susur Galur Bahasa Melayu.* Dewan Bahasa dan Pustaka, Kuala Lumpur.

Collins, J.T. 1996. *Khazanah dialek Melayu*. Penerbit Universiti Kebangsaan Malaysi, Bangi

Gandour, J.1997. Tonal Rules for English Loanwords in Thai. *Studies in Tai and Mon Khmer Phonetics and Phonology in Honour of Eugne J.A.Henderson* ed. by T.L Tongkhum, V. Panupong, P.Kullavanijaya & M.R.K. Tingsabadh. Bangkok, Chulalongkorn University Press.

Ismail, D . 1997. *Glosari Dialek Kedah.* Kuala Lumpur: Dewan Bahasa dan Pustaka.

Kamaruddin, E. 1999. *Perbandingan Fonologi Dialek Melayu Satun dengan Dialek Melayu Perlis*. Tesis Sarjana. Universiti Sains Malaysia, Penang

Luangthoangkhum. T. 2008. The Position of Non-Thai languages in Thailand. Dlm Hock Guan Lee, Leo Suryadinata (pnyt). *Language, Nation and Development in Southeast Asia.* Pp. 181-191. Singapore: LSEAS Publishing. [acessed on 12th June 2015, from http://www.google.combooks]

Hayati, L, Nor, H.J, Harishon, R , Ruslan, U . 2016. Proceeding The Tenth International Malaysian Studies Conference (MSC10): *Globalization and Regionalism: Malaysian in ASEAN*. Penerbit Universiti Kebangsaan Malaysia, Bangi.

Nor, H.J, Maizatul, F.H, Harishon, R & Junaini, K. 2017. Penyebaran Dialek Melayu di Langkawi. Analisis Geolinguistik. *GEMA Online® Journal of Language Studies.* 17 (4): 159 - 178.

Nor, H.J.2015. Penyebaran Dialek Patani di Perak: Analisis Geolinguistik. *Jurnal Antarabangsa Dunia Melayu* Jilid 8 (2): 310-330.

Nor, H.J, Zaharani. A, Harishon. R , Norlisafina. S. 2016. Varian Kata Ganti Nama Dialek di Pesisir Perak: Analisis Geographical Information System (GIS). *GEMA Online® Journal of Language Studies* Volume 16(1): 109-123.

Norizah, A & Nor, H.O. 2010. Pemeliharaan Bahasa Ibunda dalam Kalangan Masyarakat Islam di Satun, Thailand. [acessed on 12th Jan. 2018, from http://repository.um.edu.my/id/eprint/83097]

Onishi, T. 2010. Analyzing dialectological distributions of Japanese. *Dialectologia, Special issue*, I: 123-135.

Phaitoon,M.C. 2005. Perubahan autosegmen dalam dialek Patani: Fenomenatransmorfofonologisasi. Persidangan Antarabangsa Bahasa , Sastera dan Budaya. Fakulti Bahasa dan Komunikasi. Universiti Putra Malaysia. 24-25 Januari.

Ruslan. U. 2011. *Keistimewaan Dialek Melayu Patani*. Bangi: Penerbit Universiti Kebangsaan Malaysia.

Siti. N.H . 2018. Penyebaran Leksikal dan Bunyi dalam Dialek Melayu di Perak: Analisis Geolinguistik. *PhD Tesis* Program Linguistik. Bangi: UKM

Teeraroanarat, S dan Tigsabadh, K. 2011. A GIS-Based Approach For Dialect Boundary Studies. *Jurnal Dialectologia* of Universitat de Barcelona 6 : 55-75

Worawit, B. (1999). Dasar Kerajaan dan Kesannya Terhadap Bahasa Melayu di Negara Thai. Tesis *PhD Thesis*, Universiti Malaya, Kuala Lumpur, Malaysia.

Zaharani, A, Nor. H. J & Yusmaniza,Y . 2018. GIS Mapping of Dialect Variation in North Perak. *Dialectologia*. 20 (1-20).

Arabic loanwords in languages around the Indian Ocean, and what this tells us about the transmission of Arabic words

Youichi Nagato
Tokyo University of Foreign Studies

Abstract

In this article, I analyze the Arabic loanwords that exist in the languages around the Indian Ocean. In this analysis, I focus on the amount of the Arabic loanwords in each language, the forms these loanwords take, and the concepts they denote (or function they play). By doing so, I clarify the routes through which Arabic words entered these languages and the periods in time when they did so. When I use the term 'loanword,' I mainly refer to borrowed foreign words that are incorporated into the language without any morphological change. This usage corresponds to Einar Haugen's definition of loanwords — "morphemic importation without substitution." However, I also use the term to refer to 'loanblends'— words that feature a mix of native and borrowed constituents (or as Haugen puts it, "morphemic substitution as well as importation") (Haugen, 1950: 214–215).

The Indian Ocean languages I selected for my analysis include those that contain the highest amounts of Arabic loanwords—namely, Persian, Urdu, Indonesian, and Swahili. To these, I added Turkish and Uyghur, because these languages are deeply connected to Persian. One reason why these six languages contain numerous Arabic loanwords is because many of the speakers of these are Muslims.

To sample the loanwords in each language, I referred to languages surveys of the Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies (ILCAA TUFS 2000). This resource contains lists of 2,000 loanwords in Swahili, Indonesian, and Persian. I amended the lists as necessary. For the other languages (Turkish, Uyghur, and Urdu), I compiled my own loanword lists by referring to the literature.

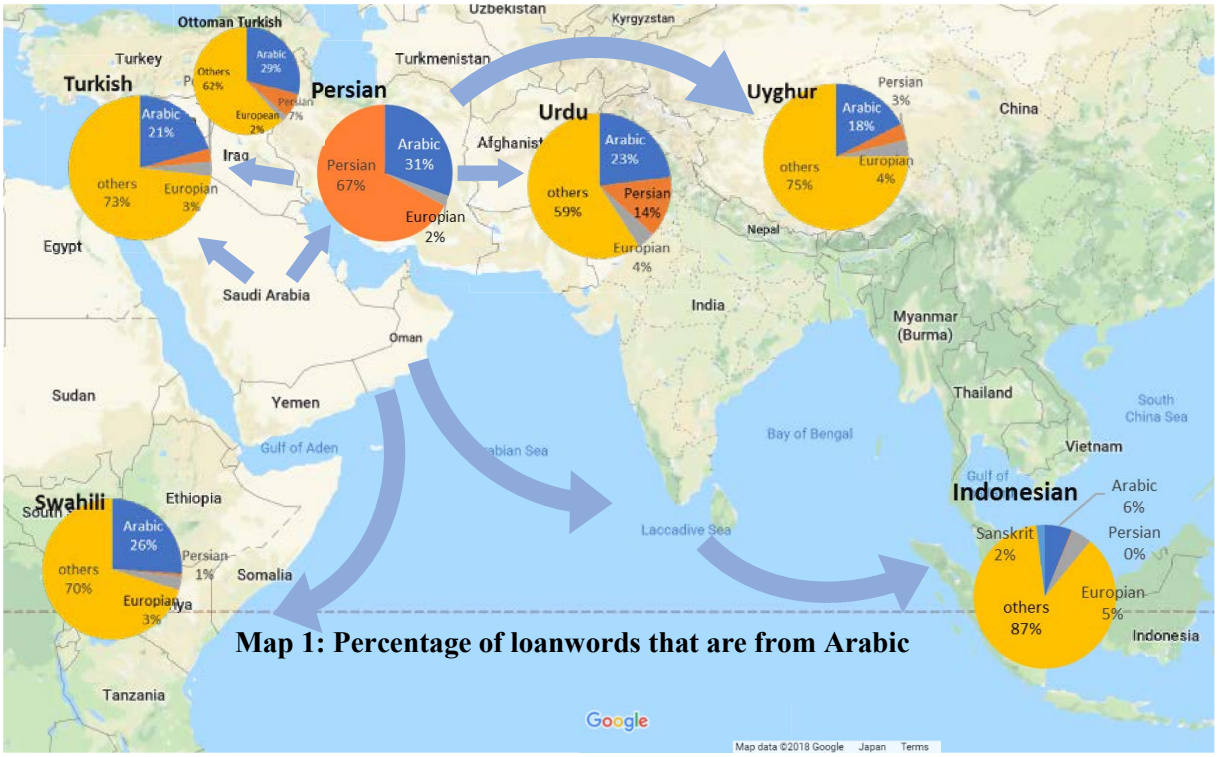
1 Numbers of loanwords

Of the six languages of Persian, Urdu, Indonesian, Swahili, Turkish and Uyghur, Indonesian borrows from Arabic the least — Arabic loanwords account for only 6% of the listed loanwords (125 out of 2,000). Persian had the highest percentage of Arabic loanwords, at 31% (618 / 2,000), followed by Swahili (26% — 519 / 2,000), Urdu (23% — 441 / 1,913), Turkish (21% — 354 / 1,705), and Uyghur (18% — 331 / 1,871). Notably, Turkish contained a higher percentage of Arabic loanwords during the Ottoman era. Of the 2,050 Ottoman-era Turkish words computed based on Katsuta 2002's lexicon, 594 (29%) have an Arabic origin. After the Turkish Republic emerged in 1923, the new government led a linguistic reform in an attempt to restore native words to the Turkish lexicon.

Why does Indonesian contain so few Arabic loanwords compared to the other language? A likely reason is that the Indonesians who sailed the Indian Ocean were mainly traders, and these traders would have had few opportunities to engage directly with Persian or Arabic speakers. On the other end of the scale, why do Arabic loanwords account for as much as 18% of the loanwords in Uyghur despite the fact that the Uyghur-speaking region is situated far from the Arabian Peninsula? A possible answer is that many of these Arabic loanwords came via Persian. The results for Swahili are notable too: the Swahili-speaking region lies adjacent to the Arabian Peninsula; yet, the language also includes loanwords originating in Indian languages, such as ***gari*** 'car'. The presence of such loanwords may be

the result of contact with trade between East Africa and the Indian subcontinent.

This guideline is adapted from the template of PACLIC 18 conference. It is an MS Word template. Users of other word processors should set their corresponding style parameters to conform to the format given here.



Map 1: Percentage of loanwords that are from Arabic

2 Categorizing the loanwords by their features

I found that certain features of the loanwords correlated with the donor language (Arabic, Persian, European language, or other language).

2.1. Arabic loanwords

Words borrowed from Arabic tend to denote concepts such as religion, language, nature, society, philosophy, or greeting, or to act as a conjunction. The words shown below are Arabic loanwords that feature in all of the six languages (the words in Arabic are shown in parenthesis. *j* = [dʒ], *y* = [j]).

Religion: **Friday** (*jumʕa*), **world** (*dunyaː*), **pilgrim** (*ħajj*), **devil** (*ʃayṭaːn*), **tomb** (*qabr, qubuːr*), **victim** (*qurbaːn*)
Language: **letter** (*ħarf, ħuruːf*), **poem** (*ʃiʕr*), **answer** (*jawaːb*), **list /table** (*jadwal*)
Philosophy: **essentials**, **condition** (*ʃarṭ*), **cause** (*sabab*)
Society: **nation**, **people**, **orphan** (*yatiːm*), **present** (*hadiyya*), **duty** (*waːjib, waðiːfa*), **right** (*ħaqq*), **poor** (*faqiːr, miskiːn*)
Nature: **nature** (*ṭabiːʕa*), **air** (*hawaː*), **season** (*mausim*)
Conjugation: **when** (*waqt*), **maybe** (*mumkin*), **finally** (*axiːr*), **for example** (*maθalan*)
Greeting: **thank** (*ʃukr*), **good day**, **Congratulations** (*mubaːrak, mabruːk*)
Others: **character**, **signal** (*iʃaːra*), **safe** (*amaːn, saːlim*), **mysterious** (*ʕajiːb*), **famous** (*maʃhuːr*)

As the above list illustrates, many Arabic loanwords pertain to cultural concepts that existed before the modern age. The Arabic loanword for 'letter' (***ħuruːf***) occupies a particularly important place in each of the six languages. Initially, this loanword possibly referred to the Arabic script (as seen in the Quran) in

particular (as opposed to the written word in general). This distinction subsequently broke down after the six languages started being written in Arabic script. Persian, for example, shifted to Arabic writing from the Pahlavi script. Urdu and Uyghur speakers similarly started using Arabic letters. As for Turkish, Swahili, and Malay (from which Indonesian is derived), these languages also used Arabic orthography, although they later replaced it with Roman characters after European colonial rule. Whatever language they speak, Muslims hold Arabic in high regard because they believe that the Almighty sent down his Quran to the Prophet Mohammed in Arabic. Hence, religious (or specifically Islamic) concepts are more likely to be expressed with an Arabic loanword than any other loanword.

I identified cases where a concept is expressed with an Arabic loanword in all six languages, but not the same Arabic loanword. Oftentimes in such cases, the languages would not borrow a new Arabic word. Rather, each language would deploy one of its existing Arabic loanwords — a loanword the language had borrowed back in pre-modern times. An example is the word 'nation.' The Arabic word for 'nation' is ***ʃaʕb***. However, rather than borrowing this Arabic word, each language deployed one of its existing Arabic loanwords.

Persian ***qoum***, Urdu ***qoum*** < Arbic ***qawm*** 'people, nation',
Indonesian ***rakya*** < Arabic ***raʕya*** 'flock, subject'
Turkish ***halk***, Uyghur ***xalq*** < Arabic ***xalq*** 'creatures, mankind'
Swahili ***watu wa taifa*** [men of nation] < Arabic ***ṭaːʔifa*** 'denomination, sect'

Map 2 'magazine'

The Urdu word ***qoum*** might have been borrowed from Persian, while the Uyghur word ***xalq*** might have originated from the Turkish ***halk***. Another example is the word 'magazine' (see Map 2). Until the 19th century, the Arabic word ***majalla*** ('codex' or 'book') was out of use, but Lebanese author Nāṣif al-Yāzijī helped return it to active use when he introduced it as the Arabic word for 'magazine' (Versteegh, 2014: 231). Thus, given that the Arabic word ***majalla*** came back into use in 19th century, we can infer that the Persian word ***majalle*** and Indonesian word ***majalah*** were borrowed from Arabic no earlier than the 19th century. Urdu, however, deployed a different Arabic loanword for magazine—namely, ***risaːla***, which in Arabic means 'missive' or 'message.'

All six languages use the Arabic word for 'Friday' (***jumʕa***) to describe this day of the week (Persian ***jom'e***, Turkish ***juma***, Urdu ***juma***, Indonesian ***jumat***, Uyghur ***jümä***, Swahili. ***ijumaa***). Islam introduced the seven-day week into these languages before the arrival of European influences. As for the other days of the week, Indonesian uses Arabic loanwords, while the other languages use either Persian loanwords or refer to days of the week with numerals, counting up from Friday.

2.2. Persian loanwords

Persian loanwords feature extensively in Turkish, Urdu, and Uyghur, but fewer exist in Indonesian and Swahili. Table 1 shows examples of Persian loanwords in each language.

Table 1: Persian loanwords (words in parenthesis are words other than Persian loanwords)

Meaning	Persian	Turkish	Urdu	Indonesian	Swahili	Uyghur
'sugar'	*ʃekar*	***ʃeker***	***ʃakar***	(*gula*)	(*sukari*)	***ʃikär***
'market'	*bāzār*	***pazar***	***baːzaːr***	***pasar***	(*soko*)	***bazar***
'if'	*agar*	***eːr*** < *eɣer*	***agar***	(*kalau*)	(*kama*)	***ägär***
'colour'	*rang*	***renk***	***rang***	(*warna*)	***rangi***	***räŋ***
'spring time'	*bahār*	***bahar***	***bahaːr***	(*musim buŋa*)	(*masika*)	***bahar***
'city'	*ʃahr*	***ʃehir***	***ʃaher***	(*kota*)	(*mji*)	***ʃähär***

The languages deploy Persian loanwords for commonplace concepts such as market, city, color, and some foods. Also of note, some these languages use a Persian loanword for the function word 'if.' In the above table, the only Indonesian example of a Persian loanword is the word for market (***pasar***). Three of the other words—***gula***, ***warna***, and ***kota***—originate from the Sanskrit ***gula***, ***varṇa***, and ***kuṭa***, while ***kalau*** is a hybrid of the Sanskrit ***kāla*** and the Arabic ***lau***. The Indonesian word for spring time (***musim buŋa***) reflects an Arabic influence—***musim*** originates from the Arabic ***mausim***, meaning 'high season.'

2.3. Loanwords from European languages

In the ILCAA TUFS 2000, loanwords from a European language account for only between 2 to 5% of the loanwords in each of the six languages. Although some European loanwords appear in basic vocabulary, they appear much more prominently in scientific and technical concepts. The six languages were exposed to European influences in the modern period. Hence, European loanwords typically denote more modern concepts. By the modern era, the metric system and the Gregorian calendar had become popular throughout the world. Accordingly, the six languages started using many European loanwords for measurements and calendar dates. Many of the month names originate from Latin. Turkish, however, derives its month names from a mixture of sources. ***mart*** (March) and ***ağustos*** (August) are from the Gregorian calendar. However, other months are taken from Arabic month names, which themselves originate in the Babylonian calendar. Uyghur is also exceptional in that it refers to months with numerals.

When the languages use a European loanword to express a concept, the specific donor language often varies between them. To take 'bicycle' as an example (see Map 3), the Turkish word for the vehicle is ***bisiklet***, which is borrowed from the French *bicyclette*. Urdu borrows from the English 'cycle' to make ***sa:ikl***. Similarly, Swahili uses the word ***baisikeli*** 'bicycle'. In Indonesian, bicycle is ***səpeda***, from the French *vélocipède*. The Uyghur word for bicycle is ***welsepit***, which originates from the Russian *Велосипед*. Persian, on the other hand, uses the word ***dočarxeh***, which is a composition of two native words—***do*** 'two' and ***čarxeh*** 'wheel'.

Map 3: Word for 'bicycle'

The following are examples of other concepts expressed with European loanwords.

Measure **:** meter, hectare, liter, gram, ton
Month **:** March, May, August
Machine **:** machine, radio, telephone, camera, bicycle (vélocipède), train, car, tramcar
Modern items **:** bank, movie, post-card, postal stamp, policeman, tobacco (16 C.)
Greeting **:** hello (on the telephone)
Others **:** Japan, cement, hotel, ticket, music

Map 4 ***gari*** 'car'

2.4. Indian loanwords in Swahili

gari, the Swahili word for 'car' (see Map 4), originates in the Urdu and Hindi word ***gāṛī***, which itself comes from the Sanskrit ***gantrī***. This and other Indian loanwords in Swahili would have been introduced as a result of contact with traders from the Indian subcontinent.

3 Function words

According to D. W. Whitney (1881; cited in Haugen 1950: 224), "nouns are most easily borrowed, then the various other parts of speech, then suffixes, then inflections, then sounds." Nonetheless, these languages use Arabic for the function words listed in Table 2 below. All six languages use Arabic loanwords for the function words 'for example' (Arabic ***maθalan***), 'maybe' (Arabic ***mumkin***), and 'when' (***waqt*** —'time'). The five languages other than Indonesian use a loanword for the function word 'but' (Arabic ***amma:. la:kin***). Given that these languages would have had their own native conjunctives, adverbs, and the like, the fact that they all borrowed function words from Arabic might have a socio-psychological explanation.

Table 2: **Function words borrowed from Arabic** (words in parenthesis are words other than Arabic loanwords)

Meaning	Persian	Turkish	Urdu	Indonesian	Swahili	Uyghur
'and'	*va*	*ve*				*vä*
'almost'	*taqrīban*		*taqri:ban*		*karibu (zote)*	*asasän*
'but'	*ammā*	*ama*	*le:kin*		*lakini*	*ämma, likin*
'(in that) case'	*(dar īn)sūrat*	*(o) hal(de)*	*su:rat*	*(dalam) hal (itu)*	*wakati*	*waqit*
'even if'				*walaw(pun)*		*mabada*
'for example'	*masalan*	*mesela*	*masalan*	*misal(nya)*		*mäsilän*
'maybe'	*momken*	*mükün*	*mumkin*	*mungkin*	*labda*	*mumkin*
'when (clause)'	*waqtī(ke)*	*vakit*	*waqt*	*waktu*	*wakati (wa)*	*waqit*

4 Borrowed forms

A loanword's morphological features can provide hints as to the word's channel of transmission. Hereunder, we examine two Arabic phonemes ***q*** and ***ḍ*** and Arabic's feminine endings.

4.1. Arabic ***q*** equivalent

The Arabic phoneme ***q*** can represent a voiceless consonant (ʔ or q) or voiced consonant (g or ɢ). In urban dialects and in Modern Standard Arabic, it represents the former. 'Coffee,' for example, is ***ʔahwa*** in Egyptian Arabic, ***ʔahwe*** in Syrian Arabic, ***qəhwa*** in Moroccan Arabic, and ***qahwa*** in Modern Standard Arabic. On the other hand, ***q*** represents a voiced consonant (specifically, g) in Arabian peninsula dialects and other Bedouin dialects. It may also have represented a voiced consonant (specifically, ɢ) in the Quranic Arabic of the 7th century. The same word, 'coffee,' is pronounced ***gahwa*** in Saudi Arabian and ***ghawa*** in Bahrain. As for the Arabic loanwords in the six languages, the ***q*** equivalents in these loanwords represent a voiceless consonant (either ***k*** or ***q***) (see Table 3).

Table 3: How the phoneme *q* is pronounced in Arabic loanwords

Meaning	Arabic	Persian	Turkish	Urdu	Indonesian	Swahili	Uyghur
'pen'	***q***alam	***q***alam	***k***alam	***q***alam	***k***alam*	***k***alamu	***q***äläm
'right'	*ħa**qq***	*ha**qq***	*ha**k***	*ha**qq***	*ha**k***	*ha**k**i*	*ha**qq***
'tomb'	***q***abr	***q***abr	***k***abir	***q***abr	***k***ubur	***k***aburi	
'east'	*ʃar**q***	*maʃre**q***	*ʃar**k***	*maʃri**q***		*maʃari**k**i*	*ʃär**q***

*Indonesian ***kalam***: Al-Qalam, name of surah 68 of the Quran. (Jones 2007: 140)

We can rule out the possibility that the six languages received the above loanwords directly and orally from the Arabian Peninsula—if they had, the ***q*** equivalents would have represented a voiced rather than voiceless consonant. Nor could they have received them directly from the urban (Mediterranean) Arabic dialects, as the speakers of these dialects did not directly engage in the Indian Ocean trade. Thus, the fact that the ***q*** equivalents represent a voiceless consonant implies that the loanwords in which they appear

came from classical literary forms of Arabic. Alternatively, given that ***q*** represents a voiceless ***q*** consonant in Iraq's sedentary dialects (Blanc 1964; Versteegh 2014: 202), this voiceless pronunciation might have been orally transmitted to Iran (Iraq's neighbor). Such an oral transmission would be consistent with the fact that the Persian equivalent of the Arabic phoneme ***ḍ*** (which I discuss in 4.2 below) represents the fricative consonant ***z***.

With regard to the loanwords in Indonesian, many of these loanwords were transmitted via writing rather than orally. As Jones (2007) argues, 'the Arabic loan-words have been consistently borrowed from the literary forms of Arabic – from Classical Arabic in the case of the earlier loans, and from Modern Standard Arabic today" (p. xxiv). As for Swahili, many of the Arabic loanwords in this language would have come from the spoken forms of Arabic used among Arab traders who operated along the east coast of Africa. On the other hand, given that the Swahili equivalent of the phoneme ***q*** represents the voiceless consonant *k*, we cannot discount the possibility that the language borrowed from classic literary forms of Arabic.

4.2. Arabic ***ḍ*** equivalent

The Arabic phoneme ***ḍ*** can represent three sounds: the plosive consonant dˤ, the fricative consonant ðˤ, or the lateral plosive consonant dlˤ. In urban dialects (such as in Rabat, Morocco) and Modern Standard Arabic, it represents dˤ. In Bedouin, Iraqi, and Arabian dialects, it represents ðˤ. In 7th century Classical Arabic, it represents dlˤ. 'Judge,' for example, is pronounced ***qaːdˤiː*** in urban dialects, ***gaːðˤiː*** in Bedouin dialects and some other Arabian Peninsula dialects, and ****ɢaːdlˤi*** in Classical Arabic. Table 4 shows the corresponding sounds in the six languages.

Table 4: Pronunciations of Arabic *ḍ* in loanwords

Meaning	Arabic	Persian	Turkish	Urdu	Indonesian	Swahili	Uyghur
ḍaʕiːf	'weak'	***za'īf***	***zayıf***	***za'iːf***		***zaifu***	
qaːḍi	'judge'	***qāzī***	***kadı*** 'Islamic judge'	***qaːziː*** 'Islamic judge'	***kadi*** 'Islamic judge'	***kazi*** 'Moslem judge'	
ħaḍir	'presence'	***hāzer***	***hazır*** 'prepared'	***haːzir***	***hadir, hadlir***		***haːzır*** 'present'
riḍaː	'contentment'	***rezā***	***rıza***	***rizaː***	***redla, réla*** 'willing'	***rizaa***	

In Arabic loanwords in Persian, the ***ḍ*** equivalent represents the fricative ***z*** consonant. It is probably pronounced this way because Iran is close to spoken Iraq dialects and Gulf Arabic. The ***ḍ*** equivalents in Turkish, Urdu, Uyghur, and Swahili loanwords represent the very same pronunciation, suggesting that these loanwords entered the languages via Persian. On the contrary, Indonesian reflects the plosive [dˤ], which is a normative literary pronunciation, as in ***kadi*** and ***hadir***

As for Indonesian, the ***ḍ*** equivalent represents ***dl*** in some loanwords and ***l*** in others. Examples include ***hadlir*** 'presence' (< Arabic ***ħaːḍir*** 'presence'), ***rela*** 'do with pleasure' (< Arabic. ***riḍaː*** 'satisfaction'), and ***pərlu*** 'need' (< Arabic ***farḍ*** 'impose'). These pronunciations reflect the 7th century classical forms of Arabic. We find a similar example in the Spanish word for 'mayor' ***alkalde*** (spelt ***alcalde*** in Spanish), which comes from ***al-qaːḍi***.

4.3. Arabic feminine endings: ***-a*** and ***-at***

In Arabic, grammatically feminine nouns end in ***-a***, or ***-at*** if the next word is a noun. In Quranic Arabic, however, all feminine nouns end in ***-at*** unless they come at the end of the sentence, in which case they use ***-a***. In Persian and Turkish, this ending surfaces as ***-e*** rather than ***-a***, which might reflect the fact that spoken forms of Syrian and Iraqi dialects use this ending. In the loanwords used in the other languages, the feminine ending takes the form of ***-a*** or ***-at***. The following table shows some examples.

Table 5: The Arabic feminine ending.

Meaning	Arabic	Persian	Turkish	Urdu	Indonesian	Swahili	Uyghur
ʿFriday’	*jumʕa*	*jom'**e***	*juma*	*juma*	*(hari)jum**at***	*ijumaa*	*jümä*
ʿsignal’	*iʃa:ra*	*eʃāre*	*iʃar**et***	*iʃa:r**a***	*iʃar**at***	*iʃar**a***	*iʃarä*
ʿpresent’	*hadi:ya*	*hadīye*	*hediye*		*hadi**ah***		*hädiyä*
ʿrest’	*istira:ħa*	*esterāh**at***			*istirah**at***		
ʿgovernment’*	*ħuku:ma*	*hokūm**at***	*hükümet*	*kuku:m**at***			*hökümat*

* ‘Government’ is not in the 2,000-word loanword lists.

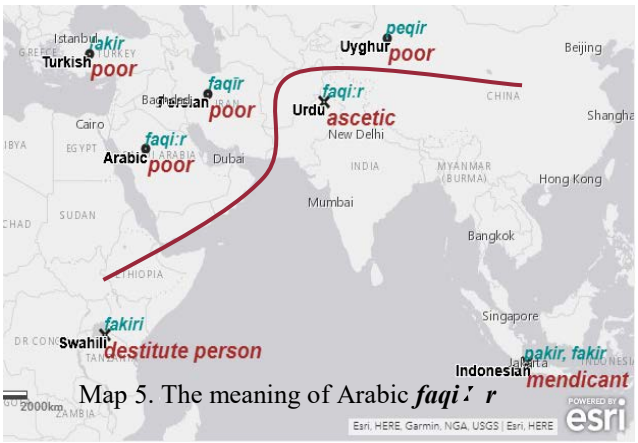
Table 6 below shows the extent to which feminine endings in Persian’s feminine Arabic loanwords match the equivalent endings in the other five languages.

Table 6: Correspondence in feminine endings (*-a* and *-at*) between the Arabic loanwords in different languages (Persia and the other five languages)

		Urdu		Turkish		Uyghur		Swahili		Indonesian	
		-a	***-at***	***-a*** (*-e*)	***-at*** (*-et*)	***-a*** (*-ä*)	***-at***(*-ät*)	***-a***	***-at***	***-a***	***-at***
		27	15	14	16	8	12	29		1	13
Persian	***-a*** (*-e*)	24	1	14	2	8	1	17		1	6
	-at	3	14		14		11	12			7

As the table shows, the feminine Arabic loanwords in Urdu, Turkish, and Uyghur are likely to use the ***-a*** ending where the Persian equivalent uses the ***-a*** (*-e*) ending. Likewise, they frequently use the ***-at*** ending where the Persian equivalent does so. This finding suggests that the Arabic loanwords entered these languages via Persian. By contrast, all feminine Arabic loanwords in Swahili, and 26 out of 29 of them in Indonesian, always use the ***-a*** ending. This finding indicates that the endings in Swahili and Indonesian do not correlate with those in Persian. From this finding, we can infer that Swahili did not acquire the Arabic loanwords via Persian. The use of the ***-a*** ending in Swahili’s feminine Arabic loanwords suggests that these loanwords might have been orally transmitted. On the other hand, the use of the ***-at*** ending in the case of Indonesian implies that the loanwords originated in literary forms of Arabic. The three loanwords in Indonesian that do use the ***-a*** ending are ***jumlah*** ‘total’, ***majalah*** ‘magazine’, and ***sejarah*** ‘history’ (< Arabic. ***ʃajara*** ‘tree’).

Thus, Persian was a stepping stone in transmission of Arabic loanwords to the adjacent languages Turkish and Urdu, and to the Central Asian language Uyghur. On the other hand, Persian was not involved in the transmission of Arabic loanwords to Indonesian or Swahili. As regards the loanwords in Indonesian, Jones (2007: 13) argues that “traders and missionaries from India were the main vectors of Arabic and Persian.”



Map 5. The meaning of Arabic ***faqi:r***

Incidentally, Perry (1991; cited in Versteegh, 2014: 323) discussed the ***-a*** (*-e*) and ***-at*** endings in Persian, and noted the kinds of nouns in which they appear. Specifically, the ***-a*** (*-e*) ending frequently appears in concrete nouns, and these nouns were transmitted orally. On the other hand, many of the nouns ending in ***-at*** are abstract nouns, and these originated in written forms. I identified similar trends in my analysis, but I also found many counter-examples.

5 The semantic features of loanwords

The meaning of an Arabic loanword can sometimes vary depending on the language that borrowed it. Such semantic discrepancies offer a further clue to the transmission routes.

I now discuss an Arabic word that holds the same meaning in its borrowed forms in Urdu, Indonesian, and Swahili. This word is ***faqiːr*** 'poor (poverty)'. Map 5 shows the meanings of the corresponding loanwords (borrowed meanings) in the six languages. The borrowed meanings in the above three languages are all 'poor person'. However, the Urdu and Indonesian versions convey a more aesthetic nuance by denoting the concepts of 'ascetic' and 'mendicant.' This finding is further evidence that Indonesian borrowed loanwords from Arabic directly or has a closer connection to Indian languages than to Persian.

Borrowed meanings can help us estimate the time when they were borrowed too. Map 6 shows the borrowed meanings of the Arabic word ***maktab***. In Arabic, ***maktab*** originally meant 'writing-school.' Nowadays, however, it typically denotes 'office' and by extension 'desk.' The corresponding loanwords in Persian, Turkish, Urdu, and Indonesian all refer to 'school,' and the loanword in Urdu refers specifically to 'Islamic school.' Thus, from these borrowed meanings, we can infer the languages borrowed ***maktab*** at a time when it still meant 'writing-school' in Arabic.

Map 6: the borrowed meaning of ***maktab***

6 Summary

I analyzed Arabic loanwords in six languages around the Indian Ocean: Persian, Turkish, Urdu, Uyghur, Indonesian, and Swahili.

When we observe the features of the Arabic, European, and Persian loanwords in the six languages, we can understand that these languages typically borrowed their loanwords from high-prestige languages and tended to borrow the same kinds of words as each other. The Arabic loanwords denote concepts related to religion (specifically, Islam), language, philosophy, and society. Loanwords from European languages, on the other hand, denote more modern concepts related to science and technology.

My analysis also highlighted cases where the borrowed meanings have deviated from the original meaning in the donor language, yet remained semantically consistent among the six languages. Such semantic consistency is particularly prevalent among the modern European loanwords. The situation is more complex, however, for words and concepts predating the Enlightenment, such as 'color' and the function word 'if'—five of the languages borrowed these words from Persian, while Indonesia borrowed from Sanskrit.

I also analyzed the phonemic structure of Arabic loanwords (the pronunciation of the phonemes ***q*** and ***ḍ***, and feminine endings). Judging from this phonemic structure, many loanwords probably came from spoken forms of Arabic by oral transmission, but the loanwords in Indonesia probably came from classical literary forms.

We also saw how the borrowed meanings can indicate the time of transmission. Prior to the modern age, the Arabic word ***maktab*** denoted 'school'. The fact that the ***maktab*** loanwords in the six languages have this very same meaning tells us that the languages borrowed the word in pre-modern times. The six languages also use some Arabic loanwords for more modern concepts like 'society.' In such cases, however, the languages did not borrow new Arabic words; rather, they deployed their existing Arabic loanwords, resulting in a mix of loanwords for the same concept.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP16K02877.

References

Blanc, Haim 1964. *Communal Dialects in Baghdad.* Cambridge, MA: Harvard University Press.
Haugen, Einar 1950. 'The Analysis of Linguistic Borrowing' in *Language* vol. 26, no. 2, pp. 210-233.
Jones, Russel ed. 2007. *Loan-words in Indonesian and Malay.* KITLV Press, Leiden.
Kokken 2007. *Loanwords and Foreign words in Japanese* (『第30回「ことば」フォーラムー日本語の中の外来語と外国語』), National Institute for Japanese Language and Linguistics.
Perry, John R. 1991. *Form and Meaning in Persian Vocabulary: The Arabic Feminine Ending*, Costa Mesa CA: Mazda Publishers.
Versteegh, Kees 2014. *The Arabic Language*, second ed. Edinburgh U.P.
Whitney, W. D. 1881. 'On Mixture in Language'. in *TAPA* 12.5-26.

Word Sources

Hayasi, Tooru. 1996.『現代ウイグル語ウルムチ方言語彙集』(*A vocabulary of the Modern Uyghur Ürümqi dialect*), TUFS AA

ILCAA TUFS, 2000. 『言語調査票2000年版』 http://www.aa.tufs.ac.jp/~mmine/kiki_gen/query/ aaquery-1.htm; 「aa_2000」 http://www.aa.tufs.ac.jp/~kmach/AA2000/AA_2000.html

Ikeda, O. & S. Takeda 1981. 『現代アラビア語小辞典』 (*Qāmūs al-ʕarabīya al-muʕāṣira*), Daisan-shokan.

Katsuta, S. 2002. 『オスマン語文法読本』(*A Grammar of Osman Turkish*), Daigaku-shorin.

Kuroyanagi, T. 1998. 『ペ日・日ペ現代ペルシア語辞典』 (*Modern Persian dictionay*), Daigaku-shorin, Tokyo.

Kuroyanagi, T. 2008. 『アラビア語・ペルシア語・ウルドゥー語対照辞典』 (*A Contrastive Dictionary of Arabic, Persian and Urdu*), Daigaku-shorin.

Perrott, D.V. 1965. *Teach yourself Swahili dictionary*. Hodder and Stoughton Ltd.

Suenaga, A. 1992. 『インドネシア語辞典』(*Kamus Terbaru Bahasa Indonesia-Jepang*), Daigaku-shorin.

Suzuki, T. & Y. Asada. 1980. 『ウルドゥー語常用6000語』 (*Urdū 6000 basic words*), Daigaku-shorin.

Suzuki, T. & Y. Asada. 1994. 『日本語ウルドゥー語小辞典』 (*Japānī-Urdū Lugat*), Daigaku-shorin.

Takeuchi, K. 1985. 『現代ウイグル語基礎1500語』 (*Uyġur Tilining Asasiy 1500 sozliri*), Daigaku-shorin.

Takeuchi, K. 1989. 『トルコ語辞典』 (*Türkçe-Japonca sözlük*), Daigaku-shorin.

Tamura, H. ed. 1980.『詳解アラビア語ー日本語辞典』(*Al-Qāmūs al-mufaṣṣal Arabī-Yābānī*), Middle East Institute of Japan.

Palatalization and hypercorrection in the history of Korean

Rei Fukui

Department of Korean Studies,

Graduate School of Humanities and Sociology,

The University of Tokyo

Abstract

This paper discusses two kinds of palatalization that have been assumed to have taken place around the 17th century in the history of the Korean language, by examining geographical distributions of lexical items affected by these phonological changes in modern dialects. It will be also shown that a kind of hypercorrection occurred as a result of the interaction between these two types of palatalization.

More specifically, the following three topics will be discussed: (1) geographical distributions of dialects with or without these two types of palatalization, based on the data collected by Ogura Shinpei (1944), (2) the reason or motivation for having forms which underwent hypercorrection, (3) the loss of initial /n/ before /i/ or /j/ which usually accompanies palatalization. As to the last topic, we have some interesting cases of dialects preserving this nasal in spite of the fact that they are typical dialects which underwent palatalization. In order to explain this anomalous phenomenon, a historical interpretation will be presented.

1 Introduction

Palatalization has been one of the most famous topics in the history of Early Modern Korean[1] phonology, because it affected not only the phonological system but also the system of orthography in many ways. Some writers tried to write according to the old tradition, but the discrepancy between written forms and actual pronunciation became more and more prominent so that the system of orthography began to fall into confusion.

There were two kinds of palatalization in the history of Early Modern Korean:

(a) Palatalization of the alveolar stop /t/ followed by either /i/ or /j/
(b) Palatalization of the velar stop /k/ followed by either /i/ or /j/

The results of these two types of changes are the same: they merged with the alveolo-palatal affricate /c/. Also, the initial nasal /n/ before /i/ or /j/ is lost in the dialects which underwent the change (a), and this change has been discussed together with the palatalization.

The central dialect (the ancestor of the present-day Seoul dialect) experienced only the (a) type palatalization but other dialects, especially southern dialects, experienced both types of palatalization. And this difference gave rise to some new forms created by hypercorrection which will be discussed later in this paper.

More specifically, the following three topics will be discussed:

(1) To clarify geographical distribution of dialects with or without these two types of palatalization, based on the data collected by Ogura Shinpei (1944).

[1] Early Modern Korean refers to the Korean language used between the beginning of the 17th century and the end of the 19th century (Lee and Ramsey (2011 : 7)).

(2) To discuss the reason or motivation for having forms that underwent hypercorrection. This has much to do with the way these two types of palatalization have occurred.

(3) It has been generally believed that the loss of initial /n/ before /i/ or /j/ accompanies the palatalization (a), but we have some interesting cases of dialects preserving this nasal in spite of the fact that they are typical dialects which underwent palatalization. In order to explain this anomalous phenomenon, a historical interpretation will be proposed.

2 Palatalization of the alveolar stop /t/ before /i/ or /j/

According to Lee Ki-Moon (1998: 208), this type of palatalization is known to have occurred at the end of the 17th century or at the beginning of the 18th century in the central dialect. In some other dialects, especially southern dialects, this seems to have occurred much earlier than the central dialect, as can be known from sporadic occurrences of palatalization in the reprinted edition of the *Tusi ŏnhae*[2] published in Kyŏngsang province in 1632 (Ahn, Pyong-Hi (1957)).

The following is a selected list of items found in Ogura (1944) for examining this type of palatalization. For each item, Middle Korean form (MK) and modern standard form (Std) is given in the Yale Romanization, followed by English gloss, and the original item name given in Japanese (Ogura 1944). Linguistic maps are shown at the end of this paper.

	MK	Std	gloss	Ogura (1944)
Map No. 1	tiph + sin	ciphsin	straw sandals	藁の草鞋 (p. 157)
Map No. 2	ilthi	ilchi	agreement	一致 (p. 511)
Map No. 3	tinakata	cinakata	to go past	過ぎ行く (p. 378)
Map No. 4	thyenti	chenci	heaven and earth	天地 (p. 15)
Map No. 5	tywohta	cwohta	to be good	善い (pp. 355–356)

For these items, only such forms that have *not* undergone palatalization are given in Ogura (1944), because for all the other areas, including Seoul, we have only palatalized forms. In such cases, he simply left blank with an additional remark that the standard form (mainly the one used in Seoul) is used 'in general', 'in many places' or 'throughout the Korean peninsula'.

As can be seen from the Map 1 through Map 5, non-palatalized forms are found in the majority of P'yŏngan province (平安道) and in the northern part of the North Hamgyŏng province (咸鏡北道) known as the 'Ryukchin' (六鎭) dialects.

In the cases of items containing /tj/ followed by a vowel other than /i/ (namely items 4 and 5), we have non-palatalized forms in P'yŏngan dialects but the glide /j/ is lost. For such items, the most conservative forms such as [t^hjɔn-di][3] for the item 'heaven and earth' and [t^hjo-t^ha] for the item 'to be good' are found in the northern part of the North Hamgyŏng dialects.

3 Palatalization of the velar stop /k/ before /i/ or /j/

This type of palatalization is not fully treated in Lee Ki-Moon (1998), because this has not occurred in the central dialect and not well documented so that we are not sure when this type of palatalization occurred.

The following is a selected list of items found in Ogura (1944) in order to examine this type of palatalization, arranged in the same way as the items shown in the previous section.

[2] *Tusi ŏnhae* (杜詩諺解) was originally published in 1481 but a reprinted version was published in 1632 in the Kyŏngsang province.

[3] Citations from Ogura (1944) are made according to the phonetic transcription he uses.

	MK	Std	gloss	Ogura (1944)
Map No. 6	kilh	kil	road	道 (p. 35)
Map No. 7	kitwong	kitwung	pillar	柱 (pp. 117–118)
Map No. 8	kye	kye	bran	糠 (pp.51–52)
Map No. 9	kyeth	kyeth	side	側 (pp.51–52)
Map No. 10	kyecip	kyeycip	woman	女 (pp.65–66)
Map No. 11	khi	khi	winnow	箕 (pp. 180–181)

These maps show that non-palatalized forms are used mainly in the north western part of the peninsula and palatalized forms in the north eastern and southern parts of the peninsula.

To sum up, non-palatalized forms are found in the following areas:

P'yŏngan (平安道)
Northern part of the North Hamgyŏng (咸鏡北道)
Hwanghae (黄海道)
Kyŏnggi (京畿道) including Seoul.

And palatalized forms are found in all other areas:

Southern part of North Hamgyŏng
South Hamgyŏng (咸鏡南道)
Kyŏngsang (慶尚道)
Chŏlla (全羅道) including Cheju (済州)
Ch'ungch'ŏng (忠清道)
Kangwŏn (江原道)

The difference between the two types of palatalization is that the k-palatalization has wider areas of non-palatalized dialects, as can be schematically shown in the following table:

Table 1. Schematic geographical distributions of t- and k-palatalization.
(NH = North Hamgyŏng, N = non-palatalized, P = palatalized)

Provinces		t-palatalization		k-palatalization	
P'yŏngan	northern NH	N	N	N	N
	southern NH		P		P
	South Hamgyŏng		P		P
Hwanghae	Kangwŏn	P	P	N	P
Kyŏnggi		P		N	
Chŏlla	Kyŏngsang	P	P	P	P
Cheju		P		P	

4 Hypercorrection

The following words are known to have a peculiar history as to the relationship between its origin and modern forms.

	MK	Std	gloss	Ogura (1944)
Map No. 12	timcoy	kimchi	pickles	漬物 (pp. 161– 162)
Map No. 13	tisay	kiwa	roofing tile	瓦 (p. 118)
Map No. 14	cilsam	kilssam	weaving	紡績 (p. 256)

One of the most famous examples is the item *kimchi* (No. 12 Korean pickles). Despite its modern standard form with an initial k, its oldest attested form was MK *timchoy* which is from a Sino-Korean word '沈菜'. If the modern standard form *kimchi* were the result of a phonetic change from the MK form, we would have to assume a change like ti > ki or ti > ci > ki, but this is certainly not the case. By the way, if we look at the Map 12, we have forms like [tʃim-tʃʰi] in many typically palatalizing dialects and it is easy to see that such a form is a normal result of the phonetic change timchʌi > cimchi [tʃim-tʃʰi]. Therefore, a different kind of explanation is required about the emergence of forms beginning with ki- and one possible way of explaining this is via hypercorrection. Speakers who spoke dialects with k-palatalization tended to have feelings like rural, unsophisticated and so on, when they became conscious of the fact the speakers of the central dialect do not have k-palatalization, so that they wrongly replaced palatalized ci- with ki- (Similar views are found in Lee and Ramsey (2011: 262)).

Another famous example is the item *kiwa* (No. 13 roofing tile) which shows similar distribution pattern to *kimchi*. Palatalizing dialects have forms beginning with ci- but non-palatalizing dialects show forms beginning with ki-. In the case of the item *kilssam* (No. 14 weaving), the distribution pattern is different from the above two cases but the data for Hwanghae, Kyŏnggi and Chŏlla provinces are lacking so that it is difficult to make an exact comparison.

Beside these items, we have some other hypercorrected words for which Ogura (1944) does not present any data. The item *kich* (feathers) is another famous example.

MK	Std	Non-standard variant	gloss
cich	kich		feathers
cywungsoyng (< '衆生')	cimsung	kimsayng	laypersons, people > beast

The next item *cimsung* (beast), a Sino-Korean word which was originally used in the context of Buddhism, has a non-standard variant *kimsayng*. The Middle Korean word *cywungsoyng* (< '衆生') had two meanings: (1) living things in general, and (2) a beast, but now it is usually used with the latter meaning.

This item is not included in Ogura (1944), nor is it well documented in any other sources so that I would like to explain my own experience about this form. When I was in Seoul in 1985, I had a chance to record a TV program (KBS2) in which the late Professor Lee Hi Sung[4] (90 years old at the time) gave a lecture on the founding myths of Korea. In this lecture, he pronounced the word *cimsung* as *kimsɛŋ* in the following contexts.

(1) 최초부텀 인간이 아니고 어느 <u>김생</u>이 변해서 사람이 됐다. (It wasn't human being from the beginning, but a beast transformed into a human being.)
(2) 오늘날같이 우리가 볼 수 있는 사람이 아니라 <u>김생</u>이다. <u>김생</u>이 변했든지 어쨌든지 해서 사람이 된다. (It wasn't a human being that we can see nowadays, but a beast. A beast transformed into a human being in some way or other.)

Next, I would like to pay attention to some minor cases found in Ogura (1944). The next two items are used only in one or a couple of places:

[4] 1896-1989, born in Kyŏnggi Kwangju.

MK	Std	gloss	Data in Ogura (1944)	Localities Ogura (1944)
tyel	cel	temple	[kjəl]	three places[5] in North Hamgyŏng
tinaka-	cinaka-	go past	[ki-na-kan-da]	Hwoyryŏng (會寧) in North Hamgyŏng

Lastly, I would like to discuss a little more on the process of hypercorrection. The theory of hypercorrection is well-motivated to explain the emergence of forms like *kimchi*, *kiwa* and so on. However, there still remain questions like the following:

(1) Who started to use hypercorrected forms?
(2) Why do we have hypercorrected forms for a particular group of lexical items?

As to the first question, it would be the speakers of the k-palatalization dialects if we take the explanation on the emergence of hypercorrection literally. However, one notable thing about the geographical distribution of items such as *kimchi, kiwa* and so on, is that the areas that have 'hypercorrected' forms almost coincide with those that *do not have* k-palatalized forms. In other words, speakers of k-palatalization dialects do not use hypercorrected forms. If so, an alternative answer would be the speakers who live at the border of palatalizing and non-palatalizing dialects (Speakers of Kyŏnggi dialects seem to be good candidates).

As to the second question, it seems to have much to do with the complexity of lexical histories of items in question. For example, the word *kimchi* is not used in all areas. In some areas, forms such as [tʃi], [ʔtʃan-dʒi] are used and the word *kimchi* is most probably an imported word (see Map 12). The word *kiwa* is also known for having a complex history as has been shown by Lee Ki-Moon (1991).

5 Initial nasal /n/ before /i/ or /j/

According to Lee Ki-Moon (1998: 209), the alveolar nasal /n/ dropped before /i/ or /j/ in the second half of the 18th century. The process of this change can be presumed something like the following:

$$[ni\text{-}] > [n^{d}i\text{-}] > [n^{3}i\text{-}] > [ʒi\text{-}] > [i\text{-}]$$

The reason why this change parallels the t-palatalization is that the change from the voiced stop-like off-glide of the nasal to an affricate is essentially the same as the t-palatalization. By the same token, Sino-Korean reading of the lái initial (来母) [r] can also be discussed together, because in Early Modern Korean it is generally rendered as /n/ before /i/ or /j/.

	MK	Std	gloss	Ogura (1944)
Map No. 15	ni	i	tooth	歯 (p. 93)
Map No. 16	li sepang	i Sepang	Mr. Lee	李書房 (p. 70)

First, we will discuss the item 16 *li sepang* (Mr. Lee)[6]. The initial n- or r- is preserved in P'yŏngan and northern part of the North Hamgyŏng as was the case with the t-palatalization. It is also interesting to note that r- is preserved not only in four places in the northern part of the North Hamgyŏng but in two places in the eastern part of the North P'yŏngan province. The preservation of the Sino-Korean initial r- may be considered as artificial, especially if we consider the modern North Korean orthography, but it seems difficult to take these dialectal pronunciations as artificial.

Next, we would like to consider the item No. 15 *i* (tooth) which involves a previously unnoticed characteristics as to the places preserving the initial n. As can be seen in the Map 15, n- is preserved not

[5] Myŏngsin (明臣), Chongsŏng (鍾城) and Kyŏngwŏn (慶源).

[6] The fact that this item is included in Ogura (1944 : 70) clearly shows that he was interested in the phonetic aspect of this item, not the semantic or lexical aspect.

only in P'yŏngan and northern part of the North Hamgyŏng as expected but in Cheju and in one place in Chŏlla province (Kwangju 光州). For Cheju dialects, this can be verified by other sources such as Hyŏn (1985), in which this item is recorded either as /ni/ or /nɯi/. Also, the homophonous word /i/ ('lice') is pronounced exactly like the word for 'tooth', e.g. /ni/ or /nɯi/.

By the way, Cheju dialects can be said to be the ones which typically underwent both types of palatalization discussed above, so that it seems difficult for such dialects to maintain the initial /ni/. Therefore, we need to postulate a different origin for the words maintaining an initial sequence /ni/, and this can be hinted at by the alternative form of this word 'nɯi' presented by Hyŏn (1985).

If we postulate two kinds of sequences /ni/ and /*nɨi/ in Proto-Korean, Cheju forms corresponding to MK /ni/ can be derived in the following way:

*nɨi > ni or nɨi (for 'tooth' and 'lice')
ni > i (for all other cases)

The fact that MK had no cases of words having the sequence /nɨi/ suggests that the distinction between /ni/ and /nɨi/ had been lost in MK and merged into /ni/. For consonants other than /n/, we have a series of distinctions such as /ki/ : /kɨi/, /ti/ : /tɨi/, /si/ : /sɨi/, etc. in MK so that it is by no means difficult to postulate the distinction between /ni/ : /*nɨi/ at an earlier stage of this language.

Lastly, I would like to point out one important thing concerning this hypothesis. Outside the Cheju island, initial /ni/ is preserved in one place in the Chŏlla province (Kwangju) as I pointed out above. If we have had such forms only in Cheju dialects, this might be taken a kind of regional peculiarity. But the existence of such a form in a place which is geographically separated from the Cheju dialects suggests that the distinction between /ni/ : /*nɨi/ at an earlier stage of this language was indeed the case.

Acknowledgements

I would like to thank Professor Mitsuaki Endo, as always, for inviting me to the 4th International Conference on Asian Geolinguistics.

References

Ahn, Pyong-Hi. 1957. Chungkan Tusiŏnhae e nathanan t ŭm kugaeŭmhwa e taehayŏ (Palatalization found in the reprinted edition of Tusiŏnhae). *Ilsŏk Yi Hisŭng sŏngsaeng songsu kinyŏm nonmunchip.* Seoul: Ilchogak.

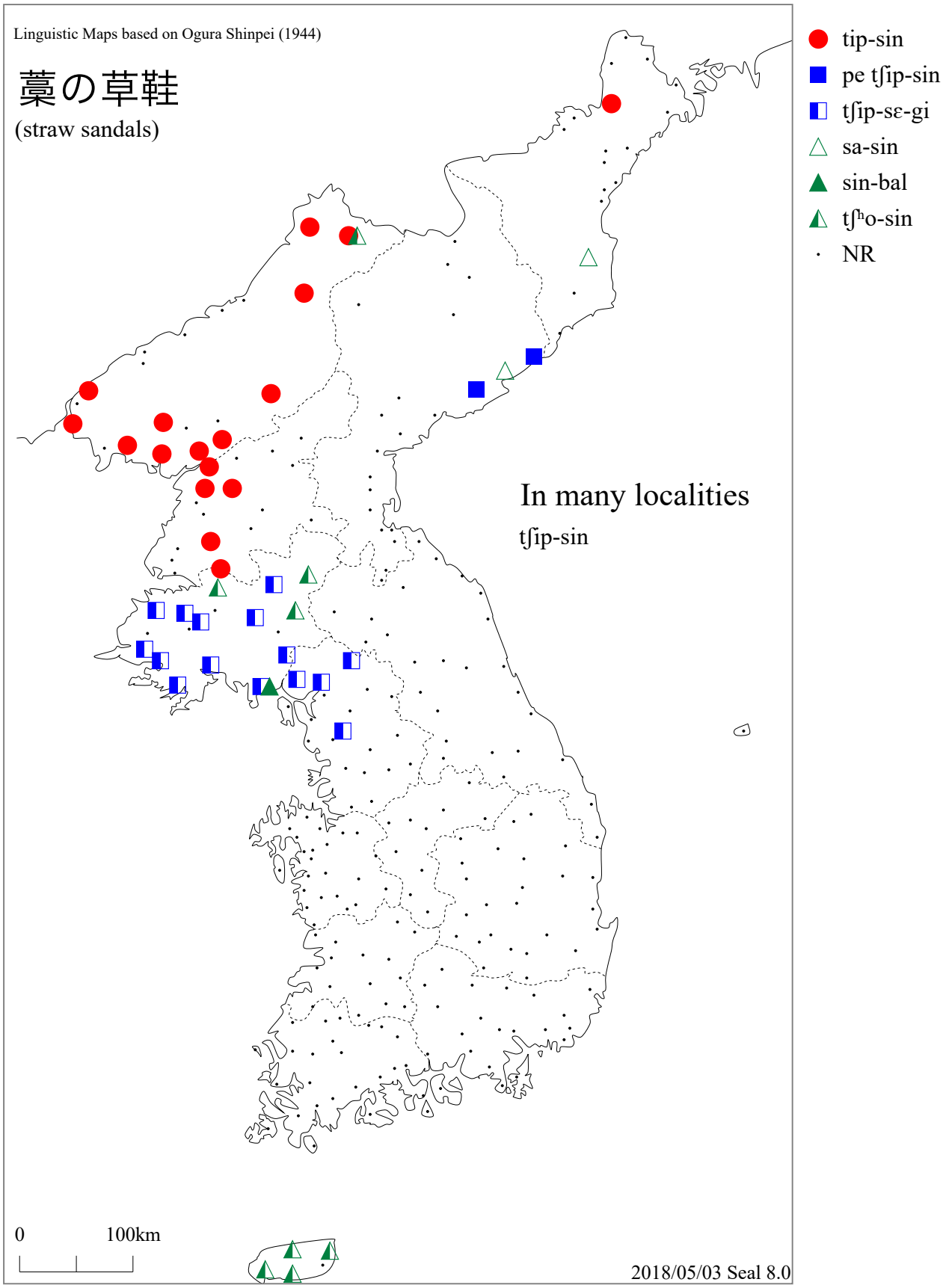
Hyŏn, P'yŏnghyo. 1985. *Chejudo pang'ŏn yŏngu* (A Study of the Cheju dialect). Seoul: Yiyu Ch'ulp'ansa.

Lee, Ki-Moon. 1991. *Kugŏ ŏhwi-sa yŏngu* (Studies on Korean vocabulary). Seoul: Tonga Press.

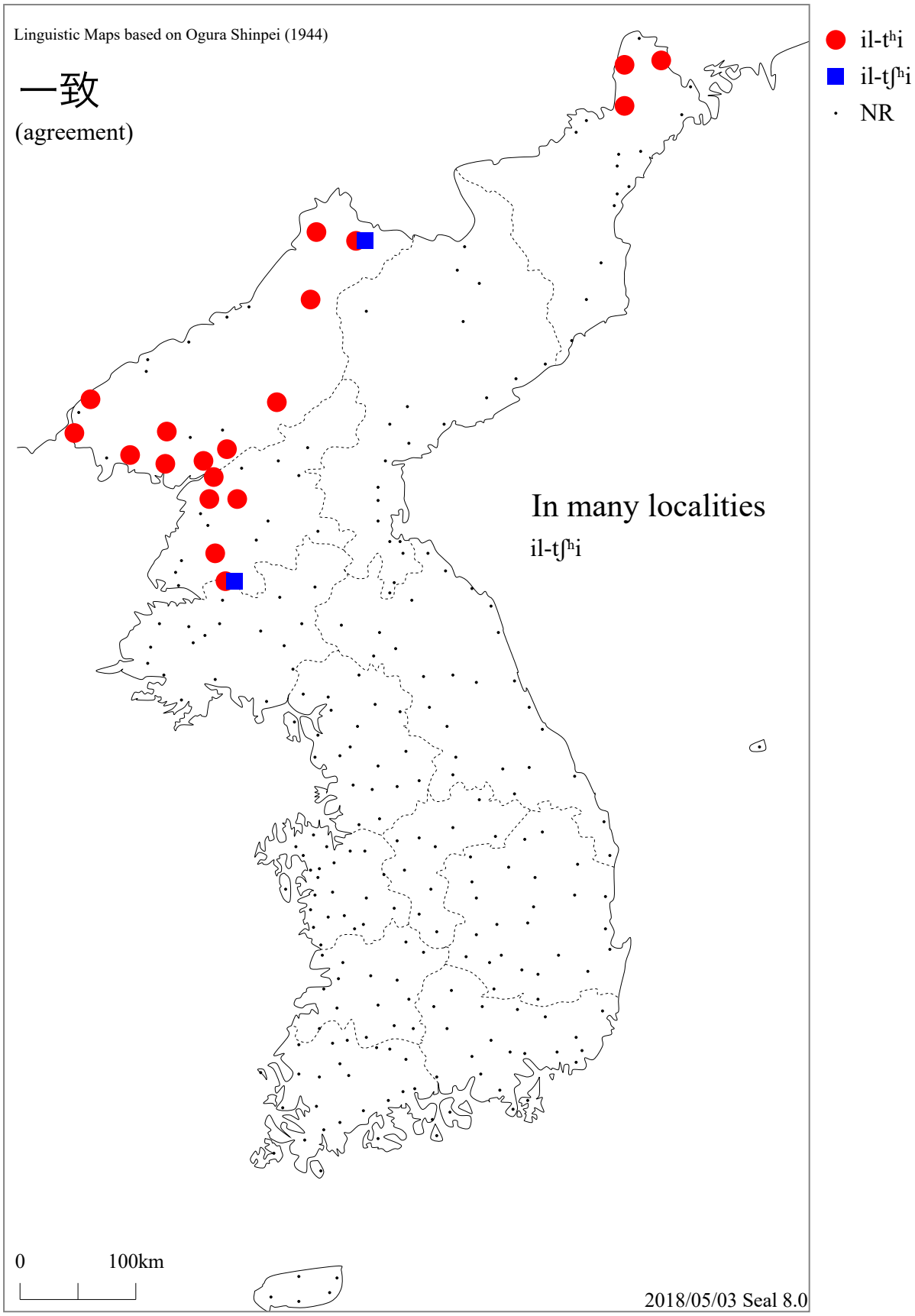
Lee, Ki-Moon. 1998. *Kugŏsa Kaeseol* (An introduction to the Korean language). Seoul: T'aehaksa.

Lee, Ki-Moon and S. Robert Ramsey. 2011. *A history of the Korean language.* Cambridge University Press.

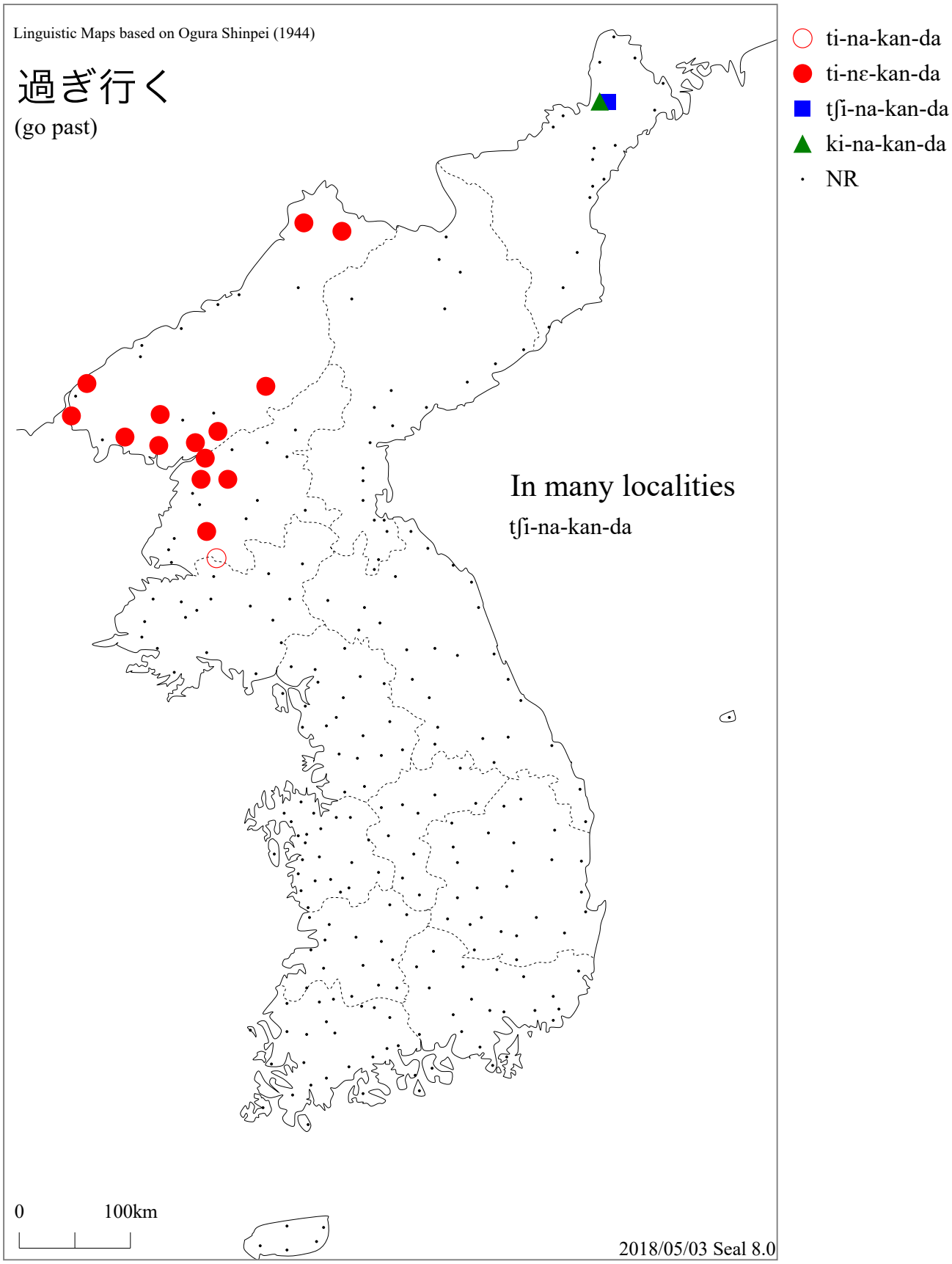
Ogura, Shinpei. 1944. *Chōsengo hogen no kenkyū* (Studies in the Korean dialects). 2 vols. Tokyo: Iwanami Shoten.


Linguistic Maps based on Ogura Shinpei (1944)
藁の草鞋
(straw sandals)
In many localities
tʃip-sin
tip-sin
pe tʃip-sin
tʃip-sɛ-gi
sa-sin
sin-bal
tʃʰo-sin
NR
0
100km
2018/05/03 Seal 8.0

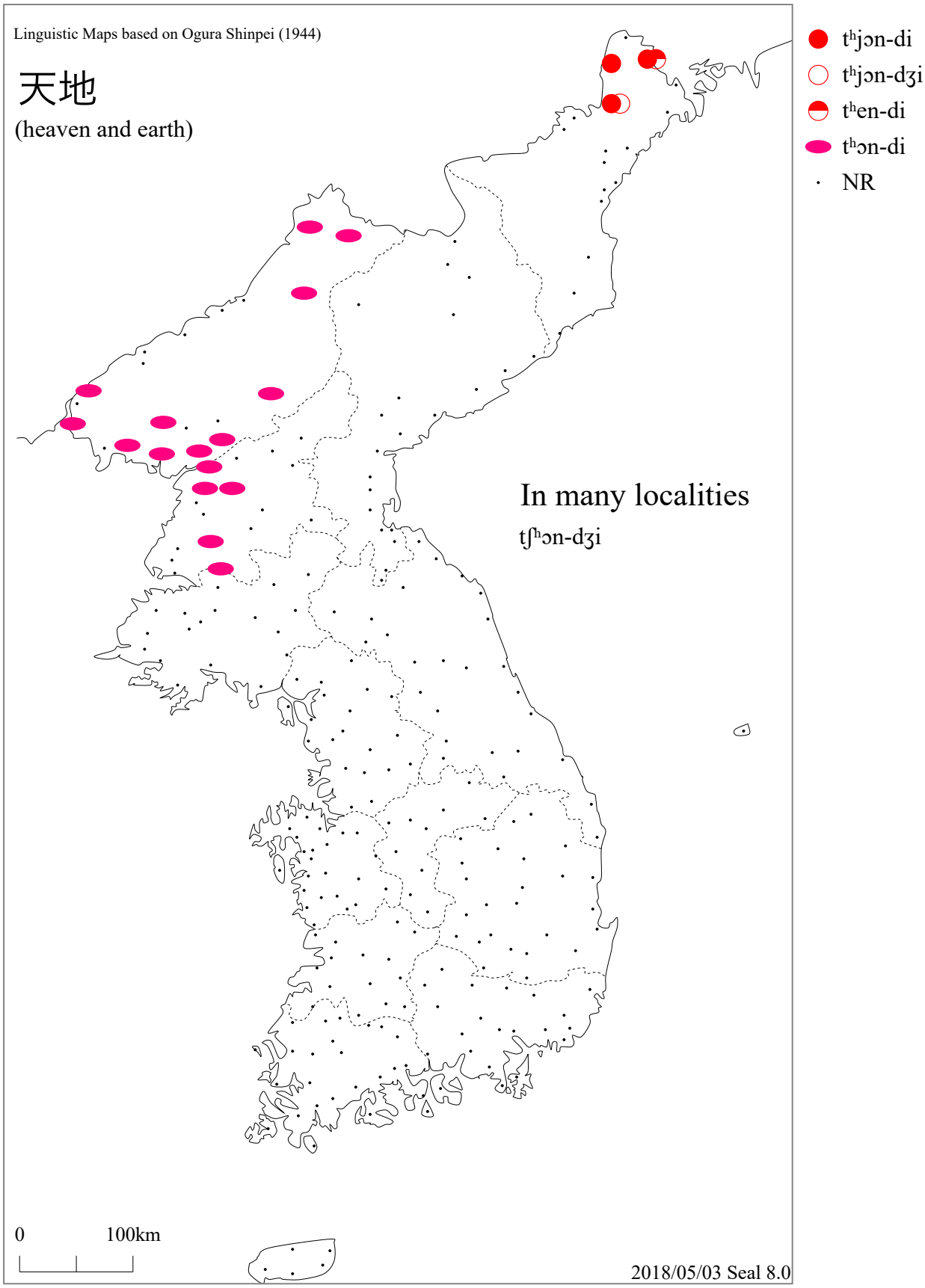
Map 1.


Linguistic Maps based on Ogura Shinpei (1944)
一致
(agreement)
il-tʰi
il-tʃʰi
NR
In many localities
il-tʃʰi
0
100km
2018/05/03 Seal 8.0

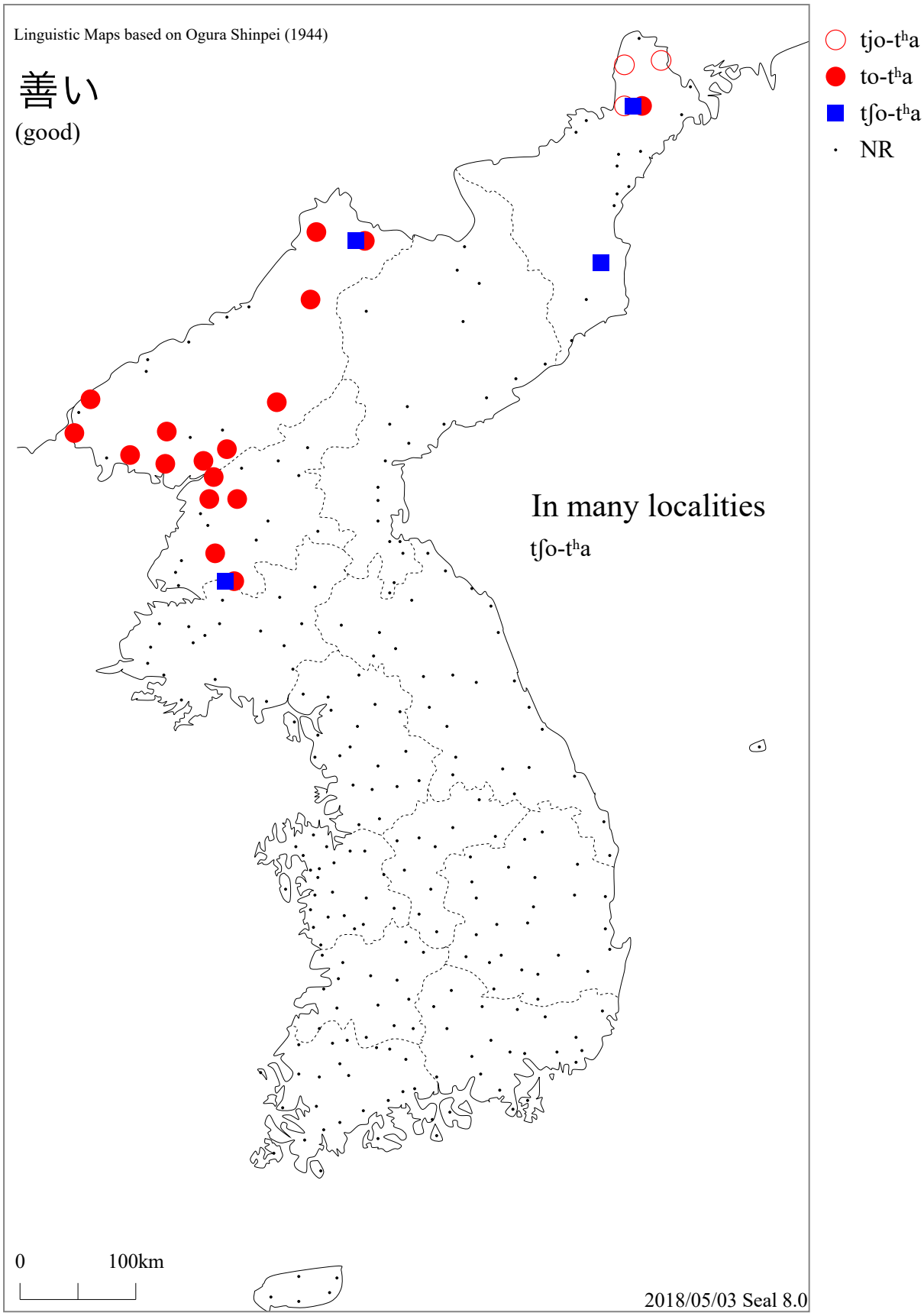
Map 2.


Linguistic Maps based on Ogura Shinpei (1944)
過ぎ行く
(go past)
In many localities
tʃi-na-kan-da
ti-na-kan-da
ti-nɛ-kan-da
tʃi-na-kan-da
ki-na-kan-da
NR
0
100km
2018/05/03 Seal 8.0

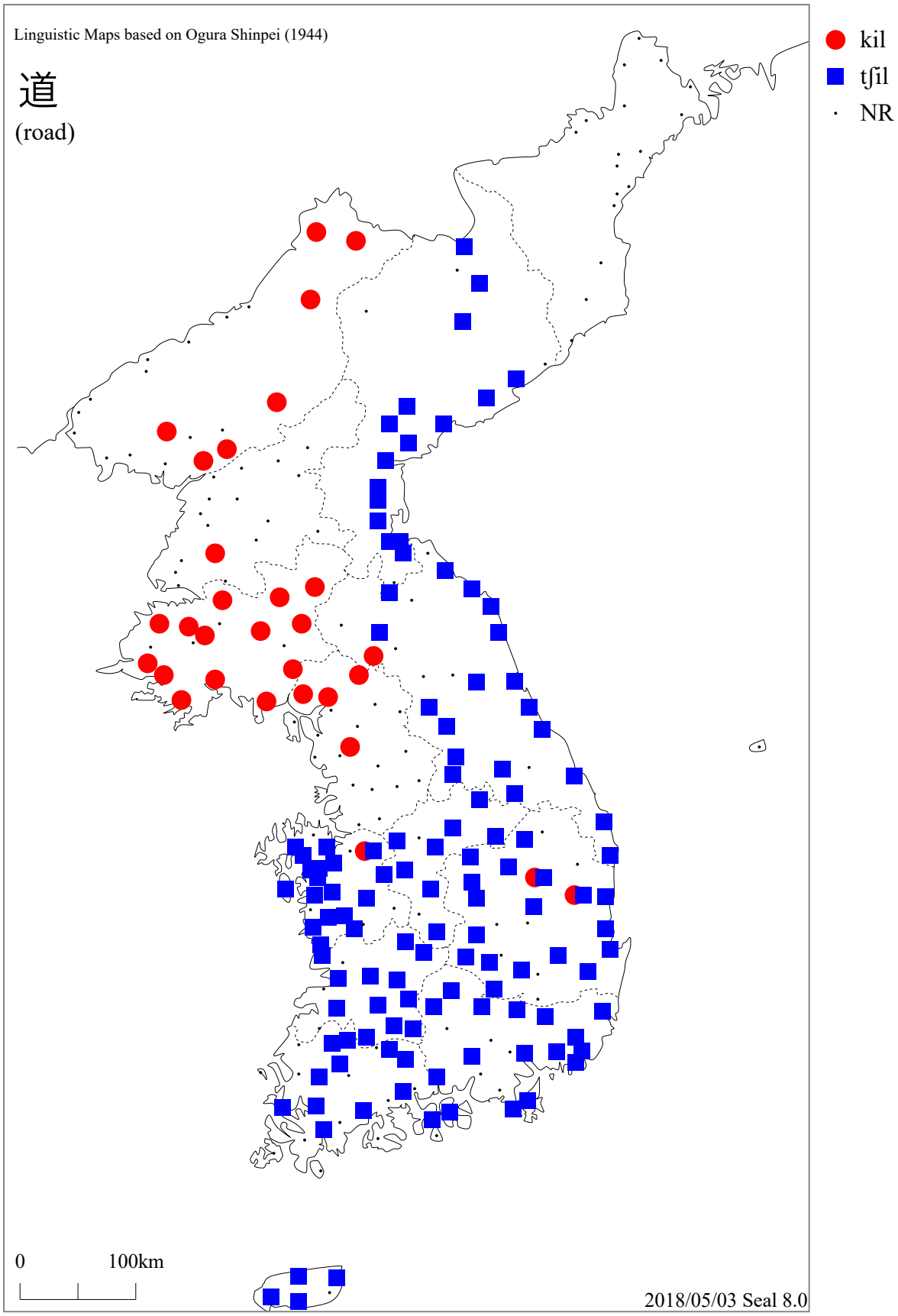
Map 3.


Linguistic Maps based on Ogura Shinpei (1944)
天地
(heaven and earth)
tʰjɔn-di
tʰjɔn-dʒi
tʰen-di
tʰɔn-di
NR
In many localities
tʃʰɔn-dʒi
0
100km
2018/05/03 Seal 8.0

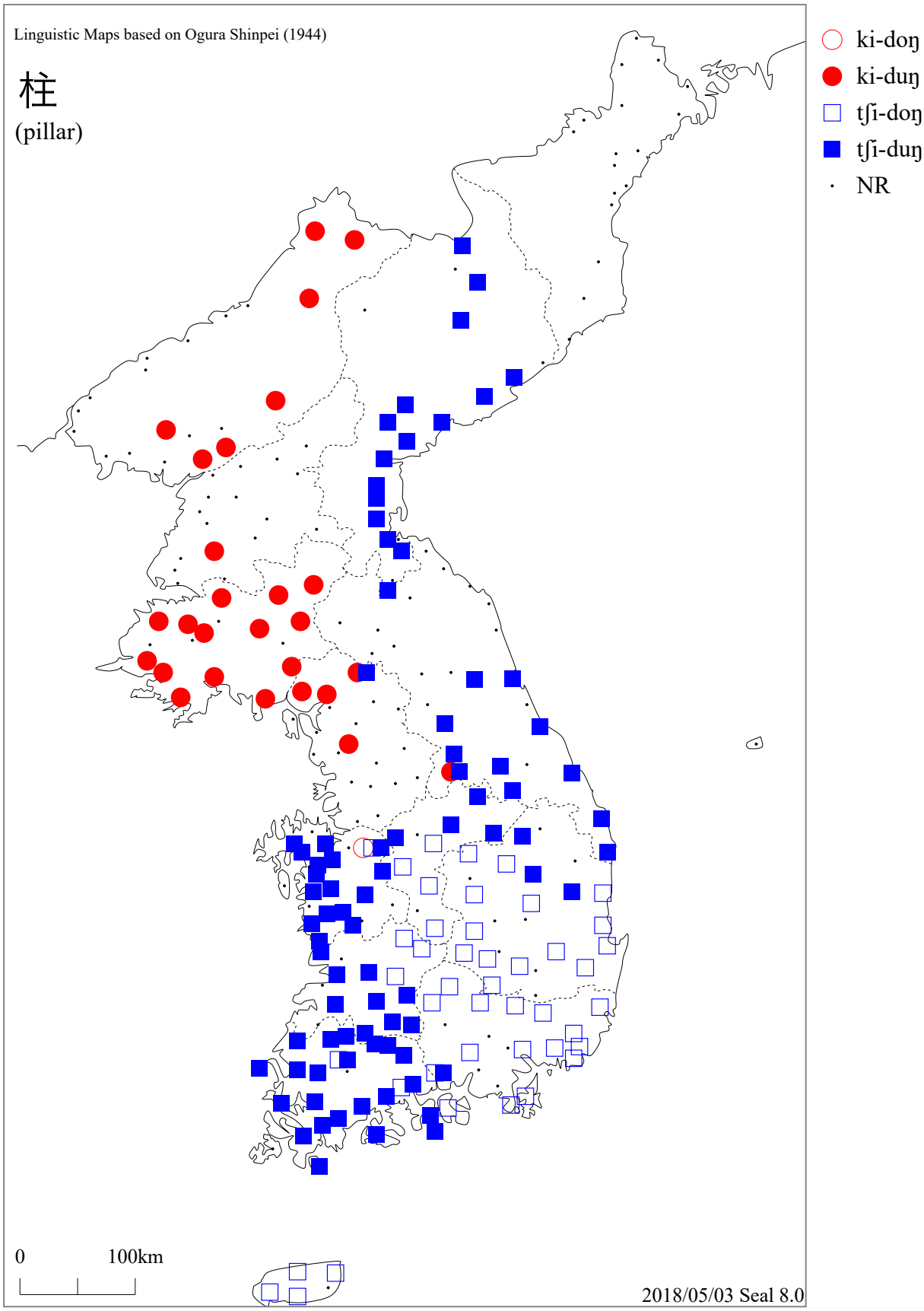
Map 4.


Linguistic Maps based on Ogura Shinpei (1944)
善い
(good)
tjo-tʰa
to-tʰa
tʃo-tʰa
NR
In many localities
tʃo-tʰa
0
100km
2018/05/03 Seal 8.0

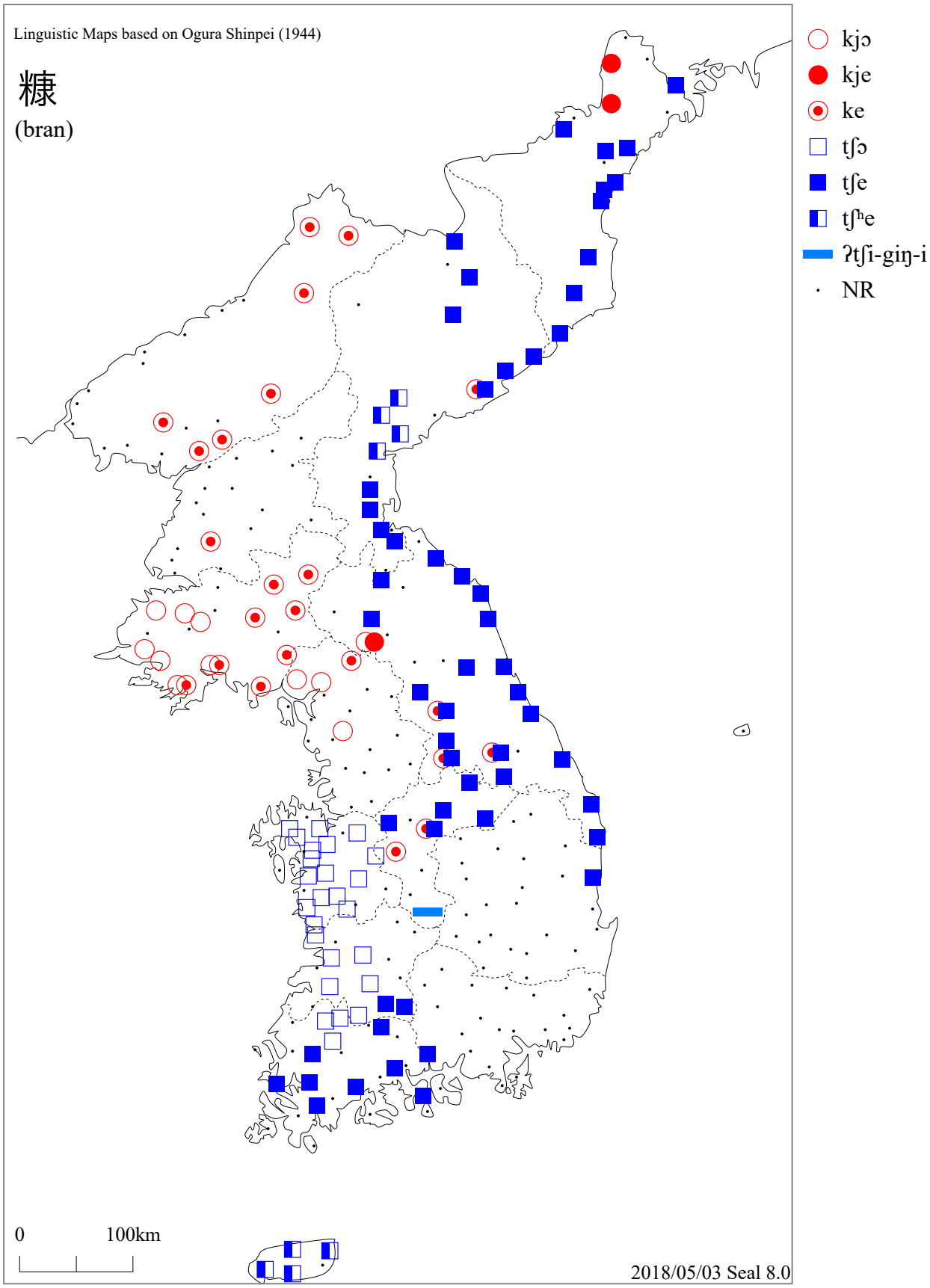
Map 5.


Linguistic Maps based on Ogura Shinpei (1944)
道
(road)
kil
tʃil
NR
0
100km
2018/05/03 Seal 8.0

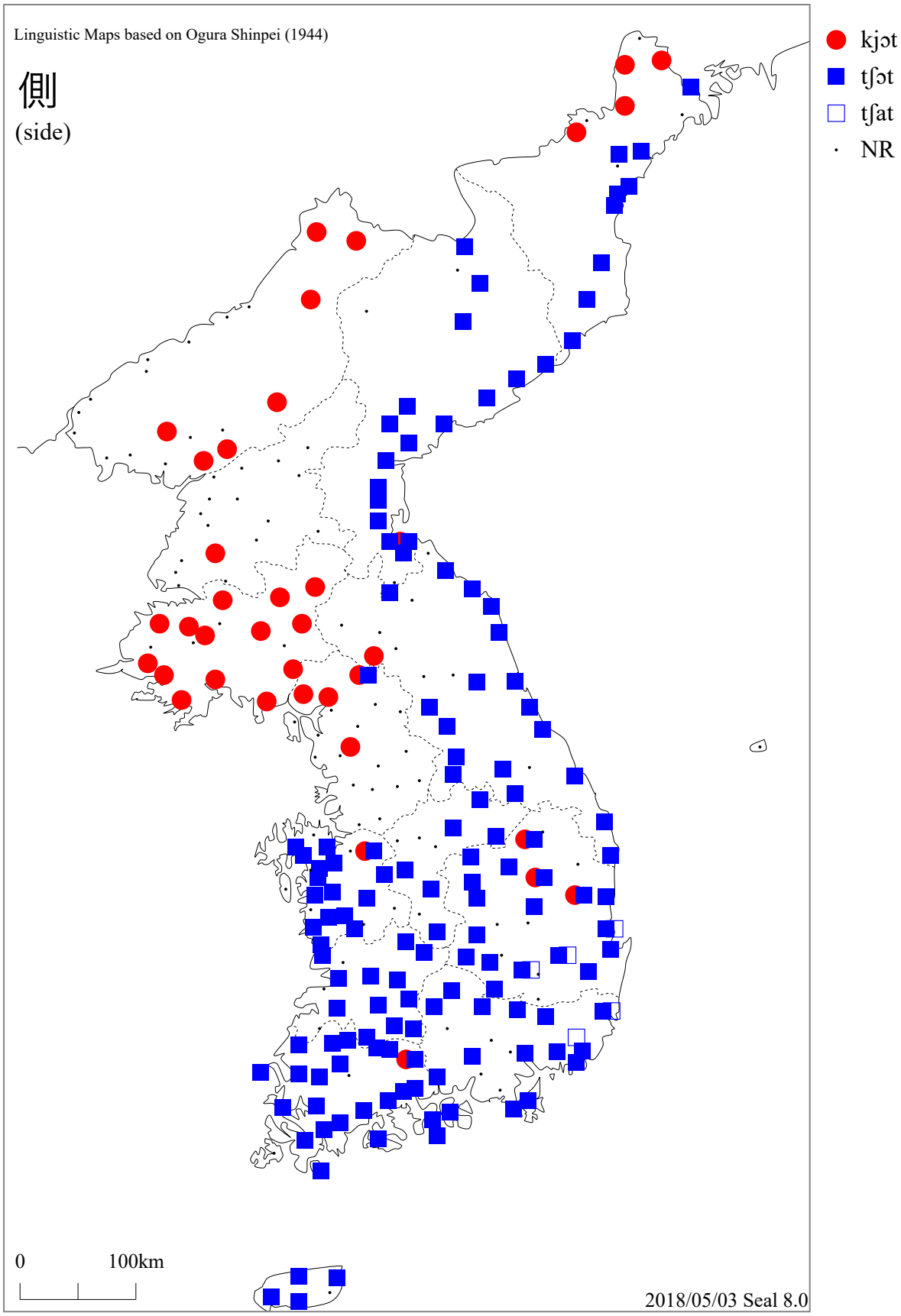
Map 6.


Linguistic Maps based on Ogura Shinpei (1944)
柱
(pillar)
ki-doŋ
ki-duŋ
tʃi-doŋ
tʃi-duŋ
NR
0
100km
2018/05/03 Seal 8.0

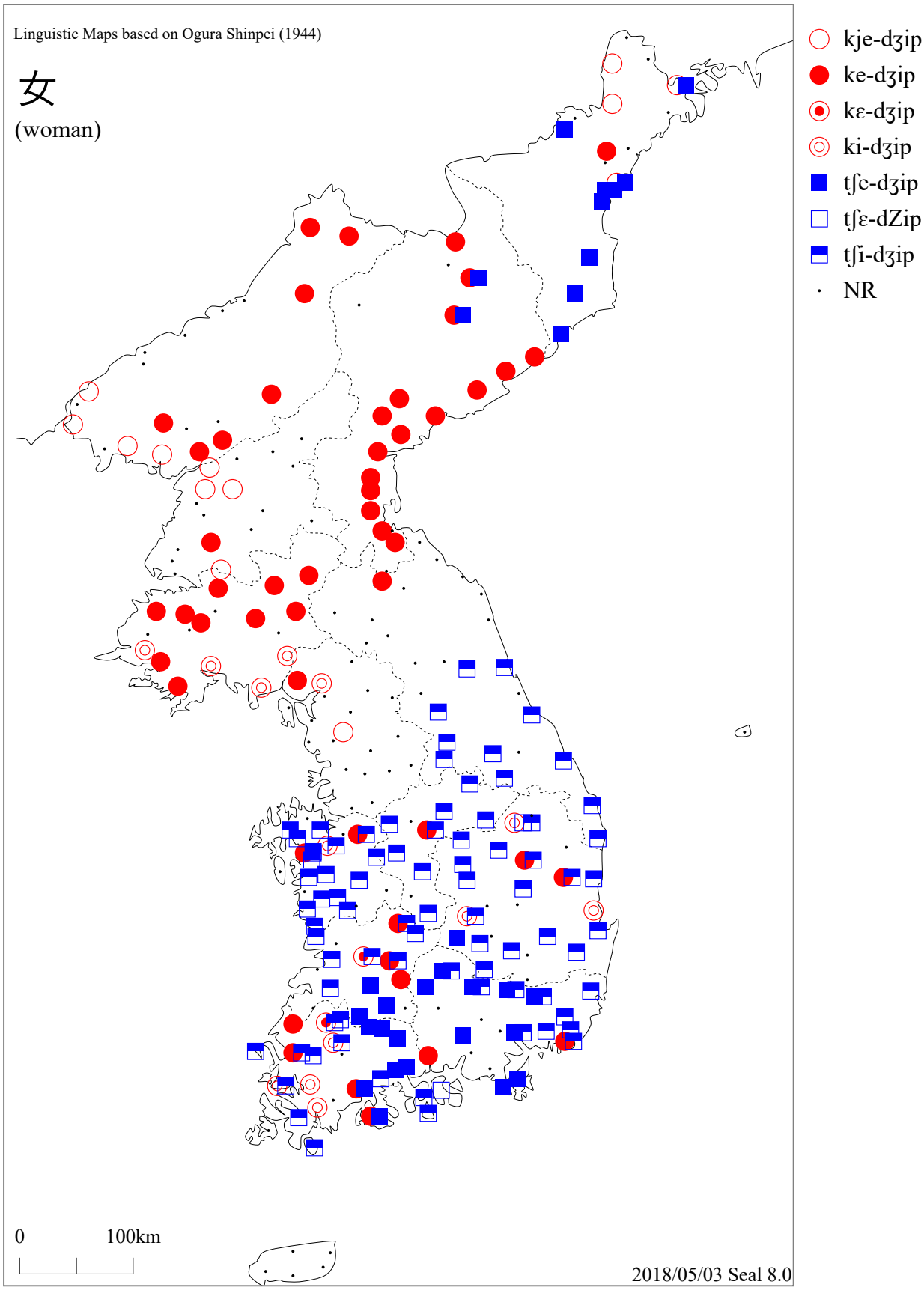
Map 7.


Linguistic Maps based on Ogura Shinpei (1944)
糠
(bran)
kjɔ
kje
ke
tʃɔ
tʃe
tʃʰe
ʔtʃi-giŋ-i
NR
0
100km
2018/05/03 Seal 8.0

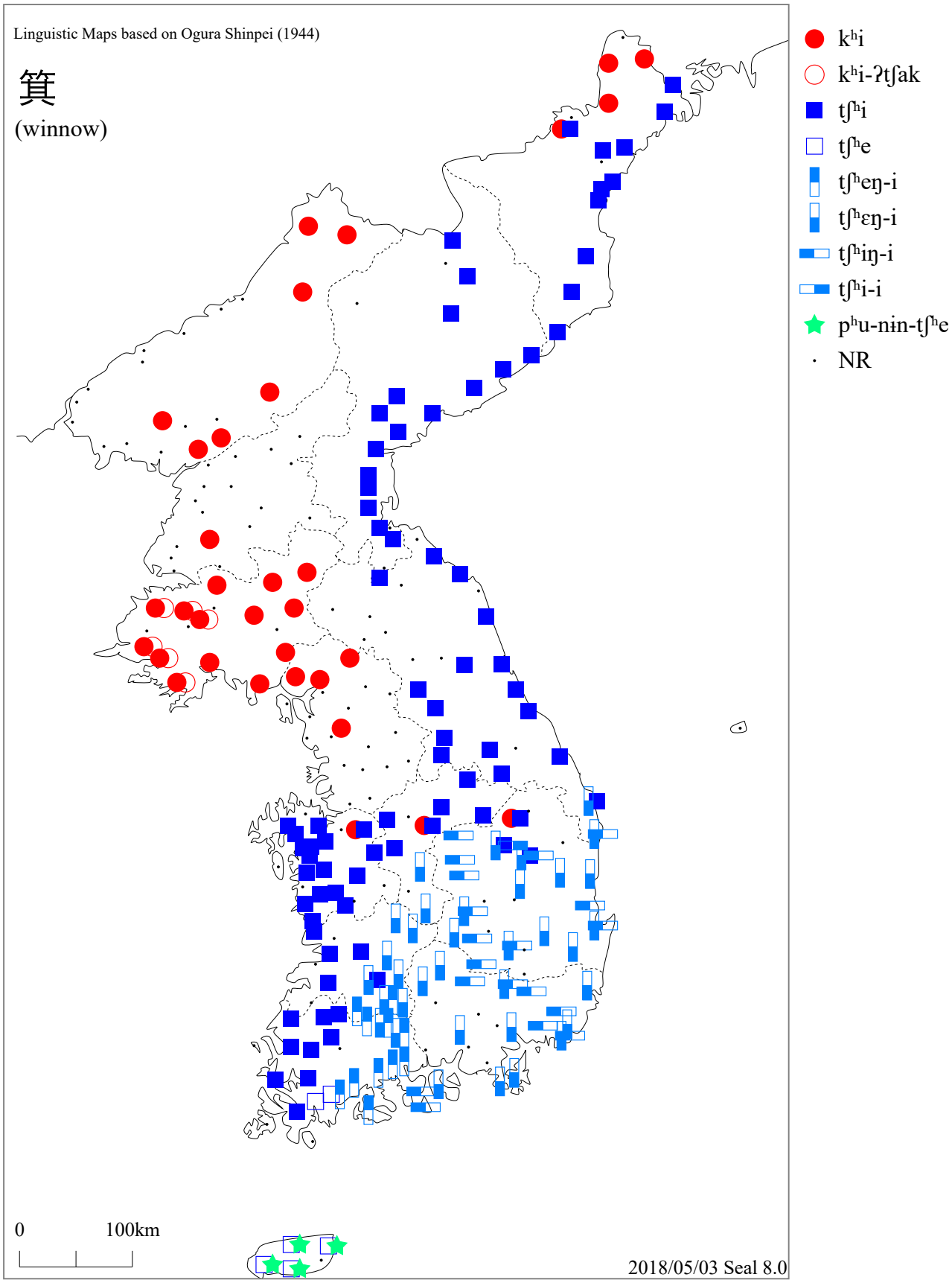
Map 8.


Linguistic Maps based on Ogura Shinpei (1944)
側
(side)
kjɔt
tʃɔt
tʃat
NR
0
100km
2018/05/03 Seal 8.0

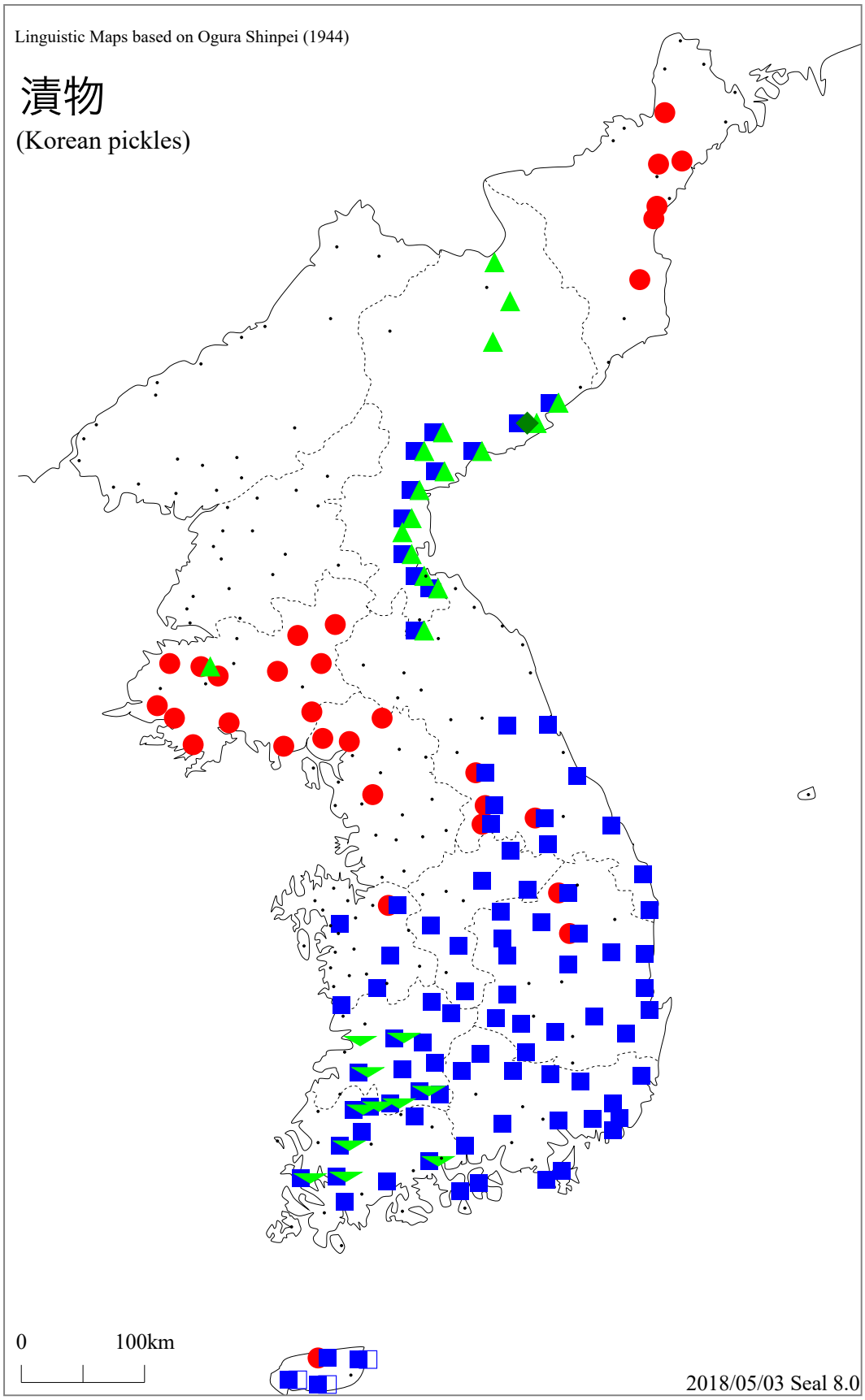
Map 9.


Linguistic Maps based on Ogura Shinpei (1944)
女
(woman)
kje-dʒip
ke-dʒip
kɛ-dʒip
ki-dʒip
tʃe-dʒip
tʃɛ-dZip
tʃi-dʒip
NR
0
100km
2018/05/03 Seal 8.0

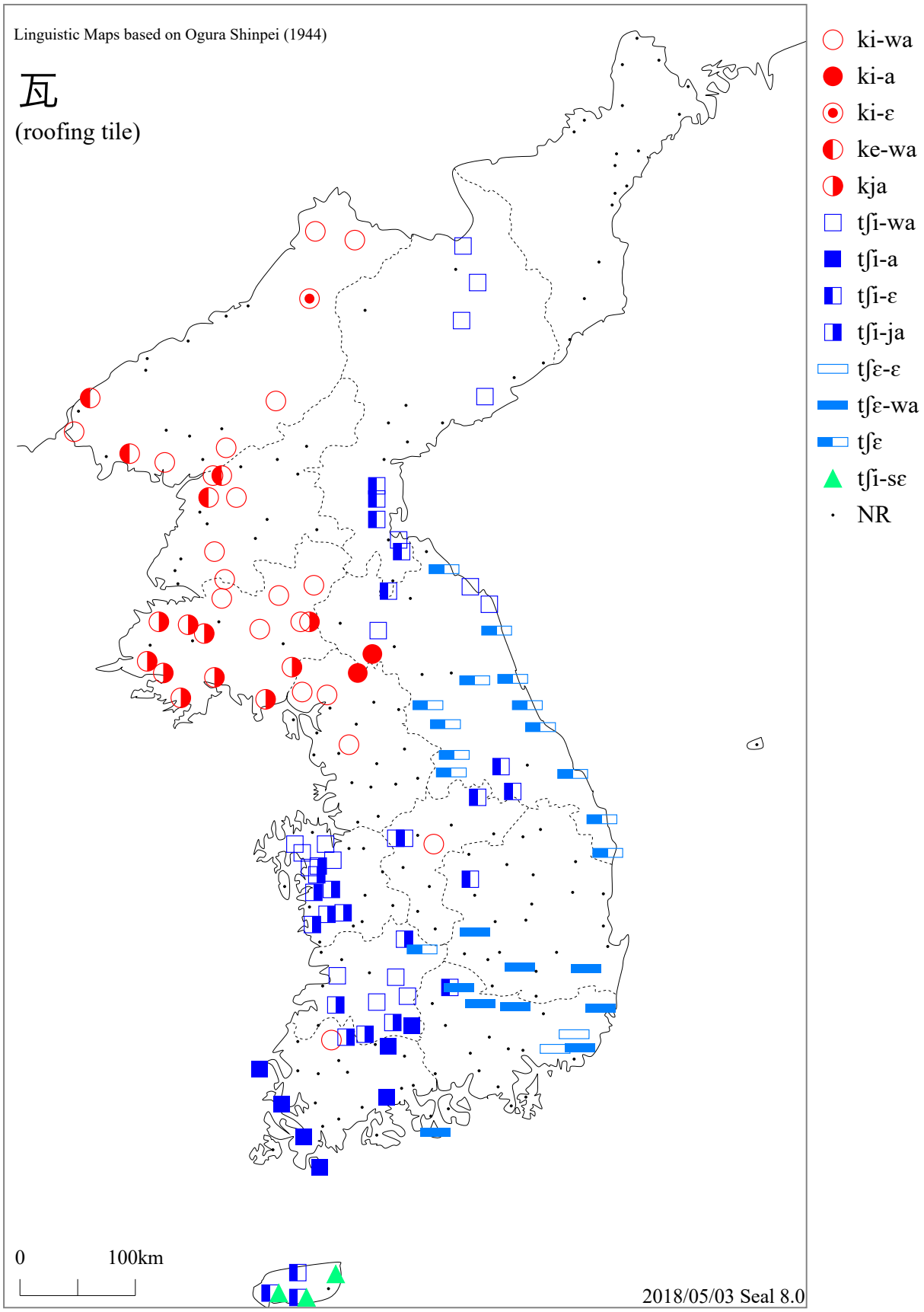
Map 10.


Linguistic Maps based on Ogura Shinpei (1944)
箕
(winnow)
kʰi
kʰi-ʔtʃak
tʃʰi
tʃʰe
tʃʰeŋ-i
tʃʰɛŋ-i
tʃʰiŋ-i
tʃʰi-i
pʰu-nɨn-tʃʰe
NR
0
100km
2018/05/03 Seal 8.0

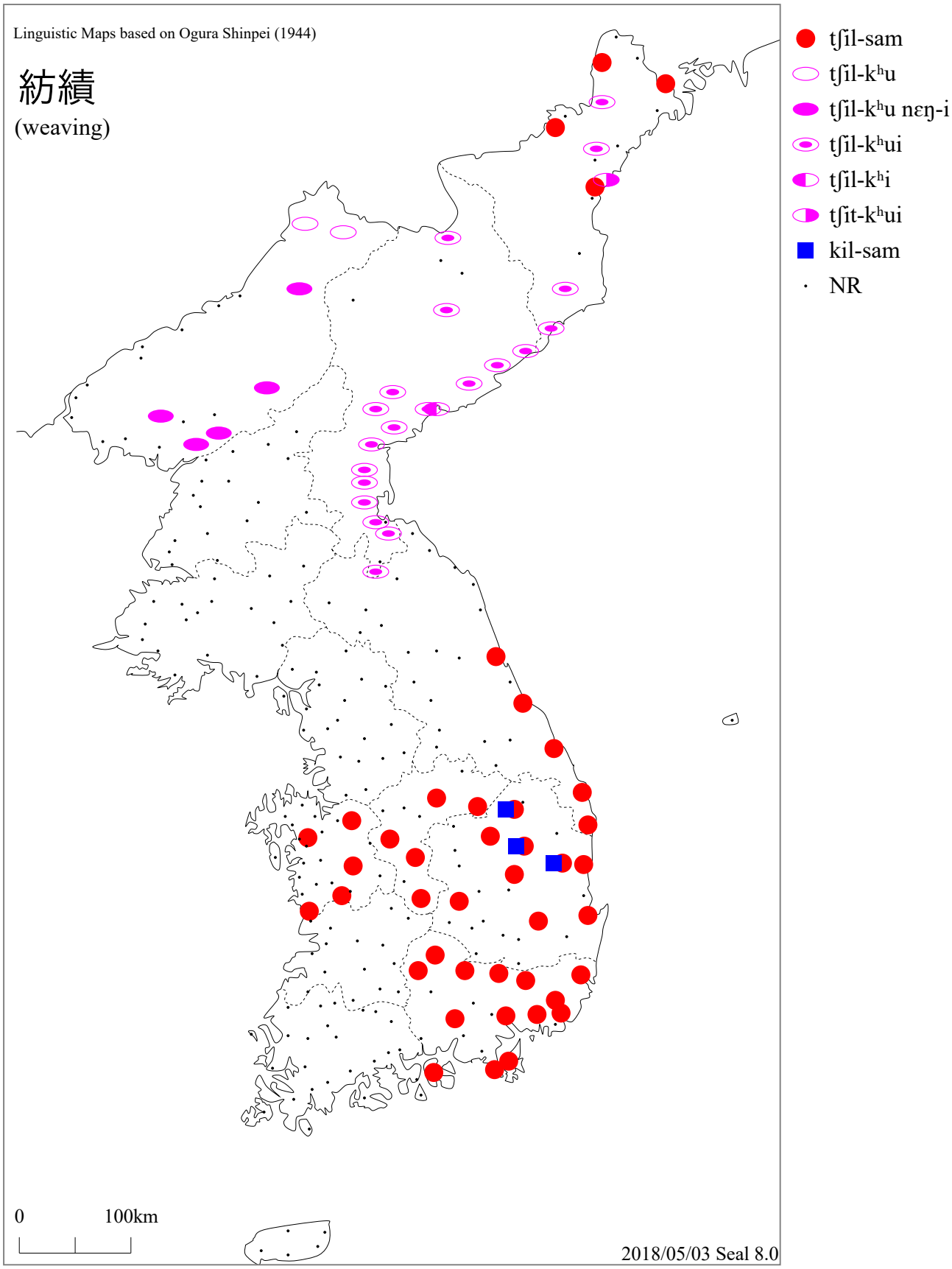
Map 11.


Linguistic Maps based on Ogura Shinpei (1944)
漬物
(Korean pickles)
kim-tʃʰi
tʃim-tʃʰi
tʃim-ʔki
tʃi
ʔtʃan-dʒi
ʔkak-tu-gi
NR
0
100km
2018/05/03 Seal 8.0

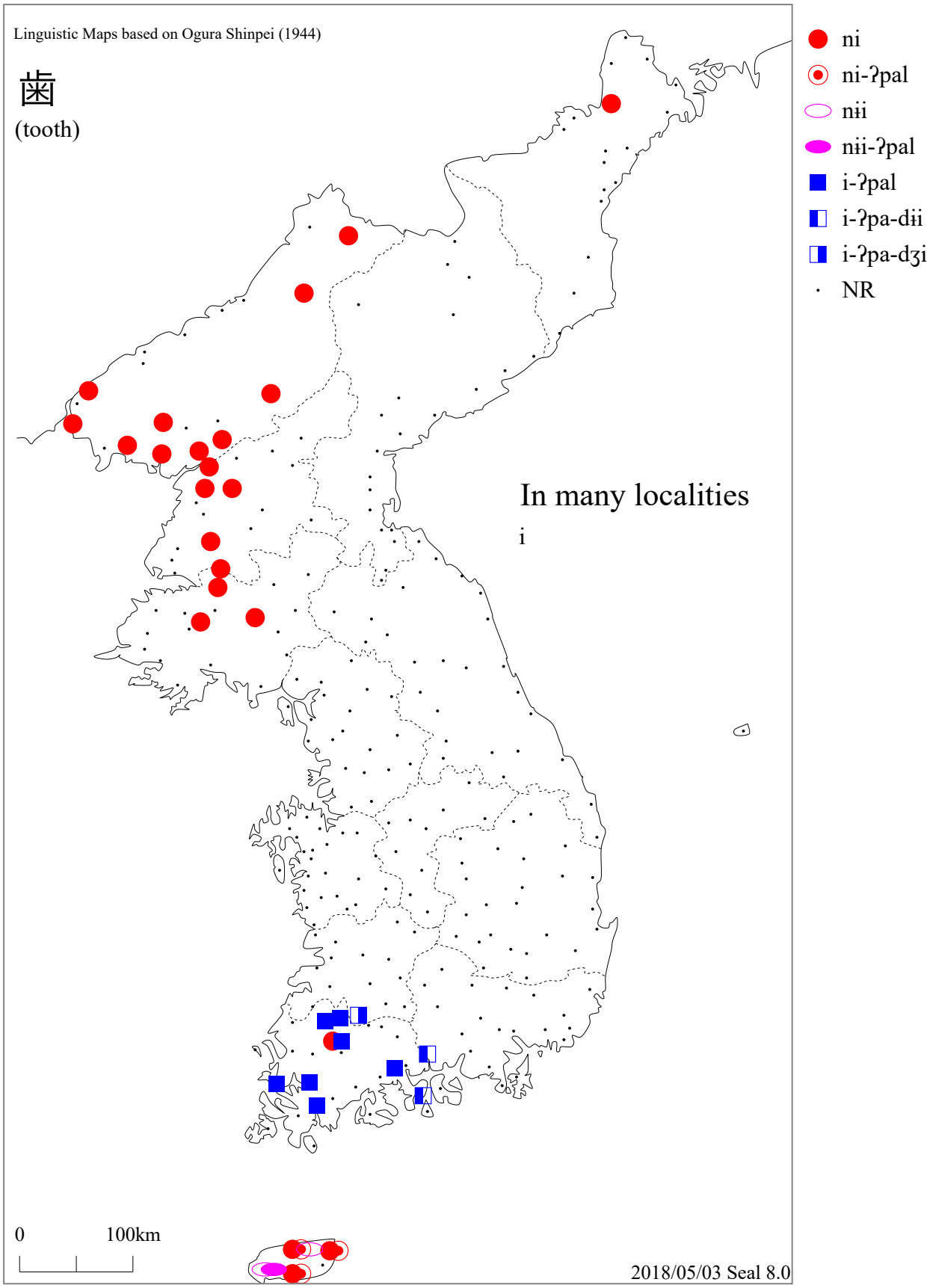
Map 12.


Linguistic Maps based on Ogura Shinpei (1944)
瓦
(roofing tile)
ki-wa
ki-a
ki-ɛ
ke-wa
kja
tʃi-wa
tʃi-a
tʃi-ɛ
tʃi-ja
tʃɛ-ɛ
tʃɛ-wa
tʃɛ
tʃi-sɛ
NR
0
100km
2018/05/03 Seal 8.0

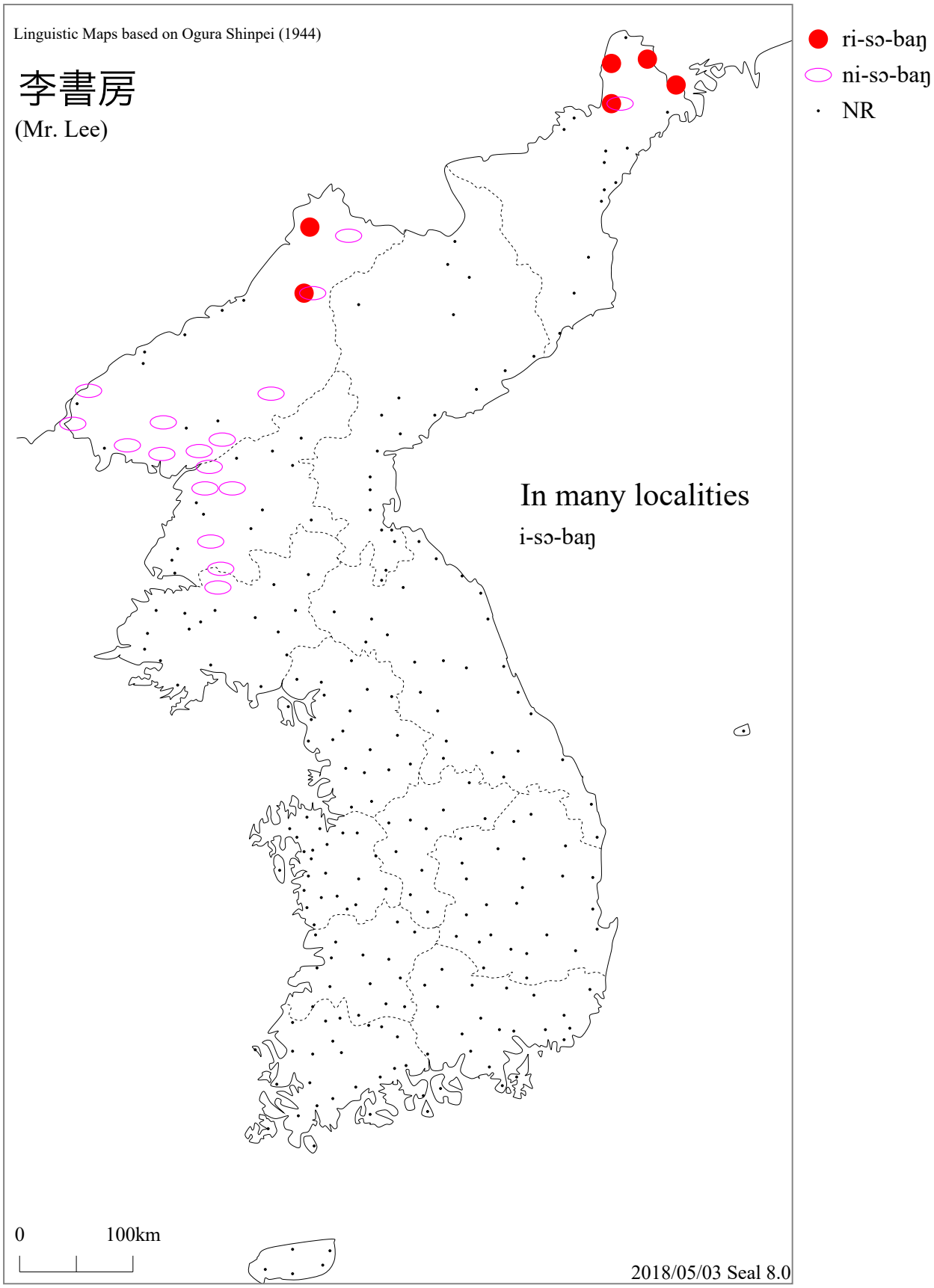
Map 13.


Linguistic Maps based on Ogura Shinpei (1944)
紡績
(weaving)
tʃil-sam
tʃil-kʰu
tʃil-kʰu nɛŋ-i
tʃil-kʰui
tʃil-kʰi
tʃit-kʰui
kil-sam
NR
0
100km
2018/05/03 Seal 8.0

Map 14.


Linguistic Maps based on Ogura Shinpei (1944)
歯
(tooth)
In many localities
i
ni
ni-ʔpal
nii
nii-ʔpal
i-ʔpal
i-ʔpa-dɨi
i-ʔpa-dʒi
NR
0
100km
2018/05/03 Seal 8.0

Map 15.


Linguistic Maps based on Ogura Shinpei (1944)
李書房
(Mr. Lee)
ri-sɔ-baŋ
ni-sɔ-baŋ
NR
In many localities
i-sɔ-baŋ
0
100km
2018/05/03 Seal 8.0

Map 16.

Correlation between onset and vowel, and the principle of "wider distribution" as revealed in the changing process of the forms for "rain" in Tai-Kadai

Mitsuaki Endo

Aoyama Gakuin University

1 Changing process of the onset

Almost all of the word forms for "rain" in Tai-Kadai are the cognate forms of the proto-Tai *fon A reconstructed by Li (1977). Aside from the forms kwən[1] in Mulao and pjən[1] in Sanjiang 三江 Dong, the changing process can be reconstructed as follows: f > ph (fortition); f > h > g or w > v (fortition). According to Li (1977: 63-65, 77-79), it is impossible to consider *ph* as the oldest form in this sound correspondence of f (Siamese): ph (Lunchow): f (Po-ai), since there is another series of sound correspondence of ph (Siamese): ph (Lunchow): ph (Po-ai), which should be reconstructed as *ph. Therefore, ph- for "rain" should be due to fortition, although such a process is rather rare. As for the geographical distribution, f- is spread throughout the area, especially in the Southwest.

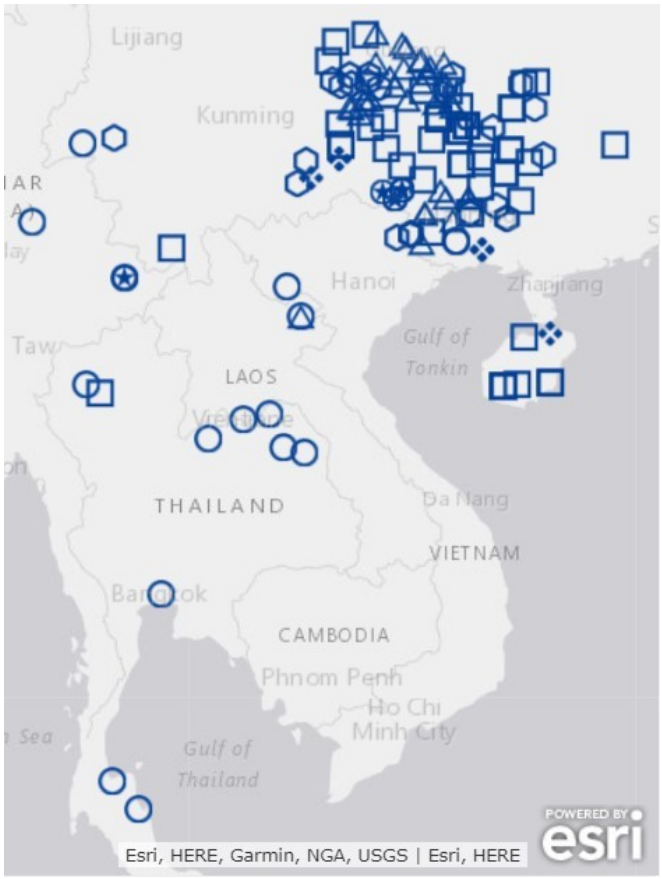
Map 1. Onset consonants of "rain" in Tai-Kadai

h- experienced delabialization from **f-*, which may be partially due to the dissimilation with the following vowel which changed from **o* to *u* in the Northern area. If lenition occurs, then *h-* may change into *g-* or *w-*.

Mainly in the Northernmost area, especially in Puyi, w- further changed into v-. Again, this is a process of fortition. It should be caused by the change of the following vowel into *ɯ* or *ə* as seen below.

Map 2. Enlarged map of the Northern area concerning the onset consonants of "rain" in Tai-Kadai

2 The changing process of the main vowel



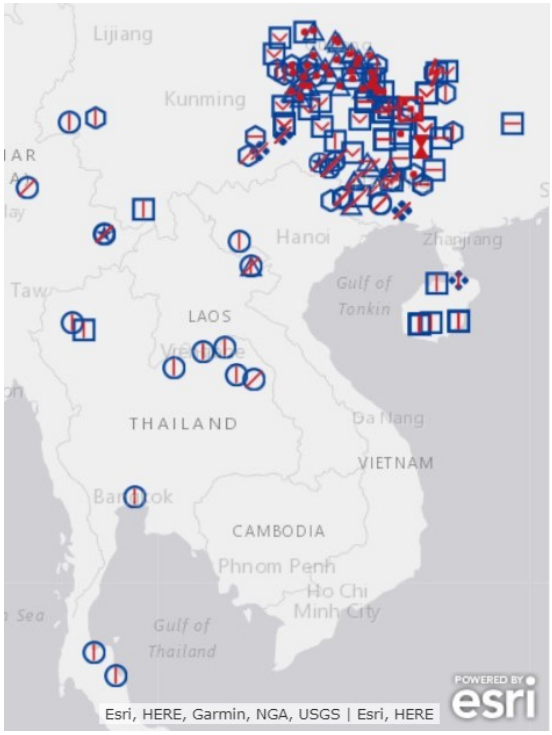
Li (1977: 272-273) reconstructed *o for the main vowel of "rain" in proto-Tai according to the sound correspondence of o (Siamese): ï [ɯ] (Lunchow): ï [ɯ] (Po-ai), since they are almost in complementary distribution with the sound correspondence of o (Siamese): u (Lunchow): ɔ (Po-ai) which was reconstructed as descending from the same *o. Li (1977: 272) says: "The final consonant of this set is dental, while the preceding set ends in a velar or a labial except in two cases where the final consonant is -n." It is impossible to consider that *u*, *ɯ*, *ə* are older than *o*, since there are the other sound correspondences for *u, *ɯ, and *ə respectively: u (Siamese): u (Lungchow): u (Po-ai) for *u; ï [ɯ] (Siamese): ï [ɯ] (Lunchow): ï [ɯ] (Po-ai) for *ï [ɯ]; a (Siamese): a (Lungchow): a (Po-ai) for *ə.

Map 3. Main vowels of "rain" in Tai-Kadai

Map 4. Enlarged map of the main vowels of "rain" in Tai-Kadai

Roughly from the South to the North, the narrowing changes o > u > ə > ɯ occurred, and fronting ɯ > i occurred in other places. A widening change o > ɔ > a occurred in the central area.

3 Correlation between onset and main vowel



Two maps concerning onset and main vowel are superimposed into one map, as seen in maps 5 and 6, respectively.

In maps 5 and 6, the correlation between onset and main vowel is easily observable. In the South, the combination f + o is dominant, and then f + u . In the Central area, f > ph and the lowering of the vowel occurred (ph + a , ph + ɔ). In the Northern area, there are many places where h + u , w + u are spreading. As pointed out above, it is possible that the cause of the changes from f to h or w is due to the dissimilation concerning the roundness of lips.

The combination of v + ɯ in the Northernmost area, especially in the Puyi language, is noteworthy. The combination v + ə is also found in the

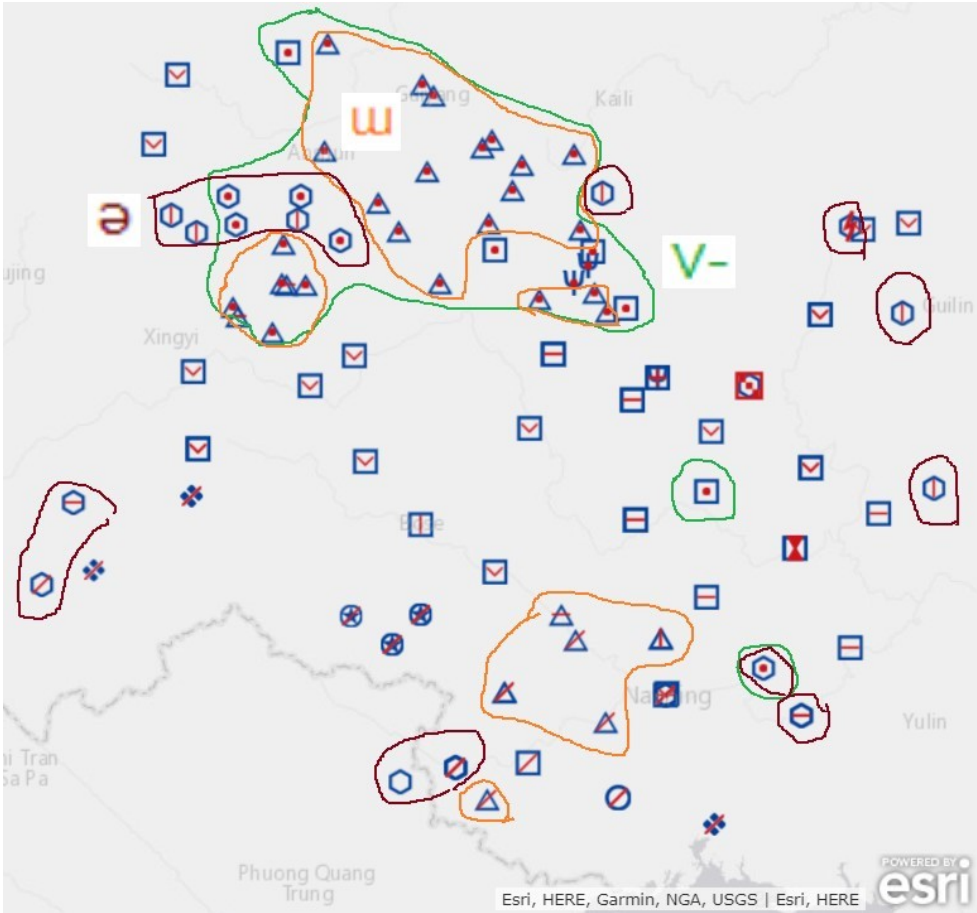
Map 5. Onset and vowel of "rain" in Tai-Kadai

Map 6. Enlarged map of onset and vowel of "rain" in Tai-Kadai

adjacent places. In these combinations, the main vowels *ɯ* and *ə* are unrounded, i.e. the edges of the lips are such that the upper teeth are likely to contact with the lower lip. Since this position of articulation is exactly that of [v], these combinations are very easily pronounced. Such a mechanism can be used for the fortition from *w* to *v*.

Map 7 shows the isoglosses of *v* (green), *ɯ* (orange) and *ə* (brown), The distributions of *ɯ* and *ə* are wider than *v*. That means, in these areas without overlapping *v*-, the changes from *u* to *ɯ* or *ə* occur without being preceded by consonant *v*-. However, there are 5 places in the peripheral areas where the combination v + u occurs as a combination v+ɯ . That means, these places accepted an onset *v*- from the neighbouring areas.

If this interpretation is right, a principle of wider geographical distribution can be postulated: Concerning two closely related linguistic phenomena, if one property is extended in a wider area, it will trigger another property.



Map 7. Isoglosses of *v* (green), *ɯ* (orange) and *ə* (brown) in the Northern area

Acknowledgements

This study was supported by ILCAA joint research project "Studies in Asian Geolinguistics", and JSPS Grant-in-Aid for Scientific Research (KAKENHI) 15K02525.

Reference

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

Variation and Change of Adjectives in Niigata Dialects

Chitsuko Fukushima

University of Niigata Prefecture

Abstract

The linguistic variation of adjectives in Niigata dialects, Japanese language is interpreted with the focus on phonological variation. Phonological, lexical and morphological variation among the young generation is described by comparing distributions in similar linguistic maps.

1 Introduction

Niigata dialects of Japanese language are located on the border of Eastern and Western dialects. Fukushima (2006) analysed the variation of adjectives with the focus on lexical variation. However, the adjectives in the dialects have peculiar phonological features. With the focus on phonological variation, the linguistic variation of adjectives in the area can be interpreted. In this presentation, first, phonological variation, and next, lexical and morphological variation among the young generation are introduced. By comparing distributions of similar linguistic maps, the changes happening in the dialects are described.

2 Adjectives in Niigata dialects

Watanabe (2001), a dialect dictionary of Northern Niigata dialects, includes dozens of entries of adjectives used in Furumachi, downtown of Niigata city. Adjectives often have a long vowel in the second last syllable:

(1) “Sweet”	Standard	*amai*	Niigata	*a:me*	{ai>e}	/a:Ce/	C=/m/

On the other hand, some adjectives have a short vowel in the second last syllable followed by double consonants:

(2) “Deep”	Standard	*fukai*	Niigata	*fukke*	{ai>e}	/uCCe/	C=/k/

See the list of other adjectives involved.

Table 1: List of adjectives used in Furumachi, Niigata city

“Ill-smelling”	Standard	*kusai*	Niigata	*kusse*	{ai>e}	/uCCe/	C=/s/
“Thick/Big”	Standard	*futoi*	Niigata	*futte:*	{oi>e}	/uCCe:/	C=/t/
“Thin”	Standard	*hosoi*	Niigata	*ho:se*	{oi>e}	/o:Ce/	C=/s/
“Dark”	Standard	*kurai*	Niigata	*ku:re*	{ai>e}	/u:Ce/	C=/r/
“Fast/Early”	Standard	*hajai*	Niigata	*ha:je*	{ai>e}	/a:Ce/	C=/j/
“Cold”	Standard	*samui*	Niigata	*sa:me*	{ui>e}	/a:Ce/	C=/m/

Whether the consonant in the last syllable is voiced or unvoiced seems to work on the choice of either pattern. See Figure 1, two pie graphs showing different patterns according to whether the consonant involved is voiced or not. If the consonant is voiced, five-sixths of adjectives show the -VVCV pattern

(e.g. *ku:re*). If it is unvoiced, more than half adjectives show either the -VCCV or -VCCVV pattern (e.g. *kusse* or *futte:*) but also has the -VVCV pattern (e.g. *ho:se*).

These patterns, -VVCV, -VCCV and -VCCVV, are distinguished with having "heavy syllables." There is another "heavy syllable" pattern, -VNCV such as *amme:* "Sweet", which is used in some Niigata dialects.

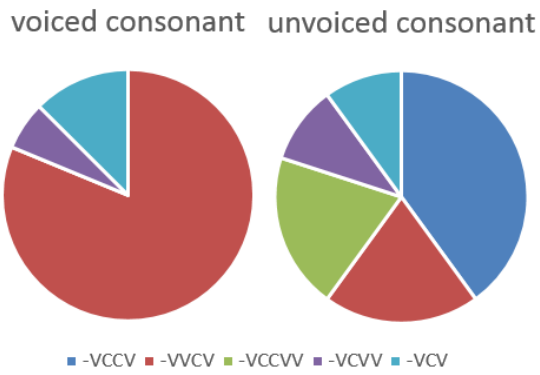


Figure 1 Phonological patterns of adjectives in Furumachi, Niigata (based on Watanabe 2001)

3 Data

The data used in this presentation are 1) the FPJD[1] data, those of the elderly surveyed in 2010-2015, and 2) the CS (College Students) data, those of the youth surveyed in 2005 and 2006. FPJD is a national survey, in which interviews were given at 21 locations in Niigata Prefecture. As for CS, 180 college students were asked to fill in a paper questionnaire on dialect during the linguistics class.

Linguistic maps are made using the GIS software SIS. College students' data are plotted at the location of a junior high school where they went.

4 Phonological variation of adjectives in Niigata

4.1 Adjective "Red": *akai*

According to Figure 2A and 2B, CS maps of "Red", two groups of wordforms are found. Group 1 includes a standard form, *akai*, and its spoken form, *ake:* {ai>e} (Figure 2A). Group 2 includes -VCCV or -VCCVV forms, *akke*, *akkai* and *akkoi*, and a -VVCV form *a:ke* (Figure 2B). Figure 2B shows typical dialectal forms still maintained by the young generation, and those words show interesting distributions. The form *akke* is widely distributed, and the form *a:ke* is found in the center of the *akke* distribution. Because of the distributions, the form *a:ke* is interpreted to be a form newer than *akke*.

4.2 Adjective "Sweet": *amai*

According to Figure 3A and 3B, CS maps of "Sweet", two groups of wordforms are also found. Group 1 includes a standard form, *amai*, and its spoken form, *ame:* {ai>e} (Figure 3A). The form *amage* is a related dialectal form. Group 2 includes -VNCVV forms, *ammai* and *amme:*, and -VVCV forms, *a:me* and *a:ma* (Figure 3B). Figure 3B shows typical dialectal forms maintained by the youth. While the form *amme:* shows a wide and thick distribution compared with *akke* (in Figure 2B), the form *a:me* has a larger distribution than *a:ke* (in Figure 2B). Again the form *a:me* is interpreted as a form newer than *amme*.

4.3 Adjective "Cold": *samui*

According to Figure 4A and 4B, CS maps of "Cold", two groups of wordforms are also found. Group 1 includes a standard form, *samui,* and its spoken forms, *sabui*, *sami:, sabi: and same:* {ui>i/e} (Figure 4A). Group 2 includes -VNCV forms, *sammu*, *sammi, sambi* and *samme*, and -VVCV forms, *sa:mu*, *sa:mi*, *sa:me* and *sa:be* (Figure 4B). Figure 4B again shows typical dialectal forms maintained by the youth. The -VVCV form s*a:me* has a thick distribution occupying the central area of peripheral -VNCV forms. Again the form *sa:me* is interpreted as a newer form spreading its distributions.

[1] FPJD is a nation-wide geolinguistic survey, "Field-Research Project for Analyzing the Formation Process of Dialects. "

4.4 Discussion

The traditional Niigata dialects would have had three patterns of older forms, such as the -VCCV pattern for voiceless consonants, the -VNCV pattern for nasal consonants, and -VVCV for other voiced consonants; then the -VVCV pattern with a long vowel in the second last syllable widened its distribution even among the words with other consonants. Maybe the change occurred first in the -VNCV pattern and next in the -VCCV pattern considering the size and density of the new distribution. Niigata city must be the center of new changes although contemporary young generation in that area tend to use standard Japanese; Figure 1 shows the results of early innovation. Figure 2A, 2B, 3A, 3B, 4A and 4B show how and why the change occurred in this area.

5 Lexical and morphological variation of adjectives in Niigata

5.1 Adjective "Interesting": *omoshiroi*

5.1.1 FPJD map

The FPJD map of "Interesting" (Figure 5) shows contrastive distributions of Eastern dialect forms and Western dialect forms. The standard form *omosjiroi* and its spoken form *omosjire* are only found in the most southern part of Niigata and Sado Island. The Eastern dialect forms *omossje* and *omsssjai* are found in the northern part of Niigata and Sado Island. The Western dialect form *omosji:* is found in the southern part of Niigata. *Omossjai* has a peripheral distribution wider than *omossje*; thus *omossje* seems to have originated from *omossjai*. Also *omosji:* is considered to be advancing to the north into Niigata city.

5.1.2 CS maps

The CS maps of "Interesting" (Figure 6A, 6B, 6C) show three groups of words, the first two of which show distributions similar to Group 1 and 2 of "Red", "Sweet", and "Cold". That is, Group 1 is a standard form *omosjiroi* and its spoken form *omosjire*, and Group 2 is *omossje* and *omosjoi*. The form *omosjoi* which is not found in the FPJD map is distributed in the center of Niigata. The form *omossjai* which is found in the FPJD map is not found in the CS map. In addition to these, there are Group 3 words, which are *omosji:* and *omoroi*, both of which are Western dialect words. My interpretation is that *omosji:* has been used in the southern part of Niigata for some time, and that it has expanded the distribution even to Niigata city.

5.2 Adjective "Interesting + past tense[2]": *omoshirokatta*

5.2.1 CS maps

The CS maps of "Interesting + past" (Figure 7A, 7B, 7C) also show three groups of words. Group 1 is a standard form *omosjirokatta*, and Group 2 includes *omossjekatta, omossjakatta* and *omosjokatta*. Unlike Figure 6B, the form *omosjakkata* is distributed in the east of Niigata city. The form *omosjokatta* shows similar distributions as *omosjoi* in Figure 6B. Also there are Group 3 words, which are *omosjikatta* and *omorokatta*, both Western dialect words. It is interesting that *omosjikatta* shows wider distributions than Figure 5C especially in Niigata city so that this word might have been introduced as this conjugation form in this area. Also, the older form *omossjai* maintains its past form in two localities.

[2] Adjectives in Japanese conjugate as verbs do. Thus a past auxiliary -*ta* is attached to adjectives.

5.3 Adjective "Interesting + negative[3]": *omoshirokunai*

5.3.1 CS maps

The CS maps of "Interesting + negative" (Figure 8A, 8B, 8C) again show three groups of words. Group 1 includes a standard form *omosjirokunai and* its spoken form *omosjirokune*, and Group 2 includes *omossje(ku)nai/ne, omossjakunai/ne* and *omosjo(ku)nai*. Compared with Figure 6B, the use of Group 2 forms does not prevail, but there is more variety. There are forms with *-ku* for every possibility but **omossjane* is not found. Also, the form *omossjakunai* is distributed in the west of Niigata city, where *omosjoi* and *omossjokatta* are found in Figure 6B and 7B.

5.4 Discussion

The Group 2 forms in Figure 6B, 7B, and 8B are as follows:

A: *omossje, omossjekatta, omo(s)sje(ku)nai/ne*
*B: * , omossjakatta, omossjakunai* (There is neither **omossjai nor* **omossjane.*)
C: omo(s)sjoi, omo(s)sjokatta, omo(s)sjo(ku)nai/ne

A has a fixed distribution, B sporadic, and C emerging. Thus, B is the oldest[4], A comes next, and C is the newest.

The group 3 forms show different distributions according to the conjugation form. The change in morphology gradually occurs through conjugation forms of each word; that is one example of "lexical diffusion."

6 Conclusion

When we see only one linguistic map, it might be difficult to interpret the change happening in the area. However, it would be possible to do so by finding similar patterns in linguistic maps. When distributions are superimposed on each other, we cannot see them well. The present-day GIS software enables us to make separate maps just by choosing forms we want to plot on the map. This method will help us to interpret the change based on the variation.

Acknowledgements

This presentation 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 Grant-in-Aid for Scientific Research (A) 23242024 and (C) 16K02688.

References

Fukushima, C. 2006 Changing Dialects of the Young Generation in Niigata, Japan, with the Focus on Adjectives. In: Guido Oebel (Ed.) *Japanische Beiträge zu Kultur und Sprache: Studia Iaponica Wolfgango Viereck emerito oblata*. Studies in Asian Linguistics 68. Lincom GmbH: Munich. 125-139

Watanabe, F. 2001. *Niigataken hogen jiten : Kaetsu-hen. [Dialect Dictionary of Niigata Prefecture: Kaetsu Region.]* Nojima Shuppan, Sanjo, Niigata.

[3] A negative auxiliary *-nai/ne* is also attached to adjectives. The form *-ne* is a spoken, dialectal form.

[4] There is a possibility that *omossjai* was created from *omossje* due to hypercorrection or back-formation because it is well-known that a standard form *-ai* is often expressed as *-e* in Niigata dialects. I appreciate a Niigata native Mitsuo Endo's suggestion about this.

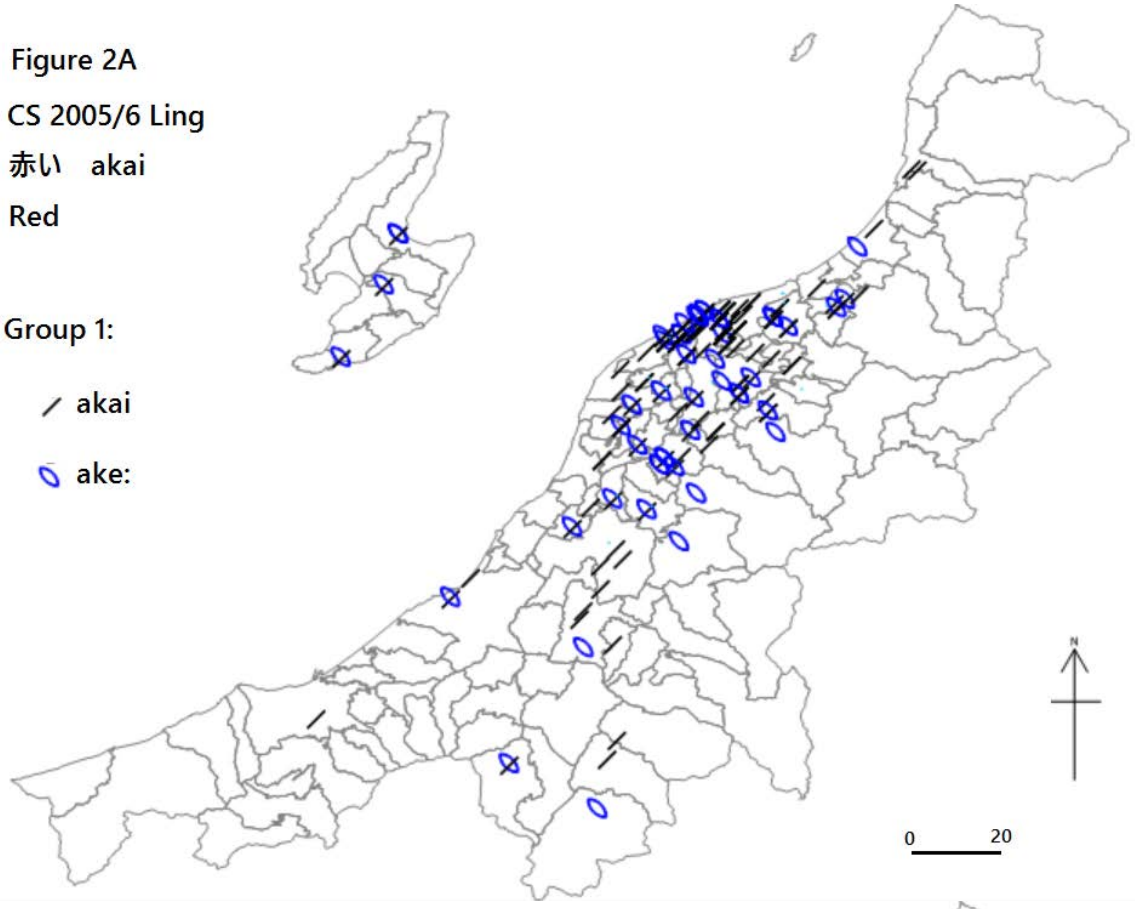


Figure 2A
CS 2005/6 Ling
赤い akai
Red
Group 1:
akai
ake:
N
0
20

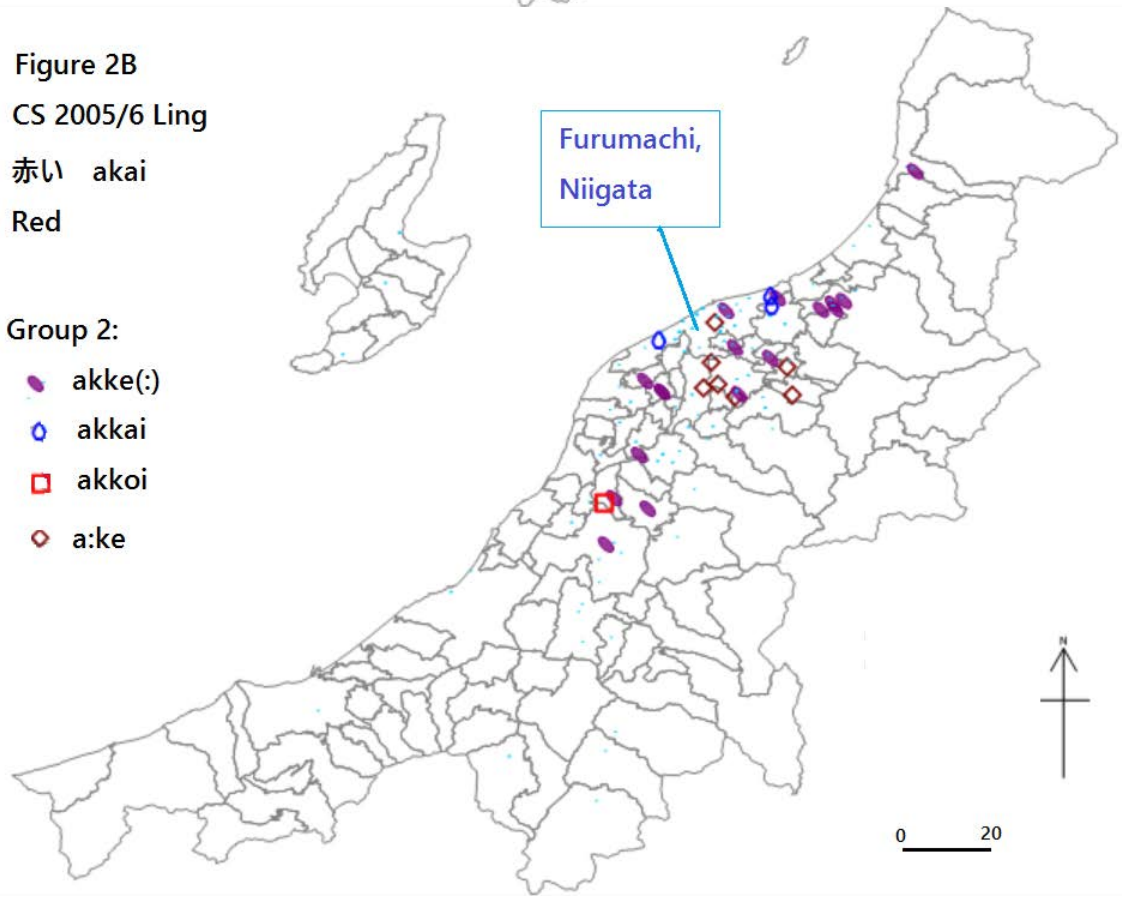


Figure 2B
CS 2005/6 Ling
赤い akai
Red
Furumachi,
Niigata
Group 2:
akke(:)
akkai
akkoi
a:ke
N
0
20

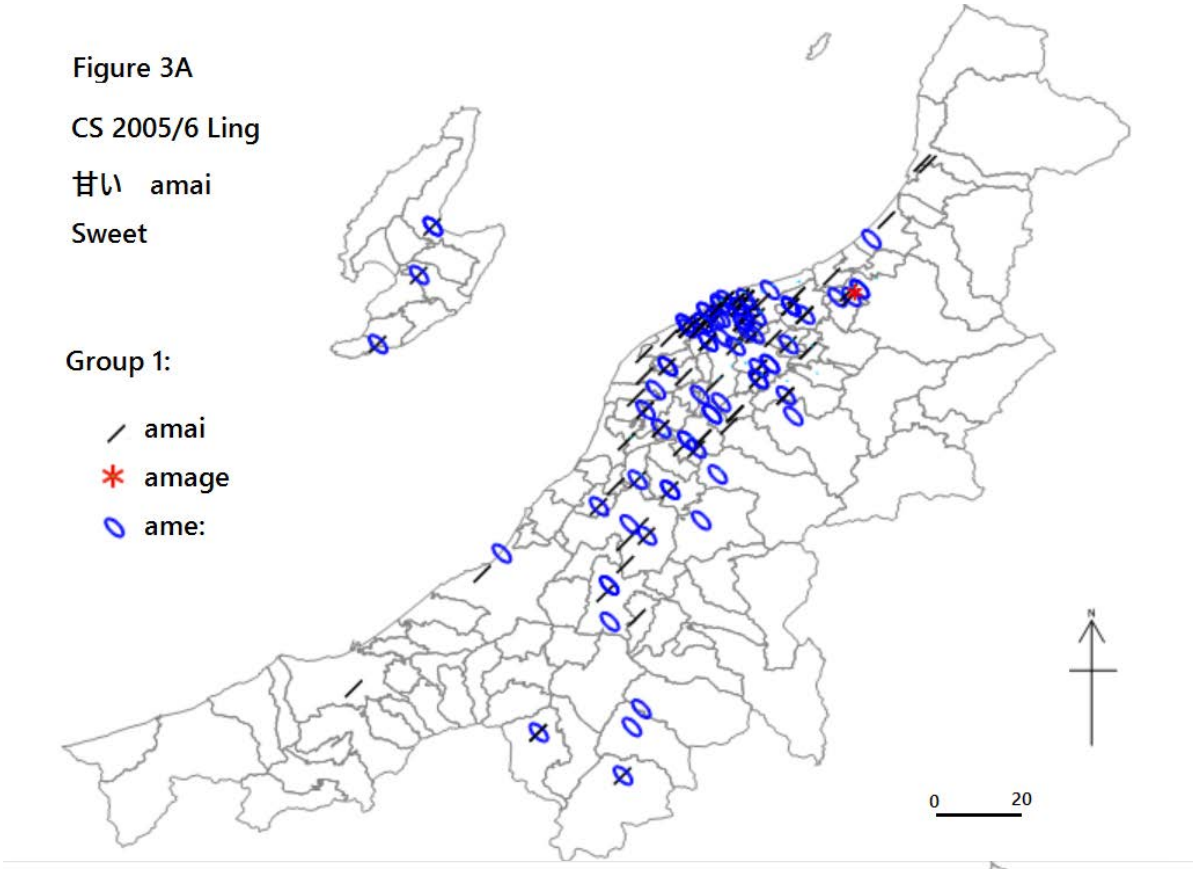


Figure 3A
CS 2005/6 Ling
甘い amai
Sweet
Group 1:
amai
amage
ame:
N
0
20

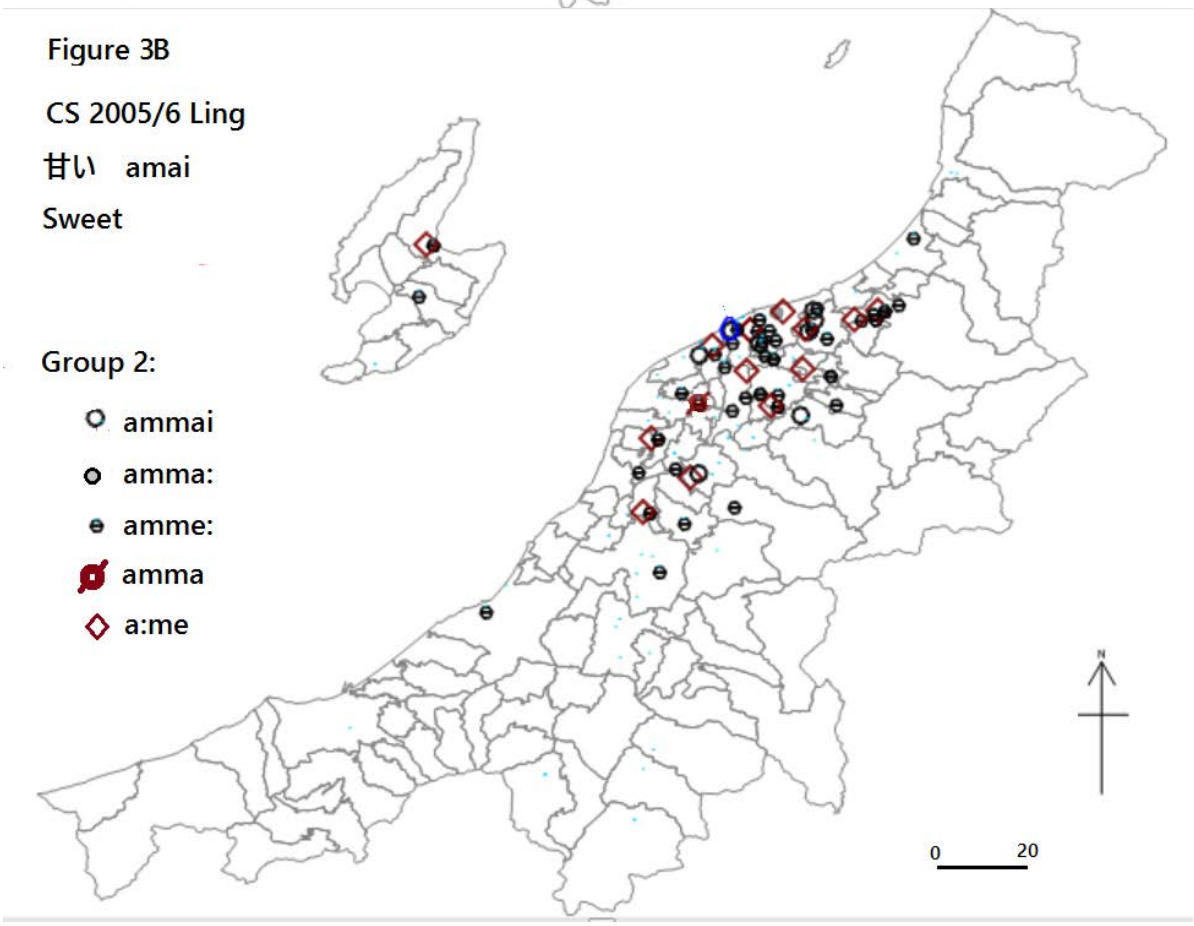


Figure 3B
CS 2005/6 Ling
甘い amai
Sweet
Group 2:
ammai
amma:
amme:
amma
a:me
N
0
20

Figure 4A

CS 2005/6 Ling

寒い　samui

Cold

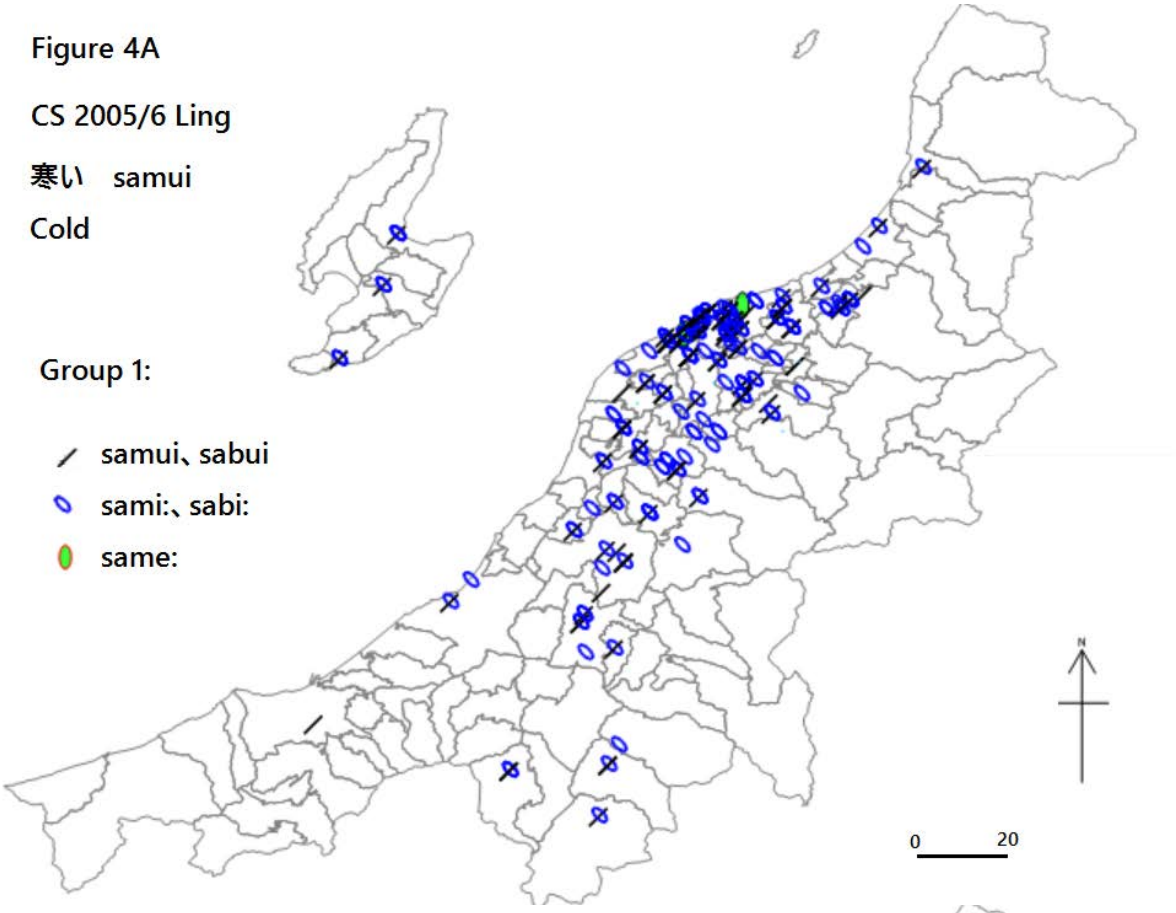


Figure 4B

CS 2005/6 Ling

寒い　samui

Cold

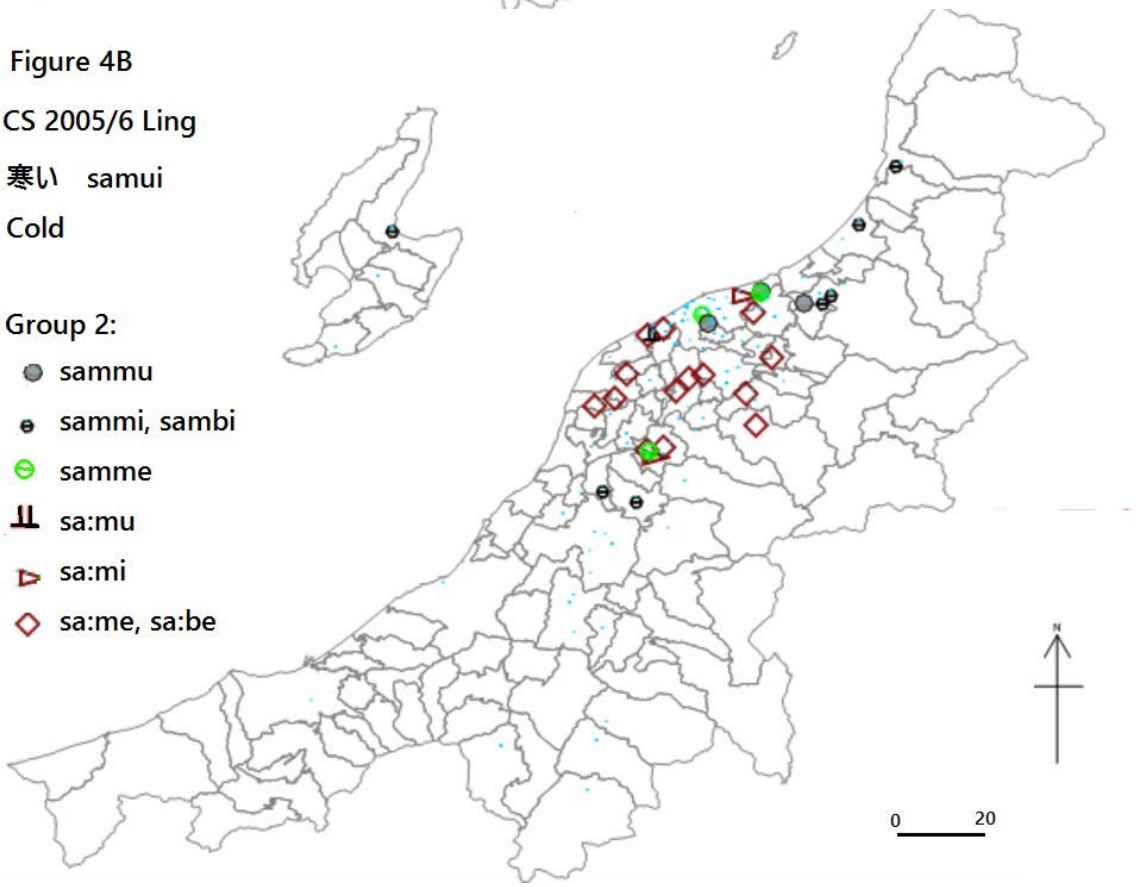


Figure 5

FPJD (2010-15 Elderly)

面白い　omoshiroi

Interesting

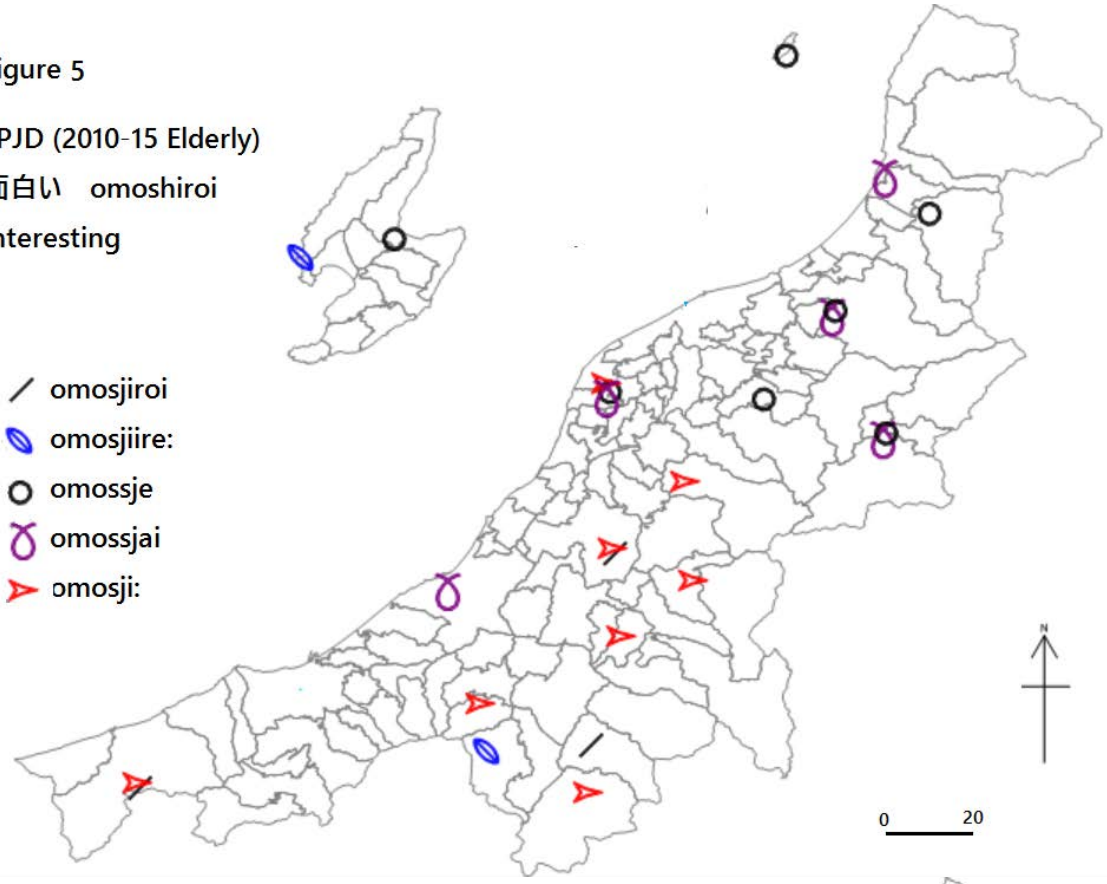


Figure 6 A

CS 2005/6 Ling

面白い　omoshiroi

Interesting

Group 1:

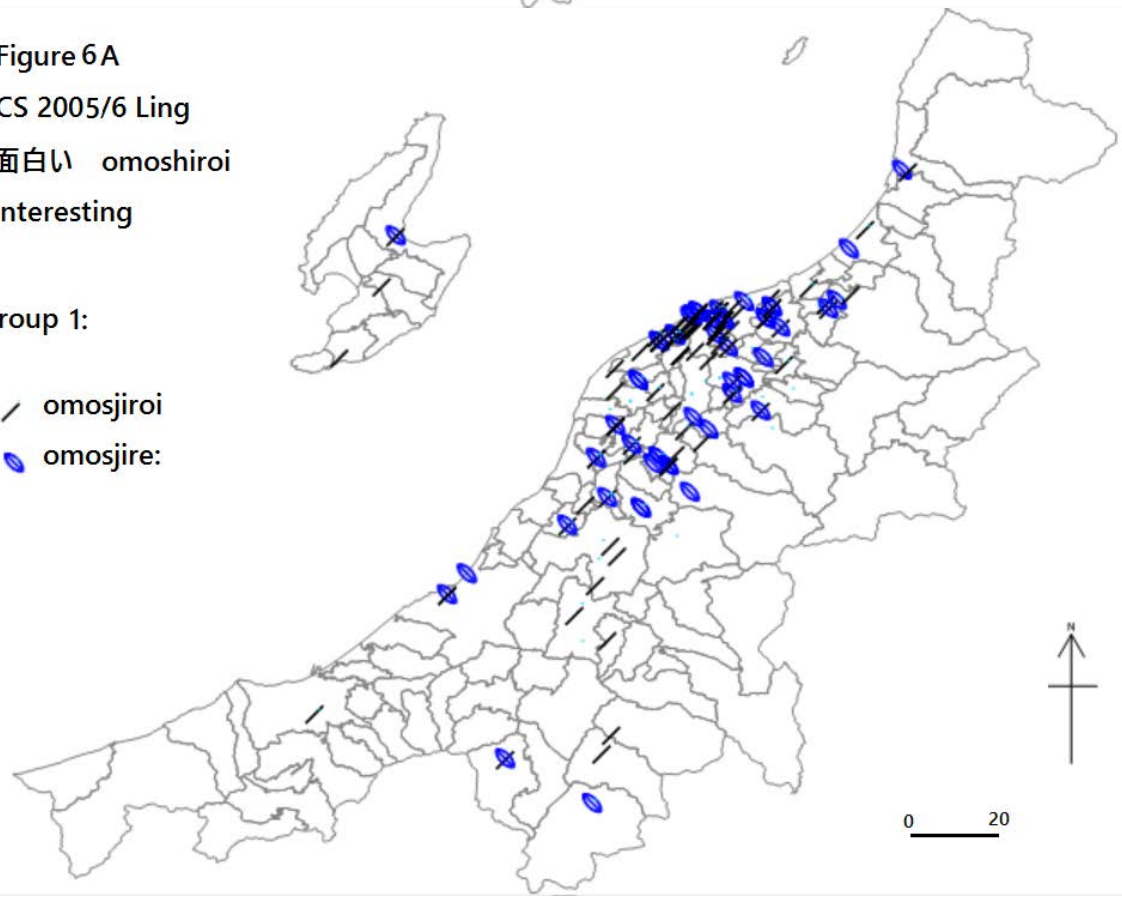


Figure 6B

CS 2005/6 Ling

面白い omoshiroi

Interesting

Group 2:

omossje omosje

omosjoi omossjoi

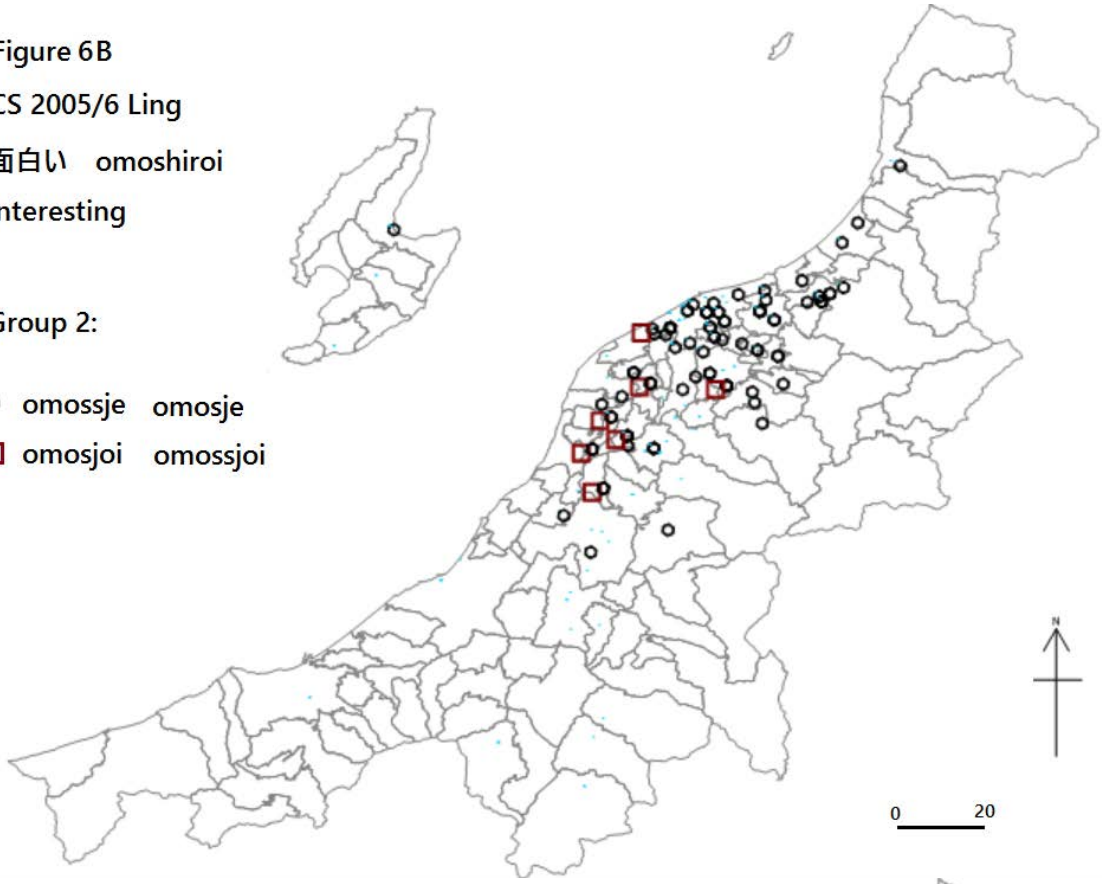


Figure 6C

CS 2005/6 Ling

面白い omoshiroi

Interesting

Group 3:

omosji:

omoroi

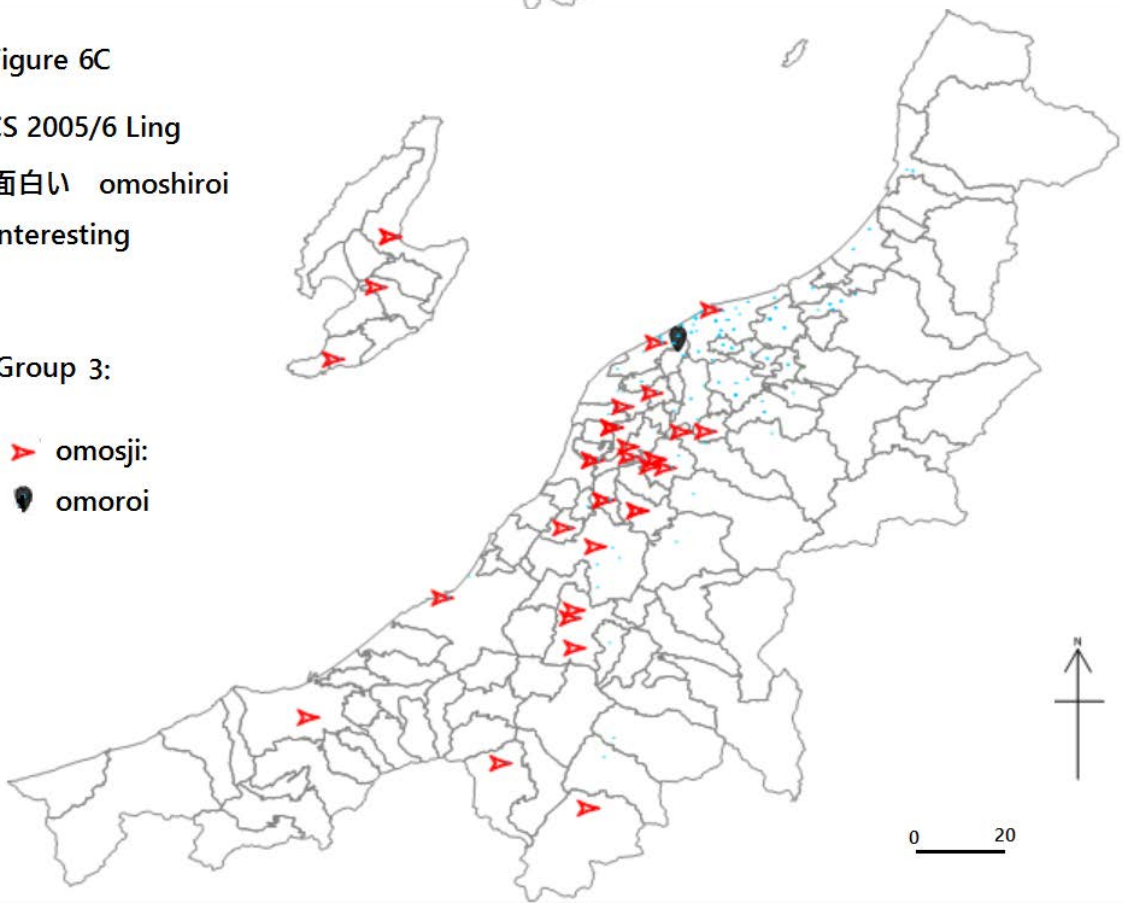


Figure 7A

CS 2005/6 Ling

面白かった

omoshirokatta

Interesting + past

Group 1:

omosjirokatta

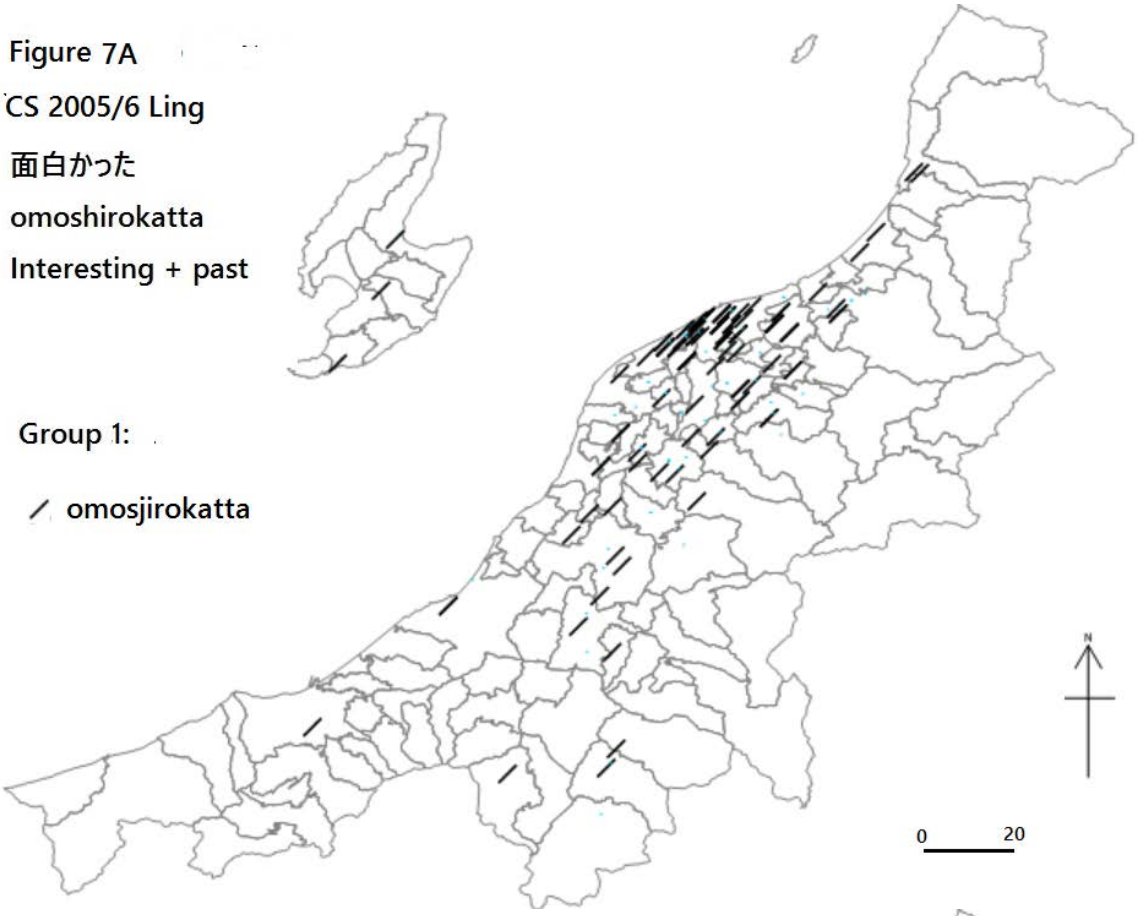


Figure 7B

CS 2005/6 Ling

面白かった

omoshirokatta

Interesting + past

Group 3:

omosjekatta

omosjokatta

omosjakatta

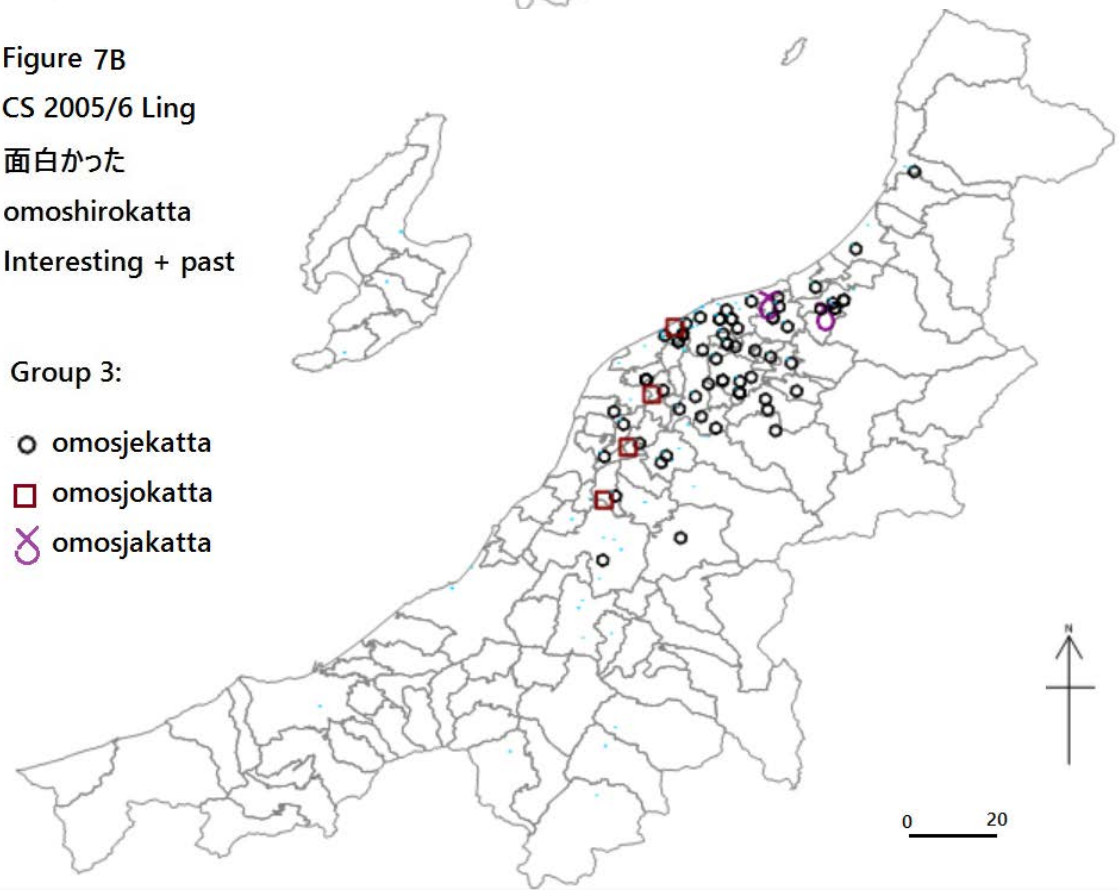


Figure 7C
CS 2005/6 Ling
面白かった
omoshirokatta
Interesting + past

Group 3:

omosjikatta
omorokatta

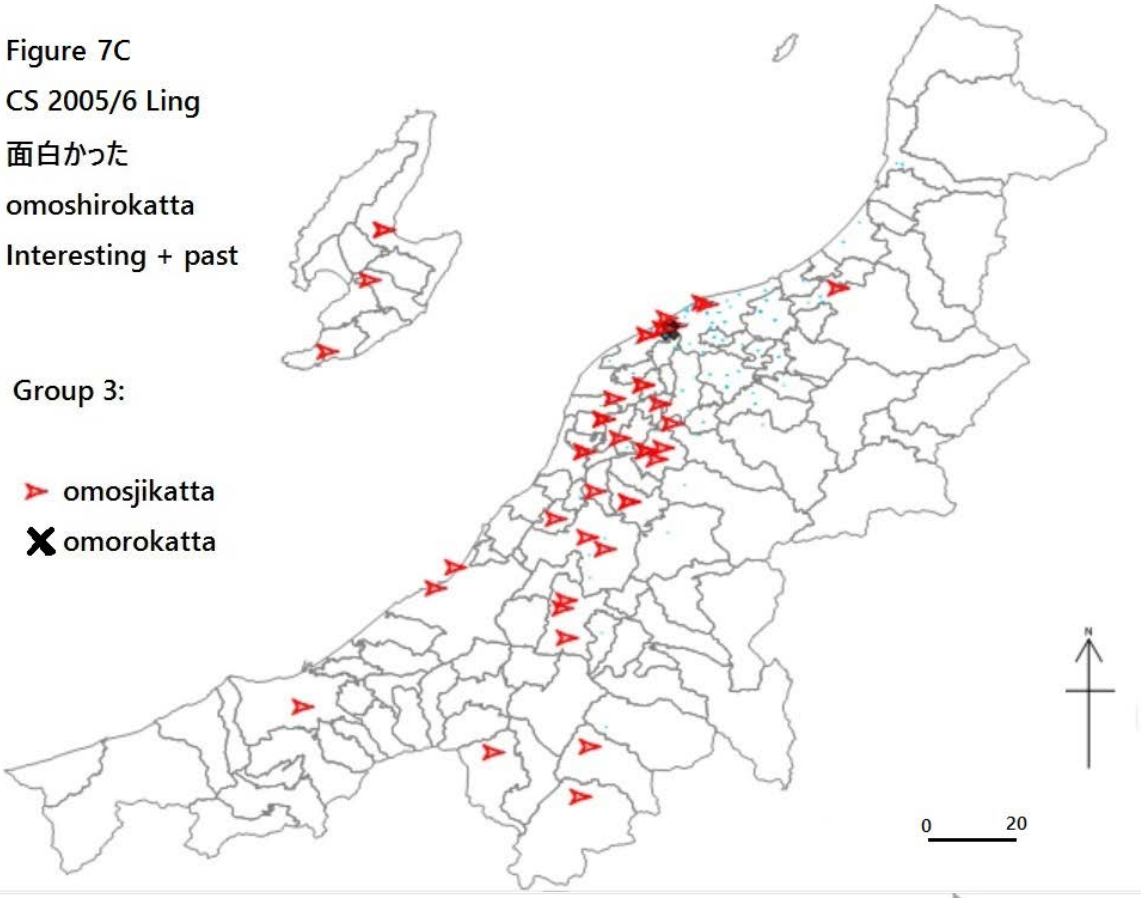


Figure 8A
CS 2005/6 Ling
面白くない
omoshirokunai
Interesting + negative

Group 1:

omosjirokunai

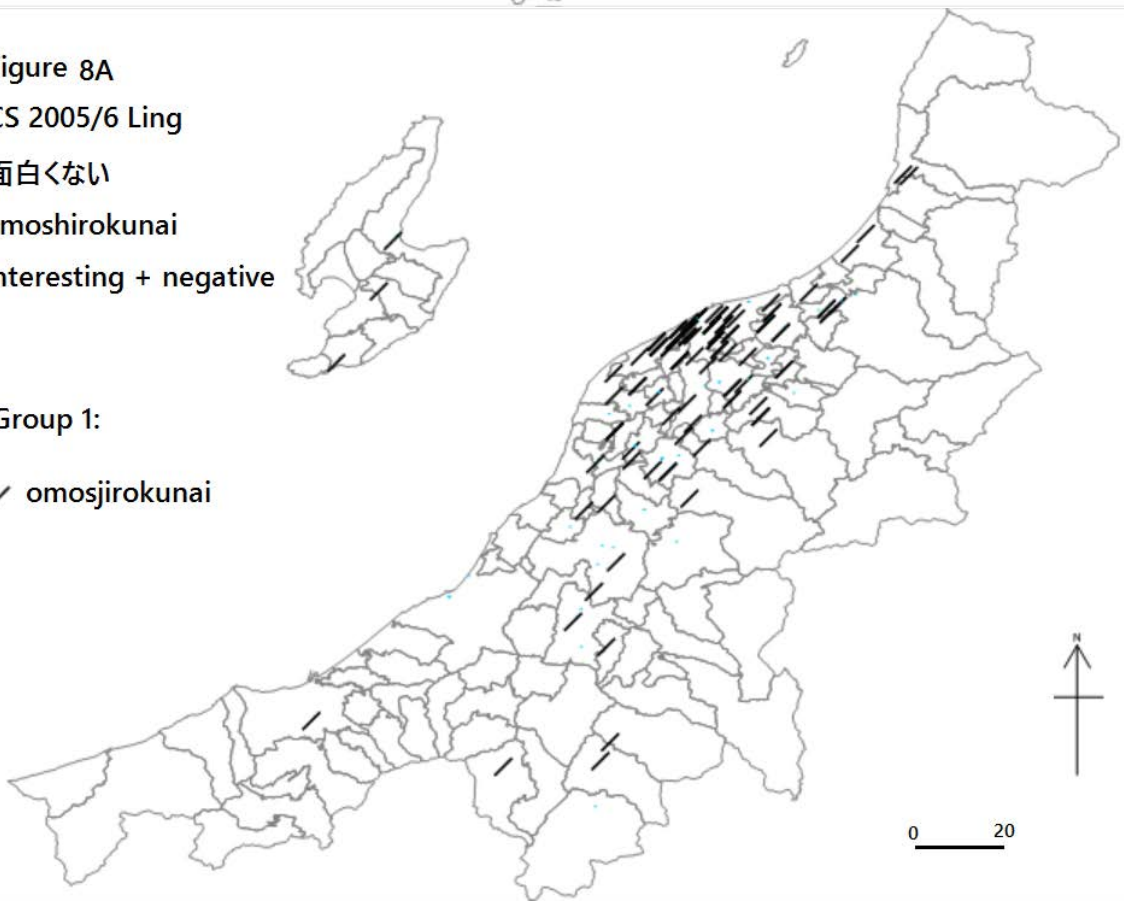


Figure 8B

CS 2005/6 Ling

面白くない

omoshirokunai

Interesting + negative

Group 2:

- omossjenai/ne
- omossjekunai/ne
- omossjakunai
- omossjone
- omossjokunai/ne

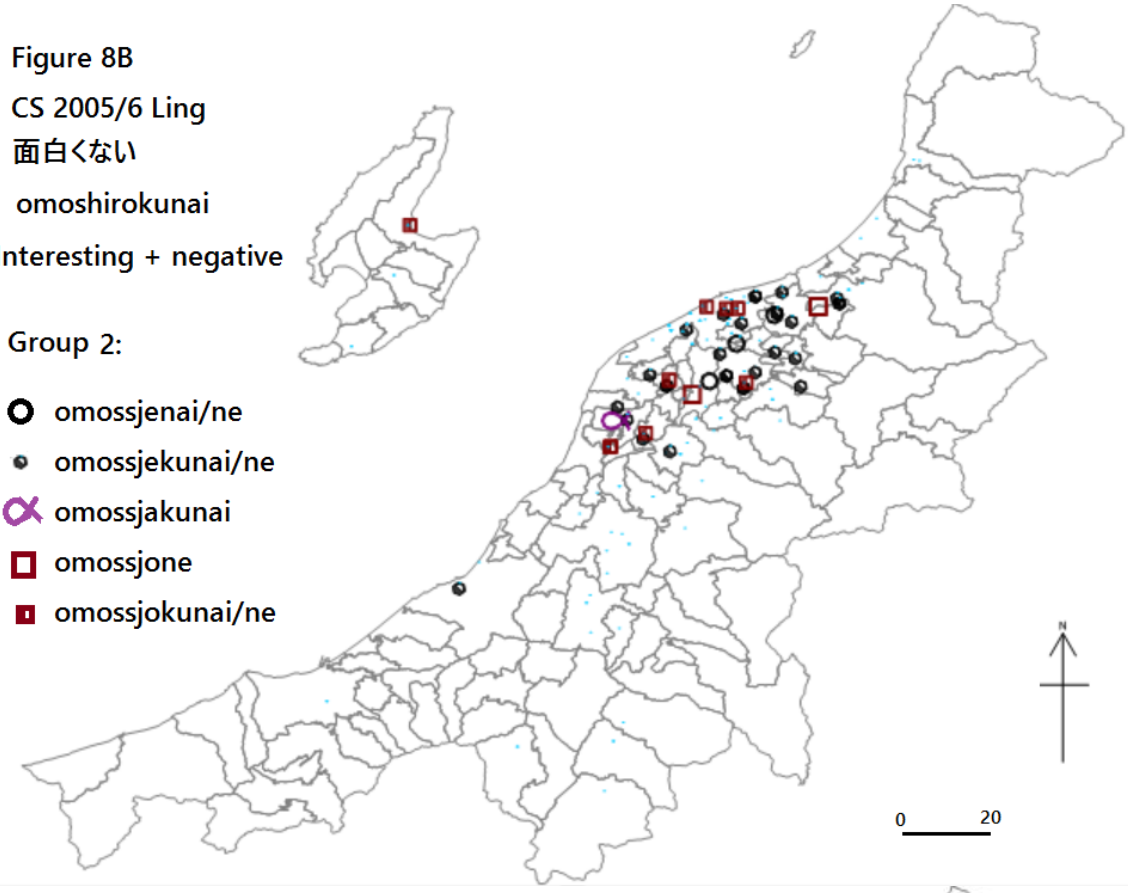


Figure 8C

CS 2005/6 Ling

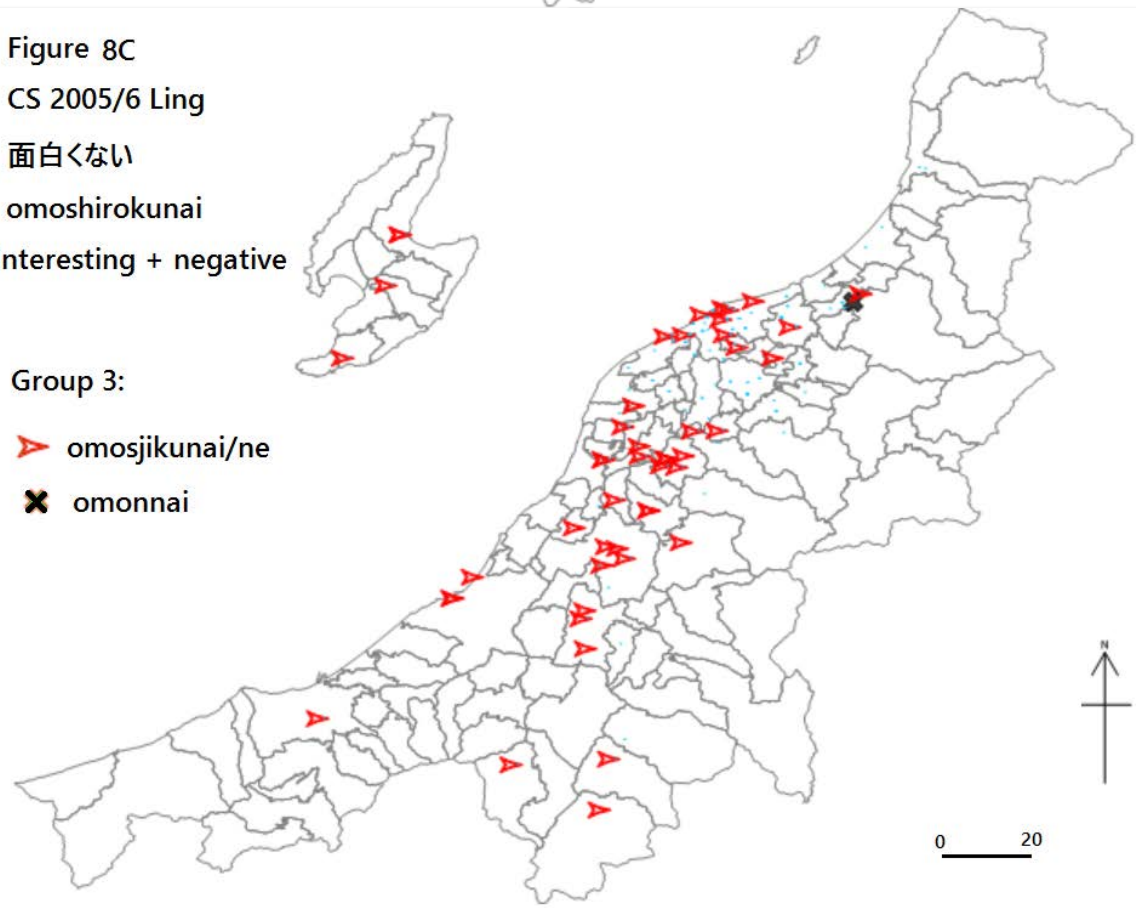
面白くない

omoshirokunai

Interesting + negative

Group 3:

- omosjikunai/ne
- omonnai



Geographical Distribution and Patterns of Basic Ainu Vocabulary in Hattori and Chiri (1960)

Mika Fukazawa

The Preparatory Office for National Ainu Museum

Abstract

In order to make 100 geographical maps for the study of geolinguistics in this article, we examined the first 100 items of the basic Ainu 200-word list found in Hattori and Chiri's "a lexicostatistic study on the Ainu dialects" (1960). Although Hattori and Chiri (1960) and Asai (1974) basically used the same data, their judgements of "similarity" among dialects were different. Here, we judge once again the cognates found in the dialectal data. The monotonous type consists of the cognate words among all of the Ainu dialects; therefore, we would be able to draw a map with the same symbols. However, we give different symbols for different accents, phonetic/phonological characters and so on, and then this article attempts to describe the historical distribution of the forms.

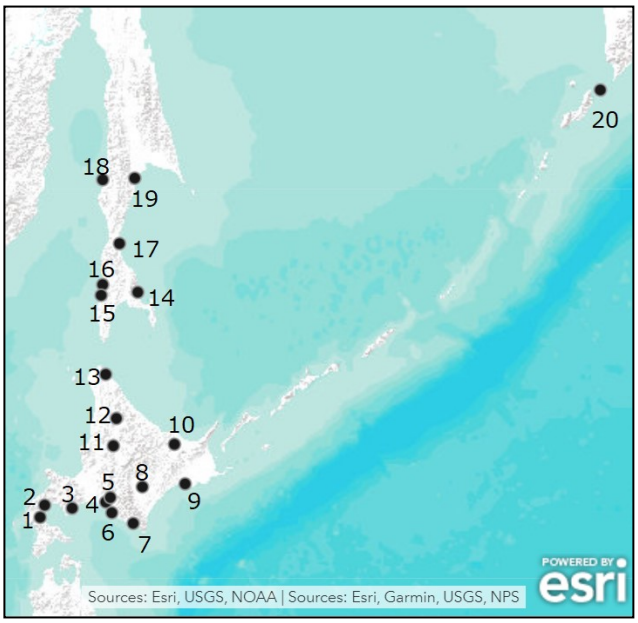
1 Introduction

In this article, we will examine the first 100 words of the basic Ainu 200-word list found in Hattori and Chiri (1960), and investigate how we can use the data to make 100 geographical maps on the study of geolinguistics. The most important work on the Ainu dialects was conducted by Hattori, Chiri, and their collaborators from 1955–1956. In the introduction of Hattori and Chiri (1960: 307), Hattori reported that some of the informants were the last native speaker or speakers of the Ainu dialect, and all of them were very old. Some informants had even died since the investigation. Hattori also said that some speakers could speak the Ainu language fluently, but others knew only several words. All of their investigation was done in Hokkaido because the southern Sakhalin region was occupied by the Soviet Union at that time, and the informants of the Sakhalin dialects lived in Hokkaido as "repatriates." However, the researchers could not have investigated the Kuril dialects, which had already disappeared in the early 20th century. At present, there are few native speakers of Ainu. In order to know the dialects and varieties in the Ainu language, we will primarily refer to Hattori and Chiri (1960), Asai (1974) and Hattori (1964).

The dataset of Hattori and Chiri (1960) is based on their investigation conducted in April 1955, and Hattori's own investigation from summer 1955. Some of the data were modified and added to by referring to later investigations with their collaborators. Map 1 shows Hattori and Chiri's investigation of 19 Ainu dialects: 13 dialects in Hokkaido and 6 from Sakhalin. Hattori and Chiri (1960) presented a list of 200 items, which was arranged according to the Swadesh wordlists. The first 100 items are exactly the same as Swadesh's later 100-word list, though the second half includes several different words from the remaining 107 words in Swadesh's original 200-word list. Hattori wrote a dictionary of the Ainu dialects (1964), which was based on the field investigations in 1955–1956 by Hattori, Chiri and some collaborators, e.g., Suzuko Tamura, who became a specialist in the Saru dialect of Ainu.

After Hattori and Chiri's works, Asai (1974) attempted to perform cluster analysis of the Ainu dialects. To the data of Hattori and Chiri's list, Asai (1974) added the Chitose (Hokkaido) dialect through his own investigation, and the Kuril dialects from written materials (Torii 1903, Murayama 1971, Pinart 1872). He also made some modifications to Hattori and Chiri's data on the Asahikawa, Obihiro, and Kushiro (Hokkaido) dialects from his own data. Asai (1974) scientifically established the

dialectal groups of Ainu; he also established the "major division" between Hokkaido, Sakhalin and the Kuril Islands in the cluster analysis, although Kindaichi (1932) had earlier suggested this.



Hokkaido:
1. Yakumo, 2. Oshamambe, 3. Horobetsu,
4. Biratori, 5. Nukibetsu, 6. Niikappu,
7. Samani, 8. Obihiro, 9. Kushiro,
10. Bihoro,
11. Asahikawa, 12. Nayoro, 13. Sōya,

Sakhalin:
14. Ochiho, 15. Tarantomari, 16. Maoka,
17. Shiraura, 18. Raichishika, 19. Nairo

The Kuril Islands:
20. Shumshu

Map 1: The Atlas of Ainu Dialects

2 Classifications in the previous studies

Hattori and Chiri (1960) and Asai (1974) basically used the same data; however, they judged "similarity" among dialects by different rules. In Hattori and Chiri (1960: 307), the cognate residues are marked with the symbol +, and the noncognates are marked with the symbol -. When one or both of the dialects have two forms and it is difficult to decide which form is more suitable for the item, the item of the forms is marked with the symbol ±. The symbol ○ stands for questionable cognacy; The symbol ? is used when there are some doubt about the record; The symbol • shows the informants had no answer; The symbol ( ) shows lacuna in the record. Hattori and Chiri (1960) calculated the similarity between dialects as the ratio of the sum of + (1.0) and ± (0.5) to the sum of + (1.0), - (1.0) and ± (1.0). The symbols ○, ?, • and ( ) were ignored.

Asai (1974) judged the questionable records by himself and ignored some items which Hattori and Chiri (1960) had marked with the symbols ?, • and ( ). For example, the item "28. skin" has two classifications of {*kap*} and {*rus*} in Hattori and Chiri (1960), although Asai (1974) omitted the item for the following reason: In Japanese, *kawa* means "bark," "skin," "leather" and less commonly "fur." There are three words in Ainu which are translated as one word *kawa* in Japanese: (a) {*kap*} for "bark," "skin" and "leather," (b) {*rus*} for "hard fur" and (c) {*ur*} for "soft fur" and sometimes "leather" (see Asai 1974: 86). Furthermore, in Hattori and Chiri (1960), words having the same root and different phonemes were judged as cognates, while in Asai (1974), they were not judged as "obviously similar or nearly similar Ainu forms." For example, the item of "42. mouth" has four classifications in Asai (1974), as {*paro*}, {*caro*}, {*cara*}, or {*caru*}, but Hattori and Chiri (1960) judged that all of them are cognates. However, *pa-* and *ca-* probably came from different roots, as Kirikae (1994) and Fukazawa (2014) mentioned.

It appears that the judgement by Hattori and Chiri (1960) is too optimistic. They omitted many items whose dialectal forms were considered to be cognates, and then used only 91 of 200 items for calculation. On the other hand, the judgement of "similarity" by Asai (1974) seems to be vague; e.g., *hotke* and *hokke* for "lie" were judged as similar words, but *apkas*, *akkas* and *ahkas* for "walk" were judged as different words (Asai 1974: 72). Asai (1974) did not mention his rules used for judgement, but gave two examples of judgement as follows: "the initial /t-/ of many forms in [the Nairo dialect] seems

to correspond to /r-/ in other dialects but they are taken as different," and "many /ar-/, for instance, in Hokkaido dialect and /-ara/ in Sakhalin dialects are believed similar."

3 Monotonous type: the items with phonological differences

This section will address the monotonous type of the first 100 items in Hattori and Chiri (1960). The monotonous type consists of cognate words among dialects that were mostly omitted in Hattori and Chiri's calculations (1960). In this article, we will call the items including the words with only phonological differences "Monotonous Type." Here, we classify the words, as in (1) and (2) below, into subitems; e.g., the item of "55. eat" is classified into the subitems of "55a. eat (V1)" and "55b. eat (V2)." The items, as in (3) and (4), are not included in the monotonous type, and the items, as in (5), vary on a case-by-case basis. Note that there may be some different meanings and uses for any given item.

(1) Single form (SG) and plural form (PL)/reduplication form (REDUP): e.g.,
a. *ráyke* (SG) and *rónno* (PL) "62. kill"
b. *kupá* (SG) and *kupápa* (REDUP for "many times") "56. bite"

(2) Argument structure of verbs (V0, V1, V2, etc.)[1]: e.g.,
a. *ipé* (V1) and *é* (V2) "55. eat"
b. *sírsesek* (V0) and *sések* (V1) "93. hot"
c. *méan* (V0) and *meerayke* (V1) "94. cold"

(3) Inclusive (INCL) and exclusive (EXCL) we: e.g.,
a. *anutári (utàr)* (INCL) and *ciókay (utàr)* (EXCL) "3. we"

(4) Proximity of demonstratives: e.g.,
a. *tánpe*, *taánpe* and *tapánpe* "4. this"
b. *toánpe* and *toónpe* "5. that"

(5) Body parts or the aspect of things: e.g.,
Item "28. skin," which Asai (1974) mentioned above

We will focus here on the items of the monotonous type as in (6).

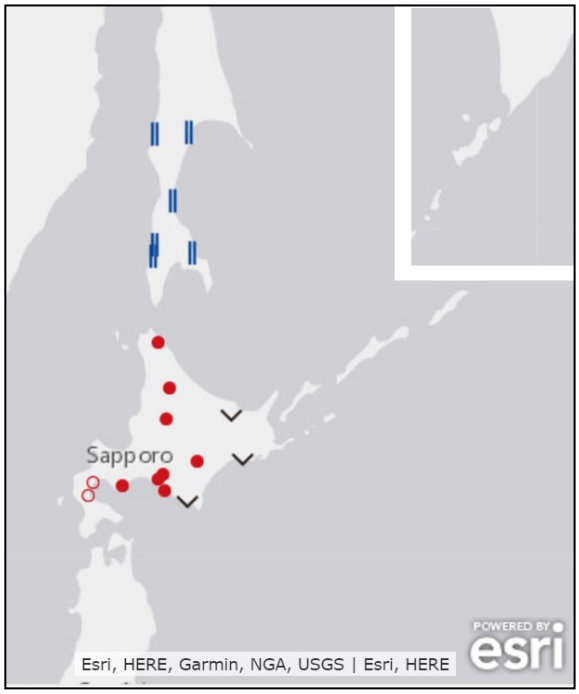
(6) Monotonous Type[2] (Hattori and Chiri 1960; Items No.1–100)

No. 1-10:	2. you [thou] / (10. many)
No. 11-20:	11. one / 12. two / 13. big / 17. man / 18. person / 19. fish / (20. bird)
No. 21-30:	21. dog / 23. tree / 24. seed / 25. leaf / 28. skin / 29. meat / 30. blood
No. 31-40:	31. bone / 32. grease / 34. horn / 36. feather / 39. ear / 40. eye
No. 41-50:	41. nose / 43. tooth / 45. claw / 49. belly / 50. neck
No. 51-60:	52. heart / 54. drink / 55. eat / 56. bite / 57. see / 58. hear / 60. sleep
No. 61-70:	61. die / 62. kill / (63. swim) / 65. walk / 66. come / 68. sit
No. 71-80:	71. say / (72. sun) / (73. moon) / 75. water / 77. stone / 78. sand
No. 81-90:	83. ashes / 84. burn / 85. path / 86. mountain / 87. red / 88. green / 89. yellow / 90. white,
No. 91-100:	92. night / 93. hot / 96. new / 97. good / 100. name

[1] These abbreviations mean the verb taking 0 argument (the zero-place verb), the verb taking 1 argument (the one-place verb), the verb taking 2 arguments (the two-place verb), and so on.
[2] The items in parenthesis may be included in the other types.

3.1 Accent

Depending on the Ainu dialect, there may be quantitative accent (QA), two types of pitch accent (PA), or no accent (NA) (See Map 2).



Pitch accent
Pitch accent on the third syllable
Quantitative accent
No accent

Map 2: The Kinds of Accents

3.1.1 Pitch accent

The majority of Hokkaido dialects have a pitch accent (PA) (c.f., Chiri 1942, Hattori 1967, Tamura 2000 etc.). In the Hokkaido dialects, the dialects of Samani, Bihoro, Kushiro and Shizunai have no accent (NA) (Hattori and Chiri 1960: 311).

Items	Hokkaido, PA	Hokkaido, NA	Sakhalin, QA	Other forms
13. big	*poró*	*poro*		c.f. (10)
18. person	*áynu*	*aynu*		
29. meat	*kám*	*kam*		
30. blood	*kém*	*kem*		
34. horn	*kiráw*	*kiraw*		
52 heart	*sánpe*	*sanpe*		
55a. eat (V1)	*ipé*	*ipe*		c.f. (55b)
77. stone	*sumá*	*suma*		
92. night	*kúnne*	*kunne*		*ukuran*[3]

Tamura (2000: 21) suggested "the rise from low to high is distinctive," and "the syllables before this rise are all low, and the syllables following it gradually fall with a certain degree of regularity." When the first syllable is open in a word with more than one syllable, the high pitch falls on the second syllable, i.e., $(C_1)\ V_1\ (C_2)\ \acute{V}_2\ (C_3...)$. When the first syllable is closed, the high pitch falls on the first syllable, i.e., $(C_1)\ \acute{V}_1\ C_2\ (C_3\ V_3...)$. In addition to these patterns, the accentual nucleus may in some cases be on the

[3] In addition to *kunne*, the Raichiska (Sakhalin) dialect has the term *ukuran*, which means "nighttime" (Hattori 1964).

open first syllable, i.e., (C_1) $\acute{V}_1$ (C_2) V_2 (C_3...), which is regarded as an exceptional accent (an irregular pattern of accent).

In Hokkaido, the accentual rules in the Yakumo and Oshamambe dialects are different from the other Hokkaido dialects. The high pitch falls on the third syllable when the first and second syllables are open in a word with more than two syllables, i.e., (C_1) V_1 (C_2) V_2 (C_3) $\acute{V}_3$ (C_4...).

Items	Hokkaido, PA		Hokkaido, NA	Sakhalin, QA	Other forms
2. you (thou)	*eaní*	*eáni*	*eani*		
23b. tree[4]	*cikuní*	*cikúni*	(no data)		c.f. (23a)
86. mountain	*nupurí*	*nupúri*	*nupuri*	*nupuru*	

3.1.2 Quantitative accent

The Sakhalin dialects have a quantitative accent, in which vowel length is distinctive. Long vowels in Sakhalin are used for the open one-syllable words and the words with an irregular accent on the first open syllable in Hokkaido.

Items	Hokkaido, PA	Hokkaido, NA	Sakhalin, QA	Other forms
23a. tree	*ní*	*ni*	*nii*	c.f. (23b)
24. seed	*pí*	*pi*	*pii*	*tané/tane, epuyke*[5]
54. drink	*kú*	*ku*	*kuu*	
55b. eat (V2)	*é*	*e*	*ee*	c.f. (55a)
58. hear	*nú*	*nu*	*nuu*	
63. swim	*má*	*ma*	*maa*	*wá, sus*[6]
68. sit	*á*	*a*	*aa*	*(móna)á, mona*[7]
71a. say (V2)	*yé*	*ye*	*yee*	c.f. (71b)[8]
83. ashes	*úna/úyna*	*una*	*uuna*	
85. path	*rú*	*ru*	*ruu*	
87. red	*húre*	*hure*	*huure*	
93a. hot (V1)	*sések*	*sesek*	*seesek/seeseh*	c.f. (93b)
100. name	*ré*	*re*	*ree/tee*	

3.1.3 Historical description

Now, we will examine Map 3, for "83. ashes." The words for "83. ashes" have an exceptional accent, that is, an accent nucleus on the first open syllable, *úna*. In Yakumo and Oshamambe, interestingly, there appears the form *úyna* for "83. ashes." The consonant *y* [j] may be the rest of vowel length in the

[4] The term {*cikuni*} means "(standing) tree" in Yakumo, Horobetsu and Saru, and "wood" in Yakumo and Saru (Hattori 1964: 196). In Raichiska, there is the form *cikurunii* "firewood" (Hattori 1964: 107), which can be analyzed as *ci-kuru-nii* 1PL.EXCL-warm_oneself_at_fire.V2-tree (c.f. Chiri 1956: 149).

[5] The term {*tane*} is originated from Japanese *tane*, and *epuyke* does not only means "seed," but also "fruit", "bud" and "flower."

[6] The Sōya dialect has the term *wá*, which means "to wade (across a river)" in the other dialects (Hattori and Chiri (1964). The term *sus* in Maoka means "to bathe in water" in Saru and Nayoro (Hattori 1964: 183).

[7] Hattori and Chiri (1960: 336) analized *mona* into *mon* "motionless" and *a* "to sit." Furthermore, in the forms of Niikapu dialect, *(móna)á*, the part of *móna* could be originated from *móno a* "to sit still" in the Saru dialect (Tamura 1996: 392-393).

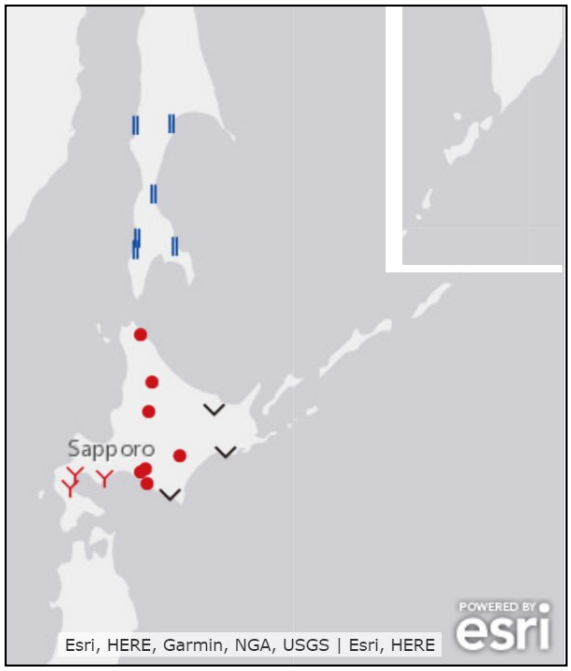
[8] Hattori and Chiri (1960: 336) reported *hawki* in Kushiro is an intransitive verb, but *ye* in the other dialect is a transitive verb, and then they assumed the Kushiro dialect also has the transitive verb *ye*. Here, we classified *hawki* as "71b. say (V1)."

process of changing to a high-pitched accent or a long vowel. Map 4 indicates the geographical distribution for "13. big" and "10. many." The word form of "many" consists of the verb *poró/poro* "big" and the adverbial suffix *-no*: e.g., the form *poróno* was also reported by Tamura (1996: 545). In the forms *porónno/poroónno/poronno* in Hattori and Chiri (1960: 314), the sound *n* may be inserted between *poró/poro* and *no*. The term *poroónno* in Oshamambe has double vowels [oo] and a high pitch on the third open syllable, while in Sakhalin, there are the long vowel form *poroono* [poroːno] and the inserted-*n* form *poronno*.

There are no audio materials of the dialects in the Kuril Islands, and so it is unknown what kind of accent these dialects had. We can only surmise Kuril accents from the materials written by some explorers from the 18th to the early 20th centuries: e.g., "porogó" (Krascheninnikov 1738) and "porohu" (Torii 1903). If we consider these terms in Kuril, the form *poroónno* in Oshamambe might go back to the proto-Ainu form **poroHó-no*[9]. Then, the historical description is as follows:

(7) Historical description of the words for "10. many"
**poroHó-no* > *poroónno* > *porónno*/*poronno*/*poroono*

Items	Hokkaido, PA		Hokkaido, NA	Sakhalin, QA	Other forms
13. big	*poró*		*poro*		
10. many	*poroónno*	*porónno*	*poronno*	*poroono*	*renkayne, okayno*[10]



● *úna*
Y *úyna*
‖ *uuna*
⌄ *una*

Map 3: "83. ashes"

"13. big"
● *poró*
⌄ *poro*

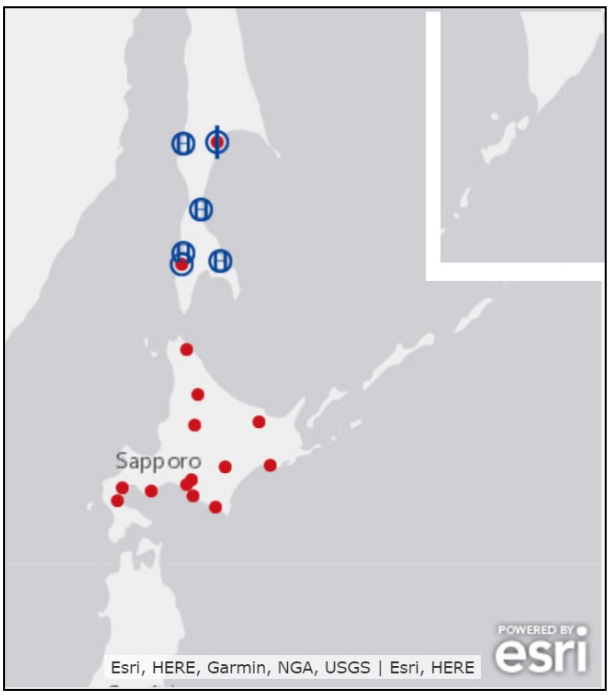
"10. many"
● *porónno*
‖ *poroónno*
‖ *poroono*
⌄ *poronno*

Map 4: "13. big" and "10. many"

[9] *H* of the proto-Ainu form might be pronounced as [x] or [ɣ].
[10] The word *okayno* is found in Tarantomari, Raichiska and Nairo, and the word *renkayne* is found in Tarantomari, Maoka and Raichiska.

3.2 Phonetic characteristics of the Sakhalin dialects

In this section, we will show the typical differences in the phonological and/or phonetic structures between the Sakhalin and Hokkaido dialects (see Chiri 1942, Tamura 2000 etc.). Map 5 shows the phonetic characteristics of the Sakhalin dialects.



(Same as the Hokkaido dialects)
rV
-h
t-

Map 5: Phonetic Characteristics of the Sakhalin Dialects

3.2.1 The syllable structure of rV in Sakhalin

In Sakhalin, the coda /-r/ has changed to the onset /r-/ of a syllable /rV/ (see Chiri 1942, Tamura 2000 etc.).

Items	Hokkaido, PA	Hokkaido, NA	Sakhalin, QA	Other forms
32. grease	*kírpu/kírpo*	*kirpo*	*kiripu/kiripo*	
39. ear	*kisár*	*kisar*	*kisara/kisaru*	
57. see	*nukár*	*nukar*	*nukara*	
60. sleep	*mokór*	*mokor*	*mokoro*	
96. new	*asír*	*asir*	*asiri*	
97 good	*pírka*	*pirka*	*pirika*	

3.2.2 *Coda -h [-x] in Sakhalin*

The codas /-p, -t, -k/ and occasionally /-r/ in Hokkaido have historically changed to /-h/ [-x] in most Sakhalin dialects. When these consonants appear after the vowel /i/ in Hokkaido, they make the sound [-is] (not [-x]) in Sakhalin (See “40. eye”).

Items	Hokkaido, PA	Hokkaido, NA	Sakhalin, QA	Other forms
11. one	*sinép*	*sinep*	*sineh*	
12. two	*túp*	*tup/cup*	*tuh*	
19. fish	*cép/ciép*	*cep*	*ceh*	

20. bird	*cikáp*	*cikap*		*cikah*	*círi*[11]
28a. skin	*káp*	*kap*		*kah*	c.f. (28b)
40. eye	*sík*	*sik*		*sis*	
66. come	*ék*	*ek*		*eh*	
72. sun	*cúp*	*cup*		*cuh*	*tonpi*[12]
73. moon	*cúp*	*cup*		*cuh*	*tonpi*
75. water	*wákka*	*wakka*		*wahka*	
17. man	*ókkayo/ókkay*	*okkay*		*ohkayo*	*okkaw*[13]
50. neck	*rekút*	*rekut*	*tekut*	*rekuh*	*onánci*[14]
65. walk	*ápkas*	*apkas*	*akkas*	*ahkas*	
93a. hot (V1)	*sések*	*sesek*	*seesek*	*seeseh*	c.f. (93b)

3.2.3 Onset t- in the Nairo dialect

The Nairo dialect of Sakhalin features the onset of the first-syllable /t-/, which corresponds to /r-/ in the Hokkaido and the other Sakhalin dialects. 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-/ of the first syllable is confused with /t-/ (or /d-/). Vovin (1993: 16-18) reconstructed Proto-Ainu **d*- when Common Ainu [r-] corresponds to Nairo [t-].

Items	Hokkaido, PA	Hokkaido, NA		Sakhalin, QA	Other forms
28b. skin	(No data)	*rus*		*tus*	c.f. (28a)
61. die	*ráy*	*ray*		*tay*	
36. feather	*ráp*	*rap*	*rah*	*tap*	*cikapráp, cikáprap*[15]
50. neck	*rekút*	*rekut*	*rekuh*	*tekut*	*onánci*
62a. kill (SG)	*ráyke*	*rayke*	*rayki*	*tayki*	c.f. (62b)
100. name	*ré*	*re*	*ree*	*tee*	
90. white	*retár*	*retar*	*tetara*		

3.3 Phonological assimilation, metathesis and so on

In this section, we will take a look at the monotonous type of items with a phonological assimilation and so on. For the geographical locations of the dialects, see Map 1 in Section 1.

3.3.1 Phonological assimilation

The words for "65. walk" and "93b. hot (V0)" are the monotonous type with a phonological assimilation across dialects, e.g., /-pk-/ > /-kk-/ and /-rs-/ > /-ss-/.

Items	Hokkaido, PA		Hokkaido, NA		Sakhalin, QA		Other forms
65. walk	*ápkas*		*apkas*		*akkas*	*ahkas*	
93b. hot (V0)	*sírsesek*	*síssesek*	*sirsesek*	*sissesek*	(no data)		c.f. (93a)

[11] The Yakumo dialect has *cikáp* and *círi*, and the Shumshu dialect (Kuril) also has *cir(i)*: *cir, čir* (Dibowski 1982), chiri (Torii 1903). The word {*ciri*} may be older than {*cikap*} since these words show an ABA distribution.
[12] The word *tonpi* occurs in Tarantomari. The lexical form of *tonpi* could be analyzed into *tom* "to sparkle; to shine" and *-pi* (< *pe*) classifier of "thing" (uncertain) (cf. Tamura 1996: 580; (See also Fukazawa 2016a).
[13] This form is found in Ochiho.
[14] The word *onánci* [onantɕi] is the borrowing word from the Japanese word *unazi* [unadʑi], which means "nape."
[15] The word {*cikaprap*} consists of {*cikap*} "bird" and {*rap*} "feather."

3.3.2 Phonological metathesis

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 (Fukazawa 2016b).

In Ainu, beside the terms *kunne* "black," *retar* "white" and *hure* "red," the term *siwnin* (*suynin* in Bihoro) is assumed to be one of the basic color terms. Tamura (1974) notes the dialectal differences in the foci of *siwnin*/*suynin*: "light blue" and "yellowish green." For example, in Biratori (Hokkaido), the prototypical *siwnin*/*suynin* color is a grass green, but in Bihoro (the northeastern Hokkaido dialect) it is a blue such as the color of the sky and the sea (Tamura 1996; Hattori 1964).

Items	Hokkaido, PA	Hokkaido, NA		Sakhalin, QA	Other forms
88. green	*síwnin*	*siwnin*	*suynin*	*siwnin*	
89. yellow	*síwnin*	*siwnin*	*suynin*	*siwnin*	*sikérpepeus*[16]/ *húre/hure*

3.3.3 The insertion or deletion of onset h-

In the Bihoro and Samani dialects (Hokkaido, NA), an onset /h-/ is often inserted or deleted. In the other dialects of Obihiro, Kushiro, Bihoro, Asahikawa and Nayoro, the onset /h-/ is often deleted, e.g., *hopuni:opuni*, *hoyupu:oyupu* "64. fly" etc.

Hattori and Chiri (1960: 336) marked the relationships between *yam* in Sakhalin and *hám/ham* in Hokkaido with the symbol ○. However, they suggested the historical description of "25. leaf": *yam* in Sakhalin may have originated as part of *niiyam* "the leaf of tree," which comes from *niiham* < *nii* "tree" - *ham* "leaf." In the Shumshu dialect (Kuril), Torii (1903) reported the same form *yam* as the Sakhalin dialects.

Items	Hokkaido, PA	Hokkaido, NA		Sakhalin, QA	Other forms
45. claw	*ám*	*ham*	*am*		
78. sand	*otá*	*hota*	*ota*		*piyóta*[17]
49. belly	*hón*	*on*	*hon*		*pise*[18]
25. leaf	*hám/ham*			*yam*	

3.3.4 Onset c- in the Samani dialect

In Samani, there is an onset /c-/ [ʧ], which corresponds to the onset /t-/ of the other dialects. However, it might be an idiolect of the informant cited, because the other speakers of Samani did not pronounce the onset /t-/ as /c-/ [ʧ] (c.f. Komatsu 2004: 181, Nakagawa, p.c., 2016).

Items	Hokkaido, PA	Hokkaido, NA		Sakhalin, QA	Other forms
12. two	*túp*	*tup*	*cup*	*tuh*	
41. nose	*etú*	*etu*	*ecu*	*etu/etupuy*[19]	

[16] As the term of the Horobetsu dialect, Hattori (1964) shows the term *sikérpepeus*, although Hattori and Chiri (1960: 89) shows *sikérpepus*. The former part of *sikérpe* means "an Amur cork tree," of which endothelium is yellow, but the latter part of *peus* or *pepeus* is unknown, that probably means "water in which an Amur cork tree is soaked."

[17] The term *piyóta* in Nukibetsu is constructed from *pi* "fine" + *ota* "sand," and it means "white volcanic ash" in Saru (Tamura 1996: 535).

[18] The term *pise* in Kushiro means "(generic) stomach" in Sakhalin, "the paunch of a bear" in Horobetsu and "the paunch of a seal or a Steller's sea lion." It also means "the bladder of a bear," and "the air bladder of a fish" (Hattori 1964: 279).

[19] The word *etupuy* is constructed with the nouns *etu* "nose" + *puy* "hole," which means "nostril(s)" in Horobetsu.

3.3.5 Vowel differences

We can find vowel differences among dialects in some items, e.g., *i:e* and *o:u*.

Items	Hokkaido, PA		Hokkaido, NA	Sakhalin, QA		Other forms
21. dog	*setá*	*sitá*	*sita*	*seta*		
32. grease	*kírpu*	*kírpo*	*kirpo*	*kiripu*	*kiripo*	
31. bone	*poné*		*pone*	*poni*		
62a. kill (SG)	*ráyke*		*rayke*	*rayki/tayki*		c.f. (62b)

3.3.6 Vowel contraction

Item "19. fish" is an example of vowel contraction: e.g., *ciép* > *cép/cep/ceh*. The form *ciép* is found in the Yakumo, Oshamambe and Horobetsu dialects, and both forms originate from *ci-e-p* 1PL.EXCL-eat.V2-thing "the thing we eat." In In the Kuril dialects, the contractive form, *cep*, was also reported: e.g., "*cep, čep*" (Dybowski 1892), "*chep*" (Torii 1903).

Items	Hokkaido, PA		Hokkaido, NA	Sakhalin, QA	Other forms
19. fish	*ciép*[20]	*cép*	*cep*	*ceh*	

3.3.7 Historical description

Item "43. tooth" is classified as a monotonous-type word. The form {*imak*} is found in Kuril (Torii 1903) as well as the Sakhalin and the eastern Hokkaido dialects; therefore, it seems to be the oldest form of this item. The forms *nimák* and *mimák* are widely distributed in the western dialects of Hokkaido, and the onset /n-/ or /m-/ of the first syllable might be copied from the nasal consonant /m/ of the second syllable, *mak*.[21]

We also suggest classifying the item "84. burn" as another monotonous type. Hattori and Chiri (1960: 336) judged that the cognateness between {*uhuy*} in Hokkaido and {*hukuy*} in Sakhalin was uncertain, and marked it with the symbol ○. In Kuril, there appears the form *uyva* "to burn" (Dybowski 1892), and the part of *va* appears to be a conjunction or a sentence-final particle in Sakhalin and Hokkaido, i.e., *wa*. If the middle /h/ were deleted, the form *uy* in Kuril may have originated from {*uhuy*}, and if the onset /h-/ were also deleted, there might have initially been **huHuy*[22] before *uhúy/uhuy* in Hokkaido and *hukuy* in Sakhalin.

Items	Hokkaido, PA		Hokkaido, NA		Sakhalin, QA	Other forms
43. tooth	*nimák/mimák*	*imák*	*imak*		*imah*	
84. burn	*uhúy*		*uhuy*	*uyuy*	*hukuy*	*eynu*[23]

(8) Historical description of the words for "84. burn"
  a. **huHuy* > *uhúy/uhuy* > *uyuy/uy*
  b. **huHuy* > *hukuy*

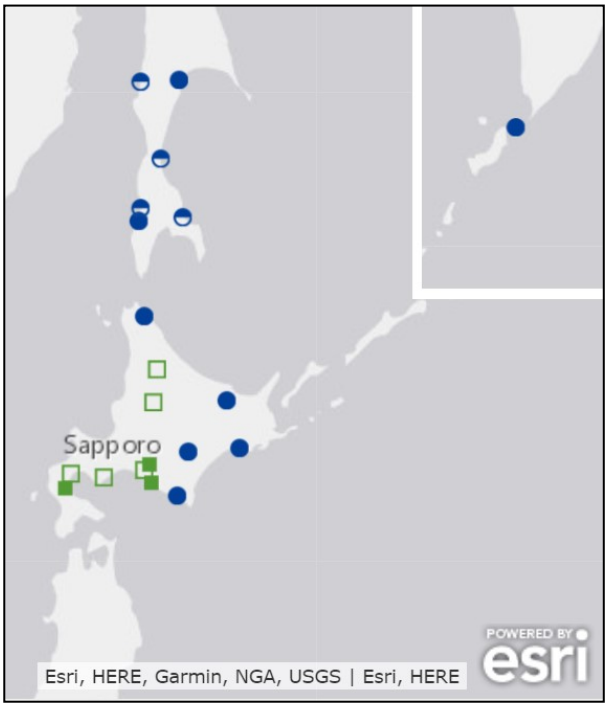
However, in the Shiraura dialect (Sakhalin), it means "nose," and when referring to "nostril(s)," we add one more *puy* to this form, i.e., *etupuypuy* (Chiri 1975: 316, 317).

[20] This form is found in the Yakumo, Oshamambe and Horobetsu dialects.

[21] You can find some words with the onsets /n-/ and /m-/ among dialects, e.g., {nimara} and {mimara} "half," and {nismu} and {mismu} "to be lonely, deserted" (Hattori 1964: 267, 285)

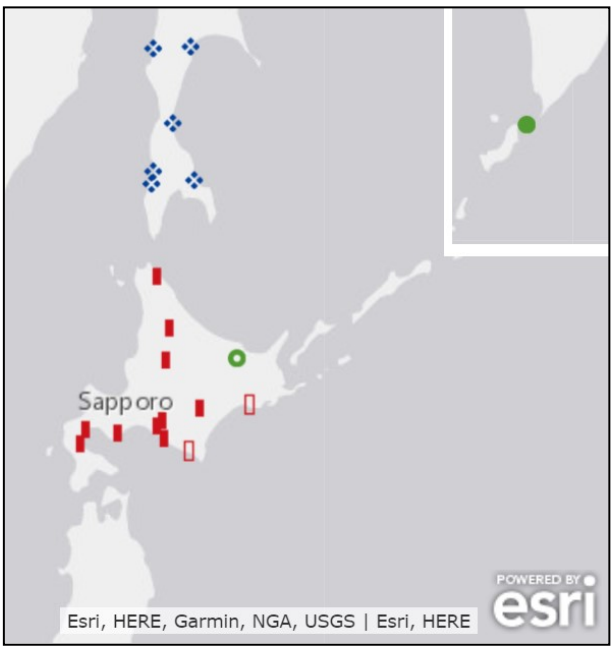
[22] Here, *H* of **huHuy* might be pronounced as [x] or [ɣ] (See also (7) above).

[23] This form is found in Bihoro.



imák, *imak*
imah
mimák
nimák

Map 6: "43. tooth"



uhúy
uhuy
uyuy, eynu
uy
hukuy

Map 7: "84. burn"

4 Conclusion

Before concluding, we will introduce a few studies on the accent of Ainu (See Section 3.2.3). Hattori (1967) proposed that Proto-Ainu had a phonological opposition of vowel length but no opposition of pitch accent, just as in the Sakhalin dialects. Recently, in opposition to Hattori's hypothesis that the shortening of long vowels produced an accent kernel, Uwano (2017: 345) suggested "it is argued that in the examples from the Ainu languages and the Ryukyuan languages the accent kernels (an ascending kernel and a falling kernel, respectively) were already produced at the stage of long vowels. The conclusion is that vowel shortening is not a sufficient condition and further specifications are necessary to apply the idea widely." Sato (2014: 133; 2016) also suggested a new idea as follows: "Ainu might

have had an unknown glide *H. For example, in 連ん本う *renhou* = *répo* 'to place a log on the fire', ん and う are likely to be the reflexes of this *H, although the modern form *rép-o* has an irregular accent on the first syllable. The assumption of this *H enables us to explain why some CVC-stems exhibit an irregular pattern of accent: *reHp-oH > *rép-o*." In Section 3.2.3, the words for "10. many," *porónno* etc., have not an irregular accent on the first syllable, however, Sato's idea inspired me to consider its proto-form. If the Kuril form, *porogó* or *porohu*, was pronounced as [*poroɣo*] or [*poroxo*], we could explain the proto-Ainu form as **poroHó-no*; Here, H was estimated to be like the sound [ɣ] or [x].

In this article, we have only investigated the monotonous type of the first 100 items in Hattori and Chiri (1960), and included the items with controversial issues, such as item "84. burn." The Kuril dialects are not included in the 19 dialects of Hattori and Chiri (1960); however, the forms found in Kuril are important to describe the geographical maps historically, e.g., item "43. tooth." When investigating the remainder of the 100 items, we will consider the Kuril dialects and review Hattori and Chiri's (1960) and Asai's (1974) judgements of "similarity" among dialects.

Abbreviations

1	1st person	REDUP	Reduplication
EXCL	Exclusive	SG	Singular
INCL	Inclusive	V0	Zero-place verb
NA	No accent	V1	One-place verb
PA	Pitch accent	V2	Two-place verb
PL	Plural		

References

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

Chiri, M. 1942. Ainu gohō kenkyū: Karafuto hōgen wo chūshin toshite [The study of Ainu grammar: Specially for Sakhalin dialects]. Reprinted: 1993 [1973]. *Chiri Mashiho chosakushū [Collected works of Chiri Mashiho],* 3. Tokyo: Heibonsha.

Chiri, M. 1956. *Chimei Ainugo shōjiten [The glossary of the place names in Ainu]*. Sapporo: Hokkaido Publication Project Center.

Chiri, M. 1975 [1953]. *Bunrui Ainugo jiten.* [Classified Ainu dictionary]. Reprinted: *Chiri Mashiho chosakushū bekkan*, 1 *[Collected works of Chiri Mashiho Separate volume],* 1. Tokyo: Heibonsha.

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

Fukazawa, M. 2014. The mystery of the phonological distribution /ca/ and /pa/ in the Eastern-Western dialects of Ainu. *Papers from the second international conference on Asian geolinguistics*, *May 2014*: 1–13. Bangkok: Pathumwan Princess Hotel.

Fukazawa, M. 2016a. "Sun" in Ainu. *Studies in Asian geolinguistics*, 1: 8–9, Endo, M. (ed.), Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Fukazawa, M. 2016b. The basic color terms in Ainu. Endo, M. (ed.), *Studies in Asian geolinguitics: Monograph series, 1; Papers from the third international conference on Asian geolinguistics, October 2016*: 102–111. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Hattori, S. 1967. Phonological structure and accent of Ainu: An attempt to reconstruct Proto-Ainu. T*he Study of Sounds*, 13. Reprinted: 2001. *Ainugokō,* 2: 739–755, Tokyo: Yumani Shobou.

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

Hattori, S. and M. Chiri. 1960. Ainugo shohōgen no kisogoi tōkeigakuteki kenkyū [A lexicostatistic study on the Ainu dialects]. *Minzokugaku kenkyū [The Japanese Journal of Ethnology]*, 24(4): 307–342.

Kindaichi, K. 1931. Ainu yūkara gohō tekiyō. [An outline of the grammar of Ainu epic poetry]. *Ainu jojishi:Yuukara no kenkyū [Ainu epic poetry: research on yukar],* 1. Tokyo: Toyo Bunko.

Kindaichi, K. 1932. Hokuōchimeikō: Ōunochimei kara mita Honshū Ezogo no kenkyū [Place names of Hokuō]. *Kanazawa Hakase kanreki kinen Tōyōgo no kenkyū [The study of the Tōyō language: The festschrift for Professor Kanazawa's 60th birthday].* Tokyo: Sanseido. Reprinted: 1993. *Kindaichi Kyoosuke Zenshu,* 6: 161–219. Tokyo: Sanseido.

Kirikae, H. 1994. Pa/ca correspondence between Ainu dialects: A linguistic-geographical study. *The proceedings of the 8th International Abashiri Symposium: Peoples and cultures of the boreal forest*, 8. Abashiri: Hokkaido Museum of Northern People.

Komatsu, K. (ed.) 2004. *Ainugo Samani hōgen jiten*, *Kaiteiban [The Samani dialect of Ainu, revised edition]*. Samani: Samani Ainugo Kyōshitsu.

Krascheninnikov, S. P. unknown. Vocabularium Latino-Curilice. Reprinted: *Kitachishima Ainugo* [*Northern Kuril Ainu*], 11–20, Murayama, S., 1971. Tokyo: Yoshikawa Kobunkan.

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

Pinart A. L. 1872. Vocaburlary in Russian and Ainu dialect of the Paramushir and Limushir Islands in the Kuriles. Transcribed and Translated: Murayama, S. 1971. *Kitachishima Ainugo [Northern Kuril Ainu]*, 134–244. Tokyo: Yoshikawa Kobunkan.

Sato, T. 2014. A Study of the Ainu vocabulary collected in 1704 by the Venerable Kunen. *Bulletin of the Hokkaido Ainu culture research center*, 20: 1–133.

Sato, T. 2016. Gengoshi no kenkyūhōhō to Ainugoshi no shomondai [Method for studying the language history and issues in the history of Ainu language]. *Historical linguistics of Japan,* 5: 79–89.

Tamura, S. 1974. Verb Suffixes *-no* and *-nu* in the Saru dialect of Ainu, *Gengo Kenkyu*, 65: 24–36.

Tamura, S. 1996. *Ainugo Saru hōgen 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.

Uwano, Z. 2017. On Hattori's Hypothesis: Does Shortening of Long Vowels Produce an Accent Kernel?. *Journal of Asian and African Studies*, 94: 343–363.

Vovin, A. 1993. *A Reconstruction of Proto-Ainu*. Leiden: E. J. Brill.

A Geolinguistic Analysis of "Rain" and "Fish" in the Western Sichuan Ethnic Corridor Languages

Satoko Shirai

Japan Society for the Promotion of Science/University of Tsukuba

Abstract

This study discusses the significance of geolinguistic approaches to the languages of the Ethnic Corridor in Western Sichuan with case studies. I first point out the difficulties of using comparative methods to investigate the languages of the target area. This is probably because long-term language contact formed a language area in which the languages share typological features. Second, I examine the words for "rain" and "fish" in the languages of the area from a geolinguistic perspective. As for "rain," two different roots show clear traces of areal diffusion: they are distributed in the north and south, respectively; moreover, compounds consisting of both morphemes are found in the intermediate area. Moreover, some of the sound correspondences are found to be rather geographically organized, which is difficult to interpret from the viewpoint of comparative linguistics. As for "fish," the distribution of lexical variations is almost consistent with a genealogical grouping. However, the sound correspondences can be best interpreted through a geolinguistic approach: peripheral spots have conservative initials but drop the glide. This study concludes that it is important to conduct a geolinguistic analysis in addition to comparative studies to understand the historical process forming this multilingual area.

1 Introduction

There is a multi-ethnic area called the "Ethnic Corridor" in western Sichuan (川西民族走廊; the Western Sichuan Ethnic Corridor, hereinafter abbreviated as WSEC),[1] which extends from northwestern Yunnan to southern Gansu (See Map 1 in Section 2). The area was first mentioned in Fei (1980), and later named by Sun (1983). From a linguistic perspective, about 30 varieties of Tibeto-Burman (TB) minority languages are spoken in this area (cf. Roche and Suzuki 2017, Sun 1982). In this study, I will refer to these languages collectively as the WSEC languages.

At present, however, the detailed genetic status and classification of the WSEC languages remains controversial, suggesting limitations of the comparative method. Although they share plenty of cognates and similar typological characteristics (cf. Sun 1982), it is difficult to find shared phonological innovations in these languages (Chirkova 2012).[2] One of the reasons for the difficulty in comparative investigations of this area is language contact between genetically close languages. For example, loanwords are found even in very basic words, such as pronouns (Shirai 2018).

In this situation, the geolinguistic method is expected to contribute to the investigation, demonstrating the historical processes underlying the formation of this multilingual area. In this study, I first survey recent comparative studies of the WSEC languages (Section 2), and second conduct a

[1] The WSEC area is almost equivalent to central part of the Tibeto-Lolo Corridor (藏彝走廊) (Shi ed. 2009) or Eastern Tibetosphere (Roche and Suzuki 2017).

[2] Matisoff (2004) points out a quasi-common sound change in Qiangic, which he calls "brightening": Proto-Tibeto-Burman *a tends to be raised or fronted, and typically becomes /i/ in Qiangic languages. Chirkova (2012: 138) mentions that brightening is "the only (phonological) innovation for the Qiangic subgroup proposed so far."

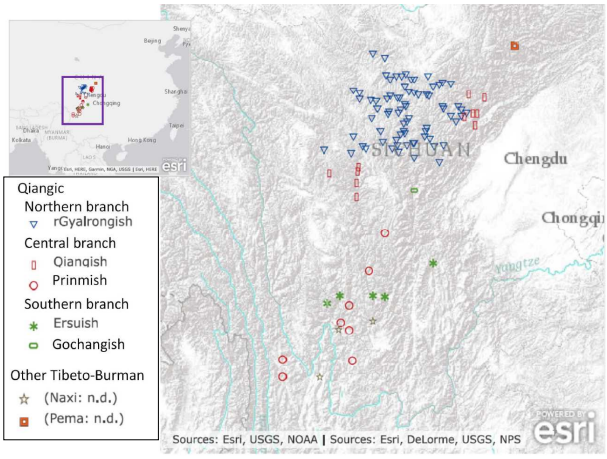
geolinguistic analysis of words for “rain” and “fish” as case studies (Section 3 and 4). Discussions are summarized in Section 5. The sources of language data in this study are listed at the end of this paper.[3]

2 Previous studies: Genealogical studies of the WSEC languages

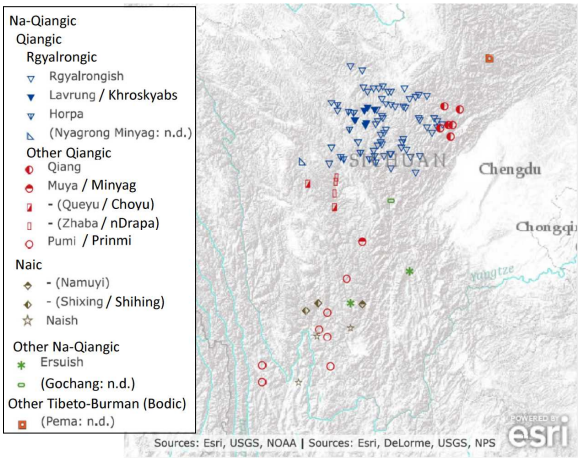
The detailed genealogical status of the WSEC languages (except for Pema, which is a Tibetic language) is still being discussed. Some studies regard the WSEC languages for the most part as a unit of a genetic group, the “Qiangic” branch of TB (Matisoff 2003, Sun 1983, 2016, etc.). Other studies regard them close to Lolo-Burmese (Jacques and Michaud 2011, Nishida 1987), Jinghpaw (Sun 2001), or languages such as those spoken in the West Himalayas that have a hierarchical person marking system (Thurgood 1985). Additionally, some studies, such as that of Chirkova (2012), are critical of arguments for treating the “Qiangic” group as a genetic unit.

Genealogical subgrouping of WSEC languages is controversial as well, except that a consensus on the rGyalrongic group, which includes rGyalrong varieties,[4] Khroskyabs, sTau, and so on, has almost been reached (Gates 2012, Sun 2000, Suzuki 2012, etc.). Here, I introduce two recently proposed hypotheses: Sun (2016: 4) and Jacques and Michaud (2011: Appendix p. 6).

Sun (2016: 4) takes a conservative view. He divides the Qiangic languages into three groups: Northern (= rGyalrongic), Central, and Southern. Sun’s (2016) Central Qiangic consists of Qiangish (Qiang, nDrapa, Choyu), Prinmish (Prinmi and Minyag), and Tangut; and Southern Qiangic consists of Ersuish (Shihing, Namuyi, and Ersu) and Gochang. However, Jacques and Michaud (2011) propose a considerably different subgrouping. They claim that Qiangic languages, together with Naish languages, belong to the Na-Qiangic subbranch of the Burmo-Qiangic branch of Tibeto-Burman. The Na-Qiangic subbranch is divided into three groups: Qiangic (the Northern and Central groups in H. Sun 2016), Ersuish, and Naic (H. Sun’s (2016) Southern group is divided into the two latter groups). Jacques and Michaud’s (2011) Qiangic consists of the rGyalrongic (Rgyalrongic) group and the following languages: Qiang varieties, Minyag varieties, Prinmi varieties, Choyu, nDrapa, and Tangut.



Map 1: The WSEC languages based on H. Sun’s (2016) classification



Map 2: The WSEC languages based on Jacques and Michaud’s (2011) classification

Maps 1 and 2 illustrate the WSEC languages’ geographical distribution with the genealogical groupings defined by H. Sun (2016) and Jacques and Michaud (2011), respectively.[5] Points of difference are found in (i) Southern Qiangic vs. Naic and Ersuish; and (ii) Qiangish and Prinmish of Central Qiangic vs. non-rGyalrongic (not subgrouped) languages in Qiangic. On the first point, I

[3] H. Suzuki advised me on language identification and their names.

[4] rGyalrong is sometimes used as a language name; however, it involves certain varieties that are mutually unintelligible (Situ, Japhug, etc.). A parallel situation is found in other “languages” such as Qiang and Prinmi. In such cases, I use the term “variety,” which may mean either language or dialect.

[5] Tangut is not discussed in this study, since it is a dead language.

tentatively follow Jacques and Michaud's (2011) classification of Naic and Ersuish, since they mainly focus on the Naic languages. Thus, the second point becomes the main point of interest of this paper: a geolinguistic approach to "Central Qiangic."

3 A geolinguistic analysis of "rain"

3.1 Classification of types

In this section, I will introduce an example that demonstrates the significance of the geolinguistic approach to understanding the historical process forming this multilingual area. Words for "rain" (noun) in the WSEC area may be classified into the following types. Representative examples are shown in Table 1.

[A] **r-məw* type: derived from Proto-Tibeto-Burman (PTB)[6] **r-məw* 'sky / heavens / clouds' (#2473)
[A'] **r-məw*-compound type: Compounds involving a morpheme derived from **r-məw*, except for [A+B].
[B] **r/s/g-wa* type: derived from PTB **r/s/g-wa* 'water / rain' (#2080)
[B'] **r/s/g-wa*-compound type: Compounds involving a morpheme derived from **r/s/g-wa*, except for [A+B].
[A+B] **r-məw* + **r/s/g-wa* type: Compounds consisting of **r-məw* and **r/s/g-wa.*
[C] Initial affricate types:
(i) **tshyar* 'rain (n.)' (#5902): Tibetan loan (WT *char pa*)
(ii) **tshyar* + *s-nak ('black'): Compound consists of Tibetan loan morphemes (WT *char nag*)
(iii) TS-type: With dental affricate or fricative initial, which are different from Tibetan; apparently similar to PTB **tshyar* or **(d)zil* 'dew' (#59), though the etymology is unclear.
[D] Other types.

Table 1: "Rain" in WSEC languages

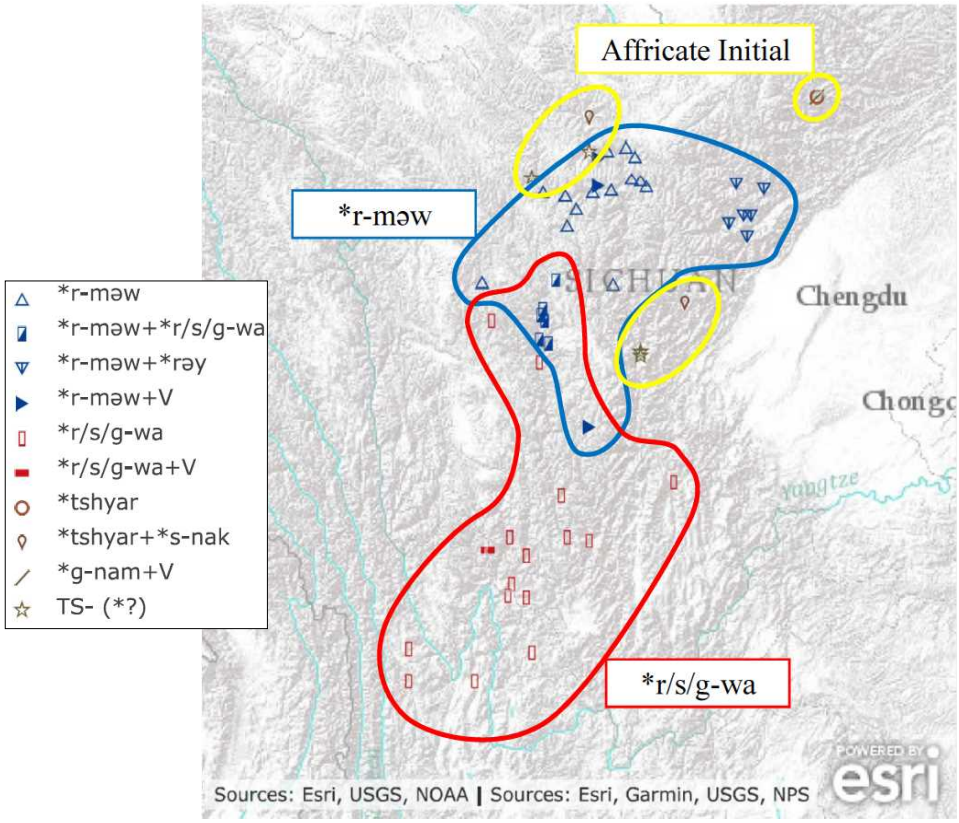
Language (*Group*)[7]	**'rain' [Type]**	**Language (*Group*)**	**'rain' [Type]**
Rongan Zbu (*G*)	tɕʰə'rnak [C-ii]	Nyato nDrapa (*Q*)	mə^{55}ku^{33} [A+B]
Shaerzong Japhug (*G*)	tɯmɯ [A]	Gala Choyu (*Q*)	hu [B]
Ribu Zbu (*G*)	təɾ'zi [C-iii]	Minyag (*Q*)	mə^{55}qhɐ̱^{33} [A']
Kangshan Tshobdun (*G*)	təmə'notoʔ [A']	Qinghua Southern Prinmi (*Q*)	gui^{55} [B]
Geletuo Zbu (*G*)	gə'rdzi [C-iii]	Upper Shihing (*N*)	xu^{53} [B]
Muerzong Khroskyabs (*G*)	'mənatos [A']	Shihing (*N*)	ɸui̱^{55}zɑ^{55} [B']
Yelong Khroskyabs (*G*)	mu^{53} [A]	Luobo Namuyi (*N*)	(ʔ)hĩ^{53} [B]
Geshitsa sTau (*G*)	mɑ [A]	Yongning Na (*N*)	hi^{11} [B]
Yoci bTsanlha (*G*)	tʃʰan^{44}nak^{44} [C-ii]	Western Naxi (*N*)	xɯ^{21} [B]
Daofu sTau (*G*)	mə qhi [A+B]	Laze (*N*)	fi^{33} [B]
Yadu Northern Qiang (*Q*)	me:ɹ [A']	Maibeng Gochang (*E*)	tshõ^{31} [C-iii]
Mawo Northern Qiang (*Q*)	məʁeɹ [A']	Lizu (*E*)	ɣuæ^{35} [B]

[6] The reconstructed PTB forms in the present paper are taken from the *Sino-Tibetan Etymological Dictionary and Thesaurus* (STEDT). The numbers with a pound sign indicate the root identification numbers in STEDT.

[7] Groups are indicated as follows: (*G*) - rGyalrongic; (*Q*) - Central Qiangic (Qiangic except for rGyalrongic); (*N*) - Naic; (*E*) - Ersuic.

3.2 Geographical distribution and interpretation

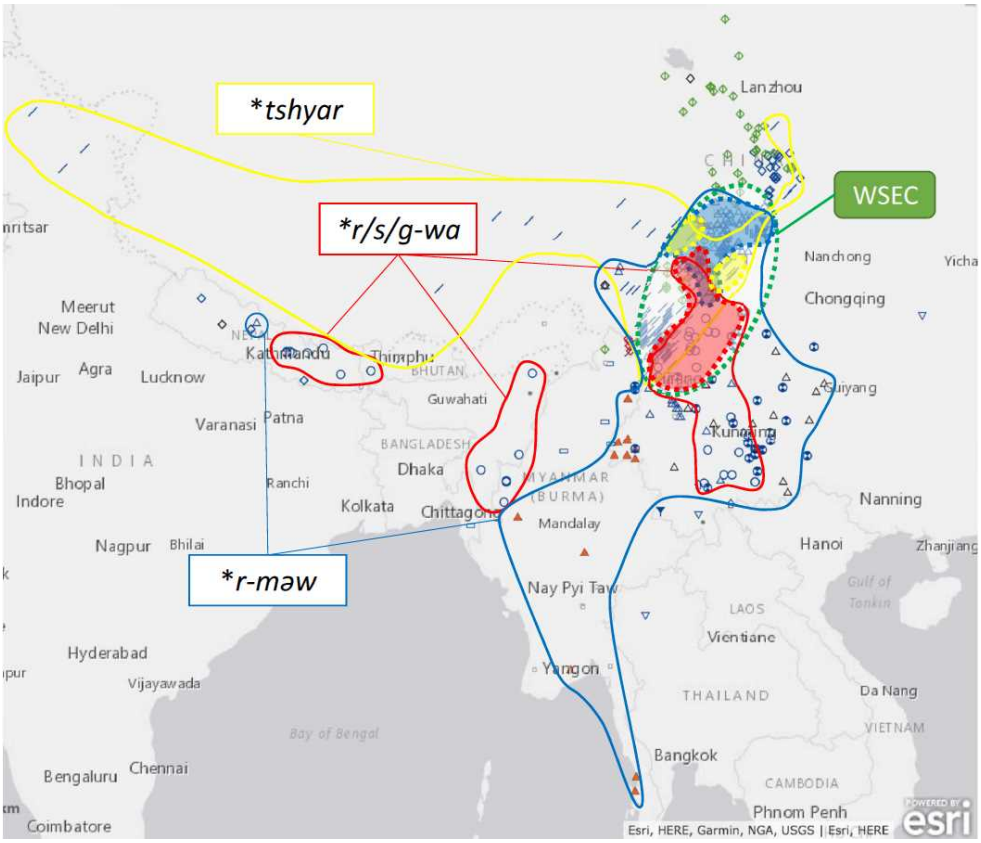
Map 3 illustrates the geographical distribution of words for "rain." Blue marks indicate [A] **r-məw* type and compound types with **r-məw*. Rectangles indicate [B] **r/s/g-wa* type and compound types with **r/s/g-wa*. Note that the blue rectangle indicates [A+B] compounds consisting of **r-məw* and **r/s/g-wa*. We find that the blue marks (**r-məw*) are clearly concentrated in the northern area (rGyalrongic varieties, Qiang, etc.), while the rectangles (**r/s/g-wa*) are mainly distributed in the southern area (Naic, Ersuish, Prinmi, etc.). The blue rectangles (**r-məw* + **r/s/g-wa*) are found between the two areas (nDrapa and sTau).



Map 3: "Rain" in WSEC

To interpret the relative chronology of **r-məw* and **r/s/g-wa* from a geolinguistic viewpoint, we need to refer to their wider distributions; see Map 4. From the perspective of Tibeto-Burman (see Shirai et al. forthcoming), the **r-məw* type is shared by many Lolo-Burmese varieties and widely found in southwestern China and Indochina. **r/s/g-wa* is also widely scattered among the Tibeto-Burman languages of Nepal, the India-Myanmar border, and southwestern China. Thus, the WSEC is the northeastern limit of both types. Moreover, in southwestern China including the WSEC area, **r-məw* has a wider distribution than **r/s/g-wa*. This fact suggests a relative chronology as follows: In the WSEC area, the **r-məw* type is older. Later, **r/s/g-wa* entered from the southern area. Afterward, compounds containing both **r-məw* and **r/s/g-wa* were formed in the area where the two types met.

Although the initial affricate types might appear to be older than **r-məw* from the distribution shown in Map 3, I do not adopt this hypothesis. The initial affricate types are apparently related to the influence of Tibetan, which is the traditional lingua franca in this area. As Sibata (1969/1977: 18) points out, "the standard language spreads like the rainfall [among local dialects]," thus, such words as Tibetan loanwords may be distributed sporadically.



Map 4: "Rain" in Tibeto-Burman and the location of WSEC

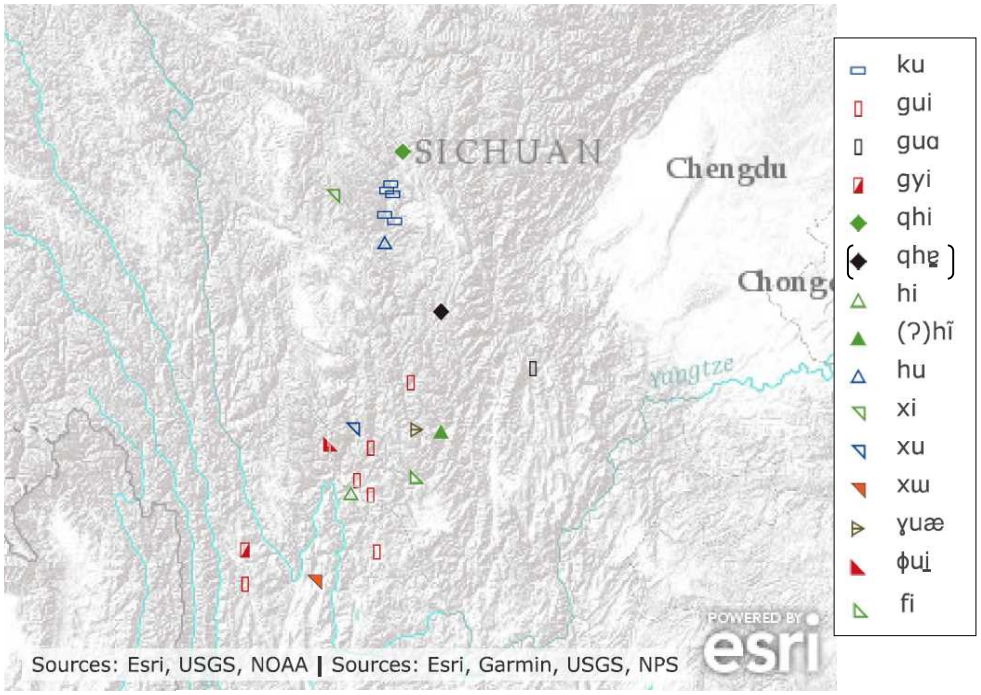
3.3 A geolinguistic analysis of the variants of **r/s/g-wa*

Next, I will examine nominal morphemes derived from **r/s/g-wa*, which are widely found in the Central Qiangic languages except for Qiang and Minyag[8] (that is, nDrapa, Choyu, and Prinmi varieties). Map 5 shows the forms' geographical distribution. Shapes and colors distinguish initial consonants and rhymes, respectively.

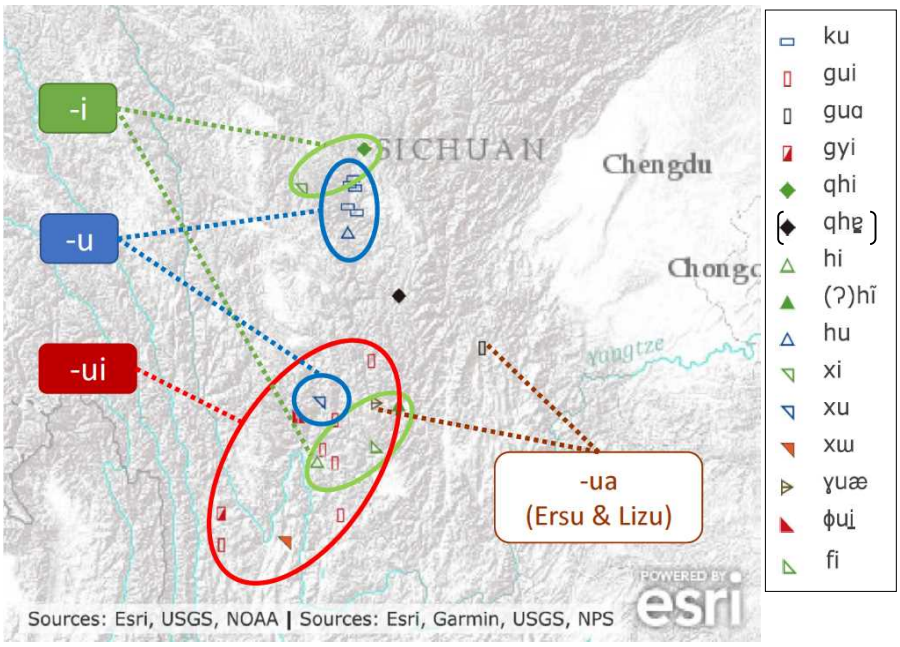
The initial consonants in the nDrapa dialects and Daofu sTau are voiceless plosives. The Prinmi varieties and Ersu have voiced plosives, while the Choyu dialects, Naic languages, and Lizu have fricatives (marked by triangles). Note that Ersu and Lizu are genetically close but differ in the initial consonants of this morpheme (Ersu *gua*^{33}; Lizu *ɣuæ*^{35}). Since Lizu is spoken among Naic languages, we can assume that it has undergone Naic influence.

Rhymes are distributed more geographically. As illustrated in Map 6, close diphthongs (*-ui*) are distributed in the southern area, back vowels (*-u*) are found in the central area, and front vowels (*-i*) occur around the area of back vowels. Here, Ersu and Lizu share the same type (diphthongs of a back vowel and an open vowel). Note that two dialects of Choyu have different rhymes (Youlaxi Choyu *xi*^{55}, Gala Choyu *hu*).

[8] I exclude from the discussion here the second syllable of *mə*^{55}*qhɐ̲*^{33} 'rain' in Minyag, which comes from a verb meaning 'to precipitate', although it may have been derived from PTB root **r/s/g-wa*. This is because in some cases verbs cannot be compared straightforwardly with nouns because of different morphology.



Map 5: **r/s/g-wa* in WSEC



Map 6: **r/s/g-wa* in WSEC: Rhymes

From the perspective of comparative linguistics, this morpheme's Proto-Central Qiangic form may be reconstructed as **gui*.[9] The following chronology could be hypothesized: The initial **g-* was preserved in Prinmish but was devoiced in other Central Qiangic varieties (nDrapa and Choyu). Afterward, it was fricativized in Choyu. The change of rhymes into single vowels occurred rather recently, since Choyu dialects have different vowels. The hypothesis based on the initials seems to be consistent with H. Sun's (2016) classification that distinguishes Prinmish from Qiangish—which

[9] It is broadly found that **a* in PTB changed into a front/close vowel in Qiangic (Matisoff 2004: 329).

includes nDrapa and Choyu—of the Central Qiangic branch. However, as the geographical distribution of rhyme types suggests, another dimension of change may have occurred geographically, that is, under language contact situations.

3.4 Summary

The points of this case study can be summarized as follows: a geolinguistic approach is appropriate for clarifying the distribution of different roots. Remember that, for example, Japhug (rGyalrongic) and Northern Qiang ((Central) Qiangic) are generally classified into different subgroups of Qiangic, while Northern Qiang and Choyu are classified into the same subgroup in Sun's (2016) classification: Central Qiangic. However, Northern Qiang has a different root from Choyu but shares **r-məw*-type words with Japhug. The **r-məw* type is found in the northern area where Northern Qiang and Japhug are spoken. These facts suggest that different roots have diffused in the context of language contact. Moreover, difficulties with certain sound changes may be solved through a geolinguistic approach, although in many cases a comparative linguistic approach is necessary to understand the phonological changes of shared roots.

4 A geolinguistic analysis of "fish"

4.1 Classification of types

"Fish" is one of a small number of cognates that are commonly found in most "Qiangic" languages. STEDT provides the Proto-Qiangic root **r-dzwa* (#5673),[10] which could have been derived from PTB **s-ŋya* (#1455).

Words for "fish" in WSEC are classified into the following types. Representative examples are shown in Table 2.

[A] **r-dzwa* type (Qiangic type)
[A'] **r-dzwa*-compound type: Compounds involving a morpheme derived from **r-dzwa*
[B] **s-ŋya* type (non-Qianic type): either Tibetan loan or cognate
[B'] **s-ŋya*-compound type: Compounds involving a morpheme derived from **s-ŋya*

Table 2: "Fish" in WSEC languages

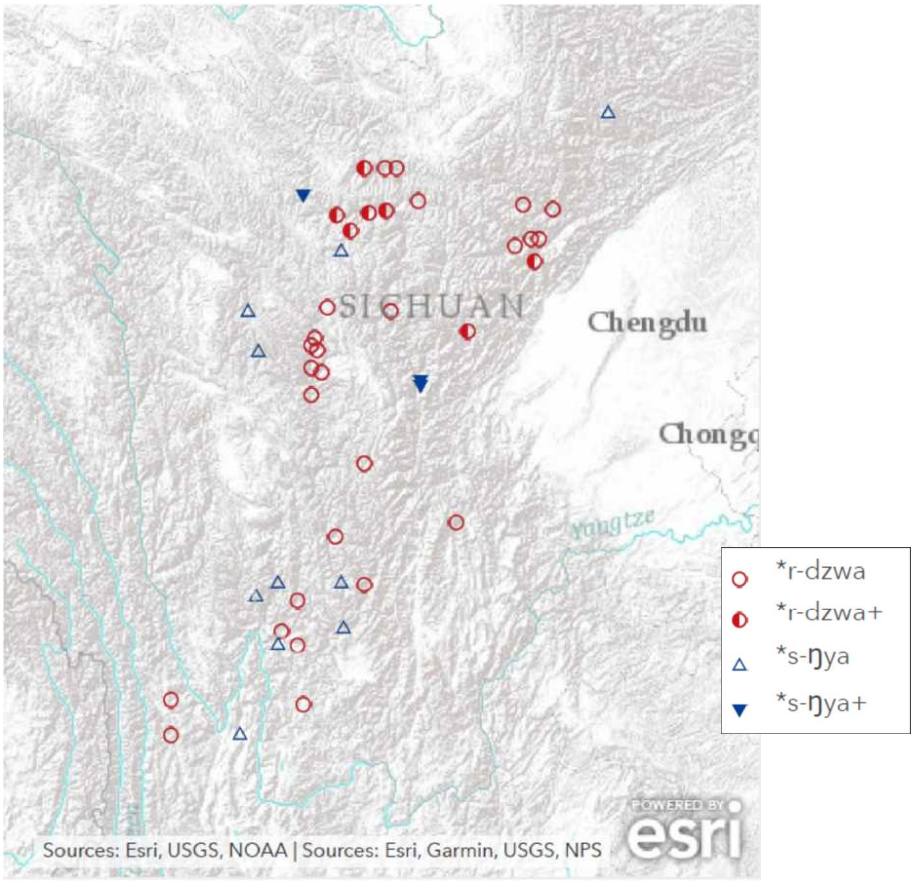
Language (*Group*)[11]	**"fish" [Type]**	**Language (*Group*)**	**"fish" [Type]**
Ganmuniao Japhug (*G*)	qaɟy [A']	Longxi Southern Qiang (*Q*)	ʁà [A]
Ribu Zbu (*G*)	ɟə'xju [A']	Mätro nDrapa (*Q*)	ɦdʑɯʒ [A]
Tshobdun (*G*)	qa^{33}gjüe^{44} [A']	Youlaxi Choyu (*Q*)	ɳa^{13} [B]
Geletuo Zbu (*G*)	ɳa'mu [B']	Minyag (*Q*)	ʁuə̠^{53} [A]
Guanyinqiao Khroskyabs (*G*)	ʁdə^{33}ju^{55} [A']	Qinghua Southern Prinmi (*Q*)	dʒə^{55} [A]
Wobzi Khroskyabs (*G*)	ʁdójú [A']	Upper Shihing (*N*)	ʔu^{55} [B]
Nyagrong Minyag (*G*)	´ɳa [B]	Shihing (*N*)	ʔo^{55} [B]
Geshitsa sTau (*G*)	ɣjə [A]	Luobo Namuyi (*N*)	zu^{55} [A]
Yoci bTsanlha (*G*)	tʃow^{44}jɯ^{22} [A']	Yongning Na (*N*)	ɲiɬzo#˥ [B']
Daofu sTau (*G*)	ʁjə [A]	Western Naxi (*N*)	ɳi^{33} [B]
Yadu Northern Qiang (*Q*)	ʁzə [A]	Maibeng Gochang (*E*)	tʃə55ni55 [A']
Mawo Northern Qiang (*Q*)	ʁdzə [A]	Lizu (*E*)	ɳæ^{35} [B]

[10] "This recently reconstructed root has so far not been found outside of Qiangic / rGyalrongic." (STEDT, https://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl/etymon/5673, accessed on June 20, 2018.)

[11] Group names are abbreviated as follows: (*G*) - rGyalrongic; (*Q*) - Central Qiangic (Qiangic except for rGyalrongic); (*N*) - Naic; (*E*) - Ersuic.

4.2 Geographical distribution and interpretation

Map 7 illustrates the geographical distribution of words for "fish." The plus mark in the legend indicates a compound type. Qiangic types ([A] and [A'] above) are in red, while non-Qiangic types ([B] and [B']) are in blue. The former types are found all around the area. Most non-Qiangic languages have the latter type: Pema, Ersuish, and Naic (except for Namuyi). Moreover, a limited number of Qiangic languages such as Geletuo Zbu, Nyagrong Minyag, and Youlaxi Choyu have non-Qiangic type words. These words, that is, **s-ŋya* in Qiangic, are loanwords from Tibetan. Note that the spots of Qiangic languages with **s-ŋya* are concentrated in the northwest of the area, which is closer to such Tibetan cultural centers as Derge.

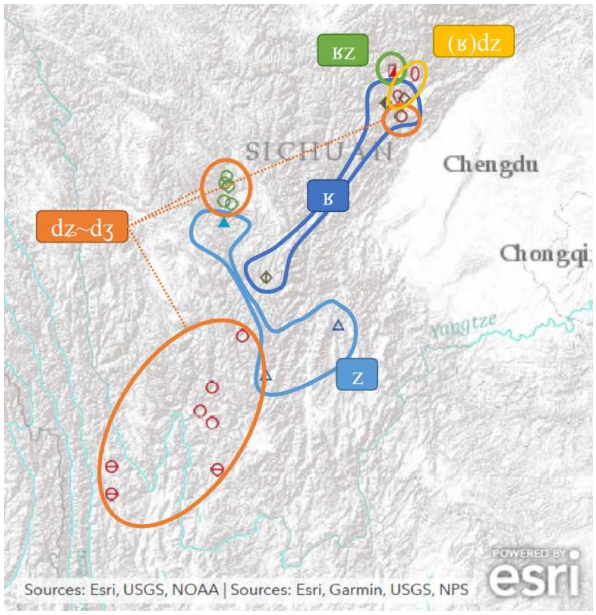


Map 7: "Fish" in WSEC

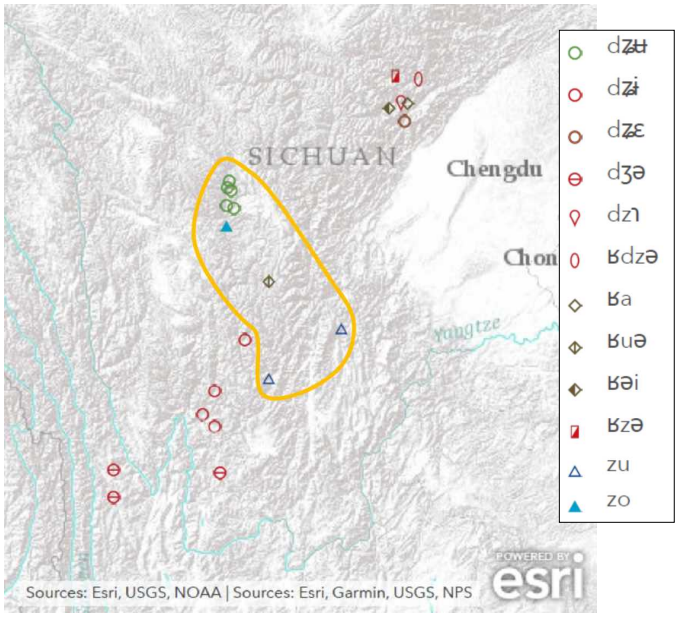
I will now examine the forms reflecting **r-dzwa* in Qiangic languages except for rGyalrongic, that is, "Southern" and "Central" Qiangic in Sun (2016). Maps 8 and 9 illustrate the geographical distribution of the forms derived from **r-dzwa*. The two maps indicate the distribution of initials and rhymes respectively, based on the same map.

As shown in Map 8, on the one hand, affricates are found in the southern and peripheral spots, while the central spots have fricatives. This indicates that the lenition of initials occurred in the central area. On the other hand, the central spots enclosed in the yellow line in Map 9 show roundedness in the rhymes. This suggests that they contain reflexes of a glide -**w*- that was dropped earlier in the peripheral spots.

The abovementioned facts suggest that**r-dzwa* spread among the Qiangic languages at a very early stage; afterwards, separate sound changes occurred within this geographical context.



Map 8: **r-dzwa*: Initials



Map 9: **r-dzwa*: Rhymes

4.3 Summary

In this case, lexical varieties are found almost entirely consistently to follow the genealogical groupings: most Qiangic languages have **r-dzwa*, while most non-Qiangic languages of WSEC (Pema, Naic, and Ersuic) do not. However, the sound correspondences of the forms derived from **r-dzwa* show areal diffusion: Affricate initials are preserved in the peripheral spots; central spots show reflexes of a glide -**w*-. Although Sun (2016) classified nDrapa (spoken in the central area) into the same subgroup with Qiang varieties (spoken in the northeastern area), they do not show a common sound change in this case.

5 Conclusion

In this study, I first pointed out the difficulties of using the comparative method alone to investigate the WSEC languages. Although only a small number of cognates are shared by all or most members, a long-term language contact situation may have formed a language area in which the languages share typological features.

Secondly, I examined the words for "rain" and "fish" in various WSEC languages from a geolinguistic perspective. As for "rain," two different roots show a clear trace of areal diffusion: **r-məw* and **r/s/g-wa* are distributed in the north and south, respectively, and compounds containing both morphemes are found in the intermediate area. Moreover, the sound correspondences of the rhymes of **r/s/g-wa* show geographical influences that are difficult to interpret from the viewpoint of comparative linguistics. The target of another case study, "fish," is one of only a small number of cognates shared by most of "Qiangic" languages. The distribution of lexical variations is almost entirely consistent with the genealogical grouping. However, the sound correspondences of this root can be better interpreted through a geolinguistic approach: peripheral spots have conservative initials but dropped the glide.

This study led me to conclude that it is important to conduct a geolinguistic analysis in addition to comparative studies to demonstrate the historical processes that formed this multilingual area.

Acknowledgements

This work is a product of the ILCAA joint research project "Studies in Asian Geolinguistics." This work was supported by KAKENHI Grant-in-Aid for JSPS Fellows (17J40087).

Data sources

Chirkova (2008) - Pema; Duo'erji (1998) - sTau (Geshitsa); Evans (2001) - Southern Qiang (Longxi and Mianchi); Gong (2007) - nDrapa (Waduo); B. Huang (2007) - Khroskyabs (Yelong and Guanyinqiao); Huang (editor-in-chief) (1992) - sTau (Daofu), nDrapa (Zhatuo), Choyu (Youlaxi), Gochang (Maibeng), Minyag, Upper Shihing, Lizu, Namuyi (Luobo), and Western Naxi; C. Huang (2007) - Southern Qiang (Puxi); Jacques (2008) - Japhug (Ganmuniao); Jacques and Michaud (2011) - Laze; Jiang (2015) - Gochang (Qianxi); LaPolla (2003) - Northern Qiang (Yadu); Liu (1998) - Northern Qiang (Mawo); Lu (2001) - Prinmi (Sanyanlong, Taoba, Tuoqi, Zuosuo, Ludian, Xinyinpan, and Qinghua); Michaud (2015) - Na (Yongning); Nagano and Prins (2013) - Japhug (Longerjia and Shaerzong), Zbu (Rongan, Ribu, Geletuo, and Zongke), Tshobdun (Caodeng and Kangshan), Situ (Songgang and Zhuokeji), Khroskyabs (Muerzong and Wobzi), sTodsde (Puxi), Choyu (Gala), Erkai, and Nyagrong Minyag; S. Shirai (ms.) Fieldnotes - Situ (Mbola), bTsanlha (Yoci), and nDrapa (Mätro, Kalakhe, and Nyato); Sun et al. (2014) - Shihing; Zangmianyu Yuyin he Cihui Bianxiezu (ed.) (1991) - Southern Qiang (Taoping, via Evans 2001) and Ersu (via STEDT).

References

Chirkova, K. 2008. On the Position of Baima within Tibetan: A Look from Basic Vocabulary. In: A. Lubotsky, J. Schaeken, and J. Wiedenhof (eds.) *Evidence and Counter-Evidence: Festschrift for F. Kortland, Volume 2: General Linguistics*. pp. 69-91. Amsterdam: Rodopi.

Chirkova, K. 2012. The Qiangic Subgroup from an Areal Perspective: A Case Study of Languages of Muli. *Language and Linguistics* 13(1): 133-170.

Duo'erji [多爾吉]. 1998.《道孚語格什扎話研究》[*A Study of Geshizha sTau*]. Beijing: Zhongguo Zangxue Chubanshe.

Evans, J. P. 2001. *Introduction to Qiang Phonology and Lexicon: Synchrony and Diachrony.* Tokyo: ILCAA, Tokyo University of Foreign Studies.

Fei, X. [費孝通]. 1980.〈関于我國民族識別問題〉[About National Identification in our Country]. *Zhongguo Shehui Kexue* 1980(1): 94-107.

Gates, J. P. 2014. *Situ in Situ: Towards a Dialectology of Jiāróng (rGyalrong)*. München: LINCOM GmbH

Gong, Q. [龔群虎]. 2007.《扎巴語研究》(中國新發現語言研究叢書）[*A Study of the nDrapa Language*]. Beijing: Minzu Chubanshe.

Huang, B. [黄布凡]. 2007.《拉烏絨語研究》(中國新發現語言研究叢書）[*A Study of Khroskyabs*]. Beijing: Minzu Chubanshe.

Huang, B. (editor-in-chief) [黄布凡(主編)]. 1992.《藏緬語族語言詞匯》[*A Tibeto-Burman Lexicon*]. Beijing: Zhongyang Minzuxueyuan Chubanshe.

Huang, C. [黄成龍]. 2007.《浦溪羌語研究》(中國少數民族語言方言研究叢書）[*A Study of Puxi Qiang*]. Beijing: Minzu Chubanshe.

Jacques, G. [向柏霖]. 2008.《嘉絨語研究》(中國新發現語言研究叢書）[*A Study of rGyalrong*]. Beijing: Minzu Chubanshe.

Jacques, G. and A. Michaud. 2011. Approaching the Historical Phonology of Three Highly Eroded Sino-Tibetan Languages. *Diachronica* 28: 468–498 (With online appendix, DOI: 10.1075/dia.28.4.02jac.additional).

Jiang, L. 2015. *A Grammar of Guiqiong: A Language of Sichuan.* (Languages of the greater Himalayan region). Leiden/Boston: Brill.

LaPolla, R. J. with C. Huang. 2003. *A Grammar of Qiang.* Berlin/New York: Mouton de Gruyter.

Liu, G. [劉光坤]. 1998.《麻窩羌語研究》(中國少數民族語言方言研究叢書）[*A Study of Mawo Qiang*]. Chengdu: Sichuan Minzu Chubanshe.

Lu, S. [陸紹尊]. 2001.《普米語方言研究》(中國少數民族語言方言研究叢書）[*A Study of Prinmi Dialects*]. Beijing: Minzu Chubanshe.

Matisoff, J. A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction.* Berkeley/Los Angeles: University of California Press.

Matisoff, J. A. 2004. "Brightening" and the Place of Xixia (Tangut) in the Qiangic Branch of Tibeto-Burman. In: Y.-C. Lin et al. (eds.). *Studies on Sino-Tibetan Languages: Papers in Honor of Professor Hwang-Cherng Gong on his Seventieth Birthday.* Language and Linguistics Monograph Series No. W-4. 327–352. Taipei: Institute of Linguistics, Academia Sinica.

Michaud, A. 2015. *Online Na-English-Chinese Dictionary* (version 1.0). <http://lacito.vjf.cnrs.fr/pangloss/dictionaries/na/dictionary_eng_mp3.pdf, accessed on June 20, 2018>.

Nagano, Y. and M. Prins. 2013. *rGyalrongic Languages Database.* <http://htq.minpaku.ac.jp/databases/rGyalrong/, accessed on June 20, 2018>.

Nishida, T. [西田龍雄]. 1987.「チベット語の変遷と文字」[The Tibetan Language: Transition and Scripts]. In: 長野泰彦・立川武蔵(編)『チベットの言語と文化』, 108-169. Tokyo: Tojusha.

Roche, G. and H. Suzuki. 2017. Mapping the Minority Languages of the Eastern Tibetosphere. In: K. Kurabe and M. Endo (eds.) *Studies in Asian Geolinguistics VI: "Means to Count Nouns" in Asian Languages*, 28-42. Tokyo: ILCAA, Tokyo University of Foreign Studies.

Shi, S. (editor-in-chief) [石碩 (主編)]. 2009.《藏彝走廊:文明起源与民族源流》[*Tibeto-Lolo Corridor: The Origin of Civilization and Nation*]. Chengdu: Sichuan People's Publishing House.

Shirai, S. 2018. The Possibility of Borrowing Basic Pronouns in Minority Languages of the Western Sichuan Ethnic Corridor. *Tokyo University Linguistic Papers* 39: 265-285.

Shirai, S., H. Suzuki, and K. Kurabe. forthcoming. Semantic Shifts in Expressions for 'it rains' in Tibeto-Burman. In: *Studies in Asian Geolinguistics VIII.*

Sibata, T. [柴田武]. 1969/1977.『言語地理学の方法』[*Methods in Linguistic Geography*]. Tokyo: Chikuma Shobo.

STEDT. *The Sino-Tibetan Etymological Dictionary and Thesaurus.* <http://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl, accessed on June 20, 2018>.

Sun, H. [孫宏開]. 1982.〈羌語支属問題初探〉[A First Approach to the Classification of Qiangic Languages]. In: 民族語文編輯部(編)《民族語文研究文集》, 189-224. Xining: Qinghai Minzu Chubanshe.

Sun, H. [孫宏開]. 1983.〈川西民族走廊地区的語言〉[Languages of the Ethnic Corridor in Western Sichuan]. In: 李鋒銘(編)《西南民族研究》, 429-454. Chengdu: Sichuan Minzu Chubanshe. (Translated into English by Jackson Sun in *Linguistics of the Tibeto-Burman Area* 13(1): 1-31, 1990.)

Sun, H. [孫宏開]. 2001.〈論藏緬語族中的羌語支語言〉[On the Language of the Qiangic Branch in Tibeto-Burman]. *Language and Linguistics* 2(1): 157–181.

Sun, H. [孫宏開]. 2016.《藏緬語族羌語支研究》(国家哲学社会科学成果文庫)[A Study of the Qiangic Branch in Tibeto-Burman]. Beijing: China Social Sciences Press.

Sun, H., D. Xu, G. Liu, and Lurongduoding [孫宏開、徐丹、劉光坤、魯絨多丁]. 2014.《史興語研究》(中國新發現語言研究叢書)[*A Study of Shihing*]. Beijing: Minzu Chubanshe.

Sun, J. T.-S. 2000. Parallelisms in the Verb Morphology of Sidaba rGyalrong and Guanyinqiao in rGyalrongic. *Language and Linguistics* 1(1): 161–190.

Suzuki, H. 2012. Ergative Marking in Nyagrong-Minyag (Xinlong, Sichuan). *Linguistics of the Tibeto-Burman Area* 35(1): 35-48.

Thurgood, G. 1985. The 'Rung' Languages: Notes on their Proto-Morphosyntax and Subgrouping. *Acta Orientalia* 46: 79–99.

Zangmianyu Yuyin he Cihui Bianxiezu (ed.) [《藏緬語語音和詞彙》編寫組(編)]. 1991.《藏緬語語音和詞滙》[*Tibeto-Burman Phonology and Lexicon*]. Beijing: China Social Sciences Press.

Geolinguistic Approach to the Route of Tibetic Loanwords in Lhagang Choyu

Hiroyuki Suzuki[a] Sonam Wangmo[b]

[a]IKOS, University of Oslo / National Museum of Ethnology
[b]IKOS, University of Oslo

Abstract

This article discusses Tibetic loanwords in Lhagang Choyu, particularly their borrowing process, by referring to Choyu dialects and surrounding Tibetic languages (Minyag Rabgang Khams, Southern Route Khams, and pastoralists' Amdo). The analysis clarifies that Lhagang Choyu has a multi-strata feature of Tibetic loans and their archaic stratum principally reflects Amdo, especially the varieties spoken in Lithang County. This means that ancestors of Lhagang Choyu speakers had some contact with Amdo Tibetan spoken in Lithang.

1 Introduction

Choyu (also known as Queyu; ISO 639-3 code: qvy) is a Qiangic language spoken by less than 10,000 Tibetans in Nyagrong, Lithang and Nyagchukha counties, Kandze Prefecture, Sichuan Province, China. Recently, Suzuki & Sonam Wangmo (2016a) reported that there is one hamlet within Lhagang Town in Dartsendo Municipality where Tibetans used to speak a Choyu-like language known as Lhagang Choyu, and they (2017) provide a Lhagang Choyu word list with forms from Thamkhas Tibetan, a dialect substituting Lhagang Choyu. This language comprises many Tibetic loanwords. However, phonetic features evident in them are quite different from those of surrounding Tibetic languages. For this reason, we examine whether we can elucidate a route of lexical borrowing from Tibetic to Lhagang Choyu language, by comparing loanwords to the original word forms found in Choyu dialects and surrounding Tibetic languages and dialects (see Map 1 for their location).

The data that will be discussed comprises Tibetic loanwords in Lhagang Choyu (Thamkhas dialect; Suzuki & Sonam Wangmo 2017). We first compare them with those in five dialects of Choyu (Lhayul, Rongpa, gYanglagshis,[1] Phubarong, and Bezi[2]) and Lhagang Choyu (Thamkhas) to examine the differences in lexical forms and phonetic realisations. Second, we examine peculiar sound correspondences demonstrated in Tibetic loanwords in Lhagang Choyu compared to examples of surrounding dialects of the Tibetic languages,[3] Khams and Amdo.[4] All the data except for gYanglagshis and Bezi Choyu was collected by the present authors. gYanglagshis and Bezi were retrieved from Huang ed. (1992) and Sun ed. (1991) respectively.[5]

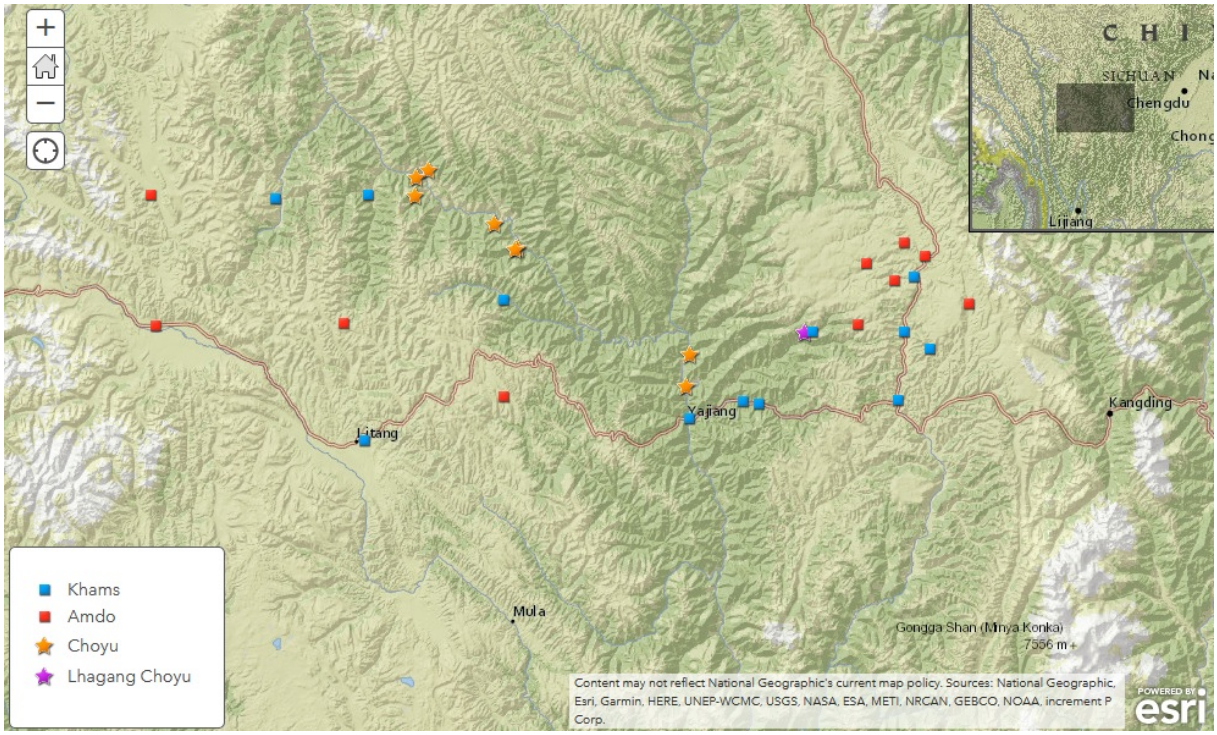
[1] See Wang (1991) for a concise description of gYanglagshis Choyu.

[2] See Lu (1985) for a short description of Bezi Choyu.

[3] See Tournadre (2014) for the definition of the term 'Tibetic'.

[4] Several vocabulary lists are on public resources. See Suzuki (2007) for Rangakha (Minyag Rabgang Khams), Suzuki & Sonam Wangmo (2015) for Lhagang (Minyag Rabgang Khams), Suzuki & Sonam Wangmo (2016c) for Shingnyag (Washul Amdo), and Suzuki & Sonam Wangmo (2017) for Thamkhas (Minyag Rabgang Khams).

[5] Bezi occupies an independent dialect branch among Choyu dialects. See Huang et al. (forthcoming).



Map 1: Location of relevant languages and dialects

Map 1 contains the following points [in order from the west to the east]:

Khams: Jowo, Lithang, Gyongpa, dGakhog, Nyagchukha, Bajiaolou (Riji), Bajiaolou, Thamkhas, mGologthog, Phagso, Lhagang, Balsrung,

Amdo: mChodrten (sDegzhongma[6]), Horra rNyingpa (gYonru), Tshonkhor (gYonru), Horlung (Othog), Shingnyag, Warnangsumdo, Nongskor, Goroma, rDorakarmo (rMewa)

Choyu: Lhayul (Gayibuli), Lhayul (Tshorong), gYanglagshis, Rongpa (Atsong), Phubarong, Bezi, Gala

Lhagang Choyu: Thamkhas

We have additional data on varieties of Khams and Amdo other than those shown on Map 1. However, the map only includes varieties that have had contact with Choyu and Lhagang Choyu. In addition, we note that non-Tibetic languages are mainly distributed mainly at the north of the Choyu-speaking region, such as nDrapa and Nyagrong-Minyag. See Roche and Suzuki (2017, 2018) and Shirai (this volume).

As Map 1 shows, Choyu and Lhagang Choyu are spoken in the mountainous area where the Nyagchu River flows through. Based on historical narratives of Lhagang Choyu-speakers, their ancestors came along this river from the western part of the Choyu-speaking region, making this a potential migration route (Suzuki & Sonam Wangmo forthcoming). Choyu-speakers are generally bilingual in Choyu and a local variety of Khams Tibetan. However, depending on the location of their communities, some might also have contact with Amdo Tibetan-speakers. Amdo-speakers' ancestors are also considered migrants; those living in Lithang County (in the west of Map 1) mostly came from the current Qinghai Lake area in Qinghai Province (Suzuki 2018a, Suzuki & Tsering Samdrup 2018), while those living around Lhagang Village (in the east of Map 1) mostly came from the current Palyul, northern Nyagrong, and Kandze counties (Suzuki & Sonam Wangmo forthcoming). Since we do not have any access to past detailed social situations of Choyu-speakers, an investigation of Tibetic loanwords in Choyu and Lhagang Choyu could aid in understanding their society to some extent.

[6] Regarding pastoralists' dialects of Amdo Tibetan, we can use another way of classification other than geographical location. See Tsering Samdrup & Suzuki (2017).

2 Examples of Tibetic loanwords in Lhagang Choyu and Choyu dialects

We present typical Tibetic loanwords in Lhagang Choyu (Thamkhas), five Choyu dialects (Lhayul, Rongpa, gYanglagshis, Phubarong, and Bezi), and Literary Tibetan (henceforth referred to as LT) in Table 1. The order is based on geographical location: from the west to the east.

Table 1: Tibetic loans in Choyu and Lhagang Choyu (n: native word; c: Chinese loan)

Meaning	Choyu/ Lhayul	Choyu/ Rongpa	Choyu/ gYanglagshis	Choyu/ Phubarong	Choyu/ Bezi	Lhagang Choyu/ Thamkhas	LT
axe	ˊⁿtsʰe /n	ˣⁿtsʰe /n	-	ˣᵉtsʰe /n	tse^{53} /n	ˣʰte ri	*sta re*
bean	ˣʱdʑa' rə mə	ˣᵐɳɛ /n	mɳye^{55} /n	ˣχõ tʰɥj /n	-	ˣʰsɛ ma	*sran ma*
book	ˣʱgə /n	ˣʱgwə /n	χpe^{55} tʃha^{55}	ˊʑi ke	dʑɯ^{35} dʑɯ^{35}	ˣʱgʷə /n	*yi ge/ dpe cha*
bridge	ˣʰtso /n	ˣʰtso /n	tso^{55} /n	ˣʰtso /n	dzã^{55}	ˊzã ᵐbe	*zam pa*
chicken	ˣʱdʑa ᵛʑa	ˊʱdʑa ᵛʑa	rdʑa^{13} bʑa^{55}	ˊʱdʑə ᵇʑwa	ʑa^{53}	ˊɸɕa	*bya*
cloth	ˊrɛ	ˊrja	riɛ^{13}	ˊri	re^{35}	ˊra	*ras*
copper	ˣxɯ rə /n	ˊrɑ	ra^{13} ɣɛ^{55}	ˊrɑ	ra^{35}	ˊzɔ̃	*zangs/ rag*
coral	ˊɸɕɯ rɯ	ˊɸɕə ʱdɥ	pɕi^{55} rdye^{33}	ˊɕɥ ʱdɥ	-	ˣɕɯ rɯ	*byu ru*
dragon	ˊᵐɖɥ	ˊᵐɖɥ	mdʑy^{13}	ˣⁿɖu	ndʑy^{53}	ˊⁿɖu	*'brug*
deity	ˣɭe	ˣɭi	-	ˣɭi	ɬi^{53}	ˣɭe	*lha*
flower	ˊmə tɥ	ˊmə tɥ	mu^{13} tye^{55}	ˣme tɥ	mo^{35} to^{53}	ˣmə to	*me tog*
forehead	ˣtʰə pə	ˣtʰə pə	thɛ^{55} pɛ^{55}	ˊtʰɛ ʱbə li	the^{55} pɐ^{55} li^{55}	ˊtʰa pe	*thod pa*
fox	ˣqa tʃwa /n	ˣᵖtʃwa /n	ptʃa^{55} /n	-	wa^{35}	ˊɣa	*wa*
glass	ˊxe	ˊxe ʱgu	-	ˣɕe ʱgu	-	ˣɕe	*shel*
gold	ˣˣsɛ	ˣŋ̊ə /n	ŋ̊ə^{55} /n	ˣɴaj	ŋ̊e^{55} /n	ˣʰsəˠ	*gser*
grandson	ˊŋo nə /n	ˊʑi	ʑi^{13} ɣə^{33} ʑi^{13} /n	ˣsu tsʰə /c	sə̃^{55} sə̃^{55} /c	ˣtsʰa wu	*tsha bo*
khatak	ˣkʰa dɑ	ˣkʰa dɑ	-	ˣqʰa dɑ	-	ˣkʰa te	*kha btags*
kidney	-	ˣʱ̊kʰe lə	rvu^{13} lo^{33} /n	ˊʱ̊kʰe lə	khe^{55} lø^{55}	ˣʱ̊kʰi ma	*mkhal ma*
lake	ˣᵐ̥tsʰo	ˣᵐ̥tsʰɥ	mtshy^{55}	ˣᵐ̥tsʰu	tshi^{55}	ˣᵐ̥tsʰu	*mtsho*
monastery	-	-	-	ˣʱgũ ᵐbi	-	ˣʱgõ ᵐbe	*dgon pa*
monk	ˣʱla ma	ˣʱla ma	pɛ^{13} ndi^{33}	ˣʱla' mo	-	ˣʷla me	*bla ma/ ban de*
new	ˣʰsa ʱbe	ˣʰsa pa	xsar^{55} pe^{55}	ˣsə ʱbi	se^{55} pi^{33}	ˣʰsa ʱbe	*gsar pa*
old	ˣʱɲi ʱbe	ˣʱɲi pa	ɲur^{13} pe^{55}	ˣʱɲĩ' ᵐbi	ɲẽ^{55} pi^{33}	ˣʱɲĩ ᵐbe	*rnying ba*
owl	-	ˣkʰu /n	khu^{55} lu^{55} xu^{33} /n	-	-	ˊɣu pa	*'ug pa*
power	ˣkʰɛ ɦõ	-	-	-	-	ˊʱʁʷɑ	*dbang*
rabbit	ˊrə vo	ˊrə kõ	ɬi^{13} /n	ˊʑi	ʑi^{35} ko^{55}	ˊrə ɣõ	*ri bong*
rice	ˣᵐɖɛ	ˣᵐɖɛ	mdʑiɛ^{13}	ˣᵐɖwa	ndʑɛ^{35}	ˣᵐɖʷa	*'bras*
sand	ˊɸɕi ma	ˊʰsa	ɕi^{13} ma^{55}	ˊɕwə ʱgɑ	ɕi^{35} ma^{53}	ˊtɕə ma	*bye ma*
shadow	-	ˊɲi ʈʰə	ɲe^{13} qo^{55} /n	ˣᵐ̥pʰə /n	na^{55} /n	ˊʈə nɑ	*grib ma*
shoulder	ˣᵐ̥pʰɛ rə /n	ˣᵐ̥pʰə rə /n	phiɛ^{55}	ˣᵐ̥pʰi /n	kho^{55} te^{53} /n	ˊᵖʈʰɑ pe	*phrag pa*
Sichuan pepper	ˊʱdʑa ʱgə	ˣsə /n	rdʑe^{13} rgo^{55}	ˣsə /n	la^{55} tsə^{53} /c	ˣʱja ma	*g.yer ma/ rgya rgod*[7]
Tibetan	ˊpə ri	ˊpe ri	pe^{55} ri^{55}	ˊpe ʱba	-	ˊpo pe	*bod*
tiger	ˣʰtɑ	ˣʰtɑ	sta^{55}	ˣʰtɑ	ta^{53}	ˣʰtɑˠ	*stag*

The native word for 'book' in Lhagang Choyu as seen in Table 1 indicates a relationship between Lhagang Choyu and Choyu. The form /ˣʱgʷə/ only appears in Lhagang Choyu, and it corresponds to /ˣʱgwə/ in Rongpa Choyu and /ˣʱgə/ in Lhayul Choyu. It is also recorded as *dgod*[8] in the Tibetan script in *Litang Xianzhi* (1996:474). This form might be maintained in dialects spoken within Lithang County

[7] The form *rgya rgod* is not a LT word but a local word form that, in fact, denotes 'chili' and not 'Sichuan pepper'. In the Lhagang dialect of Minyag Rabgang Khams, this word means 'wild onion' (Suzuki & Sonam Wangmo to appear).

[8] This spelling might be pronounced as [ʱgə] (tone unspecifiable) in a local manner. The meaning of this LT spelling is 'laugh,' which is not related to the context here.

because the dialects of gYanglagshis (Nyagrong County) and Phubarong (Nyagchukha County) use Tibetic loanwords that are different from each other, i.e., LT *dpe cha* and *yi ge*, respectively. The former word form is mainly used in Amdo Tibetan while the latter is used in Khams Tibetan. Moreover, the phonetic realisation is noteworthy. An initial uvular sound corresponding to LT *dp*, /χp/, is analysed as an archaic sound because Amdo Tibetan generally has a /χw/ sound for LT *dp*, and so do varieties spoken in Lithang. This situation implies that the form of the gYanglagshis dialect is an older borrowing. However, on the contrary, its vowel in the second syllable /a/ suggests a new loan. In any case, since the description of this dialect is not given by the present author, we cannot consider the sound form a phonetic reality. In the case of the Phubarong dialect, it is worth noting that the dialect uses a /ʑ/ sound corresponding to LT *y*. This sound correspondence is a minor case in Khams Tibetan (Suzuki 2016b, 2018a), and the same sound correspondence is merely demonstrated in the Lamdo dialect (Sems-kyi-nyila group: spoken in Lamdo hamlet of Shangri-La Municipality) within the closest place. Dialects belonging to the sPomborgang group (Suzuki 2018b) spoken near the Choyu-speaking region have a similar sound correspondence, but it is not the case for the word with a LT simplex *y*. To sum up, the word 'book' is one of the suggestive examples with which one can access the history of language contact in these languages.[9]

In addition to that, we note that Choyu dialects (and possibly Lhagang Choyu too) receive more Tibetic loanwords due to language contact, therefore experiencing a rapid language change. For example, the word for 'sun' in Lhagang Choyu is a native word /˜mi tsi/ as is the case in Phubarong Choyu /˜mə' ʰtsə/. However, Lhayul Choyu now employs /ˊɲi ma/, a Tibetic loan derived from LT *nyi ma*.

For more general discussions, we point out particular sound correspondences illustrated in the Tibetic loanwords in Lhagang Choyu:

LT initial *w* and *'* : /ɣ/ ('fox'; 'owl')
LT initial *z* : /z/ ('bridge', 'copper')
LT initial *db* and *'* : /ʁ/ ('power')
LT initial *phr* : /ᵖʈʰ/ ('shoulder')
LT vowel *a* at word-final: /e/ ('old', 'god', 'monk', 'monastery', etc.) or /a/ ('owl', 'sand', 'pepper')
LT rhyme *er*: /əʸ/ ('gold')

The features of sound correspondences illustrated in loanwords in Lhagang Choyu are not always common in Choyu dialects. Hence, it is significant to analyse how the differences occurred by comparing the data of potential origins and varieties of the neighbouring Tibetic languages.

3 Analysis of the route of loans

In order to analyse the route of loanwords, we have to find cases in which a LT form corresponds to various sounds following spatial/dialectal (synchronic) and/or temporal (diachronic) differences. The latter case can be examined through the sound correspondence of a LT rhyme *a* (single vowel without finals). As pointed out in Section 2, Lhagang Choyu has two principal sound correspondences: /e/ and /a/. We consider the first sound correspondence to be the oldest. The sound correspondence between LT rhyme *a* and lower front vowels /i, e, ɛ/ is widely illustrated in rGyalrongic and Qiangic languages. Lhagang Choyu also applies this sound change. When we deal with the issue of the route of loanwords, we have to pay attention to the differences in loanwords' strata. In this article, we mainly choose words in the older stratum of loans in order to elucidate the varieties from which the Tibetic language Lhagang Choyu has borrowed Tibetic words.

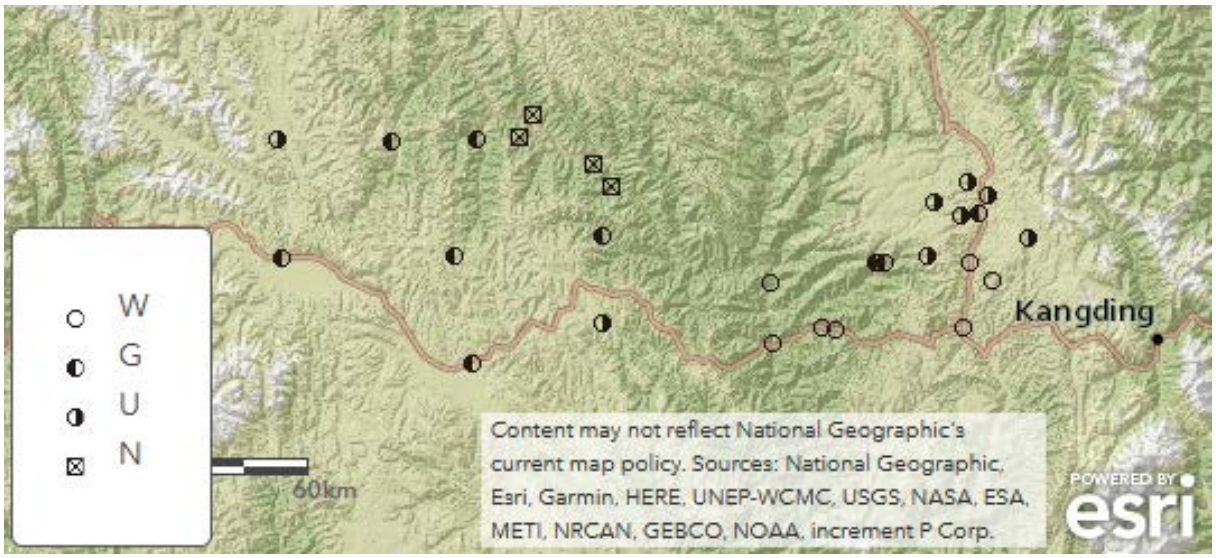
[9] Other than word forms, there is a possibility of discussing the influence from Tibetic languages regarding the semantic field and change if one examines a specific semantic change, e.g., the word form for 'rain' compared to that for 'sky.' This example is an interesting case discussed by Shirai et al. (2018) and Suzuki (2018c).

Among the items in Section 2, we deal with the following limited examples below: ‘fox’, ‘chicken’, ‘rice’, ‘bridge’, ‘glass’, and ‘shoulder’. We first discuss the variation of word forms in relevant Tibetic languages and then create a linguistic map for a geolinguistic analysis.

- ‘fox’

The word for ‘fox’ in LT is *wa*, and Tibetic languages surrounding Choyu and Lhagang Choyu employ a form corresponding to this. However, there are principally three phonetic realisations demonstrated within Tibetic languages: /wa/, /ʁa/, and /ɣa/.[10] /wa/ is widespread in Khams Tibetan; /ʁa/ is mainly found in Amdo; /ɣa/ is found in several dialects in this area, especially in Lithang, as well as in Lhagang Choyu /ˊɣa/.[11]

Lhagang Choyu has uvular sounds distinguished from their velar counterparts in their consonantism. Hence, when it borrowed the word for ‘fox’ from a Tibetic variety, the form must have been a velar initial, not an uvular one. Otherwise, Lhagang Choyu may have received an uvular sound itself and maintained its articulatory status. Therefore, we should pay attention to the distribution of dialects with a /ɣa/-form.



Legend: N=native word; U=/ʁ/; G=/ɣ/; W=/w/

Map 2: Word forms for ‘fox’ and initials

The /ɣa/-form is illustrated in both Khams and Amdo in Lithang. The /ʁa/-form also appears in some dialects in the surrounding area of the dialects with the /ɣa/-form. In Amdo especially, uvulars exist in the consonantism. Hence, a /ɣa/-form demonstrated in Amdo is noteworthy. Lhagang Choyu has borrowed from one of such varieties and maintained it to date. However, Lhagang Choyu has only borrowed this word after borrowing words with /e/ vowel corresponding to LT *a*, e.g., /ˊpo p**e**/ *bod p**a*** ‘Tibetan’ and /¯l̥**e**/ *lh**a*** ‘deity’ (see Table 1). In addition, as shown in Table 1, some dialects of Choyu maintain their native word forms. This fact suggests that Lhagang Choyu might have borrowed the Tibetic form for ‘fox’ recently.

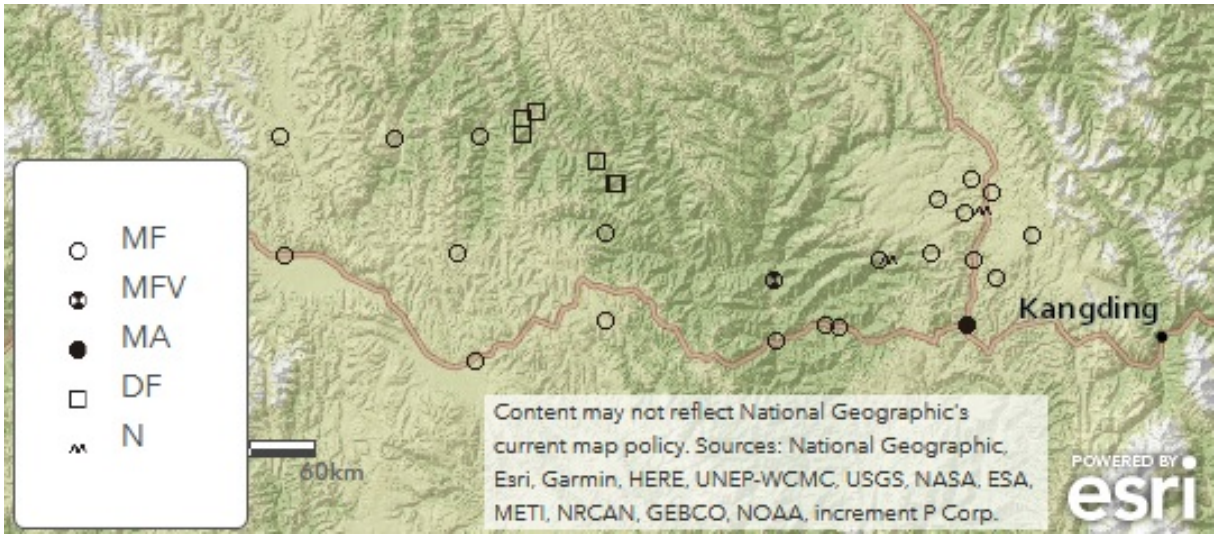
[10] Tonal signs are omitted when we do not specify a given dialect.

[11] See Hill (2006) for phonetic forms of various Tibetic languages of the word ‘fox’.

- 'chicken'

The word for 'chicken' in LT is *bya*, and Tibetic languages surrounding Choyu and Lhagang Choyu employ a form corresponding to this. However, the form applied in Choyu is dissyllabic, and it seems to correspond to LT's *rgya bya*, which literally means 'Chinese chicken'.[12] For the sound corresponding to the LT initial *by*, many Tibetic languages use a prepalatal fricative /ɕ/. Besides, several varieties of Minyag Rabgang Khams also have another correspondence: an affricate /tɕ/. However, it only appears in a few words including 'chicken', while a fricative /ɕ/ or /ʑ/ appears in other words.

Lhagang Choyu has a monosyllabic form /ˊᶠɕa/, in which the initial consists of a bilabial fricative as a preinitial and a voiceless prepalatal fricative as the main initial.



Legend: N=native word; MF=monosyllabic+/ɕ/; MFV=monosyllabic+/ʑ/; MA: monosyllabic+/tɕ/ DF=dissyllabic+/ʑ/

Map 3: Word forms for 'chicken' and initials

In Thamkhas Khams and Lhagang Khams, we find a local native form for 'chicken': /ˊkõ go/.[13] This means that Lhagang Choyu has already borrowed a word for 'chicken' from other varieties in Lhagang, but not before borrowing words with /e/ vowel corresponding to LT *a*, as in 'fox. Paying attention to voicing, we find that Choyu dialects use a voiced fricative, but Lhagang Choyu does a voiceless counterpart. As seen from the discussion on the form for 'bridge' below, devoicing of fricative series might not have occurred in Lhagang Choyu recently. Thus, Lhagang Choyu has received a voiceless form when borrowing the word. This suggests that the relative time of borrowing is neither recent nor archaic: it is highly possible that the origin of the loan is a dialect spoken in Lithang or its surrounding areas. See also the discussion on 'bridge' later.

- 'rice'

The word for 'rice' in LT is *'bras*, and Tibetic languages surrounding Choyu and Lhagang Choyu employ a form corresponding to this. The word 'rice' in most parts of the Tibetosphere can be considered a cultural word (Suzuki & Sonam Wangmo 2016b), and non-Tibetic languages spoken in the Tibetosphere often use a Tibetic loan.[14] Considering the word form in Lhagang Choyu, we pay attention to the nasal element appearing at the preinitial position because a principal difference in the word forms

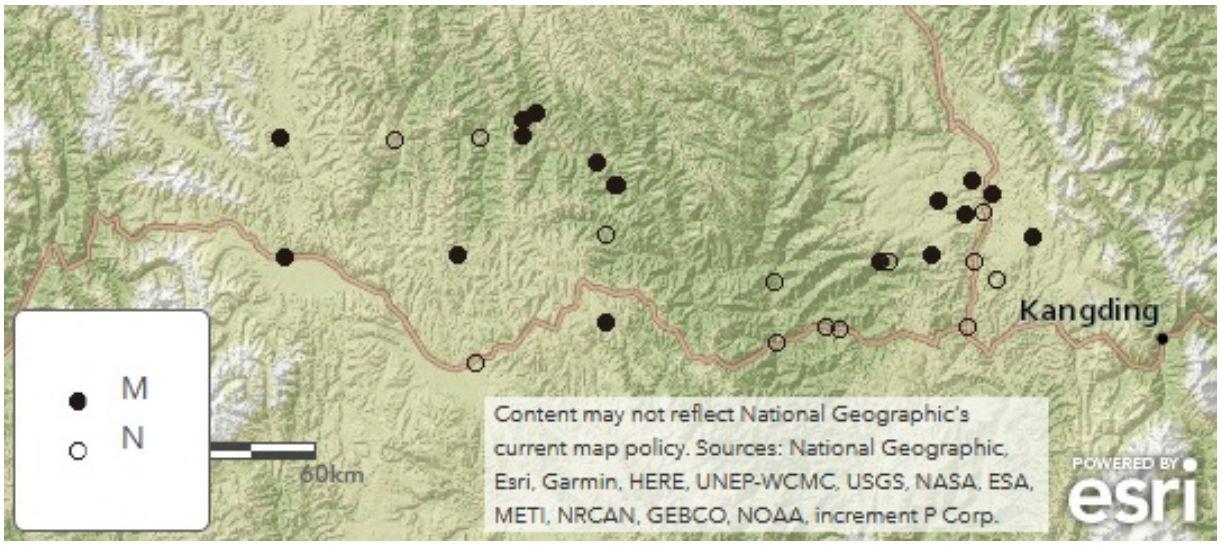
[12] However, we have never described any Tibetic languages, including Literary Tibetan, which use the form *rgya bya* for 'chicken' so far.

[13] The etymology of this form is unidentified. Another phonetic variety with uvulars also exists : /ˋqo ᴺGo/ (Suzuki & Sonam Wangmo to appear).

[14] See also Suzuki (2016a) and Suzuki et al. (2016) for the word 'rice' and its relevant words in Tibeto-Burman.

in Tibetic languages of the given area appears in this feature. We can find a dialectal difference between labial prenasal /ᵐ/ and homorganic prenasal (prenasalisation in a narrow sense).

Lhagang Choyu has a bilabial nasal preinitial, /ˉᵐɖʷa/, which reflects an older sound derived from LT *'bras*: *ᵐbras < *'bras*. Whether or not a variety can have a heterorganic labial nasal preinitial depends on the sound system. However, in a dialect that allows this heterorganic nasal to appear as a preinitial, a form with a labial nasal is considered an older type as opposed to a homorganic counterpart.



Legend: M=labial prenasal; N=homorganic prenasal

Map 4: Preinitials for 'rice'

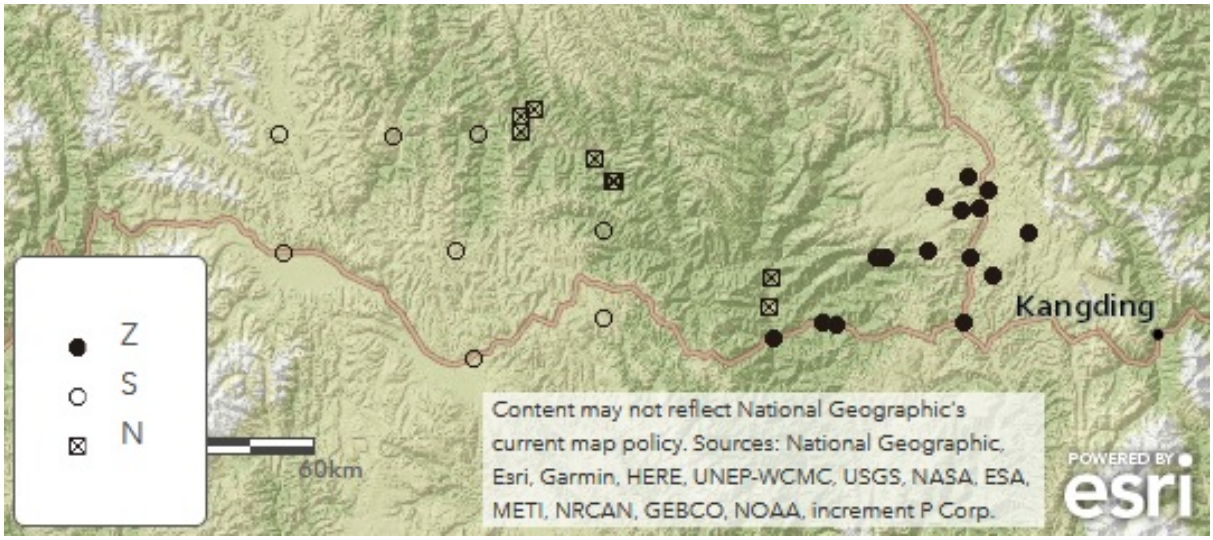
Map 4 shows that differences in the preinitial nasal merely depend on the nature of the languages and this determines whether or not a heterorganic nasal appears. In this case, we have not found any significance of geolinguistic analysis. We can say that the Lhagang Choyu form /ˉᵐɖʷa/ is loaned from a Tibetic variety which can possess a heterorganic bilabial nasal preinitial. We also note that the LT rhyme *-as* corresponds to /a/ in Lhagang Choyu, as evidenced in the word 'cloth' /ˊra/ (see Table 1). In the Rongpa dialect of Choyu, we find a similar form /ˊrja/ for 'cloth' is applied, but /ˉᵐɖɛ/ applies for 'rice'. There might be a temporal difference of the borrowing between the two words in Rongpa. In any case, the sound correspondence between LT *-as* and /a/ is of a rare type.[15] If the forms in Lhagang Choyu really reflect an archaic sound of the Tibetic languages surrounding it, they will also be useful in investigating a sound change process in Tibetic languages.

- 'bridge'

The word for 'bridge' in LT is *zam pa*, and Tibetic languages around Choyu and Lhagang Choyu employ a form corresponding to this. Regarding the word form in Lhagang Choyu, the voicing of the initial catches our attention because a principal difference in the word forms in Tibetic languages within the given area appears in this feature. We find a dialectal difference of word forms between /s/ and /z/.

Lhagang Choyu has a voiced initial that is pronounced /ˊzã ᵐbe/. Judging from the vowel of the second syllable, this word form belongs to an older stratum of the loanwords.

[15] Within the first author's field notes, only Hor Bachen dialect shows this sound corresponding within Tibetic languages.



Legend: N=native word; S=/s/; Z=/z/
Map 5: Initials in 'bridge'

Choyu's native form includes a /ts/-initial, which is probably a cognate of LT *zam* at the Proto-Tibeto-Burman level (*m-dzam, #3604, STEDT).

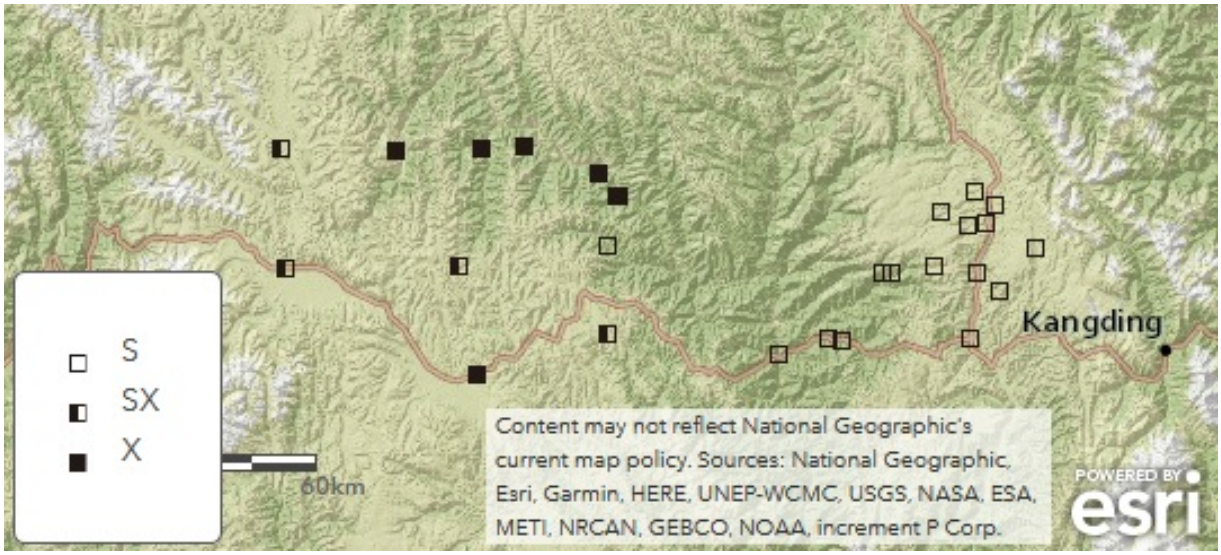
However, we cannot specify when the devoicing of Tibetic languages occurred. Presently, the Tibetic varieties spoken in Lithang have a voiceless initial /s/, but it is not guaranteed that this sound was voiceless when Lhagang Choyu received a loanword form. Based on the vocalic quality of the second syllable of the word 'bridge' in Lhagang Choyu, it should be considered a loan belonging to the old stratum. Therefore, even though there are two possibilities of the origin, a variety spoken in Lithang or one spoken in the surrounding area of Thamkhas, the former is a more potential candidate. This interpretation implies that Tibetic languages around Lithang at that time had a sound correspondence between LT *z* and /z/.

- 'glass'

The word for 'glass' in LT is *shel*, and Tibetic languages surrounding Choyu and Lhagang Choyu employ a form corresponding to this or a compound containing this. Considering the word form in Lhagang Choyu, we pay attention to an articulatory position of the initial because a principal difference in the word forms in Tibetic languages within the given area appears in this feature. We find a dialectal difference in initials between a prepalatal fricative /ɕ/, a velar fricative /x/, and a prepalatal-velar double-articulated /ɧ/[16] (regardless of the aspiration feature).

Lhagang Choyu has a prepalatal fricative initial as /˩ɕe/. Since the sounds /ɕ/ and /x/ are distinctive in the sound system of Lhagang Choyu, we can exclude the possibility that the original Tibetic form includes a velar sound. However, the interpretation of the sound [ɧ] in Lhagang Choyu is unclear, and it is also possible that it is interpreted as an allophone of /ɕ/.

[16] [ɧ] attested in Amdo Tibetan is close to a double-articulated sound of prepalatal and velar in principle, whereas [ɧ] in Swedish, it varies phonetically and it is sometimes described as a "highly rounded, labiodental, velar or velarized fricative" and a "dorsovelar voiceless fricative" (Ladefoged & Maddieson 1996:171-172; based on Lindblad 1980) in spite of the definition of International Phonetic Alphabet as a sound "simultaneous ʃ and x". As Lindblad (1980) and Ladefoged & Maddieson (1996:172) claim, the sound [ɧ] is to be distinguished from a velar fricative [x]. Additionally, [ɧ] includes various articulatory manners, and this feature is also a reason why we can apply it for the specific sound attested in Tibetic languages.



Legend: S=/ɕ/; SX=/ɧ/; X=/x/
Map 6: Initials in 'glass'

There is no appropriate way of describing the dorsal sound corresponding to a LT simplex *sh* demonstrated in some Tibetic languages, especially in Khams and Amdo. Here we must distinguish a double-articulated /ɧ/ from a mono-articulated /x/ with allophones such as [ç] and [x][17] because Tibetic languages spoken in this area distinguish these two sounds from each other. Amdo Tibetan spoken in Lithang County uses /ɧ/, whereas that spoken in Lhagang uses /ɕ/ as in the local Khams Tibetan. Varieties of Khams Tibetan spoken in Lithang County use /x/.

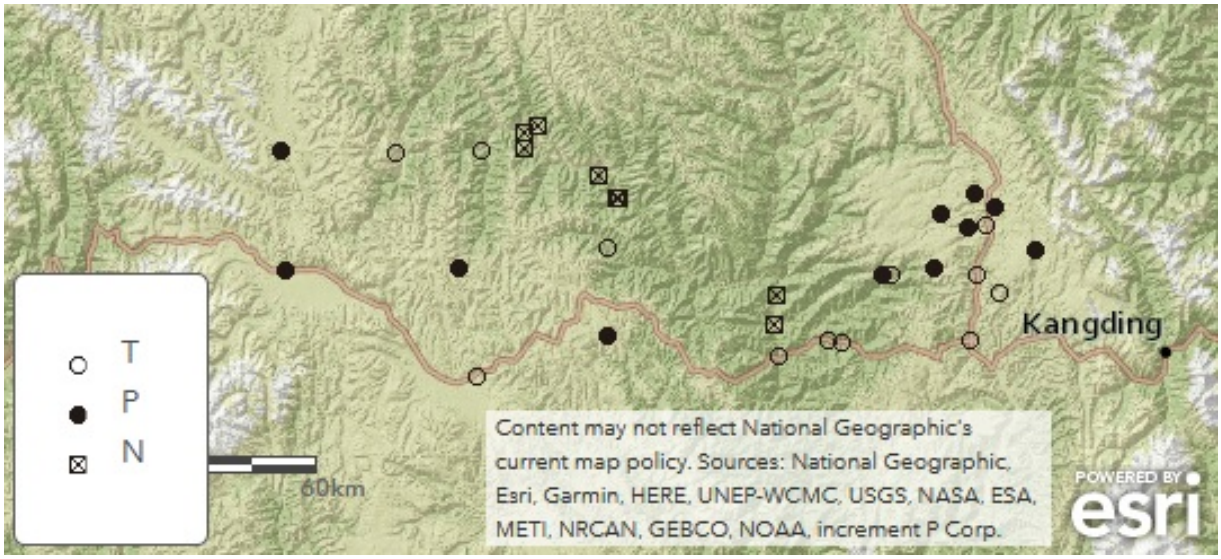
Map 6 shows that the Lhagang Choyu form is common to its surrounding varieties of Tibetic languages. Looking at the Rongpa dialect of Choyu, we find that the word 'glass' is /ˊxe ʰgu/. Hence, we assume that the word 'glass' is a recent loan. For this reason, the initial sound corresponds to that of surrounding Tibetic languages.

- 'shoulder'

The word for 'shoulder' in LT is *phrag pa*, and the Tibetic languages surrounding Choyu and Lhagang Choyu employ a word form corresponding to this. Considering the word form in Lhagang Choyu, we draw our attention to the existence of a preinitial because a principal difference in the word forms in Tibetic languages of the given area appears in this feature. We find a dialectal difference between forms with and without a labial plosive preinitial /ᵖ/.

Lhagang Choyu has an initial with a bilabial preinitial in the first syllable as /ˊᵖtʰɑ pe/, which reflects an older sound derived from a LT initial *phr-*: /ᵖtʰ-/ < *pʰr- < *phr-*. Judging from the vowel of the second syllable, this word form belongs to an old stratum of the loanwords. Whether or not a variety can have a labial plosive preinitial depends on the sound system. However, in a dialect that has this preinitial, a form with a labial nasal is considered to be an older type as opposed to being a homorganic counterpart. See 'rice' above.

[17] When one considers that /x/ has two allophones [ç] and [x], the condition is formulated as follows: [ç] /_+higher front vowel, [x] / _–higher front vowel. Even in this simple case, the phonetic value before /a/ is always problematic.



Legend: N=native word; P=preinitial /ᵖ/; T=no preinitial
Map 7: Word forms for 'shoulder' and preinitials

Whether or not a variety can have a labial preinitial depends on the sound system. Varieties of Khams Tibetan in this area cannot apply this pattern due to this restriction. This case resembles that of the example 'rice' discussed above. However, the loanword is only applied in Lhagang Choyu while Choyu dialects have a native word. The problem is that, as the Lhagang Choyu form /ˈᵖʈʰɑ pe/ suggests, it belongs to the older stratum of Tibetic loans. This situation implies that Lhagang Choyu had borrowed this form before it borrowed the word 'fox' from Tibetic varieties spoken in Lithang. It is unclear whether older varieties of Khams Tibetan allowed a labial preinitial to appear in the phonology. Hence, it is also unclear whether the loan word originated from Khams or Amdo.

4 Conclusion

This article presented an overview of the Tibetic loanwords in Choyu and Lhagang Choyu and discussed their potential borrowing route by examining six words from a geolinguistic perspective. Lhagang Choyu has at least two strata of Tibetic loanwords, and this article discussed words belonging to the older stratum. The discussion found that several phonetic features had originated from varieties of Amdo Tibetan spoken in Lithang County.

The six loanwords that we discussed principally have dialectal differences in sound and not in word form. However, as various aspects of sounds such as phonetics and phonotactics vary within the Tibetic varieties, we can analyse the borrowing route to some extent.

The article's result corresponds to the historical narratives that tell us that the ancestors of Lhagang Choyu speakers, who maybe with Amdo-speaking pastoralists, have come from the present Lithang-Nyagrong border area (Suzuki & Sonam Wangmo 2016ac, forthcoming). We can find some traces of the history in Tibetic loanwords.

Acknowledgements

The present study is funded by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science [JSPS]: "International Field Survey of Tibeto-Burman Link Languages" (headed by Yasuhiko Nagano, No. 16H02722). This article is also a part of the results of the two Grants-in-Aid for Scientific Research from JSPS: "Investigation of Undocumented Languages in the Eastern Tibetosphere and their Geolinguistic Research" (headed by the present author, No. 17H04774) and

"Geolinguistic Studies of China and Adjacent Multilingual Areas Using High-resolution and Wide-area Maps" (headed by Mitsuaki Endo, No. 18H00670).

References

Hill, N. W. 2006. Tibetan *vwa* 'fox' and the sound change Tibeto-Burman **wa* > Tibetan *o*. *Linguistics of the Tibeto-Burman Area* 29.2: 79-94. [DOI: 10.15144/LTBA-29.2.79]

Huang, B. ed. 1992. *Zangmianyuzu Yuyan Cihui*. Zhongyang Minzu Xueyuan Chubanshe.

Huang, Y., sKal-bzang Lha-mo, and bZang-po. Forthcoming. *Zhabayu Jianzhi*.

Ladefoged, P. and I. Maddieson. 1996. *The Sounds of the World's Languages*. Wiley-Blackwell.

Lindblad, P. 1980. *Svenskans sje- och tje-ljud i ett allmänfonetiskt perspektiv*. Doktorgradavhandlung, Lunds Universitet.

Litang Xianzhi Bianzuan Weiyuanhui. 1996. *Litang Xianzhi*. Sichuan Renmin Chubanshe.

Lu, Sh. 1985. Zhabayu gaikuang. *Minzu Yuwen* 2: 67-76.

Roche, G. and H. Suzuki. 2017. Mapping the linguistic minorities of the Eastern Tibetosphere. *Studies in Asian Geolinguistics VI —Means to Count Nouns—*, 28-42. https://publication.aa-ken.jp/sag6_count_2017.pdf

Roche, G. and H. Suzuki. 2018. Tibet's minority languages: Diversity and endangerment. *Modern Asian Studies* 52.4: 1227-1278. [DOI: 10.1017/S0026749X1600072X]

Shirai, S. this volume. A geolinguistic analysis of "rain" and "fish" in the Western Sichuan Ethnic Corridor languages.

Shirai, S., H. Suzuki and K. Kurabe. 2018. Semantic shifts in expressions for 'it rains' in Tibeto-Burman. *Studies in Asian Geolinguistics VIII —It Rains—*, 62-76.

STEDT=STEDT database (beta). *The Sino-Tibetan Etymological Dictionary and Thesaurus*. http://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl

Sun, H. ed. 1991. *Zangmianyu Yuyin he Cihui*. Shehui Kexue Chubanshe.

Suzuki, H. 2007. Kamutibettogo Kangding Xinduqiao [Rangakha] hoogen no onsei bunseki. *Asian and African Languages and Linguistics (AALL)* 2: 131-162. http://hdl.handle.net/10108/51094

Suzuki, H. 2016a. A geolinguistic description of terms for 'rice' in Tibetic languages of the eastern Tibetosphere. *Studies in Asian Geolinguistics II —Rice—*, 52-59. https://publication.aa-ken.jp/sag2_rice_2016.pdf

Suzuki, H. 2016b. Zangyu fangyanxue yanjiu yu yuyan ditu: Ruhe kandai "Kang fangyan". *Minzu Xuekan* 2: 1-13+92-94. http://211.83.241.172/mzxk/ch/reader/view_abstract.aspx?file_no=20160201&flag=1

Suzuki, H. 2018a. Litang xian ji qi zhoubian Zangzu yuyan xianzhuang diaocha yu fenxi. *Minzu Xuekan* 2: 35-44+106-109.

Suzuki, H. 2018b. Remarks on 'rain' in Tibetans' languages in Lithang County. *Studies in Asian Geolinguistics VIII —It Rains—*, 56-61.

Suzuki, H., K. Kurabe, K. Iwasa, S. Shirai, S. Ebihara, and I. Matsuse. 2017. Geolinguistic analysis of 'rice' in Tibeto-Burman. *Studies in Asian Geolinguistics II —Rice—*, 37-51. https://publication.aa-ken.jp/sag2_rice_2016.pdf

Suzuki, H. and Sonam Wangmo. 2015. Lhagang Tibetan of Minyag Rabgang Khams: Vocabulary of two sociolinguistic varieties. *Asian and African Languages and Linguistics (AALL)* 10: 245-286. http://hdl.handle.net/10108/85072

Suzuki, H. and Sonam Wangmo. 2016a. Lhagang Choyu: A first look at its sociolinguistic status. *Studies in Asian Geolinguistics II— Rice—*, 60-69. https://publication.aa-ken.jp/sag2_rice_2016.pdf

Suzuki, H. and Sonam Wangmo. 2016b. Cultural contexts of the expansion of a Tibetan word *'bras* 'rice' in the easternmost Tibetosphere. In Mitsuaki Endo (ed.) *Papers from the Third International Conference on Asian*

Geolinguistics, 72-79. Research Institute for Languages and Cultures of Asian and Africa. https://publication.aa-ken.jp/papers_3IC_Asian_geolinguistics_2016.pdf

Suzuki, H. and Sonam Wangmo. 2016c. Vocabulary of Shingnyag Tibetan: A dialect of Amdo Tibetan spoken in Lhagang, Khams Minyag. *Asian and African Languages and Linguistics (AALL)* 11: 101-127. http://hdl.handle.net/10108/89211

Suzuki, H. and Sonam Wangmo. 2017. Lhagang Choyu wordlist with the Thamkhas dialect of Minyag Rabgang Khams (Lhagang, Dartsendo). *Asian and African Languages and Linguistics (AALL)* 12: 133-160. http://hdl.handle.net/10108/91144

Suzuki, H. and Sonam Wangmo. to appear. Two folktales of Lhagang Tibetan (Minyag Rabgang Khams): *Three birds* and *Lark and Partridge*. *Asian and African Languages and Linguistics (AALL)* 13.

Suzuki, H. and Sonam Wangmo. forthcoming. Migration history of Amdo-speaking pastoralists in Lhagang, Khams Minyag, based on narratives and linguistic evidence. *Archiv Orientální Supplementa*.

Suzuki, H. and Tsering Samdrup. 2018. *How geolinguistics deals with pastoralists' speeches and their history: A case study on Amdo Tibetan in Eastern Tibet*. Paper presented at 9th Congress of the International Society for Dialectology and Geolinguistics (Vilnius).

Tsering Samdrup and H. Suzuki. 2017. Migration history and *tsowa* divisions as a supplemental approach to dialectology in Amdo Tibetan: A case study on Mangra County. *Studies in Asian Geolinguistics VII —Tone and Accent—*, 57-65. https://publication.aa-ken.jp/sag7_tone_2017.pdf

Tournadre, N. 2014. The Tibetic languages and their classification. In Th. Owen-Smith & N. W. Hill (eds.) *Trans-Himalayan Linguistics: Historical and Descriptive Linguistics of the Himalayan Area*, 105-129. Walter de Gruyter.

Wang, T. 1991. Queyuyu. In Q. Dai, B. Huang, A. Fu, Rig-'dzin dBang-mo, and J. Liu. *Zangmianyu Shiwuzhong* 46-63. Beijing Yanshan Chubanshe.

Distribution and diffusion of the Dialectal Formatives in the Region of Seto Inland Sea

Shinsuke Kishie[i] , Yukako Sakoguchi[ii]

[a] Graduate School of Integrated Arts and Sciences, Tokushima University
[b]Regional Education Cooperation Center, Shikoku University

Abstract

The aim of this study is to highlight the pattern of changes in the distribution of dialects recently in a linguistic map created by means of GIS (Geographic Information System) after about 60 years of the survey " Linguistic Atlas of the Seto Inland Sea Vol.1 and 2" (hereinafter abbreviated as LAS) conducted in 1974.

1. Introduction

This research is about a theme of Geolinguistics by means of GIS. Geolinguistics is one of the areas of language history. It is an area of study which deals with the language change along the geographical distribution of the pronunciation, grammar, vocabulary and so on. For example, it aims to deal with how the language change which occurs historically, and what factors and backgrounds are involved in it through the analysis of the geographical distribution of dialects.
In this study, we aim to reveal the language change in a designated area of the Seto Inland Sea by means of field survey with a research method of Geolinguistics "Method of Geolinguistics" given by Takeshi Shibata (1969).

The Geolinguistics Surveys on the Chugoku-Shikoku region, especially the *Seto Inland Sea region* include Fujiwara (1974) and "Linguistic Atlas of the Seto Inland Sea Vol. 1 and 2" (LAS) in addition to Fujiwara (1990). It will not be an exaggeration to say that the latter one being a Geolinguistics study of Japan is a unique one in the world from a birds-eye viewpoint. Because it being a linguistic map involving the older citizens and the young people which chart the geographical variation of dialect and also highlights the language change between the generations. In other words, it can be said that the difference that appears in the dialects between the two generations of young and old in "apparent time" by simultaneous representation of the geographical distribution is reflected in Linguistic Atlas.

Therefore, in this study, following the surveys by Fujiwara, we conducted a follow-up survey on the Seto Inland Sea region to investigate about how the language change has occurred and reached the present state in the Seto Inland Sea region. In the following, we will find out the trend of change in the distribution of dialects in "real time" by analyzing the change in *conjunction particle indicating the cause-reason relation in an order* as a part of this survey.

2. The Survey and Research to be Covered in this Presentation

In this study, the surveys of Fujiwara (1990), LAS (1974), Language Questionnaire Survey by Correspondence on the Seto Inland Sea (2011-2014) conducted by Sakoguchi have been taken for comparison and consideration. Accordingly, the survey conducted by Sakoguchi is referred to as "Sakoguchi's Survey" as follows.

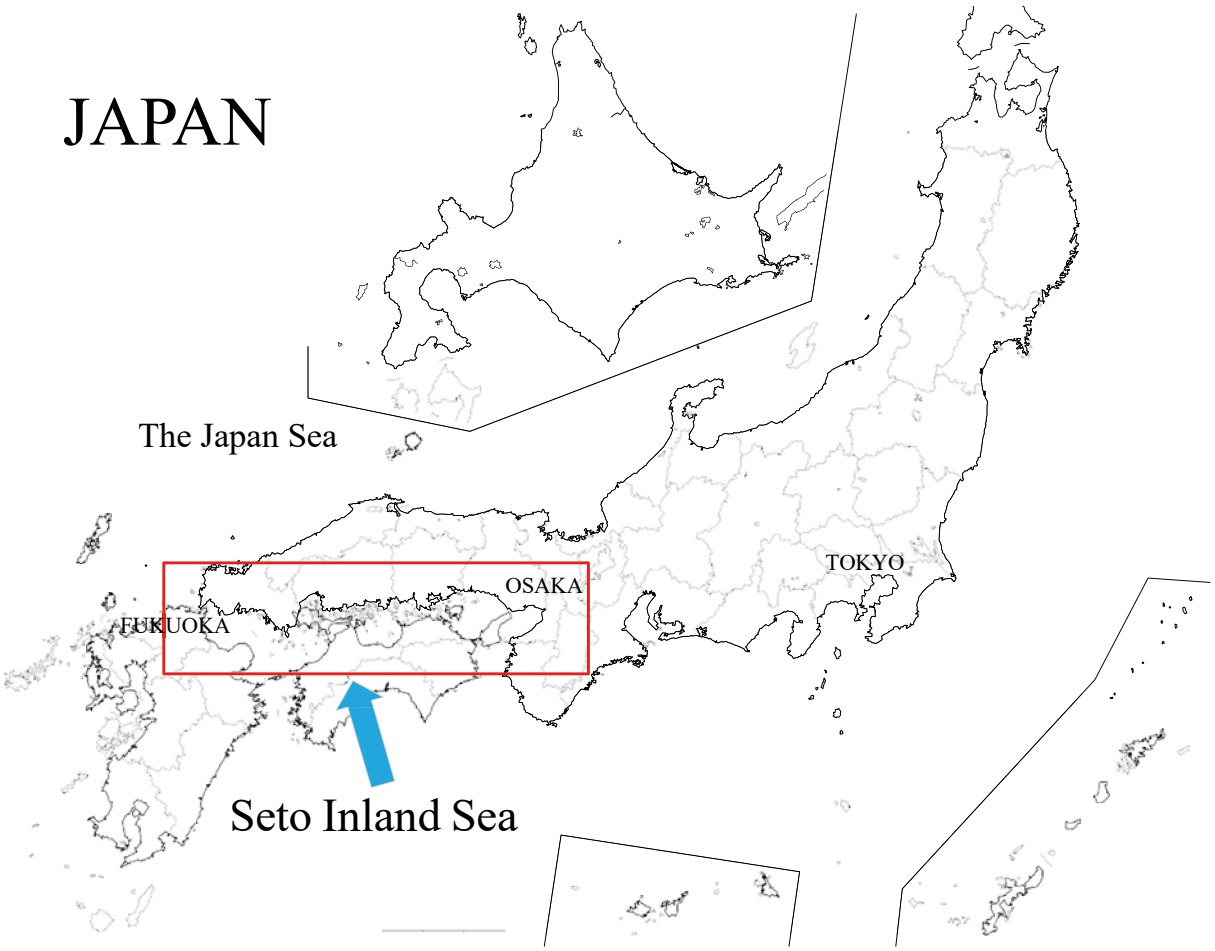


Fig 1. Seto Inland Sea

2-1. Fujiwara Survey (1990)

This study is a linguistic map of the young people female aged 15 to 18) with the survey conductedin the regions of Chugoku, Shikoku, and West Kinki. It summarizes the results of the survey by correspondence in the 833 different sites in a time period of two years from 1933.

2-2. LAS

LAS is a collection of linguistic maps comprising 251 maps which have been created out of the comparative analyses between the speech of old ladies (born in Meiji period) and girls (then junior high school students of 1st and 2nd grade) compiled in two volumes: Vol. 1 and 2. The survey through interview was conducted in a total of 925 different sites (701 in the inner islands, 141 sites in coastal areas and unidentified 83 sites) for about 5 years since 1960, and the results have been transformed into a linguistic map.

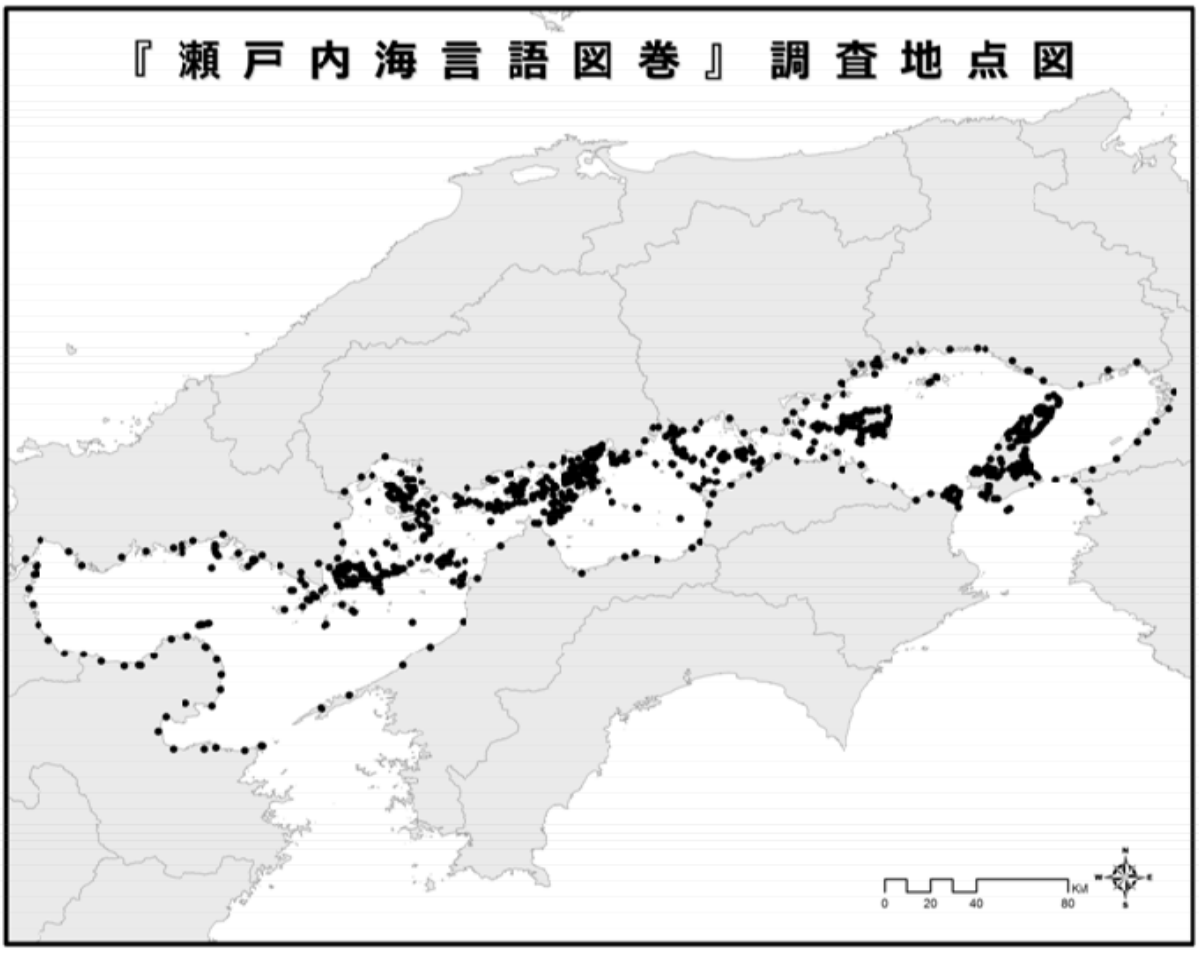


Fig 2. Survey Spots of LAS

2-3. Sakoguchi's Survey

This survey complies with the survey through correspondence and partly parallels with the survey through the interview on the Seto Inland Sea Region. Different surveys through correspondence have conducted in phases including first survey (October to December 2011), second survey (August to September 2012) and third survey (April 2013 to January 2015) which is still ongoing on.

In this survey, we have procured responses from those who were native born in the Seto Inland Sea region before 1965 in cooperation with the municipal board of education, the community hall, the fishery cooperatives etc. Accordingly, we received survey reports from the primary, secondary and tertiary survey along with the interview survey of about native-born 1,550 elderly speakers who were in their 50ties. Thus, we have created a linguistic map out of the data of 1,512 speakers for this purpose.

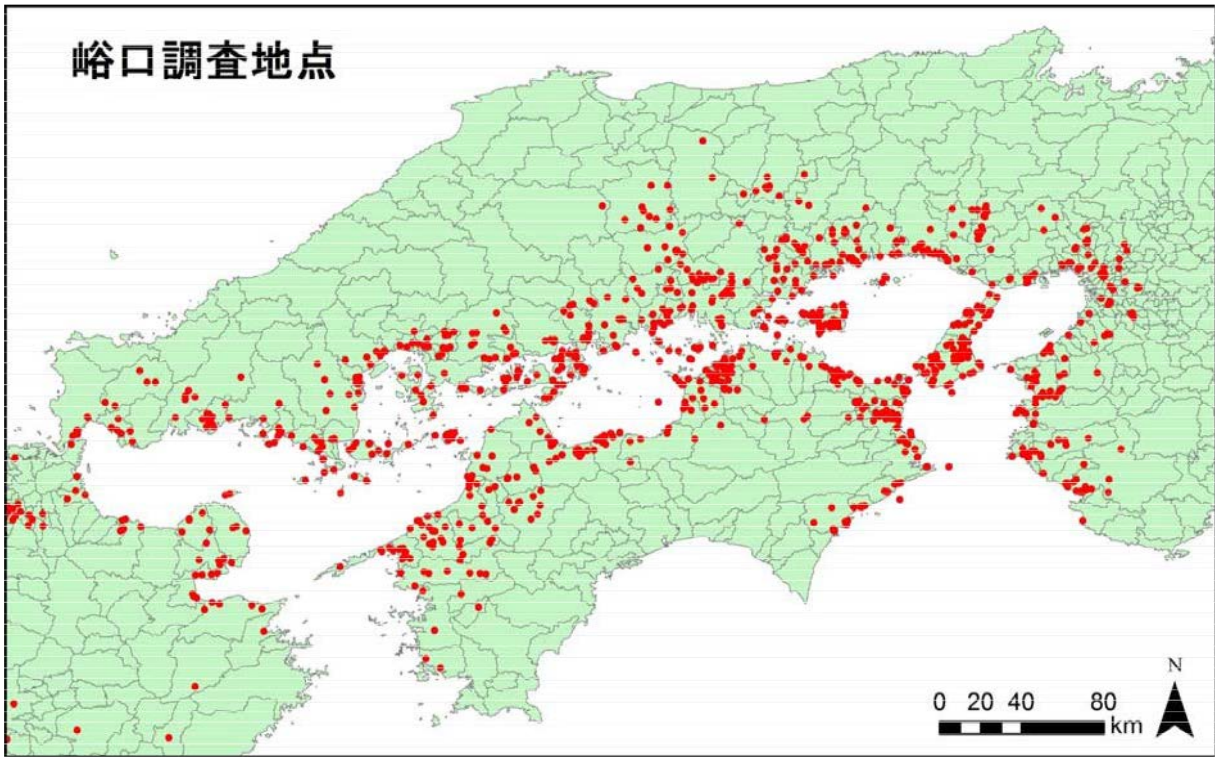


Fig 4. Survey Spots Conducted by Sakoguchi

3. Research Method

As to the organization of survey result, we have created a linguistic map and attempted to make analysis by using secondary mesh since the number of geographical sites are many. Although it is only a representation of the aggregate values, we think that making a quantitative comparison between the meshes became easy at this, because the answer format can be visualized on the map by manipulation in this map. Specifically, the grammatical form, which we have taken up here, is a conjunction particle -*kara*- indicating the cause/reason relation in an order. Here, we have found the boundary of the geographical distribution showing wide expansion and narrow retraction and also showing no changes in the LAS. Given the above distribution, we have investigated the trend of variation that reflected in the distribution over time by comparing the distribution of different eras.

4. Research Findings

The subject matter which has been taken up for the study is the distribution of conjunction particle *kara* indicating the cause/reason relation in an order in the map no.31 of LAS. Firstly, the map no.31 about older generation in the LAS shows that *ken* is distributed in Hiroshima, Ehime, Kagawa, and Tokushima, while *ke* is collectively concentrated in the Bichu Islands of Okayama Prefecture, Bigo Islands of Hiroshima prefecture, Yamaguchi prefecture and northern Kyushu etc. In the central part of Kinki, the distribution map is mainly comprised of *sakai* and *sakai ni* in Osaka, whereas *saka* and *sake* exist in opposition in the prefecture of Wakayama. In the city of Osaka and Hyogo prefecture, *yotte* and *yotte ni* are collectively distributed around the Awaji island. This trend seems to be almost the same as the distribution of *kara* as in the sentence *Ame ga futte iru kara, iku no wa yamero* [I will stop going, since it is raining (Figure 33.Vol.1)] of "Grammar Atlas of Japanese Dialects". This is nationwide atlas of Japanese dialects grammar edited by the National Institute of Japanese Language and Linguistics. As to the distribution of young generation, the trend has been seen to be similar to that of an older generation, i.e. no big difference has appeared between the two maps. In other words, it can be said that no changes occurred over time in the apparent time as to the conjunction particle *kara*.

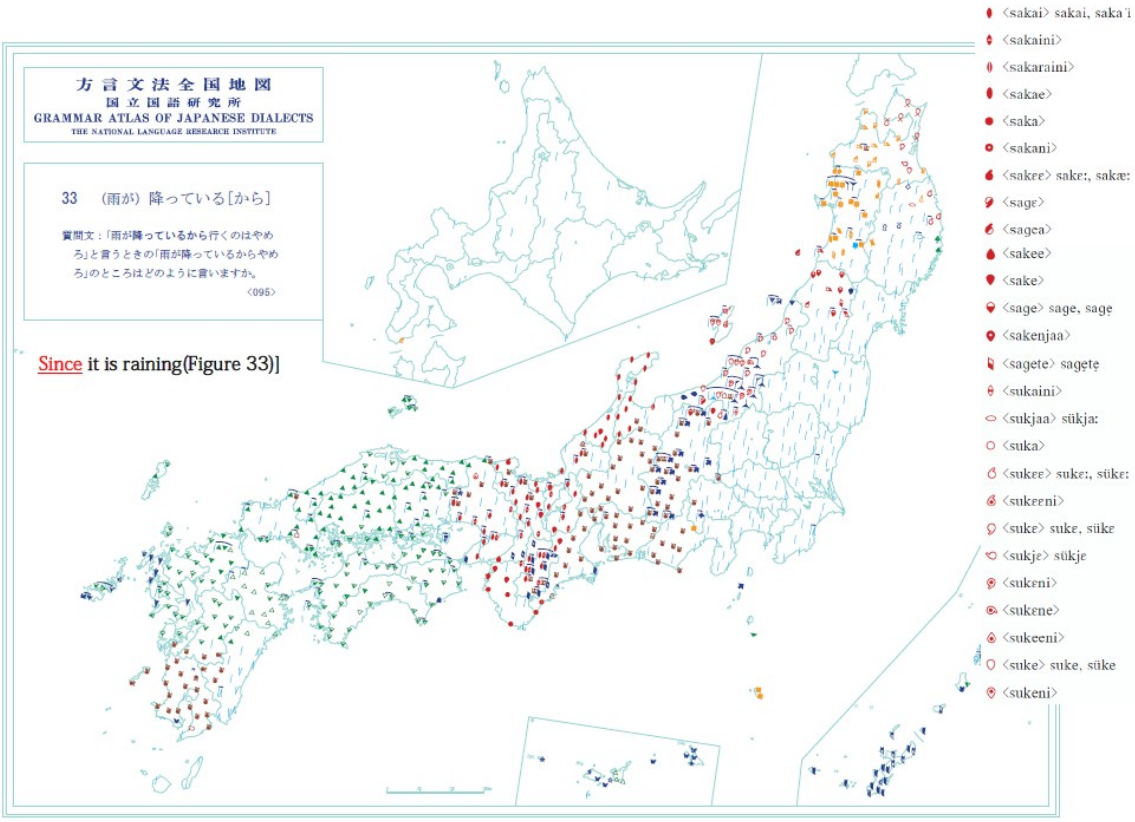


Fig 5. Since it is raining(Figure 33) of the "Grammar Atlas of Japanese Dialects"

An overview of survey results, map 1 in Sakoguchi survey report shows that *ken* is distributed mainly in Tokushima, Kagawa and Ehime prefectures in the Shikoku region, and the presence of it can also be confirmed in the prefecture of Okayama and Hiroshima of the Chugoku regions.

While the survey conducted by Fujiwara (1990) in about 80 years back shows no *ken* to be in usage, the distribution of *ken* is seen to cover the usage of *keni* or *kini* in the Sakoguchi survey. In other words, the change that occurred in the last 80 years of a time period is now reflected as *keni*> *ken*[*kini*>*kin*] which notices the diffusion of *ken*. In the survey of Sakoguchi, we can see that the usage of *keni* or *kini* is in a decrease, the trend which leads to conclude that the usage of *keni* or *kini* will be unified with *ken* from now on.

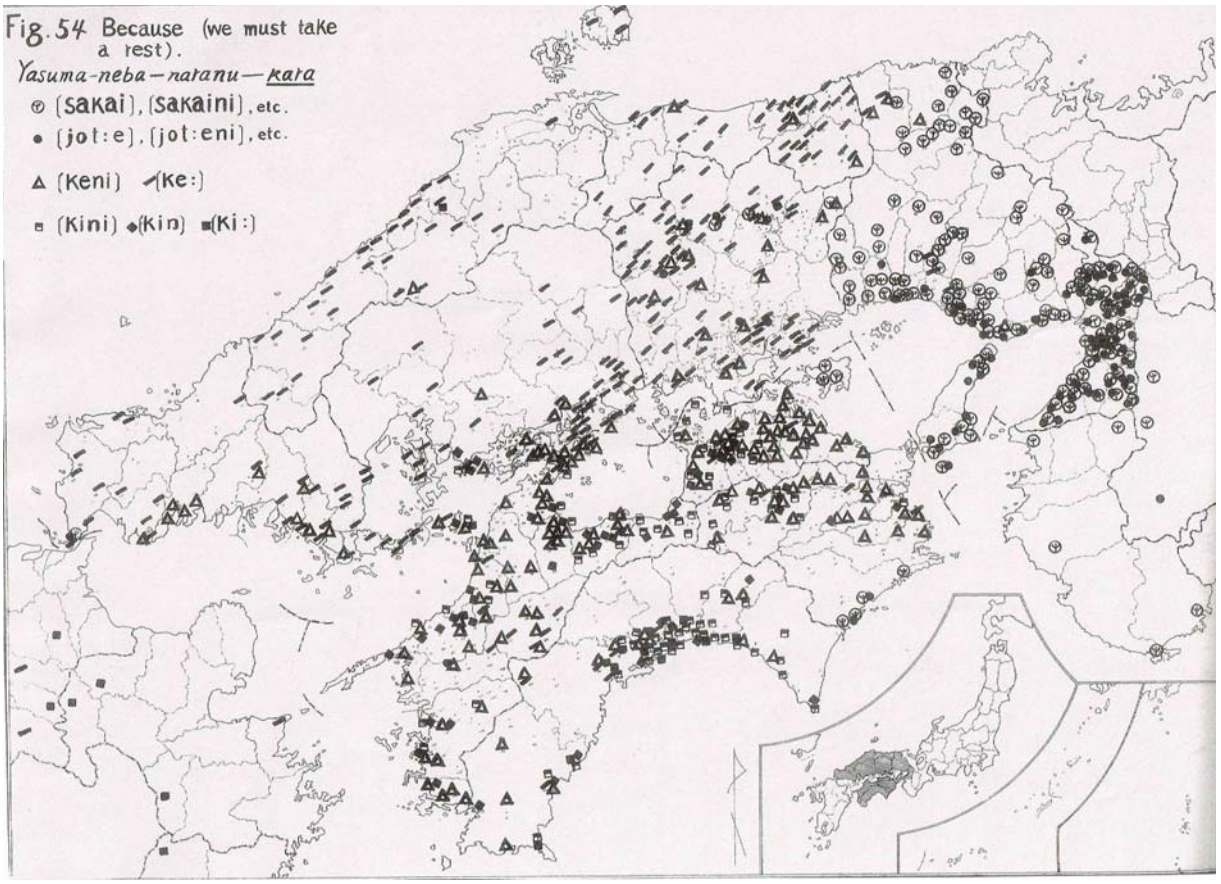


Fig 6. Fujiwara(1990)[1]

The map 2 in the LAS shows that *ke* is distributed in the Okayama prefecture, islands of Hiroshima prefecture, Yamaguchi Prefecture and northern region of Kyushu, but the Sakoguchi survey shows that the distribution of it remains predominant in Chugoku region. The Fujiwara (1990) survey also shows that the distribution is rich in *ke* . As to the *ke*, no big changes reflected in the distribution over time in the last 80 years.

For the reason of the paper width, we could not incorporate a map, but the distribution on the usage of *sakai* (*ni*) can be confirmed even by comparing the results of both LAS and Fujiwara (1990) in both the regions. In other words, for the past 80 years, it is seen that the distribution area of it did not shrink; rather it was carried out and has been used in various places.

The findings of the comparison of previous surveys show that, in general, the distribution comprised of the formatives *ke, sakai, yotte*, the Seto Inland Sea area, is still persisting with a similar distribution. With regard to *ken*, however, it did not appear at all in the survey of about 80 years ago, and it is now used in the area where *keni* or *kini* would have been once used. In the phase beginning from the survey of Fujiwara survey (1990) to LAS, we can see that *keni* or *kini* have replaced with *ken*, and now the *keni* or *kini* have almost disappeared causing to almost a merger with *ken*. The LAS survey shows that there occurred a steady decrease in the distribution area and a decline in the usage of the dialectal forms *seni, shini, shoi etd.* in various places. Given the above background, we can see a trend that various dialectal formative seen in different phases of the LAS survey has now been swallowed by *ken* with the extensive use of it in different regions.

[1] Although this was published in 1990, this survey was conducted from 1933 to 1935 by Yoichi Fujiwara.

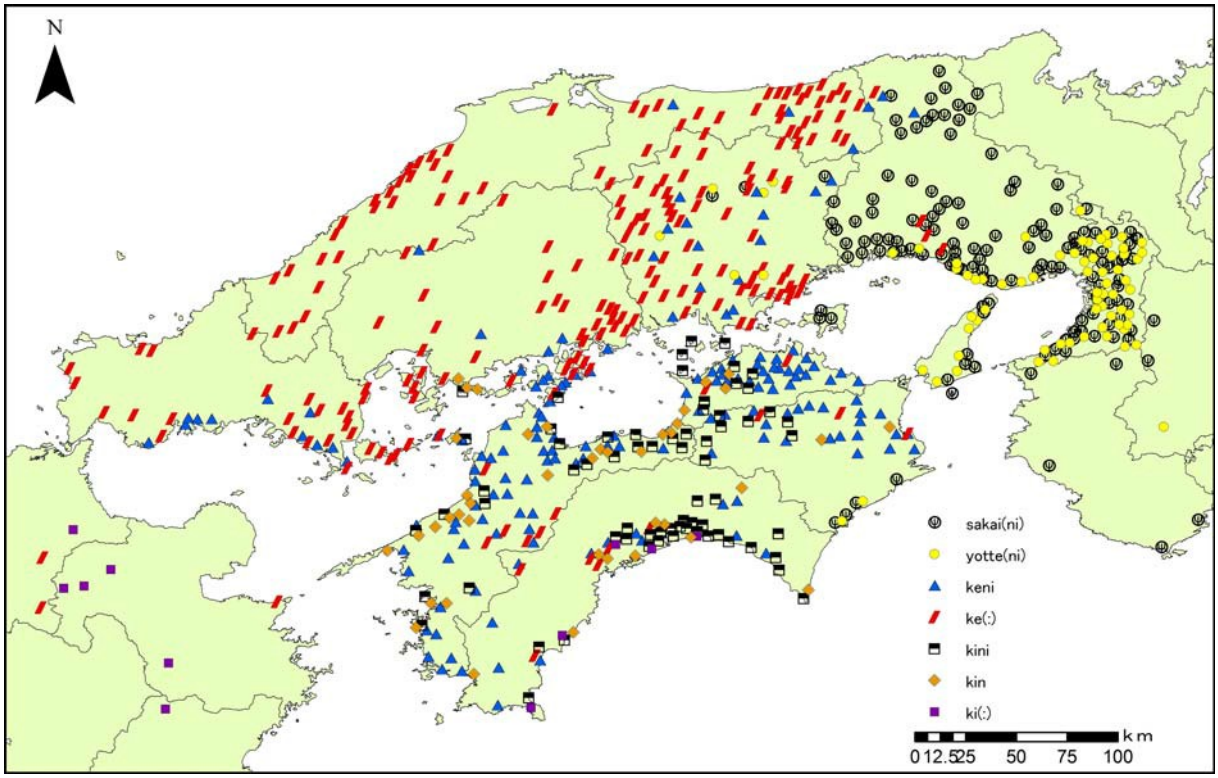


Fig. 7 Fujiwara (1990)[2]

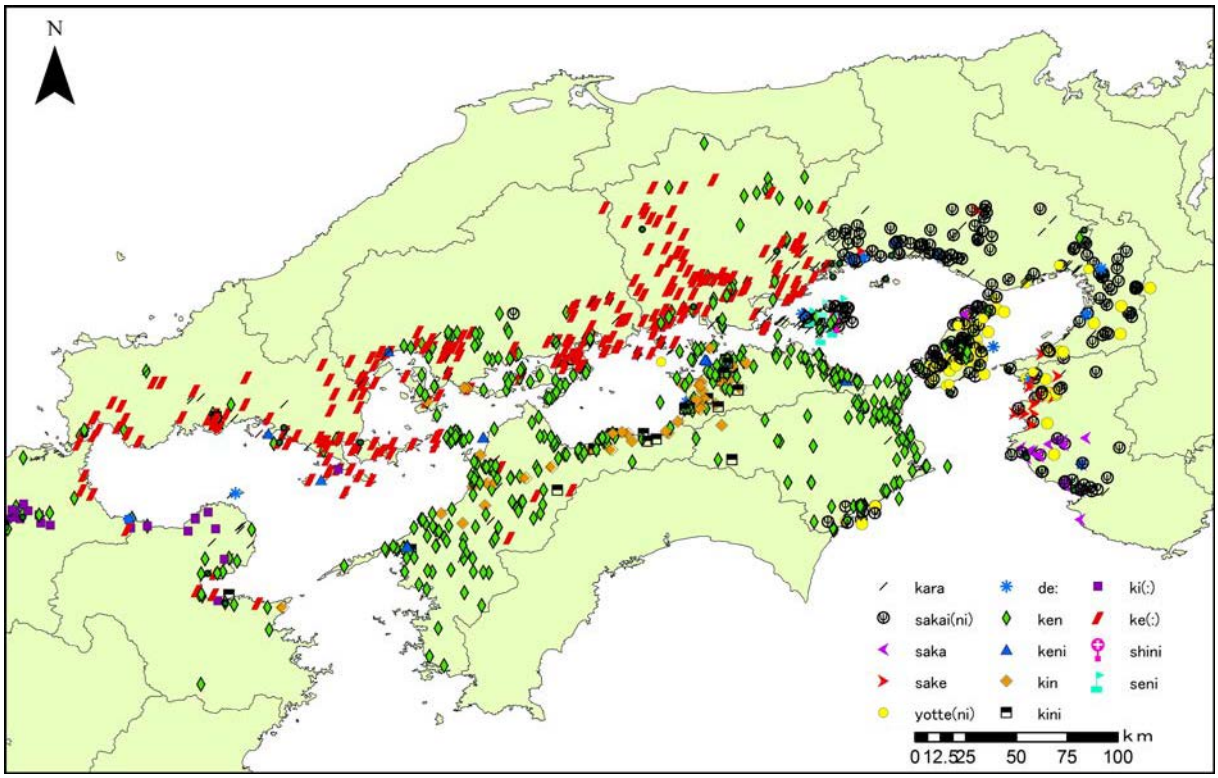


Fig. 8 Sakoguchi's survey from 2012 to 2015

[2] This map has been changed to the color version of the map in Fig. 6 in order to compare with Fig. 8.

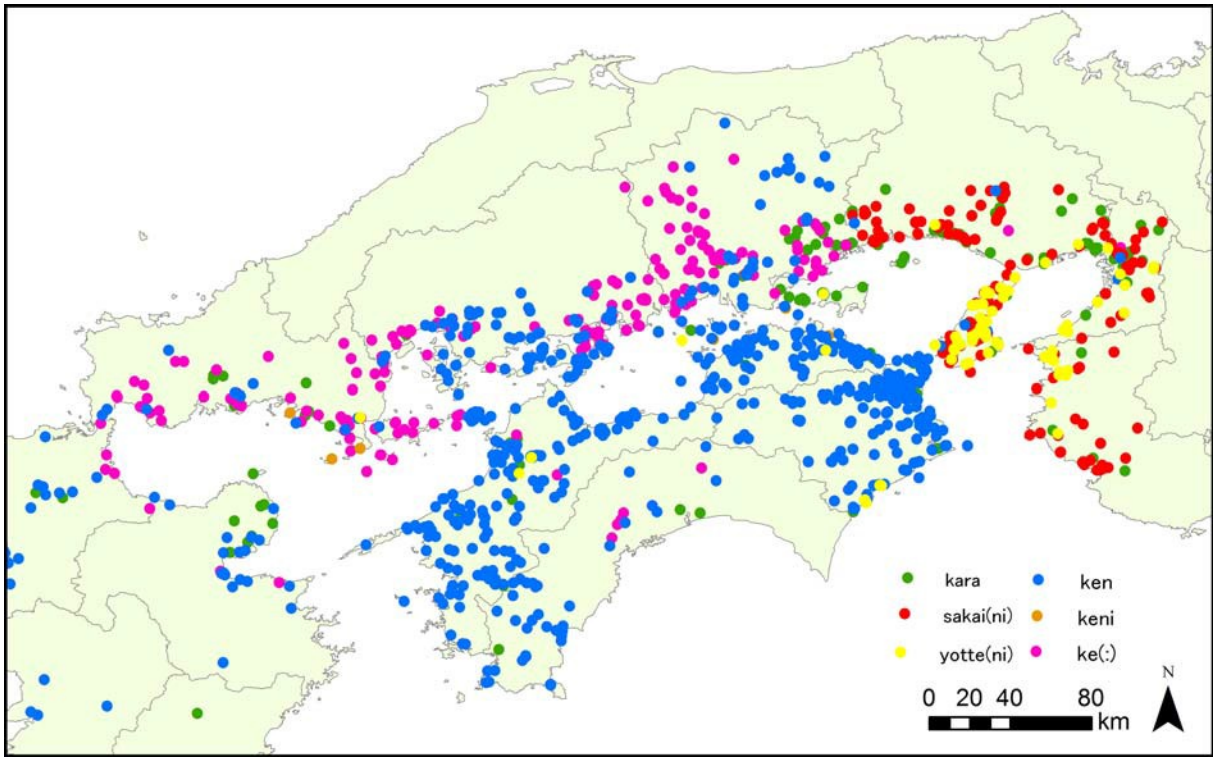


Fig. 9 The distribution of dialectal forms by Sakoguchi's survey

5. Conclusions

The findings of comparative analysis made with the previous surveys show, in general, that the formatives such as *ke, sakai(ni), yotte(ni)* which comprise the distribution in the Seto Inland Sea area, are still persisting in the same distribution area. However, with regard to the usage of *ken*, it did not appear at all in the survey about 80 years ago, but it is now used in the area where once *keni* or *kini* was used. The investigation at the phase beginning from Fujiwara (1990) to LAS shows that *keni* or *kini* have changed to *ken*, and formative *keni* or *kini* has almost disappeared causing to the merger with *ken*. In addition, the LAS survey shows the steady decrease in usage of the dialectic forms *seni, sini*, *shoi*, in addition to the decline of their usage in various places. Against this background, we can see a trend of change that various formative which are seen to be in usage during the time of the LAS survey were swallowed by the formative *ken* with a steady increase in the above-mentioned region.

References:

National Institute for Japanese Language and Linguistics. 1989. Grammar Atlas of Japanese Dialects Vol 1. Zaimusho Insatsukyoku, Tokyo. Japan.

Takeshi Shibata. 2003. Nihon Zendo no Hogen. Gengo, 32(1). Taishukan Shoten.

Yoichi Fujiwara. 1974. The Linguistic Atlas of the Seto Island Sea. Ⅰ,Ⅱ. Tokyo Daigaku Shuppan.

Yoichi Fujiwara. 1990. The Study of Linguistic Dialectology in Chugoku, Shikoku, Kinki and Kyushu. Izumi Shoin.

Sato,T. 1990. Shodoshima to Taigan Yochi no Hougen Jisho Nenso Bunpu Zushu- A Linguistic Atlas the Seto Island Sea no 25 Nen Go Tsuiseki Chosa, Dialectal Circle, Osaka Educational University.

Acknowledgements: This work supported by Grant-in-Aid for Research Activity Start-up. No.17H07293

Bhâsa Bhâbhien as a Madurese Language Dialect:
The Case of Bawean isolect on Bawean Island

Agusniar Dian Savitri

Department of Indonesian Education, Faculty of Languages and Arts, Universitas Negeri Surabaya

Abstract

The Bawean isolect debate as a language or dialect has been going on for a long time. To date, the isolect Bawean are regarded as Bawean or bhâsa Bhâbhien by its speakers. It is different for other speakers, especially Madurese speakers. For Madurese speakers, Bawean is part of the Madurese language. In relation to that, this study attempts to describe bhâsa Bhâbhien or Bawean isolect as the Madurese dialect.

Data collection method used is structured interview with informant. Data analysis using dialectometric method and bundle of isoglos. Mapping is based on lexical and phonological variations across regions that use Madurese language in East Java. In addition, there were also mapping of variations of place names in Bawean Island.

The result is, first, based on the phonological correspondence between the Bawean isolect and the Madurese language, there are areas of aspirate sound and the first syllable reduplication on Bawean Island. These two areas of sound are characteristic of Madurese language. The first reduplication is still maintained in East Madura dialect. Second, the lexical variation between Bawean isolect and Madurese languages in all areas of Madurese language in East Java is at the level of dialect differences. The isoglos line contained in fifteen semantic fields tends to group Bawean Island with Madura Island, only six isoglos maps show there is a thickening isogloss on Bawean Island. That means the lexical used in Bawean tends to be the same as the lexical Madurese language. The results show that bhâsa Bhâbhien or Bawean isolect is the same language with Madurese language, not a different language. Thirdly, the naming variation of places on Bawean Island shows much use of the Madurese language lexicon. Based on these three things, it can be concluded that bhâsa Bhâbhien is a dialect of the Madurese language or can be called Bawean dialect.

Keywords

bhâsa Bhâbhien**,** phonological correspondence, isolect status, geographical name, dialect of Madurese

1. Introduction

Bhâsa Bhâbhien or Bawean isolect is an isolect spoken on Bawean Island. The island is located in the northern part of Gresik regency. Administratively, Bawean Island is part of Gresik Regency.

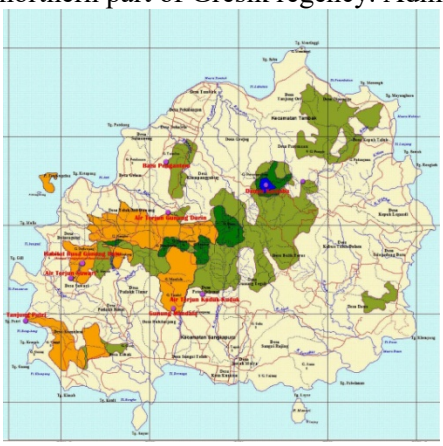


Figure 1 The Map of Bawean Island (kabgresik, 2018)

Figure 2 The Map of Bawean Island in East Java Province (DPRD Jatim, 2018)

By its speakers, bhâsa Bhâbhien or Bawean isolect is considered a different language with Javanese and Madurese language. Therefore, Bawean isolect are called bhâsa Bhâbhien 'Bawean language' by the people of Bawean Island. This is different from the Madurese speakers in Madura Island. According to them, Bawean isolect or bhâsa Bhâbhien are Madurese. Sofyan (2008: 2010) divides bhâsa Bhâbhien as a language and as a dialect of Madurese. Bhâsa Bhâbhien as a language if viewed from sociolinguistics, while bhâsa Bhâbhien as a dialect when viewed from linguistics (Sofyan, 2008, Sofyan, 2010). According to him, Madura speakers in Madura Island consider Bawean isolect as Madurese language, while isolect in Kangean is considered as different language.

The recognition of Bawean isolect as a different language from Madurese by speakers on Bawean Island is based on two things. First, geographically Bawean Island is far apart from Madura Island and Java Island. In addition, administratively, Bawean Island is part of Gresik Regency which is a Java language area. This is different from the Tapal Kuda area in East Java. Although geographically separated from Madura Island, the Tapal Kuda area is known as the Java-Madura region by the people of East Java (Savitri, 2015). Secondly, there are oral stories relating to the origin of Bawean society. There are four stories about it. First, the Bawean people come from the Malays (Palembang, Banjar, and Mandar) who transit on Bawean Island before entering the island of Java. Second, Bawean people come from Madurese fishermen who stop by on Bawean Island to avoid bad weather. They finally settled there. Third, the Bawean people came from Majapahit troops who were stranded on Bawean Island due to bad weather. Fourth, the story of the invasion of Sultan Cakraningrat (Sultan Sumenep 4th) who later occupied Bawean (Vredenberg, 1990). In relation to these two sources, it appears that the recognition of Bawean isolect as bhâsa Bhâbhien is based on the geographical and historical administrative factors of the Bawean society, not based on linguistic factors.

Ardiana's research (1997/1998) states that the language used in Bawean Island is similar to Madurese. However the statement that the language used in Bawean is Madurese has not been explicitly stated to date. The study of Bawean tends to support statements about the language of Bawean or bhâsa Bhâbhien. Based on this, this study attempted to describe bhâsa Bhâbhien or Bawean isolect as a dialect or a different language from Madurese. It needs to be done in order to know whether bhâsa Bhâbhien or Bawean isolect is a different language or the same language with Madurese language, although geographically Bawean Island far apart with Madura Island.

2. Method

Data collection method used is structured interview with informant. The list of questions used is a list of Badan Bahasa questions that have been modified. Data analysis used full permutation technique and bundle of isogloss. Mapping is based on lexical and phonological variations across regions that use Madurese language in East Java. In addition, there were also mapping of variations of place names in Bawean Island.

3. Discussion

3.1 *The Bhâsa Bhâbhien correspondence with Madurese Langugage*

The study of phonological variation bhâsa Bhâbhien with Madurese language in East Java resulted in eleven correspondence between bhâsa Bhâbhien with Madurese language. Of the eleven correspondences, the three correspondences are considered important in determining the similarity of bhâsa Bhâbhien character with the Madurese language. the correspondence is the correspondence of the aspirate sound, correspondence [i] with [ɛ], and the syllable correspondence. Example of correspondence of aspirate sound as follows.

Table 1 Correspondence of Aspirate Consonant

no	gloss	bhâsa Bhâbhien	Madurese
1	bahu	[bʰəuh], [bʰəu]	[bʰəuh], [bʰəu]

2	besan	[bʰisan]	[bʰisan], [bisan]
3	pintu	[labʰəŋ], [labəŋ]	[labəŋ]
4	belimbing	[bʰəlimbʰiŋ]	[bʰəlimbʰiŋ], [bəlimbʰiŋ], [bəliŋbʰiŋ]
5	ikat kepala	[ɔdʰəŋ]	[ɔdʰəŋ], [ɔdəŋ]
6	bisul	[budʰun]	[budʰun], [bəbudʰun]
7	ikan	[jʰukɔʔ]	[jʰukɔʔ]
8	horse	[jʰəran]	[jərən]
9	mudah	[gʰəmpaŋ]	[gʰəmpaŋ], [gəmpaŋ]
10	alu	[gʰəntɔŋ]	[gʰəntɔŋ]

The sound of aspirate is one of the characteristics of Madurese language. In Madurese, the sounds of aspirate occur in voiced plosive sounds, namely and [b], [d], [ɖ], [j], and [g]. It also happened to bhâsa Bhâbhien. The example in Table 1 shows that bhâsa Bhâbhien has an aspirate sound that occurs in the voiced plosive sounds [b], [d], [ɖ], [j], and [g]. Of all the correspondence data of aspirate sounds, bhâsa Bhâbhien tends to maintain the sound of the aspirate. In fact, some Madurese subdialects of the West Madura dialect tend to vary the sound of the aspirate with nonpirate sounds, eg [ɔɖəŋ], [gəmpaŋ], dan [labəŋ] (example in table 1).

The syllable correspondence is the correspondence between the two-syllable words with the word formed with the initial syllable reduplication so that the word has three syllables. Examples of such correspondence are listed in table 2 below.

Table 2 Syllable Correspondence

no	gloss	bhâsa Bhâbhien	Madurese
1	tiga	[tatəlɔʔ]	[təlɔʔ]
2	lima	[lɛlɛmaʔ]	[lɛmaʔ]
3	perempuan	[bəbiniʔ]	[bəbiniʔ], [bibiniʔ], [bəbinɛʔ], [biniʔ], [binɛʔ]
4	laki-laki	[lalakɛʔ]	[lalakɛʔ], [lakɛʔ]
5	kupu-kupu	[kəp:ər]	[kəp:ər], [kap:ər], [kakap:ər], [gəp:ər], [gəgəp:ər]

Similar to the sound of aspirate, silabel reduplication is one of the characteristics of the Madurese language (Pawitra, 2009: xxv). In the reduplication, the duplicated syllables are the initial syllables or the final syllables. Based on all data, the most widespread silabel reduplication is final syllable reduplication, such as [paykəp:ay] ‘fan’ and [nɔŋgunɔŋ] ‘mountain’. Early syllables duplication tends to be maintained by the Madurese language of the eastern Madura dialect which is located in the eastern horseshoe area, Sumenep, and some Sumenep islands (Savitri, 2015). According to table 2, the initial syllable reduplication is also found in bhâsa Bhâbhien. Thus, the initial syllable reduplication that exists in bhâsa Bhâbhien is the preservation of the relics by the Madurese language. That is supported also by the spread of early syllable reduplication that tends to occur in Bawean Island. It can be seen on the following map.

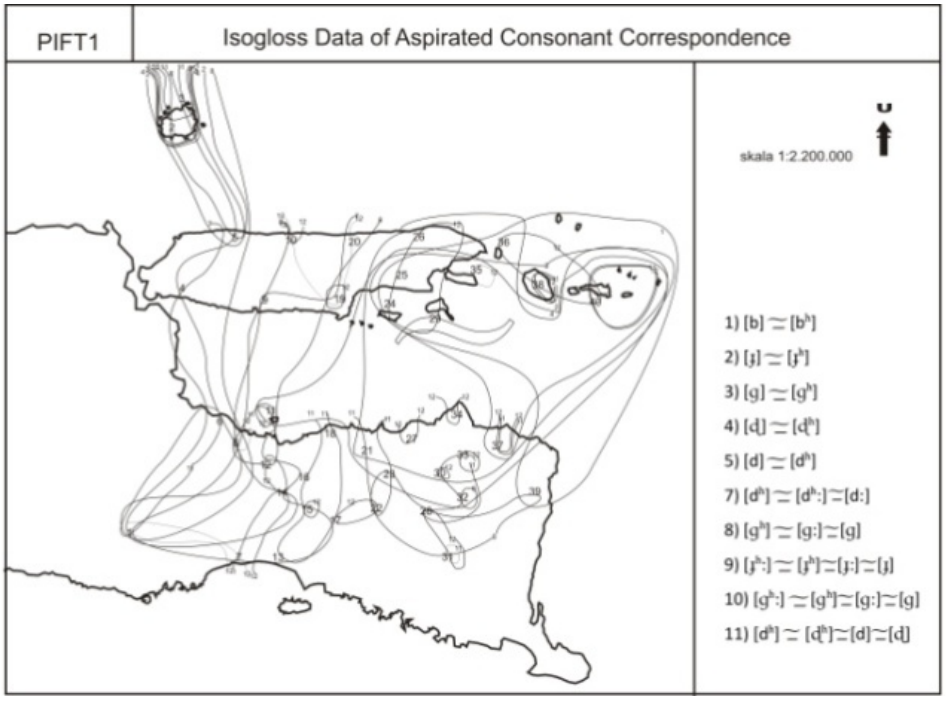


Figure 3 The Map of isogloss data of Aspirated Consonant Correspondence

Map on the fig 3 shows the grouping of aspirated consonant that tends to occur in eastern and Bawean islands. The thickening of isoglos occurs on Bawean Island. That means the bhâsa Bhâbhien tends to give rise to aspirated consonant, while other areas-which are the regions of Madurese-have varied the aspirated consonant with non-aspirated consonant. Examples of variations of aspirated consonants with nonaspirated consonants are listed in Table 1.

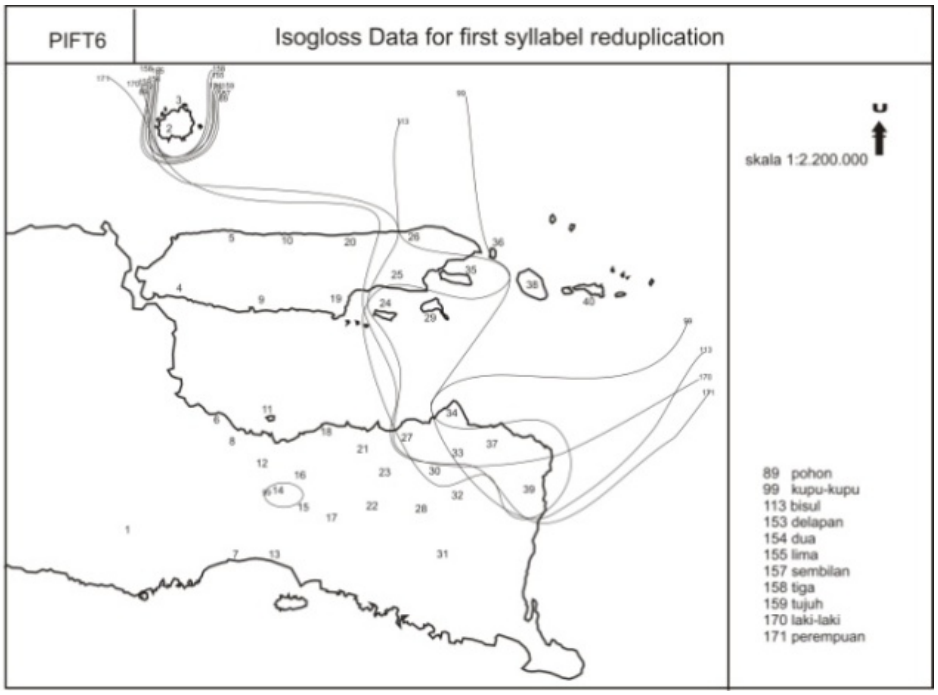


Figure 4 The Map of isogloss data of syllable reduplication

The same is true of map 2. The map shows an initial syllable reduplication grouping in the eastern region. The thickening of isoglos occurs on Bawean Island. That means the bhâsa Bhâbhien

tend to bring up early syllables. Other areas tend to vary it by eliminating the initial syllables so that the form appears two syllables. Examples of these are listed in Table 2.

The third correspondence is a correspondence [i] with [ɛ]. Vokoid [ɛ] is one of the sounds that characterize Madurese language. its tends to be in the area which use the western dialect of Madura, namely Sampang, Bangkalan, and western horseshoe area. The example of that is [kɔniŋ] 'yellow' in east Madura dialect will be uttered [kɔnɛŋ] in west Madura dialect. The vocoid correspondence [i] with [ɛ] in the eastern Madura dialect with the western Madura dialect also occurs in bhâsa Bhâbhien. That is, bhâsa Bhâbhien is part of the vocoid [ɛ] region in correspondence [i] with [ɛ].

The phonological correspondence that distinguishes bhâsa Bhâbhien and Madurese language is syllable [ŋa] with [a] and [ma] with nasal, such as [ŋabəriʔ] 'to give', [maɔḍiʔ] 'turn on' dan [maləḍʰuʔ] 'explode'. Silabel [ŋa] corresponds to [a] in Madurese. Thus, if there is [ŋalɛrɛʔ] 'glancing' and [ŋaduləŋ] 'feeding' in bhâsa Bhâbhien, the form will vary with [alɛrɛʔ] and [aduləŋ] in Madurese. Silabel [ma] in bhâsa Bhâbhien corresponds to nasal. That means, if there is [maɔḍiʔ] 'turn on' in bhâsa Bhâbhien, the form will vary with [ŋɔḍîʔi] in Madurese. Although those two distinguish bhâsa Bhâbhien with the Madurese language, the two forms—syllable [a] and [ma]—is actually a prefix in Madurese (Pawitra, 2009: xxvi). Referring to the principle of geographical diffusion (Chambers and Trudgill, 2004), the two etyma are a relics of Madurese language. Thus, the existence of the two etima actually reinforces the phonological similarity between bhâsa Bhâbhien and Madurese.

3.2 *Status of Bhâsa Bhâbhien with Madurese Language*

Determination of bhâsa Bhâbhien status was done by counting lexical variation with full permutation technique at two points on Bawean Island, namely Sangkapura and Tambak. These two points are permutated with all points using Madurese language. The result is Isolect in Sangkapura and Tambak is a Madurese language dialect. Thus, bhâsa Bhâbhien is a Madurese language. Here is a map of the permutation.

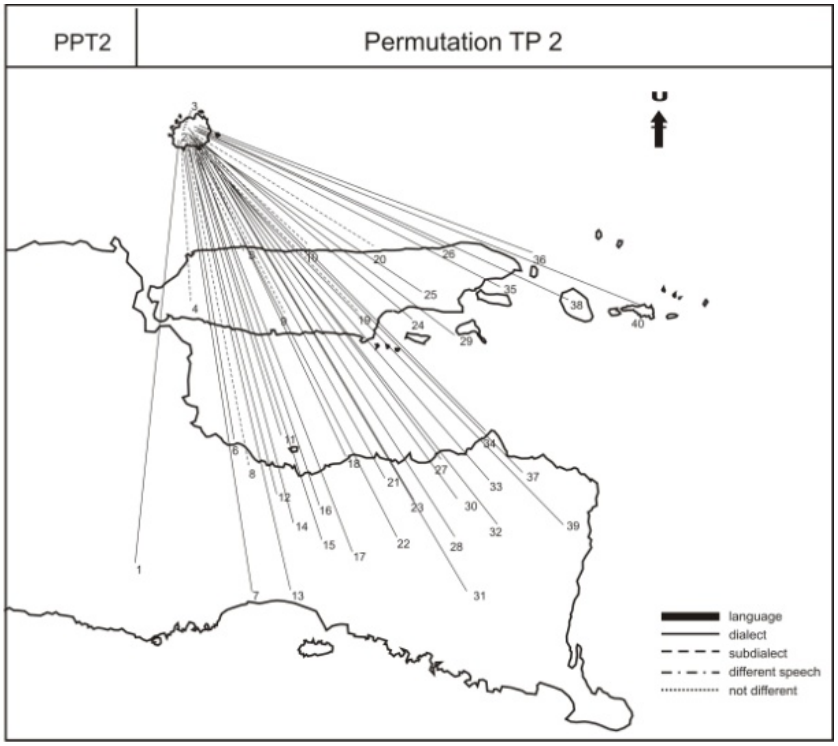


Figure 5 Permutation at Observation Point 2 (TP 2)

In figure 5 we can see that the biggest difference between observation point 2 (Sangkapura) and other observation points in East Java is dialect difference. It shows that bhâsa Bhâbhien in Sangkapura Bawean is a dialect of the Madurese language.

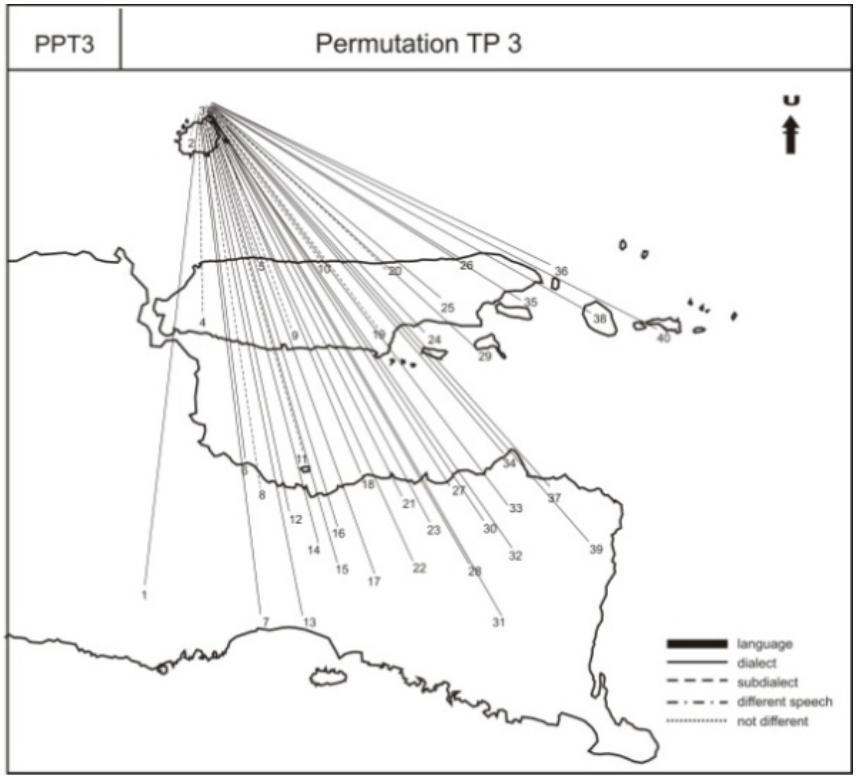


Figure 6 Permutation at Observation Point 3 (TP 3)

Similar to figure 5, in figure 6 we can see that the biggest difference between observation point 3 (Tambak) and other observation points in East Java is dialect difference. It shows that bhâsa Bhâbhien in Tambak Bawean is a dialect of the Madurese language.

The lexical variations between bhâsa Bhâbhien and Madurese tend to occur in the semantic fields of body parts; pronouns and greetings; plants and trees; animals; character, adjective, colour; motion and work. Isoglos bundles on these six semantic fields show a very noticeable thickening on Bawean Island (TP 1 and 2). Here is a map of the isoglos.

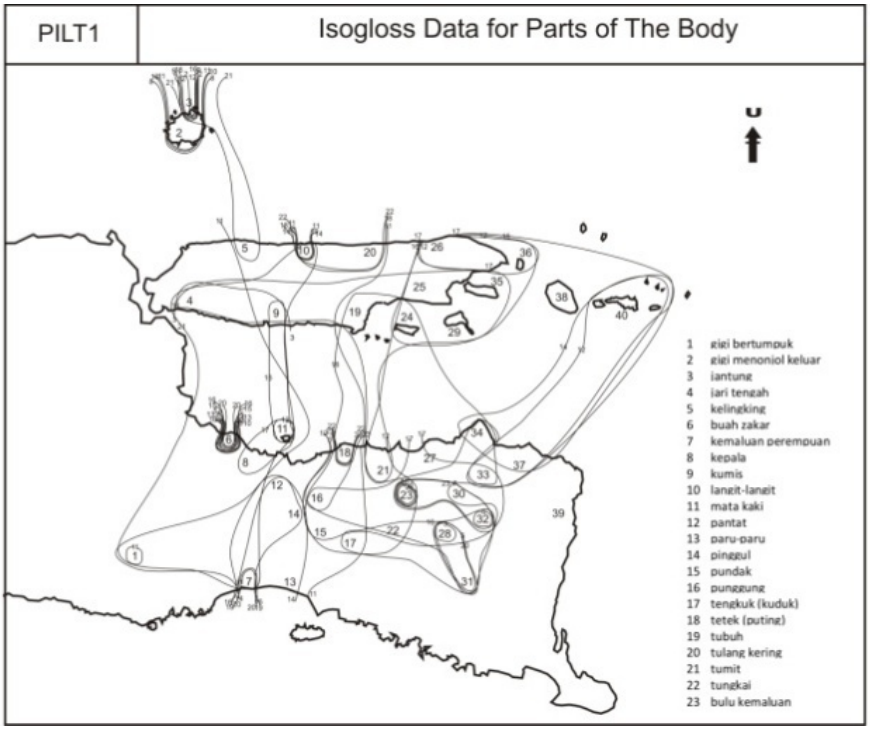


Figure 7 The Map of Isogloss Data For Parts of The Body

Figure 7 (semantic fields of body and semantic fields of pronouns and greetings) show thickening on Bawean Island (TP 1 and 2). That means there are many lexical variations on Bawean Island that are different from other TP in the two semantic fields. Examples of such variations are [rɔmpɔʔ] 'stacked teeth', in other regions [sɔlɛh] and [sɔŋil], [ɔlɔ] 'head' in other areas [cɛt:ak].

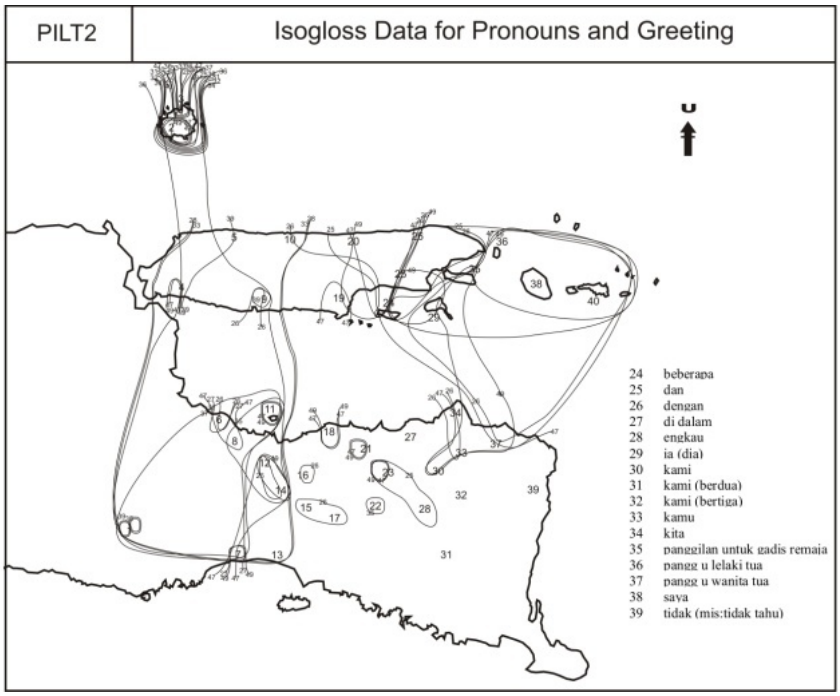


Figure 8 The Map of Isogloss Data For ponouns and Greetings

Figure 8 show thickening on Bawean Island (TP 1 and 2). That means there are many lexical variations on Bawean Island that are different from other TP in the two semantic fields. Examples of such variations are [ɛsɔn] 'I' in other areas [əŋkɔʔ] and [sɛŋkɔʔ].

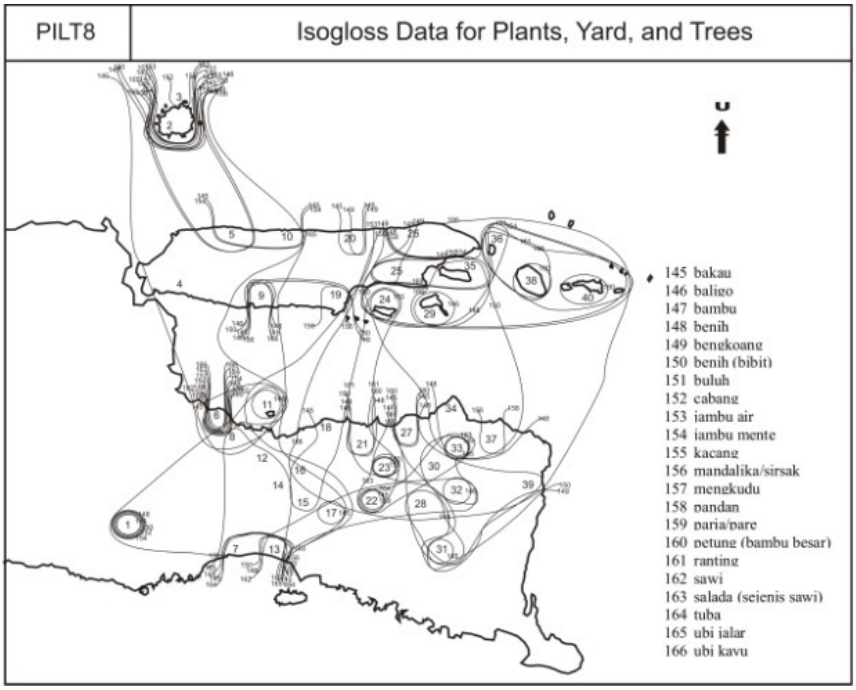


Figure 9 The Map of Isogloss Data For Plants, Yard, and Trees

Figure 9 show thickening on Bawean Island (TP 1 and 2). That means there are many lexical variations on Bawean Island that are different from other TP in the two semantic fields. Examples of such variations are [ɔrɛ] 'bamboo' in other areas [prɛŋ] and [pər:ɛŋ], [bəŋɛ] 'pandanus' in other areas [pandʰən] and [pandən].

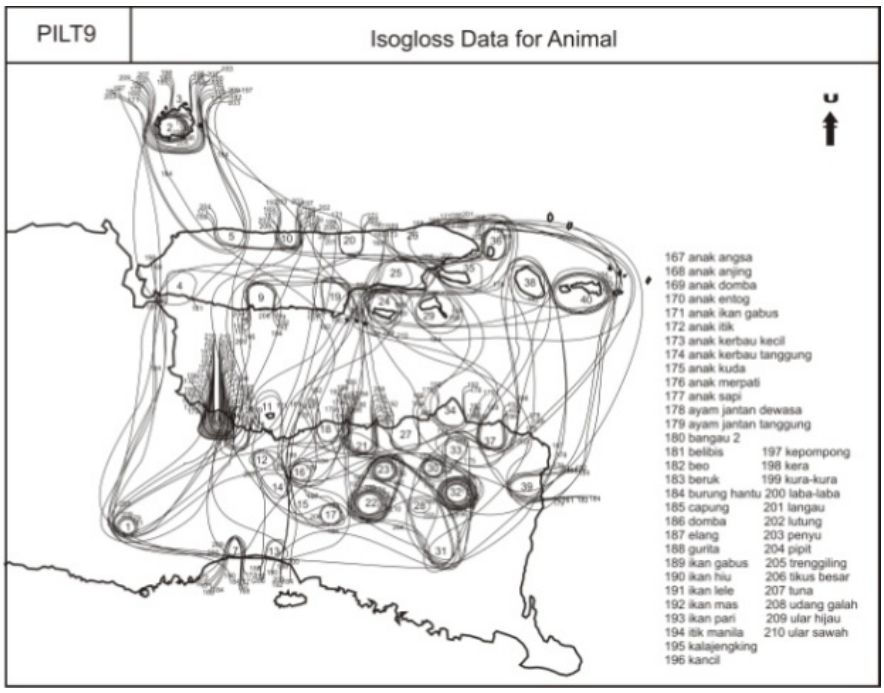


Figure 10 The Map of Isogloss Data For Plants, Yard, and Trees

Figure 10, the thickening on Bawean Island (TP 1 and 2) means there are many lexical variations on Bawean Island that are different from other TP in the two semantic fields. Examples of such variations are [bukal] ‘apes’ in other areas [mɔʈak], [ŋɔʈak], and [kəʈ:aŋ].

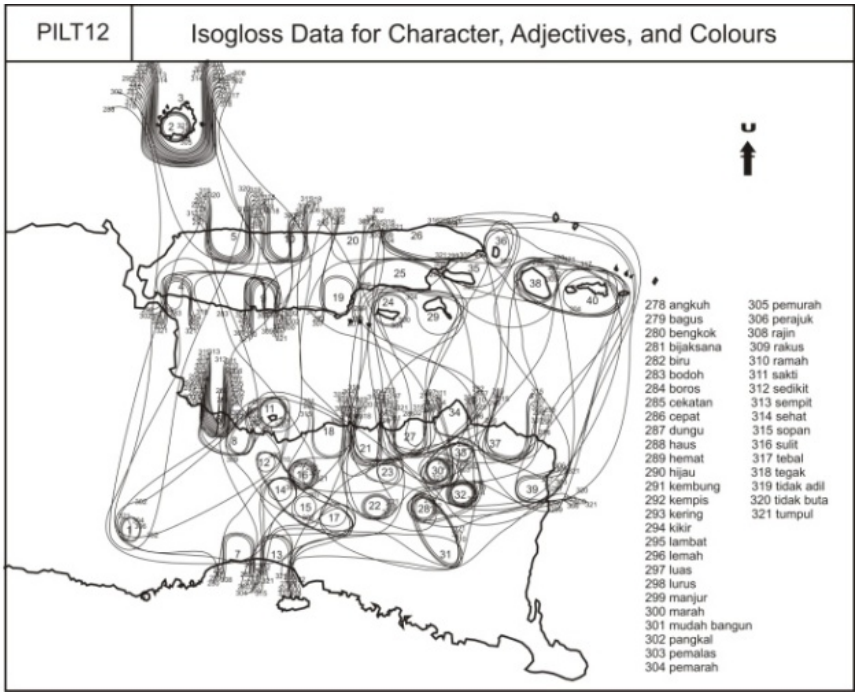


Figure 11 The Map of Isogloss Data For Character, Adjectives, and Color

Figure 11, the thickening on Bawean Island (TP 1 and 2) means there are many lexical variations on Bawean Island that are different from other TP in that semantic fields. Examples of such variations [ijo] ‘green’ in other areas [biruh]. That variation is inovation in Bawean because there is not lexical of [ijo] in Madurese. In that language, ‘green’ and ‘blue’ have a same lexical, [biruh].

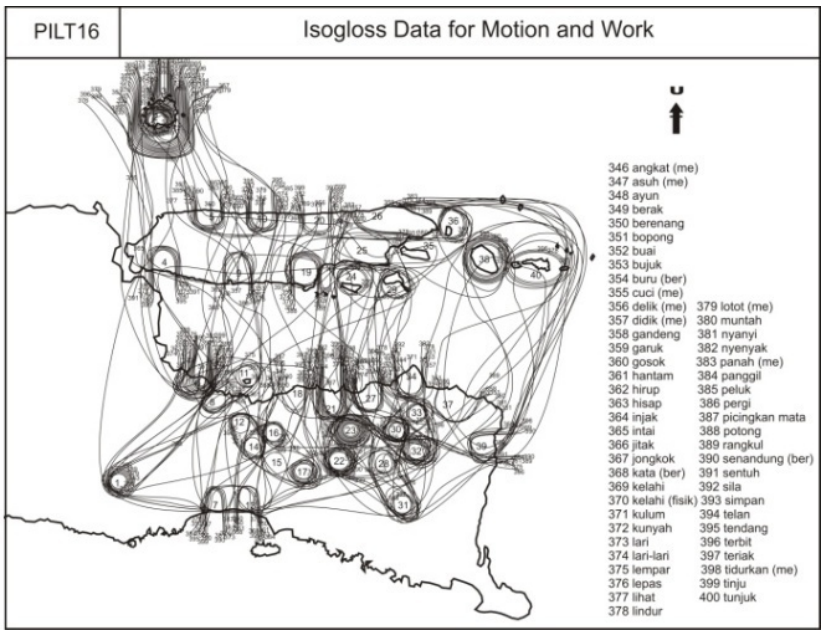


Figure 12 The Map of Isogloss Data For Motion and Work

Figure 12 shows thickening on Bawean Island (TP 1 and 2). That means there are many lexical variations on Bawean Island that are different from other TP in the two semantic fields. Examples of such variations [ɳɛrəp], [ŋɛrəʔ] 'breathe' in other regions [ɳərguʔ], [ɳərŋʊt] 'washes' in other areas [sasas:ah], [ɳas:ah]; [katɔwɛ] and [ŋɔɖɛi] 'call' in other areas [ŋɔləʔ], [ɳəl:ɔk].

Based on the isoglos bundle map of the six semantic fields, it appears that lexical variations between Bawean Island and other areas are in terms of geographical conditions. Not all plants and animals that exist on the island of Java and Madura is on the island of Bawean, and vice versa. The same is true of the semantic field of motion and work. Madurese ethnic in Java live in mountainous areas or coastal areas. It is different from Bawean society. Most of them work as sailors or work abroad (as TKI). Hence, there is a lexicon in the semantic field of motion and work which are owned Madurese language on the islands of Java and Madura but not possessed by bhâsa Bhâbhien, nor vice versa.

Although there is a lexical variation between bhâsa Bhâbhien with Madurese language which tends to occur in six semantic fields, the lexical variation does not make bhâsa Bhâbhien as a distinct language with Madurese language. It is based on a full permutation calculation that results in a 59% percentage. This percentage shows variation of bhâsa Bhâbhien with the Madurese language at the dialect level. In addition, of the 1111 gloss which asked on informants, which is a lexical variation of only 401 gloss. There are 527 gloss that vary phonologically, 25 gloss has no etyma, 168 glos have the same etyma. Thus it can be concluded that the variation between bhâsa Bhâbhien with Madura language tends to be phonological variation. It shows that bhâsa Bhâbhien tends to use the same lexical as Madurese language. Which means, bhâsa Bhâbhien or Bawean isolect is Madurese language.

3.3 *The use of Madura lexicon in geographical naming on Bawean Island*

The phonological and lexical variations that have been described have been sufficient to prove bhâsa Bhâbhien as the language of Madura. But to reinforce it, will also be seen geographical naming on Bawean Island. This needs to be done because Madurese speakers tend to use Madurese language in geographic naming, although administratively the geographical name uses other languages. An example of this is the mention of villages and towns in the Tapal Kuda area. In that area, a village that is administratively named *Sumberingin* will be called [sɔmbʰərriŋin] by Madurese speakers. Similarly the city of *Situbondo* will be called [patɔʔan] by Madurese speakers. It happens because the Madurese speakers use their language for the geographical naming. Thus, it should be seen whether in geographical naming on Bawean Island, speakers on Bawean Island using Madurese or Bawean lexical.

Geographical naming is seen from two things, namely the name of the village and the name of the place. Place name is the name of the place that formed from nature such as mountains, lakes, rivers, which is considered typical in Bawean Island. Based on these two things, three things are generated. First, there is a place name that uses Madurese language. The lexical which used is a Madurese lexical or has a Madurese language structure. Secondly, there is a place name that uses Java language. The lexical which used is a javanese lexical. Third, there is a place name that uses bhâsa Bhâbhien. The lexical used in naming the place is not a lexical of Madurese or Javanese. These three things are illustrated in the following map.

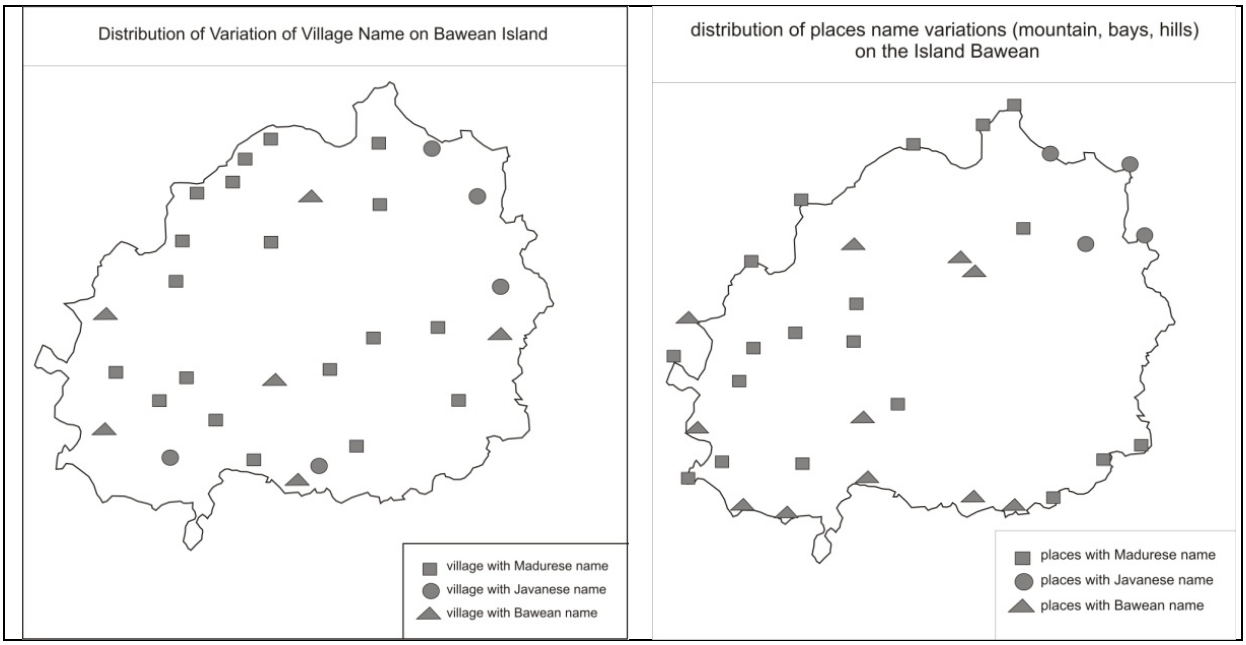


Figure 13 The Map of Distribution of village name and places names variation in Bawean Island

Of the two maps, it appears that the most extensive distribution is the name of the village and the name of the place that uses Madurese language. Examples of such names are the village of Dâun [ɖɔʊn], the village of Bululanjhâng [bululanjʰəŋ].

Based on the geographical naming it can be seen that the speakers bhâsa Bhâbhien tend to use lexical of Madurese language in geographical naming. That means the language of Madura is indeed used by speakers on Bawean Island. It supports the previous result which states that bhâsa Bhâbhien is a dialect of the Madurese language.

4. Conclusion

the correspondence of the aspirated consonant, correspondence [i] with [ɛ], and the initial syllable correspondence show that bhâsa Bhâbhien is not only the area of the aspirated consonant, the initial syllables, and the vokoid [ɛ], but the relics area of the sounds of Madurese language. The result of the lexical permutation also shows that bhâsa Bhâbhien differs dialect with the language of Madura. Both are also supported by naming villages and places on Bawean Island that tend to use Madurese language. Thus, it can be concluded that bhâsa Bhâbhien is the same language as Madurese language. It can also be said that bhâsa Bhâbhien is a dialect of the Madurese language or can be called a dialect of Bhâbhien or Bawean dialect.

References

Adrian Pawitra. *Kamus Lengkap Bahasa Madura*. Jakarta: Dian Rakyat, 2008.

Ardiana, Leo Idra, dkk. 1997/1998. “Bahasa yang digunakan di Pulau Bawean”. Hasil Penelitian. Surabaya: Proyek Pembinaan Bahasa dan Sastra Indonesia dan Daerah, Provinsi Jawa timur.

Chambers, J.K. dan Trudgill, P. 2004. *Dialectology*. Second Edition. Cambridge: Cambridge University Press, 2004.

Dewan Perwakilan Rakyat Daerah Provinsi Jawa Timur (DPRD Jatim). 2018. “Peta Daerah Pemilihan”. http://dprd.jatimprov.go.id/dapil/id/9/. Diakses 12 April 2018.

Savitri, Agusniar Dian. 2015. “Variasi Fonologis Bahasa Madura di Seluruh Jawa Timur. disertasi. Universitas Indonesia

Sofyan, Akhmad. 2008 (a). *Variasi, Keunikan, dan Penggunaan Bahasa Madura*. Sidoarjo: BBJT Pusat Bahasa Depdiknas.

Sofyan, Akhmad. 2010. “Fonologi Bahasa Madura”. *Jurnal Humaniora*. No. 2 Juni 2010, Volume 22, p. 207—218.
Vredenberg, Jacob. 1990. Bawean dan Islam. Jakarta: INIS

JAVANESE LANGUAGE IN CIREBON: THE LANGUAGE AT THE BORDERLINE

Supriatnoko
Department of Accounting
State Polytechnic of Jakarta
supriatnoko@gmail.com

ABSTRACT

Javanese Language in Cirebon exists at the borderline between Province of Central Java and Province of West Java, and grows in Sundanese Language environment. The speakers like to call the language they use as **Cerbon Language or Crebon Language**. Cerbon Language brings up the problem in geolinguistics. The problem is whether Cerbon Language as one dialect of Javanese Language shows variations because of the influence of Sundanese Language, It happens because according to the geographic, villages where the speakers of Cerbon Language live is adjacent to the villages where the speakers of Sundanese Language live on. The research uses qualitative and quantitative approach. Village is treated as a research unit. 55 villages are treated as the sample by using criterion-based selection or purposive sample technique. Informans are treated as sources of data. Data are collected by using questionnaire which contains list of questions in it, obtained through participative observation method, face-to-face conversation and recording techniques. Data are analysed by isogloss and dialectometry methods, using triangular interregional and *polygones de Thiessen* techniques. The result of analysis is presented descriptively. The research found as follows: From 558 maps, it is discovered 209 maps of lexical variations and 349 maps of phonological variations. It is also discovered 3 groups of Cerbon Language speakers, that is, the first group pronounces the open vocal phoneme /a/ as [a]. It is possibly influenced by Sundanese Language, the second group tends to pronounce the open vocal phoneme /a/ as [ɔ] like Javanese pronunciation, and the third group tends to pronounce the open vocal phoneme /a/ as [a], and [ɔ]. The achieving of lexical variations—according to Guiter (1973)—is ”Different Speech”, and the achieving of phonological variations is ”Different Subdialect”. Cerbon Language has different grammatical vocabulary from Javanese Language standard. Thus, speakers of Cerbon Language are difficult to be understood by the other speakers of Javanese Language, in the light of the frequency using grammatical vocabulary tends to be high from all dialects of Javanese Language, so it can be declared that gramatical vocabulary at Cerbon Language indicates its speakers identity in Cirebon.

Key words: *Cerbon Language, lexical variation, phonological variation, grammatical vocabulary, speakers identity*

1. Introduction

Cirebon (*municipalit*y and *regency*) is located at the border of Province of West Java with Province of Central Java. It belongs to Province of West Java region. Cirebon as the part of Province of West Java borders on Kuningan Regency, Majalengka Regency, and Indramayu Regency (see figure 1). The inhabitant of Kuningan Regency and Majalengka Regency speaks Sundanese Language and the inhabitant of Indramayu Regency, some speak Javanese, a few speak Sundanese Language.

The inhabitants of Cirebon can be divided into two speakers, namely the inhabitant who speaks Sundanese and the inhabitant who speaks Javanese. Between the two speakers, it can be found the speakers who are able to speak Sundanese and Javanese. Both of the two speakers claim that they are the Cirebon people.

Based on the information from the informers, the origin inhabitant of Cirebon is Sundanese people. Then they became the parts of Padjadjaran Empire. They spoke Sundanese Language. Padjadjaran Empire treated Sundanese Language as the official language of the empire. “Batutulis Inscription” in

Bogor and "Kawali Inscription" in North Ciamis, and "Kebantenan Inscription" in Bekasi, found in the 16 century, became the evidences that Sundanese Language used as the official language of the empire (see Ekajati, 2005[2]: 87-88; Hardjasaputra and Haris, 2011: 15-17).

At the 14 century, Javanese Language went into Cirebon. It was taken by the fishermen who came from Central and East Java and maintained a permanent residence at Muara Jati Port and at along the seashore near the port. Muara Jati was the famous port in Cirebon at that time (Supriatnoko, 2015:3). It had ever become the international port, as the port of call for international trade stripe and called "silk stripe trading" (Atja, 1986:30; see Adeng, etal, 1998:49; Buchori and Kuswiah, 2001: 19; Casta and Taruna, 2008: 8). Because Cirebon has strategic area between Javanese land and Sundanese land, Cirebon acted as a bridge for Javanese culture and Sundanese culture (Zuhdi, 1996: 4).

The interaction and communication between Sundanese people and Javanese people firstly in trading encourage the two ethnics to learn each language and then they use both languages as a means of communication in their social lives. Nowadays, the speakers of Javanese language in Cirebon call the language they use as **Cerbon Language** or **Crebon Language**. The interaction between the two languages enables the language contact, language exchange and language absorption happened. From this case, Cerbon Language brings up the problem in geolinguistics: Does it show the language variations? Do the speakers of Cerbon Language produce hybrid and local lexical? Can it include "no different language" as Seguy's band scale (1973).

Figure 1. Cirebon on the Map exists at the borderline

2. Methodology

Cirebon (*Municipality* and *Regency*) is the location of the research (see figure 2). The research uses qualitative and quantitative approach. Village is treated as a research unit. From 446 villages, it is decided 55 villages are treated as the sample of observation center/point (OP) by using criterion-based selection or purposive sample technique. Informers are treated as sources of data. The list of 55 villages of the OP (see figure 3). The numbering system starts from Guwa Lor Village and ends at Tawangsari Village (see Figure 4). Three persons were decided as the informers from each village. They had to fulfil the requirements stated before. Data are collected by using questionnaire which contains list of questions in it, obtained through participative observation method, face-to-face conversation and

recording techniques. Data are analysed by isogloss and dialectometry methods, using triangular interregional and *polygones de Thiessen* techniques. The result of analysis is presented descriptively.

3. Previous literature

The entire previous works on dialectology or geolinguistics in Cirebon analyze the primary topics about dialectology of Sundanese Language, such as Ayatrohaedi (1978), Maulina (2008), Kartika, etal (2007), Mahsun, etal (2008), Ruliah, etal (2008) did.

The only work of dialectology on Cerbon Language is Pemetaan Basa Cerbon: Distribusi Variasi Bahasa dan Penetapan Status (Cerbon Language Mapping: Distribution of Language Variations and Determining Status) by Supriatnoko (*dissertatio*n) in 2015. Up to now, there has not been found the dialectology research about Cerbon Language.

4. Language Variations

Cerbon Language (Javanese Language used in Cirebon) brings out variations both in phonology and in lexical. It happened since it is in a place which far from the land of its birth: Yogyakarta and Surakarta. The native speakers of Javanese Language claimed that Yogyakarta and Surakarta are the center of Javanese Language. Javanese Language used in the two cities recognizes as the standard language for all the speakers. The futher the places from the center the bigger the variations appear, as the weave theory ever stated.

From the other side, it can be said that the propinquity with Sundanese Language makes the use of Javanese Language in Cirebon adopt the sounds and the lexical from Sundanese Language. So, the Sundanese words mix up with the Javanese words in their speech. Thus, speakers of Cerbon Language are difficult to be understood by the other speakers of Javanese Language, in the light of the frequency using grammatical vocabulary tends to be high from other dialects of Javanese Language.

4.1 Phonological Variations

From 558 maps used as instrument to get the data in 55 observation points (OP), there are 349 maps phonological variations, included swadesh basic vocabulary, part of body vocabulary, system of kinship vocabulary, yard and timber vocabulary, animal vocabulary, action and activity vocabulary, function word vocabulary. The fact finding of phonological variations from Sundanese Language and from Javanese Language as follows.

Table 1. The fact finding of phonological variations from Sundanese Language

NO	PREVIOUS VOCABULARY	PHONOLOGICAL VARIATIONS	ENGLISH
1	upah [upah]	upa [upa], upai [upai], pai [pai], pei [pɛi]	give, pay
2	itek [itək]	itek-itek [itək-itək], tetek [tətək], tetekan [tətəkan], teken [təkən]	stick, walking stick
3	tonggar [tɔŋgar]	songgar [sɔŋgar], sanggar [saŋgar]	bucktoothed
4	suuk [suuk]	kacang suuk [kacaŋ sʊʊk], kacang suwuk [kacaŋ sʊwʊk]	peanut
5	sampeu [sampə]	sampu [sampu], campu [campu], capu [capu], kacapu [kacapu]	cassava
6	ngincer [ŋincər]	ngicer [ŋicər], nginyer [ŋiñər], kincer [kincər], kicer [kicər], kiyer [kiyər], nyiyer [ñiyər]	close one eye
7	pegat [pəgat]	pegot [pəgɔt], cegot [cəgɔt]	break off
8	dakom [dakɔm]	ndakom [ndakɔm]	lying flat,
9	sepak [sɛpak]	nyepak [ñɛpak]	kick with the heel or side of the foot

10	sabab [sabab]	sebab [səbab], sebabe [səbabɛ]	because

Table 2. The fact finding of phonological variations from Javanese Language

NO	PREVIOUS VOCABULARY	PHONOLOGICAL VARIATIONS	ENGLISH
1	Mlumah [mlumah]	mluma [mlʊma], mlumo [mlʊmɔ]	lie on
2	nglangi [ŋlaŋi]	ngelangi [ŋəlaŋi], ngilangi [ŋilaŋi], nglangen [ŋlaŋɛn]	swim
3	sira [sira]	siro [sirɔ], ira [ira]	you
4	tiba [tiba]	tibo [tibɔ], tibel [tibɛl], tigel [tigɛl], tegel [tɛgɛl], rigel [rigɛl], regel [rɛgɛl], ribel [ribɛl], rebel [rɛbɛl]	fall
5	dawa [dawa]	dawo[dawɔ], dao [daɔ]	long
6	cemeng [cəmɛŋ]	emeng [əmɛŋ], cemeng [cɛmɛŋ], cemong [cɛmɔŋ], cemeng [cəmɛŋ]	kitten
7	trenggiling [trəŋgiliŋ]	tenggiling [təŋgiliŋ], trenggalung [trəŋgalʊŋ], tenggalung [təŋgalʊŋ]	civet cat
8	piye [piyɛ], pribe [pribɛ], priben [pribɛn]	preben [prɛbɛn], prijen [prijɛn], kepriben [kəpribɛn]	how
9	ngincer [ŋincər]	ngicer [ŋicər], nginyer [ŋiñər], kincer [kincər], kicer [kicər], kiyer [kiyər], nyiyer [ñiyər]	close one eye
10	sendhe [sɛndɛ]	sender [sɛndɛr], sendeh [sɛndɛh], lende [lɛndɛ], lendeh [lɛndɛh], slendeh [slɛndɛh], nyender [ñɛndɛr], nyende [ñɛndɛ]	lean on

4.2 Lexical Variations

From 558 maps used as instrument to get the data in 55 observation points (OP), it has been discovered 209 maps lexical variations, included swadesh basic vocabulary, part of body vocabulary, system of kinship vocabulary, yard and timber vocabulary, animal vocabulary, action and activity vocabulary, function word vocabulary. The fact finding of lexical variations are hybrid and local lexical of Cerbon Language.

4.2.1 Hybrid

From 558 words used as instrument to get the data in 55 observation points (OP), it has been discovered hybrid words. Kridalaksana (2008: 82) said that hybrid is the complex word which each part of word consturcted from different language. For example the hybrid *prasejarah*, constructed from *pra* from Sanskrit and *sejara*h from Arabic. According to Lauder (2007: 25), hybrid is an intermediary language to the speakers used as self expression and as a sign of self identity.

Here are some hybrid found in Cerbon Language.

Table 3. Some Hybird Found in Cerbon Language

NO	HYBRID	CONSTRUCTED FROM		ENGLISH
		SUNDANESE WORD	JAVANESE WORD	
1	kolot lanang [kɔlət lanaŋ]	kolot [kɔlət]	lanang [lanaŋ]	grand father
2	kolot wadon [kɔlət wadɔn]	kolot [kɔlət]	wadon [wadɔn]	grand mather
3	bibi tuwa [bibi tuwa]	bibi [bibi]	tuwa [tuwa]	aunty

4	pete selong [pətɛ seloŋ]	selong [seloŋ]	pete [pətɛ]	tree planted to shade coffee or other plants
5	iwak kancra [iwak kancra]	kancra [kancra]	iwak [iwak]	goldfish
6	manuk dengan [manʊk dɛŋan]	dengan [dɛŋan]	manuk [manʊk]	owl
7	angger pisan [aŋgər pisan]	pisan [pisan]	angger [aŋgər]	often
8	ula cabe [ula cabɛ]	cabe [cabɛ]	ula [ula]	python
9	bonteng timun [bɔntɛŋ timʊn]	timun [timʊn]	bonteng [bɔntɛŋ]	cucumber
10	ula kadut [ula kadʊt]	kadut [kadʊt]	ula [ula]	python

4.2.2 Local Lexical of Cerbon Language

Cerbon Language quotes/takes of words from Sundanese Language as it lives and grows in Sundanese Language environment. Ayatrohaedi (1985: 242) said that lexical or vocabularies quoted from other language, generally related to:

1. Things that have not been ever known by local inhabitant at the previous time.
2. Things that have been ever known before, but the symbol from other language then accepted, either accepted as a substitute symbol which has been available or it has been known collectively.
3. But in the case of Cerbon Language, speakers of Cerbon Language not only quote the words from other language but also create local lexical.

Here are some local lexical created in Cerbon Language.

Table 4. Some Local Lexical of Cerbon Language

NO	LOCAL CERBON LEXICAL	SUNDANESE	JAVANESE	ENGLISH
1	glenter [glentɛr]	goler [gɔlɛr]	mlumah [mlumah]	lie on
2	lagan [lagan]	jeung [jöŋ]	karo [karɔ]	and, with
3	bagel [bagəl]	baledog [balɛdɔg]	mbalang [mbalaŋ]	throw
4	gopekna [gɔpékna]	cekel [cəkəl]	nyekel [ñəkəl], cekel [cəkəl]	take hold of
5	ungker [uŋkər]	cacaka [cacaka]	kepompong [kəpɔmpɔŋ]	cocoon
6	blesak [blesak]	goreng [gɔrɛŋ]	elek [élé'], ala [ala]	ugly
7	duwang [duwaŋ]	talaga [talaga]	tlaga [tlaga]	lake
8	dada walang [dada walaŋ]	iga [iga]	iga [iga]	flank
9	ngotrek [ŋɔtrék]	ninun [ninun]	nenun [nənʊn]	weave
10	beli [bəli], bli [bli]	teu [tö], henteu [həntö]	ora [ɔra]	no

5. Three Group of Cerbon Language Speakers

From 55 observation points (OP) , it has been discovered 3 group of Cerbon Language speakers:

1. The group who pronounces (*diucapkan*) open vocal phoneme /a/ at the final of word by the sound [a]. The group was found in 39 OP, spreaded in Northwest of Cirebon, borders on Majalengka Regency, in South of Cirebon, borders on Kuningan Regency, and in Southeast of Cirebon, borders on Kuningan Regency and Brebes Regency Central Java.
2. The group who pronounces open vocal phoneme /a/ at the final of word by the sound [ɔ]. The group was found in 8 OP, spreaded in Northwest of Cirebon, borders on Majalengka Regency and Indramayu Regency.

3. The group who pronounces open vocal phoneme /a/ at the final of word by the sound [a] and [ɔ], used by turns. The group was found in 8 OP, spreaded in Northwest of Cirebon, borders on Majalengka Regency (see figure 5).
4. Although each group has different way to pronounce open vocal phoneme /a/ at the final of word, the three of group have the same rule in writing phoneme /a/ as (a) not (o).

From 446 villages available in Cirebon it has been discovered 3 group of Cerbon Language speakers:

1. The group who pronounces open vocal phoneme /a/ at the final of word by the sound [a]. The group was found in 335 villages, spreaded in Northwest of Cirebon, borders on Majalengka Regency, in South of Cirebon, borders on Kuningan Regency, and in Southeast of Cirebon.
2. The group who pronounces open vocal phoneme /a/ at the final of word by the sound [ɔ]. The group was found in 64 villages, spreaded in Northwest of Cirebon.
3. The group who pronounces open vocal phoneme /a/ at the final of word by the sound [a] and [ɔ], used by turns. The group was found in 47 villages, spreaded in Northwest of Cirebon (see figure 6).

6. Dialectometry Result

The fact findings shown language variations (phonological variations and lexical variations), hybrid, local lexical, and group of speakers. Based on the dialectometric calculation for all of lexical, Cerbon Language gets the result of 22.42%. So it is considered in the Guiter criteria (1973) that Cerbon Language is ”Different Speech” (*beda wicar*a) from Javanese Language (see Map 1).

Based on the dialectometric calculation for all of phonology, Cerbon Language gets the result of 9.1%. So it is considered in the Guiter criteria (1973) that Cerbon Language is ”Different Subdialect” (*beda subdialek*) from Javanese Language (see Map 2).

7. Conclusion

Cerbon Language as the language in the borderline tends to be difficult to understand by the speakers from other dialects of Javanese Language since it borrows the Sundanese words in their speech. In this case, Sundanese Language is a foreign language which has becomes the biggest contributor for Cerbon Language.

Cerbon Language produces phonological and lexical variations. The feature of phonological variations is about vocabulary sound exchange and open vocal phoneme /a/ at the final of word pronounced as [a], [ɔ], and as [a], [ɔ] by turns. As the result, From 55 observation points (OP), it has been discovered 3 group of Cerbon Language speakers.

Speakers who pronounce /a/ as [a] spreads in 39 OP. All OP villages are adjacent to the villages where the speakers of Sundanese Language live on. So their influence is high contribution especially in the final sound of word which they produce the sound [ɔ] become [a]. For example: *sila* [silɔ] becomes [sila].

The main feature of lexical variations is about hybrid word and local vocabulary which they produce for the same meaning for the words found in both languages: Sundanese and Javanese.

Based on the issoglos and dialectometric calculation for phonology, Cerbon Language is considered to”Different Subdialect” and for lexical is considered to”Different Speech”.

Acknowlegement

I wish to acknowledge my special honor to the Committees of The 4th International Conference on Asian Geolinguistics (ICAG-4) who have given me an opportunity to present my paper in this prestigious event.

References

Adeng, dkk. 1998. *Kota Dagang Cirebon sebagai Bandar Jalur Sutera*. Jakarta: Depdikbud RI.

Alwasilah, A. Chaedar. 1990. *Sosiologi Bahasa*. Bandung: Angkasa.

Alwasilah, A. Chaedar. 2003. *Pokoknya Kualitatif: Dasar-dasar Merancang dan Melakukan Penelitian Kualitatif*. Jakarta: Pustaka Jaya.

Ayatrohaedi. 1985. *Bahasa Sunda di daerah Cirebon*. Indonesian Linguistics Development Project (ILDEP). Ed. W.A.L. Stokhof. Jakarta: Balai Pustaka.

Ayatrohaedi. 2002. *Pedoman Penelitian Dialektologi*. Jakarta: Pusat bahasa Departemen Pendidikan Nasional RI.

Bochari, M. Sanggupri dan Wiwi Kuswiah. 2001. *Sejarah Kerajaan Tradisional Cirebon*. Jakarta: Depdiknas RI.

Bynon, Robert.

1977. *Recherches de Lexicologie*. Paris: Editions Fernand Nathan.

Casta dan Taruna. 2008. *Batik Cirebon: Sebuah Pengantar Apresiasi, Motif, dan Makna Simboliknya*. Cirebon: Badan Komunikasi, Kebudayaan dan Pariwisata Kabupaten Cirebon.

Chambers, J.K. dan Peter Trudgill. 1980. *Dialectology*. Cambridge (Eng.); New York: Cambridge University Press.

Dubois, Jean, dkk. 1973. *Dictionnaire de linguistique*. Paris: Larousse.

Ekadjati, Edi S. 2005[1]. *Kebudayaan Sunda: Suatu Pendekatan SejarahJilid 1*. Jakarta: Pustaka Jaya.

Ekadjati, Edi S. 2005[2]. *Kebudayaan Sunda: Zaman Pajajaran Jilid 2*. Jakarta: Pustaka Jaya.

Fasol, Ralph dan Jeff Connor-Linton. 2006. *An Introduction to Language and Linguistics*. Cambridge: Cambridge University Press.

Goebl, Hans. 1981. "Elements d'Analyse Dialectometrique: Avec application a l'Atlas Linguistiques de l'Italie du Sud", *Revue de Linguistique Romanes*. Volume 45: 349—420. Strasbourg: Societe de Linguistique Romane et Centre National de la Recherche Scientifique.

Guiraud, Pierre. 1978. "Verification de loi Linguistique par Correlation", *Hommage a la Memoire de Pierre Gardette*. Strasbourg: La Societe de Linguistique Romane.

Guiter, Henri. 1973. "Atlas et Frontiere Linguistique". In Les Dialectes Romans de France, 61—109. Paris: Centre Nasional de la Recherche Scientifique.

Hardjasaputra, A. Sobana dan Tawalinuddin Haris (eds). 2011. *Cirebon dalam Lima Zaman (Abad ke-15 hingga Pertengahan Abad ke-20)*. Bandung: Dinas Pariwisata dan Kebudayaan Provinsi Jawa Barat.

Kartika, dkk. 2007. *Penelitian Kekerabatan dan Pemetaan Bahasa-bahasa di Wilayah Jawa Barat dan Banten*. Bandung: Balai Bahasa Bandung.

Kridalaksana, Harimurti. 2008. *Kamus Linguistik*. Jakarta: Gramedia Pustaka Utama.

Lauder, Multamia RMT. 1993. *Pemetaan dan Distribusi Bahasa-bahasa di Tangerang*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.

Lauder, Multamia RMT. 2007. *Sekilas Mengenai Pemetaan Bahasa*. Jakarta: Akbar Media Eka Sarana.

Lembaga Basa dan Sastra Sunda. 1985. *Kamus Umum Basa Sunda*. Bandung: Penerbit Tarate.

Mahsun. 2007. *Metode Penelitian Bahasa: Tahapan Strategi, Metode, dan Tekniknya*. Jakarta: RajaGrafindo Persada.

Mahsun, Inyo Yos Fernandez, Kisyani Laksono, Multamia RMT Lauder, dan Nadra. 2008. *Bahasa dan Peta Bahasa di Indonesia*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.

Maulina, Fajri. 2008. Kajian Dialektologi di Kecamatan Talun Kabupaten Cirebon. *Skripsi*. Jakarta: FBS Universitas Negeri Jakarta.

Meillet, Antoine. 1967. *The Comparative Methods of Historical Linguistics*. Paris: Minuit.

Nauton, Pierre. 1963. *Atlas Linguistique et Ethnographique du Massif-Central*. Volume IV. Paris: Centre National de la Recherche Scientifique.

Pusat Bahasa Departemen Pendidikan Nasional. 2007. *Kamus Besar Bahasa Indonesia*. Edisi Ketiga. Jakarta: Balai Pustaka.

Revier, Xavier. 1975. Le discours sur les isolats Pyereens et la situation linguistique du Pays de Bareges. *HPG*: 420—428.

Sasangka, Sry Satriya Tjatur Wisnu. 2011. *Bunyi-bunyi Distingtif Bahasa Jawa*. Yogyakarta : Elmatera Publishing.

Seguy, Jean. 1971. "La Relation entre la Distance Spatiale et la Distance Lexicale", *Revue de Linguistique Romane*. Paris: societe Linguistique Romane. Volume 35:335—337.

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

Sudaryanto. 1988. *Metode Linguistik. Bagian Kedua*. Yogyakarta: Gadjah Mada University Press.

Sudaryanto. 1993. *Metode dan Teknik Analisis Bahasa*. Yogyakarta: Duta Wacana University Press.

Supriatnoko. 2015. Pemetaan Basa Cerbon: Distribusi Variasi Bahasa dan Penetapan Status. *Dissertation*. Depok: Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia.

Sutopo, H.B. 2006. *Metodologi penelitian Kualitatif: Dasar teori dan terapannya dalam penelitian*. Surakarta: Penerbit Universitas Sebelas Maret.

Trudgill, Peter. 1983. *Sosiolinguistics: An Introduction to Language and Society*. 2nd Edition. Harmondsworth: Penguin Books Ltd.

Zuhdi, Susanto (Penyunting). 1996. *Cirebon sebagai Bandar Jalur Sutra: Kumpulan Makalah Diskusi Ilmiah*. Jakarta: Depdikbud RI.

Figure 2. Municipality and Regency of Cirebon

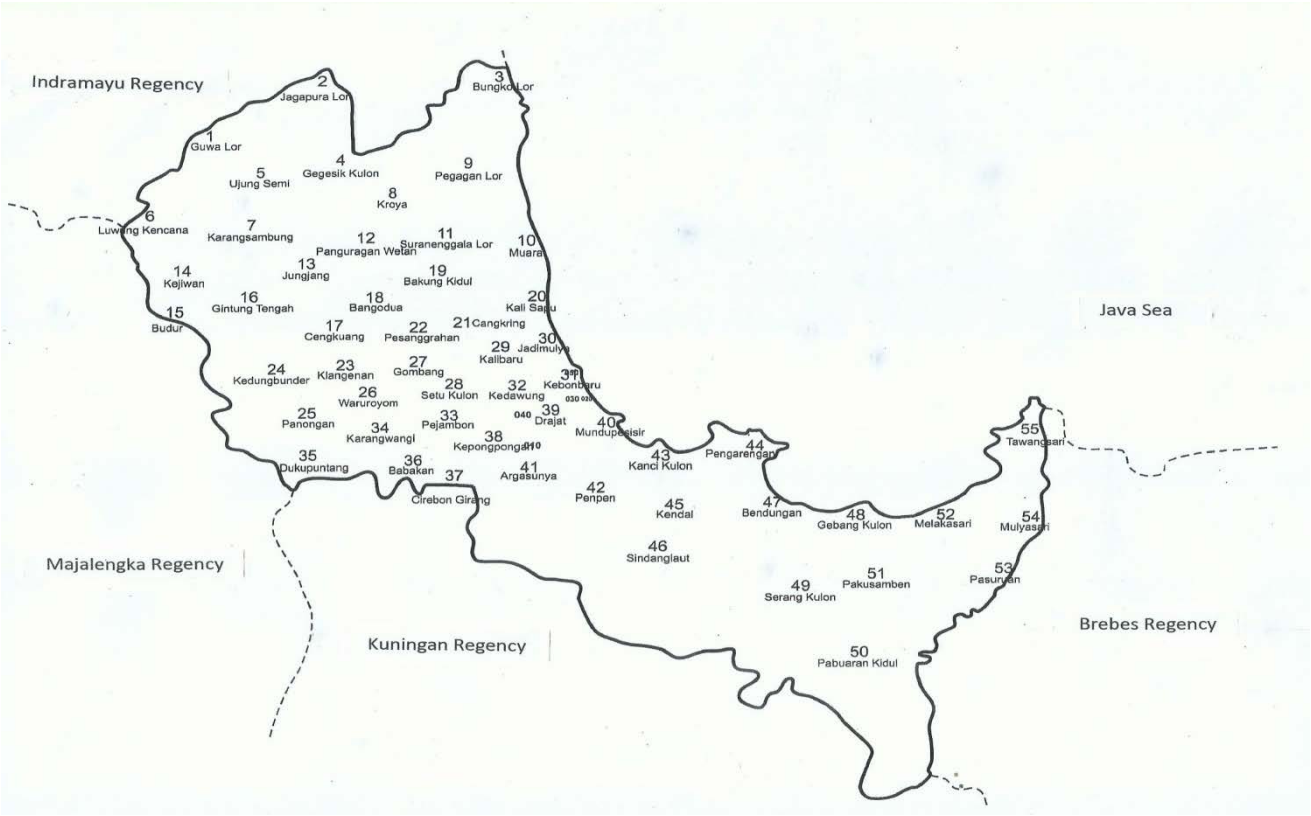


Figure 3. Villages that Become Observation Point

Figure 4. The Numbering System of Observation Point

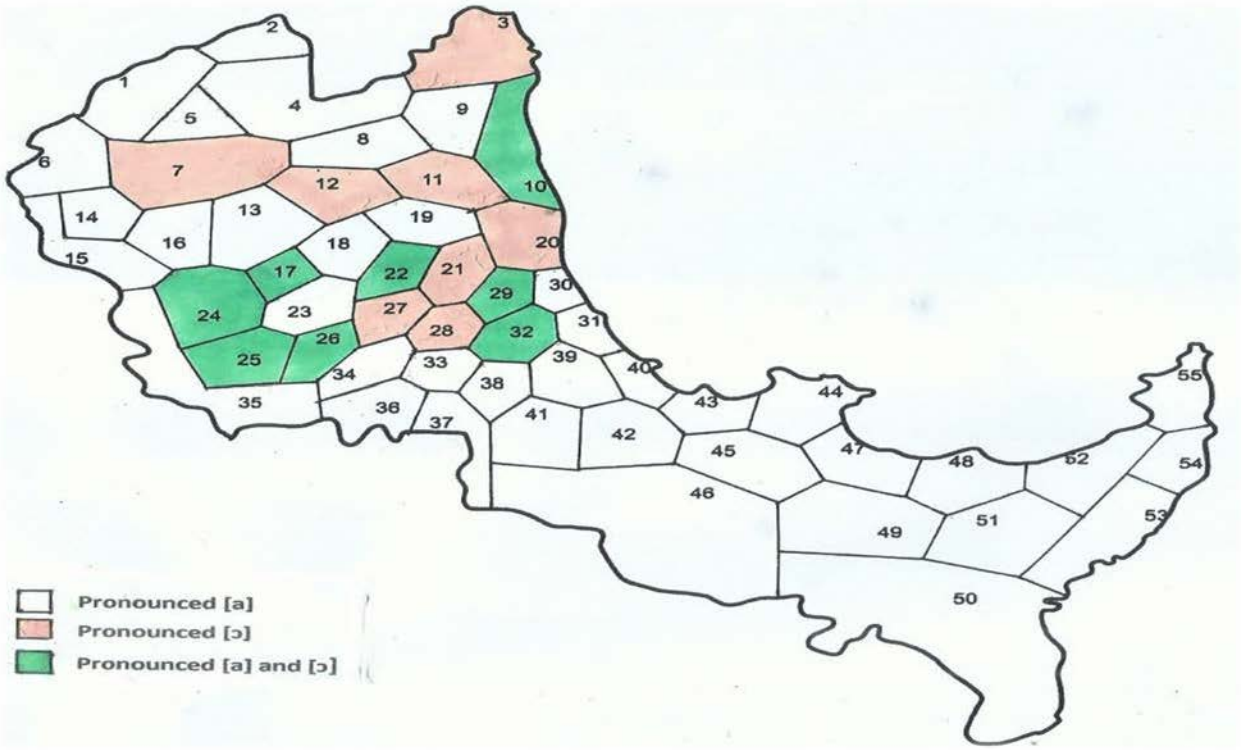


Figure 5. Geography Distribution of Variation of Open Vokal Phonemes /a/ at final from 55 villages

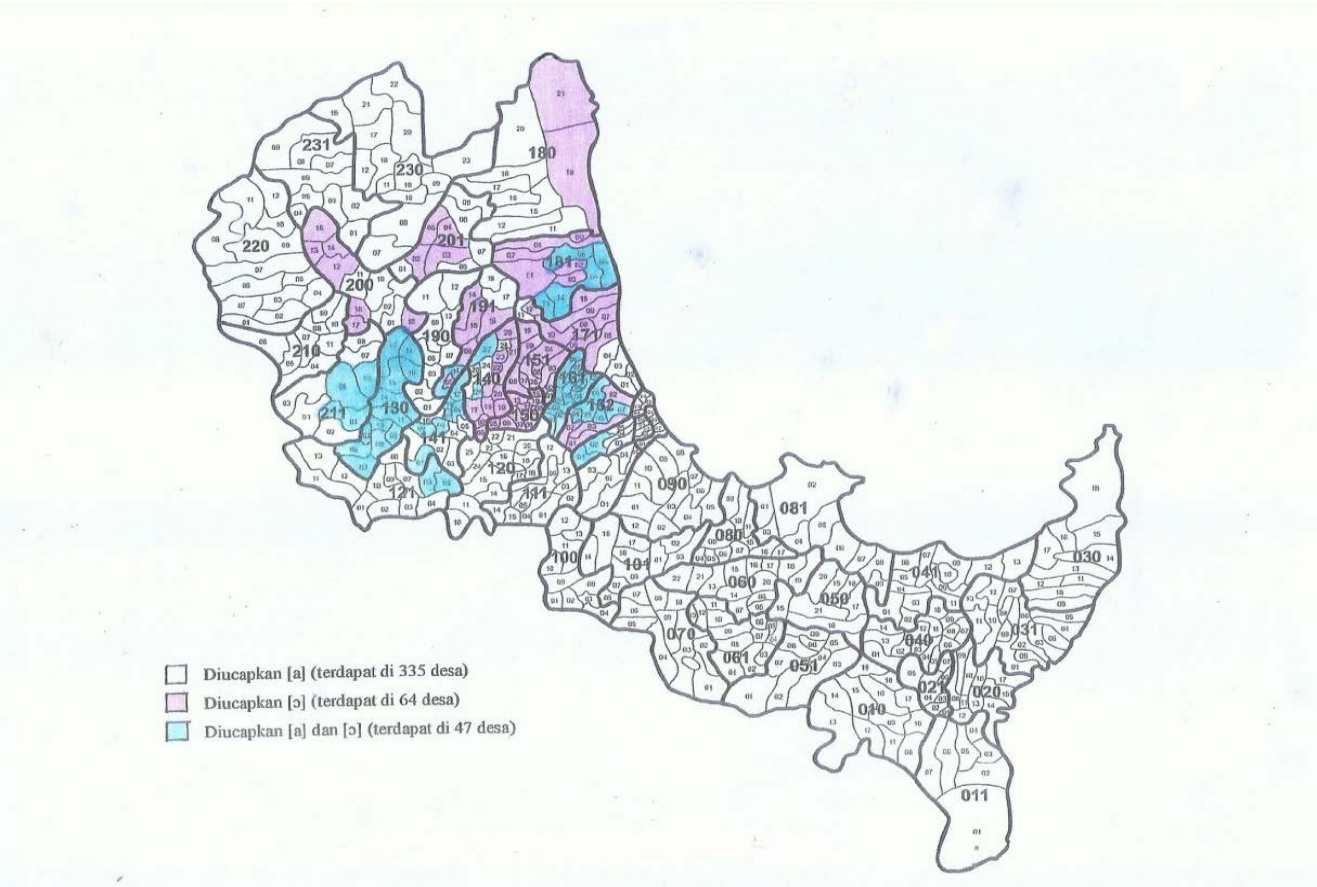
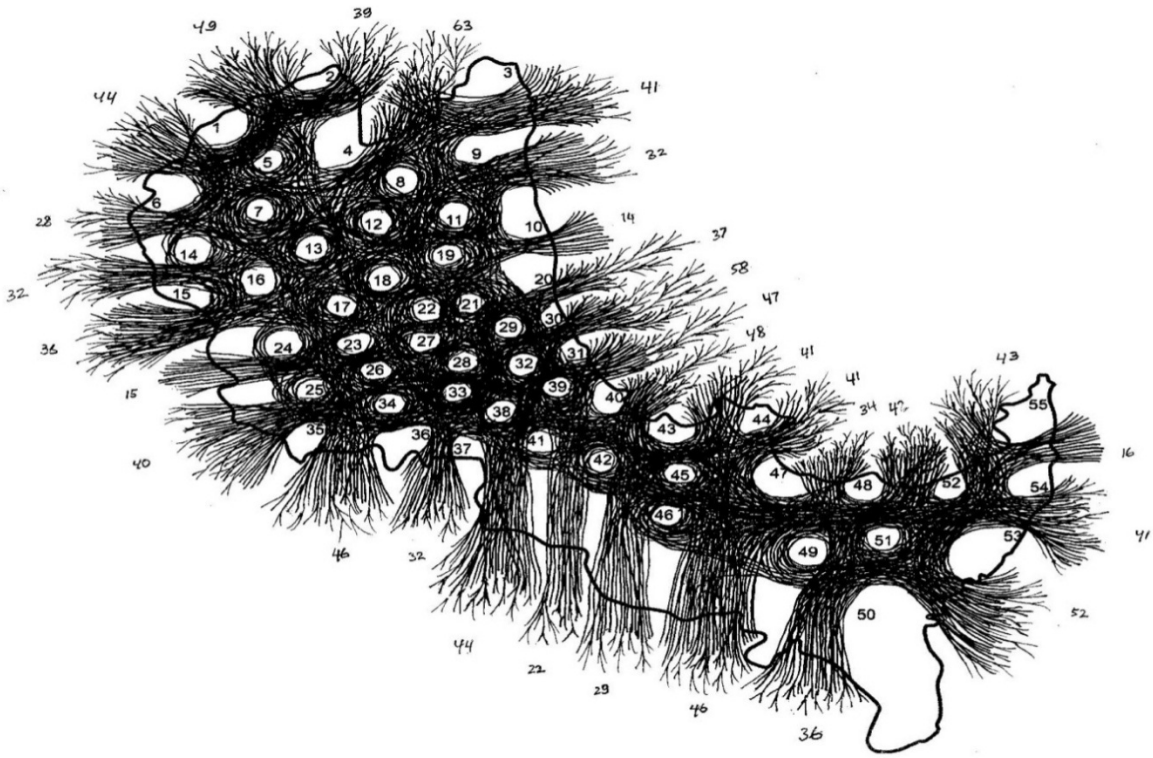
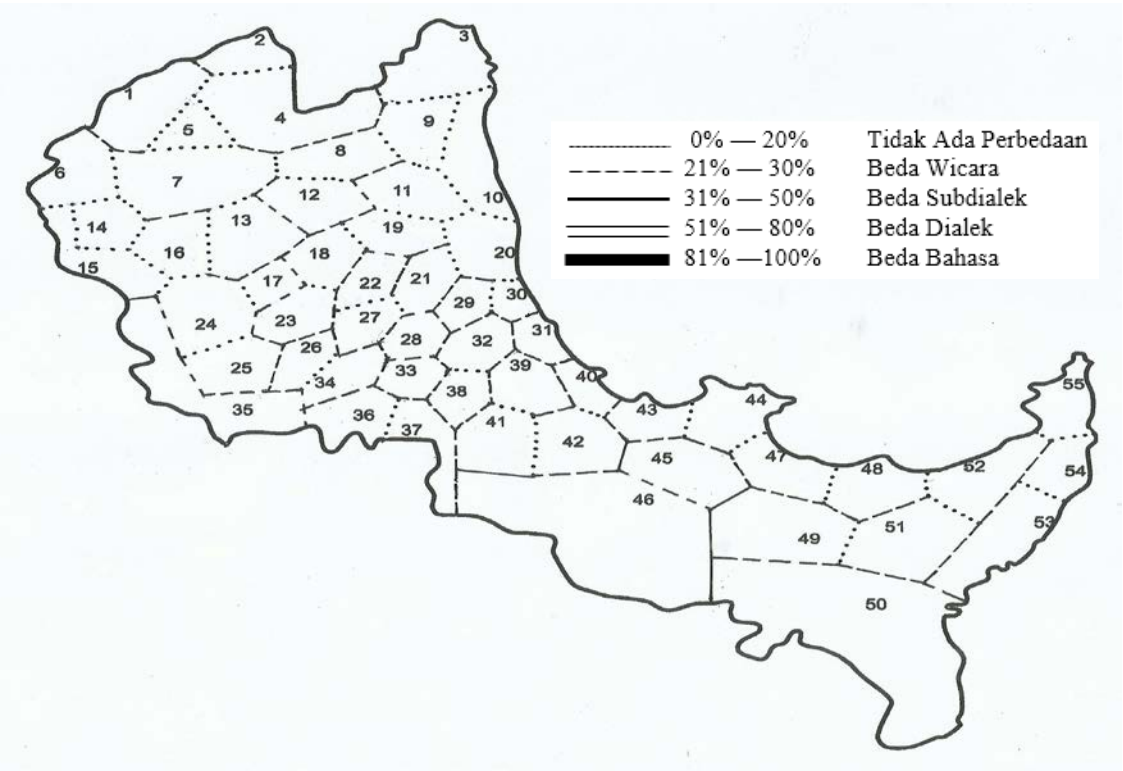


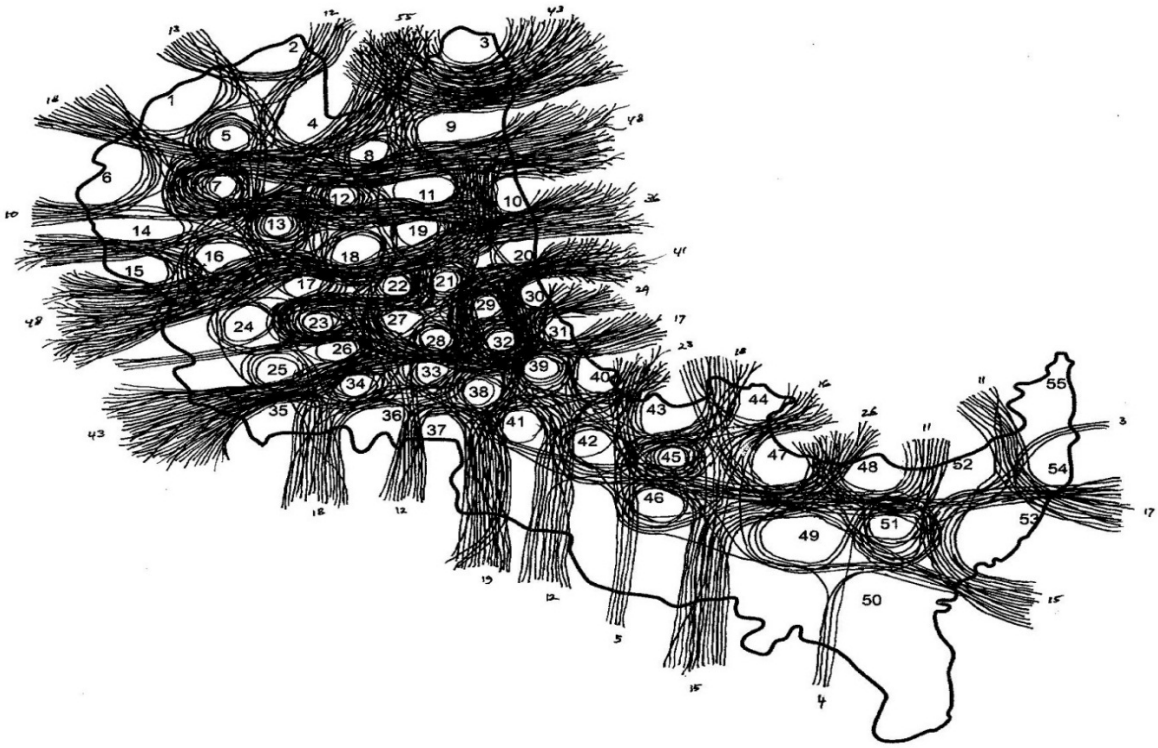
Figure 6. Geography Distribution of Variation of Open Vokal Phonemes /a/ at final from 446 villages

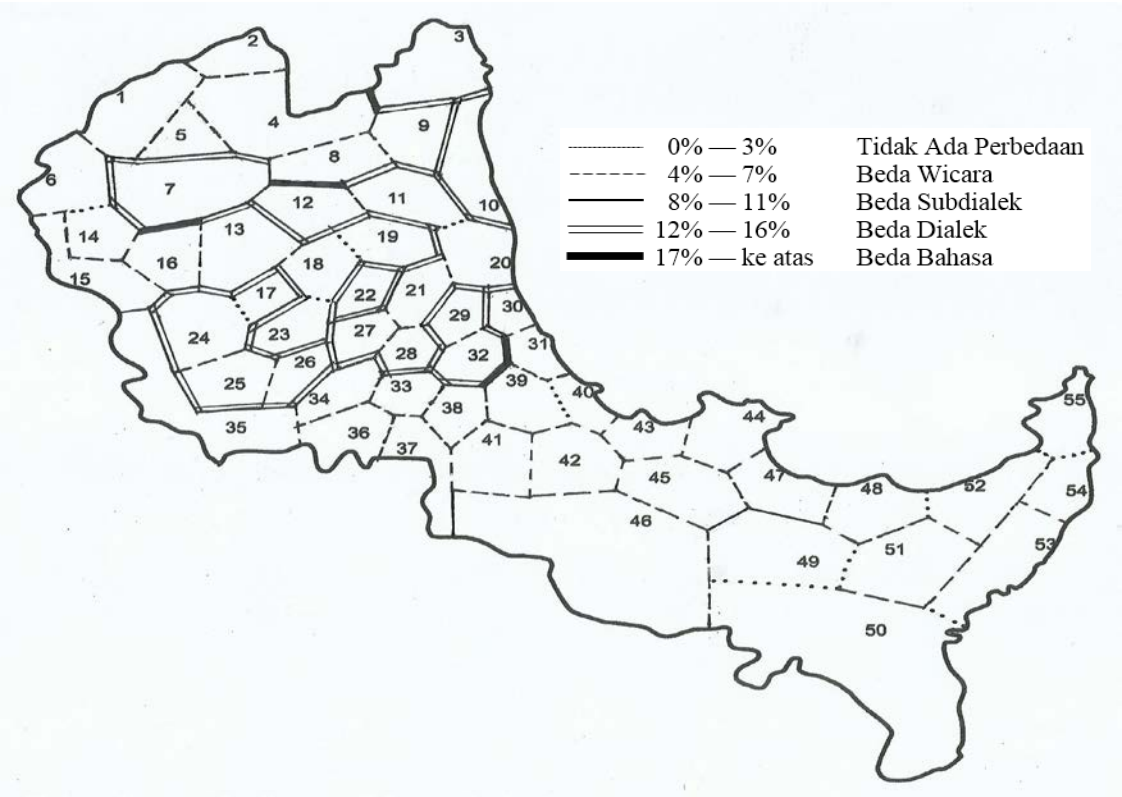
Map 1. Leksikal as a whole





Map 2. Phonology as a whole





Geographical Language Variation in Pringsewu Regency of Lampung Province

Suprayogi
Universitas Teknokrat Indonesia
E-mail: suprayogi@teknokrat.ac.id.

Abstract

Pringsewu is a regency in Lampung where many languages are widely spoken. This dialectology study was aimed at investigating language variation in Pringsewu regency, mapping the variation distribution and classifying the variation into language, dialects, subdialects and speech. There were 350 words consisting of Swadesh list and lexical field of body parts and activities used as the instrument in this study. This study revealed that there were four main languages in Pringsewu namely Lampung, Javanese, Sundanese and Semendo. The lexical varieties can be classified in 14 groups of etyma. Furthermore, the result of dialectometry calculation revealed that there were a lot of areas categorized as "different in dialect", whereas they were actually "different in language". This condition was due to the existence of language contact, and shared features of Proto Austronesian dan Proto West-Malayo Polynesian.

Keywords: dialectology, language variation, language mapping, Pringsewu

1. Introduction

Pringsewu regency in Lampung province, a home for 383,101 residents in 2015, was officially founded in 2009 (BPS Pringsewu, 2015). Pringsewu was once a small village named "Margakaya" which has been inhabitted by people of Lampung Pubian in since 1738. In 1925, people from Java island migrated to this area under Dutch's *ethiesche politiek* of transmigration (Arsip Daerah Kabupaten Pringsewu, 2015). The Javanese people started to make living in this area by converting bamboo forest into housing and farming area. From this, "Pringsewu" was named after the Margakaya's past nature condition where there were a lots of bamboo trees. In Javanese, "Pringsewu" means thousand bamboos.

People from Java and Southern Sumatera migrated to this area voluntarily since the first Javanese transmigration. Some of them were invited by the family member who came earlier convincing that the life in this area could be further better. Some others came to the area as seasonal labors to harvest coffee; in fact, many didn't want to come back, so they decided to live there. As the result, this area, that was just a small village surrounded by forest, has changed to be residential area. These migrants are believed to bring their cultures such as languages and traditions to this new place. However, what languages they speak, where these languages are spoken remain unclear. Therefore, the dialectology study is necessary to investigate the language situation in Pringsewu. Dialectology according to Lauder (2007) is a study of spatial or geographical variation as opposed to sociolinguistics that focuses on the study of social variation. This study attempted to reveal the distribution of lexical variations and the status of the variations, whether the variations are different in language, dialect, sub dialect or speech.

There have been some researches on dialectology investigating language variation in Lampung provinces. There are Lampung language map by *Lembaga Bahasa Nasional* (1972), Lampung dialect division and distribution by Kurniawati (2007) and Javanese variation in Lampung Province by Zawarnis (2009). Meanwhile, the last two researches only selected few observation locations in Pringsewu; thus, they have not represented the whole language situation in the regency.

2. Research Method

This study was conducted in Pringsewu regency of Lampung province. Per 2015, this regency consists of 131 adminisrative villages (BPS Pringsewu, 2015). There were 33 administrative villages or 25% of them selected as observation location (OL). There selection considered the areas with homogenous language, areas established earlier (old villages), areas mostly speaking non-Lampung language, and areas selected for the sake of even distribution. Those villages are Desa Pamenang (OL 23), Desa Candiretno (OL 24), Desa Sukawangi (OL 29), Desa Pujiharjo (OL 25), Desa Fajarbaru (OL 26), Desa Madaraya (OL 32), Desa Margosari (31), Desa Gunungraya (OL 30), Desa Way Kunyir OL 33), Desa Giri Tunggal (28), Desa Margakaya (9), Desa Waluyojati (OL 13), Desa Bumi Arum (OL 12), Desa Selapan (OL 16), Desa Pardasuka Timur (OL 15), Desa Kedaung (OL 16), Desa Pardasuka (OL 14), Desa Kresnomulyo (OL 18), Desa Ambarawa (OL 8), Desa Sinar Baru (OL 19), Desa Pandansari (OL 5), Desa Keputran (OL 10), Desa Siliwangi (OL 20), Desa Srikaton (OL 3), Desa Sinarwaya (OL 11), Desa Totokerto (OL 4), Desa Wonodadi (OL 2), Desa Mataram (OL 6), Desa Wates (OL 7), Desa Gadingrejo Timur (OL 1), Desa Banyuwangi (OL 27), Desa Waya Krui (OL 21), and Desa Sinar Mulya (OL 22).

One informant was selected from each OL based on Chambers and Trudgill's NORMS (nonmobile, older, rural, males) as the priority. Nevertheless, some adjustments of criteria were made, such as it can be female, it has senior high school degree at maximum, and the most important one is the informant fluently speaks the language used by the community where he/she lives.

To collect lexical data, there were 350 words used as the insrument in this study as the primary source of the data. These cover 200 Swadesh words, 52 basic words of body parts and 98 basic words of activities based on *Kuesioner Kosakata Swadesh dan Kata Budaya Dasar* (Questionaire of Swadesh List and Basic Vocabulary) from The Ministry of National Education (2013). Swadesh list contains the vocabulary that is beliveved to exist in every language in the world (Keraf, 1996), meanwhile the basic (cultural) words are believed to separate language and dialect. In addition, the data collection was conducted through direct observation as proposed by Ayatrohaedi (2002) i.e. coming to OL, having direct notetaking, and recording. Further interview was conducted when necessary. There were three methods used to ask a glos: directly asking, describing a word, using pictures or things.

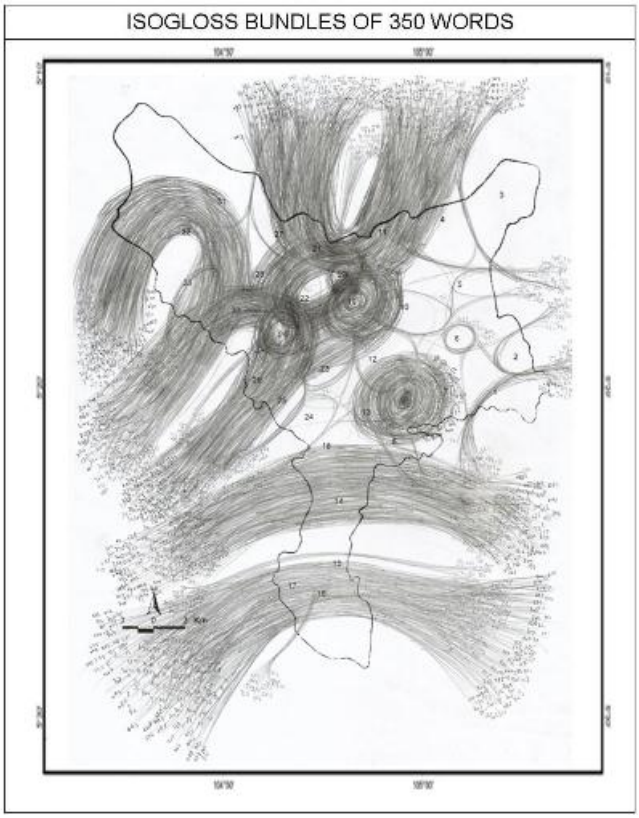
In this study, both qualitative and quantitative methods were employed. The qualitative method was used to describe and interpret lexical data in every OL. Firstly, the 350 maps were drawn, where each map presents etyma's distribution of 1 gloss. Straight line, called isogloss, in the map shows that two areas are lexically different, meanwhile dot line, called isophone, shows that two areas are lexically similar but phonologically different. Secondly, the bundle of isogloss of all words is also made to see how thick the difference in every two areas is. The quantitative method, on the other hand, was used to calculate dialectometry score. Dialectometry is the calculation to determine the degree of differences of two areas as proposed by Seguy using the following formula, where **s** is the number of differences, **n** is the number of total words, and **d** is the percentage of dialectometry.

$$\frac{s \times 100}{n} = d\ \%$$

In this study, the range of percentage from Lauder (1990 in Ayatrohaedi, 2002) was used. It states when d is more than 70%, it means different language, 51-70% means different dialect, 41-50% means different subdialect, 31-40% means different speech, under 30% means no difference.

3. Findings and Discussions

There are two points of discussion in this part. The first is the lexical variations distribution and the second one is the status of the variations. From 350 maps drawn, the lexical data were highly varied indicated by the gloss which has more than 14 etyma. There were two dominant groups of etyma, which are the group of three etyma (69 glosses) and the group of four etyma (59 glosses). The lexical variation distribution can be visually seen through the following isogloss bundles of 350 glosses.



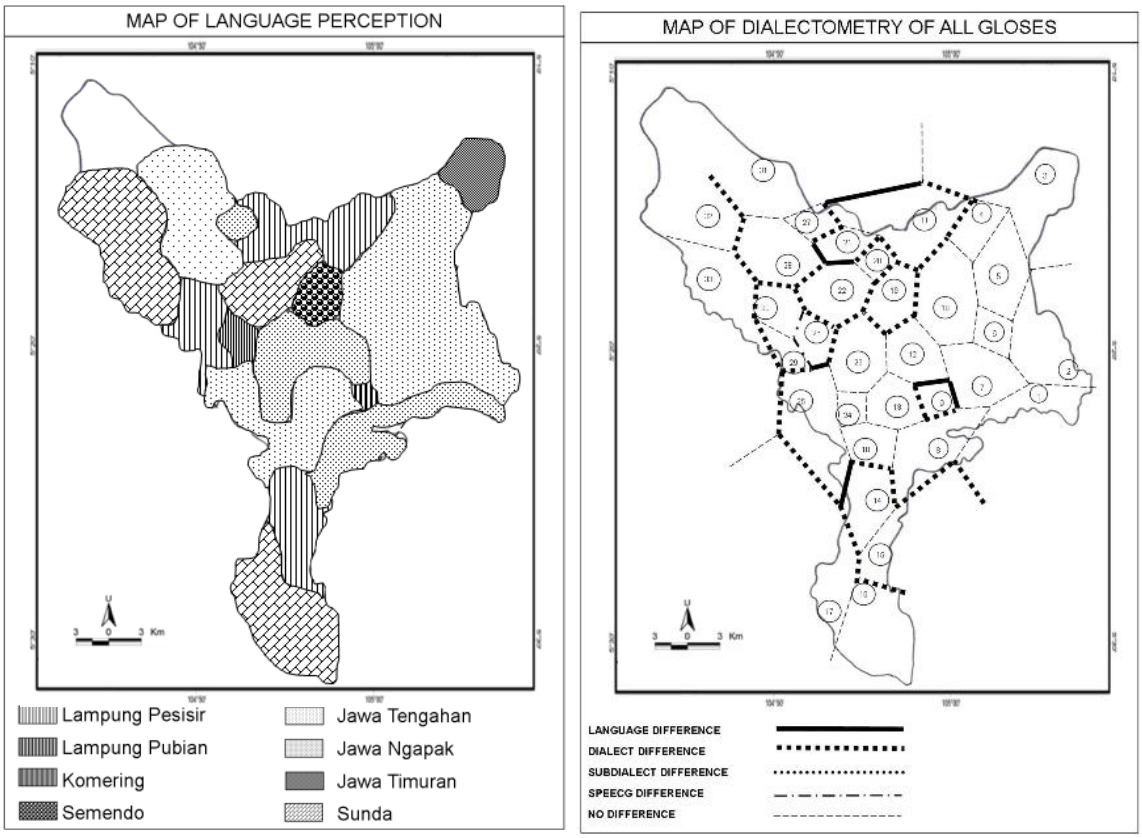
Graph 1. Isogloss Bundles of 350 Words

Graph 1 ilustrates the isoglos thickness in some areas. In the south, there are two groups of issogloses that are thick. The first is between OL 15 and OL 16-17 where OL 15 is the dominant Lampung speaking area, meanwhile OL 16 and OL 17 are Sundanese speaking areas. The second is between OL 14, the Lampung speaking area, and OL 18 and OL 8 where most of the people speak Javanese here. Furthermore, OL 9, as the people there said that this is Lampung Pubian speaking area, is surrounded by thicker group of isoglosses spearating this area and the villages of Javanese speaking communities.

In the northwest, thick isoglosses separate OL 32-33 and their surrounding. These two OL are the areas of Sundanese speaking communities and they are surrounded by Javanese speaking villages. Similar condition occured in OL 29-30, that are the areas of Lampung Pesisir speaking people, and their surroundings mostly speak Javanese. Interestingly, OL 26, the area where people there said that this village uses Komering language, has some issogloses separating this area and OL 29 and this area

and OL 30. It means that there a many similarities between two languages. As in the north, the thick issoglosses circled OL 20-22, Sundanese speaking areas, and OL 19 which is the Semendo speaking area, and it also formed curve to separate OL 11-21, the Lampung speaking areas and other areas. From these thick isoglosses in some directions, it can be seen that Pringsewu regency has high lexical variations. The finding of the two dominant group of glosses shows that approximately there are three or four languages exist in this regency.

The bundle of issoglosses can only reflect the rough differences among observation locations. The differences were made clearer by the result of dialectometry calculation and presenting the result in the map.



Graph 2. The comparison between Language Perception Map and Result of Dialectometry Map

The map on the left shows the people perception of language spoken in Pringsewu. This map was made as the preliminary study of this dialectology research using the framework of perceptual dialectology proposed by Preston (2010). The languages here according to the map are Lampung Pesisir, Lampung Pubian, Semendo, Komering, Central Javanese, Javanese Ngapak, East Javanese, and Sundanese. On the right side is the map of dialectometry calculation result where straight line, bold-dot line, dot line, dot-straight line, and small stripped line show the language difference, dialect difference, subdialect difference, speech difference and no difference respectively. According to the map of dialectometry result, the range of no difference or below/equal to 30% are OL pair of 1-2, 1-7, 1-8, 2-3, 2-5, 2-6, 2-7, 3-4, 3-5, 3-31, 4-5, 4-10, 5-6, 5-10, 6-7, 6-10, 7-8, 7-10, 7-12, 8-13, 8-18, 10-12, 12-13, 12-23, 13-18, 13-23, 13-24, 17-33, 23-24, 23-25, 24-25, 27-28, 27-31, dan 28-31. These OL pairs are the villages of Javanese community including three variaties mentioned in the perception map. It means that, three varieties of Javanese as suggested by the local there actually the same language with little difference in lexicon. What makes them name certain variety as Jawa Timuran is because their ancestor come from a region in East Java, meanwhile those who name Jawa Ngapak is actually the language spoken by the people in that region whose ancestor comes from Kebumen

regency of Central Java and surroundings. Other OL pairs in this category are 11-21, 14-15 and 29-30, Lampung speaking villages, meanwhile the OL pair of 16-17, 20-22, 32-33 are Sundanese speaking villages.

Ideally, the two areas that speak different languages, as the map of language perception has suggested, have at least or more than 70% of dialectometry result according to percentage range proposed by Lauder. However, only 5 pairs of OL have it. Other OL pairs had only difference in dialect, namely OL 18 (Javanese) – OL 14 (Lampung), OL 15 (Lampung) – OL 17 (Sundanese), OL 32 (Sundanese) – OL 28 (Javanese), OL 22 (Sundanese) – OL 26 (Komering), OL 25 (Javanese) – OL 17 (Sundanese), OL 11 (Lampung) – OL 19 (Semendo), OL 21 (Lampung) – OL 27 (Javanese), OL 14 (Lampung) - OL 17 (Sundanese), OL 15 (Lampung) - OL 16 (Sundanese), OL 29 (Lampung) - OL 33 (Sundanese), and OL 30 (Lampung) - OL 33 (Sundanese) .Furthermore OL pairs showing different speech were OL 26 (Komering) and OL 29,30 (Lampung). Here are detail dialectometry for every pair of OL.

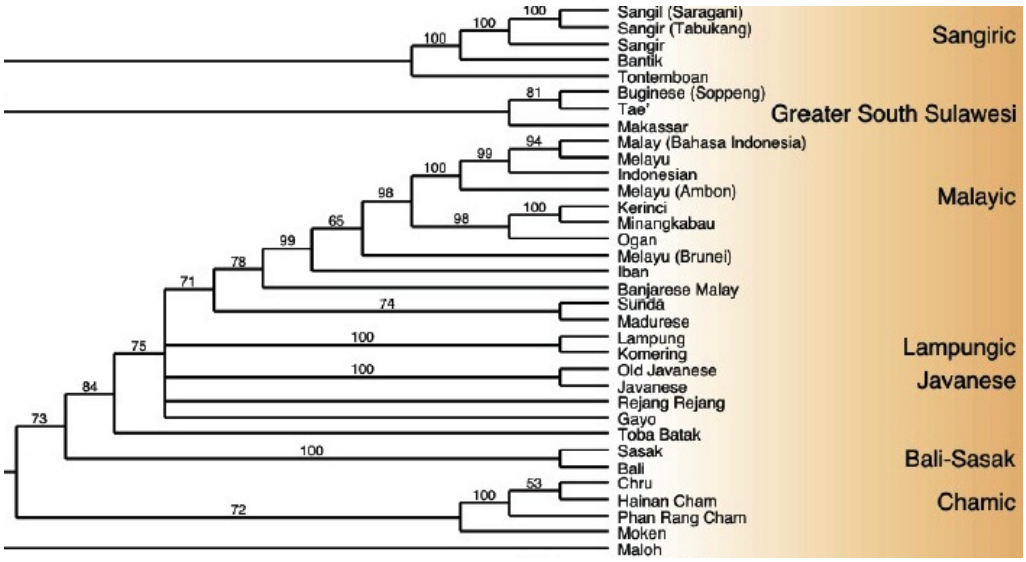
OL Pair	d	%
1—2	52	14,0 %
1—7	55	14,8 %
1—8	41	11,0 %
1—15	260	69,9 %
2—3	43	11,6 %
2—5	39	10,5 %
2—6	53	14,2 %
2—7	32	8,6 %
3—4	25	6,7 %
3—5	45	12,1 %
3—11	256	68,8 %
3—31	39	10,4 %
4—5	44	11,8 %
4—10	43	11,6 %
4—11	258	69,35 %
5—6	66	17,7 %
5—10	54	14,5 %
6—7	53	14,2 %
6—10	63	16,9 %
7—8	45	12,1 %
7—9	263	70,7 %
7—10	39	10,1 %
7—12	40	10,8 %
8—9	258	69,3 %
8—13	48	12,9 %
8—14	246	66,1 %
8—15	251	67,5 %
8—18	37	9,9 %
9—12	268	**72,0 %**

OL Pair	d	%
9—13	259	69,6 %
10—11	256	68,8 %
10—12	52	14 %
10—19	256	68,8 %
11—19	219	58,8 %
11—20	258	69,4 %
11—21	68	18,3 %
11—27	258	68,8 %
11—31	262	**70,4 %**
12—13	41	11,0 %
12—19	253	68,0 %
12—23	42	11,3 %
13—18	28	7,5 %
13—23	46	12,4 %
13—24	38	10,2 %
14—15	53	14,2 %
14—17	239	64,2 %
14---18	239	64,2 %
14—25	262	**70,3 %**
15—16	235	63,2 %
15—17	240	65,5 %
16—17	61	16,4 %
17—25	211	56,7 %
17 –33	65	17,5 %
18—24	34	9,1 %
18—25	30	8,1 %
19—20	235	63,1 %
19—22	226	60,7 %
19—23	247	66,4 %

OL Pair	d	%
20--21	257	69,1 %
20--22	57	15,3 %
21--22	266	**71,5 %**
21--27	256	68,8 %
21--28	256	68,8 %
22--23	212	60,0 %
22--26	239	64,2 %
22--28	223	59,9 %
22--30	244	65,6 %
23--24	37	9,9 %
23--25	50	13,4 %
23--26	259	69,6 %
24--25	39	10,5 %
25--26	258	69,4 %
25--29	260	69,9 %
25--33	210	56,5 %
26--29	111	29,8 %
26--30	121	32,5 %
2--28	35	9,4 %
27--31	36	9,7 %
28--30	242	65,1 %
28--31	28	7,5 %
28--32	221	59,4 %
28--33	203	54,6 %
29--30	41	11,0 %
29--33	236	63,4 %
30--33	239	64,2 %
31--32	219	58,9 %
32--33	47	12,6 %

Table 1. The Dialectometry Result in each OL Pair

. These findings raised some possible justifications. If we see the root of the language, Javanese, Sundanese, Lampung, Komering and Semendo are actually categorized in one language family, which is Proto West Malayo Polynesian (PWMP), a subdivision of Proto Austronesian (PAN). The following graph ilustrates how close the relationship among the language based on Greenhill, Drummond, dan Gray (2010).



Graph 3. Some Parts of Phylogenetic of Proto West Malayo Polynesian (PWMP)

The graph above helps us to understand why some languages have similarities in lexicon. The similar lexicon is derived from the relic words of Proto Austronesian (PAN) or Proto West Malayo Polynesian (PWMP). According to the map, the language pair that have dialect status are mostly Javanese-Lampung, Sundanese-Lampung, Lampung-Komering, and Lampung-Semendo. Here is the example of how Lampung and Semendo has similarities in their lexicon because they shared same PAN/PWMP words. The proto words are based on Austronesian Comparative Dictionary Web Edition by Blust, Robert dan Stephen Trussel (2016). Here are some example based on the findings.

No	Glos	Semendo	Lampung	PAN/ PWMP
1	*FIRE*	[api]	[apuy]	*Sapuy
2	*WET*	[basah]	[basɔh]	*baseq
3	*BLOOD*	[daxah]	[xah]	*dáRaq
4	*DUST*	[dəbu]	[haxəbu]	*ebuk
5	*TWO*	[duə]	[xua]	*duha
6	*NOSE*	[hiduŋ]	[ixuŋ]	*ijuŋ
7	*DRY*	[kəxiŋ]	[kəxiŋ]	*keRiŋ
8	*KNEE*	[əntuat]	[tuwɔt]	*tuhud

Table 1. Semendo – Lampung language relationship

The table shows that the number of similar glosses is a lot. The table does not only shows about language relationship, but also shows that both languages still preserve the archaic words from PAN/PWMP. The similar glosses also occured between Sundanese and Lampungnese, where it happened in a lot of glosses as well. At least 8 glosses reflect the similarities due to the same root of PAN and PWMP.

No	Glos	Sundanese	Lampung	PAN/ PWMP
1	*SMOG*	[hasöp]	[hasɔ?]	*qasep
2	*CORRECT*	[bənər]	[bənɔx]	*bənər
3	*FLOWER*	[kəmbaŋ]	[kumbaŋ]	*buŋaa
5	*DUST*	[dəbu]	[haxəbu]	*ebuk
6	*NOSE*	[iruŋ]	[IxUŋ]	*ijuŋ
7	*GRASS*	[jukut]	[juku?]	*zukut
8	*ULAR*	[ɔray]	[ulay]	*hulaR

Table 2. Sundanese – Lampung language relationship

A question that raised is why, if they don't share Proto Austronesian or Proto West Malayo Polinesian, they still share the similar glosses? The first justification is that it might be possible that they are actually rooted from PAN/PWMP words but the relic words don't exist in the PAN/PWMP dictionary from Blust and Trussel (2016). Second, there might be contact among the language or the speakers in the past which resulted in the shared vocabulary among them. Take a look at Javanese and Sundanese. Both languages are closely related because both languages live in one island i.e. Java island, the origin of migrated Sundanese and Javanese that now live in Lampung province. There was a very long contact between both of the ancestor speakes of both language before the next generation migrated to Lampung. It seems that the close relation occured between Javanese-Sundanese and Komering-Lampung-Semendo. In the context of Komering, Lampung and Semendo, these languages are spoken in Sumatera Island in the area of southern sumatera.

Another possible reason is that they have contact in term of education. According to Thomason (2001: 4), this type of contact is called as learned contact i.e. language contact that is caused by the process of teaching a language to some language groups so that they can communicate one another. In Indonesian context, the role of Indonesian language is very vital to bridge the unintelligibility. Indonesian language is spread through education from primary school to higher education, also from media such as television and newspaper. Due to intense casual contact towards Indonesian language, some Indonesian words are also used in daily conversation subtituting or alternating basic vocabulary.

4. Conclusion

In Lampung context, especially in Pringsewu, the language situation is complex with four major languages spreading in certain blocks and direction where Javanese is dominant and concentrated in the center tothe east ofthe region . Those four languages have their own features and some relationship with other languages surrounding due to the influence of Proto-Austronesian and Proto West-Malayo Polynesian. Therefore, it not surprising that the result of dialectormetry calculation can't really reflect the status of the language. The possible contact, either casual or non-casual contact need to be further investigated to see the clear picture of language situation in this regency.

References

Arsip Daerah Kabupaten Pringsewu. 2015. *Sejarah Pringsewu*. Pringsewu: Kantor Arsip Daerah Kabupaten Pringsewu

Ayatrohaedi. 2002. *Pedoman Penelitian Dialektologi*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.

Blust, Robert dan Stephen Trussel. 2016. *Austronesian Comparative Dictionary Web Edition*. Retrieved on March 8 2016 at www.trussel2.com/ACD

BPS Pringsewu. 2015. *Pringsewu dalam angka 2015*. Retrieved on November 13 2015 at www.pringsewuka.bps.go.id

Chambers, JK dan Peter Trudgill. 2004. *Dialectology Second Edition*. Cambridge: Cambridge University Press.

Greenhill SJ, Drummond AJ, & Gray RD. 2010. *How Accurate and Robust Are the Phylogenetic Estimates of Austronesian Language Relationships? PLoS ONE 5(3): e9573.*

Kementerian Pendidikan Nasional. 2013. *Penelitian Kekerabatan dan Pemetaan Bahasa di Indonesia, Kuesioner Kosakata Dasar dan Kata Budaya Dasar.*

Keraf, Gorys. 1996. *Linguistik Bandingan Historis*. Jakarta: PT. Gramedia.

Kurniawati, Wati. 2007. *Mendeteksi Klasifikasi Dialek Bahasa Lampung*. Unpublished materials: Tesis Universitas Indonesia

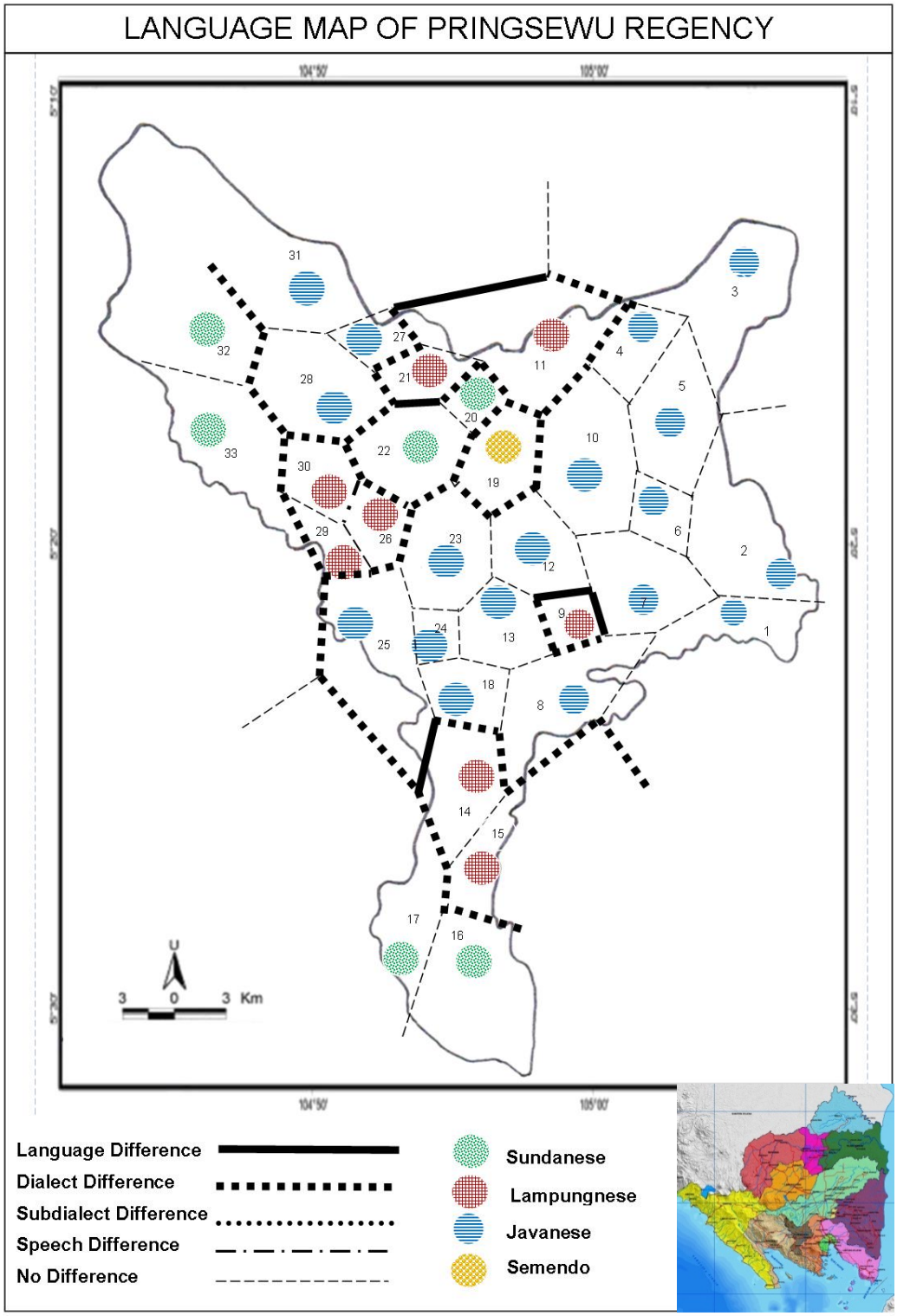
Lauder, Multamia RMT. 2007. *Sekilas Mengenai Pemetaan Bahasa*. Jakarta Timur: Akbar Media Aksara.

Lembaga Bahasa Nasional.1972. *Peta Bahasa-Bahasa Indonesia*. Jakarta: Direktorat Jendral Kebudayaan Departemen Pendidikan dan Kebudayaan.

Preston, Dennis. R. 2010. Language, People, Salience, Space: Perceptual Dialcetology and Language Regard dalam *Dialectologia 5* (2010), 87-131. Retrieved on November 15 2015 at www.raco.cat>article>download

Thomason, Sarah G. 2001.*Language Contact. Eidinburgh*: Eidinburgh University Press

Zawarnis, Yulfi. 2009. *Variasi Dialektal Bahasa Jawa di Lampung*. Unpublished materials: Universitas Indonesia.


LANGUAGE MAP OF PRINGSEWU REGENCY
104°50'
105°00'
31
32
33
28
27
21
11
3
4
5
20
22
10
30
19
29
26
23
6
2
12
9
7
1
25
24
13
18
8
14
15
17
16
U
3 0 3 Km
Language Difference
Dialect Difference
Subdialect Difference
Speech Difference
No Difference
Sundanese
Lampungnese
Javanese
Semendo

The Mandar Language Isolate in Makassar, South Sulawesi

Nadia Almira Sagitta

Badan Pengembangan dan Pembinaan Bahasa

Abstract

Makassar is the center of government, education, and economy in South Sulawesi. This becomes a public appeal for citizens to settle in Makassar. Makassar inhabited by various ethnic and can be proven by the existence of ethnic villages in Makassar such as Toraja village and Mandar village. It affects the language situation in Makassar city. Based on that, this study was conducted to describe language variation in Makassar and explain where the limits of language and dialect in Makassar. The method used in this study are qualitative and quantitative. Based on the dialectometry calculation, two languages are found in Makassar, which is Makassar language and Mandar language.

1 Introduction

Based on data released by the Summer Institute of Linguistics in 2015, Indonesia has 719 languages, of which 707 languages are alive and 12 languages are extinct. However, not all Indonesian languages have been mapped due to the number of researchers. This problem can be solved by conducting dialectological research. Dialectology research conducted in Sulawesi is quite a few only sixteen studies. Four studies of which concentrated in South Sulawesi Province, namely research on Bugis, Toraja and Massenrempulu. Until this research, no dialectology research has been found on Makassar language or dialectology research conducted specifically in Makassar. Therefore, the author is interested to examine the language situation in Makassar.

The linguistic situation in the metropolitan city - in this case the city of Makassar - is interesting to study because an urban area is known as multilingual area. Makassar is bordered by Makassar-speaking area (Gowa) and Bugis (Maros). The presence of various ethnic villages in several subdistricts also affect the language situation in Makassar as an urban city.

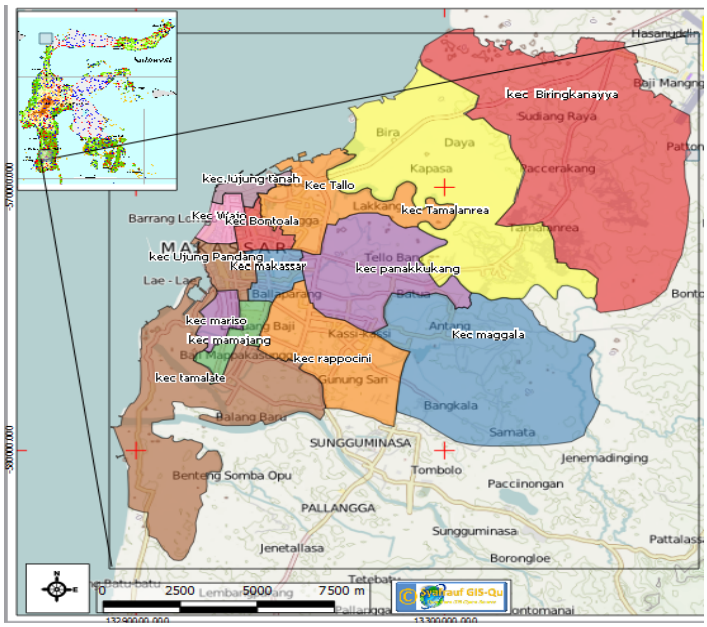
Based on this, the research problem is the variation of any language in Makassar and where the location of dialect and language boundary in the city of Makassar. The purpose of this research is to describe the variation of language found in Makassar and explain the location of dialect and language boundary in Makassar.

2 Section

The method used for this research is a field survey (Ayatrohaedi, 1983: 19). Field survey is a method of collecting data in which researchers visit the field and get information directly from informants (Lauder, 2007: 73). To conduct research in the field, the researchers must decide the observation points, the criteria used to select informants, and the design of the questionnaire which will be used to elicit information. The researchers ask all informants same questions based on the questionnaire. The response is recorded and written down using phonetic notation. The questionnaire was based on the basic word list by Swadesh (200 words), and other vocabulary in a number of semantic fields including colors, greetings, seasons, natural features and objects, and directions. The design and content of the questionnaire is based on Lauder (2007: 138—159).

Observation points are locations where data will be collected. This research gathered information in fourteen points. These are Makassar, Ujung Pandang, Wajo, Bontoala, Tallo, Panakkukang, Mamajang, Mariso, Ujung Tanah, Tamalanrea, Manggala, Rappoccini, Tamalate and Biringkanaya subdistricts. Most of the observation points are located in the cities, thus this can be considered urban.

Fourteen informants were used in the research, one in each data collection point. The criteria for their selection was as follows: (1) Age 40-60 years old; (2) Highest education level achieved is preferably elementary school; (3) The informant is indigenous to the area, being born in and having lived there until this research conducted; (4) Informant's first language is local language.



Makassar Map

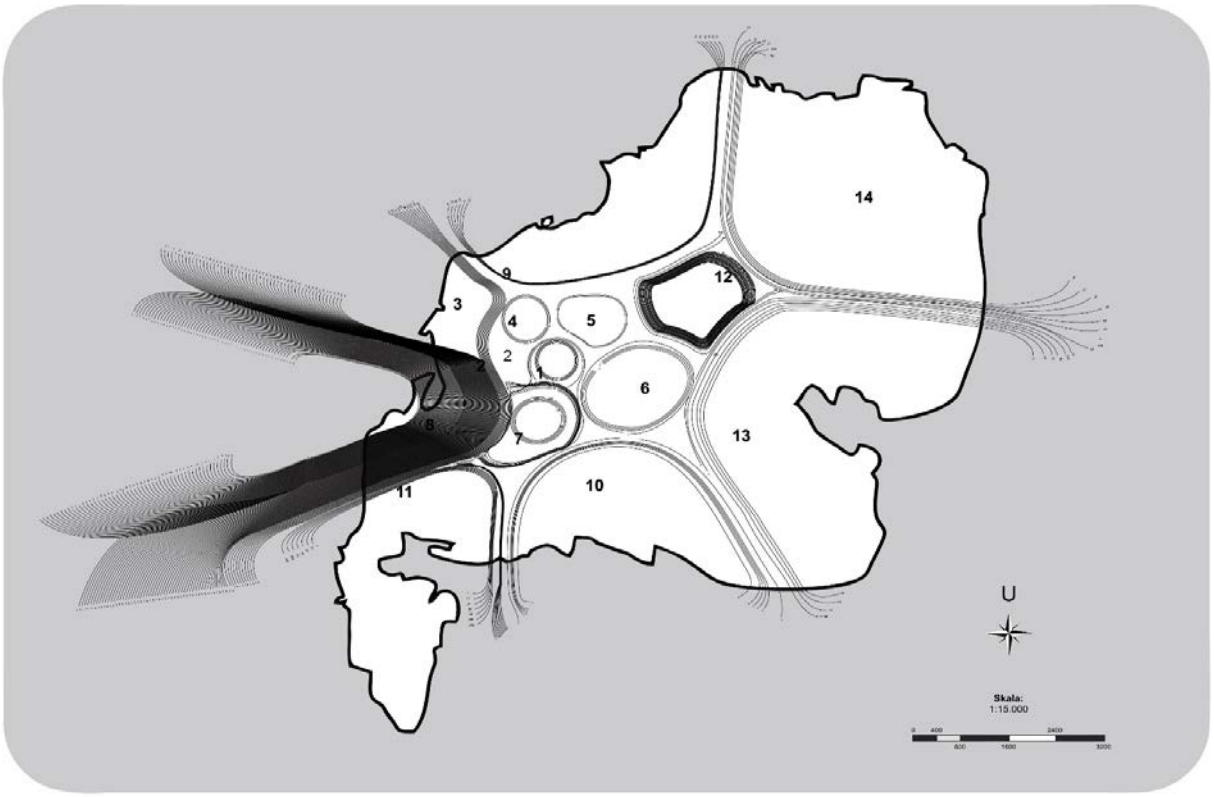
South Sulawesi Map

3 Analysis

The 248 map symbols are grouped into several etyma, one etyme, two etyma, three etyma, four etyma, five etyma, six etyma, seven etyma, and eight etyma. The group of two etyma is the group with the largest number of variants numbering 108 which represents 43.5% of the total. Group of two etyma and two symbols have 73 words, given 66 words from Swadesh. Meanwhile, 21 words belong to two etyma with three symbols group, 9 word belong to two etyma with four symbols group, 4 words belong to two etyma with five symbols group and 1 word belong to two etyma with six symbols group.

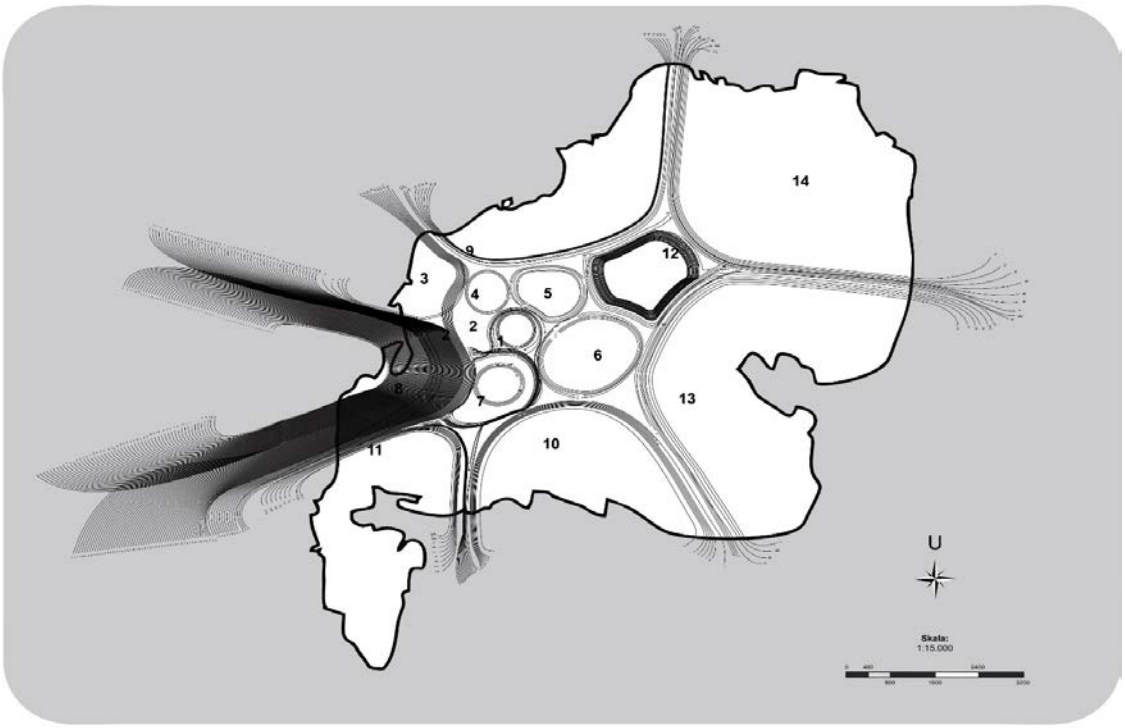
Given the vocabulary that falls into two etyma with two symbols, it can be said that there are two language differences in the research area. Two etyma two symbols group has the greatest number compared to other two etyma groups that has 29 Swadesh basic words, 1 color word, 3 greetings and 9 seasons, natural features, nature objects and directions words. The Swadesh's isogloss map is given below.

.



Swadesh basic words isogloss

Based on the visualization obtained from the Swadesh's isogloss, the buildup line occurred at the observation point 8. The thickening that occurred at point 8, namely Lette village, was caused by a different language area, the Mandar language. Lines elsewhere is not as thick as line in point 8.

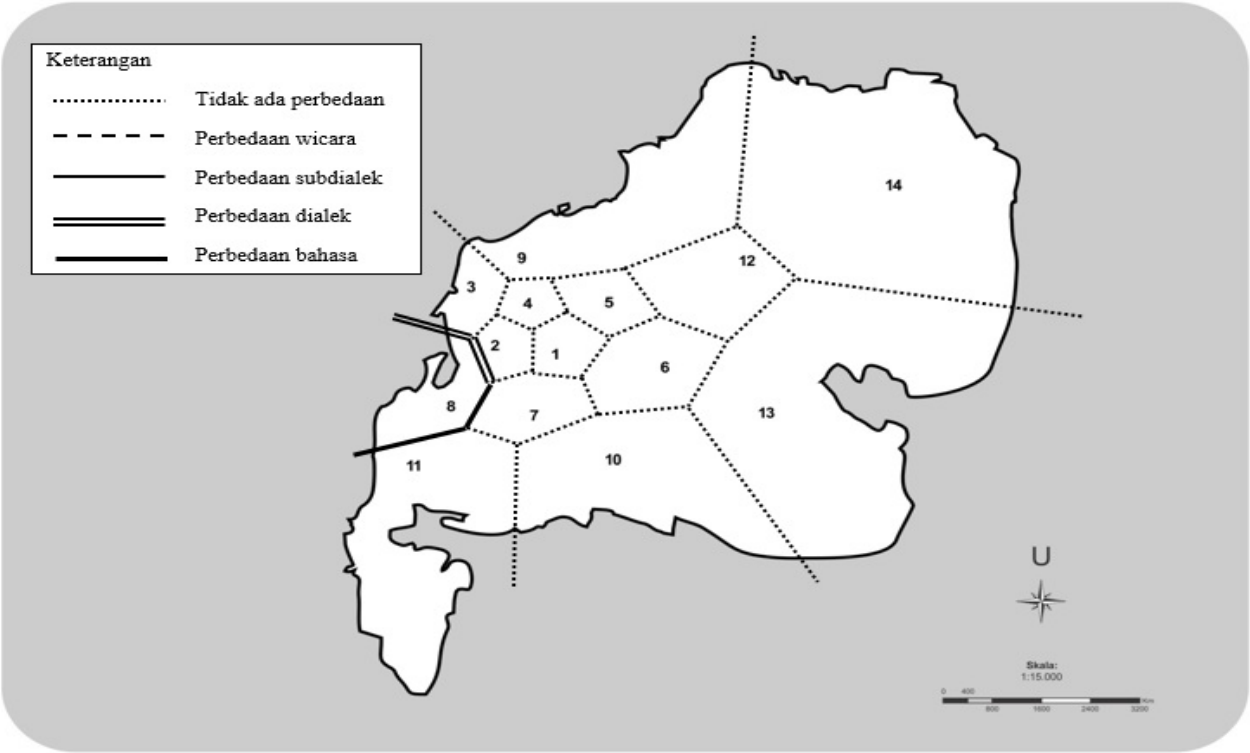


All words isogloss

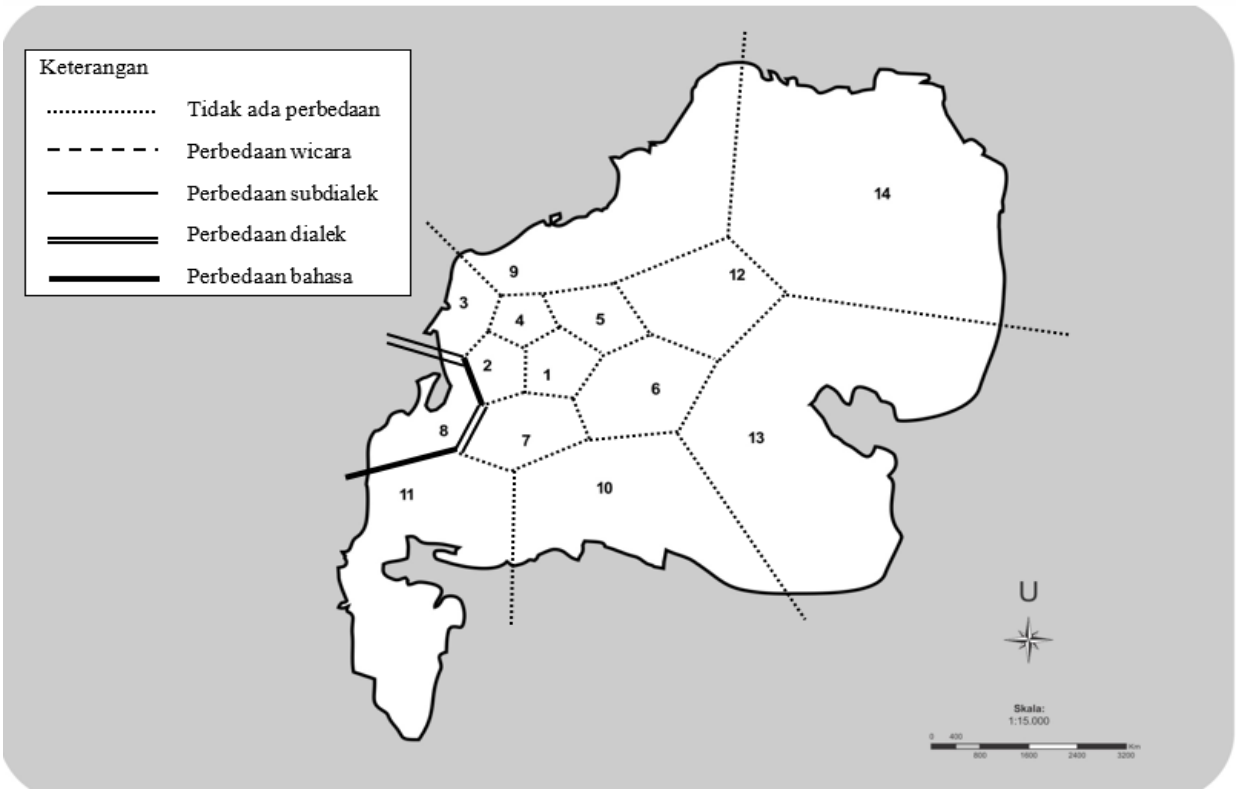
Visualization on compilation isogloss between Swadesh basic words and colours, greetings, seasons, natural objects, and direction vocabularies shows the same result. Thickening line is seen on point 8 (Mariso subdistrict). To get detailed explanation about language, dialect, subdialect, or speech differences, the dialectometry calculation is need to be done.

On Swadesh basic words, the dialectometry calculation between the point 8 (Mariso subdistrict) and 11 (Tamalate subdistrict) is 71% and point 7 (Mamajang subdistrict) and 8 (Mariso subdistrict) by 70.5%. Other observation points that connect with point 8 are point 2 and 3. The result of dialectometry calculation shows the percentage between point 8 (Mariso subdistrict) and 2 (Ujung Pandang subdistrict) is 69. Meanwhile, the dialectometry percentage between point 8 (Mariso subdistrict) and 3 (Wajo subdistrict) is 68%. The latter two belong to different dialect category.

Meanwhile, the dialectometry calculation result of all vocabularies shows results that not much different. The percentage between the point 8 (Mariso subdistrict), point 11 (Tamalate subdistrict) and point 2 (Ujung Pandang subdistrict) is 70.96%. The different percentage between point 8 (Mariso subdistrict) and 3 (Wajo subdistrict) is 68.54%. The percentage between point 7 (Mamajang subdistrict) and 8 (Mariso subdistrict) is 69.75%. The percentage elsewhere is not enough to be categorized as differences in speech, subdialect, dialect, let alone language. The spider web map below is aligned with dialectometry calculation.



Spider web of Swadesh basic words



Spider web of all words

Based on Lauder formulation (1993: 242), percentage above 70% indicates language difference. Point 8 connection with other points shows percentage above 70%. Therefore, point 8 use different language from its surroundings. Actually, there is Makassar language like the other observation points in Mariso subdistrict, but there is an isolate area, which is Lette village that use Mandar in every day communication. To strengthen the evidence of language difference, here is shown several words from two etyma group.

Num.	Gloss	Mandar	Makassar
1	WATER	way	jɛʔnɛʔ
2	FOOT	let:eʔ	baŋkɛŋ
3	FISH	bau	jukuʔ
4	TEETH	riŋŋe	gigi
5	BLACK	malɔtɔŋ	lɛʔlɛŋ
6	SHORT	map:ɔc:i	bɔdɔ
7	SHARP	mataðaŋ	taraŋ
8	TO WALK	mel:amba	aʔjap:a
9	TO LAUGH	mecawa	mak:alaʔ
10	TO STAND	mek:eʔdeʔ	am:enteŋ
11	YOU	iʔo	kau
12	I	iyaw	nak:e
13	NAME	saŋa	arɛŋ
14	NO	andiaŋ	tena
15	MANY	maiʔdiʔ	jayi

4 Difference between Mandar and Makassar Language

In addition to explain the difference between Mandar and Makassar language, this paper also explained a bit about difference from phonological and morphological aspect.

4.1 Sound correspondence between Mandar and Makassar Language

This research found sound correspondence between Mandar and Makassar language. There are four consonant correspondences and two vowel correspondences obtained from this research.

4.1.1 Consonant correspondence

Consonant correspondence identified on the first and second syllable in open and closed syllables.

Consonant Correspondence.	Gloss	Mandar	Makassar	Observation point (Makassar)
[d]~[r]	SLEEP	tindo	tinro	1, 2, 3, 4, 5, 6, 10, 11, 12, 14
	IN HERE	indini	rinni	3, 4, 6, 7, 10
	TWO	daʔduwa	ruwa	1, 2, 3, 4, 6, 7, 9, 10, 11, 13, 14
	COLD (OF WEATHER)	riŋiŋ	diŋiŋ	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14
	YELLOW	mariri	didi	1
[m]~[p]	VILLAGE	kap:uŋ	kampuŋ	2, 5, 6, 7, 9, 10, 11, 13
	FOREST	rop:oŋ	romaŋ	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14
	IF	muaʔ	pun:a	2, 3, 4, 5, 7, 9, 11, 12, 14

[n]~[t]	STAR	bit:ɔeŋ	bintəeŋ	1, 2, 3, 4, 5, 9, 11, 13
	THUNDER	gut:uri	gunturuʔ	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14

The table above shows that there are ten lexicons of Mandar which are quite similar to Makassar. The lexicons can be categorized into three sound correspondence namely [d]~[r] (4 data), [m]~[p] (3 data) and [n]~[t] (2 data). According to van der Tuuk second law, /r/, /d/ and /l/ is interchangeable in Austronesian language. There are five examples that fitted into van der Tuuk second law, e.g. [tindo] and [tinro] that shows [d]-[r] sound correspondence and [riŋiŋ] and [diŋiŋ] that refers to [r]-[d] sound correspondence. All sound correspondences located in same place of articulation. However, it is produced in different manner. When assimilation occurred, [d] merged into [n] sound. The /d/ and /r/ are alveolar sound, /m/ and /p/ are bilabial sound, /n/ and /t/ are also alveolar sound.

4.1.2 Vowel correspondence

Vowel correspondence identified on the first and second syllable in open and closed syllables.

Consonant Correspondence.	Gloss	Mandar	Makassar	Observation point (Makassar)
[i]~[e]	TIGHT	sip:iʔ	sep:aʔ	1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14
	NAME FOR LITTLE GIRL	cic:iʔ	ac:eʔ	3, 4, 9, 10, 12, 14
[a]~[e]	ALL	nasaŋ	ŋaseŋ	1, 2, 3, 4, 5, 7, 9, 10, 12, 13, 14
	KNOW	wis:aŋ	is:eŋ	1, 2, 3, 4, 6, 7, 9, 10, 11, 13, 14
	ROOT	wakeʔ	akaʔ	2, 5, 6, 9, 11, 13
	FOUR	ap:eʔ	ap:aʔ	3, 4, 6, 9, 13

4.1.3 Prefix ma- in Mandar Language

Prefix *ma-* not only to form verbs, but also adjectives. Adjectives in Mandar language can't stand alone without this specific prefix, unlike adjectives in Makassar language that does not need any prefix at all. Those adjective words are given below.

Gloss	Mandar	Makassar
BLACK	malətəŋ	lɛʔlɛŋ
BLUE	magavu	gau, biru, məncəŋ bulə, kəndəʔ- kəndəʔ
WHITE	mapute	kɛbəʔ
RED	mamea	eja, merah
WIDE	maləŋgaŋ	luwaraʔ, laʔbaʔ, ləmpə
LONG	malak:aʔ	laʔbu
SHORT	map:əc:i	bədə
OLD	mabuveŋ	təa
SLIM	manipis	nipisiʔ, nipis, tipisiʔ, tipis, caʔdi, bayaŋ

5 Conclusion

From 248 words—consist of 200 Swadesh basic words, 5 pronouns and greeting words, 7 color words, and 36 nature and direction words, two etyma group is the greatest in number which is 108 words. From both isoglosses, there is thickening line on point 8. That is aligned with spider web map. Dialectometry calculation of all words shows the largest percentage between point 8 and 11 and 2 with 70.96%. Meanwhile, the other points didn't show difference at any level of speech. Based on Lauder formula, the percentage indicates that point 8 use different language. Mandar language which located in Lette village, Mariso subdistrict, is classify as isolate language because majority language in Mariso subdistrict is Makassar language.

Glosses from Mandar and Makassar language shows consonant and vowel correspondence. On consonant correspondence, there are [d]~[r], [m]~[p] and [n]~[t]. The [d]~[r] sound correspondence align with van der Tuuk second law R-D-L. On vowel correspondence, there are [i]~[e] and [a]~[e].

Acknowledgements

The author thank Prof. Multamia Lauder for her assistance to this paper and for believing the author to write this paper. The author also thank Sri Munawarah, M. Hum for inspiring the author to do this research.

References

Ayatrohaedi. (1979). *Dialektologi: Sebuah pengantar*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.

_____. (1997). *Kamus khusus penelitian kekerabatan dan pemetaan bahasa-bahasa daerah di Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

_____. (2002). *Pedoman Penelitian Dialektologi*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.

Chambers, J.K. dan Peter Trudgill. 2004. *Dialectology second edition*. United Kingdom: Cambridge University Press.

Lauder, Multamia MT. (1993. *Pemetaan dan distribusi bahasa-bahasa di Tangerang.* Jakarta: Departemen Pendidikan dan Kebudayaan.

_____. (2007). *Sekilas mengenai pemetaan bahasa*. Jakarta: Media Eka Sarana.

Muthalib, Abdul. (1977). *Kamus Bahasa Mandar-Indonesia.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Pelenkahu, R.A., Abdul Muthalib and M. Zain Sangi. (1983). *Struktur Bahasa Mandar.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Sagitta, Nadia Almira. (2016). "Pemetaan Bahasa Mayoritas di Kota Makassar". Undergraduate thesis. Depok: Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia.

Blust, Robert Andrew. "Austronesian Languages". Britannica. Retrieved from https://www.britannica.com/topic/Austronesian-languages.

Simons, Gary F. & Fennig, Charles D. (Eds.). Retrieved from https://www.ethnologue.com/country/ID.

Contributors

Multamia R.M.T. LAUDER, Universitas Indonesia
Allan F. LAUDER, Universitas Indonesia
Sri MUNAWARAH, Universitas Indonesia
Mesiyarti MUNIR, University of Indonesia
Nor Hashimah JALALUDDIN, Universiti Kebangsaan Malaysia
Junaini KASDAN, Institute of the Malay World and Civilization
Harishon RADZI, Universiti Kebangsaan Malaysia
Youichi NAGATO, Tokyo University of Foreign Studies
Mitsuaki ENDO, Aoyama Gakuin University
Chitsuko FUKUSHIMA, University of Niigata Prefecture
Mika FUKAZAWA, The Preparatory Office for National Ainu Museum
Satoko SHIRAI, JSPS / University of Tsukuba
Hiroyuki SUZUKI, University of Oslo / National Museum of Ethnology
Sonam Wangmo, University of Oslo
Shinsuke KISHIE, Tokushima University
Yukako SAKOGUCHI, Shikoku University
Agusniar Dian SAVITRI, Universitas Negeri Surabaya
Supriatnoko, State Polytechnic of Jakarta
Suprayogi, Universitas Teknokrat Indonesia
Nadia Almira SAGITTA, Badan Pengembangan dan Pembinaan Bahasa, Indonesia

Acknowledgements

This volume is a part of the results of the project on “Studies in Asian Geolinguistics” 2015-2017 at the ILCAA, TUFS. The articles were partially supported by JSPS Grants-in-Aid for Scientific Research (KAKENHI) JP15K02525, JP16J01055, JP16K16820, JP17J40087, JP18H00670, respectively.