Studies in Asian Geolinguistics, Monograph Series No. 7

# Papers from the Workshop "Phylogeny, Dispersion, and Contact of East and Southeast Asian Languages and Human Groups"



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. 7

## Papers from the Workshop "Phylogeny, Dispersion, and Contact of East and Southeast Asian Languages and Human Groups"

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

First published 2019

Edited by Hiroyuki Suzuki, Keita Kurabe & Mitsuaki Endo

Copyright © 2019 Individual Contributors

ISBN 978-4-86337-314-3

This is an electronic publication in PDF format. This publication is offered under Creative Commons International License Attribution–NonCommercial 4.0.



http://creativecommons.org/licenses/by-nc/4.0/

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/ This work is a product of the ILCAA Joint Project Studies in Asian Geolinguistics.

#### **Table of Contents**

Program	i
Mitsuaki Endo, Introduction	. 1
Ray Iwata & Takashi Ueya, Lexical Innovation and Inter-Dialectal Distance	
in Chinese	. 3
Fumiki Suzuki, Characteristics of the Geographical Distribution of Words Denoting	
Cultural Items in Sinitic Languages	14
Kazuo Miyamoto, Prehistoric Cultural Change in East Asia and Language Groups	22
Kenji Yagi, Notes on front rounded vowels in Sinitic languages	28
Hiroyuki Suzuki, How Tibetans classify pigs in their languages in the eastern Tibetosphere:	
Revisiting the pig issue through geolinguistics	40
Ting Ji, Hanzhi Zhang, Thomas E. Currie, Andrew Meade, Mark Pagel & Ruth Mace,	
Applied Phylolinguistics in Cultural Comparative Studies: A Case Study of	
Sino-Tibetan Kinship Evolution	54

The cover collage was made of the figures retrieved from <a href="https://pixabay.com/photos/dna-biology-science-molecule-2259813/">https://pixabay.com/photos/dna-biology-science-molecule-2259813/</a> as well as Satoko Shirai, 'It rains' in Asia, in *Studies in Asian Geolinguistics*, VIII, p. 22, 2018.

#### Program

Pre-conference Workshop of the 27th Annual Conference of International Association of Chinese Linguistics (IACL-27) 国际中国语言学学会第 27 届年会 会前工作坊

Phylogeny, Dispersion, and Contact of East and Southeast Asian Languages and Human Groups 东亚与东南亚语言及人类群体的谱系、扩散和接触

Date: 9th May, 2019 Venue: Seminar Room 4, Kobe Unity

## **10:30-11:45** Chair: Endo, Mitsuaki (Aoyama Gakuin University) 主持人:远藤光晓(青山学院大学)

Endo, Mitsuaki (Aoyama Gakuin University), Introduction 远藤光晓(青山学院大学),前言

Iwata, Ray (Komatsu University), Lexical Innovation and Inter-Dialectal Distance in Chinese 岩田礼(小松大学),汉语的词汇创新与方言间距离

Suzuki, Fumiki (Nanzan University), Characteristics of the geographical distribution of words denoting cultural items in Sinitic languages 鈴木史己 (南山大学), 汉语族文化词地理分布特征

13:00-14:30Chair: Sagart, Laurent (CNRS, France)主持人: 沙加尔(法国国家科学研究中心)

Saitou, Naruya (National Institute of Genetics), In Search of Yaponesian People's Urheimats 斋藤成也(国立遗传学研究所),追寻扶桑人的原乡

Kanzawa-Kiriyama, Hideaki (National Museum of Nature and Science), Tracing movements of ancient East Asian populations using ancient human genome sequences 神泽秀明(国立科学博物馆),基于古代人类基因组跟踪古代东亚群体的迁徙

Miyamoto, Kazuo (Kyushu University), Prehistoric cultural change in East Asia 宫本一夫 (九州大学), 东亚先史文化演变

14:45-16:15 Chair: Iwata, Ray (Komatsu University) 主持人:岩田礼(小松大学)

Sagart, Laurent (CNRS), Origins of tone in Kra-Dai 沙加尔(法国国家科学研究中心), 侗台语声调的起源

Yagi, Kenji (Kokushikan University), Notes on front rounded vowels in Sinitic languages 八木坚二(国士馆大学),汉语族前圆唇元音札记

Suzuki, Hiroyuki (University of Oslo), How Tibetans classify pigs in their languages in the eastern Tibetosphere

鈴木博之 (奥斯陆大学), 东部藏区语言中藏族对"猪"的分类

16:30-18:00 Chair: Saito, Naruya (National Institute of Genetics, Japan) 主持人: 斋藤成也(国立遗传学研究所)

Ji, Ting (Institute of Zoology, CAS), Hanzhi Zhang (University College London), Thomas E. Currie (University of Exeter), Andrew Meade (University of Reading), Mark Pagel (University of Reading/ Santa Fe Institute) & Ruth Mace (University College London), Applied Phylolinguistics in Cultural Comparative Studies – A Case Study of Sino-Tibetan Kinship Evolution

季婷(中科院)、张涵之(伦敦大学学院)、汤玛斯・库里(埃克塞特大学)、安德鲁・米 德(雷丁大学)、马克・佩格尔(雷丁大学)、露丝・梅斯(伦敦大学学院),语言谱系在 文化比较研究中的应用:以汉藏亲系演化为例

Taguchi, Yoshihisa (Chiba University), Hmong-Mien: A Phylogenetic Overview 田口善久(千叶大学), 苗瑶语谱系概观

Hamada, Takeshi (Mie University), Reconstructing the past in a tree: On the challenges and validity of the cladistic approach to the history of Sinitic 滨田武志 (三重大学),过去的树状构拟:进化枝方法对汉语族史的尝试及其有效性

## Introduction

Mitsuaki Endo

Faculty of Economics, Aoyama Gakuin University

At the beginning of this workshop, we pay homage with great reverence and honour to the memory of the late professor Mantaro J. Hashimoto. I also appreciate the attendance of our two distinguished guests, Professor William Wang and Laurent Sagart at our meeting.

Professor Hashimoto was well known as a proponent of typo-geography in the East Asian region. This discipline combines dialect geography and linguistic typology. The idea of this workshop, "Phylogeny, Dispersion, and Contact of East and Southeast Asian Languages and Human Groups" is actually inspired from a section of the former International Congress of Orientalists that originated in Paris in 1873 and is now called the "International Congress of Human Sciences in Asia and North Africa."

When the meeting was held in Japan in 1983, Professor Hashimoto organized a session named "Linguistic and Cultural Ties among East and Southeast Asian People." He presented the keynote speech entitled "Origin of the East Asian Linguistic Structure----Latitudinal Transitions and Longitudinal Developments of East and Southeast Asian Languages" in the presence of linguists such as Benedict, Dyen, Egerod, and others, as well as archaeologists such as Solheim and Nitta, and a paleoclimatologist named Hideo Suzuki.

As widely acknowledged, Professor Hashimoto was a pioneer of multidisciplinary collaborations among linguistics, archaeologists, and geneticists. He planned a joint project with Professor Keiichi Omoto, an eminent geneticist and the supervisor of Professor Naruya Saito who is seated in this audience. However, this project was cancelled because of his untimely death in 1987.

After Professor Hashimoto passed away, Professor Ray Iwata organized the younger generation of scholars at that time and initiated a series of projects on Chinese dialect geography, but now all of us are about to retire or already superannuated.

Professor Iwata has played key role in each of the five 3-year projects. When Professor Hirata became the head, he invited Professor Kazuo Miyamoto who will deliver a presentation on archaeology later. When I served as a representative, Professor Naruya Saito also engaged with our project along with other archaeologist and geographer, and Professor Wang invited him to the first annual meeting of the International Association of Chinese Linguistics (IACL) in Singapore, where Professors Iwata and Lamarre made his acquaintance. In fact, Professor Wang was also Professor Hashimoto's supervisor at Ohio State University. We thus appreciate this profound affinity between the two eminent luminaries of our discipline.

The fruits of these projects were published in the internal progress reports of grant-in-aid JSPS, and finally, Professor Iwata published two volumes of *The Interpretive Maps of Chinese Dialects*. He also persuaded Professor Cao Zhiyun to begin a project on *the Linguistic Atlas of Chinese Dialects*.

Dialect geography has a long tradition in Japan. This discipline began to be practiced in Japan almost as soon as it was initiated in Germany and France. Professor Iwata applied this tradition to the domain of Chinese dialects, and recently, I began expanding the area of research to the whole of Asia with the assistance of many colleagues. Since 2012, the International Conference on Asian Geolinguistics has been held every two years. The first meeting was conducted at Aoyama Gakuin University in Tokyo, and subsequent ones in Thailand, Cambodia, and Indonesia. The proceedings of these conferences may be downloaded from the following websites: <u>https://agsj.jimdo.com/</u> and <u>https://publication.aa-ken.jp/</u>.

Moreover, a joint project at the Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies (ILCAA, TUFS) called Studies in Asian Geolinguistics was undertaken between academic years 2015 and 2017. The results of these investigations are also available as electronic publications at the above mentioned website of the ILCAA.

This project encompassed eight linguistic features, for example the sun, and its geographical distribution in the language families spoken in Asia. The density of these maps accessed about 2000 places in the whole of Asia, the map on tone comprises around 4000 places.

A MEXT project headed by Professor Naruya Saito on the Yaponesian genome is ongoing since academic year 2018. This project is a rather substantial grant-in-aid for 5 years for Scientific Research on Innovative Areas. It is constituted of four subgroups of studies by geneticists, and two subgroups of investigations by archaeologists and linguists. The aim of the project is to ascertain the origins of the Yaponesian (Japanese Archipelago) peoples and languages, mainly in terms of DNA. The Eurasian continental aspect is essential because according to recent theories of genetics, all modern human beings have originated from only one couple in Africa, and humans have expanded to other parts of the world over hundreds of thousands of years. Therefore, the Japanese definitely entered the region through the Eurasian continent. Many of us are linguists interested in Chinese or other languages of East and Southeast Asia. Hence, today's workshop is designed to broadly grasp the interrelationships of these languages and peoples. I hope that this activity widens and intensifies many aspects of this discipline and also grants us the opportunity to enter into closer collaborations with each other and with scholars of different disciplines and countries.

#### Acknowledgements

This workshop was supported by MEXT/JSPS KAKENHI Grant Numbers JP18H00670, JP18H05505, JP18H05510.

## Lexical Innovations in Chinese Basic Words a preliminary statistical measurement

Ray Iwata<sup>*a*</sup>, Takashi Ueya<sup>*b*</sup>

<sup>a</sup>Faculty of Intercultural Communication, Komatsu University

<sup>b</sup>Faculty of Foreign Studies, Kyoto University of Foreign Studies

#### Abstract

This paper is a preliminary version of our attempt to measure the degree of lexical innovation in 42 Chinese dialects. The historical depth of each form or its usage was estimated for 88 selected basic words based mainly on philological evidence, which were then identified as belonging to one of the millennium-based periods A, B, or C. Then, statistical analysis was applied by respectively assigning the numbers "0", "1", "2" to the A, B, and C classifications. The forms with unidentified etymologic forms were assigned D and the number "3." In the end, the results were shown in a figure. While generally the results conformed to common sense knowledge about Chinese dialectology, it also revealed some new evidence.

#### 1 Introduction

This study was originally associated with the research conducted by Professor Naruya Saito at the National Institute of Genetics, the aim of which was to reconstruct the phylogenetic trees of East Asian languages using a neighbor-joining method.

In Chinese linguistics research, there have been some statistical studies (Wang, 1960; Cheng, 1997) that have measured the interdialectal distance between major Chinese (Han) dialects. Wang (1960) used a lexicon from Swadesh's basic words list and Cheng (1997) used the phonological and lexical data collections compiled by Peking University in the late 1950s (Yuan 1964). These studies were focused on whether there was agreement or disagreement with two forms; that is, "+" was given to a lexical entry if the given two dialects shared a specific word form A, and "-" (minus) was given if the form A was used in one dialect while form B was used in the other. However, the problem here is that the statistical sum of these plus-minus specifications does not necessarily reveal the reality of the dialectal divergence as there could be several cases of "agreement" and "disagreement" of the lexical forms.

- Case 1 Agreement: a specific old form of A is shared by all dialects; i.e., shared retention.
- Case 2 Agreement: a specific innovative form of B is shared by all dialects; i.e., shared innovation.
- Case 3 Disagreement: both A and B are old forms, where A is shared by some dialects and B is shared some others.
- Case 4 Disagreement: A is an old form and B is an innovative form, where A is shared by some dialects while B is shared by some others.

Case 5 Disagreement: both A and B are innovative.

It is also necessary to consider the time depth of each form to identify the age of the specific old form and when and how a specific innovative form came into being and became prevalent. Therefore, detailed historical studies supported by philological evidence are required to strictly evaluate the degree of dialectal divergence between Chinese dialects.

As a preliminary to future research, this paper is confined to presenting a methodology and some early results for measuring the degree of lexical change.

#### 2 Lexical entries

Eighty-eight lexical entries are selected, of which 56 (63.6%) appear in Swadesh's 100-word or 200-word lists (Swadesh 1952). In the following list, the two asterisks indicate those that are selected from the 100-word list and the one asterisk indicates those that are selected from the 200-word list, with the number in the square bracket indicating the total number of entries in each word class. The standard Putonghua (PTH) form is appended in parenthesis to each lexical entry.

I. Nouns [52]

1. Natural phenomena [5]

sun (taiyang)\*\*, moon (yueliang)\*\*, star (xingxing)\*\*, cloud (yuncai)\*\*, rainbow (caihong)\*\*

2. Time [4]

morning (zaoshang), evening and night (wanshang, yeli)\*\*, tomorrow (mingtian), yesterday (zuotian)

3. Animals, insects [11]

dog (gou)\*\*, bird (niao)\*\*, hen (muji), egg (jidan)\*\*, snake (she)\*, louse (shizi)\*\*, flea (gezao), ant (mayi), fly (cangying), mosquito (wenzi), wing (chibang)\*

4. Plants [2]

rice plant (daozi), soybean (dadou)

5. Body parts [18]

face (lian), mouse (zui)\*\*, ear (erduo) \*\*, nose (bizi)\*\*, eye (yanjing)\*\*, tongue (shetou)\*\*, tooth (yachi)\*\*, saliva (tuomo), neck (bozi)\*\*, throat (houlong), breast (rufang)\*\*, arm (gebo), armpit (ye), belly (duzi)\*\*, navel (duqi), leg (tui), knee (xigai)\*\*, finger (zhijia)

6. Kinship [7]

father (baba)\*, mother (mama)\*, son (erzi), paternal grandfather (yeye), paternal grandmother (nainai), maternal grandfather (laoye), maternal grandmother (laolao)

7. Others [5]

thing (dongxi), rope (shengzi)\*, house (fangzi), kitchen (chufang), chopsticks (kuaizi)

II Pronoun [10]

I (wo)\*\*, you (ni)\*\*, he/she (ta)\*, they (tamen)\*, this (zhe)\*\*, that (na)\*\*, here (zheli)\*, what (shenme)\*\*, who (shei)\*\*, where (nali)\*

III Verb [13]

rain falls (xiayu), wind blows (guafeng) (\*), sleep (shui)\*\*, lie down (tang)\*\*, drink (he)\*\*, *distinction between eat and drink (chi/he)*, know (zhidao)\*\*, see (kanjian)\*\*, walk (zou)\*\*, run (pao), say (shuo)\*\*, stand (zhan)\*\*, smell (wen)\*

IV Adjective & Adverb [13]

red (hong)\*\*, black (hei)\*\*, many (duo)\*\*, small (xiao)\*\*, broad (kuan)\*, thick (hou)\*, thin (bao)\*, cold (leng)\*\*, hot (re)\*\*, fat (animal) (fei), *distinction between fat (animal) and fat(human)* (fei/pang), hungry (e), not (bu)\*\*

As can be seen, the noun percentage is higher than the rest at 59% of all entries. This is because more than 20 noun class entries are adopted that are not listed in Swadesh's list, some of which are possibly characteristic of Chinese culture (e.g., kinships and chopsticks), some of which are related with plant cultivation and housing (e.g., rice plant, soybean, house, kitchen), and all of which have been in common use in China for more than 2,000 years. Meanwhile, many entries and especially verbs and adjectives that were included in Swadesh's list are eliminated as they generally exhibit no regional word form differences; e.g., "fly", "kill", "come", "sit", "white", "new", "water", "mountain".

Swadesh's list cites two nouns "rain" and "wind," both of which have the least dialectal Chinese varieties, with the dialectal differences appearing in verb phrases such as "rain falls" and "wind blows;" therefore, these are included in the verb class.

As a tentative attempt, we include two entries which relate with the concept distinction. In the above list, these two entries are indicated in italics.

- (1) "eat" and "drink"
- (2) "fat" (animal) and "fat" (human)

For both entries, some dialects use an identical form for the two concepts while others differentiate the forms as in standard PTH.

Besides these two cases, there are also lexical entries that are related with the concept distinctions. For example, in many dialects, the concepts "sleep" and "lie down" are not distinguished in terms of word form. However, in this paper, these were treated as two independent entries. Further, the distinctions for the two concepts "evening" (early night) and "night" (late night), i.e., *wanshang* and *yeli* in PTH, could be because of recent differentiation in the semantic categories. As this type of distinction has not been widely spread, the two concepts are treated as one entry.

#### **3** Target dialects and sources

The lexical analysis sources are the 42 Dialect Dictionaries edited by Rong Li (李荣) published by Jiangsu Education Press (江苏教育出版社) in the 1990s. The list of these 42 Chinese dialects is given here. They are grouped based on the common dialect classifications used by Chinese scholars (cf. Chinese Academy of Social Sciences 2012).

- 1. Guanhua (Mandarin) group: Haerbin, Wulumuqi, Xi'ning, Yinchuan, Ji'nan, Muping, Xuzhou, Luoyang, Wanrong, Xi'an, Wuhan, Chengdu, Guiyang, Liuzhou, Yangzhou, Nanjing
- 2. Jin group: Taiyuan, Xinzhou
- 3. Wu group: Suzhou, Shanghai, Chongming, Hangzhou, Ningbo, Jinhua, Wenzhou, Danyang\*
- 4. Hui group: Jixi
- 5. Gan group: Nanchang, Lichuan, Yudu, Pingxiang\*\*
- 6. Xiang group: Changsha, Loudi
- 7. Hakka group: Meixian
- 8. Yue group: Guangzhou, Dongguan

9. Min group: Xiamen, Fuzhou, Jian'ou, Haikang (Leizhou), Haikou

10. Pinghua: Nanning

Notes:

\*Dangyang is a transient dialect on the border of Guanhua and Wu.

\*\* Pingxiang is a transient dialect on the border of Xiang and Gan.

In a few cases, as the dialect dictionary lacks the particular lexical entries being sought, other dialectal sources are consulted. However, two gaps remain unfilled: "to smell" in the Muping dialect and "navel" in the Haikang dialect. However, these gaps will not influence the statistical results.

#### 4 **Principles of the analysis**

With the focus on lexical change, the historical depth of each form or usage is estimated based on the following principles.

#### 4.1 Estimation of historical depth of lexical forms

The historical depth of each form or its usage is estimated based on historical written records. For the entries included in Swadesh's 100 basic words, a recent study by Wang (2018) is consulted because of its exhaustive philological evidence. For other entries, the Scripta Sinica Database (汉籍电子资料库) established by Academia Sinica, Taiwan, is consulted. Thus, each form or each usage in contemporary dialects is identified as belonging to one of four stages.

A: Old Chinese (OC): one millennium before the Common Era (BCE1,000  $\sim$  BCE 1)

B: Middle Chinese (MC): one millennium after CE (CE  $1 \sim$  CE 1,000)

C: Modern Chinese<sup>1</sup>: second millennium after CE (CE 1,000~present)

D: Other: mainly other forms with unknown etymologies

Although this periodization does not strictly conform to the widely shared view among sinologists (Wang, 1958: 35), the absolute time unit "millennium" may be useful in forthcoming phylogenetic studies that compare Chinese with other East Asian languages.

The following subsections examine the principles behind the judgements made on the historical depth for each lexical form or its usage.

#### 4.2. Treatment of plural forms attested in the same stage

If two forms are found to be from the OC period (stage A), the earlier or more prevalent form is referred to as *leading form* ("zhudaoci" in Wang, 2018), and labelled A1, with the other being labelled A2. If the plural forms are found to be from the MC (stages B) or modern Chinese (stage C) periods, branch numbers are added to indicate the inherited relationships. For example, B1 is the descendant of A1 and C2 is the descendant of B2. Two examples are given here. Here and in the following, the word forms are presented in Pinyin with the kanji noted in the parentheses.

#### Examples

Words for "know"

(1) As the two forms zhi (知) and xiao (晓) coexisted in OC; they are labelled A1 and A2.

<sup>&</sup>lt;sup>1</sup> The term "Modern Chinese" is a translation from the Chinese *Jindai Hanyu* (近代汉语). It is distinguished from *Xiandai Hanyu* (现代汉语), which in this paper is referred to as "Contemporary Chinese."

(2) As the present PTH form *zhidao* (知道) first appeared in MC, it is labelled B1. Another influential form *xiaode* (晓得) is distributed both in southern China and northwest China (Cao ed. 2008, Lexicon Volume, Map 154; Iwata ed. 2012, Map 42). From evidence from the distribution area and the inherited relationship from the OC *xiao*, this form *xiaode* is labelled B2, although it became prevalent in the early part of stage C (South Song Dynasty).

Words for "neck"

- (1) As in OC (stage A) the two forms *ling* (领) and *dou* (逗) were used, these are labelled A1 and A2. While the *ling* form has not been preserved in contemporary dialects as denoting "neck," it is preserved in compounds denoting related semantic categories (e.g., such as *lingzi* "the neck of clothes"). Meanwhile, the *dou* form is still in use in some dialects as denoting "neck."
- (2) The form *jing* (颈) appeared in OC; however, according to Wang (2018), it became most dominant in MC (stage B). Therefore, from this evidence and the possibility of a cognate relationship with the OC *ling* form, this is labelled B1. In MC, as another form *xiang* (项) appeared and became prevalent in some areas it is labelled B2. Note here that this form is not a descendant of A2, i.e., *dou*. In contemporary dialects this *xiang* form is seldom used independently and behaves like a bound morpheme; e.g., *boxiang* (脖项) (Wang 2018:195). Therefore, for this compound, a further principle is given in Section 4.4.2.
  - (3) The present standard form *bo* (脖) spread over the northern area in the Modern period (stage C), so this form is labelled C.

#### 4.3 Treatment of philologically unattested forms

There are two cases that are not attested in the consulted historical literature, the first of which is a monosyllabic form for which the etymology is unknown. These forms are labelled D, several examples of which are given in the following.

"drink": [lim] (啉) used in Xiamen; [iɛ] (餲) used in Jian'ou

"lie down" and "sleep": [fen](瞓) used in Guangzhou and Dongguan

"yesterday": 1st syllable of [khem iet](琴日) used in Guangzhou

The second case are polysyllabic compounds that are not covered in the consulted literature. One of the solutions for this case is to estimate the relative chronologies for the specific forms based on lexical maps (Cao ed. 2008, Iwata ed. 2009, 2012), after which their absolute chronologies are estimated in reference to the forms confirmed in the consulted literature.

#### Example

#### Words for "fly"

While the monosyllabic form *ying* (蝇) must be the oldest (A1), a bi-syllabic form *cangying* (苍蝇) is also found in the *Shijing* Poetry from the OC period (A2).

In contemporary southern dialects, the two influential forms wuying(乌蝇) and huying(胡蝇) have relatively wider distribution area than others, but are not found in historical written texts. As they are generally distributed adjacent to *cangying* (cf. Iwata ed. 2012 Map 31-3), it is assumed that the non-head *cang* in *cangying* may have been replaced by *wu* or *hu*. Here a change from *cang* "dark blue" to *wu* "black" may have been a synonymic substitution, and the change from *wu* to *hu* or vice versa may have been motivated by phonetic similarity (the semantics of *hu* are unclear, and may be related to "beard" or "the northern barbarian that grows a beard"). As the absolute chronology of *wuying* and *huying* is unknown, both forms are labelled B2.

cangying 
$$(A2) >$$
 wuying  $(B2)$ , huying  $(B2)$ 

#### 4.4 Treatment of polysyllabic forms

Chinese, which many believe is a monosyllabic language, favored polysyllabic forms in stages B and C. There are two word formation types to which different treatments are applied.

#### 4.4.1 Stem + suffix

Weight is placed on the conformity of the stem, with the use of affixes such as -zi ( $\neq$ ) and -er (/L) being generally disregarded.

#### Example

For the word "fly" mentioned above, the difference between the monosyllabic form *ying* (蝇) and the bi-syllabic *yingzi* (蝇子) is disregarded, with both forms being labelled A1.

#### 4.4.2 Stem + Stem (coordinate construction)

Compound forms of coordinate construction that have two stems originated in different historical stages are treated as intermediates between the two stages unless the use has been confirmed in the literatures as being from a specific stage.

#### Example

For the word "neck" mentioned above, there are several compounds such as *bojing* (脖颈) and *boxiang* (脖项) in the dialects. As *bo* is the stem originating in stage C, and both *jing* and *xiang* are the stems originating in stage B, both *bojing* and *boxiang* are treated as intermediates of C and B (see the statistical treatment in Section 5). Although the form *boxiang* is attested in the literatures as being from the early stage of C (Yuan Dynasty), the present treatment, though mechanical, may have a merit of being applied to other "stem + stem" construction forms if they were not well attested in written texts, with the compound *bojing* being such an example.

#### 4.5 Supplementary principles

#### 4.5.1 Split and mergers of the semantic categories

The semantic category as reflected in the lexical form could be split or merged. The occurrence of one split or one merger is counted as being equivalent to a one-time change in the word form.

#### Examples

#### "eat" and "drink"

As evidenced in written texts, the two concepts "eat" and "drink" were discriminated in terms of word form in OC as *shi* (食) and *yin* (饮). Although the forms themselves were replaced by others, typically by *chi* (吃) and *he* (喝), the semantic category distinction has been maintained in a majority of Chinese dialects; therefore, this status is labelled A. Meanwhile, two semantic categories are undistinguished in many dialects and particularly in those from the Yangtze River basin; e.g., the same form *chi* is used for both concepts. As this status is considered innovative, it is labelled B.

#### "fat (animal)" and "fat (human)"

On the contrary, as the distinction between the two concepts for "fat (animal)" and "fat (human); e.g., *fei* ( $\mathbb{H}$ ) and *pang* ( $\mathbb{H}$ ); may have appeared at a later stage, this status is labelled B; when there is no distinction, it is labelled A, irrespective of the form.

#### 4.5.2 Peculiar phonetic features

Phonetic differences that can be explained by interdialectal phonological correspondence are mostly not seen as having lexical differences. However, as there are some entries in which one particular phonetic feature reflects a clear-cut geographical distribution, these peculiar phonetic features are treated as being equivalent to one lexical feature.

#### **Examples**

#### Initial consonant for the word "son"

Northern dialects (Guanhua) and Wu dialects share the stem *er* (儿); however, the Wu dialects have preserved the nasal initial (e.g.,  $[n_i t_n]$  儿子 in Suzhou), which is considered archaic, while the northern dialects have lost the nasal initial and mostly changed to a form that has zero initial (e.g.,  $[\mathfrak{d} t_n]$  儿子 in Beijing). From Chinese phonological history, the Wu form is labelled A1 and the northern form labelled B1. Note that there is another archaic form *zai* (崽) in the southern dialects, which is labelled A2.

#### Tone for the word "nose"

All dialects are found to share the form bi(鼻). However, this form actually has two varieties that are distinguished by tone: one of Qu tone origin (A1) and one of Ru tone origin (A2), with the former being distributed in the south and the latter being distributed in the north (refer to Cao 2008, Phonetic Volume, Map 37 for the detail of distribution). The difference between the form *bi* with or without the suffix *zi* (子) is disregarded.

#### 5 Statistical Analysis

Based on the survey and analysis that identified the chronology for every lexical form or its usage, scores of either 0, 1, 2 or 3 are assigned to each form or to each semantic split or merger.

"0": stage A (OC)
"1": stage B (MC)
"2": stage C (Modern Chinese)
"3": stage D (those of unknown origin)

Here, the numbers 0, 1, 2, 3 represent the degree of lexical innovation. The following principles are used for the score assignment.

1. If two forms existed in stage A; i.e., A1 and A2; "0.5" is assigned to the A2 form and "0" is assigned to the A1 form, which is identified as a "leading form." If there were two forms, B1 and B2, that were inherited from A1 and A2, "1" is assigned to the B1 form and "1.5" is assigned to the B2 form. The case of "to know" is schematized as a typical example, with the arrows indicating the inheritance relationships from one stage to another.

2. If there was only one form in stage A, for which plural forms appeared in stage B, "0" is assigned to A, and "1" is equally assigned to B1 and B2 irrespective of the presence or the absence of an inheritance relationship. Here the example of "sun" is given. As the B2 form *taiyang* lacked an ancestor in the OC period, unlike the B2 form *xiaode* for "to know," "1" is assigned to this form rather than "1.5." The treatment of the C1 form is added here.

A1 ri (目) "0" ↓ B1 ritou (日头) "1" — B2 taiyang (太阳) "1" ↓ C1 retou (热头) "2"

- If plural forms coexist in one dialect, the added scores are divided by the number of forms. For example, for the concept "to know," the Nanjing and Wanrong dialects co-use the forms *zhidao* and *xiaode*. As the scores for these two are "1" and "1.5," the average score is 1.25; i.e., (1+1.5)/2. In another example, for the concept "moon," the Guangzhou dialect co-use *yue* (月) and *yueguang* (月 光). As the former is the sole form in stage A and the latter is one of the two forms in stage B (another B form is *yueming* 月明), the average score is 0.5; i.e., (0+1)/2.
- 4. For the treatment of compounds with two stems that were used in the different historical stages mentioned in Section 4.4.2, the score assignment principle is (X+Y)/2, where both X and Y represent the stems that constituted the particular disyllabic form. For the entry "neck," the stem *bo* in *bojing* and *boxiang* is treated as C, and the *jing* and *xiang* are treated as B1 and B2; therefore, the scores for *bojing* and *boxiang* are equally (2+1)/2; i.e., "1.5." Note here that as the B2 form *xiang* lacked an ancestor in stage A, similar to the *taiyang* form mentioned above, "1" is assigned to *xiang* rather than "1.5."
- 5. There are some entries and especially pronouns that lack the old forms identified in stage A; that is, these entries lack a "0" score in the statistics. For example, OC did not seem to use the  $3^{rd}$  person pronouns "he", "she" or "they," and it is only in the literatures in stage B (MC) that the frequent use of the present southern forms qu (渠) and yi (伊) are found, which in northern China were replaced by the form ta (他), a form originally meaning "other." As the demonstrative pronouns "this" and "that" seem to have undergone radical changes, there are no correspondences found between those in the OC and those in contemporary dialects.

#### 6 Statistics and findings

After assigning the scores to each form or to each semantic split or merger occurrence for all 88 lexical entries, the average score is calculated for each of the 42 dialects being investigated, the results for which are shown in the graph on the following page, and the findings from which are summarized in the following.

- 1. The average value for all 42 dialects is 1.079, which suggests that the Chinese basic words as a whole remained as in MC with a considerable mix of innovative and preserved features.
- 2. All five Min dialects: Fuzhou, Xiamen, Jian'ou, Haikou (Hainan Island) and Haikang (Leizhou Peninsula in Guangdong Province): are found to have the most conservative features with average scores ranging from 0.707 to 0.910.
- 3. Surrounding these Min dialects, the six non-Guanhua dialects: Meixian (Hakka), Guangzhou & Dongguan (Yue), Nanping (Pinghua), Jinhua (Central Wu) and Lichuan (Central Gan): are found to have relatively conservative features with average scores ranging from 0.928 to 1.000.
- 4. The most innovative dialects, with average scores ranging from 1.278 to 1.201, tended to be scattered in the peripheral areas within the Guanhua zone: Guiyang, Liuzhou, and Chengdu in the Southwest, Wulumuqi and Yinchuan in the Northwest, and Haerbin in the Northeast. Note that these six cities, except for Chengdu, are in areas that have undergone recent Sinicization. However, as the two Xuzhou and Ji'an dialects are located near the Yellow River and alongside the Grand Canal, the frequent linguistic innovations could also have been because of active commercial activities and population movements.

5. Between the conservative group (11 dialects) and the innovative group (8 dialects), there are 23 intermediate dialects identified. A general tendency for this group is that the Guanhua dialects are relatively innovative and the non-Guanhua dialects are conservative. However, two northern Wu dialects, Shanghai and Hangzhou, and one Xiang dialect, Changsha, exhibit relatively higher scores (1.187-1.146). Notably, within the Northern Wu and Xiang zone, more conservative dialects are found: Chongming and Suzhou for northern Wu and Loudi for the Xiang dialect: which have lower scores than the Shanghai, Hangzhou and Changsha dialects (1.047-1.028). Although this evidence may in part be related to the degree of urbanization in each of these cities, it would not be the sole factor for the enhanced lexical changes. Interestingly, Xi'an, a big city that is known as the old capital Chang'an, is found to be the most conservative of all Guanhua dialects and has even a lower score than Shanghai or Hangzhou (1.133).

#### 7 Closing remarks

This paper reported on the results of a preliminary investigation, and therefore improvements are needed. First, the number of lexical entries is not sufficient for the statistical analysis, which means that more work needs to be done to ensure that the philological survey is detailed and precise, especially for those entries Wang (2018) did not deal with. Second, the certitude of our scoring method needs to be reexamined. Finally, cross dialectal measurements and estimations are required.



## **Average Values of Lexical Innovation**

#### Acknowledgements

The contents of this paper were presented at a workshop of 24th annual conference of IACL (International Association of Chinese Linguistics). Our sincere thanks are given to Naruya Saito, William Wang, Laurent Sagart and others, all of whom gave us valuable comments.

#### References

- Cao, Zhiyun ed. 2008. *Hanyu Fangyan Dituji* [Linguistic Atlas of Chinese Dialects], 3 Volumes, Beijing, Shangwuyin Shuguan.
- Cheng, Chinchuan 1997. Measuring relationship among dialects: DOC and related resources. *Computational Linguistics and Chinese Language Processing*, Vol. 2, No. 1, pp. 41-72.
- Chinese Academy of Social Sciences 2012. Zhongguo Yuyan Dituji [Language Atlas of China], New Edition, Beijing, Shangwuyin Shuguan.
- Iwata, Ray ed. 2009. Interpretative Maps of Chinese Dialects, Vol. 1, Tokyo, Hakuteisha

Iwata, Ray ed. 2012. Interpretative Maps of Chinese Dialects, Vol. 2, Tokyo, Kobun Shuppan.

- Swadesh, Morris 1952. Lexicostatistic Dating of Prehistoric Ethnic Contacts. Proceedings of the American Philosophical Society, Vol. 96, No. 4, pp. 452-463.
- Wang, Li 1958. Hanyushigao [Studies in Chinese linguistic history], Second Edition, Kexue Chuban.
- Wang, Weihui 2018. *Hanyu hexinci de lishi yu xianzhuang yanjiu* [History and present status of Chinese basic words], Beijing, Shangwuyin Shuguan.
- Wang, Yude 1960. Chugoku Godai Hogen no Bunretsu Nendai no Gengonendaigaku teki Kenkyu [The Lexicostatistic Estimation of the Time Depths of the Five Main Chinese Dialects], *Gengo Kenkyu* (Journal of the Linguistic Society of Japan), Vol. 38, pp.33-105

Yuan, Jiahua ed. 1964. Hanyu Fangyan Cihui [Lexicon of Chinese Dialects], Beijing, Wenzi Gaige Chuban.

## Characteristics of the Geographical Distribution of Words Denoting Cultural Items in Sinitic Languages

Suzuki, Fumiki

Faculty of Foreign Studies, Nanzan University

#### Abstract

This paper discusses the distributional characteristics of word forms denoting cultural items in Sinitic languages, based on linguistic maps included in "The Interpretative Maps of Chinese Dialects" (Iwata, 2009) and "Studies in Asian Geolinguistics" (ILCAA Joint Research Project 2015–2017). The items we refer to include iron, rice plant, sorghum, and corn. The geographical distribution of word forms denoting these items is closely related to the process of transmission or the degree of connection to people's lifestyles. Furthermore, this paper analyzes the different types of geographical distribution caused by time lags in transmission.

#### 1 Introduction

In this paper, we discuss the distributional characteristics of word forms denoting cultural items in Sinitic languages, based on linguistic maps included in "The Interpretative Maps of Chinese Dialects" (Iwata, 2009) and "Studies in Asian Geolinguistics" (ILCAA Joint Research Project 2015–2017). In this paper, sections 2 to 5 analyze linguistic maps of iron, rice plant, sorghum, and corn, respectively, and section 6 presents a brief conclusion from the perspective of the different types of geographical distribution resulting from time lags in transmission.

#### 2 The case of iron: Archaic forms preserved in neighboring languages

Suzuki et al. (2017) pointed out that the word forms denoting "iron" in Sinitic languages essentially show a unanimous distribution of *tie*铁. The phonetic forms of *tie* are consistent with the phonological systems of each dialect.



Map 1 "Iron" in Sinitic (Suzuki et al., 2017)

The reconstruction of Middle Chinese for "iron" by Schussler (2007) is \*t<sup>h</sup>iet. According to Baxter and Sagart (2014), the Old Chinese form is \*l<sup>c</sup>ik. Traces of these archaic forms are observed in neighboring languages. Taguchi (2018) pointed out, "In East Asia, most language groups—Korean, Ainu (via Japanese), Hmong-Mien, and Tai-Kadai—borrowed Chinese (Sinitic) words from several different sources. Hmong-Mien and Tai-Kadai tend to express more archaic word forms than Korean or Ainu. This suggests that Hmong-Mien and Tai-Kadai have had a longer contact history with the Chinese language in terms of iron metallurgy than the other languages have."

The proto Tai form for "iron" is reconstructed as \*hlek D by Li (1977). According to the map of Tai-Kadai created by Endo (2017), this form is preserved in A lek type. Sagart (1999) pointed out that this proto Tai form is close to the Old Chinese form \*ahlik; therefore, it was incorporated into Tai as well as other Southeast Asian languages following the beginning of iron manufacturing in this area. This occurred not earlier than 700-600 BCE, before the regular change \*hl- > th- and \*-ik > -it took place before Middle Chinese age. B khjak type preserves coda [-k]; C khət type can be considered as an intermediate form, in which the initial changed from \*hl- to kh-, while the final consonant changed from \*-k to -t; D thi:t type is the borrowed form from the Yue or Hakka dialects. Types A, B, C, and D are all considered loan forms from Chinese of different chronological strata and areas.



Map 2 "Iron" in Tai-Kadai (Endo, 2017)

#### 3 The case of rice plant: Semantic field subdivided in the central areas of cultivation

Rice is mainly cultivated in the southern part of China. The north-south dividing line for rice cultivation is known as the Qinling-Huaihe line. In the central area for rice cultivation, different terms are used for rice plant, husked rice, unhusked rice, steamed rice, and so on. However, outside of this area, different terms are not always used.

According to the map created by Yagi and Ueya (2016), A1 Dao  $\overline{R}$  type is mainly distributed in the northern area of China and the lower reaches of the Yangzi river. A2 Tiu  $\overline{R}$  type, distributed in Fujian province, is characterized by the glide -i- in the root. B He  $\overline{K}$  type is distributed in other southern areas, including Hunan, Jiangxi, Guangdong and Guangxi provinces. C1 Gu  $\overline{C}$  type is distributed in the upper reaches of the Yangzi river. C2 Daogu  $\overline{R}$   $\overline{C}$ , the contaminated form of dao  $\overline{R}$  and gu  $\overline{C}$ , is distributed in the upper reaches of the two types.

A conceptual model of distribution of the rice plant (RP) referring to "unhusked rice" (UR) and "Italian millet" (IM) is shown in Table 1. Yagi and Ueya (2016) identified two features that indicated the south-north opposition: First, in the southern area, "rice plant" and "unhusked rice" are distinguished; however, this is not the case in the northern area. Second, in the northern area, gu is used for "Italian millet", while the same form is used for "unhusked rice" in the southern area. The Yangzi basin,



Map 3 "Rice plant" in Sinitic (Yagi and Ueya, 2016)

reflecting its geographical location, presents intermediate features. For example, in Wuhan, rice plant and unhusked rice are denoted by identical terms, but gu? is used to denote unhusked rice. Gu? was originally used as a generic term for "grain," and therefore the use of this term may indicate the superiority or the importance of the referent as grain. Northern people rarely see planted rice, and hence they do not need a separate term to represent rice as a plant or as grain.

		RP	UR	IM	
Northern	太原	稻子 tau tsə?	谷 kuə?		
type	西安	稻子 thau tsy	谷子 ku tsy		
Yangzi	武漢	谷 ku	粟谷 ciou ku		
basin	合肥	稻 to	谷子 kuə? tsə		
Southern	蘇州	稻 dɛ	谷 ko?	粟 so?	
type	梅県	禾vo	谷 kuk	粟仔 siuk e	

Table 1 Conceptual model of "rice plant" (Yagi and Ueya, 2016)

#### 4 The case of sorghum: Lexical changes occur in the central areas of cultivation

Sorghum is a type of grain in the rice family. It is primarily cultivated in the northern area, including the Dongbei district, Hebei, and Shandong; its southern limit of cultivation roughly corresponds to the Huai River. Although it remains unclear whether sorghum is native to China, it is certain that sorghum has occupied an important position among grains since ancient times. Generally, objects that have existed since ancient times are denoted using exclusive monosyllabic forms, as observed for "iron" (*tie*铁) and "rice plant" (*dao*稻). However, most of the word forms denoting sorghum, including historical literature and modern dialects, are disyllabic. This suggests that words for sorghum were formed later than for other grains.

According to the map created by Suzuki (2009a), shushu秫秫 (B-3) is distributed in the central



Map 4 Overall sketch of "sorghum" (Suzuki, 2009a)

areas of sorghum cultivation: Shandong, Henan, and Anhui provinces. This form is disyllabic reduplicates. It is noted that *shushu*秫秫, *shushu*蜀黍, and *shushu*蜀秫 have almost the same phonetic shapes in modern north China except for the difference of tone. It is possible that *shushu*蜀黍 or *shushu*蜀秫 is an older form than the reduplicated form, and its etymology can be "millet introduced from the Shu蜀 district." After phonetic changes, such as the disappearance of the Ru $\lambda$  tone, the first and second syllable of *shushu*蜀黍 or *shushu*蜀秫 became almost equal in the phonetic shapes and changed into the reduplicated form *shushu*秫秫 can be regarded as an unmotivated form. In the sorghum cultivation area, the word form *shushu*秫秫 and its referent are related closely because of frequent usage; however, in the non-cultivation area, they are difficult to accept (see Suzuki, 2018).

The forms having lu芦 (A-2-1), such as lushu芦黍, luji芦穄, lusu芦粟, and lushu芦秫, are distributed to the south of the Huai River. Lu芦 means reed; shu黍 and ji穄 means millet; su粟 and shu秫 means Italian millet. Sorghum, millet, and Italian millet are not planted in these areas. Iwata and Nakayama (2004) pointed out that south of the Huai River, there is no need to distinguish these grains, and people are not familiar with them; therefore, lu芦 can be regarded as a type of motivating component to reinforce the connection of word forms and their referents.

Gaoliang高粱 (A-1) is the most frequently used form of all, distributed into the Dongbei district, Hebei province, and western area of China. Standard Chinese adopts gaoliang高粱 as the form for sorghum, and therefore people in the western area, where sorghum is not cultivated, also use this form. Gaoliang高粱 possibly originated in the northeastern area of China. Liang粱 was a type of Italian millet; however, its cultivation had declined. In other words, it was not necessary to distinguish sorghum and liang粱; therefore, gao高 was also a motivating component. Although it is difficult to establish continuity with gaoliang高粱 and the other word forms for sorghum, it is probable that the unmotivated form *shushu*秫秫 was used in this area and replaced by the motivated form *gaoliang*高粱 because of its relevance to daily life (see Suzuki, 2018).

#### 5 The case of corn: Overlapping with words denoting native plants

Corn was introduced into China in the 16th century. Similar to sorghum, corn is primarily cultivated in the northern area, including the Dongbei district, Hebei, and Shandong. The majority of word forms denoting "corn" comprise names of native plants together with prepositional components, namely, stems and modifiers.



Map 5 Modifiers of "corn" (Suzuki, 2009b)

Concerning modifiers, A-1  $yu \equiv$  type meaning "jade" and A-2  $bao \boxdot$  type meaning "wrapped" have the highest usage frequency, followed by B-1 bang type meaning "stick." To the north of the Huai River,  $yu \equiv$  and bang tend to exhibit an east-west opposition; these two northern forms contend with bao D that has a broad distribution in south and west China.

Concerning stems, word forms for "corn" and "sorghum" share their components. For example, in



Map 6 Northern forms of "corn" and "sorghum" (Suzuki, 2016)

the eastern area of Shanxi山西, *jiaozi*茭子 denotes "sorghum," and *yujiaozi*玉茭子 denotes "corn"; from the western area of Shanxi山西 to Shanxi陕西, *taoshu*稻秫 denotes "sorghum," and *yutaoshu*玉稻秫 denotes "corn"; in Henan province, *shushu*秫秫 denotes "sorghum," and *yushushu*玉秫秫 denotes "corn."

In Jiangsu, Anhui, Zhejiang, and Fujian provinces, forms having lu芦, such as lushu芦秫 and luji芦穄 are used for "sorghum" and forms for "corn" are the composites of lu芦 and modifier yu玉 or *bao* 包, such as yulu玉芦, *baolu*包芦; in Jiangxi and Guangdong provinces, su栗 or *gaoliangsu*高粱粟 denotes "sorghum" and *baosu*包粟 denotes "corn." Thus, word forms for "corn" adopted the names of "sorghum" as their stems.

After its introduction in China, the corn plant replaced sorghum because of its usefulness. Against the background of this non-linguistic factor, "corn" took over the word forms denoting "sorghum," which brought about the risk of homonymic collision. In some areas of Shanxi山西 and Henan provinces, "corn" took over the word forms denoting "sorghum." Therefore, to avoid homonymic collision, "sorghum" adopted other word forms or proposed the modifier *hong*红, meaning red. For example, in HenanJiyuan河南济源, *jiaojiao*茭茭 denotes "corn" and *hongjiaojiao*红茭茭 denotes "sorghum"; in ShanxiLinyi山西临猗, *taoshu*稻秫 denotes "corn" and *hongtaoshu*红稻秫 denotes "sorghum."



Map 7 Hong红 type forms (Suzuki, 2016)

In Anhui province, da type forms denoting "corn" share stems with the forms for "sorghum"; they are distinguished through a difference in modifiers: modifier xiao h is added to words for "sorghum." Da and xiao h are derived from the size of the fruit or the flower's ovary. Moreover, since da and xiao h are antonymous, this promotes systematization of the lexicon. In relation to the formation of da and xiao h types, two interpretation can be proposed: (1) "corn" first adopted the word forms for "sorghum" as the stem together with the prepositional component da  $\pm$ ; thereafter, the modifier xiao h was added to words for "sorghum" to adjust the lexical system; (2) "corn" took over the word forms denoting "sorghum," and the modifier xiao h was added to words for "sorghum" to avoid homonymic collision; thereafter, the modifier da was added to words for "corn" to adjust the lexical system.

*Hong*  $\leq$  type forms denoting "sorghum" and da  $\gtrsim$  type forms denoting "corn" can be regarded as the traces of homonymic collision; however, both types are temporary forms. In Map 8, da  $\gtrsim$  type is used in ten more places, while the map included in the "Linguistic Atlas of Chinese Dialects" (Cao,

2008) only has six places using  $da \pm type$ . This suggests that these forms are a transient phenomenon and are easy to decline.



Map 8 Da大 type forms for "corn" and xiao小 type forms for "sorghum" (Suzuki, 2016)

#### 6 Conclusion

The preceding analysis elucidates that the geographical distribution of word forms denoting cultural items is closely related to the degree of connection to people's lifestyles.

Both iron and rice plant have existed since ancient times. Ironware was essential in all the areas, including neighboring areas of China, and the technology of ironware was transmitted along with names for it, whereas rice plant was mainly limited to rice cultivation areas; therefore, words for iron exhibit a unanimous distribution of the same form, whereas division of the semantic field for rice plant can differ according to whether it is being used in a cultivation area or not.

Sorghum has also existed since ancient times; however, the word structures suggest that words for sorghum were formed later than for other grains such as rice and millet. Word forms denoting sorghum experienced lexical changes caused by the degree of connection of word forms and their referents.

Corn was introduced and transmitted to China in modern times. The distribution of stems of the word forms have overlapped with the names for native grains, especially sorghum. The broad distribution of its word structure suggests a high speed of transmission, which is characterized as a phenomenon of the modern era.

#### Acknowledgements

This paper is supported by JSPS Grants-in-Aid for Scientific Research (KAKENHI) Grant Number JP18H00670, JP19K13187.

#### References

Baxter, W.H. and L. Sagart. 2014. Old Chinese: A New Reconstruction. Oxford University Press, Oxford, England.

Cao, Zhiyun. 2008. Linguistic Atlas of Chinese Dialects. The Commercial Press, Beijing, China.

Endo, M. 2017. Iron: Tai-Kadai. Studies in Asian Geolinguistics V-IRON-:15-16.

Iwata, R. ed. 2009. The Interpretative Maps of Chinese Dialects. Hakuteisha, Tokyo, Japan.

- Iwata, R. (岩田礼) and S. Nakayama. (中山里美) 2004. 「コウリャン」関連語幹をめぐる諸説. 歴史文献 データと野外データの綜合を目指した漢語方言史研究(3):漢語方言地図集(稿)第4集,平成13-15 年度科学研究費総合研究(B)研究成果報告書-第3分冊(研究代表者:太田斎):140-144.
- Li, Fanggui. 1977. A Handbook of Comparative Tai. Oceanic Linguistics special publication No. 15. University of Hawaii Press, Honolulu, USA.
- Sagart, L. 1999. The Roots of Old Chinese. John Benjamins, Amsterdam, Netherlands.
- Schussler, A. 2007. ABC Etymological Dictionary of Old Chinese. University of Hawaii Press, Honolulu, USA.
- Suzuki, F. 2009a. Map 25 Sorghum. Iwata, R. ed. 2009. *The Interpretative Maps of Chinese Dialects*:162-165. Hakuteisha, Tokyo, Japan.
- Suzuki, F. 2009b. Map 26 Corn. Iwata, R. ed. 2009. *The Interpretative Maps of Chinese Dialects*:166-171. Hakuteisha, Tokyo, Japan.
- Suzuki, F. (鈴木史己) 2016. 言語地理学的手法をもちいた漢語語彙史研究―新来事物を表す語を手がかりに. 京都大学博士論文.
- Suzuki, F., T. Ueya and K. Yagi. 2017. "Iron" in Sinitic. Studies in Asian Geolinguistics V-IRON-:7-8.
- Suzuki, F. (鈴木史己) 2018. 試論表名物詞多樣化的成因—以表〈高粱〉義詞為例. 中國語學研究『開篇』, 36:179-192.
- Taguchi, Y. 2018. "Iron" in Asia. Studies in Asian Geolinguistics VIII:9-10.
- Yagi, K. and T. Ueya. 2016. "Rice" in Sinitic languages. Studies in Asian Geolinguistics II-RICE-:9-10.

## Prehistoric Cultural Change in East Asia and Language Groups

Kazuo Miyamoto

Faculty of Humanities, Kyushu University

#### Abstract

Following a period of cooler climatic conditions, around 3000 BC, East Asia became divided into four areas according to differences in subsistence activities: agricultural societies in mainland China, herding societies in northern Asia, and two secondary agricultural areas. The existence of these two secondary agricultural areas suggests that hunter-gatherer societies were influenced by these agricultural societies from the Chinese mainland, in turn becoming agricultural societies themselves.

At the end of the Neolithic period, around 2000 BC, bronze production techniques spread through the Eurasian grasslands to the Central Plains of China. It was on the Central Plains of China that a technique emerged for creating ritual bronze vessels using clay molds. On the other hand, the Northern Bronze of the Eurasian grassland area developed into bronze weapons, particularly daggers.

As discussed by Prof. Peter Bellwood concerning the origins of language groups in East Asia, it is probably the case that the Austronesian language group spread to Taiwan from southeastern China along with rice and millet agriculture at around 3000 BC. On the other hand, agricultural societies in the Central Plains are thought to have belonged to the Sino-Tibetan language group. I infer that Proto-Japonic and Proto-Koreanic originated in Manchuria based on the hypotheses of Janhunen (1996) and Robbeets (2017).

Pianpu culture with millet agriculture spread from western Liaoxi district to Liaodong district in Manchuria at around 2400 BC following the emergence of the Great Wall region. Pianpu culture influenced the Konggwiri type, which triggered the emergence of Mumun culture, characterized by banded pottery. The genealogical relationship between the same pottery production techniques in the Pianpu, Mumun and Yayoi cultures is thought to indicate the spread of Proto-Japonic. During the 6<sup>th</sup> and 5<sup>th</sup> centuries BC, the rolled vessel culture of the Liaoning type dagger culture spread from the Lioaxi district to Liaodong district and Korean Peninsula during invasion by the Yan state. During this process, Proto-Koreanic spread from the western Liaoxi district to the Korean Peninsula.

#### 1 Introduction

The Late Palaeolithic in East Asia is divided into two genealogical lines, based on the blade techniques of northern China and pebble stone tool techniques of southern China in the Pleistocene. At the beginning of the Neolithic period, millet agriculture developed in the middle and lower Yellow River basin of northern China, and rice agriculture developed in the middle and lower Yangtze River basin of southern China. These two forms of agriculture developed based on different geographical and biological conditions, and consisted of genealogically different stone tools techniques. Both types of agriculture gradually mixed to form agricultural societies on the Chinese mainland in the Hypsithermal period of the Holocene.

Areas of grassland became increasingly fertile in the north of the Great Wall region from 3500 BC to 3000 BC, a time of cooler climatic conditions. Following this, it became divided into four areas in East Asia according to differences in subsistence activities: agricultural societies in mainland China, herding societies in northern Asia, and two secondary agricultural areas (Fig. 1). The existence of these two secondary agricultural areas suggests that hunter-gatherer societies were influenced by these agricultural societies from the Chinese mainland, in turn becoming agricultural societies themselves.



Fig. 1 – Regional division based on subsistence activities in East Asia

One of the secondary agricultural areas is northeast Asia, which includes southern Far Eastern Russia, the Korean Peninsula and the Japanese archipelago (Miyamoto 2014a, 2015). The other is southern China and southeast Asia.

It is thought that such cultural units based on subsistence activities formed the units of social groups. Some social groups may have been related to language groups. In addition issues concerning to the Indo-European language group, archaeological explanations on the movement or processes of change of material culture groups should be connected with linguistic explanations of language groups. paper attempts to provide This an archaeological explanation for the movement of language groups in Prehistoric East Asia.

#### 2 Prehistoric cultural change in East Asia

At the end of the Neolithic period, around 2000 BC, bronze production techniques spread through the Eurasian grasslands to the Central Plains of China. Not only bronze products but also wheat and barley spread to the Central Plains of China via the Eurasian grasslands from western Asia. It was on the Central Plains of China that a technique emerged for creating ritual bronze vessels using clay moulds. In this way, the Central Plains developed a unique bronze culture with a mixture of agricultural practices, leading to the establishment of early states like the Xia, Shang and Zhou dynasties. On the other hand, the Northern Bronze of the Eurasian grassland area developed into bronze weapons, particularly daggers. Northern Bronze of this kind spread not only to northeast Asia—i.e. the Liaoning dagger culture of Liaoxi, Liaodong and the Korean Peninsula—but also to the eastern Tibetan Plateau and Yunnan area (Fig. 2). Bronze daggers featuring a three-pronged design influenced by the Northern Bronze culture could be found in the eastern Tibetan Plateau and Yunnan area (Miyamoto 2014b).

Iron products also spread in the same way. Wrought iron techniques spread through the Eurasian grasslands to the Great Wall region from western Asia during the 9<sup>th</sup> and 8<sup>th</sup> centuries BC. Wrought iron techniques arrived in the Central Plains of China, and subsequently developed into iron casting techniques during the 6<sup>th</sup> and 5<sup>th</sup> centuries BC. The Central Plains of China developed under political centralization to gain new farming land with iron tools in the Warring States period, such as with the Yan and Chu states.

Following this, the Qin and Han dynasties extended their territorial area to include the Korean Peninsula, and to Vietnam with an iron product culture. On the other hand, wrought iron techniques spread not only to northern Asia but also to the eastern Tibetan Plateau and Yunnan area. It is probable that wrought iron techniques spread from the Yunnan area to southeast Asia.

#### **3** Spread of language groups in East Asia

Prof. Peter Bellwood discussed the origins of language groups in relation to archaeological cultures in Neolithic East Asia (Bellwood 2005). However, archaeology does not provide us with the evidence needed to indicate which language groups originated with which cultures in prehistoric times. On the other hand, the territory of current language groups indicates the probability that archaeological cultural groups in Neolithic times spread along with language groups.

In the case of prehistoric East Asia, the most important thing to bear in mind is the establishment of two cultural groups—agricultural societies and herding societies—from 3500 BC to 3000 BC due to cooler and drier climate conditions. It is believed that these two areas developed two different language



Fig. 2 – Bronze culture groups and genealogical relationship between them in East Asia

groups: Sino-Tibetan and Altaic groups (Transeurasian language group). These two areas became two different states in ancient times: the Han dynasty and the Xiongnu herding state. Characters on oracle bones indicate that they spoke a Sino-Tibetan language. The Shang dynasty was established by descendants of the Yangshao culture of the middle Neolithic in the middle Yellow River basin. It is believed that the language of the Yangshao people in the Central Plains of China belonged to the Sino-Tibetan language group. In the late Neolithic, the Great Wall region established the same cultural aspects as stone wall settlements. At the same time, the Northern Bronze culture spread east from the western Eurasian grasslands to reach the Great Wall

region. In this area, Xiongnu herding states emerged after the Northern Bronze Age. And people in these Xiongnu herding states spoke languages belonging to the Altaic language group (Beckwith 2011). The Great Wall region of the Northern Bronze Age also belonged to the Altaic language group.

Domesticated rice spread from the lower and middle Yangtze River basin to southern China from 3500 to 3000 BC due to cooler climatic conditions. It is believed that the secondary agricultural areas—northeast Asia and southern China—started at the same time, around 3500 BC to 3000 BC, due to cooler climatic conditions (Miyamoto 2014a 2015). Recently, foxtail millet dating to 4500 BC was found at Shangshan Site, Zhejiang province in the lower Yangtze River basin (Zhao & Jiang 2016). Not only domesticated rice but also millet spread from the lower Yangtze River basin to southeast China, including Fujian province. In this area, rice and millet were also found during the late Neolithic (Deng et al. 2018). In Taiwan, rice and millet were discovered at Nankuanli Site dating to around 3000 BC (Tsang Chen-Hwa et al. 2017). Therefore, rice and millet agriculture spread from the lower Yangtze River basin via the Fujian area to Taiwan around 3000 BC.

It is probably the case that the Austronesian language group spread to Taiwan from southwestern China along with rice and millet agriculture at around 3000 BC (Sagart et al. 2018). This archaeological evidence suggests that a secondary agricultural area was established in southern China and southeast Asia. It is also probable that Austroasiatic languages spread at this time from the middle Yangtze River basin to southern China and southeast Asia along with painted pottery (Bellwood 2005).

#### 4 Archaeological explanation of Proto-Japonic and Proto-Koreanic

The relationships between archaeological cultures changed following cooler and drier conditions, around 3000 BC. Similarities appeared between the pottery cultures of the Laohushan and Xiaheyan cultures in the Great Wall region. Herding societies developed in the Great Wall region following cooler and drier climatic conditions. These areas had the same Northern Bronze culture after around 2000 BC. On the other hand, Neolithic culture groups in the middle and lower Yellow River and the middle and lower Yangtze River maintained a mutual relationship through jade goods. These areas constituted the early states of the Xia, Shang and Zhou dynasties, particularly with ritual bronze vessels in the Bronze Age.

A clear distinction emerged between herding societies and agricultural societies, not only by subsistence activities but also by different bronze cultures. While even agricultural societies received bronze techniques from the Eurasian grassland area, including the Great Wall region, agricultural societies developed a unique bronze culture, such as ritual bronze vessels. Northern Bronze culture is conjectured to belong to the Transeurasian language group. Agricultural societies in the central plains are thought to have belonged to the Sino-Tibetan language group. I infer that Proto-Japonic and



Fig. 3 – Genealogical relationship of pottery between Pianpu culture and Mumun culture

Proto-Koreanic originated in Manchuria based on the hypotheses of Janhunen (1996) and Robbeets (2017).

Pianpu culture with millet agriculture spread from the western Liaoxi district to Liaodong district at following around 2400 BC the emergence of the Great Wall region. Pianpu culture influenced the Konggwiri type, which triggered the emergence of Mumun culture, characterized by banded (Fig. 3). A pottery. genealogical relationship existed between both cultures, which utilized the same unique pottery production techniques. This genealogical relationship between the same pottery production techniques in the Pianpu and Mumun cultures is thought to indicate the spread of Proto-Japonic. (Miyamoto 2016).

The emergence of Yayoi culture was influenced by Mumun culture

during the 9<sup>th</sup> century BC. A handful of people from the Mumun culture on the southern Korean Peninsula immigrated to northern Kyushu, mixing with the majority indigenous Jomon people (Miyamoto 2019). A genealogical relationship existed between Yayoi pottery and Mumun pottery, with both cultures utilizing the same unique pottery production techniques. This situation indicates the spread of Proto-Japonic from the Mumun culture of the Korean Peninsula to Yayoi culture in northern Kyushu. Itazuke type pottery based on the pottery production techniques of the Mumun culture was established in northern Kyushu during the early Yayoi period. Following this, Itazuke pottery spread to the whole of western Japan. This also signifies the spread of Proto-Japonic to western Japan during the early Yayoi period, the 6<sup>th</sup> to 5<sup>th</sup> centuries BC (Miyamoto 2016).

At the same time, the rolled vessel culture of the Liaoning type dagger culture spread from the Lioaxi district to Liaodong district and the Korean Peninsula during invasion by the Yan state. The rolled vessel culture spread from Liaodong to the Korean Peninsula according to the pottery chronology (Miyamoto 2016). This indicates the spread of Proto-Koreanic. The rolled vessel culture led to the development of the Three Kingdoms societies through the Proto-Three Kingdoms period. Therefore, the spread of rolled vessel culture means the spread of Proto-Koreanic.

Proto-Japonic exists earlier than Proto-Koreanic in Unger's chronological scheme (2009). Proto-Japonic was spoken on the Korean peninsula according to Koguryo toponymical data (Whitman 2011). Therefore, Proto-Japonic spread eastward from the western Liaoxi and Liaodong areas to the Korean Peninsula with the Pianpu culture at around 2400 BC, and finally reached northern Kyushu at the beginning of the Yayoi culture, 9<sup>th</sup> century BC. On the other hand, Proto-Koreanic of the Liaoning type dagger culture spread eastward from Liaoxi district to the Korean Peninsula during the 6<sup>th</sup> and 5<sup>th</sup> centuries BC during invasion by the Yan state. The pottery styles and production techniques of Mumun and Yayoi pottery originated from the Pianpu type of the western Liaoxi and Liaodong districts in the late Neolithic. This area, therefore, constitutes the homeland of the Japanese language family (Miyamoto 2016).

#### 5 Conclusion

Herding societies developed in the Great Wall region following cooler and drier climatic conditions, around 3000 BC. These areas had the same Northern Bronze culture after around 2000 BC. Northern Bronze culture is conjectured to have belonged to the Transeurasian language group. On the other hand,

agricultural groups in the middle and lower Yellow River and the middle and lower Yangtze River maintained a mutual relationship and constituted early states, with ritual bronze vessels during the Bronze Age. Agricultural societies in the central plains of China are conjectured to have belonged to the Sino-Tibetan language group.

Proto-Japonic and Proto-Koreanic, part of the Transeurasian language group, originated from Manchuria, located at the eastern edge of the Great Wall region as part of the Northern Bronze culture. The bronze products of the Mumun culture on the Korean Peninsula and of Yayoi culture on the Japanese archipelago belonged genealogically as part of the Northern Bronze culture.

In the process, new pottery production techniques originating from the Pianpu culture in the western Liaoxi district of Manchuria spread after 3000 BC. Proto-Japonic is believed to have spread from the western Liao-xi district to the Korean Peninsula through the Liaodong district accompanying the same pottery production techniques. Based on the process of diffusion of these pottery production techniques, Mumun culture with irrigated agriculture was established on the Korean Peninsula. At the beginning of the Yayoi, immigrants from southern Korea brought Proto-Japonic with them. It is believed that the transitional time from Jomon to Yayoi is in accordance with the traditional timeframe for the transition from Jomon language to Proto-Japonic in Northern Kyushu, because pottery production techniques were learned through language rather than custom.

The Yan state in the early Iron Age extended its territory and indirectly occupied the western Liaoxi district of the Northern Bronze culture during 6<sup>th</sup> and 5<sup>th</sup> centuries BC. This political interference triggered the movement of the rolled vessel (Jeontodae pottery) culture of Manchuria to the Korean Peninsula. The Proto-Koreanic language group is thought to indicate rolled vessel culture with slender dagger culture and early iron products. During the process of territorial expansion by the Yan state, Proto-Koreanic spread from the western Liaoxi district to the Korean Peninsula. This Proto-Koreanic language group led to the establishment of the Three Kingdom entities on the Korean peninsula during the early Iron Age of Korea.

#### References

Bellwood, Peter. 2005. First Farmers: The Origins of Agricultural Societies, Blackwell Publishing, Malden USA.

- Beckwith, Christopher. 2011. Empires of the Silk Road: A History of Central Eurasia from the Bronze Age to the Present.
- Deng Z., Huang H., Fan X., Huang Y. and Lu H. 2018. The ancient dispersal of millet in southern China: New archaeological evidence. The Holocene 28 (1): 34-43.
- Janhunen, J. 1996. Manchuria: An Ethnic History. Helsinki: Suomalais-Ugrilainen Seura.
- Robbeets. 2017. The Transeurasian Languages. Raymond Hickey ed., *The Cambridge handbook from Part II Case Studies for Areal Linguistics*. pp. 586-626, Cambridge University Press.
- Isabelle, P. and L. Bourbeau. 1984. TAUM-AVIATION: its technical features and some experimental results. *Computational Linguistics*, 11(1):18-27.
- Miyamoto, Kazuo. 2014a. Human Dispersal in the Prehistoric Era in East Asia. Takahiro Nakahashi and Fan Wenquan, ed. Ancient People of the Central Plains in China. pp.63-83, Kyushu University Press.
- Miyamoto, Kazuo. 2014b. The Emergence and Chronology of Early Bronzes on the Eastern Rim of the Tibetan Plateau. Edited by Anke Hein, *The "Crescent-Shaped Cultural-Communication Belt": Tong Enzheng's Model in Retrospect, An examination of methodological, theoretical and material concerns of long-distance interactions in East Asia.* BAR International Series 2679, pp. 79-88, Archaeopress, Oxford, England.
- Miyamoto, Kazuo. 2015. The Initial Spread of Agriculture into Northeast Asia. In *Asian Archaeology* Vol.3 pp.11-26, Research for Chinese Frontier Archaeology of Jilin University, Science Press, Beijing.
- Miyamoto, Kazuo. 2016. Archaeological Explanation for the Diffusion Theory of the Japonic and Koreanic Languages. In *Japanese Journal of Archaeology* Vol.4 No.1, 53-75
- Miyamoto, Kazuo. 2019. The spread of rice agriculture during the Yayoi period: From the Shandong Peninsula to Japanese Archipelago via Korean Peninsula. In Japanese Journal of Archaeology Vol. 6 No.2, 109-124.
- Sagart L., Hsu T., Tsai Y., Wu C., Huang L, Chen Y., Chen Y., Tseng Y., Lin H., and Hsing Y. C. 2018. A northern Chinese origin of Austronesian agriculture: new evidence on traditional Formosan cereals. Rice 11:57

- Tsang C., Li K., Hsu T., Tsai Y., Fang P. and Hsing Y. 2017 Broomcorn and Foxtail millet were cultivated in Taiwan about 5000 years ago. Botanical Studies 58:3
- Unger, J. Marshall. 2009. The Role of Contact in the Origins of the Japanese and Korean Languages, University of Hawaii Press, Honolulu.
- Whitman Johon. 2011. Northeast Asian Linguistic Ecology and the Advent of Rice Agriculture in Korea and Japan. In Rice (4): 149-158.
- Zhao, Zhizhun & Jiang, Leping. 2016. Analysis on the Plant Remains by Flotation Unearthed in the Shangshan Site, Pujiang County, Zhejiang Province. Cultural Relics in Southern China (3): 109-130 (In Chinese).

## Notes on front rounded vowels in Sinitic languages

Kenji Yagi

Kokushikan University

#### Abstract

The geographical distribution of front rounded vowels in the world is "relatively concentrated" in the central to "northerly parts of the Eurasian land mass," suggesting an areal influence on the occurrence of these vowels in the vowel inventory (Maddieson 2013). In Maddieson's map from 2013, date points are too few to conduct geo-linguistic analysis. However, there is a certain tendency of distribution. In the central area of distribution, front rounded vowels are used as both mid and high vowels, while only one of these is used in the peripheral area, especially in Eastern Eurasia. China is located at the border of the distribution; in Southern China, a lot of dialects do not possess front rounded vowels (=y), although many of the Northern dialects do. Many "y" vowels found in Chinese dialects are derived from the "iu" type diphthongs (Division-III,closed  $\Xi \begin{subarray}{l} \oplus \mbox{c} \oplus \mbox{$ 

#### 1 Introduction

Worldwide geographical distribution of "front rounded vowels" (FRVs) is "relatively concentrated" in the middle and "northerly parts of the Eurasian land mass," suggesting the areal influence on the occurrence of FRVs in the vowel inventory (Maddieson 2013). There are insufficient data points to conduct a geo-linguistic analysis in the map created by Maddieson (2013). However, a tendency can be observed regarding distribution: in the center of the distribution area, FRVs are found both in mid and high vowels, however, only one form of the FRVs is used in peripheral areas, particularly in eastern Eurasia.

China is located on the edge of the distribution of FRVs; in southern China many dialects do not possess FRVs (≒y) but many northern dialects do (Cao [曹志耘] 2008[语音卷]: 116-117). Many of the "y" vowels found in Chinese dialects derive from the "iu"-type diphthongs (Division-III, closed [三等 合口]) of Middle Chinese (Wang [王力]1985). The change from [iu] to [y] might be connected to other phonetical or phonological changes, such as dentilabialization (Pulleyblank 1984: 86).

Cao [曹志耘] (2008) addressed only limited forms of FRVs, this paper presents maps depicting the occurrence of many varieties of FRVs found in Chinese dialects and other languages in China and investigates FRVs from a geo-linguistic perspective.



Map 1. Distribution of front rounded vowels worldwide (Source: Maddieson 2013, WALS Feature 11)

## 2 Methods

This paper investigates the presence of FRVs using published literature. FRVs found in Chinese dialects are classified as below. Central rounded vowels (CRVs) are also included in this classification because they are important for discussing the formation of FRVs. Categorization of FRVs and CRVs is based on phonetic criteria and is not phonologically processed.

Type a: Close front rounded vowels: [y], [Y] Type b: Close-mid front rounded vowel: [Ø] Type c: Open-mid front rounded vowel: [œ] Type d: Close central rounded vowels: [ʉ], [ų], [ų] Type e: Close-mid central rounded vowel: [Θ] Type f: Others

## **3** Maps of FRVs in Chinese Dialects



Map 2. Numbers of FRVs (including CRVs)

Map 2 shows the distribution of FRVs (including CRVs) in Chinese dialects. Many dialects possess one or more FRVs. Dialects possessing only one type of FRV are the most common, followed by those with two or three types, and some dialects possess four types. Tentative numbers of FRV and the number of locations in which they occur are shown in Table 1.

Table 1. Number of locations in which specific quantities of FRVs (including CRVs) are present (tentative)

Number of FRVs	0	1	2	3	4
Number of locations	171	332	132	36	7

Dialects possessing more than two FRVs are mainly distributed south of the Yangtze River and are especially abundant in the coastal area ranging from Zhejiang to Fujian and Guangdong to Jiangxi. Many of northern dialects only possess one type of FRV, usually the "y"-type. However, from Guangxi, Guangdong to Jiangxi, many dialects do not have FRVs. Thus, there are sometimes situations in which FRV- rich dialects and dialects without FRVs are adjacent.

This paper examined 678 Chinese dialects, and FRVs are observed in 507 (74.8%) of them. The most frequently occurring FRV was Type a: [y] (including [y]), identified in 480 locations (94.7% of all observed FRVs). In 308 dialects, only one kind of close FRV, [y] or [y] was present; in five locations (1%), two kinds ([y] and [y]) were present. Other types of FRV were mainly distributed in the southern area, and many of them co-occurred with [y].

Type b: [ø] appears in 90 locations (17.8% of all FRVs), and it co-occurs with [y] or another FRV except in four locations (4.4%) don't co-occur with [y] or other FRV.

Type c:  $[\alpha]$  appears in 74 locations (14.6% of all FRVs) and co-occur with [y] or another FRV in only two locations (2.7%).

Similarly, Type d: [u][u][u] occurs in 43 locations (8.5% of all FRVs), and 12 locations (27.9%) alone; Type e:  $[\Theta]$  occurs in 36 locations (7.1% of all FRVs) and six locations (16.7%) alone.

For each type of FRV, the number of location in which it occurs and its rates of co-occurrence are shown in Table 2.

	Type a: [y][Y]	Type b: [ø]	Type c: [œ]	Type d: [ʉ][ਪ][ਪ]	Type e: [ $\theta$ ]
<u> </u>	480 locations	90 locations $95 (04.40/)$	74 locations	$\frac{43 \text{ locations}}{20 ((7.49/))}$	30   00   00   00   00   00   00   00
with Type a	_	83 (94.4%)	70 (94.3%)	29 (67.4%)	29 (80.0%)
Co-occurrence	85 (17.7%)	—	26 (35.1%)	9 (20.9%)	5 (13.9%)
with Type b					
Co-occurrence	70 (14.6%)	26 (28.9%)	—	4 (9.3%)	2 (5.5%)
with Type c					
Co-occurrence	29 (6%)	9 (10%)	4 (5.4%)	—	5 (13.9%)
with Type d					
Co-occurrence	29 (6%)	5 (5.6%)	2 (2.7%)	5 (11.6%)	—
with Type e					

Table 2. Numbers of locations and ratios of co-occurrence (tentative)

In Table 2, 85 locations (94.4%) in which Type b occurred also featured Type a, and 70 locations (94.5%) in which Type c occurred also featured type a. Although ratios of co-occurrence with Type a decreased somewhat for CRVs (type d: 67%, type e: 80.6%), it is possible to suggest that the occurrence of type a is a requirement for the occurrence of another FRV.

We examine the distribution of the four types of FRVs other than type a below.



Map 3. Type b: [ø]

Map 3 shows the distribution of Type b:  $[\emptyset]$  (Close-mid front rounded vowels). Type b FRVs are distributed widely in the southern area, from southern Jiangsu to Guizhou, Guangxi. The distribution of Type b is relatively wide —90 locations feature this vowel, and its number of occurrence is the largest after Type a.



Map 4. Type c: [œ]

Map 4 shows the distribution of Type c:  $[\alpha]$  (Open-mid front rounded vowels). The prevalence of type c follows Type b (map 3): 74 locations feature this vowel. The distribution area is somewhat smaller than that of Type b, however, like Type b, Type c distributed widely in the southern area, and distributions of type b and type c are adjacent.



Map 5. Туре d: [ʉ] [ų] [ų]

Map 5 shows the distribution of Type d: [u] [v] [v] [v] (close central rounded vowels). The distribution area of this vowel is mainly around the lower Yangtze River and is scattered throughout the northern and southern inland area. Type d is the most northerly distributed type other than Type a.



Map 6. Type e:  $[\Theta]$  and others

Map 6 shows the distribution of Type e:  $[\Theta]$  (close-mid central rounded vowel). The distribution area of Type e is similar to that of Type d (Map 5), mainly near the lower Yangtze River and central Jiangxi to the northern Guangdong area.

Map 7 combines the above maps (Maps 4 to 7). The distribution of each vowel appears somewhat sporadic; however, they are distributed throughout the southeastern area of China. Central rounded vowels tend to be distributed in the northern area.



Map 7. Unified map of FRVs (excluding Type a) and CRVs

## 4 Relationship with Middle Chinese

Places	The Sixteen Rhyme Groups	Example	Sources
Yong'an [福建永安]	Liu [流]	头走沟厚	Chen et al.1991
Sanyuan [福建三元]	Liu [流], Xian [咸]	头走沟厚咸	Chen et al.1991
Suzhou [江蘇蘇州]	(Liu [流]), Xian [咸], Shan [山]	剖南甘占餐川渊	Ye 1988
Wuyuan	Zhen [臻], Zhi [止], Xie [蟹],	术恤儿耳锐阅越	Hirata et al.1998
[安徽婺源]	Shan [山]		
Gao'an	Yu [遇], Xie [蟹], Zhi [止], Tong	女緒書制世税智	Liu 1999
[江西高安]	[通]	支児玉	
Zhongshan	Guo [果], Yu [遇], Xie [蟹], Liu	多他贺模初菜待	Xie 2007
[广西钟山]	[流], Dang [宕], Jiang [江]	搂鄂桌	

Table 3. Middle Chinese sources of [ø]

Table 3 shows how the mid-close front rounded vowel  $[\emptyset]$  (type b) in some dialects points to a Middle Chinese sources for the rhyme. The Middle Chinese sources are quite variable, and no rhyme groups are found consistently throughout all the dialects. This situation suggests at least some instances of  $[\emptyset]$  were formed rather recently.

#### 5 FRVs in Non-Chinese Languages

Map 8 shows the distribution of FRVs in non-Chinese languages in China. In general, northern languages possess more FRVs than southern languages. FRVs sometimes only exist in Chinese loanwords. However, in central areas such as Sichuan province, some languages possess three types of FRV.

Map 9 shows the distribution of Chinese dialects and other languages which do not feature FRVs. The absence of FRVs in Chinese dialects and other languages show a clear geographical succession (In Map 9, small purple circles indicate Chinese dialects, and large white circles indicate non-Chinese languages).



Map 8. Distribution of FRVs in non-Chinese languages in China



Map 9. Chinese and non-Chinese languages that lack FRVs

Table 4. Absence of	1.17 6 2	amon	ig iang	guage	group	s m c	mma							
	彝	景	缅	羌	侗	苗	蒙	南	南	印	混	满	突	臧
		颇			台	瑶	古	岛	亜	欧	合	通	厥	
Number of	7	6	2	2	15	4	2	10	7	1	1	0	0	0
languages lacking														
FRVs														
Total number of	15	9	6	12	22	7	7	16	9	1	5	6	9	4
languages														

Table 4 Absence of FRVs among language groups in China

(Source: Sun et al. 2007)

Table 4 shows the distribution of languages lacking FRVs in languages in China. Lacking of FRVs is found in almost all languages in southern area.

#### 6 Conclusion

This paper investigates the distribution of FRV of both Chinese and non-Chinese languages in China. FRVs are regularly found in supra dialects or languages; therefore, it is important to examine the formation of FRVs from a geo-linguistic perspective. China is located at the border of the eastern Eurasian FRV distribution; in southern China, many dialects do not possess front rounded vowels [y], although many of the northern dialects do. Many of [y] FRVs found in Chinese dialects derive from the "iu"-type diphthongs (Division-III, closed  $[= \Leftrightarrow ch ]$ ) of Middle Chinese (Wang 2008).

The distribution of FRVs may help to examine Mantaro Hashimoto's theory that Chinese languages have changed via the influence of the northern languages (Hashimoto [橋本] 1978).

#### Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP16K16820, JP18H00670.

#### References

Cao, Zhiyun [曹志耘]. 2008. 《汉语方言地图集》语音卷. 商务印书馆.
Chen, Zhangtai [陈章太] 李如龙. 1991. 《闽语研究》. 语文出版社.
Hashimoto, Mantaro [橋本萬太郎]. 1978. 『言語類型地理論』. 弘文堂.
Hirata, Shoji [平田昌司] 趙日新 劉丹青 冯爱珍 木津祐子 溝口正人. 1998. 《徽州方言研究》. 好文出版.
Liu Lunxin [刘纶鑫] 主编. 1999. 《客赣方言比较研究》. 中国社会科学出版社.
Maddieson, Ian. 2013. Front Rounded Vowels. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary

Anthropology. (Available online at http://wals.info/chapter/11, Accessed on 2019-03-11.) Pulleyblank, E.G. 1984 *Middle Chinese*. University of British Columbia Press. Sun, Hongkai [孙宏开] 胡增益 黄行 (eds.). 2007. 《中国的语言》. 商务印书馆. Wang, Li [王力]. 1985. 《汉语语音史》. 中国社会科学出版社(1998年重印). Xie, Jianyou [谢建猷]. 2007. 《广西汉语方言研究》. 广西人民出版社. Ye, Xiangling [叶祥苓]. 1988. 《苏州方言志》. 江苏教育出版社.

## How Tibetans classify pigs in Tibetic languages in the eastern Tibetosphere: Revisiting the *pig issue* through geolinguistics

Hiroyuki Suzuki<sup>a</sup>

<sup>a</sup>IKOS, University of Oslo / National Museum of Ethnology

#### Abstract

This article primarily describes the morphological variation of lexical forms denoting domestic pigs in Tibetic languages spoken in the eastern Tibetosphere and examines the relationship between the geographical distribution and the pig-breeding habit from a geolinguistic perspective. In Literary Tibetan, many words for domestic pigs are derived from the root *phag*, meaning 'pig', whereas in spoken languages, various forms that both include and exclude the root are attested. The geolinguistic analysis shows that the lexical complexity related to pigs is strongly connected to the pig-breeding habit.

#### 1 Introduction

Most words regarding the category of domestic pigs in Literary Tibetan (henceforth LT) are derived from the root *phag*, e.g. *pho phag* 'boar' (literally 'male'+'pig'), *mo phag* 'sow' (literally 'female'+'pig'), and *phag phrug* 'piglet' (literally 'pig'+'child').<sup>1</sup> However, in spoken varieties of Tibetic languages and dialects (see Tournadre 2014 for the definition of 'Tibetic'), we find various lexical forms and even differences in the categorisation. We notice the existence of languages, for example, which distinguish 'male piglet' from 'female piglet', and which distinguish 'sow with her piglets' or 'sow without them'. These languages are generally found in the eastern Tibetosphere, where many minor non-Tibetic (Tibeto-Burman, as well as Sinitic, Mongolic, and Turkic) languages are spoken (Roche & Suzuki 2017, 2018).

According to Yang et al. (2011:6), Tibetan pigs originated in the Tibetan highlands from a genetic viewpoint. This means that Tibetans have not been strangers to pig-domestication and breeding since prehistoric times. However, from the viewpoint of linguistics, the pig has not played a crucial role in Tibetans' lives, unlike cattle, where rich lexical forms are used to distinguish different types from each other (cf. 'Brug-mo-mtsho 2002; Sung & lHa-byams-rgyal 2005; Shao 2018; Ebihara 2019). Additionally, Sagart et al. (2019) placed the origin of Sino-Tibetan as north-eastern China and Sagart (2019) reconstructed two word forms of 'pig'. This means that the word forms attested in Tibetic languages are also related to the Sino-Tibetan *Urheimat*.

The pig-breeding habit in the Tibetosphere generally exists in rural, agricultural areas; it is rarely practised in city areas such as Lhasa or pastoral areas. In the eastern Tibetosphere, we frequently encounter pigs in agricultural areas, and we also find various ways of breeding them; for example, raising them in the basement or ground floor of a house, letting them roam grasslands or even a forest, and 'pig pastoralism', i.e. letting them graze under human surveillance. Pigs are raised for food, i.e. pork. Recently, there have been Tibetans who raise pigs for commercial purposes. Their business model is to sell pork with local branding. The pig also functions as one of the zodiac signs: the year of the pig.

Twelve years ago, Suzuki (2007) discussed the word forms in Tibetic languages of the eastern Tibetosphere based on limited data. This article revisits his analysis with more data (286 dialects in total) and more useful software to produce geolinguistic maps: ArcGIS online. For phonetic notation, I follow the method defined by Suzuki (2005, 2016) and Zhu (2010) for segmental description. Suprasegmentals are, however, omitted unless necessary.

<sup>&</sup>lt;sup>1</sup> The LT *phag* includes several species other than domestic pigs. Its semantic category is similar to Sus.

#### 2 Variations of 'pig'

The word form of 'pig' in Tibetic languages in the eastern Tibetosphere is mostly stable, and a form corresponding with the LT *phag* is widely employed. There are several other forms reported in small areas, and they are classified into two categories: one is a form containing a word derived from the LT *phag*, and the other is not. Map 1 displays the distribution of the word forms for 'pig'.



A: *pnag* A+G: *pnag rgan* B: */ R* Map 1: Distribution of word forms for 'pig'

Map 1 demonstrates the following:

- the geolinguistic variation on the word form for 'pig' is simple;

- most word forms are derived from *phag*, 'pig', i.e. types A and A+G. Type +G, *phag rgan*, literally denotes 'old pig'; and

- there is an exception: Type B /?a gu/ (Serpo dialect; Khromjekhog<sup>2</sup>).

Other than the features mentioned above, we find /lu lu/ (Hua and Klu-'bum-rgyal 1993) as a form in the Sogwo dialect;<sup>3</sup> however, this is not reflected in the map as there is no countercheck.<sup>4</sup> Outside of the eastern Tibetosphere in principle, we find a pig 'with a tail': *phag pa*.<sup>5</sup> The form *phag lu* is also attested in some dialects of Amdo Tibetan.

A geolinguistic analysis of Map 1 tells us that the exceptions (A+G, B) appear alone in specific varieties. I am still unsure how the exceptions were generated; however, a new form might be needed in order to distinguish an animal from a year (cf. dialects using *phag pa* 'pig' in Central Tibet and *phag lu* 'pig' in some places in Amdo). Type A+G, *phag rgan*, is used as a humilific form by pastoralists (Amdo Tibetan),<sup>6</sup> although this function is perhaps different from Type A+G and its distribution is connected with Amdo. The geographical distribution of Type B cannot be solved using a geolinguistic approach.

#### 3 Variations of 'boar'

The following description is divided into two parts: lexical variation and geographical variation.

#### Lexical variation

We find the following word forms<sup>7</sup> for 'boar' in Tibetic languages in the eastern Tibetosphere:

- P+R type (=corresponding to LT *pho phag*): there are several phonetic realisation types.
- P+WA type: a form like /p<sup>h</sup>o wa/; the first syllable is related to LT *pho* 'male'.
- R type (LT *phag* only).
- R+J type: LT phag followed by a syllable /ja?/.
- R+P type (=corresponding to LT *phag pho*).
- R+S type (=corresponding to LT *phag gseb*).
- R+L type: LT *phag* followed by a syllable /la/.
- R+T type: LT *phag* followed by a syllable /t<sup>h</sup>u?/.
- S type: LT gseb followed by a suffix /wa/.
- GL type: a form like /<sup>h</sup>gu lu/.
- R+PA type (=corresponding to LT *phag pha*).
- R+ZH type: LT *phag* followed by a syllable /za/.
- LC type: a form like /li <sup>h</sup>tcu/.
- PE+T type: a form like /p<sup>h</sup>e tut/.
- BR type: a form like /<sup>h</sup>bə-/.
- BR+L type: a form like /<sup>h</sup>bə- lə/.
- J+TS type: a form like /ja tsə/.
- JR type: a form like /ja rə/.
- P+G type: a form like /po yo/.
- PA+R type (=corresponding to LT *pha phag*).
- PE+C type: a form like  $/p^{h}e tc^{h}\tilde{o}/.$

<sup>&</sup>lt;sup>2</sup> See Suzuki (2009) for the linguistic classification.

<sup>&</sup>lt;sup>3</sup> It is not always appropriate to designate a variety of Amdo by using a toponym (Tsering Samdrup & Suzuki 2017). Here, I follow the original description.

<sup>&</sup>lt;sup>4</sup> I have some data on the Sogwo dialect in which 'pig' corresponds to the LT phag.

<sup>&</sup>lt;sup>5</sup> Tibetan dictionaries, such as Jäschke (1881:339) and Zhang ed. (1985:1699–1700), describe *phag* as a name of the year and a morpheme denoting 'pig' in compounds, and *phag pa* as the animal 'pig'. If we follow the definition of this description, the use of the LT *phag* as 'pig' does not correspond to the literary meaning. However, my classification is not based on the meaning of LT, but the sound correspondence with LT.

<sup>&</sup>lt;sup>6</sup> See Tsering Samdrup and Suzuki (2019) for humilifics in Amdo Tibetan (Mabzhi Amdo).

<sup>&</sup>lt;sup>7</sup> In the following listed types, 'R' denotes 'root', designating the LT *phag*.

- PE+R type: forms /p<sup>h</sup>e/ followed by LT *phag*.
- PJE+P type: forms /p<sup>h</sup>je/ followed by LT *pho*.
- PJE+T type: a form like /p<sup>h</sup>je tx/.
- R+B type: a form like  $/p^{h}a?^{h}b$ ?/.
- R+D type: LT *phag* followed by a syllable /dɔ̃/.
- R+G type: LT *phag* followed by a syllable /gə/.
- R+R+C type: LT *phag* followed by /rə <sup>h</sup>tci/; the first syllable might be a class term.<sup>8</sup>
- R+TR type: LT *phag* followed by a syllable /t<sup>h</sup>o?/.
- R+TS type: LT *phag* followed by a syllable /ts<sup>h</sup>ə/.
- R+W type: LT *phag* followed by a syllable /wo/.
- S+R type (=corresponding to LT gseb phag).
- TS type: a form like /htsa htsu?/.
- TS+R type: a form like  $/ts^{h} = / followed by LT phag$ .
- TS+J type: a form like /ts<sup>h</sup>ə ja/.

Based on the morphological criterion, I classified the various types listed above into the following groups:

- A: LT-R group, including: Types P+R, R, R+P, R+PA, PA+R, PJE+P
- B: LT-S group, including: Types S, R+S, S+R
- C: R+affix group, including: Types R+J, R+L, R+T, R+ZH, PA+R, PE+R, PE+T, PJE+T, PE+C, R+B, R+D, R+G, R+R+C, R+TR, R+W
- D: TS group, including: Types R+TS, TS, TS+R, TS+J
- E: others

Each of the types and groups above is referred to in Maps 2 and 3, respectively.

#### Geographical variation

Map 2 displays a distribution of the types mentioned above, only based on my descriptions.

<sup>&</sup>lt;sup>8</sup> The term 'class term' denotes, for example, the LT *bya* 'bird' used as a part of the words for birds such as *bya de bo* 'rooster' and *bya khrung* 'crane'. It functions to indicate a category (here 'bird') despite the words *de bo* and *khrung* possessing their meaning as specific species of birds ('rooster' and 'crane', respectively). See Tournadre and Suzuki (forthcoming) for details.



Map 2: Distribution of word forms for 'boar' (following types of word form)

Some general observations of Map 2 are as follows:

- the form corresponding to LT (P+R=*pho phag*; black dot) is broadly attested regardless of the geography;

- high lexical variation is found in the easternmost and southernmost areas;

- some dialects just use *phag*, a simple form (R);

- the main difference is in the inclusion of the morpheme *phag*;

- in the easternmost area, there are forms including a /s/-initial (probably derived from *gseb* 'stallion; uncastrated' or *sos* 'breed'); and

- there are forms with a /l/-initial, of an unclear origin, in both the easternmost and southernmost areas.

A geolinguistic analysis of Map 2 tells us that lexical variation is prominent in the southernmost and easternmost areas where pig-breeding is practised extensively. I will focus on several word forms that are indicated by the black symbols. There are two literary forms, *pho phag* (P+R) and *phag pho* (R+P), of which the former appears more widely. Another form corresponding to *phag pha* 'pig-father' (R+PA) is potentially analysed due to the morphological analogy parallel to *phag ma* 'sow'; see Suzuki (2019) for details.

To summarise the word forms, I classified them into five groups (A to E), as shown on Map 3.



Map 3: Distribution of word forms for 'boar' (classified)

Map 3 still shows the peculiarity of the word forms in the easternmost and southernmost areas, where Tibetans indeed raise pigs. One can notice that Groups C and E are both distributed over a wide region of the Tibetosphere of Yunnan. As Group E is a collection of various minor word forms, Group C is significant as an areal lexical feature. It can be interpreted that Group C is a new form and has expanded

from the rGyalthang area. It has been considered as the centre of this region (see Suzuki 2018; Wang 1995); hence, this interpretation is valid from the geolinguistic viewpoint. Group E, although a collection of various word forms, is principally located in the outer area of Group C; hence, it could reflect archaic forms.

#### 4 Variations of 'sow'

The following description is divided into two parts: lexical variation and geographical variation.

#### Lexical variation

We find the following word forms for 'sow' in Tibetic languages in the eastern Tibetosphere:

- R+MA type (=corresponding to LT *phag ma*): there are several types of phonetic realisation.
- M+R type (=corresponding to LT *mo phag*): there are several types of phonetic realisation.
- R+M type (=corresponding to LT *phag mo*): there are several types of phonetic realisation.
- J+M type: forms such as /ji ma/.
- R+MA/M+WA type: there is a subclassification on 'sow': R+MA type for 'sow with piglets' and M+WA type for 'sow without piglets'.
- M+WA type: forms such as /mo wa/.
- P+J+M type: forms analysed as LT *phag* followed by /ji ma/. The first *phag* might function as a class term which categorises 'pig'.
- R+P+M type: forms analysed as LT *phag* followed by LT *phag mo*. The first *phag* might function as a class term which categorises 'pig'.
- J+TR type: a form like /je t<sup>h</sup>o?/.
- JA+M type: a form like /ja mo/.
- MA+R type (=corresponding to LT ma phag).
- R type (LT *phag* only).
- R+G type: a form like LT *phag* followed by /ga/.
- R+J+K type: forms corresponding to LT *phag* followed by /jut ku/. The first *phag* might function as a class term that categorises 'pig'.
- R+MA+M type: forms corresponding to LT *phag* + *ma* + *mo*. The second syllable *ma* might be a part of the word *phag ma*, and the third, *mo*, a feminine suffix.
- R+MA+MA type: forms corresponding to LT *phag* + *ma* + *ma*. The first *ma* might be a part of the word *phag ma*, and the second, a feminine suffix.

Based on the morphological criterion, I classified the various types listed above into the following groups:

- A: LT-R group, including: Types R+MA, M+R, MA+R, R+M, R
- B: LT-R' group, including: Types R+P+M, R+G, R+MA+M, R+MA+MA
- C: J group, including: Types J+M, P+J+M, JA+M, P+J+K
- D: others

Each of the types and groups above are referred to in Maps 4 and 5, respectively.

#### Geographical variation

Map 4 displays a distribution of the types mentioned above, based solely on my descriptions.



Map 4: Distribution of word forms for 'sow' (following types of word form)

Some general observations on Map 3 are as follows:

- forms corresponding to LT (R+MA=*phag ma*; M+R=*mo phag*; R+M=*phag mo*) are broadly attested regardless of the geography;

- high lexical variation is found in the easternmost and southernmost areas, similar to 'boar';
- most word forms are derived from phag 'pig'; and
- the main difference is in the inclusion of the morpheme phag.

A geolinguistic analysis of Map 4 tells us, like Map 2, that lexical variation is prominent in the southernmost and easternmost areas where pig-breeding is practised extensively. I will focus on several word forms that are indicated by the black symbols. There are three literary forms, *phag ma* (R+MA), *mo phag* (M+R), and *phag mo* (R+M); the first appears most often, while the second appears most widely. The forms including a 'J' syllable (J+M, J+TR, JA+M; of unclear origin) are probably related

to each other and are mainly found in the southernmost area. In some restricted areas, a semantic subclassification has occurred: 'sow without piglets' and 'sow with piglets'. In this case, *phag ma* 'pigmother' is used for the latter (Tshering Yangdron & Suzuki 2019).

To summarise the word forms, I classified them into five groups (A to D), as shown on Map 5.



Map 5: Distribution of word forms for 'sow' (classified)

The peculiarity of the word forms in the southernmost area is still visible on Map 5. Additionally, the distribution of Groups C and D displays geographical continuity. As Group D is a collection of various minor word forms, Group C is significant as an areal lexical feature. It can be interpreted that Group C has an archaic form in this area based on the same historical and social backgrounds mentioned in the interpretation of Map 3.

#### 5 Variations of 'piglet'

The following description is divided into two parts: lexical variation and geographical variation.

#### Lexical variation

We can find the following word forms for 'piglet' in Tibetic languages in the eastern Tibetosphere:

- R+PR type (=corresponding to LT *phag phrug*): there are several types of phonetic realisation.
- PJE type: forms with a glide such as /p<sup>h</sup>je/.
- R+L type: a form like LT *phag* followed by /la, li, lə, lu/.
- PU type (=mostly corresponding to LT *phag gu*): including monosyllabic and disyllabic forms.
- PJE+G type: a form like  $/p^{h}je/$  followed by LT diminutive gu or 'u.
- P+K type: a form like  $/p^{h}u k^{h}a/$ .
- P+W+TS type: a form like  $/p^ha$ ? wo tsə/.
- PJE+J type: a form like /p<sup>h</sup>je ji:/
- R+GU type: a form like LT *phag* followed by /ɣu/.
- PE type: forms without a glide such as /p<sup>h</sup>e/.
- PE+J type: a form like /p<sup>h</sup>e ja/.
- R+G type: a form like LT *phag* followed by /ga/.
- AG type: a form like /?a gu/.
- PE+PR type: a form like /p<sup>h</sup>e/ followed by LT *phrug*.
- PJE+L type: a form like /p<sup>h</sup>je/ followed by /li, lə/.
- R+KR type: a form like LT *phag* followed by /t<sup>h</sup>a?/.
- PJE+PR type: a form like  $/p^{h}$  je/ followed by LT *phrug*.
- R+CC type: a form like LT *phag* followed by /tc<sup>h</sup>ə/.
- SH+R type: a form like /cə/ followed by LT phag.
- AM type: a form like /?a mu/.
- ANG type: a form like /?a ŋu/.
- E+PR type: a form like /?e/ followed by LT phrug.
- GD type: a form like /go di/ (for 'male piglet').
- PW type: a form like /p<sup>h</sup>ow yu/.
- R+CK type: a form like LT phag followed by /tcə <sup>fi</sup>gə/.
- R+GE type: a form like  $/p^{h}i$  gɛ/.
- R+J type: a form like LT phag followed by /ji:/.
- R+MM type: a form like LT *phag* followed by /me me/ 'small'.
- R+RU type: a form like LT *phag* followed by /ru?/.
- R+TI type: a form like LT *phag* followed by /<sup>h</sup>ti?/.
- R+TR type: a form like LT phag followed by / thi?/.
- R+TSK type: a form like LT phag followed by /tsə ke/ 'small'.
- R+W type: a form like LT phag followed by /wo/.

Based on the morphological criterion, I classified the various types listed above into the following groups:

- A: LT group, including: Type R+PR only
- B: LT-diminutive group, including: Types PJE, R+L, PU, PJE+G, R+GU, R+G, PE, PE+PR, PJE+L, PJE+J, PJE+G, PJE+PR
- C: PR group, including: Types R+PR, PE+PR, PJE+PR, E+PR, R+RU
- D: R+adjective group: Types R+MM, R+TSK, R+CK
- E: others

Each of the types and groups above are referred to in Maps 6 and 7, respectively.

#### Geographical variation

Map 6 displays a distribution of the types mentioned above, only based on my descriptions.



Map 6: Distribution of word forms for 'piglet' (following types of word form)

Some general observations on Map 6 are as follows:

- the form corresponding to LT (R+PR=*phag phrug*; black dot) is broadly attested regardless of the geography; and

- high lexical variation is found in the easternmost and southernmost areas, similar to 'boar' and 'sow'; common forms are used in both of these areas.

A geolinguistic analysis of Map 6 tells us that the use of the literary form is attested in the areas where Tibetans are less interested in pig-breeding; conversely, in the pig-breeding areas, particular words for 'piglet' are found. I will focus on several word forms that are indicated by the black symbols. Two greater types can be found, K-type (R+G, R+GU, P+K, PU, PJE+J, PE, etc.) and L-type (R+L, PJE+L); the former is derived from + iu, and the latter from + le, both of which are LT diminutive

markers. In some north-eastern areas, a semantic subclassification of gender has occurred for 'male piglet' and 'female piglet', which is not reflected on the current map.

To summarise the word forms, I classified them into five groups (A to D), as shown on Map 7.



Map 7: Distribution of word forms for 'piglet' (classified)

Map 7 shows that the peculiarity of the word forms is distributed along the borderline of the eastern Tibetosphere like a chain. Interestingly, the distribution pattern differs between the easternmost and southernmost areas. The majority of dialects in the easternmost area use Group E (a collection of various minor forms), whereas those in the southernmost area use Group B. Since the morphology of Group B is also attested in LT, its scattered distribution (easternmost, eastern, and southernmost) potentially originates from a LT word formation.

#### 6 Conclusion

We can find a rich lexical set of 'pigs' in the easternmost and southernmost areas, where pig-breeding is widely practised. It has been considered that Tibetans' lexicon on pigs is monotonous, corresponding to the literary words. However, the situation in the eastern Tibetosphere implies that enriching the vocabulary depends on the necessity of classifying something. This relationship is simply predicted due to the potential need for lexical differences; no substantial evidence has been provided, but this is still the prevailing view. Although biologically Tibetan pigs originated in the Tibet Plateau, without their extensive breeding the variation of word forms for 'pig' would not have grown.

The lexical variation appears not only in derivation but also in the utilisation of other stems than *phag*, such as *gseb*, /li/, /ja/, and /ji/. Additionally, we also find the usage of the syllable corresponding to the LT *phag* as a class term.

#### Acknowledgements

The present study is funded by three Grants-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. JP16H02722), 'Investigation of Undocumented Languages in the Eastern Tibetosphere and their Geolinguistic Research' (headed by the present author, No. JP17H04774), and 'Geolinguistic Studies of China and Adjacent Multilingual Areas Using High-resolution and Wide-area Maps' (headed by Mitsuaki Endo, No. JP18H00670).

#### References

'Brug-mo-mtsho. 2002. Maqu Zangyu Yanjiu. Minzu Chubanshe.

- Ebihara, Sh. 2019. Amudo-Tibetttogo bunpoo. Hituzi Syoboo.
- Hua, K. & Klu-'bum-rgyal. 1993. Anduo Zangyu Kouyu Cidian. Gansu Minzu Chubanshe.
- Jäschke, H. A. 1881. A Tibetan-English dictionary with special reference to the prevailing dialects. Kegan Paul.
- Roche, G. & H. Suzuki. 2017. Mapping the Linguistic Minorities of the Eastern Tibetosphere. *Studies* in Asian Geolinguistics VI—"Means to Count Nouns" in Asian Languages—, 28-42. https://publication.aa-ken.jp/sag6\_count\_2017.pdf
- Roche, G. & H. Suzuki. 2018. Tibet's minority languages: Diversity and endangerment. *Modern Asian Studies* 52(4), 1227-1278. doi: 10.1017/S0026749X1600072X
- Sagart, L. 2019. *The position of Chinese in Sino-Tibetan, and of Sino-Tibetan in East Asia*. Keynote speech at the 27<sup>th</sup> Annual Conference of International Association of Chinese Linguistics (Kobe)
- Sagart, L., G. Jacques, Y. Lai, R. J. Ryder, V. Thouzeau, S. J. Greenhill & J.-M, List. 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences* 116(21): 10317-10322. doi: 10.1073/pnas.1817972116
- Shao, M. 2018. Hexi Zoulang Binwei Zangyu Dongnahua Yanjiu. Zhongshan Daxue Chubanshe.
- Sung, K.-M. & IHa-byams-rgyal. 2005. *Colloquial Amdo Tibetan: A complete guide for adult English speakers*. Zhongguo Zangxue Chubanshe.
- Suzuki, H. 2005. Tibettogo onsetu koozoo no kenkyuu. *Journal of Asian and African Studies* 69: 1-23. http://hdl.handle.net/10108/20212
- Suzuki, H. 2007. Senseiminzokusooroo Tibettogo hoogen ni okeru 'buta' wo arawasu go. *Kyoto* University Linguistic Research 26: 31-57. <u>https://doi.org/10.14989/57308</u>
- Suzuki, H. 2009. Introduction to the method of the Tibetan linguistic geography——a case study in the Ethnic Corridor of West Sichuan——. In Yasuhiko Nagano (ed.) *Linguistic Substratum in*

*Tibet—New Perspective towards Historical Methodology (No. 16102001) Report* Vol.3, 15-34. National Museum of Ethnology. <u>http://hdl.handle.net/10502/4341</u>

- Suzuki, H. 2016. In defense of prepalatal non-fricative sounds and symbols: towards the Tibetan dialectology. *Researches in Asian Languages* 10: 99-125. <u>http://id.nii.ac.jp/1085/00002195/</u>
- Suzuki, H. 2018. *100 Linguistic Maps of the Swadesh Word List of Tibetic Languages from Yunnan*. Research Institute for Languages and Cultures of Asia and Africa. <u>https://publication.aa-ken.jp/sag\_mono3\_tibet\_yunnan\_2018.pdf</u>
- Suzuki, H. 2019. *Tibetic terms on 'sow' and 'boar' in the eastern Tibetosphere: (Mis-)interpretation and analogy around 'pig-mother' and 'pig-father'*. Paper presented at 1st Annual Meeting of Japan Geolinguistics Society (Tokyo)
- 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.
- Tournadre, N. & H. Suzuki. forthcoming. *The Tibetic Languages: An Introduction to the Family of Languages Derived from Old Tibetan* (with the collaboration of Xavier Becker and Alain Brucelles for the cartography)
- Tsering Samdrup & 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. <u>https://publication.aa-ken.jp/sag7\_tone\_2017.pdf</u>
- Tsering Samdrup & H. Suzuki. 2019. Humilifics in Mabzhi pastoralist speech of Amdo Tibetan. *Linguistics of the Tibeto-Burman Area* 42.2, in press.
- Tshering Yangdron & H. Suzuki. 2019. *Dongwang Zangyu yuyan ditu ji qi yu zhoubian tuhua de guanxi*. Paper presented at 2019 salon of geolinguistics on China's minority languages (Beijing).
- Wang, H. 1995. Diqing Zangzu Shehuishi. Zhongguo Zangxue Chubanshe.
- Yang, Sh., H. Zhang, H. Mao, D. Yan, Sh. Lu, L. Lian, G. Zhao, Y. Yan, W. Deng, X. Shi, Sh. Han, Sh. Li, X. Wang & X. Gou. 2011. The local origin of the Tibetan pig and additional insights into the origin of Asian pigs. *PLoS ONE* 6(12): e28215. doi: 10.1371/journal.pone.0028215
- Zhang, Y. ed. 1985. Bod-rgya tshig-mdzod chen-mo. Minzu Chubanshe.
- Zhu, X. 2010. Yuyinxue. Shangwu Yinshuguan.

## Applied Phylolinguistics in Cultural Comparative Studies – A Case Study of Sino-Tibetan Kinship Evolution

Ting Ji<sup>*a*</sup>, Hanzhi Zhang<sup>*b,c*</sup>, Thomas E. Currie<sup>*d*</sup>, Andrew Meade<sup>*e*</sup>, Mark Pagel<sup>*e*,*f*</sup>, and Ruth Mace<sup>*b,c*\*</sup>

<sup>a</sup> Institute of Zoology, Chinese Academy of Sciences

<sup>b</sup> School of Life Sciences, Lanzhou University

<sup>c</sup> Department of Anthropology, University College London

<sup>d</sup> Department of Biosciences, University of Exeter

<sup>e</sup> School of Biological Sciences, University of Reading

<sup>f</sup>Santa Fe Institute

#### Abstract

It is widely believed in China that matrilineal kinship was the ancestral human kinship system, and patrilineal systems evolved more recently. Genetic and archaeological evidence is mixed on whether matrilineal system is ancient or recent in the evolution of the Sino-Tibetans. Here we use cultural phylogenetics to address this question. We construct a phylogenetic tree of Sino-Tibetan languages using Bayesian MCMC inferences, and then use a variety of ethnographic and historical sources to identify dispersal strategies across cultures. We compare the likelihood of different ancestral states of kinship systems, based on the current distribution of kinship descent and residence norms across cultures. We find the ancestral states at the root of the Sino-Tibetan cultural phylogeny reflect male-centred kinship (i.e. female dispersal and patrilineal descent). Our results suggest that female-centred kinship is relatively recent among Sino-Tibetan groups.

#### 1 Introduction

One of the most remarkable features of human societies is the extraordinary extent of cultural diversity manifested in the form of stable and consistent behavioural variation among ethnolinguistic groups. Major differences exist in marital system, subsistence, political organisation, and other social institutions among cultures. Our study compares the likelihood of different ancestral states of kinship systems among Sino-Tibetan populations, using methods that were originally developed by evolutionary biology to study mechanisms that generate biological diversity.

#### 2 Background

By mapping ethnographic data on kinship systems onto linguistic phylogenetic trees, phylogenetic comparative methods can be used to make inferences about the pattern of cultural evolution (Mace and Holden 2005), including ancestral states and patterns of historical change in kinship along the branches of the tree. It has been showed that Austronesian kinship traits - including single-community patrilines, kin terminology, matrilineal descent, unilineal descent, brideprice, and cousin marriage - are best predicted by nearest phylogenetic neighbour instead of nearest geographical neighbour (Mace and

Jordan 2011). Correlation analyses also found that family and kinship norms among ethnolinguistic groups in Africa are best predicted by population history (Guglielmino et al. 1995; Hewlett, DeSilvestri, and Guglielmino 2002). Indeed, kinship norms can be stabilized even in the presence of intergroup marriage, as immigrants tend to adopt norms of the new group (Richerson and Boyd 2005); hence they generally show strong phylogenetic signal. Phylogenetic comparative methods have been applied in many cultural regions to infer the evolution of kinship systems including Bantu, Indo-European, and Austronesian populations (Opie et al. 2014; Fortunato 2011; Jordan et al. 2009). Nevertheless, cultural practises in different regions may also evolve in different trajectories. A recent phylogenetic study (Moravec et al. 2018) showed that kinship systems in five language families (Austronesian, Bantu, Indo-European, Pama-Nyungan and Uto-Aztecan) evolved in different directions and speeds. To this date, there is no cultural comparative study that examines Sino-Tibetan kinship evolution with the evolutionary paradigm explicitly. Here we use cultural phylogenetics to address this question.

#### 3 Methods

#### 3.1 Phylogenetic Reconstruction

We use Bayesian Markov-chain Monte Carlo inferences (Huelsenbeck, Rannala, and Masly 2000; Pagel and Meade 2004) that produce a large sample of possible phylogenies to infer the ancestral relationships between languages, weighted according to their likelihood. Cognate data for basic vocabulary in Sino-Tibetan languages are taken from the Tower of Babel project (Peiros and Starostin 1996).

#### 3.2 Ethnographic and Historical data

Written records on ancestral kinship systems were found in imperial annals, envoys' reports and administrative correspondence from 220 B.C. to the 18th century. Information from these historical records were used to inform phylogenetic reconstructions of ancestral nodes, in addition to the present-day kinship status extracted from Murdock's Ethnographic Atlas (Murdock 1967; Gray 1999), government censuses, and contemporary ethnographies (Wang 1994; Academia-Sinica 2009; Yang and Ding 2003).

#### 3.3 Comparative analyses of residence and descent

We inferred the ancestral states of residence, descent and rates of transitions between two states using Bayesian RJMCMC methods as implemented in BayesTraits (www.evolution.reading.ac.uk) (Pagel and Meade 2004) which accounts for uncertainty in the phylogeny by integrating models over the posterior. We employed the "most-recent-common-ancestor" approach (Pagel, Meade, and Barker 2004) to find the node in each tree in the posterior sample that contains a given set of descendant taxa, and calculates the ancestral value at that node. Internal nodes can be fixed as either state where historical records give independent information about the ancestral state, by applying the "Fossil" command in BayesTraits v3. This function is also a useful way to test if one state receives better support than the other at various ancestral nodes. Support for different RJMCMC models were compared using log Bayes Factor (log BF) (Pagel and Meade 2006).

#### 4 Finding

We find no evidence to support the widespread belief that ancestral Sino-Tibetans had female-centred kinship, i.e. matrilineal descent or matrilocal residence. Switches from male-centred kinship systems to female-centred systems are relatively recent. Furthermore, no uni-directional model of kinship evolution received support in our analyses, suggesting changes in kinship norms in both directions occurred during the evolution of the Sino-Tibetans. Our methods take into account both the uncertainty in the tree and in the patterns of cultural evolution on the tree and use all the results to work out the likelihood of the various possible evolutionary pathways.

#### 5 Conclusion

Here we use cultural phylogenetics to investigate the ancestral state of Sino-Tibetan kinship system. We construct a phylogenetic tree of Sino-Tibetan languages using Bayesian MCMC inferences, identify current distribution of kinship systems across ethnolinguistic groups, and compare the likelihood of different ancestral states of kinship systems, based on the. We find the ancestral states at the root of the Sino-Tibetan cultural phylogeny reflect male-centred kinship (i.e. female dispersal and patrilineal descent).

#### References

- Academia-Sinica. 2009. "中研院歴史語言研究所集刊論文類編: 民族與社會編 A Collection of Acaememia Sinica History and Language Studies Theses 1930-2000: Ethnic and Societies Volume." In, edited by Academia Sinica. 中華書局 Zhonghua Publishing House.
- Fortunato, L. 2011. 'Reconstructing the history of marriage and residence strategies in Indo-European-speaking societies', *Hum Biol*, 83: 129-35.
- Gray, J Patrick. 1999. 'A corrected ethnographic atlas', World Cultures, 10: 24-85.
- Guglielmino, C. R., C. Viganotti, B. Hewlett, and L. L. Cavalli-Sforza. 1995. 'Cultural variation in Africa: role of mechanisms of transmission and adaptation', *Proc Natl Acad Sci U S A*, 92: 7585-9.
- Hewlett, BarryS, Annalisa DeSilvestri, and C Rosalba Guglielmino. 2002. 'Semes and genes in Africa', *Current Anthropology*, 43: 313-21.
- Huelsenbeck, John P, Bruce Rannala, and John P Masly. 2000. 'Accommodating phylogenetic uncertainty in evolutionary studies', *Science*, 288: 2349-50.
- Jordan, F. M., R. D. Gray, S. J. Greenhill, and R. Mace. 2009. 'Matrilocal residence is ancestral in Austronesian societies', *Proc Biol Sci*, 276: 1957-64.
- Mace, R., and C. J. Holden. 2005. 'A phylogenetic approach to cultural evolution', Trends Ecol Evol, 20: 116-21.
- Mace, R., and F. M. Jordan. 2011. 'Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation', *Philos Trans R Soc Lond B Biol Sci*, 366: 402-11.
- Moravec, Jiří C, Quentin Atkinson, Claire Bowern, Simon J Greenhill, Fiona M Jordan, Robert M Ross, Russell Gray, Stephen Marsland, and Murray P Cox. 2018. 'Post-Marital Residence Patterns Show Lineage-Specific Evolution', *Evolution and Human Behavior*.
- Murdock, George Peter. 1967. 'Ethnographic atlas'.
- Opie, C., S. Shultz, Q. D. Atkinson, T. Currie, and R. Mace. 2014. 'Phylogenetic reconstruction of Bantu kinship challenges Main Sequence Theory of human social evolution', *Proc Natl Acad Sci U S A*, 111: 17414-9.
- Pagel, M., and A. Meade. 2004. 'A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data', *Syst Biol*, 53: 571-81.
- Pagel, M., A. Meade, and D. Barker. 2004. 'Bayesian estimation of ancestral character states on phylogenies', *Syst Biol*, 53: 673-84.
- Pagel, Mark, and Andrew Meade. 2006. 'Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo', *The American Naturalist*, 167: 808-25.
- Peiros, Ilia, and Sergej Starostin. 1996. A comparative vocabulary of five Sino-Tibetan languages (The University of Melbourne, Department of Linguistics and Applied Linguistics: Parkville, Australia).
- Richerson, Peter J, and Robert Boyd. 2005. "Not by genes alone." In.: Chicago: University of Chicago Press.
- Wang, ZH. 1994. 'History of nationalities in China', China Social Science Press, Beijing (in Chinese).
- Yang, Shengmin., and Hong. Ding. 2003. *The Study of Chinese Ethnography* 中国民族志 (Beijing: Central University for Nationalities Press).

## Contributors

Mitsuaki ENDO	Aoyama Gakuin University
Ray IWATA	Komatsu University
Takashi UEYA	Kyoto University of Foreign Studies
Fumiki SUZUKI	Nanzan University
Kazuo MIYAMOTO	Kyushu University
Kenji YAGI	Kokushikan University
Hiroyuki SUZUKI	University of Oslo / National Museum of Ethnology
Ting JI	Institute of Zoology, CAS
Hanzhi ZHANG	Lanzhou University / University College London
Thomas E. CURRIE	University of Exeter
Andrew MEADE	University of Reading
Mark PAGEL	University of Reading / Santa Fe Institute
Ruth MACE	Lanzhou University / University College London