

ISSN 2436-6471

地理言語学研究  
Studies in Geolinguistics

3



日本地理言語学会  
Geolinguistic Society of Japan  
2023

ISSN 2436-6471

地 理 言 語 学 研 究

第3号

*Studies in Geolinguistics*

No. 3

2023

日本地理言語学会編集委員会 | Editorial board of Geolinguistic Society of Japan

コーディネーター | Coordinator  
斎藤純男 (Yoshio SAITÔ)

委員 | Editors

鈴木博之 (Hiroyuki SUZUKI)  
岩田 礼 (Ray IWATA)  
白井聡子 (Satoko SHIRAI)

2023年10月13日発行

doi: <https://doi.org/10.5281/zenodo.8437140>

© 2023 Individual contributors

編集・発行 日本地理言語学会  
学会ウェブ <https://geolinguistics.sakura.ne.jp/>  
事務局 東京都渋谷区渋谷4丁目4-25青山学院大学G1305研究室内

## 目 次

### 論文/Research articles

HEERINGA, Wilbert and Fumio INOUE

Exploring the Japanese dialect geography dialectometrically:

Division and continuity ..... 1

井上史雄

外行語の計量地理言語学 Google trends データの因子分析と初出年..... 45

SUZUKI, Hiroyuki

Geolinguistic analysis of word forms for ‘today’ in Tibetic languages

in Yunnan ..... 92

NAKAO, Shuichiro

‘Rice,’ ‘sorghum,’ ‘pearl millet’ and ‘finger millet’ in Nilo-Saharan ..... 99

半沢康

東北地方における当為表現の変化..... 106

TRẦN, Thị Hồng Hạnh

A further study of kinship terms for ‘mother’ in Vietnamese ..... 137

特集「シナ=チベット語族への地理言語学的アプローチ」

SHIRAI, Satoko

Geolinguistics and Sino-Tibetan: An introduction ..... 146

EBIHARA, Shiho and Yoshio SAITÔ

Language contact between Tibetic and Mongolic, with a special focus

on the Amdo Linguistic Area ..... 152

SUZUKI, Hiroyuki, Kenji YAGI, and Fumiki SUZUKI

Geolinguistic approach to Sino-Tibetan: Lexical relationship

in some animal and plant terms ..... 164

IWASA, Kazue

Current issues and perspectives on Yi characters:

From the past and into the future ..... 180

彙報：2022年度..... 199



## Contents

### Research articles

HEERINGA, Wilbert and Fumio INOUE

Exploring the Japanese dialect geography dialectometrically:

Division and continuity ..... 1

INOUE, Fumio

Computational geolinguistics of lendword:

Factor analysis of Google Trends data and year of first appearance ..... 45

SUZUKI, Hiroyuki

Geolinguistic analysis of word forms for ‘today’ in Tibetic languages

in Yunnan ..... 92

NAKAO, Shuichiro

‘Rice,’ ‘sorghum,’ ‘pearl millet’ and ‘finger millet’ in Nilo-Saharan ..... 99

HANZAWA, Yasushi

Changes in deontic expression in the Tohoku region ..... 106

TRẦN, Thị Hồng Hạnh

A further study of kinship terms for ‘mother’ in Vietnamese ..... 137

### FEATURED THEME: Geolinguistic approach to Sino-Tibetan

SHIRAI, Satoko

Geolinguistics and Sino-Tibetan: An introduction ..... 146

EBIHARA, Shiho and Yoshio SAITÔ

Language contact between Tibetic and Mongolic, with a special focus

on the Amdo Linguistic Area ..... 152

SUZUKI, Hiroyuki, Kenji YAGI, and Fumiki SUZUKI

Geolinguistic approach to Sino-Tibetan: Lexical relationship

in some animal and plant terms ..... 164

IWASA, Kazue

Current issues and perspectives on Yi characters:

From the past and into the future ..... 180

Memoranda and personalia 2022 ..... 199

Research article

## Exploring the Japanese dialect geography dialectometrically: Division and continuity

HEERINGA, Wilbert  
Fryske Akademy

INOUE, Fumio  
Tokyo University of Foreign Studies

**Abstract:** In this paper we explore the Japanese dialect by applying PMI Levenshtein distance to 2400 localities and 37 items of the Linguistic Atlas of Japan Database (LAJDB). Ward's clustering and t-distributed stochastic neighbor embedding (t-SNE) were applied to the average Levenshtein distances among the 2400 localities. Using these techniques we obtained an area map which divides the dialect landscape in areas, and a RGB map that visualized the dialect landscape as a dialect continuum. On the basis of this huge data set we are able to generate very detailed results. Using this methodology we do not need to make subjective choices but provide objective results. We found five optimal groups representing the Tohoku dialects, Eastern dialect, Western dialect, Kyushu dialect and Ryukyuan dialect. We especially focused on the the Ryukyuan varieties and found a division in three groups: Amami dialects, Okinawan dialects and Southern Ryukyuan dialects. Both on the global level and on a more detailed level the results mainly confirmed results from earlier studies.\*

**Keywords:** dialectometry; Japanese dialects; Japanese dialectology; dialect classification; dialect areas

### 1. Introduction

A “dialect landscape”<sup>1</sup> refers to the geographical distribution and variation of dialects within a particular region. There are several reasons why people may want to visualize the dialect landscape. First, visualizing this diversity helps linguists, researchers, and the general public understand the linguistic richness of their area. Second, visualizations of dialect landscapes can reveal historical migrations, internal culture divisions, trade routes, and interactions that have shaped linguistic variation. Third, governments and educational institutions may use dialect visualizations to make

---

HEERINGA, Wilbert & Fumio INOUE. 2023. Exploring the Japanese dialect geography dialectometrically: Division and continuity. *Studies in Geolinguistics* 3: 1–44. doi: <https://doi.org/10.5281/zenodo.8437044>

\* The first author thanks Prof. Dr. Takuichiro Onishi (NINJAL) for introducing him to the Linguistic Atlas of Japan, and Ph.D. candidate Takamasa Seimiya for pointing him to the website that contains digitized data of 38 items from the LAJDB.

<sup>1</sup> In this paper the term 'dialect landscape' refers to the distribution and variation of dialects or regional language varieties within a particular geographic area. We use the term in the same way as Grosse (1955) used the term 'Sprachlandschaft' and Ivić (1962) used the term 'linguistic landscape'.

informed decisions about language policy, curriculum development, and language education. Fourth, knowledge of the dialect landscape may help broadcasters to tailor their broadcasts more precisely to their audience.

For a long time the primary tool for revealing dialect areas within the dialect landscape was the isogloss. An isogloss is a line on a map dividing areas whose dialects differ in some specific respect (Matthews 1997). Using the isogloss method, isoglosses of different phenomena are drawn on a map. Isogloss bundles represent dialect boundaries. See also Inoue (1996, pp. 72–74).

However, isoglosses do not always coincide, but may run parallel or may cross each other; by selecting only coinciding isoglosses the method becomes subjective. A problem mentioned by Kessler (1995) is that isoglosses do not neatly bisect the language area. Often variants do not neatly line up on two sides of a line, but are intermixed to some degree. Another problem which Kessler pointed out is the fact that in case of a dialect continuum with very gradual changes, it seems arbitrary to draw major dialect boundaries between two villages with very similar speech patterns.

The problems that are identified with the isogloss method are eliminated when we use so-called dialectometric methods. Dialectometry means literally ‘the measure of dialect’. This term was coined by Jean Séguy (Chambers & Trudgill 1998: p. 137). He and his associates published the *Atlas linguistique de la Gascogne* that consists of six atlas volumes (Séguy 1973). In these volumes maps are included in which single answers were plotted (Chambers & Trudgill 1998: p. 137). However, Séguy looked for a way to analyze the maps in a more objective way than was possible with traditional analytic methods. For each pair of contiguous sites Séguy and his research team counted “the number of items on which the neighbors disagreed.” The number of disagreements between two neighbors was expressed as a percentage, “and the percentage was treated as an index score indicating the linguistic distance between any two places” (Chambers & Trudgill 1998: p. 138).

The work of Hans Goebel and Edgar Haimlerl is strongly related to the work of Séguy, although independently developed. Instead of distance they measure similarity between two dialects, calculated as the number of items on which two dialects agree, expressed in a percentage (Goebel 1982). They refer to their measure as Relative Identity Value (RIV) (Goebel 2006; 2010).

Similar to Goebel, Kumagai (2016) measured similarity between Japanese dialects using data from the Linguistic Atlas of Japan Database (LAJDB). He calls this measure NC and describes this as the “number of common linguistic features.”

Although the methodology of Séguy, Goebel and Kumagai can be applied to any linguistic level, it is complicated to apply this directly to phonetic transcriptions. After all, several dialect atlases such as the Linguistic Atlas of Japan provide phonetic

transcriptions. A solution was introduced by Kessler in 1995. He measured linguistic distances among Irish Gaelic dialects using Levenshtein distance. The basic idea of this algorithm is that it calculates the cost of changing one string of characters into another. The algorithm was introduced by Vladimir Levenshtein in 1965.

The Levenshtein distance was applied by Huisman et al. (2019) and Huisman (2021) to Japanese dialects, using a newly compiled “comparative dataset based on the 100-item Swadesh List” (Swadesh, 1952). This list includes basic concepts, such as body parts and everyday actions.

In this paper, we research the Japanese dialect landscape using the best of both worlds. Like Kumagai we use the LAJDB, a large data set that includes 2400 locations in Japan, and like Huisman et al. (2019) and Huisman (2021) we will use the Levenshtein distance, supplemented by statistical techniques like cluster analysis and multidimensional scaling. To the best of our knowledge, Levenshtein distance has never been applied to the LAJDB before. While Huisman et al. (2019) and Huisman (2021) used classical Levenshtein distance, we follow Wieling et al. (2009) and Wieling (2012) by using PMI Levenshtein distance.

Additionally, we will especially focus on the Okinawa prefecture which includes the islands southwest of the mainland Japanese area. It consists of several hundred islands that stretch over a distance of 1200 kilometers between Kyushu and Taiwan.

Our paper is structured as follows. In Section 2 we describe the LAJDB, the Levenshtein distance and several statistical techniques for visualizing the dialectal variation. In Section 3 we apply the methodology to Japanese dialects. First, we present the results from all over Japan, and then we zoom in on the Ryukyuan languages. We end with discussion and conclusions in Section 4.

For measuring Levenshtein distances and visualizing the measurements we use LED-A throughout this paper. LED-A is an online web application for calculating linguistic distances with Levenshtein distance and visualizing them. The app is freely available at <https://www.led-a.org>.

## 2. Methods

### 2.1. Data source

The description of the data source in this section is based on Kawaguchi & Inoue (2002), Onishi (2002) and Kumagai (2016).

The Linguistic Atlas of Japan (LAJ) is based on a survey that was conducted from 1957 to 1965 by the National Language Research Institute (NLRI) which was the predecessor of today’s National Institute for Japanese Language and Linguistics (NINJAL). 2400 localities were surveyed, and 285 - mainly lexical - items per locality.

The items include nouns, verbs and adjectives. Usually “one male informant (speaker) born before 1903 and born at the given location (or, at least, who spent time there without interruption from the age of 3 to 15) was chosen as informant.” (Kumagai 2016, p. 335)

The LAJ was published from 1966 to 1974 in six volumes (Kokuritsu Kokugo Kenkyujo (NLRI) 1966–1974).

In 1999 professor Yasuo Kumagai and his research team began constructing the LAJDB (Kumagai 2013a, 2013b). The LAJDB includes both the original survey card images and the geographical distribution data of word forms. The linguistic forms shown on the maps of the LAJ are the result of the editor’s classifications of varieties recorded by the fieldworkers.

Kumagai (2016) wrote that “currently, 119 items have been completed, corresponding to 43% of the number of surveyed items (285 items).” We use data from the LAJDB which was available at:

[https://www.lajdb.org/lajdb\\_data/LAJDB\\_data\\_withCord\\_download002\\_v20180328\\_rev.html](https://www.lajdb.org/lajdb_data/LAJDB_data_withCord_download002_v20180328_rev.html). On this site, data is found for 38 items numbered as 005, 006, 007, 012, 031, 032, 036, 047, 048, 051, 052, 063, 068, 076, 081, 104, 105, 110, 111, 118, 127, 188, 190, 191, 192, 194, 195, 231, 233, 235, 237, 240, 241, 250, 264, 268, 280, and 284.

For each of the items an Excel table is provided. For our analysis we used the column 語形 ‘word form’. The transcriptions in that column represent word forms that may either lexically or phonemically vary, except for item 280 that presumably represents tonemic variation and which we excluded so that 37 items remained. The transcriptions are written in original type of Roma-ji. We made the following adaptations:

- when multiple variants were given like CINNYA(N) they were split like CINNYA / CINNYAN;
- spaces were removed;
- the following characters and character sequences were removed: =, その他, [g], [N\\\*], +, -, ', \_, [A@] ;
- in item 081 the following character sequence was removed: << 卑称 >>;
- in items 012 and 110 the following character was removed: [;
- in item 111 @ was replaced by ə;
- in item 104 we removed the following character sequences: [ti], [tui], [tsy:], [tsi:], [tsi], [zi], [t0sui], [t0si:], [t0si], [ts0wi:], [ts0wi], [t0wi], [to@:];
- ~ was replaced by / which is a separator between multiple pronunciations.

The items are listed in Table 1. Not all items were recorded in all locations. But for each of the localities transcriptions of at least 25 items are available.<sup>2</sup> Figure 1 shows the number of locations in which each item was recorded.

---

<sup>2</sup> Actually we found 2402 different locations in the item files at [https://www.lajdb.org/lajdb\\_data/LAJDB\\_data\\_withCord\\_download002\\_v20180328\\_rev.html](https://www.lajdb.org/lajdb_data/LAJDB_data_withCord_download002_v20180328_rev.html). However, we left out locations 3776.83 and 3787.35 since they have transcriptions for only 7 items. Thus 2400 locations remain which is exactly the number of locations mentioned by Kumagai (2016).

Table 1 37 items from the LAJDB

code	Japanese	English
5	かたつむり	snail
6	なめくじ	slug
7	おたまじゃくし	tadpole
12	とかげ	lizard
31	あたま	head
32	つむじ	whirlpool
36	ものもらい	sty
47	くちびる	lips
48	した (舌)	tongue
51	<塩味が>うすい	lightly salty
52	あまい (甘い)	sweet
63	おやゆび	Thumbelina
68	しもやけ (凍傷)	frostbite
76	うろこ	scales
81	おんな (女)	woman
104	おととい	day before yesterday
105	さきおととい	the day before yesterday
110	しあさって	the day after tomorrow
111	やのあさって	Yano (the day after tomorrow)
118	つゆ (梅雨)	Tsuyu (rainy season)
127	こおる (水が凍る)	Kooru (water freezes)
188	さつまいも	sweet potato
190	とうもろこし	corn
191	かぼちゃ	pumpkin
192	すみれ	violet
194	つくし	horsetail
195	すぎな	too much
231	はげる (禿げる)	bald
233	くるぶし (踝)	ankle
235	すてる (捨てる)	throw away
237	おそろしい	scary
240	ひまご (曾孫)	Himago (great-grandson)
241	やしやご (玄孫)	Yashago (great-grandson)
250	<虹が>きれいだ	the rainbow is beautiful
264	かつぐ (材木)	Katsugu (timber)
268	いる (居る)	there is
284	とんぼ	dragonfly

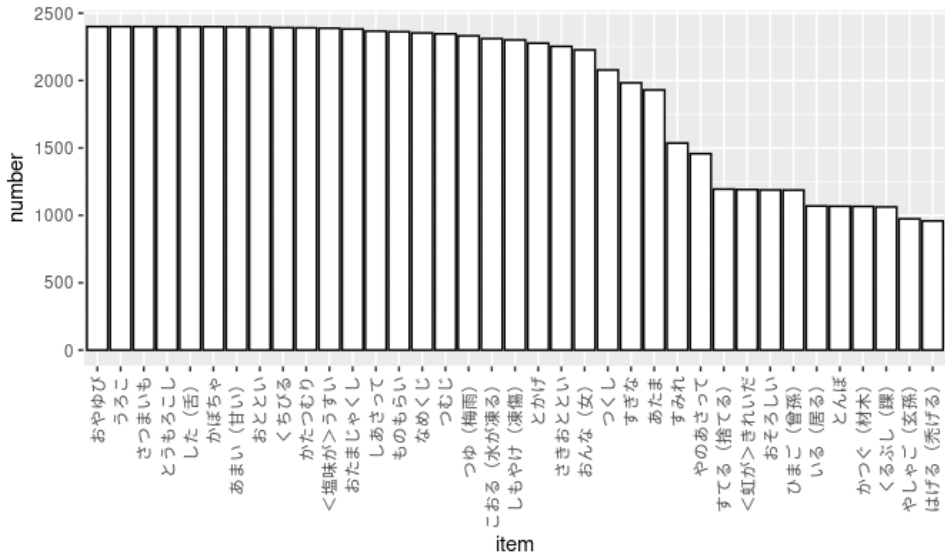


Fig. 1 For each of the 37 items the number of locations where the item has been recorded is given.

## 2.2. Levenshtein distance

The Levenshtein distance was introduced in dialectology by Kessler (1995) who used this measure for the comparison of Irish Gaelic dialects. Later it was applied to Dutch, Sardinian, Norwegian, American English, German and Bulgarian dialects, and Bantu languages by others. The algorithm calculates the cost of changing one string of characters into another. In our case the characters are phonetic segments. For example, milk is pronounced as [mɛlɐk] in the Dutch dialect of Haarlem and as [mɔlkə] in the Frisian dialect of Grouw. Now we can change the phonetic transcription of the first realization into the other as is shown in Figure 2. The alignment in Figure 2 shows that the realization of the dialect of Haarlem can be changed into the realization of the dialect of Grouw by substituting the [ɛ] by [ɔ], by deleting an [ə] and by inserting an [ə].

						total	
Haarlem	m	ɛ	l	ə	k		
Grouw	m	ɔ	l		k	ə	
actual cost		1		1		1	3
maximum cost	1	1	1	1	1	1	6
normalized cost							3/6

Fig. 2 Alignment of the realizations of milk in the local dialects of Haarlem and Grouw.



Since [mɛlək] can be changed into [mɔlkə] by three operations, the Levenshtein distance is equal to 3. Many different operation sequences map [mɛlək] → [mɔlkə], but the Levenshtein algorithm always gives the cost of the cheapest mapping.

We also calculate the cumulative maximal cost which we will use to normalize distances. The advantage of normalizing Levenshtein distances is that they can be expressed as a percentage. Percentages are more easily interpreted than ‘raw’ Levenshtein distances. Since the maximum cost of each operation is 1 in our example, and the alignment consists of six slots, the cumulative maximal cost is 6. Then the normalized Levenshtein distance is  $3/6$  is 0.5. This means that the two realizations of milk in the local dialects of Haarlem and Grouw differ from each other by 50%.

When the weights of the three operations – substitution, deletion, insertion – are the same (as in our example), it may happen that for some word pairs multiple alignments may yield the (same) minimum cost. In that case the Levenshtein distance is normalized by dividing it by the cumulative maximal cost of the longest alignment. The longest alignment has the greatest number of matches and is linguistically most plausible.<sup>3</sup>

To accord with syllabification in words, the Levenshtein distance should be based on an alignment with plausible matches. In our implementation of the algorithm the basic rule is that a vowel may only match with a vowel, and a consonant only with a consonant. However, the [w] and the [j] may also match with vowels, and the [u] and the [i] may also match with consonants. The schwa may match with a sonorant. This approach was introduced by Bolognesi & Heeringa (2002) – see also Heeringa (2004) – and was coined as VC-sensitive by Wieling et al. (2009).

Kessler (1995) distinguished between the same-word approach where the Levenshtein distance is only used for comparing word forms that are phonetic or phonological variants of each other, i.e., cognates, and the all-word approach where word forms are also compared to each other when they are lexical variants of each other. Since the word forms of an item in the LAJDB can be both phonetic/phonological variants of each other and lexical variants of each other, we use the all-word approach.

Variation among dialects is usually not measured on the basis of a single word, but on the basis of a set of words. When we compare two dialects by comparing  $n$  word pairs with Levenshtein distance, we need to calculate the aggregated distance. In Table 2 we show the procedure by means of an example in which the aggregated distance between Grouw and Haarlem is calculated on the basis of six word pairs. In this example the aggregated distance is  $2.35/6 = 0.39$  or 39%.

---

<sup>3</sup> When insertions and deletions have a weight that is the half of the weight of a substitution, all possible alignments of a word pair will have the same cumulative maximal cost. In that case any alignment can be chosen.

Table 2 Normalized Levenshtein distances are calculated between Grouw and Haarlem on the basis of six word pairs. The aggregated distance is  $2.35/6 = 0.39$  or 39%.

	Grouw	Haarlem	cost	max. cost	norm. cost
work	ʋʏrk	ʋɛrək	2	5	0.4
ship	skɪp	sxɪp	1	4	0.25
finger	fɪŋər	vɪŋər	1	5	0.2
wine	ʋɪn	ʋɛɪn	1	4	0.25
house	huz	hœys	3	5	0.75
milk	mɔlkə	mɛlək	3	6	0.5
					2.35

As we mentioned in Section 2.1 not all items were recorded in all locations. So when comparing two local dialects on the basis of 37 word pairs, it may happen that for some pairs, for one or both localities a realization is missing. Since the aggregated distance is calculated as the average of the complete word pairs, the effect of missing transcriptions is discounted.

In Section 2.1 we mentioned that multiple realizations may be given per item. We process all of them. When a location has one realization for a particular item, and another location has  $n$  realizations, the  $n$  realizations are compared to the realization of the first location, and the  $n$  Levenshtein distances are averaged. When one location has  $n1$  realizations, and the other location has  $n2$  realizations, the  $n1$  realizations are multiplied  $n2$  times, and the  $n2$  realizations are multiplied  $n1$  times. Subsequently, pairs are formed so that the average Levenshtein distance is minimized. For details about this procedure see Heeringa (2004, pp. 134–135).

In the examples above all operations bear a ‘cost’ of, for instance, 1. When we compare the realization of milk in the dialect of Haarlem with the realization of the same word in the dialect of Grouw, the substitution of the [ɛ] by [ɔ] has a cost of 1, and the substitution of the [ɛ] by [e] would cost the same, since the phonetic affinity between phones is ignored. In this paper we apply a more sensitive approach where gradual distances between segments are used as operation weights. We follow Wieling et al. (2009) and Wieling (2012) who introduced Pointwise Mutual Information (PMI) Levenshtein as a method for comparing dialects.

Using this method dialects are initially compared by using non-gradual weights, as we did in the examples above. Then new weights are found by analyzing the alignments: the more frequently segments co-occur in a slot within an alignment, the smaller the distance weight. These two steps are repeated, where in the first step the weights that were newly calculated in the second step of the previous iteration are used. The two steps are repeated until the weights do not (or hardly) change any more. PMI

refers to the degree of dependence between aligned segments.<sup>4</sup> In Figure 3 the algorithm is given in pseudocode. Wieling et al. (2009) found that alignments made by PMI Levenshtein distance improved compared to alignments made by ‘classical’ Levenshtein distance.

```

repeat

  compare each variety to each variety by using
  Levenshtein distance; the first time
  substitution=ins=del=1, later times the weights
  calculated in the previous iteration are
  used;

  find new weights by analyzing the alignments:
  the more frequently segments co-occur in
  an alignment, the smaller the distance weight

until weights do not change any more

```

Fig. 3 Pseudocode of the PMI Levenshtein distance algorithm.

### 2.3. Measuring consistency

It is important to check whether the number of items in the data set is large enough to extract a reliable signal. A well-known measure of reliability is the Cronbach's  $\alpha$  coefficient, which is derived from the inter-item correlation of the items (Cronbach (1951), see also Heeringa (2004, pp. 170–173) for more details). The value of Cronbach's  $\alpha$  ranges between zero and one. The higher the value of Cronbach's  $\alpha$  the greater the consistency between the items. A widely accepted threshold is 0.70 (Nunnally 1978, Nunnally & Bernstein 1994).

### 2.4. Visualizing dialect distances

Aggregated Levenshtein distances can be visualized by means of a beam map. Beam maps were introduced by Goebel (1983), see also Inoue (1996), Goebel (2010), and Kumagai (2016). Localities are connected to neighboring localities by a straight line. The darker the line, the smaller the distance. Lighter lines connect more remote dialects. In order to find neighboring localities, Delaunay triangulation was used. A network of triangles is constructed so that the vertices are the locality points and that there is no

<sup>4</sup> We used LED-A for calculating the PMI Levenshtein distances. The implementation of PMI-Levenshtein in LED-A differs in details from Wieling (2012). It is beyond the scope of this paper to go into this.

other locality point in the circumscribed circle of a triangle. Delaunay triangulation is named after the Russian mathematician Boris Delaunay (Delaunay 1934).

It is also possible to connect each locality with each locality regardless of whether the locations are adjacent to each other, see for example Heeringa et al. (2010), figures 3a and 4a. However, drawing beam maps on the basis of Delaunay triangulation like Goebel (1983) and others did, produces clearer images.

## 2.5. Finding dialect areas

In order to find areas in the dialect landscape, we use hierarchical agglomerative cluster analysis (HAC). Jain & Dubes (1988, p. 55) define cluster analysis as ‘the process of classifying objects into subsets that have meaning in the context of a particular problem.’ The goal of clustering is to identify the main groups in complex data. Those groups are called clusters. When using hierarchical agglomerative cluster analysis, a hierarchy of clusters is obtained, where each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. HAC was introduced by Goebel (1982) in dialectometry, see also Inoue (1996) and Kumagai (2013, 2016).

The general scheme used for the type of cluster analysis we use here is called Johnson’s algorithm. Jain & Dubes (1988) mention that the scheme was suggested by King (1967) and formalized by Johnson (1967). We will demonstrate the algorithm on the basis of Table 3. This table includes aggregated Levenshtein distances among five local dialects in The Netherlands.

Table 3 Aggregated Levenshtein distances among five local dialects in the Netherlands.

	Grouw	Haarlem	Delft	Hatterm	Lochem
Grouw	0	<b>41</b>	<b>44</b>	<b>45</b>	<b>46</b>
Haarlem	41	0	<b>16</b>	<b>34</b>	<b>36</b>
Delft	44	16	0	<b>37</b>	<b>38</b>
Hatterm	45	34	37	0	<b>20</b>
Lochem	46	36	38	20	0

Since the matrix is symmetrical (for example, the distance between Grouw and Haarlem is the same as between Haarlem and Grouw), we focus only on the values that are shown in bold. We start by looking for the smallest distance in the matrix. The smallest distance is 16, which is found between Haarlem and Delft. Therefore, we consider ‘Haarlem & Delft’ as one cluster. Both Haarlem and Delft are removed from the matrix, and a new cluster Haarlem & Delft is inserted. To iterate, we have to assign a distance from the newly formed cluster to all other points. For example, the distance between Grouw and Haarlem & Delft is calculated as the average of the distance

between Grouw and Haarlem (41) and the distance between Grouw and Delft (44) which is  $((41+44)/2=)$  42.5.

After calculating the distances between Hattem and Haarlem & Delft and between Lochem and Haarlem & Delft as well, we get the distance matrix that is shown in Table 4.

Table 4 Aggregated Levenshtein distances among four local dialects in the Netherlands and the cluster ‘Haarlem & Delft’.

	Grouw	Haarlem & Delft	Hattem	Lochem
Grouw	0	<b>42.5</b>	<b>45</b>	<b>46</b>
Haarlem & Delft	42.5	0	<b>35.5</b>	<b>37</b>
Hattem	45	35.5	0	<b>20</b>
Lochem	46	37	20	0

In each iteration the matrix is reduced in size. The iterations are repeated until no elements are left which can be fused to a new cluster. The final result is a complete hierarchical grouping of varieties. This grouping is visualized as a dendrogram, a tree in which the leaves are the varieties and the lengths of the branches correspond with the distances. In our example we get the dendrogram that is shown in Figure 4.

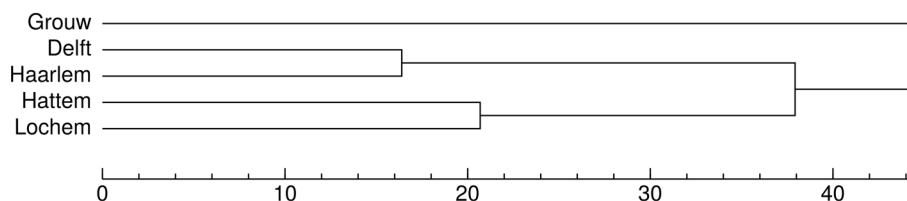


Fig. 4 Dendrogram obtained on the basis of the distance matrix in Table 2 using hierarchical agglomerative cluster analysis.

In our example the distance of an element –either a local dialect or a cluster– to a newly formed cluster is calculated by averaging. We calculated the average of the distance between Grouw and Haarlem and the distance between Grouw and Delft, thus we obtained the distance between Grouw and the cluster Haarlem & Delft. This type of clustering is known as Unweighted Pair Group Method using Arithmetic averages (UPGMA).

Several alternatives exist as well. In later work of Hans Goebel the Ward’s method was used, see, for example, Goebel (2006, 2007, 2010). This method minimizes the variance in the clusters (Ward, 1963). At each stage those two clusters are found whose merger gives the minimum increase in the total within group error sum of squares.

Gere (2023) extensively validated several cluster methods, including the Ward’s method. He concludes that using the “Ward’s method seems a safe and reasonable

choice.” The dendrograms produced by the Ward’s method have well distributed clusters, i.e. they are “symmetric and seem to have evenly distributed clusters.” As for a division into two groups, the Ward’s method produced clusters that were most significantly different.

The dendrogram provides valuable insight in the relationships among the local dialects. Additionally, we can derive dialect areas from the tree structure, and draw these areas in a map which we call area map.

Areas can be derived from the dendrogram by drawing a vertical line somewhere in the dendrogram and counting the horizontal lines matching it. Local dialects connecting to the matching horizontal lines will then belong to the same cluster (or area).

In order to find the optimal number of clusters, we have to find the right position of that vertical line in the dendrogram. This position is found by looking at the distances between the successive nodes in the dendrogram. In Figure 5 the nodes are numbered from 1 to 4. Each node marks the fuse of elements (local dialects or dialect clusters), and the number indicates the order in which the elements were fused. Then we have to find the largest horizontal difference between two successive nodes. In Figure 5 it can be seen that the largest distance between two successive nodes is between nodes 2 and 3. Now we draw a vertical line in the middle of these two nodes. The result can be seen in Figure 6. The number of horizontal lines that is intersected by that vertical line is the optimal number of clusters. In Figure 6 this vertical line intersects three horizontal lines which connect to three groups, namely Grouw, Delft and Haarlem, and Hattem and Lochem. Therefore, the optimal number of clusters is 3 in our example.<sup>5</sup>

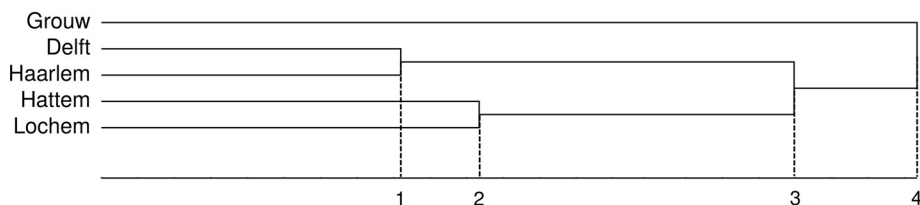


Fig. 5 The dendrogram has four nodes which are numbered from 1 to 4. Each node marks the fuse of elements (local dialects or dialect clusters), and the numbers indicate the order in which the elements were fused.

<sup>5</sup> For more information about this procedure see <https://statsandr.com/blog/files/Hierarchical-clustering-cheatsheet.pdf>. This procedure is implemented in LED-A.

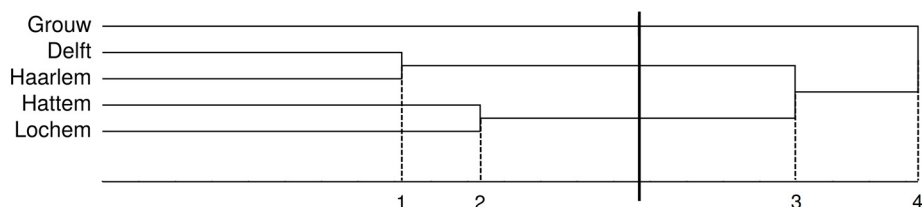


Fig. 6 The largest distance between two successive nodes is between nodes 2 and 3. A vertical line is drawn in the middle of these two nodes and intersects with three horizontal lines that connect with three groups.

When the dialect areas that are represented by the cluster groups are drawn in a map, each area gets a unique color. The way colors are assigned to the areas is described at the end of Section 2.6.

## 2.6. Showing the dialect continuum

Multidimensional scaling was introduced in dialectometry by Embleton (1993) (see also Embleton et al. (1993)). By using multidimensional scaling (MDS) we are able to visualize the dialect landscape as a continuum. When applying MDS to our example set of five local Dutch dialects, we obtain the plot shown in Figure 7. In this figure the five local dialects are put on a map so that the distances in two-dimensional space reflect the distances in the matrix as closely as possible. Like dialects are plotted nearby and unlike dialects are plotted distant to each other.

Fig. 7 Using MDS the five dimensions are reduced to two dimensions. X-coordinates represent the first, and Y-coordinates represent the second dimension.

MDS exists in several variants. Togerson (1952) proposed the first metric method, known as classical MDS. Kruskal extended this to non-metric MDS, which uses the ranks of the distances (Kruskal & Wish 1978). The plot shown in Figure 7 is made using Kruskal’s non-metric MDS.

In this paper we use t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton, 2008; Van der Maaten, 2014). Van der Maaten and Hinton (2008) write that “t-SNE is capable of capturing much of the local structure of the high-dimensional data very well, while also revealing global structure such as the presence of clusters at several scales” and that t-SNE “produces significantly better visualizations by reducing the tendency to crowd points together in the center of the map” which we found a justification for using this method in this paper.

In LED-A the function `Rtsne` of the R package `Rtsne` is used (Krijthe, 2015). This function has a perplexity parameter which is related to the number of nearest neighbors considered when placing each data point. When using lower values more attention is paid to the local aspects of the data, and the use of higher values increases the impact of more distant neighbors and global structure. Van der Maaten and Hinton (2008) write that “typical values are between 5 and 50”. The perplexity value is set to 30 by default. When the number of localities is smaller than 91, this would result in an error message: “Perplexity is too large”. In order to get the perplexity value as close as possible to the default value when the number of localities is smaller than 91, we calculate:

$$\text{perplexity} = (n - 1) \text{ DIV } 3$$

where  $n$  is the number of localities.

Using MDS it is possible to scale multidimensional data to two dimensions (like we did in Figure 7) or to three or more dimensions. When scaling to three dimensions, each local dialect is represented by three values, i.e., a value for  $x$ ,  $y$  and  $z$ . If we let  $x$  be the intensity of red,  $y$  the intensity of green and  $z$  be the intensity of blue, each dialect gets its unique color. In this way, each dialect point is associated with a mix of colors depending on its coordinates in the three-dimensional MDS solution, and a map is obtained that visualizes the dialect landscape as a continuum. The first MDS color map was published by Nerbonne & Heeringa (1998). Shortly afterwards the same map was also published by Nerbonne et al. (1999). We refer to this kind of maps as RGB maps.

In a RGB map each locality has its unique color. We also assign unique colors to the areas in an area map (see Section 2.5) by averaging the colors of the individual localities within each area.



## 2.7. Visualizing the dialect landscape from the perspective of a reference point

A reference point map is a map that shows the similarity or difference of local dialects compared to a reference point. The reference point can be one of the local dialects, a supraregional variety, a standard language or the proto-language. Reference point maps were introduced by Goebel (1981), see also Inoue (1996). While the black and white maps published in 1981 use different line patterns in order to express the similarity compared to the reference points, in later publications colors are used. Goebel (2002), for example, uses a rainbow scheme – red, orange, yellow, green, blue –, where dialects most similar to the reference point are red, and most distant dialects are blue. In LED-A a rainbow scheme is used as well. Kumagai (2016) varies the dot size: the larger the dot the more similar a variety is to the reference point.

## 3. Application to Japanese dialects

In this section we present the results of measuring aggregated normalized PMI Levenshtein distances among 2400 localities on the basis of 37 items from the LAJDB. In Section 3.1 we present the results of Japan, and in Section 3.2 we focus on Ryukyuan languages. We will compare our results to the other maps of Inoue & Yoshioka (2003) (see Map 1a and Map 1b) and to the map of Huisman et al. (2019), p. 8, Figure 1. The latter map is also found in Huisman (2021), p. 68, Figure 3-2.

### 3.1. Japanese dialects

First the conclusive explanation will be given.

The measurements for the full set of 2400 localities and 37 items appeared to be extremely consistent. We found a Cronbach's  $\alpha$  value of 0.9978 which indicates that 37 items are enough to provide a reliable results (see Section 2.3).

The clear correspondence between preceding studies of Japanese dialectologists in Maps 1a and 1b, and our results in Maps 2 to 11 is impressive.

#### 3.1.1. Dialect division theory and concentric distribution theory

An overview of trends in the study of Japanese dialects over the past 100 years reveals a conflict between the dialect division theory (*kukaku-ron*) and concentric distribution theory (center versus periphery theory, *shuken-ron*). Map 1a is a composite of a number of figures, superimposed on each other.

(1) The dialect division theory is concerned with boundaries, which Tojo (1927 and 1953) focused on. The isoglosses of grammar, phonology, pitch accent and lexicon (words) became the basis, and the East-West difference and the positions of Kyushu and Okinawa were the main themes.

(2) Yanagida (1943) proposed a theory of dialectal concentric distribution based on the dialectal names of snail and pointed out the distant agreement between periphery areas. He was not interested in the dialect division. Haruhiko Kindaichi (1955) pointed out the distant agreement mainly on the basis of pitch accent, but it should be called “reverse concentric distribution theory” because it was based on maintenance at the center and modification at the periphery, shown in Map 1a by the density of the brown color of the regions.

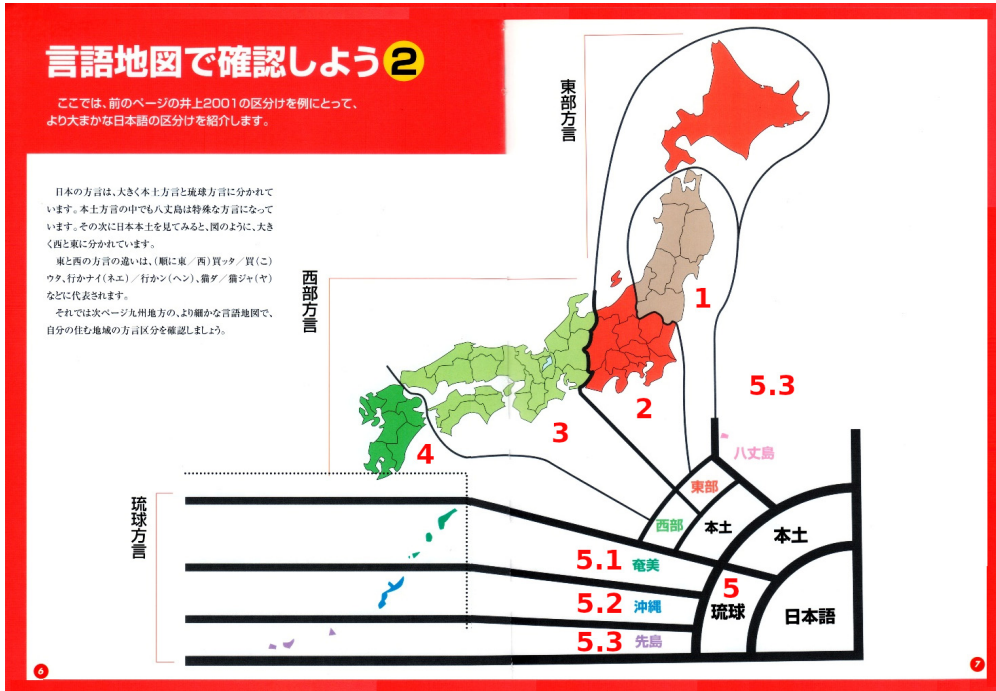
(3) Fujiwara (1990) pointed out the continuity of dialects, shown in Map 1a by the pattern and angle of straight and wavy lines in the region.



Map 1a Dialect areas in Japan (Inoue & Yoshioka 2003). The map is supplemented with the numbers of the five clusters, for the sake of comparison with the results of our analysis, since they are almost identical.

1 Tohoku dialect, 2 Eastern dialect, 3 Western dialect, 4 Kyushu dialect, and 5 Ryukyuan dialect (or Ryukyuan language(s), Okinawan dialects).

Map 1b shows the conclusions of Map 1a and Inoue (2001). As in Map 1a, the numbers of the five clusters have been supplemented. (This is in good agreement with the cluster analysis in Map 3 below.)



Map 1b Japanese dialects according to Inoue &amp; Yoshioka (2003).

### 3.1.2. UNESCO list of languages in danger

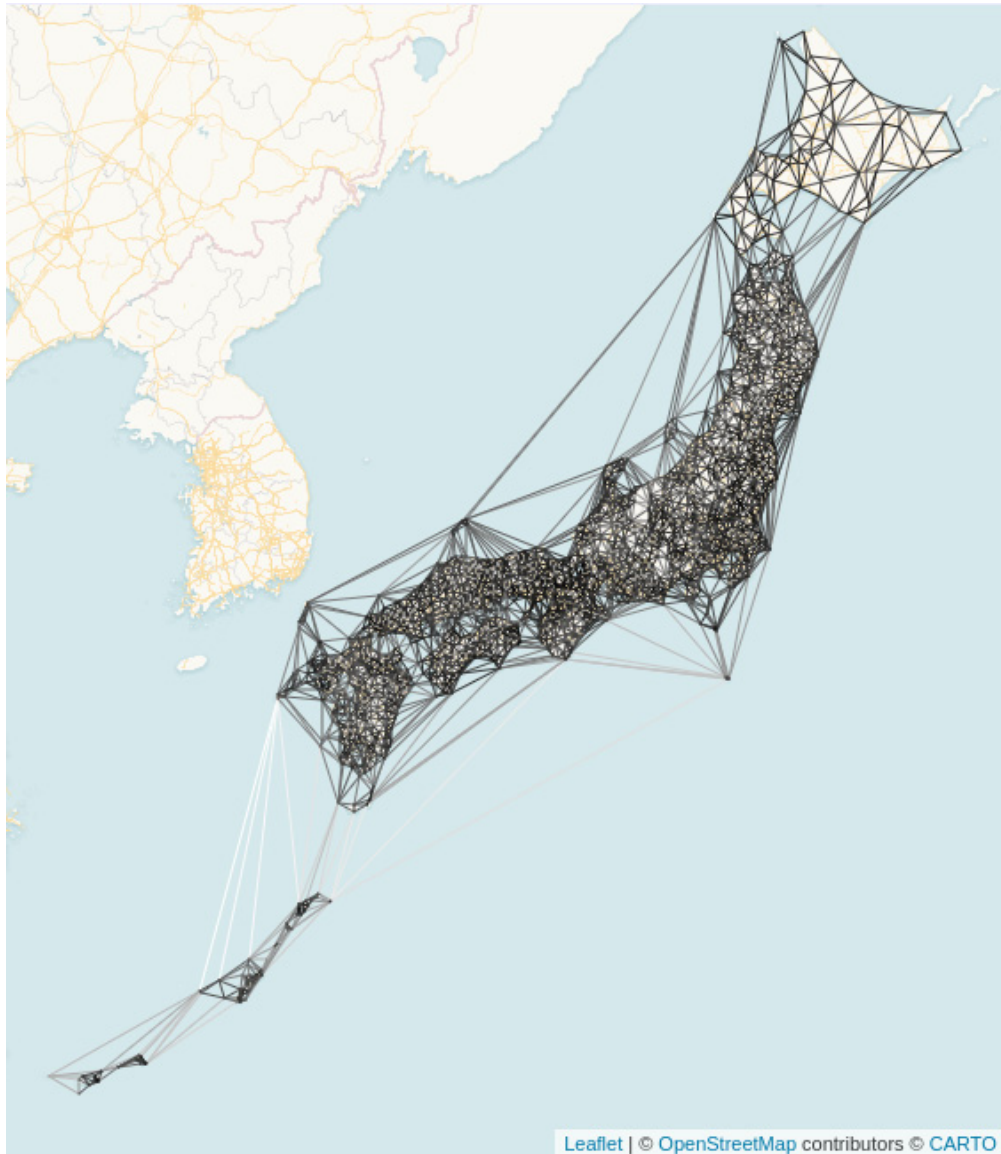
Map 1b differs from the eight languages in danger of UNESCO 2009. Dialectologists have diverse opinions about the subdivisions of Ryukyuan languages. Map 1b shows Ryukyuan languages divided into three groups: Amami (5.1), Okinawa (5.2), and Sakishima (5.3), while UNESCO 2009 divided in more detail into six groups as Amami (5.1), Kunigami (5.2), Okinawan (5.2), Miyako (5.3), Yaeyama (5.3) and Yonaguni (5.3). The criteria of UNESCO 2009 are unclear, both in terms of (linguistic) mutual intelligibility and in terms of extra-linguistic factors (former political boundaries). The Hachijo dialect, one of the languages in danger<sup>6</sup>, is placed outside the mainland dialects in Map 1b.

### 3.1.3. Overview of results of computational analysis

In Map 2a beam map is shown. In the map dark lines represent small distances and light lines large distances. Only adjacent localities are connected by lines (see Section 2.4). The map shows mainland Japan as a large continuum without any clear division

<sup>6</sup> It is considered to be a lineage of the ancient Eastern dialects.

in groups. In the southwest we find the Ryukyuan languages clearly distinguished from mainland Japan which are in turn distinguished in a northern and southern group.



Map 2 Beam map of 2400 Japanese localities.  
Dark lines represent small distances and light lines large distances.

#### **3.1.4. Dendrogram in cluster analysis**

We obtain a sharper and more detailed classification by means of Ward's clustering. The dendrogram is shown in Figure 8. The Dendrogram shows the branching process and hierarchy of the cluster analysis. The blue vertical line on the left suggests a division in five groups, which is the optimal number of groups (see Section 2.5). The blue vertical line on the right suggests a more detailed division in 15 groups. The two vertical lines show the reference points on the overall branching process. The five optimal groups are shown in Map 3, and the 15 groups are shown in Map 4.

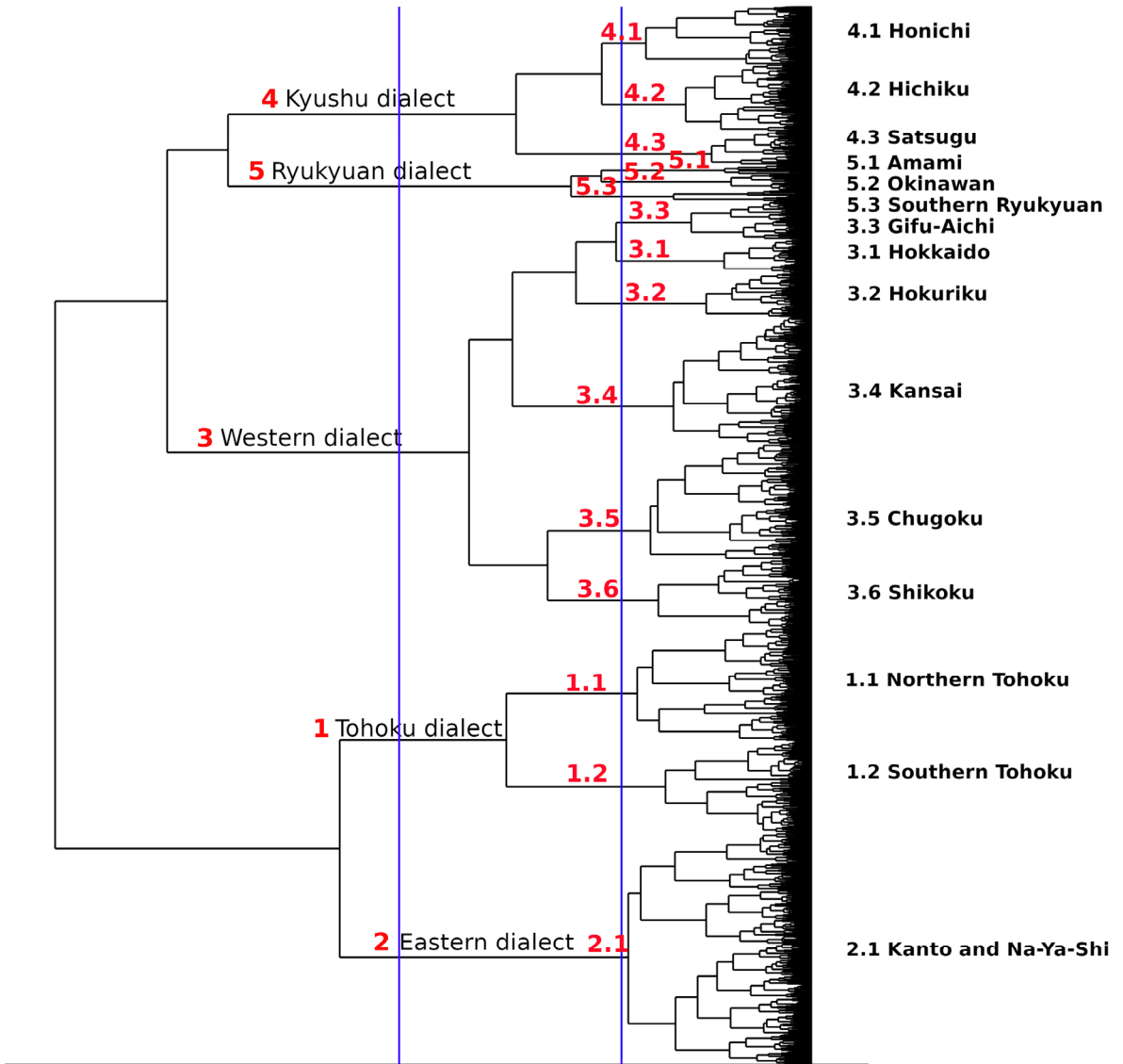


Fig. 8 Dendrogram.

The dendrogram was obtained by Ward's clustering on the basis of the aggregated normalized Levenshtein distances among 2400 Japanese localities. The blue vertical line on the left suggests a division in five groups, which is the optimal number of groups (see Section 2.5). The blue vertical line on the right suggests a more detailed division in 15 groups.

### 3.1.5. Map display of cluster analysis

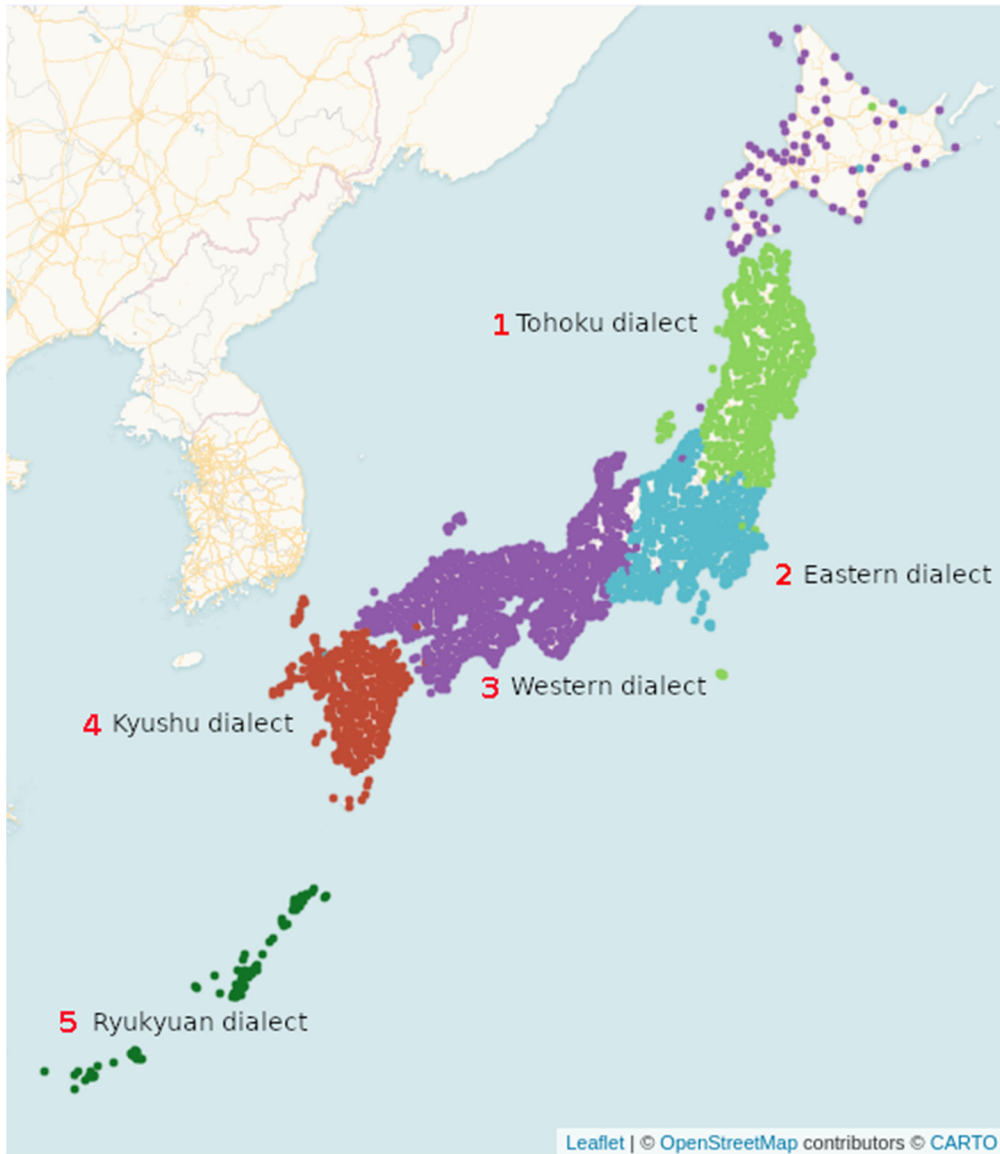
When comparing our Map 3 with Map 1b, we find in both maps the Ryukyuan languages (5), the Kyushu dialects (4), and a Western Japanese group (3). Our Eastern Japanese group (2) includes the Chubu dialects and the Kanto dialects. The group of Tohoku dialects (1) in our map mainly overlaps with Tohoku dialect group. In our map the Hokkaido dialects belong to the Western Japanese group (3), with some exceptions that belong to Eastern Japanese. This is the same as in Map 1b. Note also the green dot south of the Eastern Japanese group (2) which represents a Hachijo variety and which is classified among the Tohoku dialects.

The five clusters in Map 3 are plausible results, easy to understand and consistent with the preceding theory. The clusters are again numbered from north to south. The five clusters are consistent with Map 1b. The five clusters are also bisected into east (1 2) and west (3 4 5) clusters.

However, compared to Figure 8, the order of bifurcation (tree hierarchy) is different. What is different is that Ryukyu (5) and Kyushu (4) do not separate as early as in Figure 8. The boundaries in Map 3 correspond almost exactly to the prefectural boundaries, but there are some enclaves. We will see this in detail later. The study of Inoue (2001) was based on data that were averaged per prefecture, so minute enclaves were not visible. Yarimizu (2007) is the result of a multivariate analysis of the data per locality, and similar enclaves were observed.

In Map 4 a more detailed division into 15 groups is provided. Again Ward's clustering was used. Here we find especially that the further division of the Ryukyuan languages and the Western Japanese group agrees with Map 1a.

Due to the difference in detail, it is more difficult to compare the division of the mainland with the map of Huisman et al. (2019).



Map 3 Clustering of five optimal groups.  
Ward's clustering was applied to the average normalized Levenshtein distances among the 2400 Japanese localities. Five optimal groups were found which are shown in this map.





Map 4 Division into 15 groups.  
2400 Japanese local dialects into 15 groups. Ward's clustering was applied to the average normalized Levenshtein distances among the localities.

### 3.1.6. A more detailed discussion of dialect groupings

In this section we discuss the dialect groupings in more detail. We view the groups from the north to the south on Map 4 and check them against the dendrogram in Figure 8.

#### *3.1 Hokkaido dialect*

In Figure 8, Hokkaido is a part of Western dialects (3) in the dendrogram, belonging to the same cluster as the subcluster Gifu-Aichi (3.3). This is the only case where there is no geographical continuity on the map. This is because of mass migrations in the settlement frontier since the early modern period<sup>7</sup>.

#### *1 Tohoku dialect*

The dendrogram in Figure 8 divides the area into two parts: Northern Tohoku (1.1) and Southern Tohoku (1.2). This is in close agreement with the conventional dichotomy based on pitch accents. In Map 4 Northern Tohoku includes Sado Island.

Southern Tohoku is almost the same as the “no (monotypic) pitch accent region”, but there is a discrepancy with the boundary of Ibaraki and other prefectures (there is an enclave). It differs from the conventional division in that it includes Hachijo Island.

Dialects of remote (peripheral) areas are not prominent here. The pronunciation of Otori in Yamagata Prefecture and Hinoemata in Fukushima Prefecture are different from Tohoku dialectal, but are not prominent in this analysis.

#### *2 Eastern dialect*

In this cluster, the Kanto dialect and a large region of Na-Ya-Shi dialect (Nagano, Yamanashi, and Shizuoka prefectures) are grouped together and not separated. However, according to Figure 8, if the vertical line is shifted a little more to the right, the number of clusters becomes 16, and Kanto dialect and Na-Ya-Shi dialect are divided into two clusters of approximately the same size.

In Map 4 this is consistent with the east-west boundary line (Itoigawa-Hamanako line). The important east-west boundary, the North Alps, is exceptional. Blank areas with no inhabited settlements indicate the steepness of the natural terrain.

#### *3 Western dialect*

In Figure 8 three Western dialects cover a large area and are subdivided into many regions.

3.1 Hokkaido has already been described.

3.2 Hokuriku dialect

---

<sup>7</sup> It reminds us of interpretation of ALF enclaves by Goebel (2010).

Hokuriku includes the Toyama-Ishikawa-Fukui prefectures, which belong to the Western dialects. It is consistent with geography and administrative divisions, but does not include Niigata prefecture, which belongs to the Eastern dialect.

### 3.3 Gifu Aichi (Gi-A) dialect

This belongs to the west side of the east-west boundary of the Northern Alps. In Map 4, most of the localities in Aichi prefecture fall into the Kanto dialect and Na-Ya-Shi dialects (2.1), while the actual localities are almost exclusively in Gifu Prefecture.

### 3.4 Kansai dialect

Kansai is a large area of Kinki Region that includes Kyoto and Osaka, which were the center of Japanese cultural transmission in the pre-modern era. Other quantitative analysis shows a high degree of internal homogeneity in the Kansai dialect.

There is an enclave in the southern Kii Peninsula (Wakayama prefecture) which is similar to Gifu-Aichi dialect, and the connection to the enclave in the pitch accent is of interest.

### 3.5 Chugoku dialect

In Figure 8 there is a stage where Chugoku-Shikoku dialect is separated from the other clusters and is coming together as a cluster.

In Map 4 the Izumo-Inaba dialect (in the Shimane and Tottori prefectures) is not separated. The Northern San-in region is continuous with the southern San-yo region, and the area along the Seto Inland Sea is continuous. Furthermore, there are many enclaves of Kansai dialects in the central part of the Chugoku Mountains.

### 3.6 Shikoku dialect

In Figure 8 Shikoku is clearly separated from the others. Map 4 shows the distribution of enclaves of the Kansai dialects in Shikoku.

## 4 *Kyushu dialect*

Kyushu dialects (4) are grouped together. It is consistent with the conventional theory that Kyushu dialect is divided into three parts. However, there are enclaves in the Kyushu cluster.

Honichi dialects (4.1) of Oita and Miyazaki prefectures are located on the east coast of Kyushu, including Tsushima island (belonging to Nagasaki prefecture but with significant influence from Fukuoka prefecture in terms of transportation and economy). There is an enclave in Hiroshima prefecture.

Hichiku dialect (4.2) in Fukuoka, Saga, Nagasaki and Kumamoto prefectures is located on the west coast of Kyushu, with enclaves in the mountainous areas along its boundaries.

Satsugu (Satsuma Osumi) dialect (4.3) extends beyond Kagoshima prefecture in terms of its boundaries into the territory of Kumamoto prefecture.

### *5 Ryukyuan languages*

As for the Ryukyuan languages (5), the Southern Ryukyuan languages in our map are divided in Yaeyama Ryukyuan and Miyako Ryukyuan in the map of Huisman et al. (2019). Both in the map of Huisman et al. (2019) and in our map we find a distinction between the Okinawan language and the Amami language, but the boundary between the two groups is found more to the south in our map, making the groups of Amami varieties larger and the groups of Okinawan languages smaller.

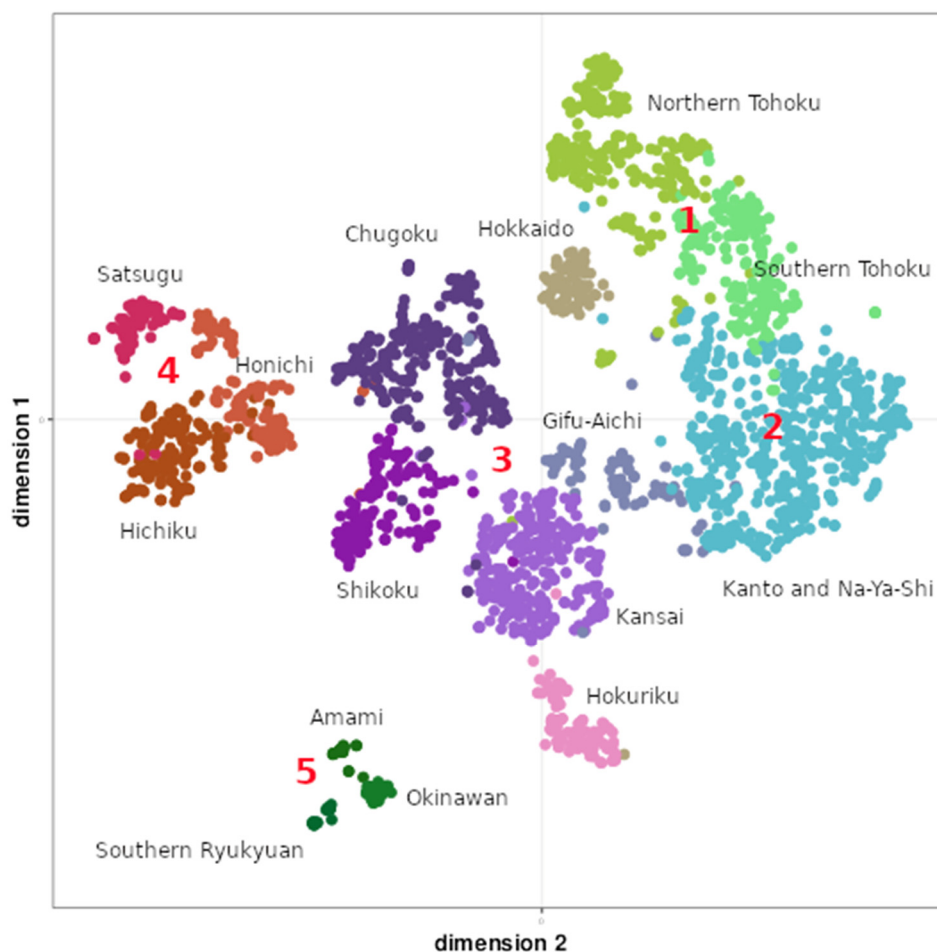
Ryukyuan languages (5) are grouped together and have no enclaves, and the trichotomy into Amami, Okinawan, and Southern Ryukyuan is consistent with the preceding theories. In the dendrogram of Figure 8, Southern Ryukyuan (5.3) is trisected into three sub-divisions. In Map 4 there are no color differences among the three divisions, but they are subdivided in the RGB map of Map 11.

#### **3.1.7. Application of multidimensional scaling**

In Map 5 a multidimensional scaling (MDS) plot is shown. Using t-SNE the distances among the 2400 localities were scaled to two dimensions. The colors correspond with the colors in Map 4. The groupings suggested by the plot itself and by the cluster analysis agree relatively well with each other. However, the Northern Tohoku dialects 1 (yellowish green), Southern Tohoku dialects 1 (green) and Kanto and Na-Ya-Shi dialects 2 (lighter blue) which are distinguished by the cluster analysis look more like a continuum in the plot itself.

In Map 5, the five clusters were numbered, and the locations of the regions resemble the east-west configuration of the Japanese archipelago. However, among the three western regions, Hokuriku (3.2) and Gifu-Aichi (3.3) are reversed from north to south.

The graph reflects the geographical arrangement of the three regions in the results of many previous multivariate analyses. This indicates the reliability of the multivariate analysis.



Map 5 Multidimensional scaling plot.

Multidimensional scaling plot obtained on the basis of the average normalized Levenshtein distances among the 2400 Japanese localities. Using t-SNE the distances were scaled to two dimensions. The colors correspond with the colors in Map 4.

### 3.1.8. Visualizing the continuum in a RGB map

The RGB map in Map 6 shows the Japanese dialect landscape as a continuum. Especially the Ryukyuan varieties 5 and the Kyushu varieties 4 in the most southern part of mainland Japan are well distinguished. In the Western Japanese group 3 a tripartite division in Chugoku 3.5, Shikoku 3.6 and Kansai 3.4 varieties can be recognized. The Hokuriku dialects 3.2 are clearly distinguished within the Western Japanese group 3. Within the group of Tohoku varieties 1 a division between northern varieties 1.1 and southern varieties 1.2 can more or less be recognized.

### *Continuity*

The RGB Map shows the continuity of the various dialects. In conclusion, this map depicts the reality of all dialects in the Japanese archipelago as continuous, as opposed to cluster analysis. In Map 6 R Red is for Kyushu (4), G Green is for Ryukyuan (5), B Blue is for Eastern (2).

### *Boundaries*

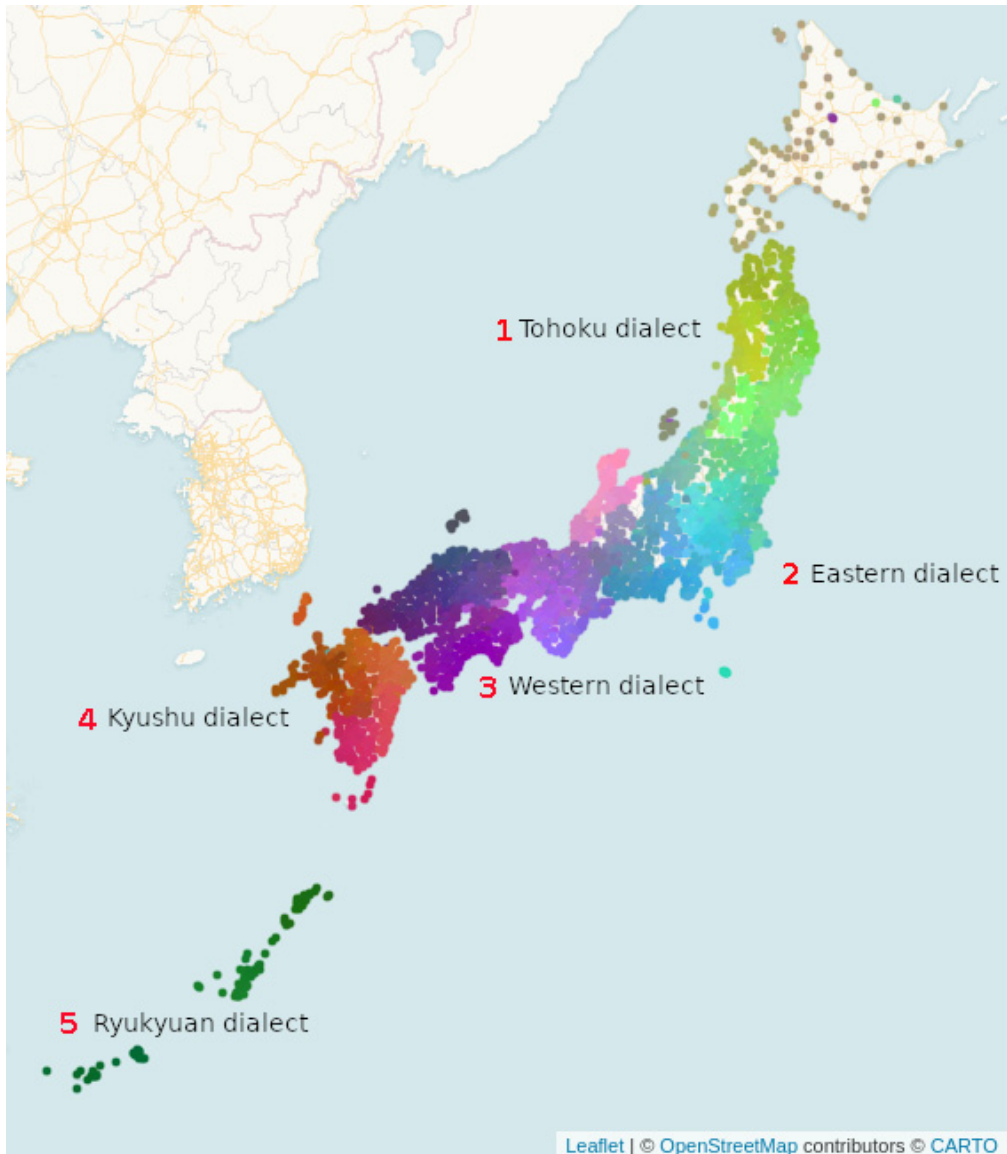
The RGB Map should show continuity, but the boundary line is clearly visible. The more southern, the more pronounced, especially between Ryukyu (5) and Kyushu islands (4), and between Kyushu islands (4) and western dialects (3) of Chugoku and Shikoku. The other dialects without ocean border are continuous, as in Chugoku and Kansai dialects (3).

The difference between the east-west boundary of the steep Northern Alps is large at the northern end, Hokuriku, while the Tokaido side at the southern end is continuous.<sup>8</sup> Eastern (2) and Tohoku (1) also appear to be continuous.

Compared to the Dutch RGB Map, Map 6 shows clear differences in coloration. There is a difference with Swiss German even in the same Germanic languages.

---

<sup>8</sup> This phenomenon corresponds to the Rhenish fan in the German dialects.



Map 6 RGB map of all Japan

RGB map obtained on the basis of the average normalized Levenshtein distances among the 2400 Japanese localities. Using t-SNE the distances were scaled to three dimensions. The first dimension is mapped inversely to red, the second inversely to blue and the third inversely to green. Thus each location gets its own unique color and the dialect landscape is visualized as a continuum.

### 3.1.9. Reference point map with Tokyo as reference point

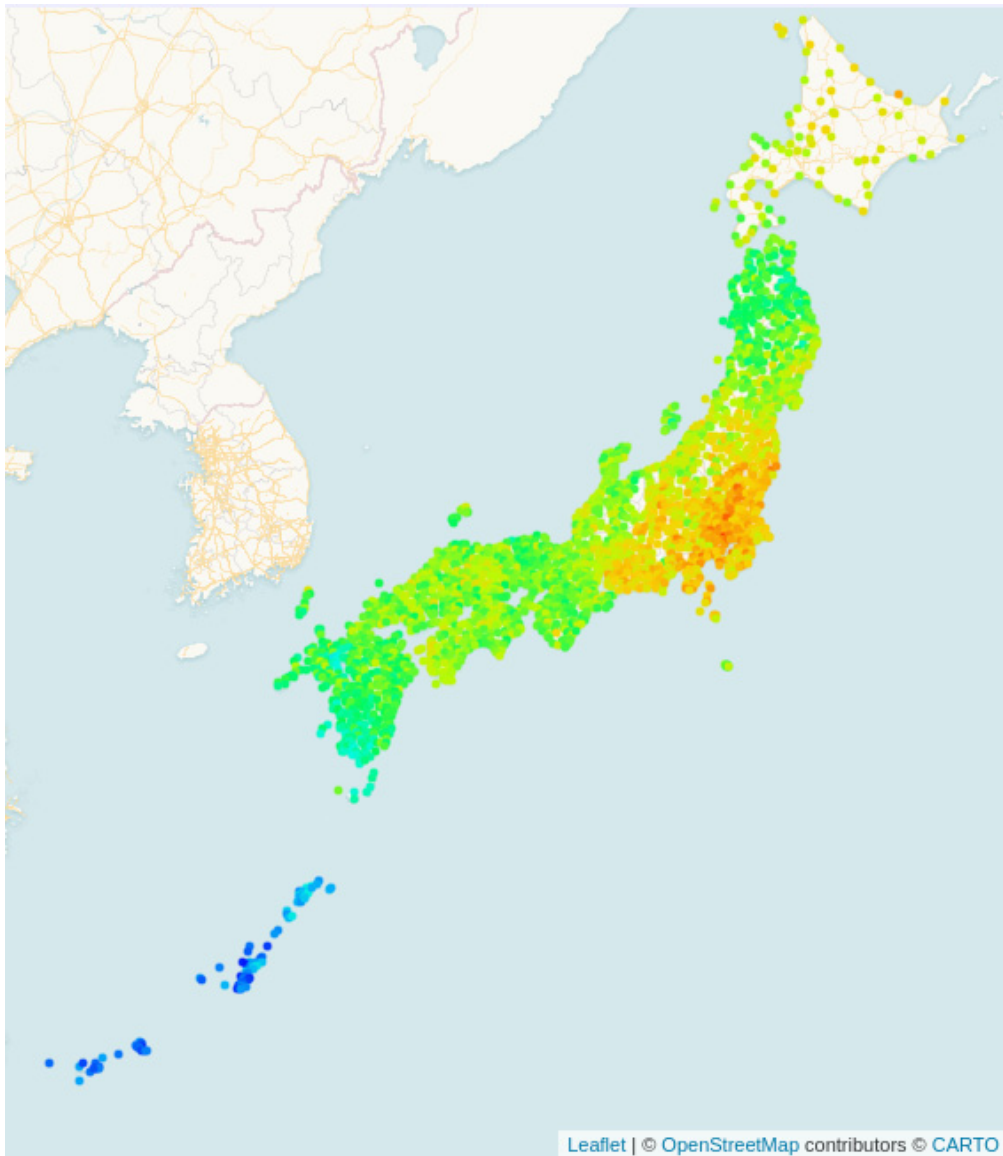
In Map 7 the Japanese local dialects are compared to the dialect spoken in or close to Tokyo.<sup>9</sup> Distances are visualized by colors according to a rainbow scheme. Red means that a local dialect is identical to the dialect of location 5698.69, and darker blue means that a local dialect is maximally different from location 5698.69. In general, the greater the geographical distance from Tokyo, the greater the aggregated normalized Levenshtein distance. The most distant local dialects are found in the Ryukyuan group.

There is an older attempt at degrees of agreement with the standard form or Tokyo form using prefecture-by-prefecture averages (Inoue 2001), and there is also an analysis based on the grammatical phenomena of GAJ 800 localities data by Yarimizu (2007). Comparing the data from the standard (map headword form) or Tokyo respondents to each prefecture or each informant is almost proportional to distance. Using the rail distance from the old capital Kyoto and the new capital Edo Tokyo, a link between pre-modern and post-modern transportation routes was ascertained (Inoue 2020). It can be inferred from this map.

---

<sup>9</sup> In the LAJDB all localities are represented by codes. We determined location 5698.69 closest to Tokyo just by eyeballing.





Map 7 Reference point map

Reference point map showing the distances of 2399 local Japanese dialects relative to point no. 5698.69 (Tokyo). Distances are visualized by colors according to a rainbow scheme. Red means: local dialect is identical to the dialect of location 5698.69, darker blue means: the local dialect is maximally different from location 5698.69.

### **3.2. Ryukyuan languages**

In this section we focus on the relationship between the various localities inside the Ryukyuan languages.

The LAJDB includes 83 local dialects that belong to the Ryukyuan languages (5, see Map 4). The items with the numbers 194 and 195 were not found for the localities in this area, therefore our measurements are obtained on the basis of 35 items. When calculating the average normalized Levenshtein distances among the localities, we used the PMI segment distances (see Section 2.2) that were calculated on the basis of the complete set of 2400 localities.

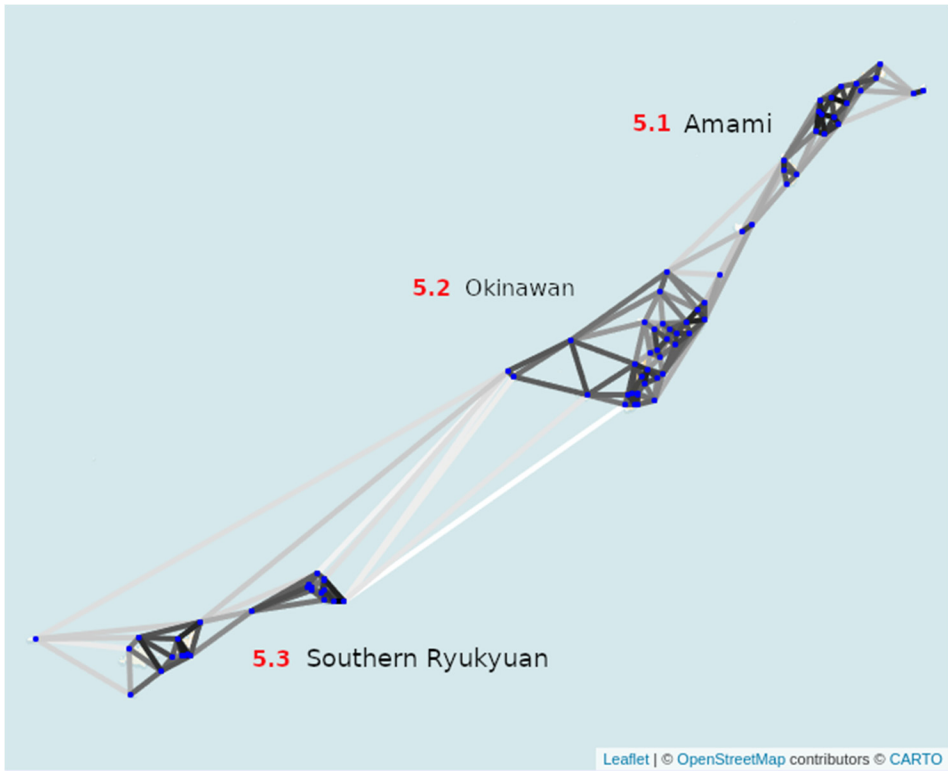
We measured Cronbach's  $\alpha$  and found a value of 0.9348, which is much higher than the widely accepted threshold of 0.7. Therefore, we can be confident that our results are reliable.

#### **3.2.1. Overview of metrological analysis results**

In Map 8 a beam map shows the relationships of adjacent localities to each other. When applying Ward's clustering to the 83 local dialects, the optimal number of clusters (or groups) is three.

The dendrogram is shown in Figure 9 and the area map in Map 9. For the coloring of the groups we used again the procedure that is described at the end of Section 2.6. The dendrogram shows the branching process. In the dendrogram two blue vertical lines are drawn. The one on the left divides into the top three clusters, and the one on the right divides into the top nine clusters. Southern Ryukyuan is conventionally divided into three clusters (also by UNESCO). If a vertical line is drawn at that position, Okinawan is divided into two clusters, and Amami is divided into four clusters, together dividing it into nine clusters.

Map 9 shows the division into the top three clusters on a map. The three clusters are neatly divided geographically. A map showing the division in nine clusters is omitted.



Map 8 Beam map of 83 local dialects that belong to the Ryukyuan languages.  
Dark lines represent small distances and light lines large distances.

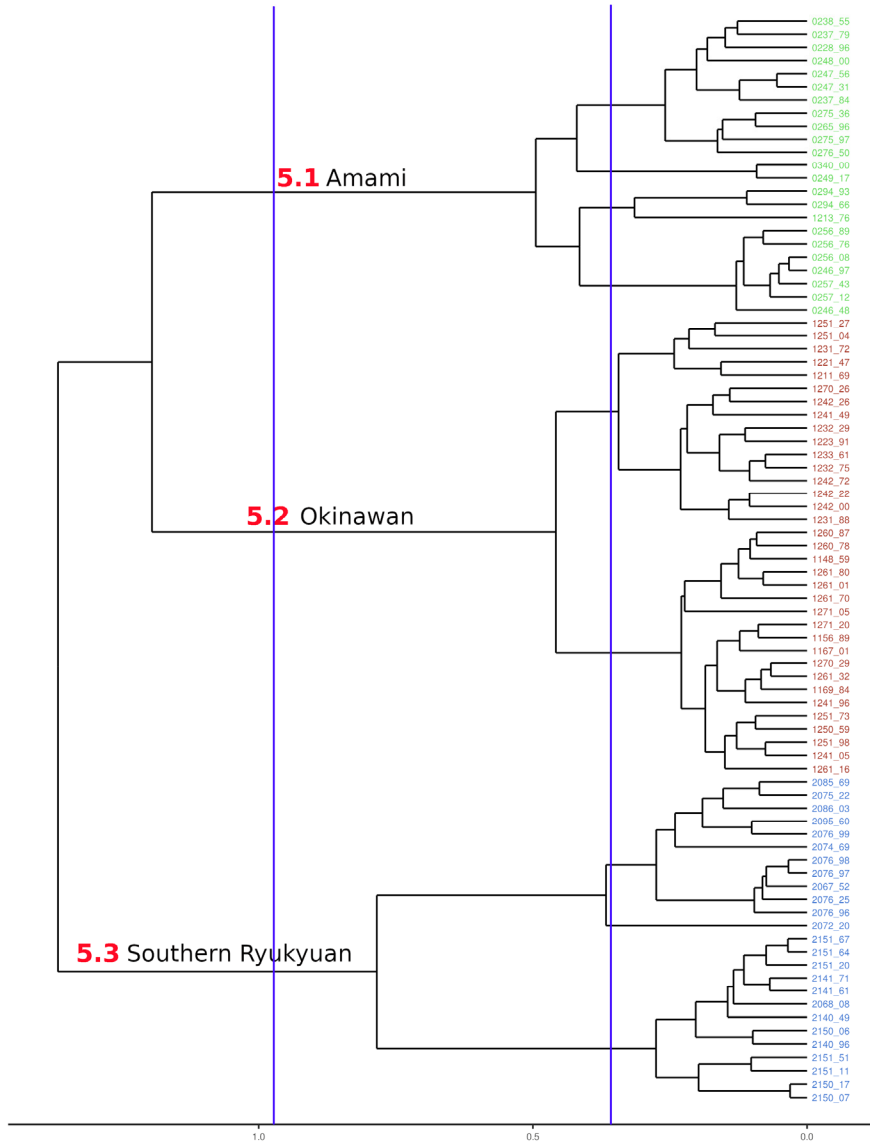
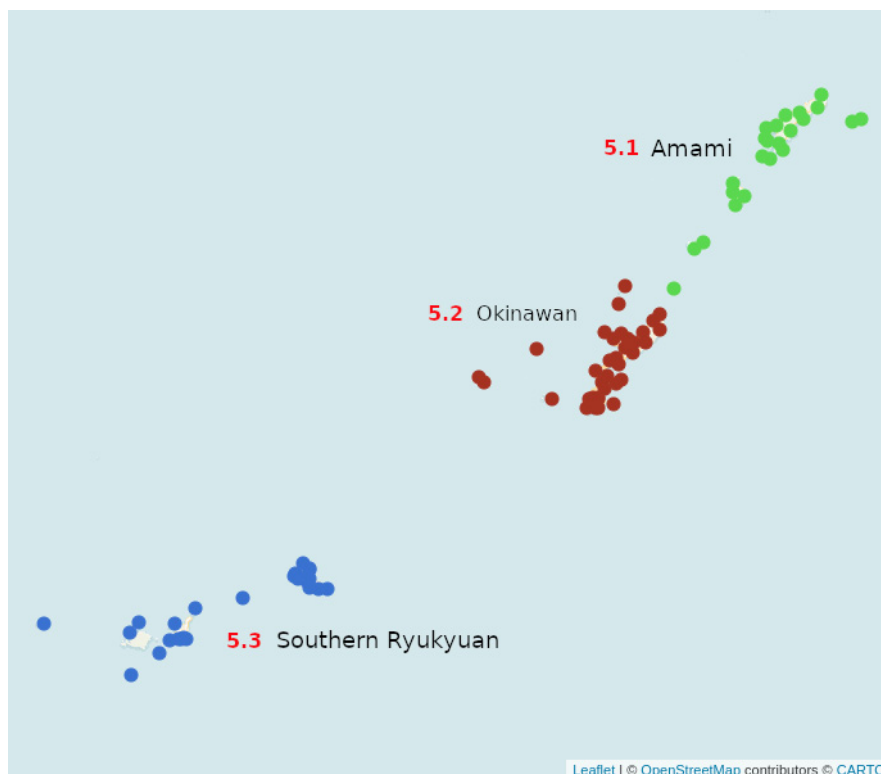


Fig. 9 Dendrogram

The dendrogram was obtained by Ward's clustering on the basis of the aggregated normalized Levenshtein distances among 83 local dialects that belong to the Ryukyuan languages. The blue vertical line on the left suggests a division in three groups, which is the optimal number of groups (see Section 2.5). The blue vertical line on the right suggests a more detailed division in nine groups.



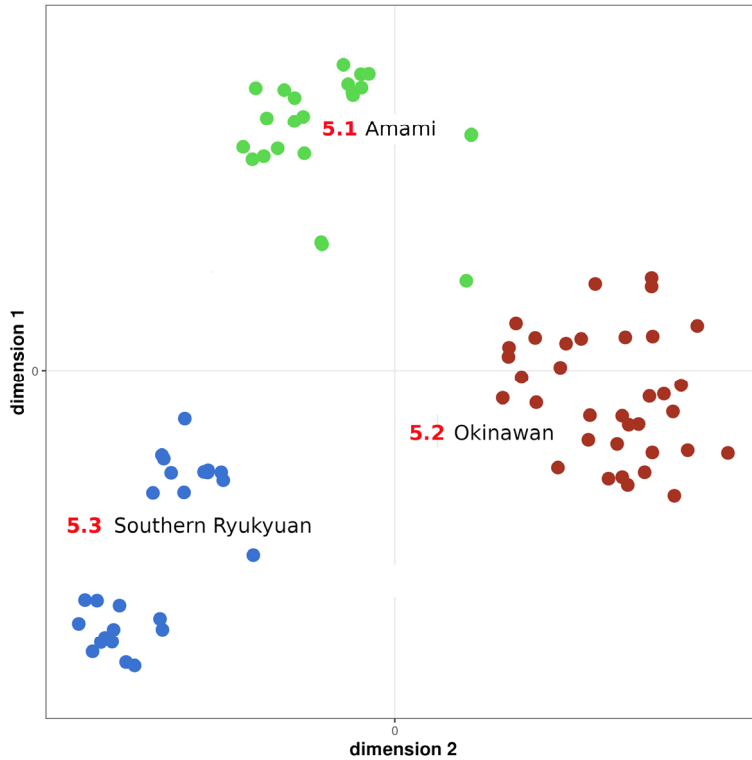
Map 9 Clustering of three optimal groups.

The 83 local dialects that belong to the Ryukyuan languages divided into three groups. Ward's clustering was applied to the average normalized Levenshtein distances among the localities.

### 3.2.2. Application of multidimensional scaling

In Map 10 a multidimensional scaling plot can be found. Using t-SNE the distances among the 83 localities were scaled to two dimensions. The colors correspond with the colors in Map 9. The groupings suggested by the plot itself and by the cluster analysis agree relatively well with each other.

The results of application of the MDS were mapped. It is clearly divided into three parts, with an intermediate point between Amami and Okinawa, but no enclaves were observed. This is a major difference between the mainland, which is almost completely connected by land, and the Ryukyu Islands, which have many isolated islands.



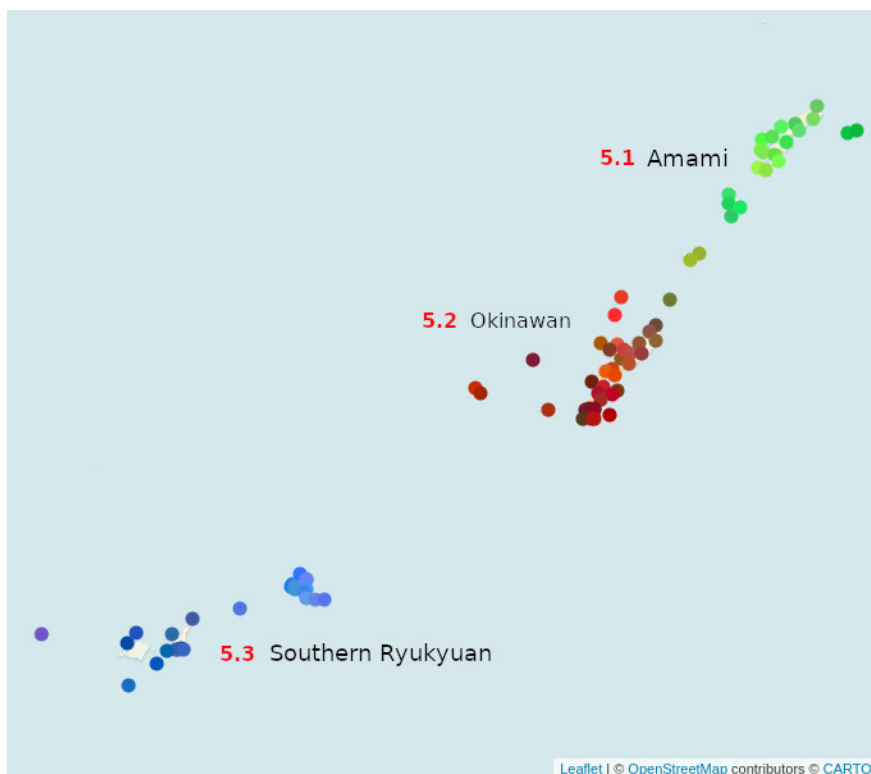
Map 10 Multidimensional scaling plot

Multidimensional scaling plot obtained on the basis of the average normalized Levenshtein distances among 83 local dialects that belong to the Ryukyuan languages. Using t-SNE the distances were scaled to two dimensions.

### 3.2.3. Visualizing the continuum in a RGB map

Map 11 shows the RGB map. The map suggests a division in three groups as well, although some heterogeneity is found within the groups, especially within the groups of the Okinawan dialects.

When looking closer to the beam map in Map 8, the dendrogram in Figure 9 and the multidimensional scaling plot in Map 10, we find two subgroups within the Southern Ryukyuan group 5.3 which correspond with the Yonaguni and Yaeyama dialects in the southwest and the Miyako dialects more to the northeast. In the RGB map, however, the colors of those two subgroups are not strongly distinguished. Rather we find the Yonaguni variety – represented by the most western purple dot – distinguished from the other Southern Ryukyuan varieties.



Map 11 RGB map

The map was obtained on the basis of the average normalized Levenshtein distances among the 83 local dialects that belong to the Ryukyuan languages. Using t-SNE the distances were scaled to three dimensions. The first dimension is mapped inversely to red, the second to green and the third to blue.

The RGB Map shows continuity. In conclusion, the entire area of Ryukyuan dialects is expected as continuous in reverse to the cluster analysis. In Map 11, R Red is for Okinawan (5.2), G Green is for Amami (5.1), and B Blue is for Southern Ryukyuan (5.3). In terms of color, the differences among the three clusters are significant, as shown in Map 9. There is a continuity of tints between Okinawan (5.2) and Amami (5.1), but there is a clear difference between Okinawan (5.2) and Southern Ryukyuan (5.3). Its internal difference is small in terms of tint. This is a trend not observed in the whole Japan scale RGB Map (Map 6).

Since Ryukyuan dialects are internally diverse, 37 items of our data were not sufficient. It is desirable to treat quantitatively with a larger number of items. However, as in Nakamoto (1983, p.166), there are cases in which even a small number of items could show the internal concentric distribution of Ryukyuan dialects in the former Ryukyu Kingdom. The choice of items is also important.

## 4. Discussions and conclusions

### 4.1. Direct results

Using the dialectometric methodology as presented in this paper and implemented in LED-A we were able to measure and visualize lexical and phonetic/phonological variation in word forms of 2400 local dialects and 37 items from the LAJDB. By using the dialectometric approach, with PMI Levenshtein distance as our workhorse, we did not suffer from the limitations of the traditional isogloss method, i.e. we could utilize all the information that is contained by the transcriptions without making subjective choices. By using different visualization techniques – beam maps, areas maps, RGB maps and reference point maps – we were able to visualize dialect variation from different but complementary angles and at different levels of detail.

The results that we obtained are plausible and consistent with the preceding studies such as those from Fujiwara (1990), Inoue (2001) and others. They reinforce, rather than contradict, the results of previous Japanese dialectologists. This outcome also confirms the validity, convenience, and reliability of the Levenshtein distance.

In lexicostatistics proposed by M. Swadesh in the 1950s, the question was not whether the word forms were similar but whether they were cognates. The LAJ was concerned with derivational relations between word forms, and the choice of map symbols reflected derivational relations. The GAJ, on the other hand, was interested in the grammatical conjugation system, and sometimes treated words with similar pronunciations as grammatically distinct. Both were based on the researcher's academic knowledge and were virtuoso and subjective. In contrast, the Levenshtein distance is an objective criterion and reliable.

Methodologically speaking, our approach is very similar to Huisman's (2019). As to the differences we note the following:

1. We measured all-word distances which include both lexical and phonetic/phonological variation.

2. Huisman et al. (2019) used 100 items from the Swadesh list which includes basic concepts. Our word list contains some basic concepts as well, but also less basic concepts like *tadpole*, *dragonfly* and *whirlpool*. The difference in concepts may explain the difference in results. Huisman et al. (2019) may have measured all-word distances as well, but they do not mention this explicitly.

3. We do not expect that the difference in the number of items (100 versus 37) plays a role. We found Cronbach's  $\alpha$  values larger than 0.9 which shows that our measurements are very reliable (see Section 2.3, 3.1 and 3.2).



4. Huisman et al. (2019) used ‘classical’ Levenshtein distance, whereas we used PMI Levenshtein distance, which may also explain the difference between their and our results.

5. The analysis of Huisman et al. (2019) is obtained on the basis of 58 Japanese and 32 Ryukyuan varieties. The LAJDB includes 2317 Japanese and 83 Ryukyuan varieties. This allowed us to zoom in more deeply on the results (see, for example, Maps 4 and 5).

6. Like Huisman et al. (2019) we applied Ward’s clustering to the Levenshtein distance matrix. However, in addition to this we used t-SNE for plotting dialect relationships in two-dimensional space and for creating RGB maps. A reference point map was added as well.

#### **4.2. Macroscopically and microscopically**

##### *Macroscopically*

On the large scale, the dialect divisions obtained were in close agreement with the prevailing theory. In particular, the east-west boundary, which has been the focus of interest in Japanese dialectology, was in agreement in terms of the geographical location of the border.

However, there were discrepancies in the hierarchical structure of the dendrogram which did not show that the Ryukyuan dialects (and Kyushu dialects) are far removed from the mainland dialects. On the other hand, Map 7 shows that the Southern Ryukyuan dialects and the Okinawan dialects are most distant to the dialect of Tokyo. This agrees with previous studies in which the usage rate of standard forms was found to show a neat hierarchy of dialects from different regions in the proximity of Tokyo.

##### *Microscopically*

When looking at the data in detail, enclaves were observed. In previous studies that focused on the average use rate of words in the prefectures it was not possible to detect those enclaves.

The enclaves we found may be the result of migration. According to introductory books on dialect in the first half of the 20th century, there were enclaves of feudal lords and their vassals in various parts of Japan. There were regional differences with social stratification due to the relocation of the feudal lords of the Edo shogunate. Thanks to a large-scale study on the (new) dialect formation in mass settlements in Hokkaido, the hypothesis that regional differences fade away after three generations has been

proposed. This tendency seemed applicable to derivational relations. On the other hand, in mountainous areas throughout Japan, there are legends of medieval samurai who were defeated in battle and migrated. In the mountainous areas, there are villages that use a different language from that of the surrounding area. Whether or not the enclaves found in this paper are the result of emigration should be examined on an individual basis in the future.

The individual differences that suggest the existence of enclaves may also be due to mere coincidence. Whether it is a coincidence or not can be determined by comparing our results with the results of an analysis on the basis of other items.

#### **4.2. Future developments**

The first data ever treated quantitatively were standard word forms of 82 LAJ items each of them being averaged per prefecture (Inoue 2001). It was conjectured at that time that other items and approaches show similar distribution patterns. With this paper, our earlier desire for a quantitative study of the entire LAJ 2400 localities data has been fulfilled. Although the number of items in this study is small, it agrees well with the previous analysis and demonstrates the reliability of both approaches.

In the previous analysis of the standard word forms for the 82 LAJ items, multivariate analysis was also applied to the word dimension of the data matrix, and succeeded in confirming relationships with the semantic field, frequency of use, and year of first appearance in the literature. This method can also be applied to the 37 items that were used in this study.

The large language difference between mainland Japan and Ryukyu (Okinawa) was striking. The chain of mutual intelligibility was not established in Japan, and the language difference was comparable to that of Germanic languages in northern Europe.

In our study we focused on lexical and phonetic/phonological variation in word forms. But we are aware that measures focusing on morphological, syntactic and prosodic variation may yield different results. Useful future research may be to study these levels as well.

Future quantitative work may also include application to ALE data and other European dialectology data; application to the Swadesh list and parts of Bible, as well as to discourse, which is already underway.

Some quantitative studies have added GIS geographic information to look at the relationship between topography, transportation, economics, and history. These new approaches will allow generalizations to be made about the rules and regularities of dialect distribution in general.

## References

- Bolognesi, Roberto and Wilbert Heeringa (2002) De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten [The influence of dominant languages on the lexicon and phonology of Sardinian dialects]. *Gramma/TTT: tijdschrift voor taalwetenschap* 9(1): 45–84. URI: <http://www.wjheeringa.nl/papers/rom01.pdf>
- Cronbach, Lee J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297–334. URI: [http://cda.psych.uiuc.edu/psychometrika\\_highly\\_cited\\_articles/cronbach\\_1951.pdf](http://cda.psych.uiuc.edu/psychometrika_highly_cited_articles/cronbach_1951.pdf)
- Delaunay, Boris (1934) Sur la sphère vide. *Bulletin de l'Académie des Sciences de l'URSS, Classe des Sciences Mathématiques et Naturelles* 6: 793–800. URI: [http://galiulin.narod.ru/delaunay\\_.pdf](http://galiulin.narod.ru/delaunay_.pdf)
- Embleton, Sheila (1993) Multidimensional scaling as a dialectometrical technique: outline of a research project. In: R. Köhler & B. Rieger (eds.), *Contributions to Quantitative Linguistics*, 267–276. Dordrecht: Kluwer.
- Embleton, Sheila, Dorin Uritescu & Eric S. Wheeler (1993) Defining dialect regions with interpretations: advancing the mds approach. *Literary and Linguistic Computing* 28(1): 13–22. doi: <https://doi.org/10.1093/lilc/fqs048>
- Gere, Attila (2023) Recommendations for validating hierarchical clustering in consumer sensory projects. *Current Research in Food Science* 6: 100522. doi: <https://doi.org/10.1016/j.crf.2023.100522>
- Goebel, Hans (1981) Éléments d'analyse dialectométrique (avec application à l'ais). *Revue de linguistique romane* 45: 349–420.
- Goebel, Hans (1982) *Dialektometrie; Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie* [Dialectometry; principles and methods of using numerical taxonomy in the field of dialect geography]. Philosophisch-Historische Klasse Denkschriften 157. Vienna: Verlag der Österreichischen Akademie der Wissenschaften. With assistance of W.-D. Rase and H. Pudlatz.
- Goebel, Hans (1983) Parquet polygonal et treillis triangulaire: les deux versants de la dialectométrie interponctuelle. *Revue de linguistique romane* 47: 353–412.
- Goebel, Hans (2002) Analyse dialectométrique des structures de profondeur de l'ALF. *Revue de linguistique romane* 66(261–262): 5–63.
- Goebel, Hans (2006) Recent advances in Salzburg dialectometry. *Literary and linguistic computing* 21(4): 411–435. doi: <https://doi.org/10.1093/lilc/fql042>
- Goebel, Hans (2007) On the geolinguistic change in Northern France between 1300 and 1900: a dialectometrical inquiry. In: *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, 75–83. URI: <https://aclanthology.org/W07-1310/>
- Goebel, Hans (2010). Dialectometry: Theoretical prerequisites, practical problems, and concrete applications (mainly with examples draw from the "Atlas linguistique de la France", 1902-1910). *Dialectologia: revista electrònica* 63–77. URI: <https://raco.cat/index.php/Dialectologia/article/view/242102/324714>
- Grosse, Rudolf (1955) *Die meissnische Sprachlandschaft: Dialektgeographische Untersuchungen zur obersächsischen Sprach- und Siedlungsgeschichte*. Halle: M. Niemeyer.
- Heeringa, Wilbert (2004) Measuring Dialect Pronunciation Differences using Levenshtein Distance. Unpublished doctoral dissertation, University of Groningen. URI: <http://www.wjheeringa.nl/thesis/>
- Heeringa, Wilbert, John Nerbonne and Peter Kleiweg (2010) Dialectvariatie, dialectafstanden en dialectindelingen. Een formele analyse van uitspraakverschillen [Dialect variation, dialect distances and dialect divisions. A formal analysis of pronunciation differences]. In: Onno Boonstra and Anton Schuurman (eds.), *Tijd en ruimte. Nieuwe toepassingen van GIS in de alfawetenschappen*, 88–101. Utrecht: Uitgeverij Matrijs.

- Huisman, John L.A., Asifa Majid and Roeland van Hout, R. (2019) The geographical configuration of a language area influences linguistic diversity. *PLoS ONE* 14(6): e0217363. doi: <https://doi.org/10.1371/journal.pone.0217363>
- Huisman, John L.A. (2021) Variation in form and meaning across the Japonic language family: With a focus on the Ryukyuan languages. Doctoral dissertation, Radboud University Nijmegen. URI: [https://pure.mpg.de/rest/items/item\\_3311139/component/file\\_3311141/content](https://pure.mpg.de/rest/items/item_3311139/component/file_3311141/content)
- Inoue, Fumio (1996) Computational dialectology (1). *Area and Culture Studies* 52: 67–102. doi: <https://doi.org/10.15026/23628>
- Inoue, Fumio (2001) *Keiryoteki Hogen Kukaku* [Quantificational Dialect Classification] Tokyo: Meiji Shoin.
- Inoue, Fumio (2020) Standard language distribution in LAJ and railway distance --- Geographical and historical interpretation of limestone cave model. *Dialectologia: revista electrónica* 24: 111–156. URI: <http://www.publicacions.ub.edu/revistes/dialectologia24/documentos/1602.pdf>
- Inoue, Fumio & Yasuo Yoshioka (eds.) (2003) *Kyushu no Hogen*. Tokyo: Yumani Shobo.
- Ivić, Pavle (1962) On the structure of dialectal differentiation. *Word* 18(1–3): 33–53. URI: <https://doi.org/10.1080/00437956.1962.11659764>
- Jain, Anil K. and Richard C. Dubes (1988) *Algorithms for Clustering Data*. Englewood Cliffs, New Jersey: Prentice Hall.
- Johnson, Stephen C. (1967) Hierarchical clustering schemes. *Psychometrika* 32: 241–254. URI: [http://cda.psych.uiuc.edu/psychometrika\\_highly\\_cited\\_articles/johnson\\_1967.pdf](http://cda.psych.uiuc.edu/psychometrika_highly_cited_articles/johnson_1967.pdf)
- Kawaguchi, Yuji and Fumio Inoue (2002) Japanese dialectology in historical perspectives. *Revue belge de Philologie et d'Histoire* 80(3): 801–829. doi: <https://doi.org/10.3406/rbph.2002.4642>
- Kessler, Brett (1995) Computational dialectology in Irish Gaelic. In: *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, 60–67. Dublin: EACL. URI: <https://arxiv.org/pdf/cmp-1g/9503002.pdf>
- King, Benjamin (1967) Step-wise clustering procedures. *Journal of the American Statistical Association* 69: 86–101. doi: <https://doi.org/10.1080/01621459.1967.10482890>
- Kokuritsu Kokugo Kenkyujo (NLRI) (1966) Nihon gengo chizu kaisetsu: Hôhō [Introduction to the linguistic atlas of Japan: Methodology]. In: Kokuritsu Kokugo Kenkyujo (NLRI) (ed.), *Nihon gengo chizu* (Linguistic atlas of Japan). Tokyo: Printing bureau, Ministry of Finance.
- Krijthe, Jesse (2015) *Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation*. <https://cran.r-project.org/web/packages/Rtsne/>. R package version 0.13.
- Kruskal, Joseph. B. and Myron Wish (1978) *Multidimensional Scaling*. Sage University Paper Series on Quantitative Applications in the Social Sciences 07–011. Newbury Park: Sage Publications.
- Kumagai, Yasuo (ed.) (2013a) Daikibo hôgen dêta no takakuteki bunseki seika hôkokusho: Gengo chizu to hôgen danwa shiryô [Analyzing large-scale dialectal survey data from multiple perspectives]. Tokyo: Kokuritsu Kokugo Kenkyujo (NINJAL).
- Kumagai, Yasuo (2013b) Development of a way to visualize and observe linguistic similarities on a linguistic atlas. *Working papers from NWAV Asia-Pacific* 2. URI: <https://www2.ninjal.ac.jp/past-events/nwavap02/Kumagai-NWAVAP2-2013.pdf>
- Kumagai, Yasuo (2016) Developing the Linguistic Atlas of Japan Database and advancing analysis of geographical distributions of dialects. In: Marie-Hélène Côté, Remco Knooihuizen, John Nerbonne (eds.), *The Future of Dialects. Selected Papers from Methods in Dialectology XV*, 333–362. doi: <https://doi.org/10.17169/langsci.b81.159>
- Matthews, Peter Hugoe (1997) *The Concise Oxford Dictionary of Linguistics*. Oxford Paperback Reference. Oxford: Oxford University Press.
- Nakamoto Masachie (1983) *Ryukyu goishi no kenkyu* [Study on Ryukyuan lexicological history] Tokyo: San'ichi shobo

- Nerbonne, John and Wilbert Heeringa (1998) Computationale vergelijking en classificatie van dialecten [Computational comparison and classification of dialects]. *Taal en Tongval, Tijdschrift voor Dialectologie* 50(2): 164–193. URI: <https://www.let.rug.nl/~nerbonne/papers/tetv99.pdf>
- Nerbonne, John, Wilbert Heeringa and Peter Kleiweg (1999) Edit distance and dialect proximity. In David Sankoff and Joseph B. Kruskal (eds.), *Time Warps, String Edits, and Macromolecules. The Theory and Practice of Sequence Comparison*, 2nd, v–xv. Stanford: CSLI.
- Nunnally, Jum C. and Ira H. Bernstein (1994) *Psychometric theory* (3rd ed.). McGraw-Hill.
- Onishi, Takuichiro (2002). Dialect research at the National Institute for the Japanese Language. *Dialectologia et Geolinguistica (DiG)* 10: 87–96. doi: <https://doi.org/10.1515/dig.2002.2002.10.87>  
Also available at: <https://www2.ninjal.ac.jp/takoni/English/DRNIJLA.pdf>
- Séguy, Jean (1973) La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane* 37: 1–24.
- Togerson, Warren S. (1952) Multidimensional scaling: I. Theory and method. *Psychometrika* 17(4): 401–419. doi: <https://doi.org/10.1007/BF02288916> Also available at: <http://www.galileoco.com/literature/togerson52.pdf>
- Swadesh Morris (1952) Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96(4): 452–463. URI: <https://www.jstor.org/stable/3143802>
- van der Maaten, Laurens (2014) Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* 15(93): 3221–3245. URI: <https://jmlr.org/papers/v15/vandermaaten14a.html>
- van der Maaten, Laurens and Geoffrey Hinton (2008) Visualizing Data Using t-SNE. *Journal of Machine Learning Research* 9(86): 2579–2605. URI: <https://www.jmlr.org/papers/v9/vandermaaten08a.html>
- Ward Jr, Joe H. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58: 236–244. doi: <https://doi.org/10.2307/2282967>
- Wieling, Martijn, Jelena Prokić and John Nerbonne, J. (2009) Evaluating the pairwise alignment of pronunciations. In: Borin, L. and Lendvai, P., editors, *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education, (LaTeCH - SHELTeR 2009). Workshop at the 12th Meeting of the European Chapter of the Association for Computational Linguistics. Athens, 30 March 2009*, 26–34. URI: <https://aclanthology.org/W09-0304.pdf>
- Wieling, Martijn (2012) A quantitative approach to social and geographical dialect variation. Doctoral dissertation, University of Groningen. URI: <https://hdl.handle.net/11370/cd637817-572f-4826-98c1-08272775fb64>
- Yanagita, Kunio (1943) *Kagyu Ko* [Reflections on Snail] Tokyo: Sogensha
- Yarimizu, Kanetaka (2007) Hogen bunpo zenkoku chizu ni okeru kyotsugoka no jokyo [The situation of common language in GAJ]. *Nihongogaku* 26 (11): 112–119.

## Publication history

Date received: 26 September 2023

論文

## 外行語の計量地理言語学 Google trendsデータの因子分析と初出年

井上史雄  
東京外国語大学

### Computational Geolinguistics of Lendword Factor Analysis of Google Trends Data and Year of First Appearance

INOUE, Fumio  
Tokyo University of Foreign Studies

**Abstract:** This paper presents a comprehensive analysis of the global trends of "lendwords" (Gaiko-go), which are words that have gone abroad from Japan. In the first half of the paper, the global distribution and historical transition of the four words "kimono, shogun, sakura and sukiyaki," which were surveyed in the "International Census on Attitudes toward Japanese (ICATJ)" were confirmed by adding subsequent information from the Internet. We reconfirmed that the item selection was appropriate for understanding global distribution at present.

In the second half of the paper, we analyzed data acquired from Google Trends on the use of more than 100 Japanese lendwords in 180 countries and regions around the world to uncover the current status of the entire world. The data were graphed and discussed word by word and country by country.

Factor analysis was applied to the numerical matrix from the graphs to examine lexical and national trends. The first factor, traditional culture, was found in English-speaking countries, the second factor, modern pop culture, was found in Latin American Spanish-speaking countries, and the third and fourth factors, modern culture, were found in Southeast Asian countries. Drawing the year of first appearance from OED (Oxford English Dictionary), the older words that first appeared before the 19th century are more common in English-speaking countries, while the newer words that first appeared in the 20th century are more common in countries in South America and Southeast Asia. This corroborates the tendency over all the world's lendwords for word borrowing to occur in lexical groups as a result of cultural exchange.\*

**キーワード：**外行語；借用語；言語間方言学；グーグルトレンド  
**Keywords:** lendword; loanword; interlingual dialectology; google trends

---

井上史雄(2023)「外行語の計量地理言語学 Google trendsデータの因子分析と初出年」『地理言語学研究』3: 45-91 doi: <https://doi.org/10.5281/zenodo.8437054>

\* 因子分析には柳村裕・半沢康両氏のお世話になった。結果の解釈および文章化の責任は筆者にある。本研究はJSPS科研費基盤研究(B) 16H03420 井上史雄「公共用語の地域差に関する社会言語学的総合研究」の助成を受けた。

## 1. 外来語と外行語の先行研究

### 1.1. 外来語・外行語研究史

借用語 borrowing には、方向性の視点の置き方によって、外来語 loanword のほかに外行語 lendword がある。この論文では外行語を取り上げる。言語の借用のうち、日本語の外来語の研究は従来から盛んだった(煤垣 1963、石綿 1985、Daulton 2022)。少数例を羅列し、個別的に論じる著もあるが、多数例で大きな傾向を把握する努力もなされていた。計量的研究も早くからあり、視覚化も進んでいた(林 1982)。量だけでなく質として意味分野、階層構造(上位概念)に着目した研究もある(金 2014)。

日本語が外に出る現象は「外行語」と名付けられた(三輪 1977)。便利なのでその後も用いられた(井上 2001、2007.10、2012.10、2013.3.21、2013.3.31、2013.12、2016.3、井上・柳村 2015.3、2015.9、石綿 2005、エシュバハ・サボー 2005、中川 2005、永島 2013、小林 2014)。英語では貸与語 lending, lend word または輸出語 export word と呼べる。外行語の先行文献としては以下があり、「国際日本語」(原口・原口 1998)、「外国語になった日本語」(加藤・熊倉 1999)、「英語に入った日本語」(早川 2003)、「英語になった日本語」(早川 2006)、「日本語からの借用語」(彭 2013)など、様々な名称がそれぞれ先行研究を十分に参照せずに用いられている(Cannon 1996、Hiramoto 2010、ハナシロ・島田 2011、Long & Imamura 2013)。研究史については、井上(2012.10、2013.3)がある。

日本語外行語の世界への普及状況を広く見渡すことは、従来は不可能に近かった。転換点になったのは、「日本語観国際センサス」(新プロ「日本語」総括班編 1999、国立国語研究所 2002)における日本を含む世界 28 の国と地域での面接調査の結果で、「キモノ、ショーグン、サクラ、スキヤキ」4語の世界分布が明らかになった。数表で元データが公開されており、井上(2001.8)に世界地図の形で紹介された(図 1-1 に再掲する)。外行語が欧米先進国に多いこと、日本中心に地理的近接効果(世界的な周圏論)が見られることが示された。同じ規模の面接調査を繰り返すのは、費用から言って、困難である(オムニバス調査として同じ4語を入れるだけなら可能性がある)。

しかしその後のインターネット検索技術の向上によって、個人でも(もっと大量の)データを作成することが可能になった。年代順にたどろう。

井上(2011.2)では、Google map による世界地図や日本地図を示した。

井上(2012.2)では、Google insights を用いて「日本語観国際センサス」の「キモノ、ショーグン、サクラ、スキヤキ」4語の世界地図を示し、81語 124 国家の外行語の数値行列を作成して、単語順と国家順でソートしてグラフ化した。数表作成技法も紹介している。

井上 (2012.10) では研究史を略述するとともに Google map と Google insights の 120 語のデータを紹介し、OED 初出年と照合して、明治初期初出の語が多いことを確かめた。

井上 (2013.3.21) では Google insights の 120 語のデータを 117 語に精選し、2004 年から 2011 年にかけての 7 年間の推移 (トレンド) をグラフ化した。企業名、商品名の使用率が桁外れに多く、全体として増える傾向がとらえられた。言語間方言学にも言及した。

井上 (2013.3.31) ではハワイ (日系人) の外行語を多様な資料によって位置付け、方言的背景などを指摘した。

Inoue (2013.4) では日本語からの外行語だけでなく、英語、スペイン語、中国語、韓国語からの外行語 (輸出語) のデータを提示したが、論文としては未発表である。

井上 (2013.12) では外行語研究法を、(1) しみの目 (アームチェア)、(2) 虫の目 (フィールドワーク)、(3) グラフと地図 (世論調査) に分け、さらにインターネット検索技法を (1) 蠅の目図 Google street view、(2) 鳥の目図 Google maps、(3) 魔女の目図 Google trends、(4) 歴史家の目図 Google Ngram viewer に分類して、実例を紹介した。

井上・柳村 (2015.3) では Google trends による 114 語 145 国家の外行語の数値行列を 59 語×125 国家に精選して因子分析を適用し、第 1 因子=現地定着・食品因子 (中南米) と第 2 因子=現代若者文化因子 (西欧) を取り出した。また代表的な 18 語の世界地図を提示し、分布が多様でありながらいくつかのパターンに分かれることを示した。

井上・柳村 (2015.9) はその発展で、各国の貿易輸出額のデータと外行語の数が比例関係を示すことを見た。また因子分析の結果を国ごとに第 10 因子まで細分して分析した。

Inoue (2022.1) では、Google trends でダウンロードしたデータを貿易輸出額と対比させ、文字と比べ、因子分析にかけた結果を国ごとにまた単語ごとに散布図にして、各因子の意味を探った。また OED との照合にも言及した。一連の日本語論文の英語版要約と言える。

世論調査を含め、インターネット検索結果や各地の景観写真など、複数の外行語調査の分布パターンはほぼ対応しており、各種の調査方法の信頼性を相互に補強する (井上 2017.10、2018.8、2021.12) 。

本稿は以上の基盤の上に立つものとして執筆された。



## 1.2. 日本語観国際センサスの外行語

「日本語観国際センサス」は、1997年1月から1998年8月にかけて世界28か国で面接調査として実施された。以下のインターネットサイトで基本情報が公開されている。[https://mmsrv.ninjal.ac.jp/n\\_census/research/](https://mmsrv.ninjal.ac.jp/n_census/research/)

外行語に関する質問文の基本は以下の日本語だが、日本では調査されていない。

Q38 次の日本語のことばを見たり、聞いたりしたことがありますか。次の中から、見たり、聞いたりしたことがあるものすべてお答えください。(MA)

1. すきやき 2. さくら 3. きもの 4. 将軍 5. いずれもない 無回答  
(上記のことばを、各国のことばの表記に直してください。)

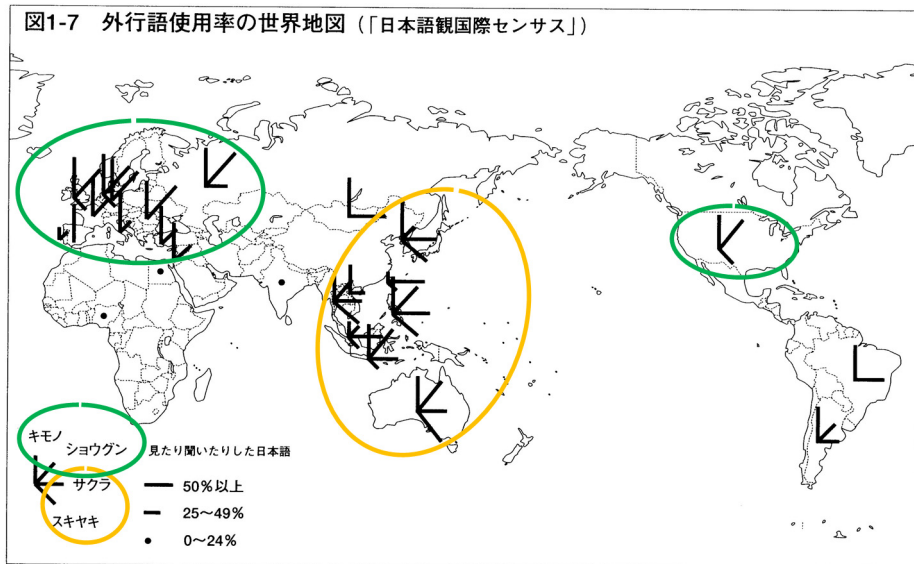


図1-1 4語の世界的分布 日本語観国際センサス

図1-1に「日本語観国際センサス」4語の世界地図を増補して再掲する(井上2001:8)。使用率を3段階に分け、棒の方向で4語を判別した。欧米と東アジアに多いが、キモノ、ショウグンは欧米に多く、サクラ、スキヤキは東アジアに多い傾向がある。アフリカ、西アジアには普及していない。

「日本語観国際センサス」の元データを使うと、さらに細かい分析ができる。図1-2abcdに4語の国家別、世代別の折れ線グラフを掲げる。図1-3に示すように、普及の新古、程度によってパターンが違うので、古くから普及していたと思われる語を最初に掲げる。

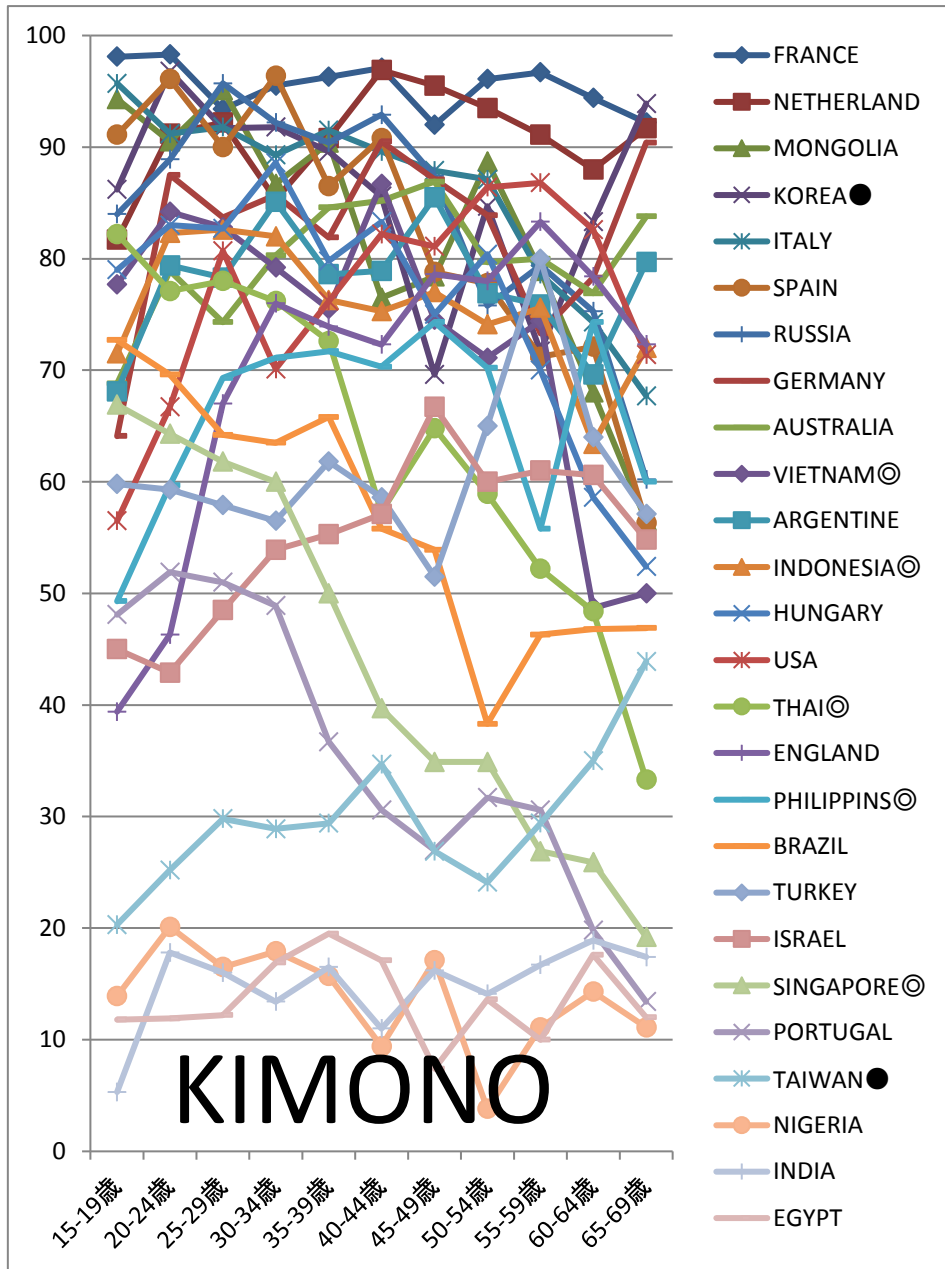


図1-2a キモノ 日本語観国際センサス国家別の世代別分布

**図1-2a**キモノは、古くから普及していたと見られ、欧米の諸国家では8割以上に使われる。一方世代差は、国によって違うが、若い世代に多い国家も目立ち、現在も生命力がある。

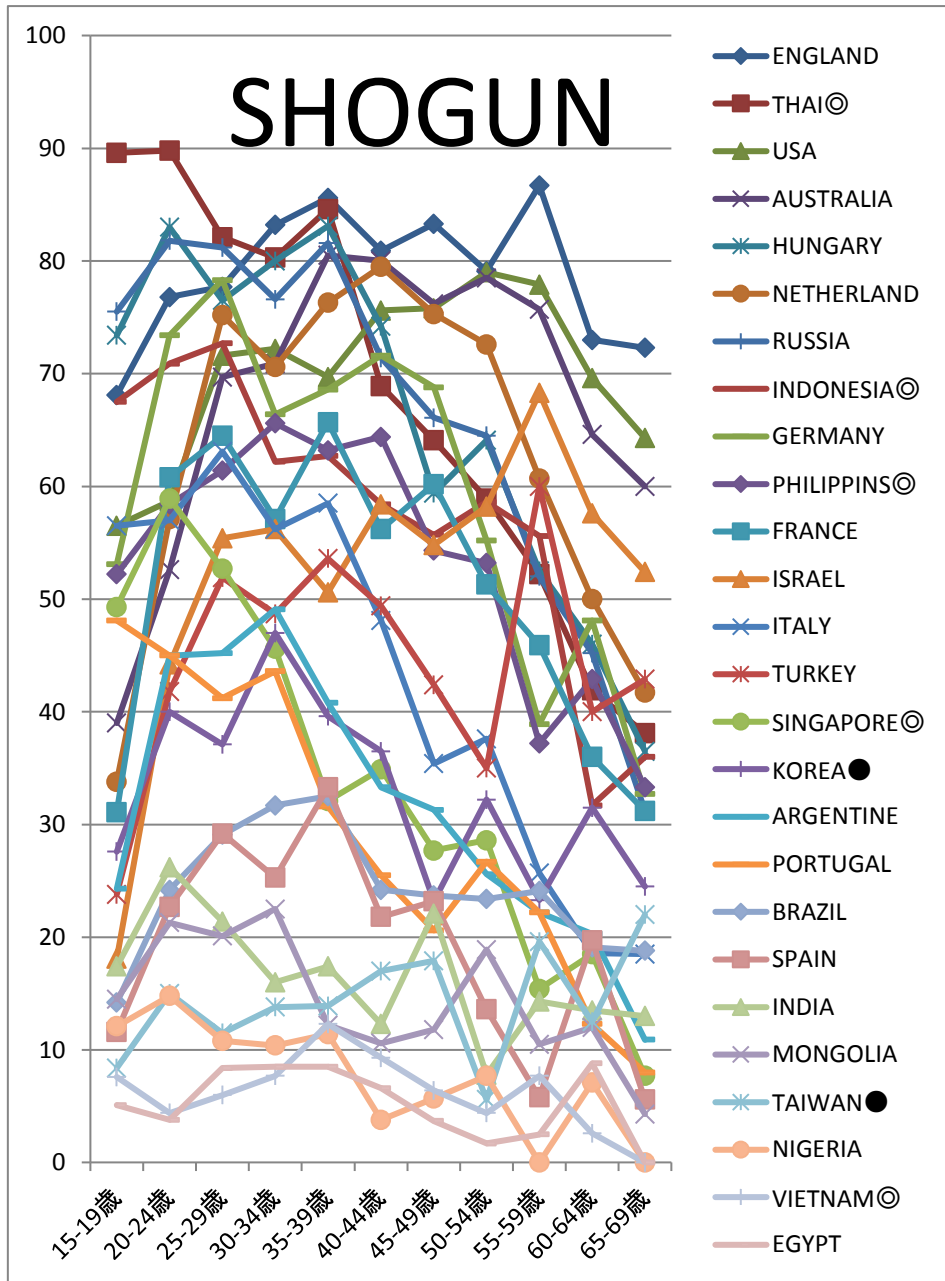


図1-2b ショーゲン 日本語観国際センサス国家別の世代別分布

**図1-2b** ショーゲンは、欧米の諸国家では8割以上に使われる。世代差は、国によって多様だが、壮年世代に多い国家もある。【後述のように、小説・映画の題名になったために、時代・世代による違いが大きい。】

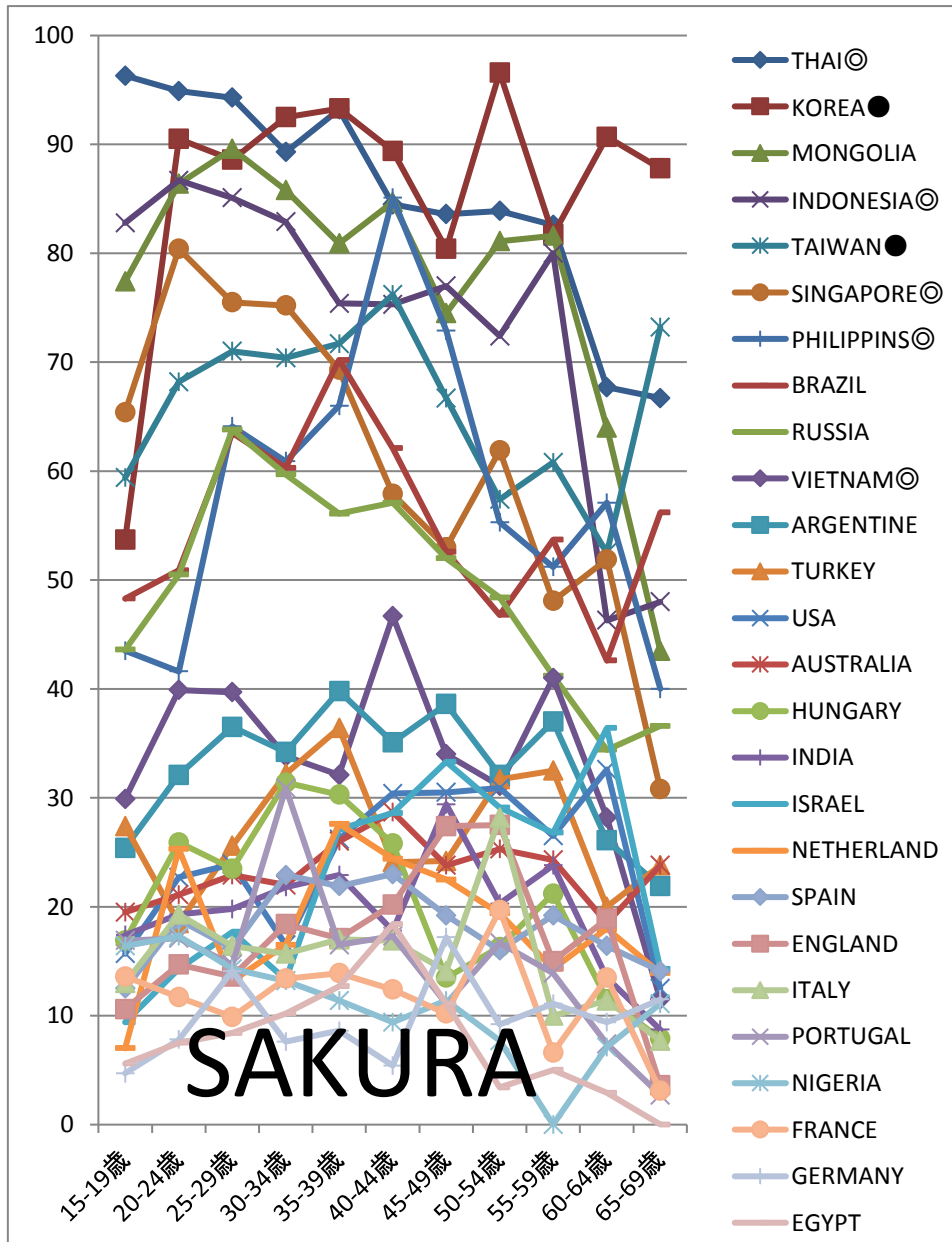


図1-2c サクラ 日本語観国際センサス国家別の世代別分布

**図1-2c** サクラは、2割近くに使われる国家が多い。実物の移植とともに名が知られた可能性がある。世代差は、壮年世代に多い国家もある。【後述図3-4によると、ポップカルチャーの語に近く位置付けられる。】

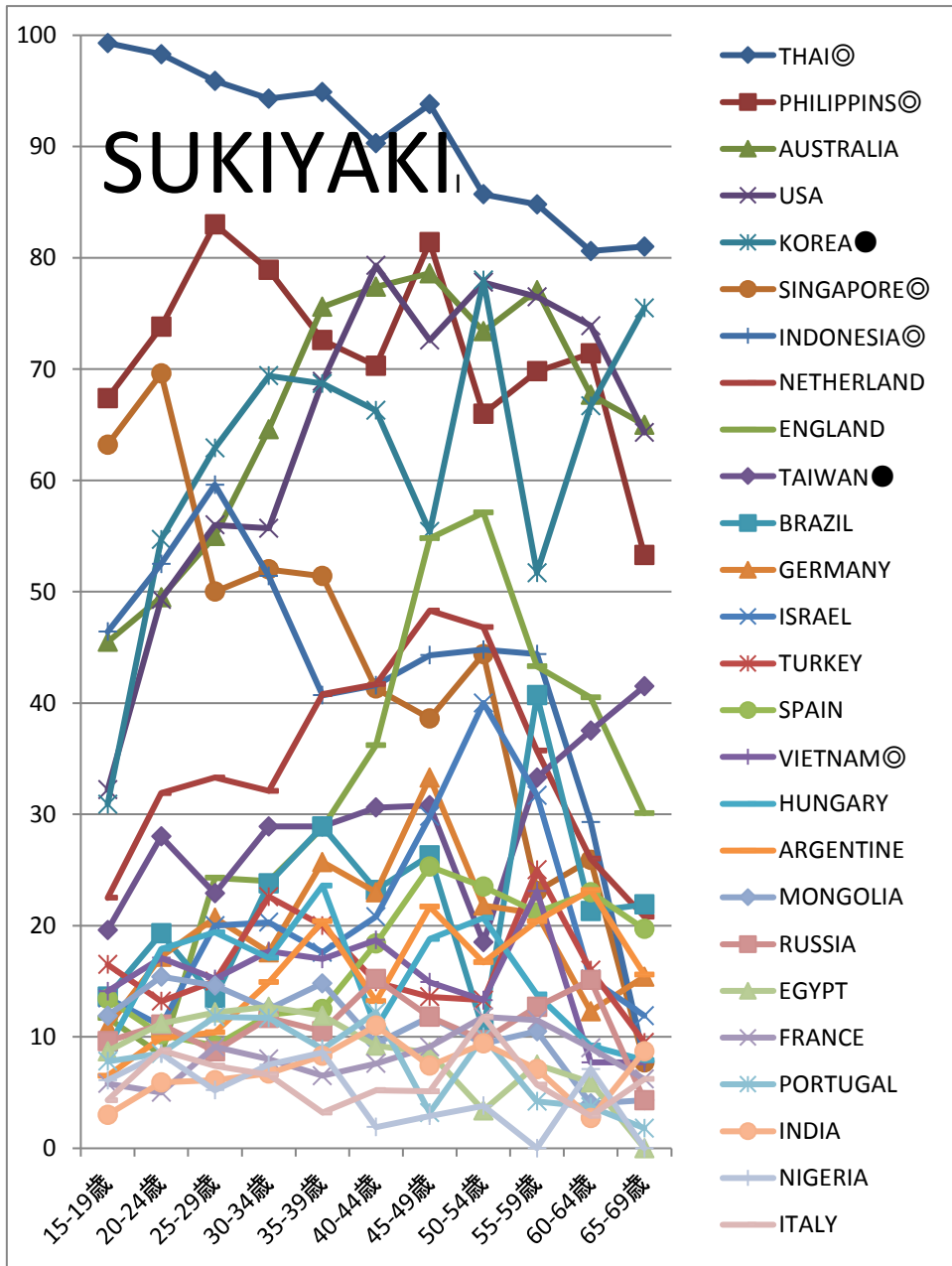


図1-2d スキヤキ 日本語観国際センサス国家別の世代別分布

**図1-2d** スキヤキは、1割近くに使われる国家が多い。東南アジアに目立つ。世代差は、壮年以上に多い国家もある。【後述のように、近代に生まれた調理法で、商業的要因による違いが大きい。】

# 日本語観国際センサス census 年齢別集計 age differences 全国家 all the nations

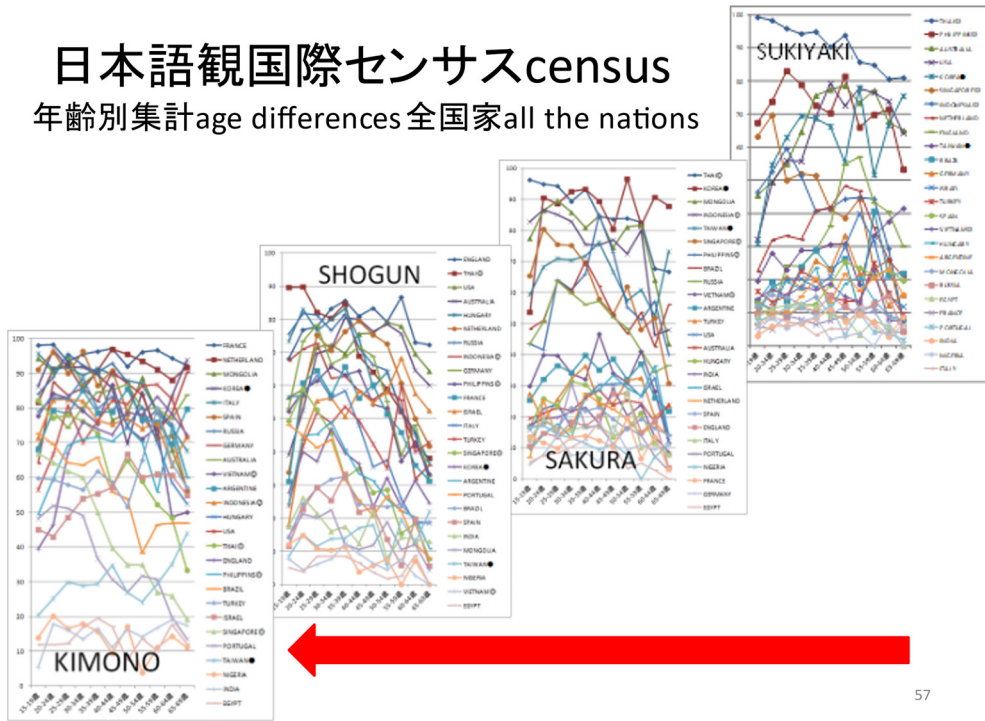


図1-3 4語の世代別分布 日本語観国際センサス国家別

図1-2abcdの、国家別、世代別の折れ線グラフによると、キモノ、ショーグン、サクラ、スキヤキの順に普及したと、考えられる。普及の新古、程度によってパターンが違うので、古くから普及していたと思われる語を図1-3に示すように、右から並べる。矢印で示すように、最初は右端sukiyakiのように一部の国家に薄く普及する程度だったが<sup>1</sup>、やがてもっと多くの国家に広く使われ、さらに多くの国で盛んに使われるようになる。

矢印を逆に左から右にたどると、普及しつくした語から途中の段階の語をたどることになる。さらに左には全国家全国民に普及した語がある。SONY HONDA TOYOTAなどの企業名がこれに近い（井上2013.3.21）。すべての語が定速で広がるわけではなく、分野により、実物の普及に影響されて、急速に広がることもある。逆に右端にも至らない、外行語になっていない日本語の単語は数多い。

<sup>1</sup> スキヤキは文明開化以降に牛鍋とともに広がった新しい食文化である。中国地方ではスキの（部品の）方言へカを使ったためにへカ焼き、へカ鍋とも言った。広島方言の影響の大きいハワイの日本語方言でもheka heka hekanabeが使われた（Inoue 1991.3、井上2013.3.31）。日本語方言起源の外行語の一例である（井上2011.1）。後述sukiyakiのOEDの初出1920年より早い可能性がある。

### 1.3. ウェブによる言語研究

以上のように、「日本語観国際センサス」により、外行語の全世界の分布を確認できた。日本経済の好調だった時期に計画された大規模面接調査の成果である。

インターネット情報の活用により、外行語についてグローバルな視野が開けた。ウェブの検索システム発達により、言語研究の視界が広がった。ウェブの情報を言語研究のコーパスとして扱う技法については、荻野（2014）があり、新しい研究資料として活用が勧められている。もっとも学問的ではないという評価があり、「ウェブ検索の利用が何となく安直でアマチュア的な印象を与える」（荻野・田野村2011、第6巻 p.90）。しかしグーグルトレンドGoogle trendsについては、荻野・田野村（2011）の講座7冊のいくつかの章で言及がある。また医学・経済学で実用としての活用が盛んで<sup>2</sup>、マーケティングでも活用されており、展望論文でも確認できる（Mavragani et al. 2018）。グーグルNグラムビューアーによる文化史も新知見を提供した（Michel et al. 2010、Younes and Reips 2019）。日本語についても、グーグルトレンドのいくつかの適用例がある（井上2012.2、2012.10、2013.3.21、2013.12、2016.12、上村2014、有働2016）。限界として、元データがグーグルトレンドでの検索語であるために、人々の興味のある方を反映し、実際の日常生活での使用を直接に反映するわけではない<sup>3</sup>。そもそも「検索エンジンは、基本的には「事柄」を検索する」（荻野・田野村2011、第6巻 p.59）。しかしビッグデータであるが故の利点があり、同一のデータ間での比較では問題が少ない。

インターネット検索の機能拡大により、外行語の世界的規模の研究が可能になった。世界の全体像を把握するために、グーグルトレンド（当時の名称Google insights）の2011年12月にダウンロードされた数値データを様々な形で扱うことができた（井上・柳村2015.3、2015.9）。本論文での118語は、井上（2012.2）のリストを拡充し、先行研究で指摘された語を集めたものである<sup>4</sup>。これまでに国ごとまたは単語ごとの合計値が扱われて、国家の経済条件などとの関連が解明された。

データの性格を示すために、「日本語観国際センサス」と同じ4語の世界地図を掲げよう<sup>5</sup>。4語をGoogle trendsで 2022.04.23にダウンロードし、地図と約20年間の変遷（トレンド）を図1-5以下に掲げる。図1-2abcdの「日本語観国際センサス」の図と照合すると、様々な違いが浮かび上がる。

<sup>2</sup> インフルエンザについての検索が患者数に比例し、商品についての検索がその売り上げと連動するなど、実際に活用されており、日常語での使用を反映すると考えられる。辞書代わりに意味用法やつづりを知るために検索することもあるが、分かりきった語については、日常の使用頻度数よりも低くなる。ただし欠陥の恐れがあるからと言って適用しない状態からは進歩がない。

<sup>3</sup> Google trendsや Google Ngram viewerに限らず、新しいデータは実際に分析してみて、欠陥に気づき、さらなる可能性が開ける。

<sup>4</sup> ダウンロードした125語からSONY TOYOTAなどの固有名詞を除き、スペイン語動詞活用形と語形が一致するsoyと誤植のshinkanseiを省き、118語を扱う。

<sup>5</sup> Google insights と呼ばれていた時期（2011.12.10）にダウンロードしたものである。



#### 1.4. 外行語4語の世界的分布

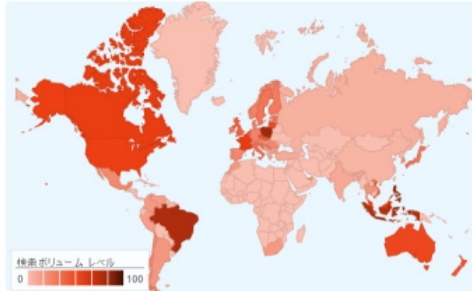


図1-4a キモノの世界的分布 Google insights 2011.12.10

Kimonoは東南アジアとブラジルおよび欧米に多く使われている。国ごとの差が大きい。【**図1-1**、**図1-2a**、**図1-5ab**でも同様である】



図1-4b ショウグンの世界的分布 Google insights 2011.12.10

Shogunは東南アジアと中南米でやや多く使われている。【**図1-1**、**図1-2a**でも同様である。**図1-5ab**では国ごとにずれがあり、20年の推移が見られた。】

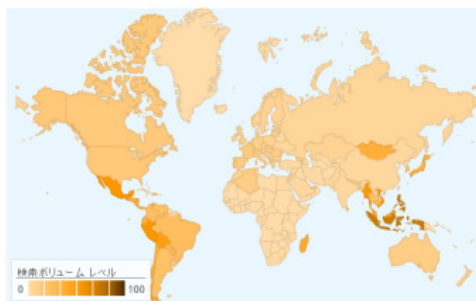


図1-4c サクラの世界的分布 Google insights 2011.12.10

Sakuraは東南アジアと中南米でやや多く使われている。アフリカ、西アジアでも使われることが他の3語と違う。【**図1-1**、**図1-2a**でも同様である。**図1-5cd**では国ごとの違いが薄れ、20年の推移が見られた。】



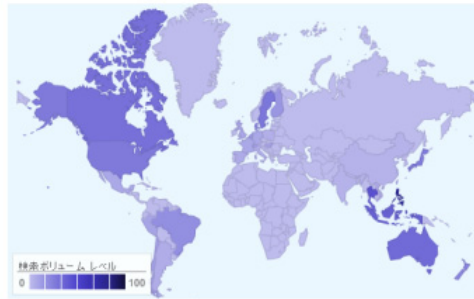


図1-4d スキヤキの世界的分布 Google insights 2011.12.10

Sukiyakiは東南アジアと北米でやや多く使われている。国による違いが大きく、ヨーロッパでは一部で使われるのみである。アフリカ、西アジアでは使われない。【図1-1、図1-2aでも同様である。図1-5cdでは国ごとの違いが薄れ、20年間の普及が見られた。】

図1-5では4語のGoogle trendsにおける推移を示す。2004年から2022年までの約20年間の推移（世界全体）が分かる。手間を惜しまなければ、国ごとに20年間の推移が分かり、国内の地域差も分かり、例えばドイツ国内の大きな地域差が見える（Inoue 2022.1）。

世界全体として一番普及しているのはsakuraである。人口の多い国家で使われるからか。大きな波が数年ごとにあり、2021年に最大のピークがある。Covid-19で観光が衰退した時期にあたり、今のところメカニズムは不明である。

Shogunは小説と映画の題名として広がり、ミュージシャンの名にも使われた。Shogunは2011年3月ころにピークがあり、東日本大震災との関係が考えられる。



図1-5 4語の推移 Google trends 2022.04.23

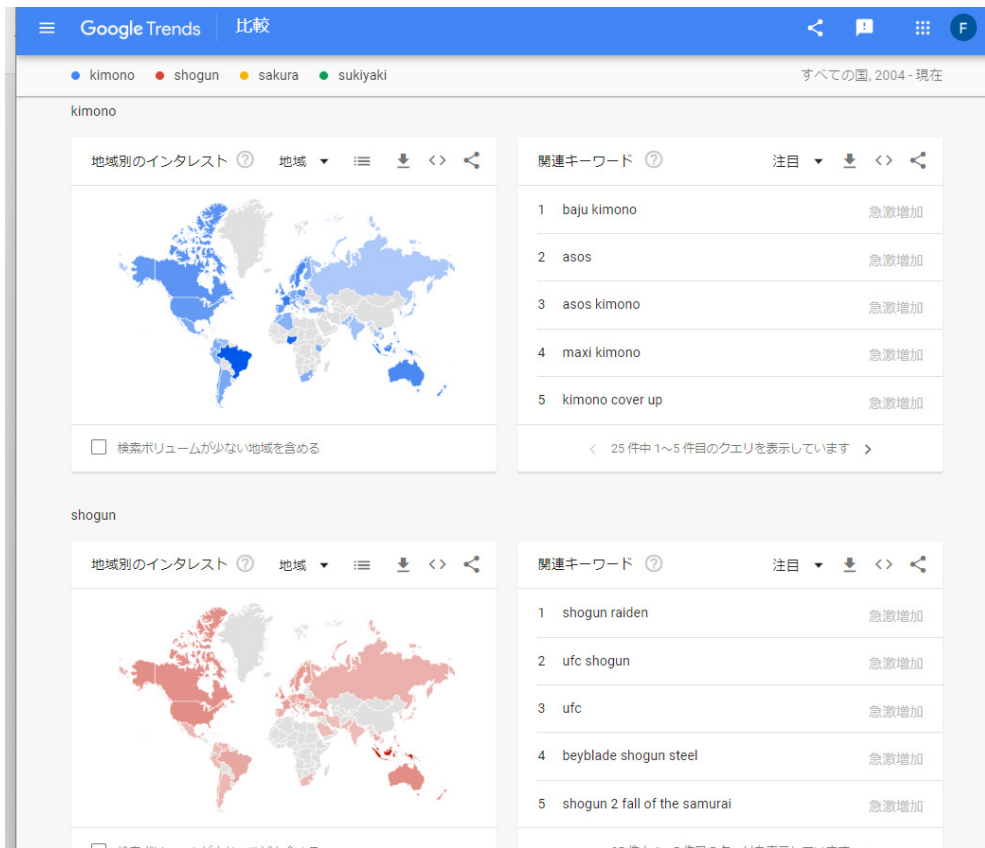


図1-5ab キモノ・ショーグンの世界的分布 Google trends

Kimonoはshogunと同程度の使用率だが、季節変動を伴いつつ、増えている。Sukiyakiは世界全体の使用率が小さい。図1-2dでも読み取れる。しかし図1-4dの地図の濃淡（国による違い）からは別の印象を受ける。

以上、2004年からのトレンド・推移について考察した。以下では2011年から10年後の世界的分布の違いを考察する。ただしすでに【 】内で指摘済である。

図1-5abではキモノ・ショーグンの世界的分布を示す。キモノで欧米とアジア・アフリカとの違いが大きいのは、10年前と同様である。ヨーロッパやアフリカで個々の国家の使用率の変遷が見られる。

ショーグンでも欧米とアジア・アフリカとの違いが大きいのは、10年前と同様である。ただインドネシアが目立たなくなった。

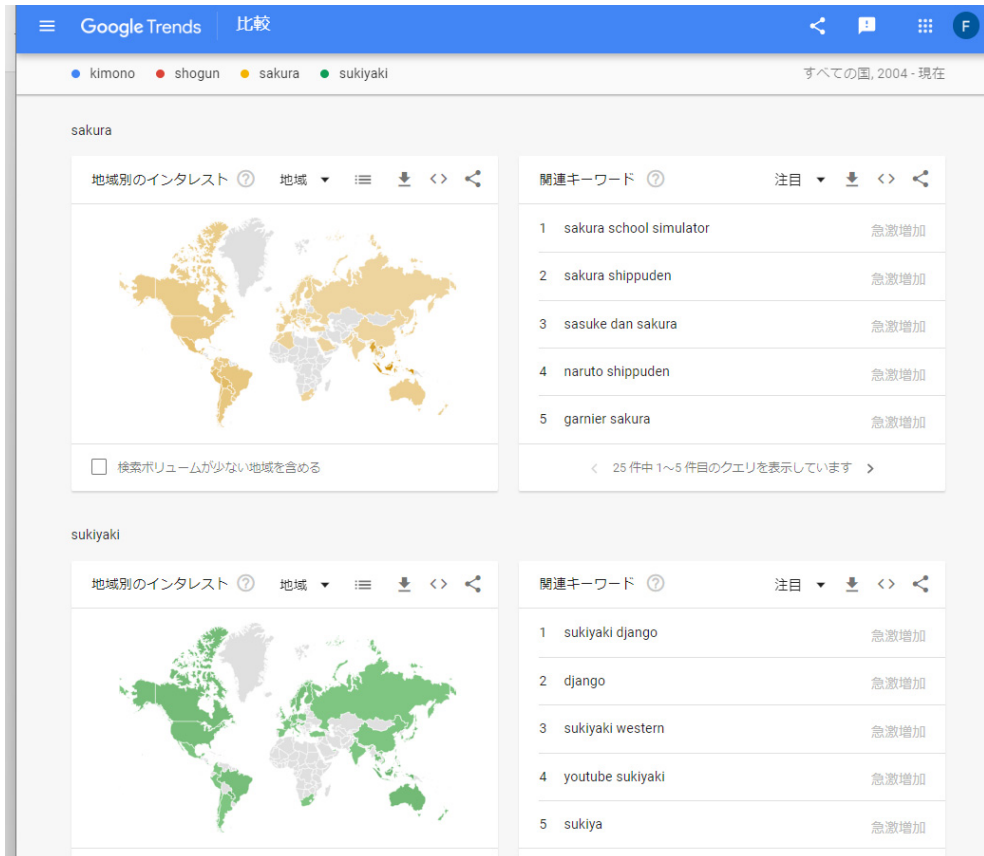


図1-5cd サクラ・スキヤキの世界的分布 Google trends

図1-5cdにサクラ・スキヤキの世界的分布 (Google trends 2022.04.23) を示す。サクラで欧米とアジア・アフリカとの違いがあるが、10年前に比べて個々の国家の使用率の差が目立たなくなった。

スキヤキで欧米とアジア・アフリカとの違いが大きい、10年前に比べて、ロシアやヨーロッパ各国に広く普及した。背景として焼肉店が世界各地に出店したことが考えられるが、10年間の推移は1-5では目立たず、単独のグラフで見ても10年間にやや増えているだけである。

まとめると、10年ほど経ってGoogle trends 2022.04.23で再検索し、2004~2022の20年弱の合計で見ると、世界の地域差は薄れた。時代差は固有の変動があり、キモノ・サクラは着実に増えた。ショーグンは一時増えたが、スキヤキは横這いである。たった4語であるが、それぞれの語に固有の歴史があるように見える。ただし方言地理学で多数の語を計量的に扱くと、一定の傾向が浮かび上がる。個別に注意を払うか、全体の流れを見るかは、研究の目的や立場による。両方が必要であろう。

## 1.5. Google Ngram viewerによる4語の200年間の普及

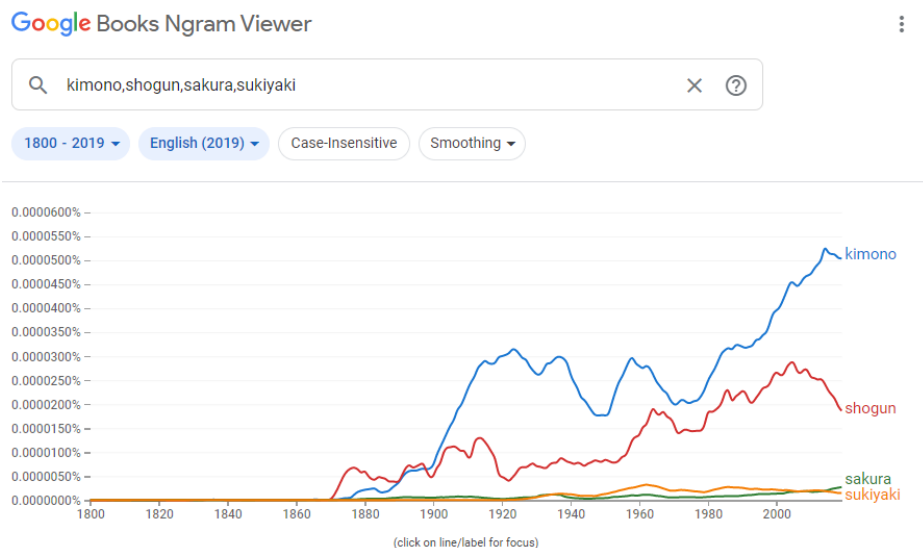


図1-6a Google Ngram viewerによる4語の普及（英語）

外行語使用の歴史的変動について、別の方法で確かめることができる。Google Ngram viewerによれば200年間（または500年間）のGoogle booksで電子化された文献内の用例が分かる。図1-6aに4語について英語文献全体を示す。明治維新1868年ころからkimono shogun sakura sukiyakiの順に使われはじめ、使用率の順番もほぼ一致する（sakura sukiyakiは途中で交代を見せるが）。一般論として、早く使われた語ほど多く使われるようになる。1960年前後と2010年代にピークがあるのは、日本への関心の（一時的）拡大と関連がある。江戸時代以来の風俗・制度・風土・習慣に関わる直訳不能の「特有語」（井上2000.2、2011.11）がまず文献に表れた。2.3で述べるOEDの初出年の傾向とも一致し、借用語・外来語の一般傾向とも一致する。

英語と米語の違い、フィクションだけのグラフも出るが、全体像は変わらない。フランス語、イタリア語、スペイン語、ドイツ語での使用状況も分かり、大勢は英語と似る<sup>6</sup>。SAE (Standard Average European) (Whorf 1956)というとらえ方は、外行語にも適用されそうだが、個々の言語による違いもある。図1-6bに示すように、ドイツ語だけはsakuraの使用が1850年代から見られる点で、他のヨーロッパの諸言語と違う。書籍における日本文化の紹介（言及）が言語によって違うことを意味する。

この4語の順番は、図1-5のGoogle trendsの最近の順番（sakura kimono shogun sukiyaki）と一致しない。Sakuraが検索で急上昇したためである。文献の中の用法とGoogle検索の必要とは違うのだろう。この点はGoogle trendsの限界を示す。

<sup>6</sup> しばらくはデータに大きな変更がなく、希望者は確認できるので、ここには示さない。

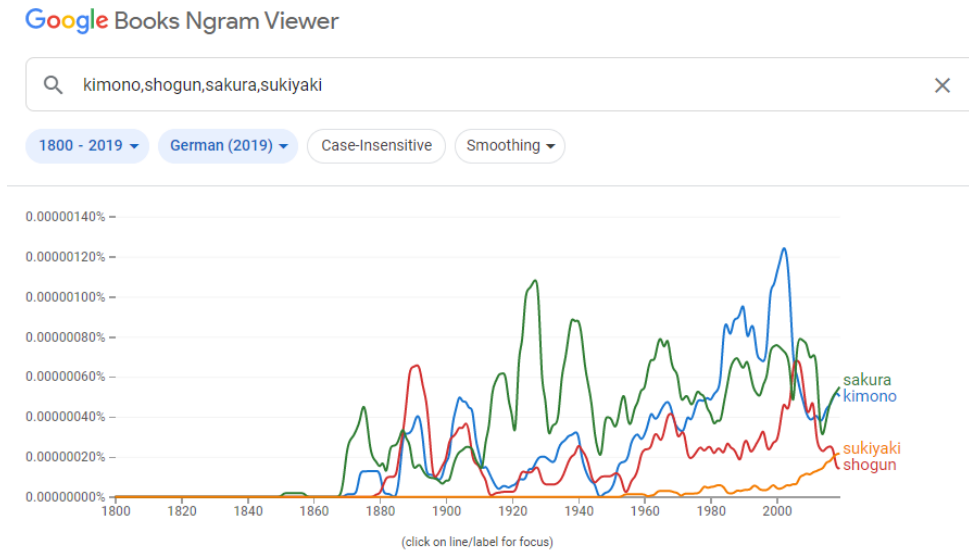


図1-6b Google Ngram viewerによる4語の普及（ドイツ語）

### 1.6. グーグルトレンドの外行語の計量地理言語学

以上では、「日本語観国際センサス」で取り上げた4語について、インターネット検索によって20年間および200年間の推移と世界の地域差について考察した。世界の使用状況を知るのに適切な項目選択だったことを再確認した。

以下ではGoogle trendsによって、多数の外行語の数値データをダウンロードした結果を紹介する。ダウンロードの技法、個々の語の世界分布図の一部、データの数表、単語ごとの順番のグラフ、国家ごとの順番のグラフなどについては、先行文献を見られたい（井上2012.2）。

## 外行語 lend word 114語 3/14/25/145国 Google trends

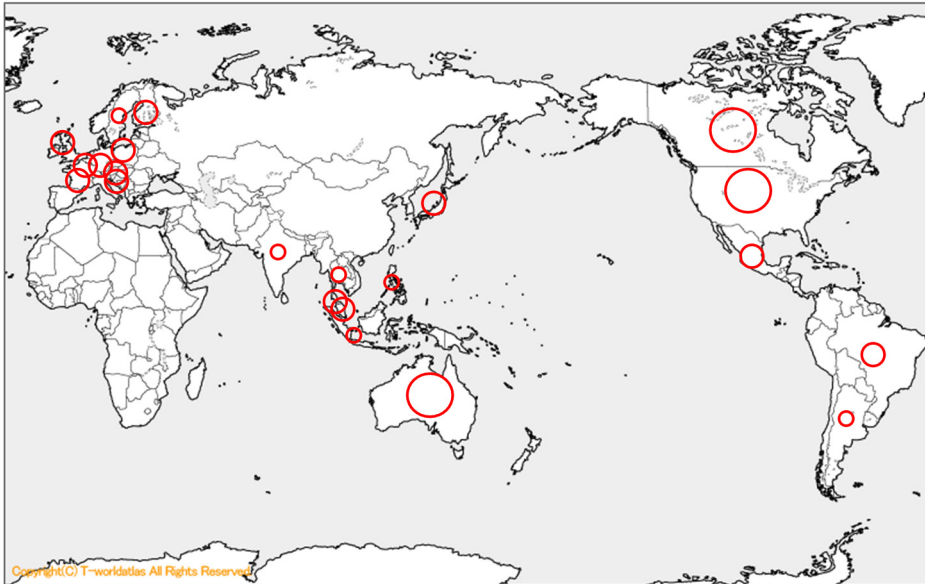


図1-7 Google trendsによる国ごとの118語合計値の世界分布

まず図1-7で外行語114語の全体像を世界地図の形で提示する。国家ごとに外行語の数の多さの順に配列し、最大の3か国、次の14か国、その次の25か国に3種の大きさの円を付けた。その他の145か国は無印である。最大の3か国は(英)米語圏と見て良い、次の14か国はヨーロッパ、東南アジア、中南米に分布する。その次の25か国は、その周辺に広がる。これまでに見た「日本語観国際センサス」の4語での世界分布パターンとほぼ一致する。4語の選択の適切さを補強し、インターネット情報の有効性を物語る。

世界全体として、北米、オセアニアなどの英語国に多く、また東南アジア、ヨーロッパや南米がそれに次ぐという地理的パターンが見えた。本来の意味の計量地理言語学の成果にあたる。地理的距離にもよるが、貿易量などの経済交流も作用し、一方アルファベット対民族文字などの表記上の問題、音声的な借用語を用いるか対意味論的な翻訳語を用いるかなどの言語(政策)上の問題も、考えられる。影響関係を確認するのに、世界地図を照合する技法も考えられるが、外行語のデータは数表になっているので、散布図などによって、相関係数を知る技法も考えられる(井上・柳村2015.9)。さらに進んで、全体に多変量解析法を適用して、パターンを把握する手法もある。本稿ではこれらの計量地理言語学の分析結果を紹介する。

分析対象の国家(地域)の数値はGoogle提供の基準による。また調査語は先行論文(井上2012.2など)および表4-1に紹介されている。



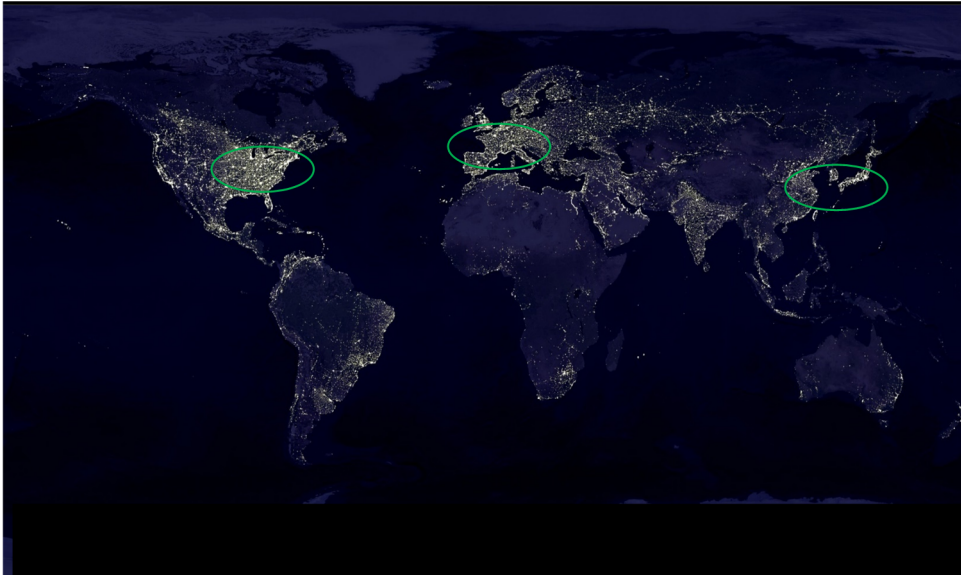


図1-8 夜の明るさの世界分布Earth lights

世界全体として、3か所が様々な現象で立ち表れることは、すでに各所で指摘済みである。その背景に経済的要因があることも論じた（井上2021.12）。世界地図レベルで印象的に示すのが、**図1-8**夜の明るさの世界分布Earth lightsである。人口密度が高く、GDP、GNPで分かるように、経済的に豊かで、文化度、平均寿命など多くの指標が、いわゆる先進国であることを示す。相互の貿易量が多く、人的交流（旅行者、在住者）も多い。多言語表示が多く、日本語からの外行語も多い。夜の明るさは豊かさの指標とも見られ、豊かな社会では言語的交流も盛んになると解釈できる。

明るさと都市化、共通語・公用語の普及などについて、因果関係・相互影響関係のメカニズムについては、井上（2015.5、2021.12）参照。

## 2. ゲーグルトレンドの外行語の因子分析

以上では Google trends による数値データ全体像（の一部）を単純な集計によって見た。第2節ではこのデータに多変量解析法を適用して、内部構造を見る。データに語らせるという発想である。さらに第2節後半では各語形の初出年の情報を追加して、外行語個々の歴史的背景を調べ、論拠を確実にする。

## 2.1. 外行語の国別因子得点Factor score

以下では、多変量解析法によって、どんな国でどんなことばが検索されているかを、総合的に見る。色々な手法があるが、古典的な因子分析にかける<sup>7</sup>。表2-1に4因子の基本情報を掲げる。結論を先取りする形で4因子の解釈も記す。のちの説明文でも【 】内に、後述OED初出年との対応を見た結果を提示する。

元のデータはこれまでに分析したGoogle trendsによる外行語118語の、世界180の国家および地域の（検索に使われた）数値行列である<sup>8</sup>。前述のように59語×125国家に精選して、因子分析を適用した。

表2-1：因子の意味・属性（59語×125国家）

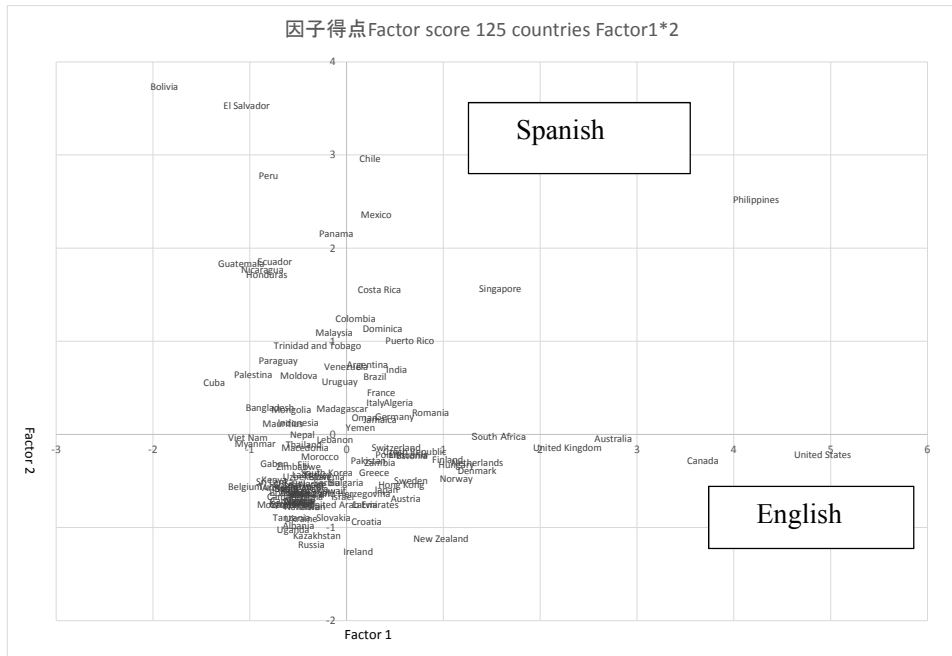
	Factor1	Factor2	Factor3	Factor4
因子得点	.688	.512	.600	.412
因子寄与	9.046	6.501	4.689	4.505
使用国家	英語国	南米スペイン語圏	インドネシアマレーシア	東南アジア
語彙特性	伝統語彙	新語	新古混在	新古混在
初出年ピーク	19 c	20 c	19 c	19 c

まず因子得点を手がかりに、データ行列の国家の次元の構造を把握しよう。**図2-1**に第1因子と第2因子を組み合わせた国別の散布図を示す。横軸**第1因子**プラス側に北米と英国、オセアニアの英語国が固まった。その近くにはドイツ・オーストリア・スウェーデンなどのゲルマン諸語の国がある。マイナス側と原点付近には、アジア・アフリカ諸国が固まるが、字が重なって読み取れない（拡大図は略す）。従来の外行語の研究は、英語についてのものが多く、英語辞書が電子化され、検索が楽になったおかげで、多くの実例が集められた。それらを元に118語を選定し、59語に精選したから、当然の結果と言える。英語国民は似たことに興味を持って検索したわけである。**【図2-5】**によれば、これらは英語国における19世紀的な日本伝統文化を示す語である。】

<sup>7</sup> 因子分析にも細かな手法の差がある。ここでは最小二乗法、バリマックス回転を適用した。また欠測値（欠損値）や少数例の省き方も、色々あり、単語も200語のうち少数の国家の答えしかない語は省いた。データ行列の行に適用するか列に適用するかの違いもある。数表の行列の単語の次元を因子分析にかけて、その数値を元に国家の平均値を計算することもでき、井上・柳村（2015.3, 2015.9）で報告した。結果はほぼ同様である。

<sup>8</sup> Shinkanseiなどの誤植の語を省いた。また回答の数値の低い（少ない）単語と国・地域を省いた。多変量解析法において少数回答が全体パターンをゆがめる傾向を避けた。井上・柳村（2015.3）と、第1因子・第2因子の順番が違い、中南米が第1因子で取り出されたのは、スペイン語ではsoyが醤油以外にbe動詞の活用形としても使われるためと思われる。ただし井上・柳村（2015.3）図C16によれば、soyの使用率は北米とオセアニアに多い。本節ではsoyを削除したので、本稿の分析のほうが世界の状況を反映している。





**第2因子**のプラス側（図の上方）には中南米スペイン語諸国が分布し<sup>9</sup>、ポルトガル語を使うブラジルはやや離れる。ヨーロッパのスペインとポルトガルも離れて、原点付近にあり、この図2-1では字が読み取れない。マイナス側と原点付近にはその他の国々が集中する。南米の人々は似た事柄（語彙グループ）についてグーグル検索を試みたことを意味する。【図2-6によれば、20世紀初出の語が大きな値を取る。】

まとめると、第1、第2因子では、使用言語によって国家が分類された。英語、スペイン語・ポルトガル語などの言語が大きな分類基準になるが、地理的位置も関係する。

<sup>9</sup> 単語ごとの世界地図を通覧した段階では中南米の特性・共通性に気づかなかった。図1-7で分かるように、国ごとの使用率が低かったせいもある。ヨーロッパと違う種類の少数の語が卓越していたのであろう。事情は東南アジアでも同様である。

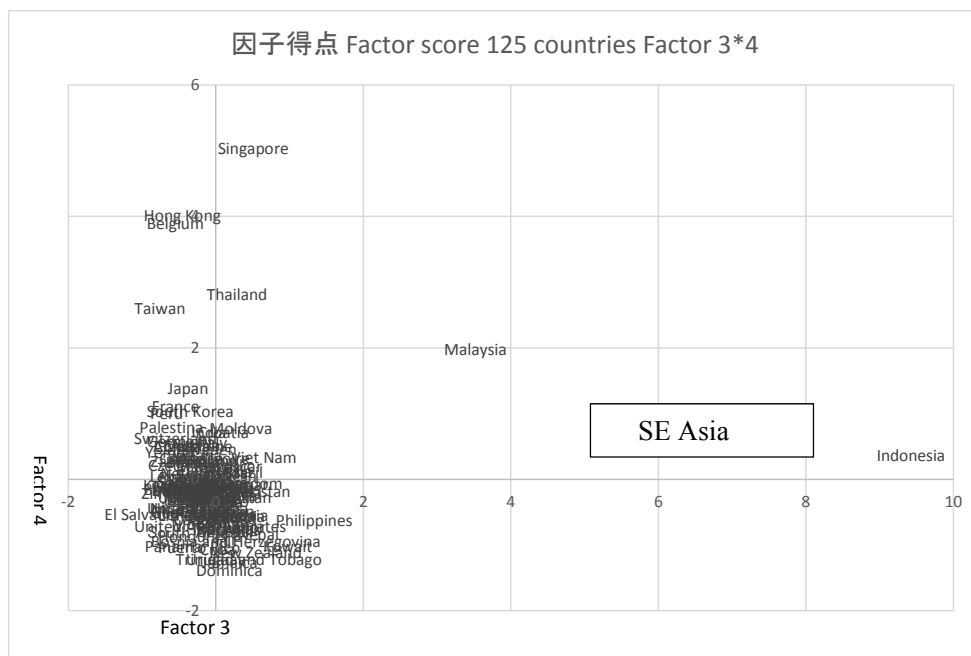


図2-2：外行語の国別使用パターン（3 4因子）

図2-2に国別の第3因子と第4因子の散布図を示す。横軸**第3因子**プラス側でインドネシアが大きな値をとり、マレーシアが次ぐ。両国の公用語（の一つ）はインドネシア語で、実質的に同一と見てよい。ともにアルファベットを用いる。

【図2-7によれば、18、19、20世紀初出の少数の語kaki geisha ninjaが大きな値を取る。】

縦軸**第4因子**プラス側も、シンガポール、タイ、香港、台湾など東南アジア諸国が目立つ。中国の漢字文化圏の周辺部に位置し、民族文字でも検索するが、わざわざ英語つづりで検索したことを意味する。

この中にベルギーが混じるので、その公用語の一つ、オランダ語との関連が考えられる。オランダはかつてインドネシアの宗主国だった。しかし、オランダ自体は図2-2の第2象限、原点近くの集団の中に埋もれており、ベルギーとの強いつながりは考えにくい。【図2-8によれば、18世紀初出のnippon moxaと20世紀初出のkawaii ramenなどが大きな値を取る。】

第3因子と第4因子ともに東南アジアの外行語を取り出す作用を示す。第5因子以下は因子寄与が少ないので、考察を省く。

以上因子分析により、国家の性格を考察したところ、使用言語が大きな影響を与えることが読み取れ、英語・スペイン語と東南アジア諸国が取り出された。

## 2.2. 外行語の単語別因子負荷量Factor load

次に、データ行列の単語の次元を考察する。図2-3の散布図に第1因子と第2因子の因子負荷量を示す。横軸の**第1因子**プラス側には、haiku shogun teriyaki tofuなど食べ物や伝統的な日本文化を表す特有語が多い。コインの裏表にあたる図2-1によれば、英語圏に目立つ語である。現代のグーグル検索の傾向が、古くからの欧米人の日本への興味と似ていることを示す。【図2-5によれば、これらは英語国における19世紀的な日本伝統文化を示す語である。】

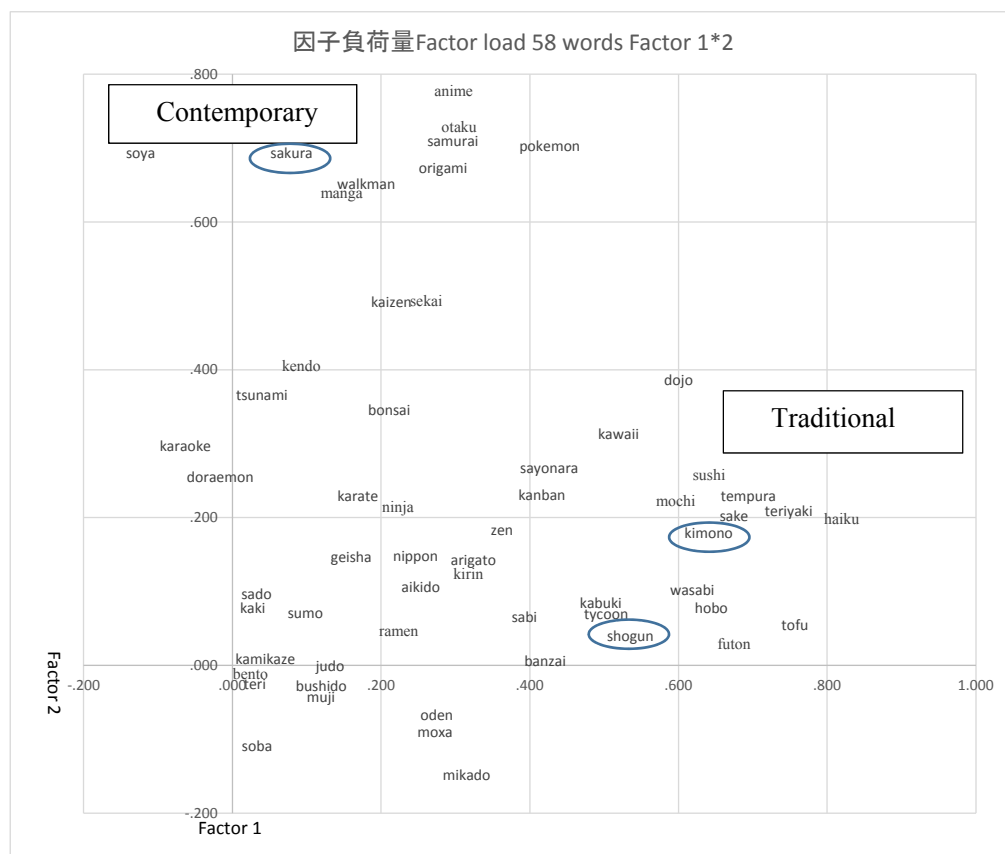


図2-3：外行語の単語別使用パターン（1 2因子）

なお日本語学習の動機が、かつては教養としての日本伝統文化への興味だったが（井上2001.8）、20世紀後期には実用としての日本経済への興味に移り、21世紀にかけては娯楽としてのアニメで代表されるようなポップカルチャーに移ったとされる<sup>10</sup>。グーグル検索にも反映されるはずだが、グーグルトレ

<sup>10</sup> Google Ngram viewerによれば、200年（以上）にわたる出版物における語の使用状況の推移が分かる。井上（2013.12）に一部を示した。

ドは2004年以降のデータに基づくので、時系列のトレンドは明瞭に出ない<sup>11</sup>。Google Ngram viewerで丹念に検索すれば確認できる。

図2-3上方の**第2因子**プラスの語は、anime otaku pokemon kaizenなど現代の日本文化に関わる語である<sup>12</sup>。ただしsoya(大豆、しょうゆ) samurai origami sakuraなどの古くからの外行語も近くにある。図2-1によれば、南米スペイン語圏に目立つ語である。これらの国々では、最新の日本文化について検索することが多い。【図2-6によれば、20世紀初出の語が大きな値を取る。】

原点付近には特徴の少ない語が固まる。特に新しくもなく古めかしくもない外行語で、kamikaze judo sumo bentoなどである。特定の国に偏らず、全世界から検索を受けている。

「日本語観国際センサス」で調査された4語のうち、因子分析にかけたshogun 1875; sakura 1884; kimono 1886には○を付けた(後述のOEDの初出年も添える)。shogun kimono は伝統的語彙にプロットされ、sakuraは現代的語彙にプロットされた。初出年とは違ったメカニズムが働いたようである。

図2-4に第3因子と第4因子の散布図を示す。横軸の**第3因子**プラス側では、geisha ninja shogun kakiのような古くから日本で使われていた語とdoraemon mangaのような新しい語が共存する。図2-2ではインドネシア語に目立つ語だった。こども向けのポップカルチャーと、派生としてのエキゾチックな文化についての検索が多い。【図2-7によれば、18、19、20世紀初出の少数の語kaki geisha ninjaが大きな値を取る。つまりOED初出の古い語が復活して、東南アジアで検索を受けている。】

縦軸の**第4因子**プラス側では、nippon muji kirin ramenのような新しい語が目立つ。図2-2ではシンガポール、タイ、香港やベルギーに目立つ語だった。muji kirin ramenは、戦後の商品や経済活動と関わりがある。検索傾向から見ると、東南アジアにおける日本は基本としての安定した位置を確保したとは言にくい。【図2-8によれば、18世紀初出のnippon moxaと20世紀初出のkawaii ramenなどが大きな値を取る。】

第3因子以下は、少数の国家の少数の語における特別な傾向を取り出す働きを示す。これは井上・柳村(2015.3、2015.9)でも見られた傾向である。図2-3と2-4の結果をまとめると、因子分析によって、国家の言語が分離され、伝統文化を示す語と現代日本文化ポップカルチャーに関わる語が分離された。20世紀末期以降に広がった外行語が、新たに交流の広がった国家に受け入れられたのだろう。日本からの外行語に接した時期の違いに関わる可能性がある。

<sup>11</sup> 本データは2011年12月にダウンロードしたもので、約8年間全体の数値を考察する。

<sup>12</sup> 井上・柳村(2015.3) 図C13によれば、kaizenの使用率はインドと南米で大きい。

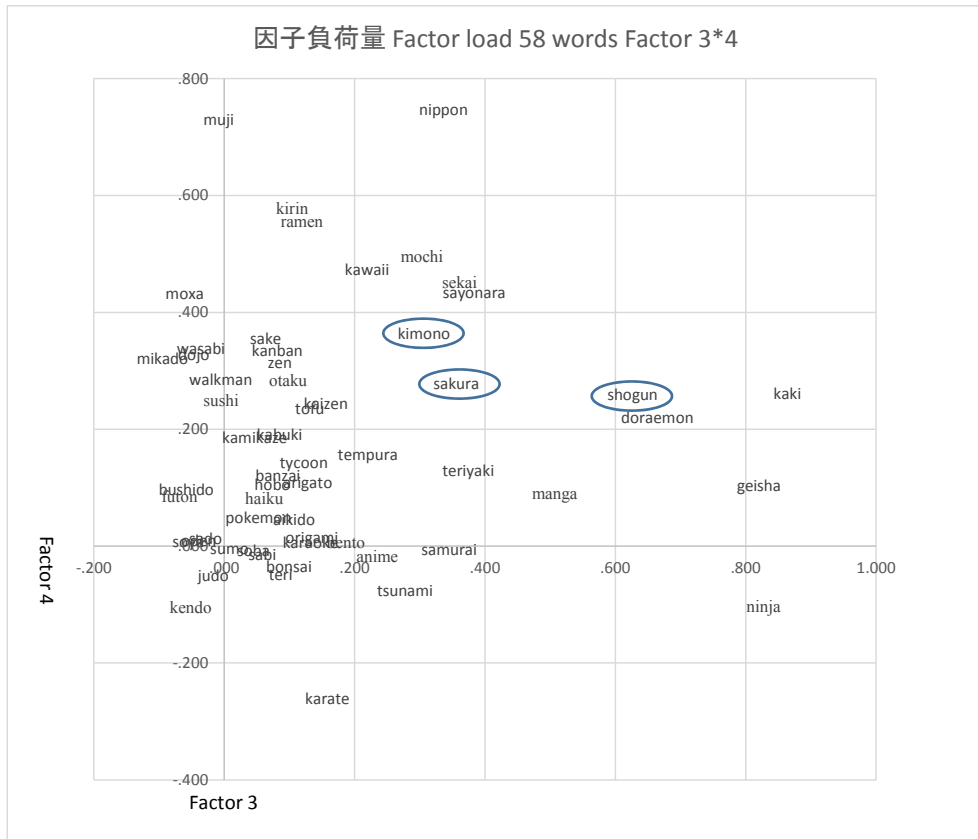


図2-4：外行語の単語別使用パターン（3 4因子）

実際にそのとおりなのかは個々の語の世界分布図で確かめることができる（井上・柳村2015.3）。確かに伝統文化に関わる語は英語国に多く、また現代のポップカルチャーに関係する語は南米に多い。南米は日本との経済交流・貿易のためには距離が遠かった。しかし20世紀末期から日系人の子孫がデカセギなどで来日することもあり、庶民的交流が増えた。これが因子分析によって取り出されたのだろう。何かのきっかけである単語が受け入れられ、広がると、近隣諸国にも広がるのだろう。個々の外行語が個別的に孤立して広がるのではなく、語彙グループとして広がるのが推定される。これがグーグル検索に反映された。

「日本語観国際センサス」で調査された4語のうち、因子分析にかけた3語○印のshogun kimono sakuraは近くに固まった。東南アジアにおいては似た位置付けを受けたようである。

### 2.3. 外行語のOED初出年と4因子

そもそもある「事柄」に興味を抱いても、それが (kimonoのような) 外行語 = 借用語の形で受け入れられるか、(中国語「和服」のような) 翻訳語として受け入れられるかで、検索のパターンは異なる。第3因子と第4因子の中国周辺の東南アジアについては、留意する必要がある(第1因子と第2因子の英語・スペイン語についてはその問題が少ない)。いずれにせよ、グーグル検索で多用されるからには、当該の語はその(人の)言語で十分に確立していると判断できる。

上記図2-3、図2-4では、単語の意味を手がかりに伝統文化とポップカルチャーを判別した。これらの語の発生年代は、日本語ネイティブならある程度直感で判断できる。また『日本国語大辞典』により推定できる。また「日本語歴史コーパスCorpus of Historical Japanese」(CHJ)により、使用状況が分かる。しかし外国語に採用されはじめた時期は、日本語内部での初出年とは食い違う。日本からの外行語を受容した時期に関しては、英語での初出年を照合すれば手がかりが得られる。そこで調査した語すべてについて、OED電子版で初出年を確認した<sup>13</sup>。ちなみに「日本語観国際センサス」で調べた4語のOED初出年は shogun 1875; sakura 1884; kimono 1886; sukiyaki 1920であり、時期が似ている。

ただし各国が英語経由で日本語を採用したわけではない。例えばjudoでなくjiudoのつづりが、(恐らくドイツ語を通じて)いくつかの地域で用いられ<sup>14</sup>、そこではjudoの検索が少なくなる。Byombu bonzu moxa mebos hoboなどについても、ポルトガル語経由など多様な語源説がある。Kanban kaizenも英語を通さず、トヨタの現地企業などから世界各地に直接普及した可能性がある。ただしスペイン語圏でもca-のつづりを使わないのは、日本語ローマ字つづりを介したからだろう。しかし、個別の木を見て、惑わされて、大勢としての森を見逃す可能性があるのも、今は考察しない。

以下のグラフでは縦軸に初出年を(下から古い順に)示す。横軸は因子分析の因子得点を示す。

**図2-5**にOED初出年と**第1因子**の散布図を示す。これまでの分析によれば、第1因子は英語国、伝統語彙と結びつくものだった。【OEDによる初出年を調査した結果の図2-1と照合すると、これらは英語国における19世紀的な日本伝統文化を示す語である。】

縦軸のOED初出年によると、外行語は江戸時代前半の17、18世紀に少し見られ、明治維新後の19世紀後半に急に増え、20世紀に入ってやや落ち着く。この趨勢は図2-10早川(2003)のOED所載語データの年表とも一致する。

横軸の第1因子との対応を見ると、19世紀末期初出の語が大きな値を取る。近似直線はやや右上がり、近代の語ほど第1因子の貢献が大きいことを示す。第1因子は寄与率も大きく、全体の傾向を強く反映する。英語に採用された外

<sup>13</sup> OED、Oxford English Dictionaryに載らない語は、no dateとして扱う。

<sup>14</sup> ベルツがドイツに帰国した後出版した著はDas Kano Jiu-Jitsu (Judo) 1906で、ハンガリー語でもJudoが使われる。なおjudoのOED初出年は1889年である。

行語が、OED初出年と相関を示す。図2-9で見ると、初出の古い語は全体の使用率が低く、新しい語は使用率が高い。

「日本語観国際センサス」で調査された4語のうち、因子分析にかけたshogun 1875; sakura 1884; kimono 1886には○を付けた。shogun kimono は第1因子（伝統的語彙）の値が大きく、sakuraは小さい。OEDの初出年とは違ったメカニズムが働いたようである。

図2-6で横軸の**第2因子**との対応を見る。これまでの分析によれば、第2因子は南米スペイン語圏、新語と結びつくものだった。【図2-1と照合すると、南米スペイン語圏における日本ポップカルチャーを示す語である。】20世紀初出の語が大きな値を取る。近似直線は右上がり、現代の語の貢献が大きいことを示す。第1・第2因子の近似直線は右上がり、新しい語ほど関係が強いことを示す。例外的に18世紀初出のsamurai soya（醤油）が大きい値を示す。現代南米スペイン語圏で復活したと考えられる。

「日本語観国際センサス」で調査された4語のうち、sakuraは第2因子（現代的語彙）の値が大きく、shogun kimonoは小さい。OEDの初出年とは違ったメカニズムが働いたようである。

図2-7で横軸の**第3因子**との対応を見る。これまでの分析によれば、第3因子はインドネシア語、新古混在と結びつくものだった。【図2-2と照合すると、インドネシア語に目立つ語である。】18、19、20世紀初出の少数の語kaki geisha ninjaが大きな値を取る。現代インドネシアで復活した単語と考えられる。近似直線は右下がり、古くから使われていた語の貢献が大きい。英語初出の古い語が復活して検索を受けたためである。

「日本語観国際センサス」で調査された4語のうち、sakura shogun kimonoはほぼ重なる。第3因子以降では似た性格を示すようである。

図2-8で横軸の**第4因子**との対応を見る。これまでの分析によれば、第4因子は、東南アジア諸国、新古混在と結びつくものだった。【図2-2によれば、シンガポール、香港、インドネシアに目立つ語である。】18世紀初出のnippon moxaと20世紀初出のkawaii ramenなどが大きな値を取る。井上・柳村（2015.3）図B9によれば、nipponの使用率は確かにインドネシアで大きい。他の欧米諸国では、中国語起源のジパングJapan系統の国名が確立しているためである<sup>15</sup>。第3因子と同様、近似直線は右下がり、古くから使われていた語の貢献が大きい。英語初出の古い語が復活して検索を受けたためである。

「日本語観国際センサス」で調べたsakura shogun kimonoはほぼ重なる。第3因子以降では似た性格を示すようである。

<sup>15</sup> マルコ・ポーロ『東方見聞録』のイタリア語版（1300年頃）「Zipangu」（日本国）に基づくとされる。



図2-5 : 外行語の初出年 (第1因子)





図2-6：外行語の初出年（第2因子）



図2-7 : 外行語の初出年 (第3因子)

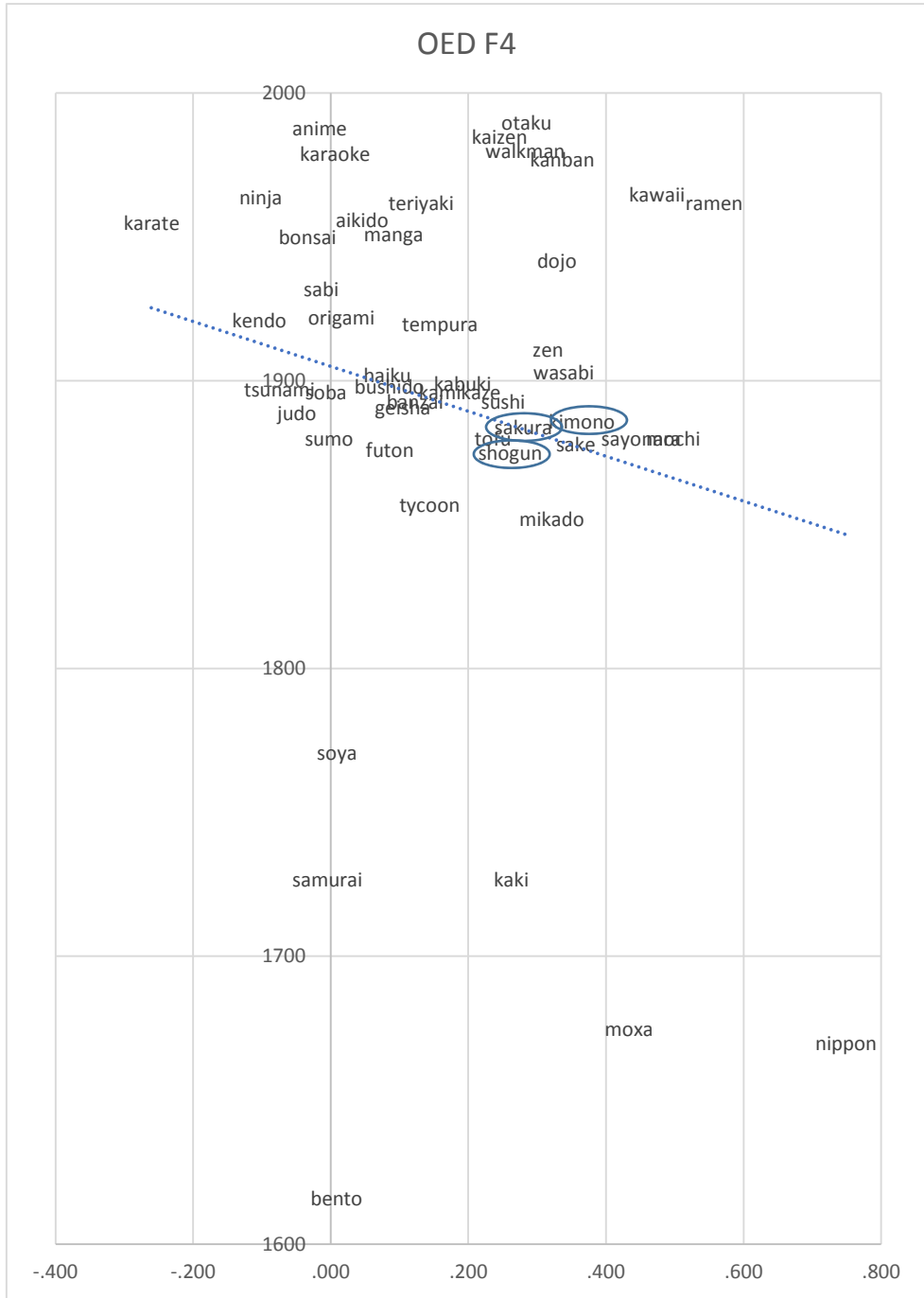


図2-8：外行語の初出年（第4因子）

## 2.4. 外行語のOED初出年と使用頻度数

ここでOED初出年の性格について、別の目で確認しよう。図2-9で横軸の118調査語について、世界全体の使用率との対応を見ると、18世紀以前初出の語は使用率が低く、20世紀初出の語は使用率が高い。右上がりの近似直線で示されるように、左1900年前後初出の多くの語の使用率（検索率）の低さが大きく働いた。また1900年以前初出の語は使用率が低く、1950年以降初出の語は着実に高い使用率を示す。右下に例外的なtycoonとtsunamiがあるが、tsunamiは災害のために国際的に認定された語で、例外的である。Tycoonは意味を一般化させて、使用頻度を上げた。

図2-9全体としては、近似線が示すように、近代に新しく出た語ほど多く使われ、初出の古い語は廃れる傾向を表す。最近の外行語は、インターネットの発達、国際交流の拡大などによって、世界全体に早く普及すると考えられる。また新語の多くがのちに必要性または新鮮さを失って、使用頻度を減らすのと、同じメカニズムが働いたと見られる。古い外行語が使われなくなる傾向は、グーグルNグラムビューアー (Michel et al. 2010)とグーグルトレンドで個々の語の推移を数100年、数10年単位で観察しても確認された (井上2013.12)。本稿図1-6aの4語の使用率増加でも示された。

「日本語観国際センサス」で調べた4語のうち3語shogun 1875; sakura 1884; kimono 1886は、図2-9のグラフでの配置が類似していた。世界各国で一定の使用率がありそうな語を選んだためである。調査企画当時は、外行語の世界全体での位置、地位について、ほとんどデータがなかった。本稿のように様々なデータを相互参照して、様相が把握できた。今後はもっと効率的に多様性を把握した上で、規則性を見出すことができるだろう。

これまでに外行語の新古については、日本語の中での使われ方を元に推定して、考察してきた。しかし例えばninja samuraiが古くから海外各国に普及していたかは、確証がなかった。時代の手がかりとしてOEDによる初出年を調査したところ、因子分析の結果と対応関係が見られた。近代発生の語については、大まかには日本語の中での成立と一致するが、近世（江戸時代）初出の語については、20世紀以降の復活によるものか、個別に検討する必要がある。ことにポップカルチャーにまつわる語の中に、江戸時代の語の新採用が混じる。

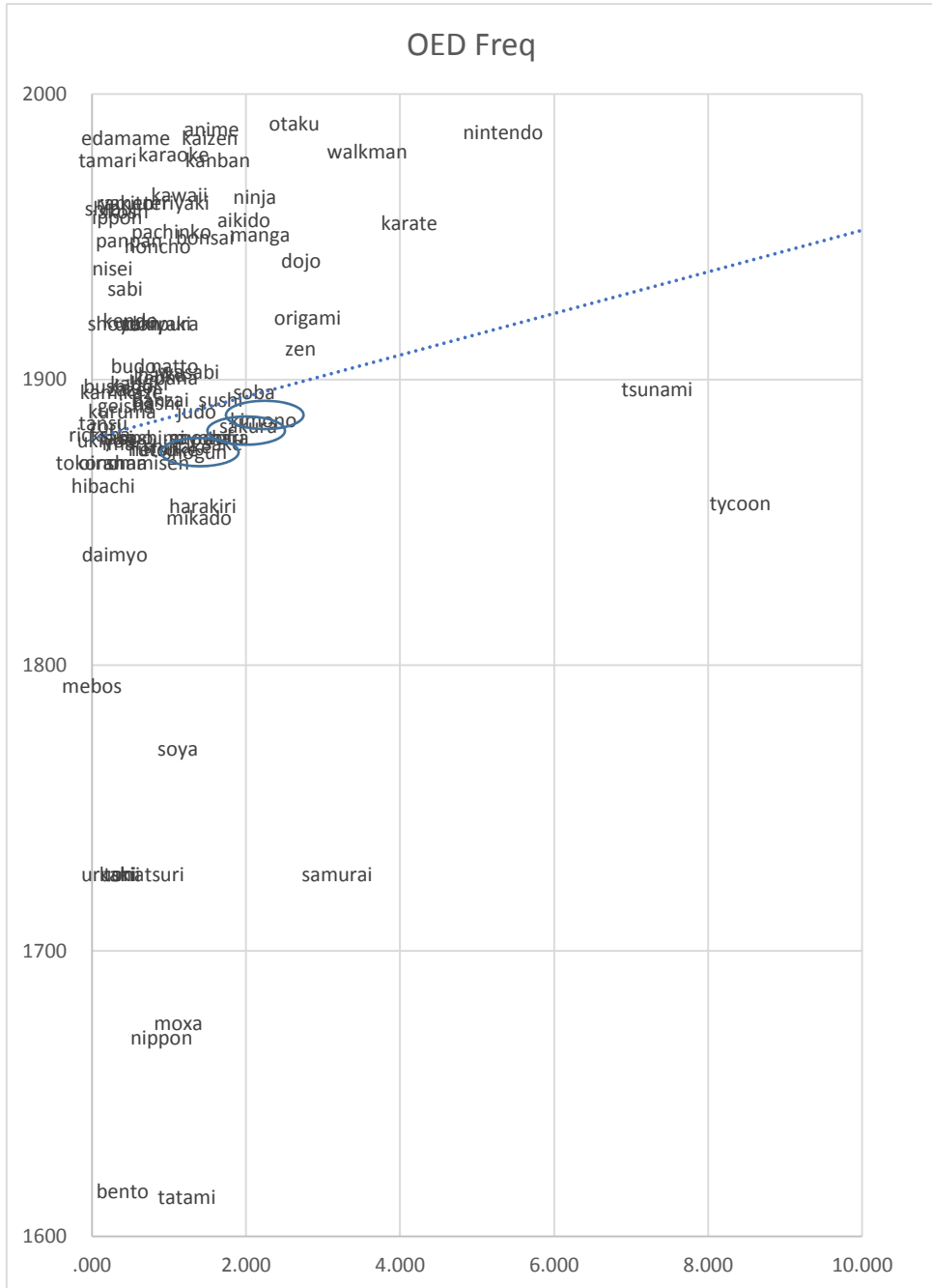


図2-9：外行語の初出年と使用頻度

## 2.5. 外行語905語のOED初出年

英語への外行語については、多くの研究がある。そのうちHayakawa (2003)の研究では、905語についてOEDまたは他の辞書や原典による初出年がリストになっている。初出年10年ごとに語数を数えると色々なことが分かる。図2-10に示した。江戸時代(1603-1868)の直前から文献記録に出るが、江戸時代は10年単位の違いが大きい。1720年代はKempfer、1790年代はThunberg、1860年代はHepburn、1880年代はChamberlain その他大勢の著作のおかげである。明治期(1868-1912)には日本紹介の本が多く出て、外行語も増えた。しかし1910年代第1次世界大戦の10年間には、減った。ヨーロッパのことで頭が一杯だったのだろう。これが1940年代まで続く。第2次世界大戦では、harakiri banzai kamikazeが欧米で使われたが、その初出年はもっと早いので、このグラフには表われない。Tojoが(首相の名前と)ずれた意味で使われたが、総数としては多くない。1940年代初出の語は柔道用語や科学用語である。戦争は相手国への関心を高めるが、借用語の全体としてはむしろ排除される傾向がある。日本の側では「敵性語」の排除を行ったが、英語の側にも外行語を受け入れない傾向が深層心理としては、あったと考えられる。この勢いは1950年代に少し回復するまで続いた。

その後は年代が近すぎて、辞書に載せるのが間に合わず、また定着するまで辞書への掲載を待つ傾向があるので、数が少ない。1980年代以降に出現した語も多いので、全体としての使用は増えている。

外行語初出年のデータと貿易量、輸出量の歴史的推移との関係を、インターネットからダウンロードした各種統計グラフと比べて考察したが、長期間の統計データは少なく、あっても、外行語の推移との対応は期待ほどにきれいではなかった。しかし1940年代には貿易量が減り、外行語も減る点など、戦争は借用語を抑え、経済交流、貿易は多くすると、一般化できそうである(井上2015.5)。外行語の実際の使われ方の減少は、過去200年についてはGoogle Ngram viewerで、21世紀以降なら後述のGoogle trendsで見当が付く。今後の課題である。

図2-10と図2-9を照合すると、共通性が読み取れる。図2-9の項目は従来の様々な研究で取り上げられた外行語で、初出年に関して特に配慮したわけではない。しかし江戸時代の語が少なく、明治になって19世紀後半に急に増えることは、共通である。20世紀前半に減るのも、両グラフともに見られる。20世紀後半に図2-10でさらに減るのは、新語導入に用心深い辞書の宿命と言える。それに対し図2-9では、最新の外行語情報を求めて諸文献を渉猟したので、かなり多くの語が取り上げられ、それらの使用率は大きい。ちょうど時流に乗った段階をとらえることができた。

新語・流行語は一定の年月が経つと衰えることがあり(上村2014)、他方一般語として確立することもある。辞書編纂者にとっては、予測できる法則性を知りたいだろう。Google trendsが手がかりになる。本稿の(新しい)外行語も、追跡調査により、何らかの規則性発見の貢献ができるだろう。

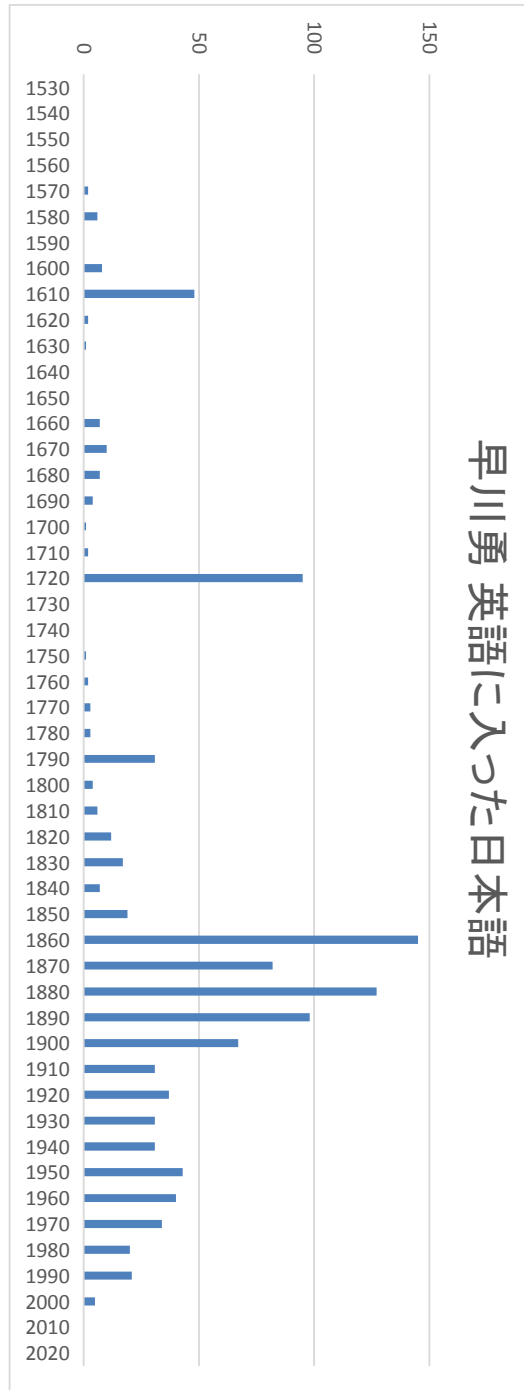


図2-10 OEDの外行語初出年

### 3. グーグルトレンドの外行語111語の因子分析

#### 3.1. 外行語の125国家の111語のデータ

第2節では語数と国家数を精選し、かつ soy を削除したデータに因子分析を適用した。しかし語数と国家数を減らさずに得られたデータ全体を扱ってみたい。また soy については、以前の稿で確認したところ（井上・柳村 2015.3 図 C-16 soy）中南米の使用率は小さい。地図を図 3-1 に再掲する（Google 2012 の出力結果による）。スペイン語 be 動詞の活用形 soy が Google 検索で使われることは、soy bean, soy sauce などが英語国で検索される機会に比べればまれなのだろう。

第2節で中南米が第1因子で取り出されたのは、スペイン語で soy が醤油以外に be 動詞の活用形としても使われるためと考えたが、再考の余地がある。そこで 125 国家の 111 語の大きなデータに因子分析を適用した。

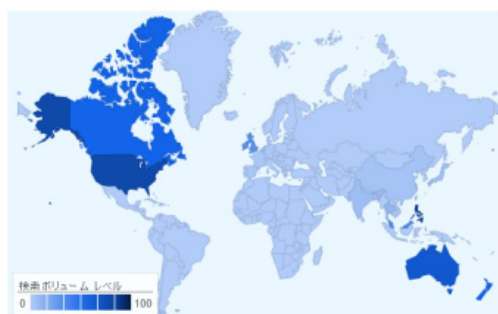


図3-1 soy の世界的分布

表3-1 に111語の4因子全体の基本情報を掲げる。結論を先取りする形で4因子の解釈も記す。表2-1の59語に比べて、第1因子の因子寄与が大きい。第1因子が英語圏に結びつく点は同じだが、59語では伝統語彙と関連が見られ、111語ではポップカルチャーの新語との関連が見られた。

表3-1：因子の意味・属性（111語×125国家）

	Factor1	Factor2	Factor3	Factor4
因子得点	.795	.352	.167	.211
因子寄与	12.205	9.194	7.792	7.779
使用国家	英語圏	中南米	東南アジア	アジア
語彙特性	ポップカルチャー	伝統的活動	産業	日常生活
新古	新語	伝統的	新語＋古語	伝統的



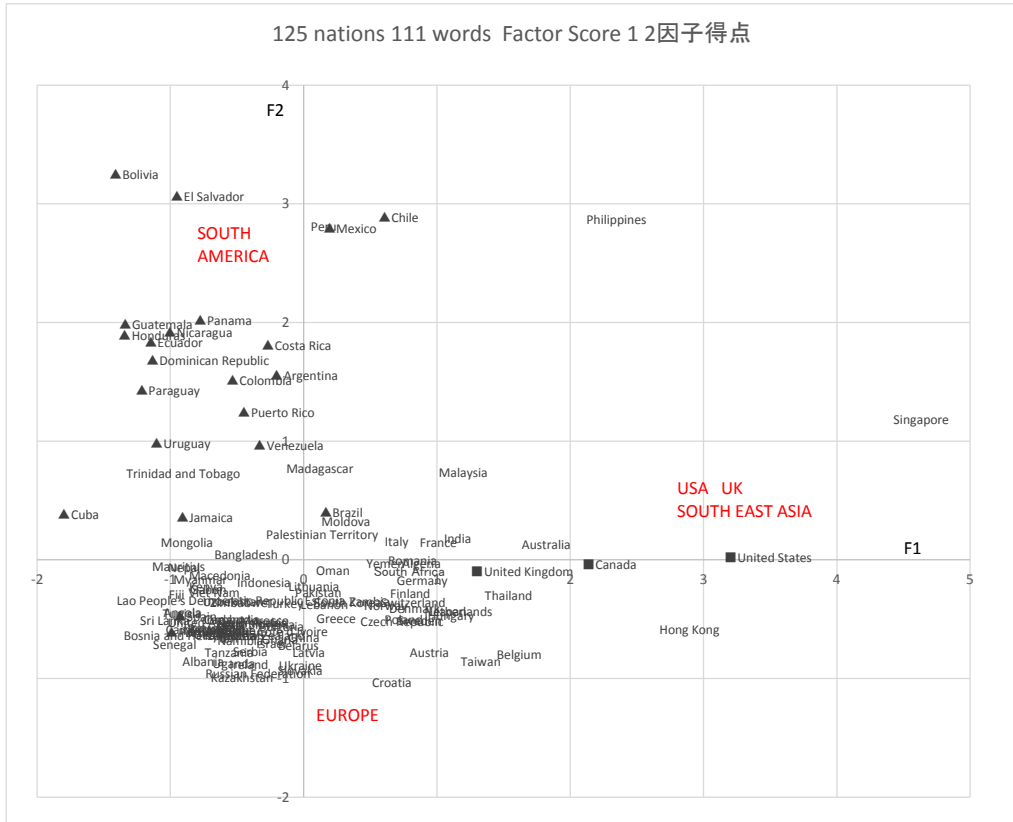


図3-2 125国家 第1、2因子

### 3.2. 外行語の125国家のデータ

結果を順に見よう。まず図3-2に国別の第1因子と第2因子を組み合わせた散布図を示す。横軸は**第1因子**で、一番重要な因子である。プラス側に■印アメリカ・カナダと英国とオーストラリアが固まった。シンガポールが例外的に大きな値を取り、香港、マレーシア、フィリピンも近くに分布する。東南アジアを含めた「英語圏因子」と名づけよう。

縦軸は次に重要な**第2因子**で、プラス側に中南米諸国▲がプロットされた。「中南米因子」と名づけよう。ヨーロッパのスペインとポルトガルは離れて、むしろ米国に近い。また南米でも日系人の多いブラジル、アルゼンチンなどよりは、中米の小国が固まって大きなプラス値を示す。「南米スペイン語圏」と呼びたい。原点付近には、ヨーロッパ諸国、東アジア諸国が固まり、国名が読み取れない。

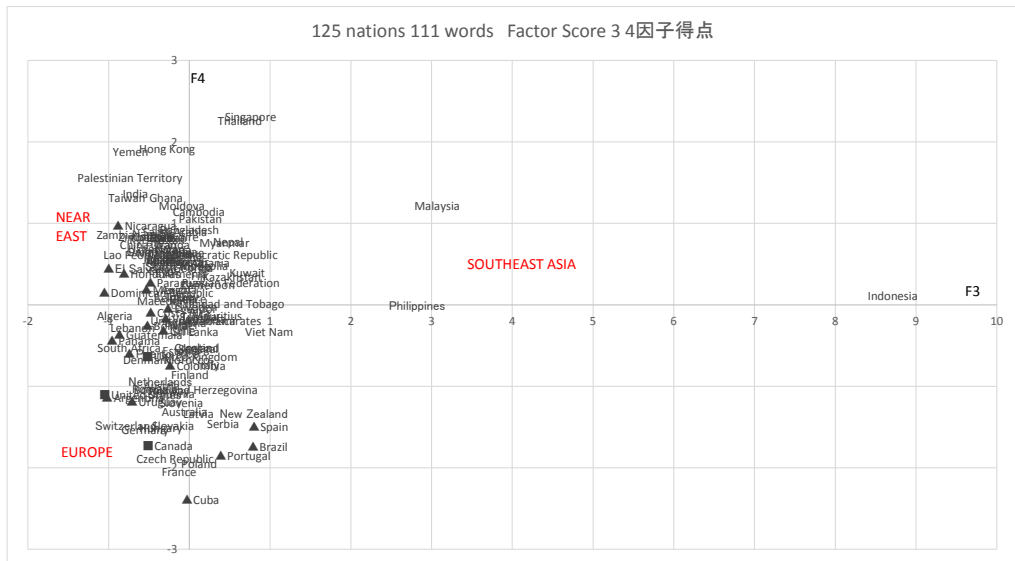


図3-3 125国家 第3、4因子

これまでの個々の外行語の分布から見て、西欧とアジアの違いが出ることは、予想していた。確かにきれいに分離されたが、第2因子の中南米は想定外だった。このグラフからはもっと細かく読み取れた。

図3-3に国別の第3因子と第4因子を組み合わせた散布図を示す。横軸**第3因子**は、インドネシア、マレーシア、フィリピンが大きなプラス値をとる。「東南アジア因子」と呼ぶ。

縦軸**第4因子**で、プラス側に東アジア諸国、中東諸国などが分布する。「アジア因子」と呼ぶ。中南米諸国▲は原点付近にプロットされた。原点付近には、他の多くの諸国が固まり、国名が読み取れない。英語圏■は第3、4軸マイナス値で、アジアと対立する傾向を示す。

図2-1、2-2と照合すると、国の配置が一致しない。一部の国家でしか使われないために省かれた語の受容が多様なためであろう。しかし第1～4因子の意味としては同一である。

### 3.3. 外行語の111語のデータ

因子分析では、以上のような国別の値を元に、単語ごとの平均値が因子得点として計算される。図3-4に第1、2因子の散布図を示す。図3-2とコインの裏表の関係にあり、英語圏因子にあたる。20世紀末期以降に広がった外行語が、昔伝わった外行語と異なった傾向を示す。OEDでは20世紀末期以降の語が載ら

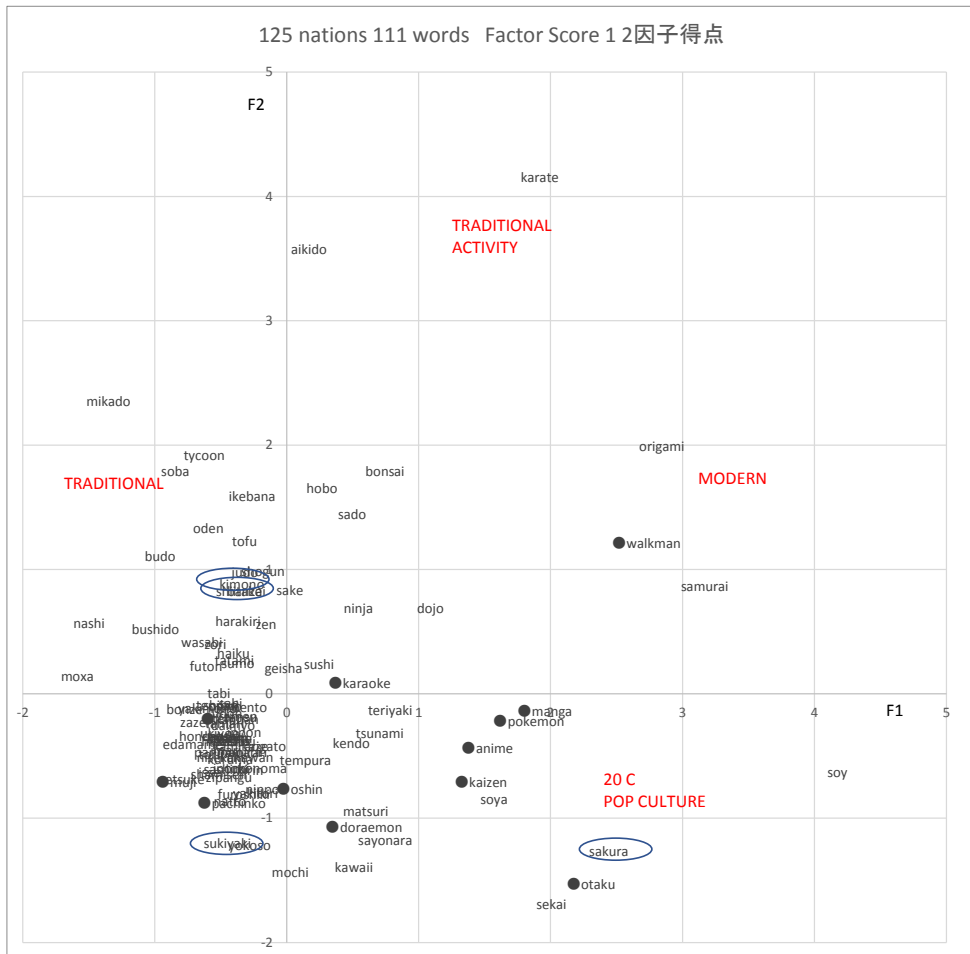


図3-4 111語 第1、2因子

ないので初出年データが得られない。日本語自体での初出年を手がかりに、新語12語に●印を付けた（表4-1参照）。

図2-3と照合すると、語の配置が一致しない。一部の国家でしか使われないために省かれた語の受容が多様なためであろう。

図3-4横軸の**第1因子**では、soyとsoya（大豆、しょうゆ）が右端の大きな値を取る。Soy bean、soy sauceの普及のおかげだろう。第2因子の中南米と結びつかないから、スペイン語be動詞のせいではない。なおtofuは左上の第1因子マイナス、第2因子プラスの位置にあるから、同じ原料によるものを別系統の外行語で呼んでいる。Samurai origami sakura sekaiのような昔からの語も第1因子プラスである。

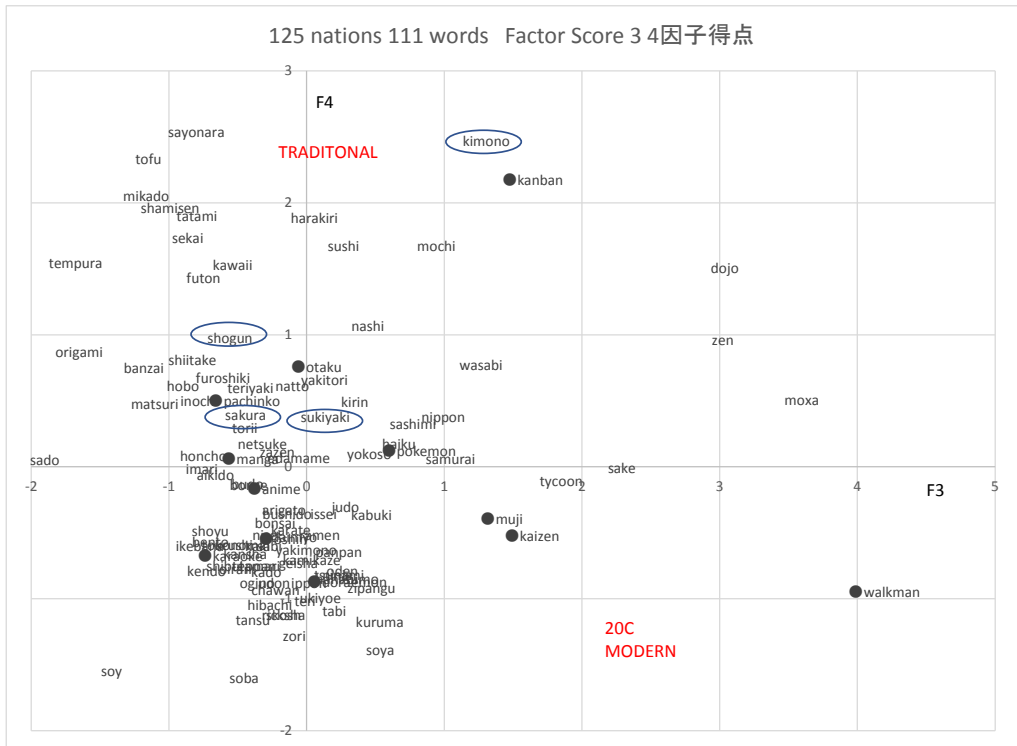


図3-5 111語 第3、4因子

●印の walkman manga pokemon anime otaku kaizen が第1因子で大きな値を取る。「ポップカルチャー」の単語である。新たな日本文化の表れとして、古くからの外行語が復活して導入されたのだろう。しかし●印のすべてが第1因子プラスではなく、原点付近に多く重なっている。世界全体で万遍なく（偏りを持たずに）検索された語である。

**第2因子**、図3-4の上の語は「中南米因子」で、中南米スペイン語圏に多い語である。Karate aikido origami bonsai sado ikebana budo dojo ninjaなど伝統的（身体的）活動に関わる語が固まり、mikado tycoon hoboやsoba oden tofuも近くにある。

図3-5に第3、4因子の散布図を示す。図3-3とコインの裏表の関係にあり、東南アジア因子にあたる。

**第3因子**、図3-5の右の語は、●印の walkman kaizen kanban muji など産業に関わる新語と、moxa zen dojo sake tycoon kimono wasabi mochi など身近な生活に関わる伝統的な語が混在する。

**第4因子**、図3-5の上の語は、sayonara kimono mikado shamisen tatamiのような雑多な意味分野の語や tofu tempura sushi mochi nashiなどの食品名である。

第3, 4 因子ともに、庶民の日常生活で実際に目にふれそうな具体的な語が多い。若い世代よりは、働きざかりの人が使いそうな語で、東南アジアにおける現代日本文化の位置付けに関係するのだろう。

第3 節では、125 国家の 111 語の大きなデータを用いたために、「日本語観国際センサス」で扱った 4 語すべての位置付けができた。○で囲んである。

**図 3-4** 第1, 2 因子では、第2 節で省かれていた sakura が第1 因子プラス、つまり英語圏因子、ポップカルチャー因子に近く位置付けられた。Kimono shogun は第2 因子プラス、つまり中南米因子、伝統文化因子に近く位置付けられた。Sukiyaki はこれらから離れて、第1, 2 因子マイナス側に孤立している。図 1-1 だけからこの傾向を見つけることは困難である。4 語が異なった性格の語であることを意味し、選択が適切だったと判断できる。

**図 3-5** 第3, 4 因子では、kimono が第1, 2 因子ともにプラスで他の3 語と離れ、東南アジア因子と深く結びつくことを示す。アジアの多様な民族衣装の一つと位置付けられているのだろう。Sukiyaki sakura shogun は原点付近にプロットされ、伝統文化の語の中に位置付けられる。

第2 節で精選した 59 語によって、世界の外行語を大きくとらえようとした。第3 節では、大勢を把握した上で、国家数と語数を増やし、細かい部分までとらえようと試みた。第1 因子以下の、取り出された国家の傾向は同じである。英米と中南米と東南アジアが特徴的な因子だった。

それぞれの地域に特徴的な単語も取り出すことができた。20 世紀後半に生まれた語(●印)が、英米語圏に目立った。それに対し、中南米では伝統的(身体的)活動に関わる語がグループをなし、東南アジアでは庶民の日常生活に関わる語が目立った。

分析の語数を増やした効果はあった。少数の語に頼るよりは、根拠が確実に became。外行語が個々に取り入れられるのではなく、社会的・文化的背景をもとに、グループとして入りこむことが分かった。因子分析の手法も多様だが、分析するデータ(国家数、語数)によっても結果が違ふことが分かった。この節での結論は今後追試されてしかるべきである。

## 4. 議論と結論

### 4.1. 外行語の全体的傾向の把握

従来の借用語の研究は、個々の単語がどの国でどう使われているかを個別的に論じるものが多かった。世界全体の地理的分布を見渡す研究は、実行不可能に近かったが、インターネット検索の発達によって視野が開けた。多数の単語の大きな傾向も把握する必要があるが、インターネット情報のおかげで可能に

なった。グーグルマップで個々の外行語の世界地図を考察すると、多様性が目立つが、多くの世界地図を通覧すると、似たパターンが浮かび上がる（井上2011.2、2012.2、井上・柳村2015.3、2015.9）。方言地理学では「単語は独自の歴史を持つ」と言われるが（Jaberg 1908）<sup>16</sup>、独立に変化するのではない。似た性格の語彙は似た地理的分布を示す。日本国内の方言地図で観察された一般傾向だが、世界全体にも通じる。

社会言語学的にグローバルに見渡すと、地球上の3か所（合衆国東部、西ヨーロッパ、東アジア）が特徴的であることは、以前に気づかれていた（井上2001.8、2011.11、2015.5、2018.1、Inoue 2005、2012.1）。「日本語観国際センサス」（新プロ「日本語」総括班編1999、国立国語研究所2002）で日本語の評価が高く、インターネット検索で多言語使用が観察され、多言語景観についての様々なデータとも一致する。衛星写真で夜明るい世界の3か所とも一致し、人口密度、GDPが高く、交通網・通信網・教育程度など、生活水準の高い地域でもある（井上2018.1）。

本稿第2節、第3節で多数の外行語に多変量解析を適用して、データの内部構造を分析した結果が、この世界の3か所と一致することを期待したが、一部ずれが見られた。ヨーロッパとアジアで外行語の受け入れ方が違うことは、交流の歴史から言って予測できた。また西アジアから東アジアにかけては漢字と各種民族文字が使われ、検索でも英語つづりがあまり用いられないことにより別の傾向が出ることは予測していた。しかし南米スペイン語圏の特性は想定外だった。

南米の（または大西洋やアフリカなど世界各地の）スペイン語の違いは一種の方言差と考えるべきだろう（上田1987、2013、Ueda 1995、Fernández & Ueda 2018）。南米の中でも単語により国によって濃淡（使用・検索の比率）は多様である。グーグル検索における国家間の差は、世界のスペイン語の間の方言差としてもとらえられる。言語間方言学interlingual dialectology（Weijnen 1978、井上2001.2）にふさわしい研究テーマである。借用は言語間で研究されることが多く、Bloomfield（1933）で、文化借用cultural borrowing、密接借用intimate borrowing、方言借用dialect borrowingの3種が指摘されているにも関わらず、方言差が考慮されることは少なかった。しかし最近と同系の諸言語間や、隣接地域での借用、移民によるホスト言語への借用など、interlingual dialectologyにあたる研究が増えている（Oja 2002、Saramandu & Nevaci 2019、ハナシロ・島田2011、島田2018）。

本稿では、因子分析を適用することにより、国家を分類でき、単語も分類できた。単語の借用されはじめた時期については、これまでは断片的な情報に頼

<sup>16</sup> ちなみにこの表現は言語地理学の開祖Jules Gilliéronのものではなく、Bloomfield（1933）pp. 328, 520, 533では、出典としてKarl Jaberg（1908）*Sprachgeographie*を挙げる。インターネット検索Google Booksで原典にあたれたが、その後たどれなくなった（参考文献Jabergの項）。“Jedes Wort … hat seine besondere Geschichte”が一番近い。元はJakob Grimm（1819）*Deutsche Grammatik I*の序文冒頭の“Jedes Wort hat seine Geschichte”と思われる（風間1985 p.132）。フランス語の表現“Chaque mot a son histoire”と対応する。

って推測していたが、OEDによる初出年を調査した結果で、因子分析による推測の適切さが確認された。

#### 4.2. 言語の統合と分岐

一般に、前近代には言語の分岐Divergenceの動きのみがあり、近代には言語の統合Convergenceの動きのみがあったと考えられている。つまり、前近代では、言語の分岐・分離の動きが大きく、方言の差が大きくなり、各国の標準語の違いが強調されるようになった。近代になると、言語の統合・収束の動きが強くなり、各地の方言が標準語化され、国家間の交流により借用語（外来語・外行語）が増加した。

しかし、調査によると、古くから分岐と統合の両方の動きがあった。統合の動きは古代から都市の標準語のことばの伝播・拡大の形で続いている。分岐の動きは現代でも見られ、例えば、新方言が生まれ（井上 2012.1）、ウクライナ語、スロバキア語、クロアチア語などの造成言語Ausbauspracheでは借用語を排除し、意図的に言語間の差異を大きくする動きもある。図4-1は、その大まかな傾向を示している。

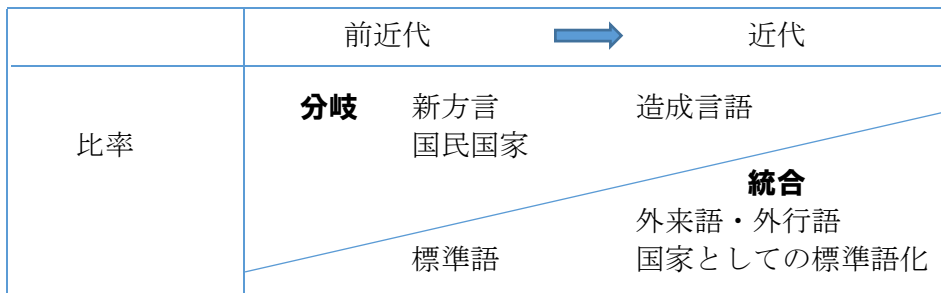


図4-1：言語の分岐と統合

このように、インターネット情報は、現代の標準化だけでなく、新方言や気づかない方言の広がりも分岐として示している。つまり、言語変化が進行している様子を直接観察することができる。方言学は言語学の一分野として研究されてきたが、方言と言語の境界は連続的である。外来語・外行語はその典型的な例である。方言学の原則は、一般言語学にも適用される。言語間方言学は、この2つの分野をつなぐ境界領域である。グローバル化の進展に伴い、地理言語学・言語間方言学はますます盛んになることが予想される。

#### 4.3. 地理言語学・言語間方言学の新たな地平

英語の世界変種に関する研究が進んでいる。発音や文法の世界的研究としてuptalk: rising intonationについては、オーストラリア、ニュージーランド、アメリカ・カリフォルニアだけでなく、国内規模の研究も行われている（井上 2008.5）。英語の変種は、国によってイングランド英語、アメリカ英語、オー

ストラリア英語などに分けられ、さらにそれぞれは、内的特徴によって、北部・南部、西部・東部に分けられる。方言地図は国家内で作成され、方言区画・分割の試みに役立っている。一方、sneaked/snuk、yinz (複数形専用のyou)、torpedoといった新しい地域差も指摘されている。今後、言語間方言学の実践として、世界の様々な言語について外行語の研究が行われるだろう。

#### 4.4. 外行語研究の視野拡大

外行語の研究は世界の諸言語について可能である。日本語からの外行語だけでなく、英語、スペイン語、中国語、韓国語から全世界への外行語（輸出語）でも、興味深い結果が得られた（Inoue 2013.4、梁yang2005、2007）。なおグーグルの機能を活用するともっと色々なことが分かる。グーグルマップを使うと外行語による店名などの世界分布が分かる。またグーグル・ストリート・ビューを使えば、店名の表示そのものや周囲の環境も分かる（井上2018.1）。グーグルNグラムビューアーにより1800年（または1600年）から2008年までの長期の（書籍の中での）言語使用が分かる。数年後、数10年後には、同じ（または更に拡充された）データを使ってreal timeで追跡調査を行い、変化を実証できる。今後世界の諸言語について外行語の研究が行われるだろう。さらに外行語と外来語を合わせた借用語について、国内外の方言差を把握し、世界全体を把握する研究が進むことを期待する。新研究手法による言語間方言学interlingual dialectologyの発展である。文化交流、経済発展、言語構造、言語イデオロギー（言語純化運動＝外来語排斥運動）などの要因がどう働くかの研究も、実証的データを元に、行われると期待される。

今後の課題として、語彙論的な意味分野、使用場面、使用者のバックグラウンドなどの解明がある。例えば「専門家アクセント」（外来語アクセントの平板化）は、個人の所属集団や趣味に左右される（井上2008.5）。日本国内の外来語の裏返しとして外行語をとらえると、言語ごとの語彙論的な（文化的な）守備範囲の偏りが浮かび出る可能性がある<sup>17</sup>。日本ではフランス語、イタリア語、ドイツ語などが料理や音楽やスポーツなどの限られた分野でその道の人に盛んに使われる。個々の単語の背後と周辺には類義語と関連語彙が控えており、機会があれば外行語が増殖する。また親しさ・なじみ度が増すと「特有語」から一般語に変身する（Tycoonやhoboがその例とされる）。個々の語史を見るだけでなく、大量のデータを通覧して規則性・法則性を探る態度が必要である。

外来語の裏返しの現象が世界各国の日本語からの外行語に見られるだろう。今回の因子分析でも英語圏、中南米、東南アジアで使われる単語に違いが見られた。言語（文化）接触の時期の違いにもよるが、意味分野、使用者の社会層などにもよる。かつての東洋のエキゾチックな国から、凶暴な侵略国家、猿真似による安物の輸出国、すぐれた技術製品の国、アニメなどの大衆文化の国と、この200年間にイメージの変遷があった。それが各国の外行語の有無、多少、

<sup>17</sup> 井上・柳村（2015.3）図B11、B12によれば、judoとsumoの使用率はモンゴルで大きい。



意味分野の違いに反映している可能性がある。個々の語の考察（語史）の積み重ねでも分かるが、本稿のような大量データの解析で、証明力、説得力が増す。社会言語学、経済言語学、地理言語学の新たな研究領域の拡大につながるだろう。

選択	111語	<input type="checkbox"/>	kabuki	●	oshin		tatami
<input type="checkbox"/>	aikido		kado	●	otaku	<input type="checkbox"/>	tempura
●	anime	●	kaizen	●	pachinko		teppan
<input type="checkbox"/>	arigato	<input type="checkbox"/>	kaki		panpan	<input type="checkbox"/>	teri
<input type="checkbox"/>	banzai	<input type="checkbox"/>	kamikaze	●	pokemon	<input type="checkbox"/>	teriyaki
<input type="checkbox"/>	bento	●	kanban	<input type="checkbox"/>	ramen	<input type="checkbox"/>	tofu
<input type="checkbox"/>	bonsai		kansha		ricksha		tokonoma
	bonze	●	karaoke	<input type="checkbox"/>	sabi		torii
	budo	<input type="checkbox"/>	karate	<input type="checkbox"/>	sado	<input type="checkbox"/>	tsunami
<input type="checkbox"/>	bushido	<input type="checkbox"/>	kawaii	<input type="checkbox"/>	sake	<input type="checkbox"/>	tycoon
	chawan	<input type="checkbox"/>	kendo	<input type="checkbox"/>	sakura		udon
	daimyo	<input type="checkbox"/>	kimono	<input type="checkbox"/>	samurai		ukiyoe
<input type="checkbox"/>	dojo	<input type="checkbox"/>	kirin		sashimi		urushi
●	doraemon		kuruma	<input type="checkbox"/>	sayonara	●	walkman
	edamame	●	manga	<input type="checkbox"/>	sekai	<input type="checkbox"/>	wasabi
	furoshiki		matsuri		shamisen		yakimono
<input type="checkbox"/>	futon	<input type="checkbox"/>	mikado		shibui		yakitori
<input type="checkbox"/>	geisha	<input type="checkbox"/>	mochi		shiitake		yokoso
<input type="checkbox"/>	haiku	<input type="checkbox"/>	moxa	<input type="checkbox"/>	shogun		zazen
	hanabi	●	muji		shoyu		zen
	harakiri		nashi		skosh		zipangu
	hibachi		natto	<input type="checkbox"/>	soba		zori
<input type="checkbox"/>	hobo		netsuke		soy		
	honcho	<input type="checkbox"/>	ninja	<input type="checkbox"/>	soya		
	ikebana	<input type="checkbox"/>	nippon	4	sukiyaki		
	imari		nisei	<input type="checkbox"/>	sumo	●	新語 111語中12語
	inochi	<input type="checkbox"/>	oden	<input type="checkbox"/>	sushi	<input type="checkbox"/>	精選58語
	ippon		ogino		tabi	4	国際センサス4語
	issei		oiran		tamari	無印	精選58語もれ53語
<input type="checkbox"/>	judo	<input type="checkbox"/>	origami		tansu		

表4-1：因子分析の項目一覧（58 or 111語×125国家）

表4-1に因子分析の項目一覧を掲げる。111語全語形をアルファベット順に並べ、精選58語、国際センサス4語、OEDで初出年判定不可能の12新語などに記号を付けた。無印の一部の国家でしか使われないために省かれた語は多様で、海外の言語景観として目にしたもの、文献で接したものも混じる。外行語の受容の多様性を示す。

## 参考文献/References

- Bloomfield, Leonard (1933) *Language*. New York: Holt.
- Cannon, Garland (1996) *The Japanese Contributions to the English Language, An Historical Dictionary*. Wiesbaden: Harrassowitz.
- Daulton, Frank E. (2022) Japanese loanwords and lendwords. In Yoshiyuki Asahi, Mayumi Usami, Fumio Inoue (Eds.) *Handbook of Japanese Sociolinguistics*, 509–537. Boston/Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9781501501470-018>
- エシュバハ・サポー, ヴィクトリア (2005) 「ドイツの日常のなかの日本語 (特集= 日本語に入ったことば, 日本語から出たことば)--」(「外行」語--ソトへ出た日本語)国文学解釈と鑑賞70(1): 105–116.
- Fernández, Francisco Moreno and Hiroto Ueda (2018) Cohesion and Particularity in the Spanish Dialect Continuum. *Open Linguistics* 4(1): 722–742. <https://doi.org/10.1515/opli-2018-0035>
- ハナシロ・チャド, 島田めぐみ (2011) 「ハワイにおける日本語語彙の認知に関する研究」『日本語教育世界大会予稿集』(天津) .
- 原口庄輔・原口友子編訳 (1998) 『新「国際日本語」講座』洋販出版.
- 早川勇 (2003) 「英語に入った日本語語彙の初出年調査」『日本語科学』13: 79-108.
- 早川勇 (2006) 『英語になった日本語』春風社.
- 林大監修 (1982) 『図説日本語』角川書店.
- Hiramoto, Mie (2010) Dialect contact and dialect change among plantation immigrants from northern Japan in Hawai'i. *Journal of Pidgin and Creole Linguistics* 25(2): 229–262.
- 彭広陸 (2013) 「中国語の新語に見られる日本語からの借用語」『日本語学』32(13): 14–25.
- Inoue, Fumio (1991.3) *A Glossary of Hawaiian Japanese* (private edition)
- 井上史雄 (2001.8) 『日本語は生き残れるか—経済言語学の視点から』PHP新書.
- Inoue Fumio (2005) Econolinguistic Aspects of Multilingual Signs in Japan, Changing Language Regimes in Globalizing Environments: Japan and Europe. *International Journal of the Sociology of Language*, 175/176: 157–177.
- 井上史雄 (2007.10) 「外行語カキ」『日本語学』26–10(2013.9『ことばの散歩道』明治書院に再録)
- 井上史雄 (2008.5) 『社会方言学論考---新方言の基盤---』明治書院
- 井上史雄 (2011.11) 『経済言語学論考—言語・方言・敬語の値打ち—』明治書院.
- 井上史雄 (2011.2) 「Google言語地理学入門」『明海日本語』16: 43–52.
- Inoue, Fumio (2012.1) Improvements in the sociolinguistic status of dialects as observed through linguistic landscapes. *Dialectologia* 8: 85–132.
- 井上史雄 (2012.2) 「日本語世界進出のグーグル言語地理学—グーグルインサイトにみる外行語総合分布—」『明海日本語』17: 29–42.
- 井上史雄 (2012.10) 「日本語の世界進出—グーグルでみる外行語—」. 陣内他編『外来語研究の新展開』おうふう.
- 井上史雄 (2013.3.21) 「日本語の国際進出—グーグルインサイトにみる外行語のトレンド—」『明海大学外国語学部論集』25: 1–15.
- 井上史雄 (2013.3.31) 「ハワイと世界の外行語」『応用言語学研究』15: 93–103.

- Inoue, Fumio (2013.4) *Google Linguistic Geography and Worldwide Borrowings — Interlingual dialectology in Japanese and Spanish — (Special Lecture) Universitat de Barcelona.*
- 井上史雄 (2013.12) 「外行語と外来語」『日本語学』32(12): 70–79.
- 井上史雄 (2015.3) 「外行語の世界分布地図—Google Trendsデータの因子分析—」『明海日本語』20: 1–10.
- 井上史雄 (2015.5) 「言語景観の表層と深層—政治と経済—」『日本学The Ilbon-Hak』40: 85–107.
- 井上史雄 (2016.3) 「データの視覚化 (4) — Excel散布図のグラフ・地図への応用 —」『計量国語学』30(4): 216–233.
- 井上史雄 (2016.12) 「オリンピックの経済言語学—グーグル検索と言語景観—」『日本語教育』165: 3–17.
- 井上史雄 (2017.10) 「経済言語学と言語景観」『ことばと社会』19: 26–58
- 井上史雄 (2018.1) 「言語景観与語言經濟」『中国語言戰略 China Language Strategies』6(1): 6–18.
- 井上史雄 (2018. 8) 「東アジアの日本語景観と外行語の地域差」『東アジア日本語教育日本文化研究学会2018年度国際学術発表大会発表要旨』p.96
- 井上史雄 (2021.12) 『ことばの格差』kindle
- Inoue, Fumio (2022.1) *Geolinguistics and Interlingual Dialectology of Lendword …Global borrowing from Japanese in Google Trends.* In Manuela Nevaci, Irina Floarea, Ioan-Mircea Farcaș (eds.) *Ex Oriente Lux In Honorem Nicolae Saramandu*, 617–632. Alessandria: Edizioni dell'Orso.
- 井上史雄・柳村裕 (2015.3) 「外行語の世界分布地図—Google Trendsデータの因子分析—」『明海日本語』20: 1–10.
- 井上史雄・柳村裕 (2015.9) 「外行語世界分布の国別因子分析— Google Trendsによる傾向 —」『計量国語学』30(2): 73–97.
- 石綿敏雄 (1985) 『日本語の中の外国語』岩波新書.
- 石綿敏雄 (2005) 「江戸時代までの外行語 (特集=日本語に入ったことば, 日本語から出たことば) — (「外行」語—ソトへ出た日本語) —」『国文学解釈と鑑賞』70(1): 89–95.
- Jaberg, Karl (1908) *Sprachegeographie* (Aarau: Sauerländer) (p.6 p.22)
- 加藤秀俊・熊倉功夫 (1999) 『外国語になった日本語の事典』岩波書店.
- 金愛蘭 (2014) 「文章構成機能からみた外来語の基本語化」『計量国語学』29(6): 211–226.
- 小林孝郎 (2014) 「英語辞典の「日本語借用語」に見る日本語文化の指標語句」『拓殖大学日本語紀要』
- 国立国語研究所 (2002) 『東アジアにおける日本語観国際センサス』国立国語研究所.
- Long, Daniel & Keisuke Imamura (2013) *The Japanese Language in Palau*. Tachikawa: NINJAL
- Mavragani, Amaryllis, G Ochoa, KP Tsagarakis (2018) Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. *Journal of Medical Internet Research*. <https://doi.org/10.2196/jmir.9366>
- Michel, Jean-Baptiste et al. (2010) *Quantitative Analysis of Culture Using Millions of Digitized Books* Downloaded from www.sciencemag.org on December 19, 2010
- 三輪卓爾 (1977) 「外行語の昨日と今日—海を渡った日本語—」『言語生活』312: 49–57.
- 永島恭子 (2013) 「グーグルインサイトに見る外行語の分布」『明海日本語』第18号オンライン論文集: 337–343.

- 中川裕 (2005) 「アイヌ語にくわわった日本語 (特集= 日本語に入ったことば, 日本語から出たことば)-」 (「外行」語--ソトへ出た日本語) 国文学解釈と鑑賞 70(1): 96-104.
- 荻野綱男 (2014) 『ウェブ検索による日本語研究』朝倉書店.
- 荻野綱男・田野村忠温 (2011) 『講座ITと日本語研究』1-7巻. 明治書院.
- Oja, Vilja (2002) Words for flax in the Finnic languages, *Linguistica Uralica* 38(2): p. 81-
- Saramandu, Nicolae and Manuela Nevaci (2019) Greek terms in daco-romanian sub-dialects (based on ALRR. Sinteză. Vol. I-II) 337, *Balkan Area Altaic Studies General Linguistics In Memoriam Albina H. Girfanova (1957-2018)*, 337-343. Russian Academy of Sciences.
- 島田めぐみ (2018) 「ハワイにおける借用語habut-の使用実態について—ハッシュタグ検索を用いて—」『学芸国語国文学』50: 294-285. [https://doi.org/10.24672/gkokugokokubun.50.0\\_294](https://doi.org/10.24672/gkokugokokubun.50.0_294)
- 新プロ「日本語」総括班編 (1999) 『日本語観国際センサス 単純集計表 (暫定速報版)』国立国語研究所. [http://www.ninjal.ac.jp/archives/n\\_census/result/](http://www.ninjal.ac.jp/archives/n_census/result/)
- 有働ななえ (2016) 「ブランド品消費のナウキャストイング—Google Trendを使った予測」『跡見学園女子大学マネジメント学部紀要』22: 117-137.
- Ueda, Hiroto (1995) Zonificación del Español Urbano. *VARILEX 3* (非売品)
- 上田博人 (1987) 『スペイン語の語彙の頻度と拡がり: 資料集 = Frecuencia y dispersión del vocabulario español』東京外国語大学語学研究所.
- 上田博人 (2013) 「広域スペイン語語彙バリエーション研究における新しい数量化の試み—日本語計量方言学の方法に学ぶ—」『日本語・日本学研究』3: 59-90.
- 上村健太郎 (2014) 「新語・流行語の使用の経年変化: Google Trendsと新聞記事データベースを用いて」『明海日本語』19: 11-20.
- 榎垣実 (1963) 『日本外来語の研究』研究社.
- Weijnen, Antonius (1978) *Outlines for an Interlingual Dialectology*. Assen: Van Gorcum.
- Whorf, Benjamins Lee (1956) *Language, Thought and Reality: Selected Writings of Benjamins Lee Whorf*. Edited by John B. Carroll. Cambridge, Mass.: The M.I.T. Press.
- 梁敏鎬(ヤンミンホ) (2005) 「[日本語観国際センサス]から見た外行語の認知度」『일본어학연구』135-151.
- 梁敏鎬(ヤンミンホ) (2007) 「外来語をめぐる意識に関する日韓対照研究」『国語学研究』46: 73-85.
- Younes, Nadja and Ulf-Dietrich Reips (2019) Guideline for improving the reliability of Google Ngram studies: Evidence from religious terms. *PLoS ONE* 14(3): e0213554 <https://doi.org/10.1371/journal.pone.0213554>

## 出版情報

投稿受理日: 2022年4月28日

投稿採択日: 2022年10月4日

Research article

## Geolinguistic analysis of word forms for ‘today’ in Tibetic languages in Yunnan

SUZUKI, Hiroyuki  
Kyoto University

**Abstract:** This article describes the word forms for ‘today’ in Kham Tibetan varieties spoken in Yunnan, China, and analyses their distribution. It confirms two types of word forms: (1) /tV/ as the first syllable and (2) /ʔa/ as the first syllable. The distribution of (1) is observed in the Sems-kyi-nyila group, while that of (2) is in the other groups. Despite rGyalthang (the Sems-kyi-nyila group) being the governmental, economic, and religious centre, the second form exhibits a higher expansion power.\*

**Keywords:** Tibetic; Kham Tibetan; lexical distribution; expansion; syllabic coalescence

### 1. Introduction

This article first describes word forms for ‘today’ in Kham Tibetan varieties in the Tibetosphere of Yunnan Province, China, and then examines how they are distributed from a geolinguistic viewpoint. Figure 1 shows the most recent dialectal classification of these varieties (Suzuki 2022c), which is a revised version of Suzuki (2018).

The morphological features of ‘today’ in various languages have been discussed by works such as Iwata (2009, 2021) for Sinitic. However, scholars have not investigated the word form for ‘today’ in Yunnan Tibetan and even in Tibetic languages from a geolinguistic perspective. Its description will provide the first account of day ordinals (cf. Bradley 2007: 138–139; 2013) in Tibetic languages. In addition, I will examine the geolinguistic features of the word form for ‘today’ in Yunnan Tibetan.

---

SUZUKI, Hiroyuki. 2023. Geolinguistic analysis of word forms for ‘today’ in Tibetic languages in Yunnan. *Studies in Geolinguistics* 3: 92–98. doi: <https://doi.org/10.5281/zenodo.8437061>

\* An earlier version of the article was presented at the fourth annual meeting of Geolinguistic Society of Japan. I should thank Ray Iwata, Mitsuaki Endo, and Satoko Shirai for their insightful comments. The work was supported by the ILCAA Joint Research Project ‘Studies in Asian and African Geolinguistics’ and JSPS KAKENHI Grant Numbers JP17H04774, JP18H00670, and JP18H05219.

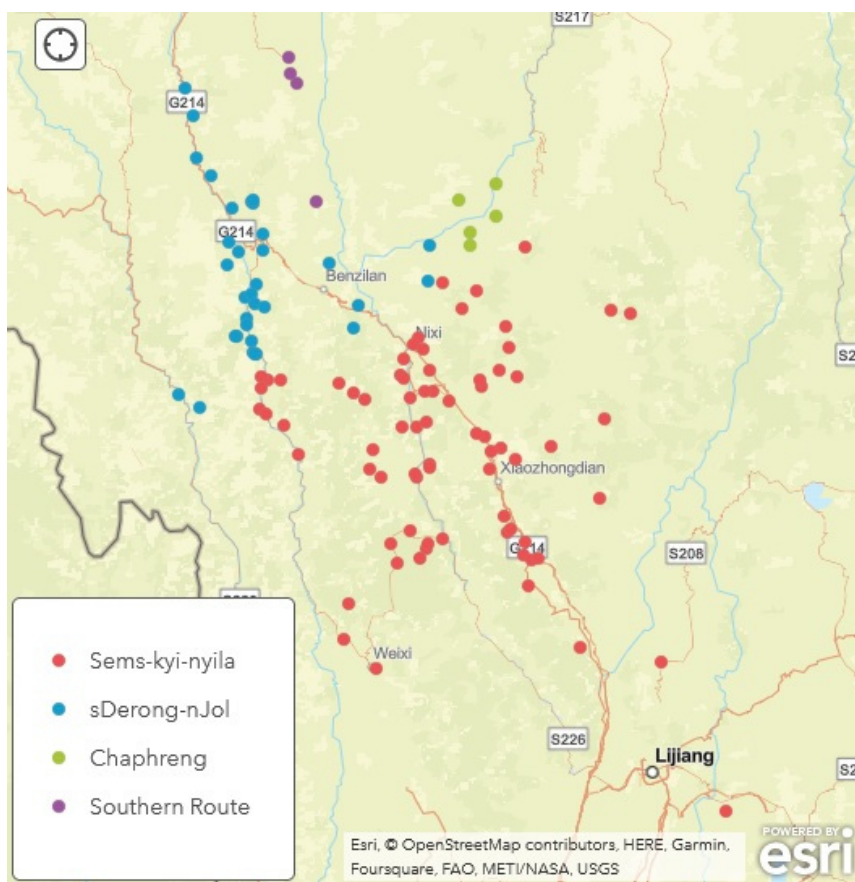


Fig. 1 Classification of Yunnan Tibetan.

## 2. Word forms

The Literary Tibetan (LT) word for ‘today’ is *de ring* (hypothetical literal meaning: ‘that’ + ‘long’), and the spoken forms in the target varieties are classified into four:

- (1a) a disyllabic form corresponding to the LT form, e.g., /tə rī/ and /tə rēj/;
- (1b) a disyllabic form of which the first initial consonant is /t/, partially differing from the LT form, e.g., /<sup>h</sup>teŋ ŋu/ and /tə nu:/;
- (1c) a monosyllabic form of which the initial consonant is /t/, e.g., /tā<sup>s</sup>/; and
- (2) a disyllabic form of which the first syllable is /ʔa/, e.g., /ʔa ri:/ and /ʔa rēj/.

Types (1a–c) are related to LT *de ring*, in which (1c) can be derived from either (1a) or (1b). The second syllable of Type (1b) does not correspond to that of the LT form, *ring*, because its initial is a nasal (/n/ or /ŋ/), which is probably related to LT *nub* ‘west,

sunset’. This word is used as a part of compounds such as *sang nub* ‘tomorrow evening’ and *kha nub* ‘last night’.

Meanwhile, Type (2) uses /ʔa/ instead of the form corresponding to the LT first syllable *de*. The syllable /ʔa/ is often connected to the vicinal demonstrative, ‘this’, although it does not function as a demonstrative alone. Hence, the hypothetical LT form *a ring* would literally mean ‘this’ + ‘long’.

### 3. Geographical distribution and its analyses

Figure 2 shows the distribution of the four types mentioned above.

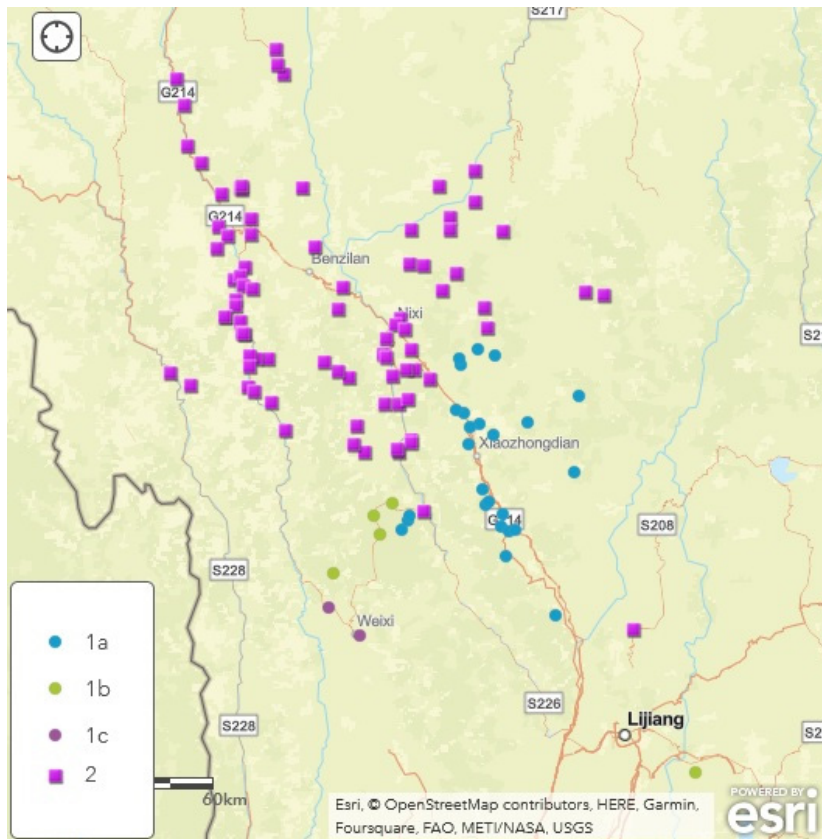


Fig. 2 Word forms for ‘today’.

Comparing this with Figure 1, which reflects the genetic classification based on the shared innovations of the sound change process, we find the broadest distribution of Type (2) whereas Types (1a, b, c) appear in limited varieties of the Sems-kyi-nyila dialect group.



From Figure 2, we can interpret the origins of Types (1b) and (1c), which are only found in the Melung subgroup in the present dataset, as also shown by Daan, a variety belonging to the said subgroup (Suzuki 2009). Suzuki (2022b) proposed a model of Sems-kyi-nyila Tibetan varieties that displays a branching of ‘proto-Melung’ and ‘proto-rGyalthangic’ at an earlier stage. If Suzuki’s (2022b) branching model is accurate, Types (1b) and (1c) substituted Type (1a) after the Melung group separated from other Sems-kyi-nyila Tibetan varieties.

Considering the phonetic realisation of Type (1c), we observe the existence of the pharyngealised feature as in /t̪ʰˤ/. Suzuki (2011) argued that pharyngealisation in the Zhollam variety originated from either the /r/-sound preceding the vowel or coalescence of two syllables. Hence, Type (1c) can originate from either (1a) or (1b) depending on the analysis. Nevertheless, no evidence has demonstrated the Zhollam variety’s independent development in the lexical features within the Melung subgroup. Hence, it is more probable that Type (1c) originates from a coalescence of the disyllabic form of Type (1b).

The word forms of Type (2) are considered recent forms since we find no LT form that corresponds to that type. They appear in all dialectal groups spoken in Yunnan. The centre of Type 2’s distribution is in the north-west area of Figure 2, and the distribution is extended to a geographically contiguous area from the north-west to the south-east, based on the Yunnan Tibetosphere. Figure 3 illustrates a combination pattern of the dialectal classification and the word forms.

As Figure 3 shows, the varieties of the Sems-kyi-nyila group spoken in the connecting zone with the other dialectal groups accepted Type (2). They belong to the East Yunling Mountain, dNgo, and Lamdo subgroups, with the last two exhibiting a degree of commonality in phonological development process with the sDerong-nJol and Chaphreng groups. Hence, it is reasonable that they share lexical features with these non-Sems-kyi-nyila groups.

However, we consider it a striking phenomenon that most East Yunling Mountain subgroup varieties systematically received the lexical item for ‘today’ from another dialect group. From this perspective, the varieties spoken in mBacug Village noticeably use Type (1a) between the distribution areas of Type (2) (Qidzong; East Yunling Mountain subgroup) and Type (1b) (mThachu; Melung subgroup). Based on the phonological shared innovations, the mBacug varieties are classified in the East Yunling Mountain subgroup under the Sems-kyi-nyila group. Hence, we can argue that the word form for ‘today’ *has not been substituted* by Type (2). One factor of such retention may be the location of mBacug, which is on the hill remote to the main traffic route.



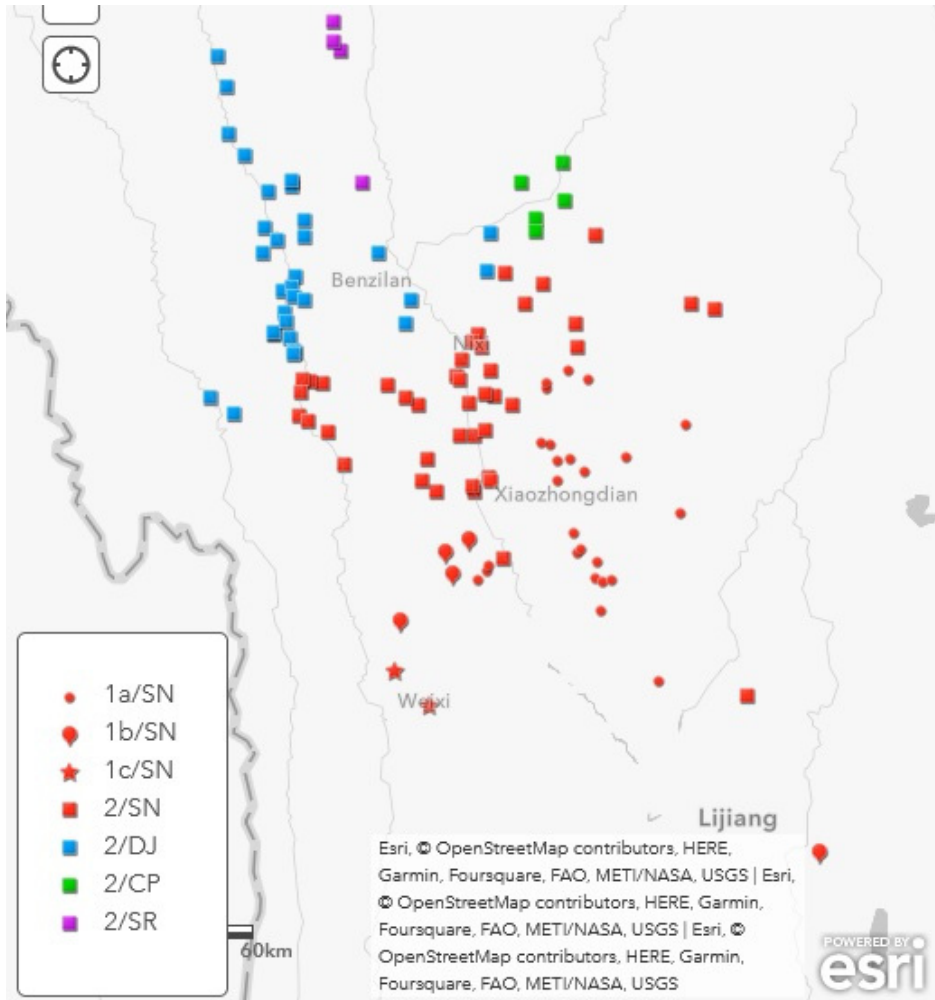


Fig. 3 Relation between the classification and the word forms for 'today'.

For the entire Tibetic languages, the Sems-kyi-nyila group is distributed at the south-eastern periphery of the Tibetosphere (Roche and Suzuki 2018; Tournadre and Suzuki 2023); thus, the case of Sems-kyi-nyila seems to be retention of the archaic form following the concentric theory. However, Type (1a) is still used in the central area of the Tibetosphere. Simply put, Type (2) is a local form developed in the region centred on the north-western part of Yunnan, and Figure 3 shows a situation in which Type (2), a newcomer, gradually expands its power beyond the dialect groups.

#### 4. Conclusion

This article examined the lexical forms for ‘today’ in Tibetic languages in the Yunnan Tibetosphere. Two greater forms are observed: (1) derived from LT *de ring* and *de nub* and (2) derived from another form, *a ring*. The distribution of (1) is confirmed in the Sems-kyi-nyila group, while that of (2) is seen in the other groups. However, form (2) gradually expands its power to the periphery of the Sems-kyi-nyila group, substituting form (1). Contrary to the governmental, economic, and religious centre being in rGyalthang (the Sems-kyi-nyila group), form (2) demonstrates more power of expansion.

#### References

- Bradley, David (2007) What elicitation misses: dominant languages, dominant semantics. *Language Documentation and Description* 4: 136–144. doi: <https://doi.org/10.25894/ldd264>
- Bradley, David (2013) Time ordinals in Tibeto-Burman. Paper presented at the 23th meeting of South-eastern Asian Linguistic Society (Bangkok). URI: <http://jseals.org/seals23/bradley2013timep.pdf>
- Iwata, Ray (2009) Ditu 6 Jintian: Zonghe ditu [Map 6 today: Synthesised map]. In Ray Iwata (ed) *Hanyu fangyan jieshi ditu*, 84–87. Tokyo: Hakuteisya. Also available at: <http://chinesedialectgeography.jp/chinesedg/doc/Maps%206%20%20Today.pdf>
- Iwata, Ray (2021) Parallelisms in lexical changes across languages: Analogical changes in Chinese and French time words. *Studies in Geolinguistics* 1: 123–141. doi: <https://doi.org/10.5281/zenodo.5529335>
- Roche, Gerald and Hiroyuki Suzuki (2018) Tibet’s minority languages: Diversity and endangerment. *Modern Asian Studies* 52(4): 1227–1278. doi: <https://doi.org/10.1017/S0026749X1600072X>
- Suzuki, Hiroyuki (2009) Naxi bunkaken no Tibettogo Yongsheng ken Daan [Daan] hoogen no hoogen syozoku [Dialectal affiliation of the Daan dialect of Tibetic spoken in Yongsheng County in the Naxi cultural area]. *Bulletin of National Museum of Ethnology* 34(1): 167–189. doi: <https://doi.org/10.15021/00003919>
- Suzuki, Hiroyuki (2011) Gagatang Zangyu de yanhua yuanyin yu qi laiyan [Pharyngealised vowels of Gagatang Tibetan and their origin]. *Language and Linguistics* 12(2): 477–500.
- Suzuki, Hiroyuki (2022a) *Geolinguistics in the eastern Tibetosphere: An introduction*. Tokyo: Geolinguistic Society of Japan. doi: <https://doi.org/10.5281/zenodo.5989176>
- Suzuki, Hiroyuki (2022b) Shaping rGyalthangic: A historical account of Yunnan Khams. Paper presented at the 55th International Conference of Sino-Tibetan Languages and Linguistics (Kyoto). doi: <https://doi.org/10.13140/RG.2.2.14370.91849>
- Suzuki, Hiroyuki (2022c) Dialectal affiliation of Tibetic varieties in gYagrwa within Yunnan Tibetan. *Kyoto University Linguistic Research* 41: 43–68. doi: <https://doi.org/10.14989/281541>
- Tournadre, Nicolas and Hiroyuki Suzuki (2023) *The Tibetic languages: An introduction to the family of languages derived from Old Tibetan*. Villejuif: LACITO Publications. URI: <https://lacito.cnrs.fr/the-tibetic-languages/>

## **Publication history**

Date received: 20 August 2023

Research article

## ‘Rice,’ ‘sorghum,’ ‘pearl millet’ and ‘finger millet’ in Nilo-Saharan

NAKAO, Shuichiro  
Osaka University

**Abstract:** This study overviews the geographical distribution of the terms for four major types of crops cultivated in Africa, ‘rice,’ ‘sorghum,’ ‘pearl millet’ and ‘finger millet’ in Nilo-Saharan languages. Most Nilo-Saharan languages lack a native term for ‘rice,’ but a number of languages attest a loanword from Arabic or Swahili. On the other hand, these languages attest a (possibly) native terms for ‘sorghum,’ ‘pearl millet’ and ‘finger millet,’ and we will analyze them by heuristically reconstructing proto-forms.\*

**Keywords:** rice; sorghum; pearl millet; finger millet; Nilo-Saharan

### 1. Introduction

This study aims at surveying terms for crops to supplement related articles in *Language Atlas of Asia* (Endo et al. eds. 2021), which includes articles for ‘rice plant,’ and *Language Atlas of Asia and Africa II* (Suzuki et al. eds. 2023), which includes articles for ‘broomcorn millet,’ ‘foxtail millet’ and ‘barnyard millet’. Since broomcorn millet, foxtail millet and barnyard millet are not widely cultivated in Africa, we analyze terms for sorghum, finger millet and pearl millet that are more common in the regions (see Fukushima et al. eds. 2023 for the reference of linguistic data used for this study). In this study, heuristically reconstructed roots are marked by #.

### 2. ‘Rice’

Most Nilo-Saharan languages do not attest a native term for ‘rice’, but a number of them attest a loanword from a local *lingua franca*, namely Arabic (Type A, mostly via

---

NAKAO, Shuichiro. 2023. ‘Rice,’ ‘sorghum,’ ‘pearl millet’ and ‘finger millet’ in Nilo-Saharan. *Studies in Geolinguistics* 3: 99–105. doi: <https://doi.org/10.5281/zenodo.8437074>

\* This study was supported by JSPS KAKENHI Grant Number 18KK0009 and 19K13160. This study is a result of the ILCAA Joint Research Project “Studies in Asian and African Geolinguistics” (jrp000256).

Sudanese Arabic *ruzz* or Chadian Arabic *rizz*, itself a Wanderwort related to Indian roots, cf. Sanskrit *vrihi*) and Swahili (Type B, the standard form is *mchele*, an Indo-Aryan loanword, cf. Hindi/Urdu *cāval*). Saharan languages Kanuri and Tudaga attest a form similar to the major Chadic (Afroasiatic) language of the area, Hausa *shìnkāfā* (Type C), which is possibly in fact a borrowing from Hausa.

Three Western Nilotic languages attest a native term heuristically reconstructed as *#lab* (Type D, Dinka *lop*, Shilluk *alabo* ‘rice grain’ and Anuak *alumo* ‘wild rice’), which resembles Nuer (Western Nilotic) *laap* ‘wheat’, Kuliak Ik *reba* ‘finger millet’ and the possible root for ‘sorghum’ in some other Western Nilotic and Surmic languages (*#labi*). Shilluk is unique among Nilo-Saharan in that it distinguishes between the term for ‘rice grain’ (*alabo*) and ‘rice plant’ (*peth*).

As for the other native terms (Type E), Dar Daju Daju in Chad attests *kasalange* and Zarma in Niger attests *mo*.

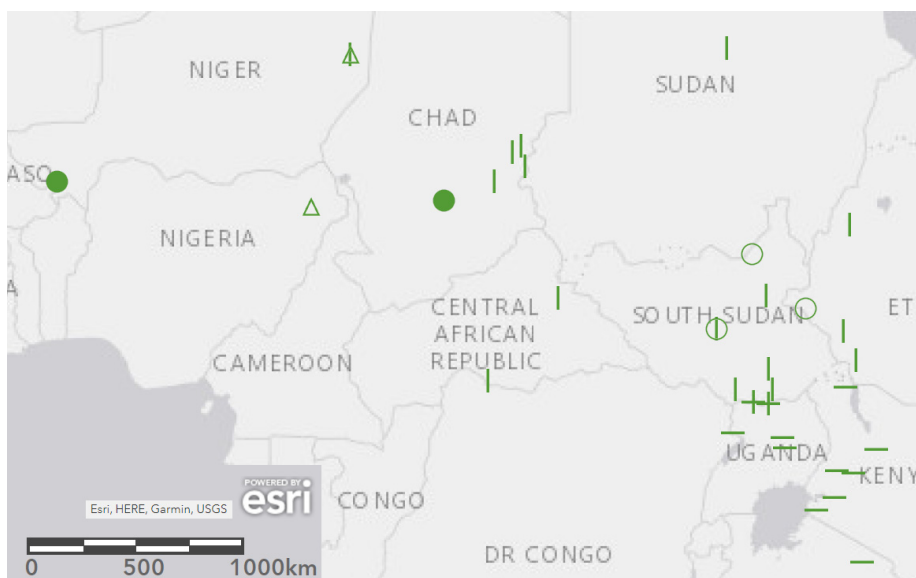


Fig. 1 ‘Rice’ in Nilo-Saharan

Table 1 Signs for Fig. 1

- | A: Arabic borrowing
- B: Swahili borrowing
- △ C: possible Hausa borrowing
- D: *#lab*
- E: the other native terms

### 3. ‘Sorghum’

Sorghum (or guinea corn, sometimes glossed *dura*, a Sudanese Arabic term) is one of the main crops in Sudanic Africa and is attested with various native terms.

Type A #*bel* is commonly attested in Western Nilotic (Reel *beel*, Kumam *bel*, Shilluk *byel*) and could be related to Dinka (Western Nilotic) *bel* ‘durra cane’. Notably, two Central Sudanic languages attest it (Kara *bila*, Yulu *abeel*) probably due to contact.

Type B #*mwa* is common in Eastern Nilotic (Bari *kima*, Otuho *nema*, Turkana *emwae*). Songhay Zarma also attests a similar form *haamo* (compare with Zarma *mo* ‘rice’), but its resemblance with Eastern Nilotic might be coincidental.

Type C #*kari* is commonly attested in Taman (Tama *kari-t*, Abu Sharib *koriñen*) and is also attested in a Saharan language (Zaghawa *gure*).

Type D #*labi* is attested in Surmic (Murle *labitot*, Suri *liway*, Kwegu *ruub*) and Western Sudanic (Dinka *rap*). Compare with Kuliak Ik *reba* ‘finger millet’, Western Nilotic Nuer *laap* ‘wheat’ and the possible Western Nilotic root #*lab* ‘rice’.

Three other types are attested in only one branch. Type E #*jana* (adopted from \**jana* by Otero 2019) is commonly attested in Koman, Type F #*mony* in Central Sudanic (Bongo *mony*, Beli *manya*, Jur-Modo *nyonyu*) and Type G #*kunj* in Daju (Dar Daju *kuñje*, Shatt *kuc*). The other terms (Type H) are attested in only one language.

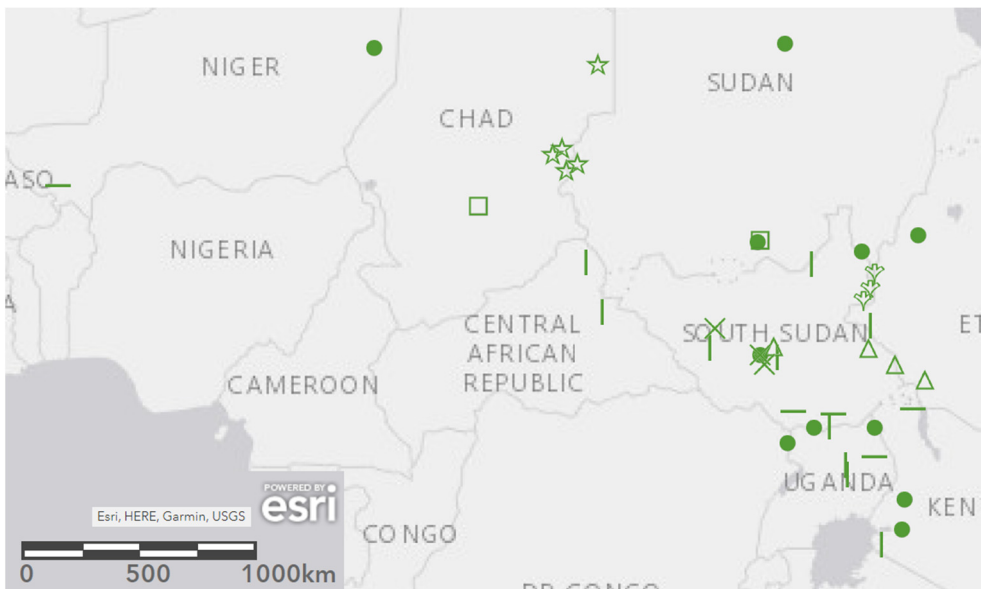


Fig. 2 ‘Sorghum’ in Nilo-Saharan

Table 2 Signs for Fig. 2

	A: # <i>bel</i>
—	B: # <i>mwa</i>
☆	C: # <i>kari</i>
△	D: # <i>labi</i>
✧	E: # <i>jana</i>
×	F: # <i>mony</i>
□	G: # <i>kunj</i>
●	H: the other native terms

#### 4. ‘Pearl millet’

As for pearl millet (bulrush millet, Sudanese Arabic *duxun*), four roots can be heuristically reconstructed.

Type A #*ta’ba* is attested in a Central Sudanic language Yulu (*ta’ba*) and two closely related Eastern Nilotic Maa languages (Chamus *n-tapa*, Samburu *ntapa*). Their resemblance could be merely coincidental.

Type B #*kurwec* is attested in the Daju branch (Dar Daju Daju *kuruce*, Shatt *gurwec* ‘millet’) and the Eastern Nilotic language Bari (*kureja*). Compare with Saharan Zaghawa *gure* ‘sorghum’ and the common Taman root #*kari* ‘sorghum’.

Type C #*kel* is attested in two Central Sudanic languages (Bongo *kel*, Baka *keli*). It resembles the root #*kal* ‘finger millet’ attested in Southern Lwo (Western Nilotic), but their etymological relation is obscure.

Type D #*rau* is attested in Eastern Nilotic Turkana (*erau*) and Western Nilotic Lwo (Acoli *raa*, Belanda Bor *raw* ‘millet’, Pāri *rau* ‘millet’).

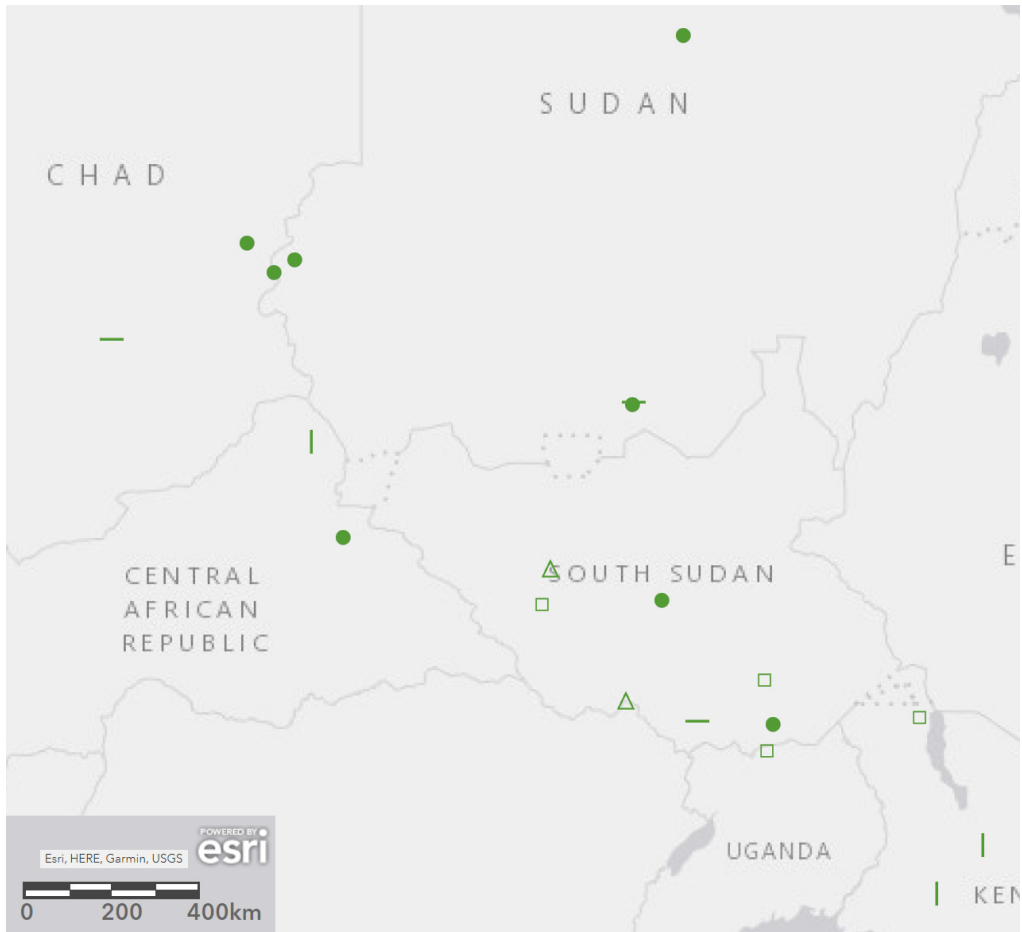


Fig. 3 'Pearl millet' in Nilo-Saharan

Table 3 Signs for Fig. 3

- | A: #ta'ba
- B: #kurwec
- △ C: #kel
- D: #rau
- F: the other native terms



## 5. ‘Finger millet’

As for finger millet (eleusine, Sudanese Arabic *talabūn*), three possible roots are given on the map. Each one is attested in only one branch.

Type A *#k(p)aya* is attested in Central Sudanic (Baka *koyo*, Jur-Modo *kpaya* ‘petit mil’, Yulu *kayə* ‘petit mil’).

Type B *#kima* is attested in Eastern Nilotic (Turkana *akimait*, Karamojong *a-kima-it*). Another Eastern Nilotic language, Bari, attests *kima* ‘sorghum’.

Type C *#kal* is attested in Southern Lwo (Kumam *kal*, Acoli *kal*), which resembles Central Sudanic *#kel* ‘pearl millet’.



Fig. 4 ‘Finger millet’ in Nilo-Saharan

Table 4 Signs for Fig. 3

- | A: *#k(p)aya*
- B: *#kima*
- △ D: *#kal*
- F: the other native terms

## References

- Endo, Mitsuaki, Makoto Minegishi, Satoko Shirai, Hiroyuki Suzuki and Keita Kurabe (eds.) (2021) *Language Atlas of Asia*. Tokyo: Hituzi Syobo.
- Fukushima, Chitsuko, Satoko Shirai, Mika Fukazawa, Hiroyuki Suzuki and Mitsuaki Endo (eds.) (2023) *Language Atlas of Asia and Africa III*. Tokyo: Geolinguistic Society of Japan.
- Suzuki, Hiroyuki, Kohei Nakazawa and Mitsuaki Endo (eds.) (2023) *Language Atlas of Asia and Africa II*. Tokyo: Geolinguistic Society of Japan. doi: <https://doi.org/10.5281/zenodo.7754469>

## Publication history

Date received: 23 August 2023

論文

## 東北地方における当為表現の変化

半沢 康  
福島大学

### Changes in deontic expression in the Tohoku region

HANZAWA, Yasushi  
Fukushima University

**Abstract:** This paper examines the changes in deontic expressions in the Tohoku region. By comparing *The Grammar Atlas of Japanese Dialects* and *New Linguistic Atlas of Japan*, real-time changes were ascertained. The comparison showed that the use of contractions was widespread throughout the Tohoku region. The contraction of deontic expressions was influenced by the progress of grammaticalization.

Glottogram data were analyzed to determine dialectal changes after World War II. There were no major discrepancies with the results of the linguistic map. In the region along the Tohoku-Line, the common language was widely used by the generation born in the 1980s and after, with no dialectal change. However, in the Tsugaru region of Aomori Prefecture, the younger generation continued to use dialectal forms. Even in the areas along the Ou-Line and Uetsu-Line, dialectal forms remained relatively common among the younger generation.

A glottogram survey conducted in the southern Tohoku region showed the existence of an intermediate form between the common language and the dialect. The front element was replaced by the common language form, whereas the rear element retained the dialect form. A survey conducted in Miyagi Prefecture suggested that this intermediate form was a dialectal form that was misidentified with the common language.

Minutes of local government assemblies revealed that this intermediate form was often encountered in the Miyagi and Iwate prefectures and was relatively common in the Hokkaido and Niigata prefectures. These four prefectures adopted the common language at an early stage, which may have led to the emergence and spread of the intermediate form.\*

**キーワード：**経年比較；グロットグラム；中間形；気づかない方言；地方会議録

**Keywords:** real time comparison; glottogram; intermediate forms; unnoticed dialectal forms; minutes of local government assemblies

---

半沢康 (2023) 「東北地方における当為表現の変化」 『地理言語学研究』 3: 106–136. doi: <https://doi.org/10.5281/zenodo.8437084>

\* 本稿はJSPS科研費 JP20K00642の助成による研究成果の一部である。

## 1. はじめに

本稿では東北地方における当為表現の変化について検討する。「しなければならぬ」など義務をあらわす二重否定形式当為表現の中央語における変遷については、田中章夫(1967)、田中章夫(1969)、渋谷勝己(1988)などによって江戸・東京語の様相が確かめられ、また近年、矢島正浩(2013)が関西・大阪の状況も含めた中央語全体の変化をまとめた。湯浅彩央(2019)は関東の文献資料に加えて各種方言地図のデータをも援用し、当為表現の近代の変化について考察している。

これら中央語の変化過程において生じた諸形式の一部は現代に引き継がれ、共通語でもシナケレバ-ナラナイ、シナクテワ-イケナイなど複数の形式が併存する。さらに各地方言に目を転じると、より多様な形式が全国に分布することが『方言文法全国地図』(以下 GAJ と略記)206~208 図、『新日本言語地図』(以下 NLJ と略記)80 図から知られる。日本語全体として非常に多くの変異形を有する表現といえる。

こうした方言量の豊富さは、関東、関西圏以外の諸地域においても、中央語と同様にさまざまな史的変容の生じたことをうかがわせる。既述のとおり、中央語については当為表現の言語史の様相が十分に明らかにされてきているが、その他の地域における変化の状況はほとんど不明のままといつてよい。

もちろん近世以前の豊富な文献資料を活用し、中央語と同等のレベルで変化過程を把握することは困難であるが、それでもこれまでに収集された方言調査データを用いれば、少なくとも近代以降、現代までの進行中の変化の様相を把握することは一定可能となろう。

本稿ではこうした問題意識にもとづき、東北地方の方言を対象として、当為表現の変化概況を明らかにすることを試みる。東北地方では、GAJ、NLJのほか、多数のグロットグラム調査が行われており、地域差に加え、各地の世代差にかんするデータも十分に蓄積されている。

以下、まず GAJ と NLJ のデータを用いて分布の実時間比較を行い、方言が活力を保っていた時代の変化を把握する。さらに前世紀末から今世紀初頭にかけて実施されたグロットグラム調査データを用い、高度経済成長期、共通語の影響を大きく受けた時期の変容を探る。その間、方言衰退が進む一方、共通語の影響を受けた中間形が生じていた状況についても言及する。

## 2. 当為表現分布の経年変化

当為表現の方言形式については、GAJ206~208 図「行かなければならぬ」の全国分布が示され、さらに国立国語研究所による共同研究プロジェクト「方

言の形成過程解明のための全国方言調査」(以下 FPJD と略記)によってその追跡調査が行われた。ただし FPJD のデータにもとづいて公刊された NLJ では、GAJ206 図と 207 図に対応する地図、当為表現を前部要素と後部要素とに分けての作図はなされず、GAJ208 図のような、表現全体の地図(NLJ80 図)のみが示されるにとどまる。

当該表現は、前部、後部要素いずれについても変異形が豊富であり、それらの組み合わせにより全体形の方言量はさらに増える。そのため NLJ80 図では、多様な形式を統合し、さらに出現頻度 4 以下の語形は「その他」として扱われ、同一の記号が与えられている。したがって厳密な経年比較を行うには元データを参照し、NLJ80 図にて「その他」としてまとめられた少数語形の状況にも目を配らなければならない。

GAJ と NLJ を用いた当該項目の方言分布経年比較は、すでに湯浅彩央(2019)によって試みられているが、湯浅彩央(2019)は刊行された地図のみをもとに比較を行っていて、元データに戻っての確認は行われてはいないように見受けられる。

本節では GAJ と FPJD の元データを用い、新たに前後部要素に分けた地図を再描画して経年比較を行う<sup>1</sup>。

## 2.1. 前部要素

当為表現の前部要素(「(行か)なければ-」に該当する部分)の分布図を図 1-1, 1-2 に示す<sup>2</sup>。前者は GAJ, 後者は FPJD データにもとづく。本稿の目的が東北地方における変化をとらえるものであることから、東日本(静岡, 長野, 新潟県以東)の結果のみを示した。FPJD は GAJ に比べて調査地点の密度が低い<sup>3</sup>ため、一見すると方言形が減少したかのようにみえてしまう点、注意が必要である。

分布状況を比較すると、湯浅彩央(2019)が「東日本ではそれほど大きな変化は見受けられない」(p.182)と指摘するとおり、東北地方の 2 回の調査結果は大局的には一致する。図 1-2 にて各地で確認される語形はおおむね図 1-1 でもその近辺に分布している。福島県西南端の南会津郡檜枝岐村には図 1-2 にのみイカザラニ(♥)という語形が確認されるが、これは大島一郎(1959)にも報告されている当地域の古い形式で新形の発生とはみなせない(半沢康 2018a)。なんらかの事情により GAJ 調査時にはたまたま捕捉されなかったものであろう。

<sup>1</sup> 本稿の言語地図はいずれも GIS ソフト MANDARA10 ver10.0.1.6 を使用して作成した(谷謙二 2018)。

<sup>2</sup> 音声の変異形など類似の形式を適宜「類」として統合し、記号化した。中には 1 地点のみで回答された形式もあるが、他と統一しこれも「類」として示す。前部要素については、動詞語幹部分の違いを捨象し、否定辞相当部分を基準にまとめた。一部を除き、凡例には代表形としてイカ-を示している。本文中で言及する場合には基本的に動詞語幹を示さない。なお前後部の分節が不明瞭な一部の形式については [ ] で括って前部要素の地図に示している。

<sup>3</sup> 調査地点数は、GAJ807 地点、FPJD554 地点。

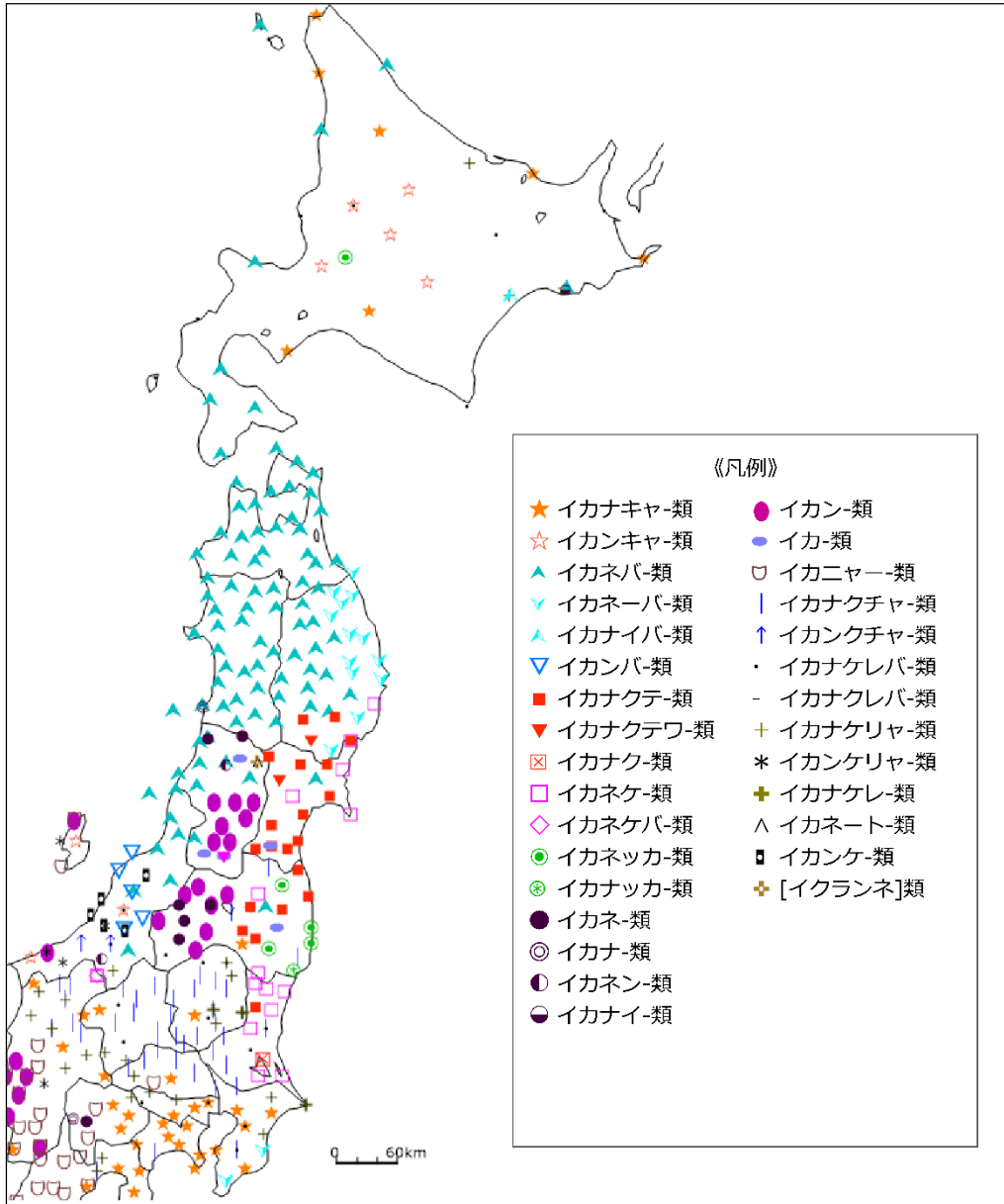


図 1-1:前部要素(行かなければ-)[GAJ]

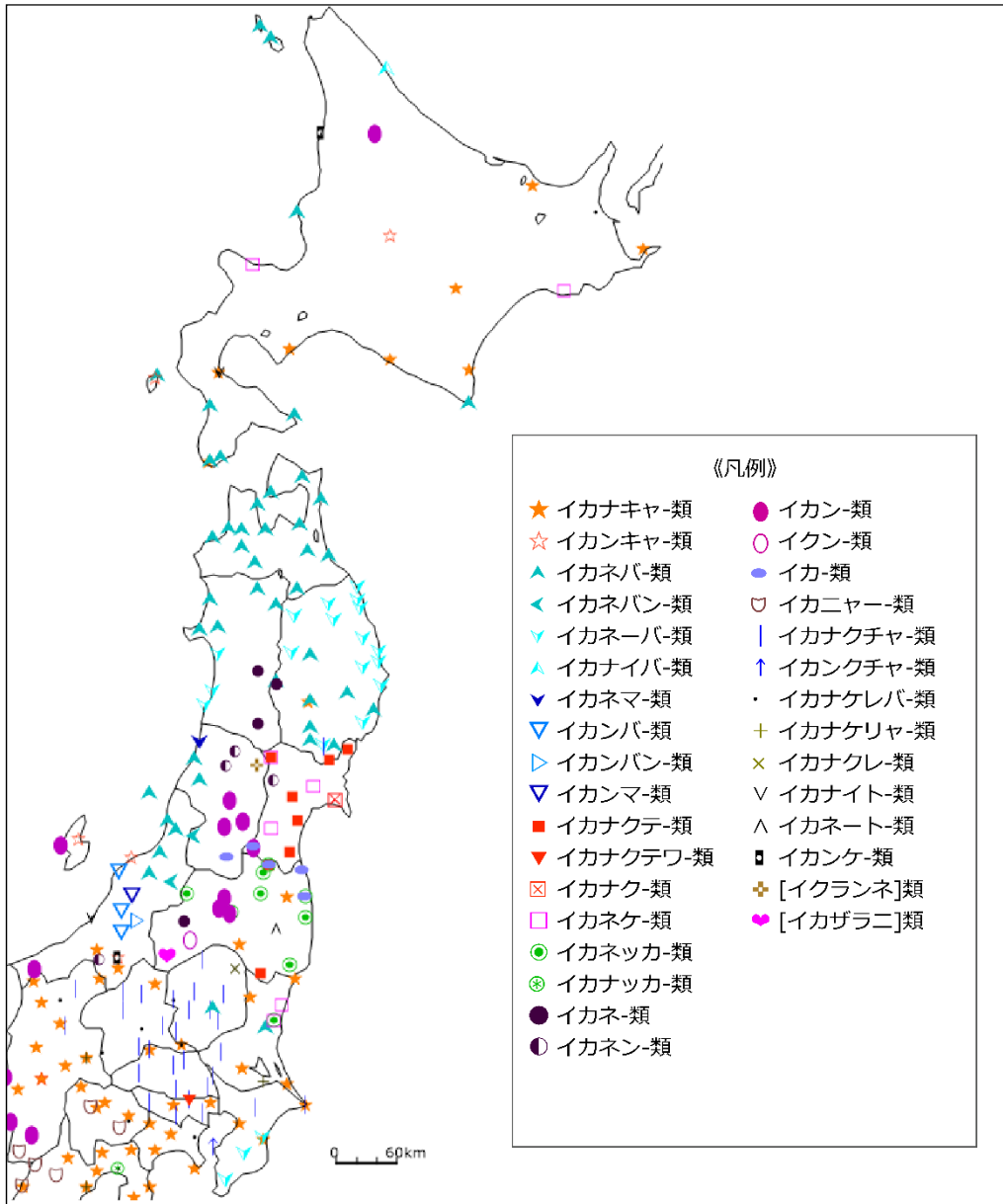


図 1-2:前部要素(行かなければ-)[FPJD]

その一方で、分布変動とおぼしき事象も各地に観察される。図 1-1 において、宮城県を中心に東北南部太平洋側に分布していたナクテ-類(■), ネケ-類(□)はその分布域を縮小した<sup>4</sup>。GAJ 調査時点では岩手県南部(旧伊達藩領域)にもナクテ-類, ネケ-類の使用が一定認められるが、図 1-2 には観察されず、岩手県内に主流の形式ネバ-類(▲)へと置き換わっている。

かつて旧藩境と方言境界とが一致していた地域で、都道府県を単位とした新たな方言分布領域が形成される事例は他にも観測されている<sup>5</sup>。また大西拓一郎(2017)は GAJ, FPJD 両データの比較から、方言の分布変動が都道府県という行政区画をベースに生起することを指摘している(pp.5-8)。近代以降、都道府県を基準とした生活が定着する中で、人々の交流域が変化したことが反映したものと考えられる。

ナクテ-類, ネケ-類はまた福島県内にもみられなくなり、その結果、東北地方ではおおよそ宮城県専用の形式となった。福島県の場合は、岩手県とは異なり、県全域が共通の形式に統一されたわけではないが、GAJ に比して、FPJD データではネッカ-類<sup>6</sup> (●)の勢力がやや強まっている。1980 年代初頭に、福島県から東京にかけて実施された東北線沿線のグロットグラム調査(SF グロットグラム, 井上史雄 1985:p.24)の結果においても、福島県内の調査地点では、高年層(おおよそ 1910 年以前生まれ世代)に多いナクテ-類が、共通語形のナケレバ-類やナキヤ-類のほか、ネッカ-類へ置き換わるという世代差が確認され、図 1-1, 1-2 の比較結果に符合する。

一方で、図 1-1 で秋田、山形県境付近(最上地方)にみられるネ-類(●), ネン-類(●)が、図 1-2 では隣県の秋田県や宮城県へと染み出したかの様子も認められる<sup>7</sup>。後掲図 2-1, 2-2 と照合すると、これらはイカネ-ネ, イカネン-ネという

<sup>4</sup> 東北地方諸方言では助詞ワが未発達であり(日高水穂2023:p.83)、当該形式にもワが用いられないため、関東平野などにみられる縮約形式ナクチャ-類( | )は、方言形としては基本的に存在しない。図1-1, 1-2ではナクテ-類とナクチャ-類を別類とみなし、異なる記号を与えている。

<sup>5</sup> 宮城、福島県境付近における常磐線沿線グロットグラム調査の結果を報告した半沢康(2001)によれば、近代以降に生じたとおぼしき宮城県方言形ジャス(ジャージ)は、旧伊達藩領である福島県相馬郡新地町へはほとんど普及せず、その伝播は宮城、福島県境にとどまる(p.5)。一方、福島県の方言形トツチェ(取りたい)は若い世代に向けて北進傾向を示すが、やはり伝播は福島県内に限られ、県境を越えて宮城県側へは広がらない(p.6)。櫛引祐希子(2012)は、宮城県内に新しく生じたイキナリの程度副詞用法が旧伊達藩領の岩手県南部沿岸地域に認められず、その使用域が宮城県内のみであることを指摘している(p.144)。

<sup>6</sup> 菅野宏(1982)は、このネッカ-類の-ッカを-クワに由来するものと解し、形容詞の条件表現として当為表現に限らず福島県内に広く使用された旨記述している(p.386)。GAJ, FPJDともに他の条件表現ではネッカ-類が確認されないところから、現在では当為表現の専用形式として固定化、残存したようである。当為表現に古い形式が保存されやすいことは、次節の後部要素にも認められる傾向である。

<sup>7</sup> GAJでは、他に山形県庄内地方(飽海郡八幡町:現酒田市)にも1箇所ネ-類が確認される。当該地点の具体的な回答語形は、公開されたデータ([https://www2.ninjal.ac.jp/hogen/dp/gaj\\_all/gaj\\_all.html](https://www2.ninjal.ac.jp/hogen/dp/gaj_all/gaj_all.html))によれば[eganemane]となっている。GAJではeganemaneとmanemaneに分節して記号を与えているが、この回答は周囲に広く分布しているイカネバ-ネの音変異形とみることも可能なように思われる。FPJDデータにもとづく本稿の図1-2, 2-2および後掲のグロットグラム(図4-1, 4-2)ではそのように解釈して記号を与えた。図1-1, 2-1ではGAJの分節にしたがって記号化し、図1-1に●を示しているが、この地点のネ-類は、あるいは最上地方のそれとは異なるものかもしれない。



形式であることがわかる<sup>8</sup>。二重否定形式の当為表現については、文法化にともなう短縮化が各地で生じており、当該形式もその一種とみなせる(国立国語研究所 2002:p.15)。1970年代に山形県最上地方で行われたグロットグラム調査(井上史雄他 1980:p.115)では、シネン-ネ、スネン-ネ(しなければならない)が斜めの統語線を描いて北上する様子がとらえられている。比較的近年に発生した新形式の、周辺への拡大が続いていたようである<sup>9</sup>。

## 2.2. 後部要素

「ならない」に該当する後部要素についても、前部要素同様、GAJとFPJDの状況はおおむね一致するが、やはり局所に変化が生じている(図 2-1, 2-2)。

まず秋田県南部で図 2-1 に広く確認される-デキネ類(✚)が、図 2-2 では全くみられなくなった。

秋田県南部には、GAJ 図 172 「行っただめだ」、図 226 「行ってはいけない」にも-デキネ類が確認されるが、その回答は図 207 に比べて少ない<sup>10</sup>。湯浅彩央(2019:p.186)の指摘のとおり、定型化した当為表現に古い形式が保存されたもので、各用法間に含意性が認められる。FPJD の調査期には、古形が残りやすい当為表現にすら-デキネ類が回答されず、禁止表現(NLJ 図 79 「行ってはいけない」)でも、秋田県内に-デキネ類は当然確認されない<sup>11</sup>。30年間でほぼ消滅したものと目される。井上史雄他(1980)でも1920年代生まれ世代に広く使用されている-デキネ類が、世代が下がるとともに減少する傾向が報告されている(p.116)。

東北地方に特有の-ワカンネ類(◆)も図 2-2 で回答が減少する傾向にあるが、それでも-デキネ類に比べればなお命脈を保つ。秋田県南部の-デキネ類と同様、GAJ 図 172, 図 226 に使用が確認され、さらに禁止表現についてはFPJD調査期でも残存している。

<sup>8</sup> 図2-2では記号が重なってみえないが、元データを確認すると岩手、秋田県境付近の岩手県沢内村ではイガネバ-ワガンネとともにイガネ-ネが回答されている。

<sup>9</sup> 本稿の考察の対象外ではあるが、関東平野に目を転じると、内陸部北部にはナクチャ-類、南部にはナキヤ-類(★)が分布し、その大局は2図で大きくは変わらない。独立した2回の調査で同様の傾向がとらえられていることから、関東南北の本来の方言差と認められよう。なお、図1-1に比べて図1-2では東京都内にナクチャ-類が増加しており、関東北部方言の流入が想起される。田中章夫(1969)などの先行研究によれば、文献史上、江戸・東京ではナクテワ-類の出現がナケレバ-類に遅れるという。ナクテワ-類、ナクチャ-類の発生、普及が北関東で先行し、その後東京へ持ち込まれたと考え、文献史の結果と図1-1, 1-2の分布変動の状況とが整合する。ただし大橋勝男(1975)の当該項目地図(Map.100)ではすでに都内にナクチャ-類が確認されることから、GAJとFPJDの結果の異なりは、あるいは調査誤差に起因する可能性も残る。またGAJ調査期に関東平野の周縁に分布していたナケリヤ-類(+ )は、FPJD期には消滅傾向にある。

<sup>10</sup> 秋田県教育委員会編(2000)には禁止表現としてデギネアが使用されることが示されている(p.112)。

<sup>11</sup> FPJDの元データも確認した。

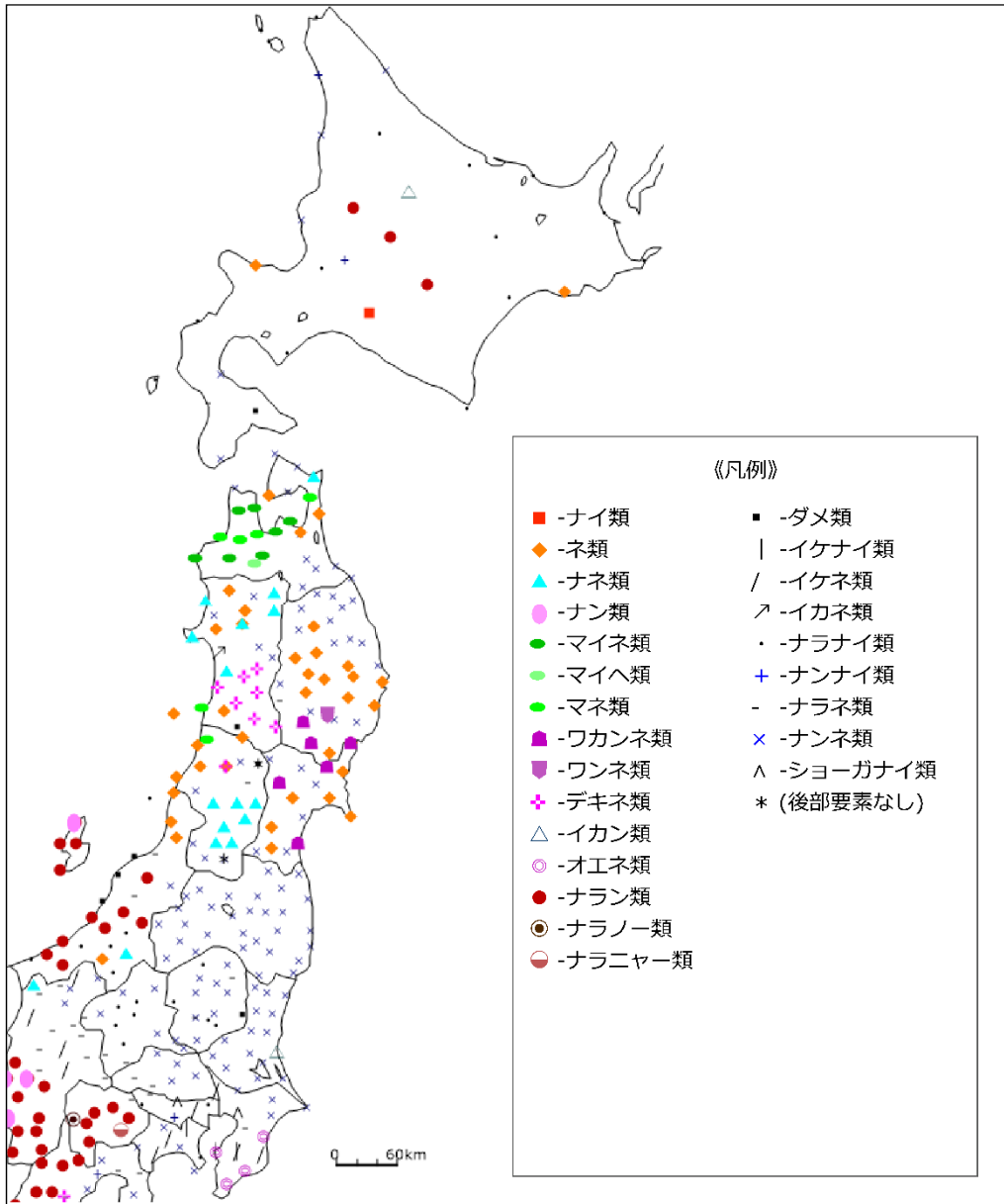


図 2-1:後部要素(-ならない)[GAJ]

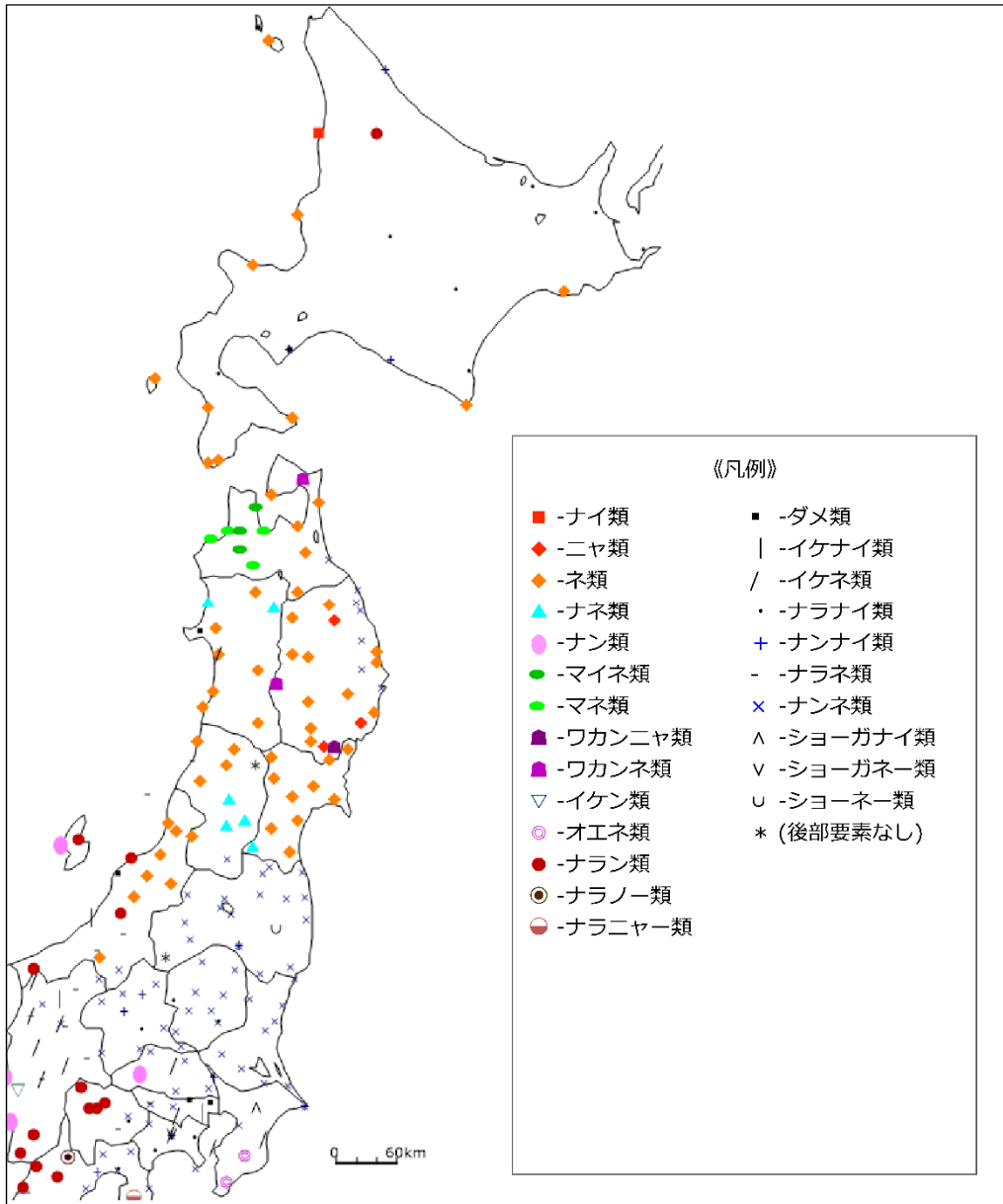


図 2-2:後部要素(-ならない)[FPJD]

こうした伝統的方言形衰退の様相が確認される一方、-ネ類(◆)は30年間でその分布域を拡張した。GAJ 調査期には分布のなかった岩手県北部から青森県太平洋側の南部にまで-ネ類が広がり、もともとの-ナンネ類(×)は太平洋沿岸北部のごく狭い範囲まで退縮している。この結果、宮城県から青森県旧南部藩領まで、太平洋沿岸域を広く-ネ類が覆ったことになる。-デキネ類が消滅した秋田県南部へも-ネ類が進出し、秋田全県へ広がった。東北地方以外でも、新潟県において、GAJ 期には下越地方にのみ確認された-ネ類が、FPJD 期には中越域にまで達している。他に北海道でも沿岸部を中心に-ネ類が増加している。

東日本各地において、30年間でその分布域が拡大していることから、東日本の-ネ類は古い形式の残存ではなく、-ナンネ類などが短縮化して生じた比較的新しい形式とみてよいと思われる<sup>12</sup>。2.2 節にて拡張傾向が確認されたイカネ-ネ、イカネン-ネもこの後部要素を持つ。さらに山形県村山地方のイカン-ナンネ、福島県会津地方のイカン-ナンネ、宮城、山形、福島県に散見されるイカン-ナンネなど、前部要素が短縮化した形式もまた東北地方に目立っている。

シナケレバ-ナラナイなど共通語の当為表現については、機能語として一語化が進行したことが知られるが(花蘭悟 1999)、東北地方をはじめとした諸方言では、その変化がさらに形態にまで及んでいる。

### 3. グロットグラム調査データとの比較

GAJ および FPJD のデータによって当為表現の分布図を再描画し、東北地方における30年間の分布変動を確認した。GAJ のインフォーマントは基本的に1925年以前、FPJD のインフォーマントは1940年以前に生まれた話者をそれぞれ対象としている<sup>13</sup>。いずれも、いまだ各地の方言が健在だった戦前期に出生、生育した方々である。東北地方ではその後、戦後の高度経済成長期以降に劇的な方言変容の時期を迎え、共通語化が急速に進行した。たとえば近世期に編まれた山形県庄内藩の方言集『浜荻』掲載語彙の残存状況をみると、1930年代生まれまでは大きな変化がなく、1940年代から1960年代生まれ世代にかけて使用率が大きく低下する(Inoue & Hanzawa 2021:p.124, 半沢康 2021:p.56)。

当為表現についても同様に、戦後生まれの世代で大きな変化が起きている可能性がある。2 節にみた諸変化が当該世代ではどうなっているか知りたいところであるが、残念ながらより若い世代の地理的分布データは存在しない。そこで、以後の時代の変容について、その概要をグロットグラム調査のデータを援

<sup>12</sup> 湯浅彩央(2019)は、九州中央部の後部要素-ンを東日本の-ネ類と同種の否定辞のみの「同類事項」として「周囲分布」とみなすが(p.187)、少なくとも東北地方の-ネ類は古形の残存とは考えにくい。東北と九州の一致は、それぞれ独自に生じた偶然の結果とみたほうがよさそうである。

<sup>13</sup> GAJ の話者生年の平均は1911.0年(899人)、FPJD の平均は1934.9年(554人)である。

用して把握する。東北地方では、20世紀末から今世紀初頭にかけて多くのグロットグラムが作成されており、主要鉄道沿線地域を中心に、現代の若年層世代をも含めた方言の状況を知ることができる。

### 3.1. THグロットグラム

東北地方の太平洋側を走る東北線福島駅から、津軽線を経由して北海道の函館線滝川駅まで、駅周辺の地域 69 地点を対象に広域のグロットグラム調査が実施された(井上史雄他編 2003)。調査期は 2000 年から 2002 年。東北・北海道のイニシャルをとって TH グロットグラムと呼称される。

この調査では当為表現項目として「しなければならぬ」が設定されている。2 節の言語地図同様、前部要素と後部要素に分け、全調査域の結果を再描画した(図 3-1, 3-2)。記号の与え方は図 1-2, 2-2 に準じている<sup>14</sup>。

図 3-1 前部要素について、グロットグラム調査の結果は、図 1-1, 1-2 の比較によって把握された変化傾向をおおむね再現しているといえよう。1930 年代以前に生まれた高年層では、岩手県中北部より北にネバ-類が広がり、宮城県を中心にナクテ-類がみられる。ナクテ-類は県境を越え、岩手県南部(旧伊達藩領)と福島県北部にも分布しているが、その下の世代(1940~70 年生まれ)では、宮城県内を除いて退縮し、岩手県内ではネバ-類、福島県内ではネッカ-類が主流となる。先にふたつの方言地図を比較して把握された県内部での均質化が、GAJ, FPJD とは独立に実施された TH グロットグラム調査でも同様に確認される。

もっともこうした分布変動は 1980 年代以降に生まれた世代には引き継がれない。青森県の津軽地方では、それでもまだ最若年層に伝統的なネバ-類が確認されるものの、他地域ではナキヤ-類、ナクチャ-類、ナケレバ-類(・)など首都圏でも使用される形式が一様に分布し、上の世代にみられた分布動態をかき消してしまっている。

図 3-2 後部要素の地域差、世代差も、2.2 節に示した方言地図の状況とは矛盾しない。1930 年代以前生まれの高年層に-ナンネ類が分布する岩手県北部地域へ中南部の-ネ類が普及し、一部は青森県内東部にも及ぶ。また岩手、宮城県境にまとまっていた-ワカンネ類は消失傾向を示す。

最若年層世代に共通語形が広まり、方言形が衰退する点も図 3-1 前部要素と同様である。また図 3-1 で、この世代に方言形ネバ-類が確認された青森県津軽地方では、後部要素についても同様に方言形式-マイネ類(●), -マネ類(●)が健在である。鑑水兼貴(2003)は TH グロットグラムデータをもとに共通語形普及の状況を整理し、さらにデータに多変量解析を適用して、東北線沿線地域の共通語化の地域差と世代差を把握した。調査対象の他地域に比べ、青森県西部、

<sup>14</sup> 前部要素は、図 1-1, 1-2 と同様、基本的に動詞語幹部分の違いを捨象して語形を統合し、記号化している。凡例には代表形シ-を示す。

半沢「東北地方における当為表現の変化」

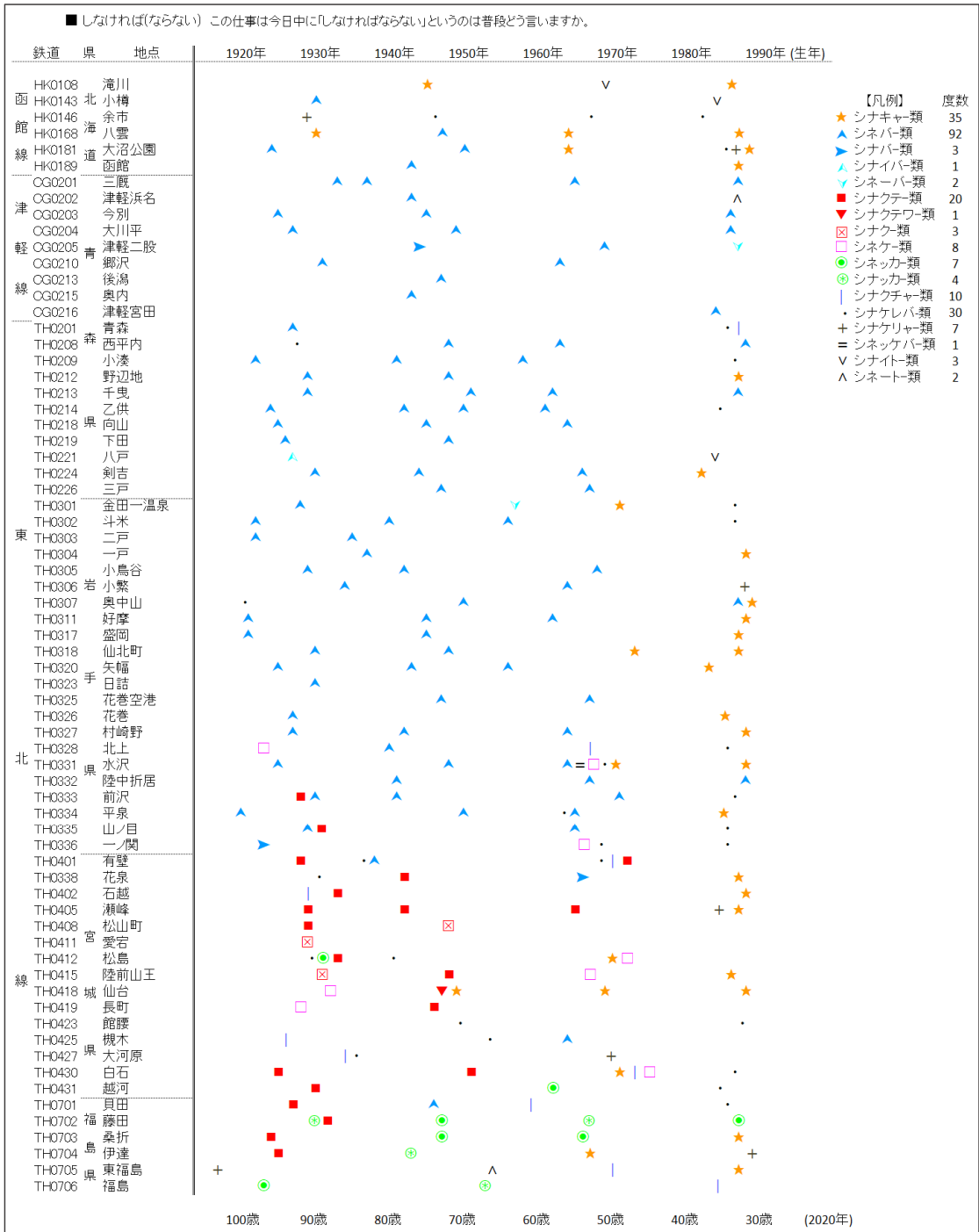


図 3-1:前部要素(しなければ)-[TH グロットグラム]

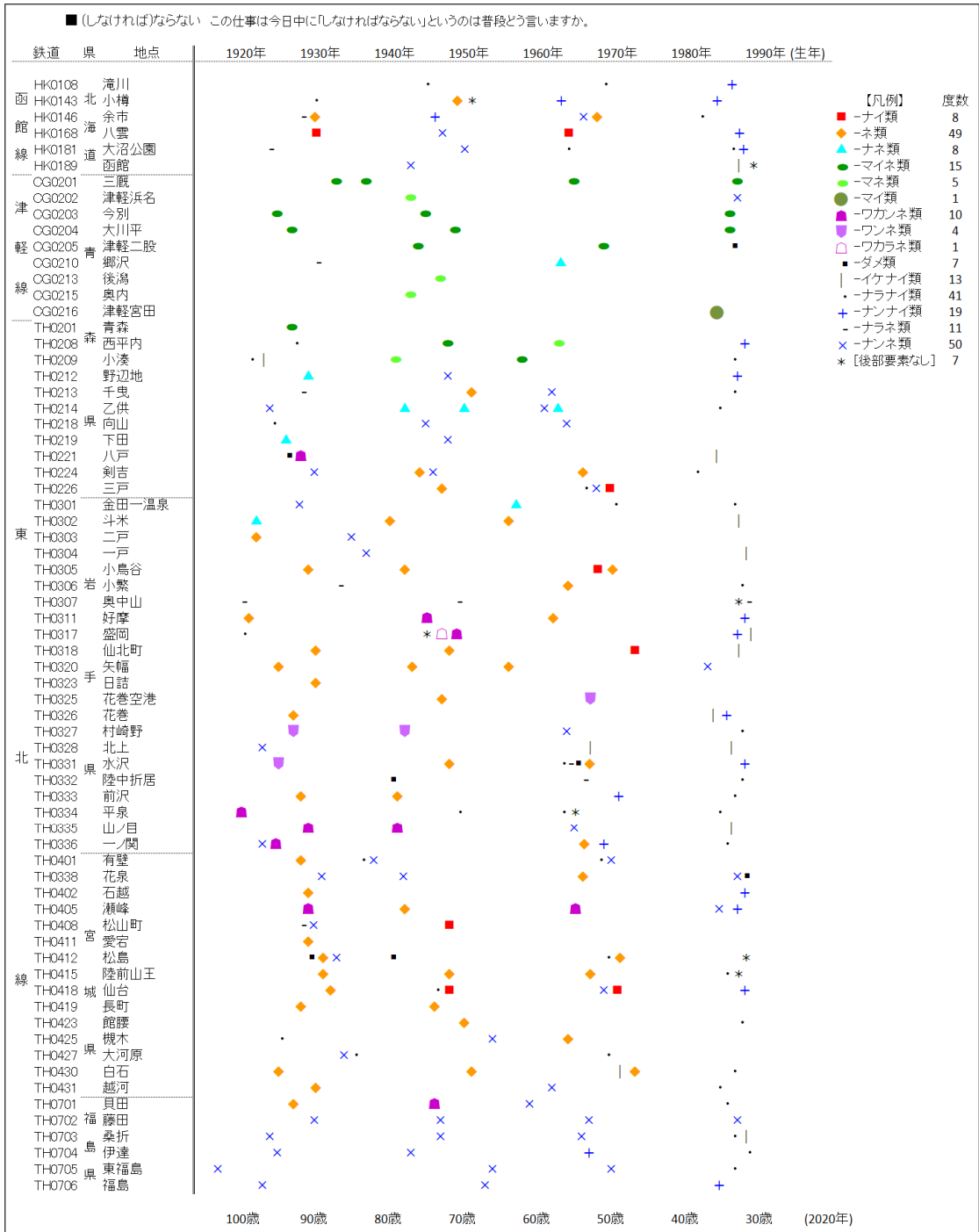


図 3-2:後部要素(-ならない)[TH グロットグラム]



旧津軽藩領域で共通語化の遅れが際立つという。当為表現の変化もこの一般傾向に準じている。

### 3.2. 日本海グロットグラム

東北地方日本海側の地域では、奥羽線、羽越線沿線の調査が行われている。最終的には新潟県内の白新線、信越線を経由して京都まで広域のグロットグラムが作成された(井上史雄編 2008, 井上史雄他編 2011)。ここでは奥羽線浪岡駅から信越線越後岩塚駅まで、青森～新潟県 52 地点の結果を図化する。この地域の調査は 2004 年から 2005 年にかけて実施された。当為表現項目は、TH グロットグラム調査と同様「しなければならない」が採用されている。

図 4-1 前部要素について、東北地方に広く分布するネバ-類は新潟県下越地方にまで及んでおり、高年層の結果は図 1-1, 1-2 と齟齬がない。東北地方のネバ-類はさらに若い世代にも引き継がれ、1980 年代以降生まれの世代にも目立つ。TH グロットグラムではこの世代に広く共通語形が普及していたのとは対照的である。TH グロットグラムデータに日本海グロットグラムデータを加え、東北 6 県の共通語化の傾向を、多重対応分析によって把握した半沢康(2018b)によれば、東北地方の共通語化には地域差がみられる。総じて太平洋側南部で共通語化が先行し、最北部の青森県および日本海側の秋田県、山形県で変化が遅れる。当為表現についてもこの指摘がおおむねあてはまる。

なお、さらに南部の新潟県中越地方には、図 1-1, 1-2 と同様、ンバ-類(▽)、ンキャ-類(☆)が確認される。図 1-1 と図 1-2 の比較ではこの地域で特段の分布変動は認められなかったが、日本海グロットグラムの結果から、1960 年代以降生まれ世代で下越域に向けてンバ-類が伝播しつつあった様子が見えてくる。中心都市である新潟市の形式を周辺が受け入れたものであろうか。調査時の最若年層に共通語形が広がり、近年生じかけたこの分布変動がかき消されてしまっている点は、東北地方太平洋側の地域に通ずる。

図 4-2 後部要素についても、2 枚の言語地図の比較結果と変わらず、全体に-ネ類の進展、北上が目立つ。80 年代以降生まれ世代の共通語化には県による違いがみられる。青森県や山形県では方言形が一定継承されている一方で、秋田県や新潟県には共通語形が広まった<sup>15</sup>。

<sup>15</sup> -デキネ類の消滅という大きな変動がとらえられた秋田県内陸部が調査地点に含まれないため、日本海グロットグラムのデータから当該地域の状況を確認することはできないが、先にみたように井上史雄他(1980)のグロットグラムでは、-デキネ類の衰退が確認される。



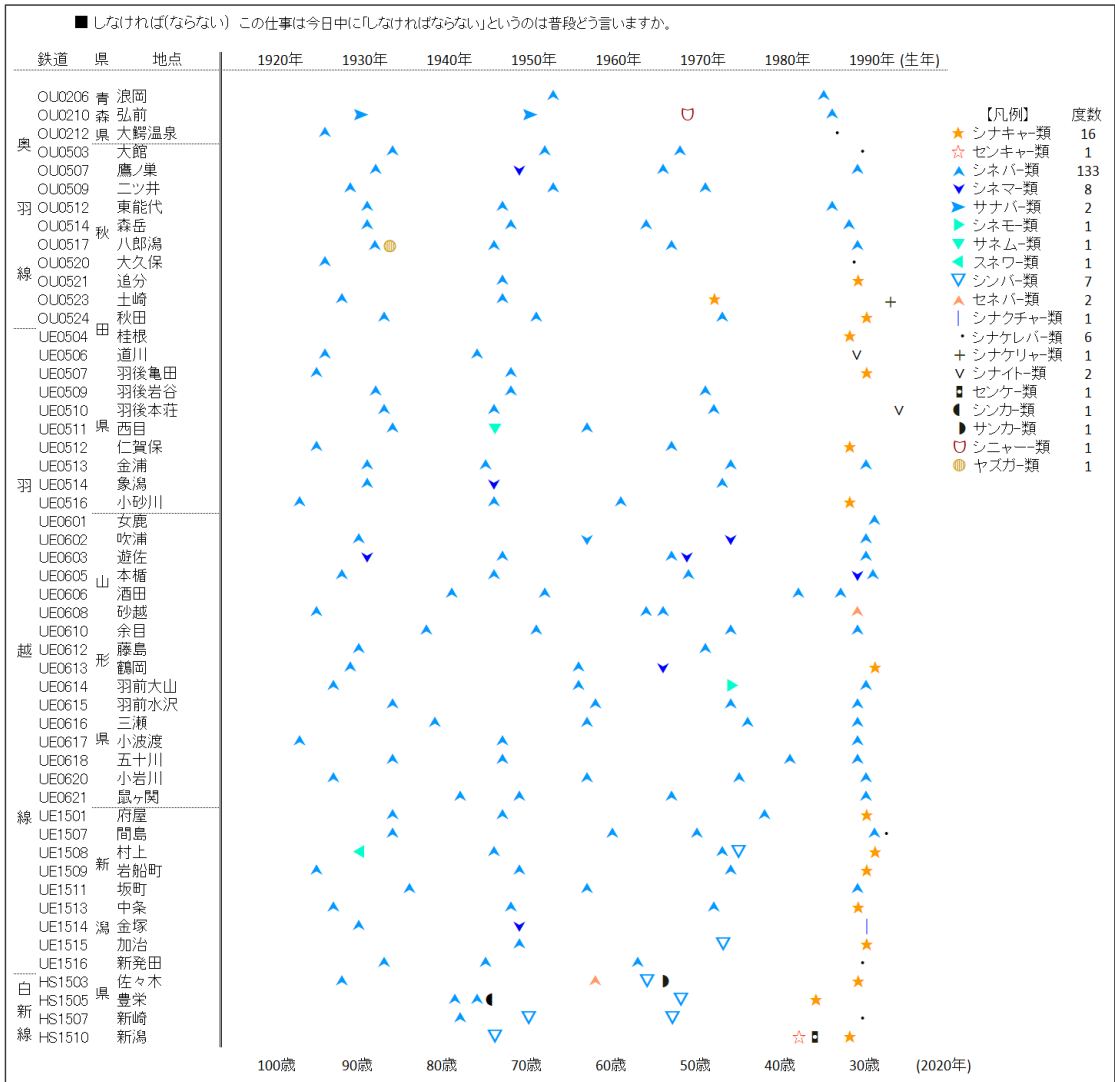


図 4-1:前部要素(しなれば-) [日本海グロットグラム]

半沢「東北地方における当為表現の変化」

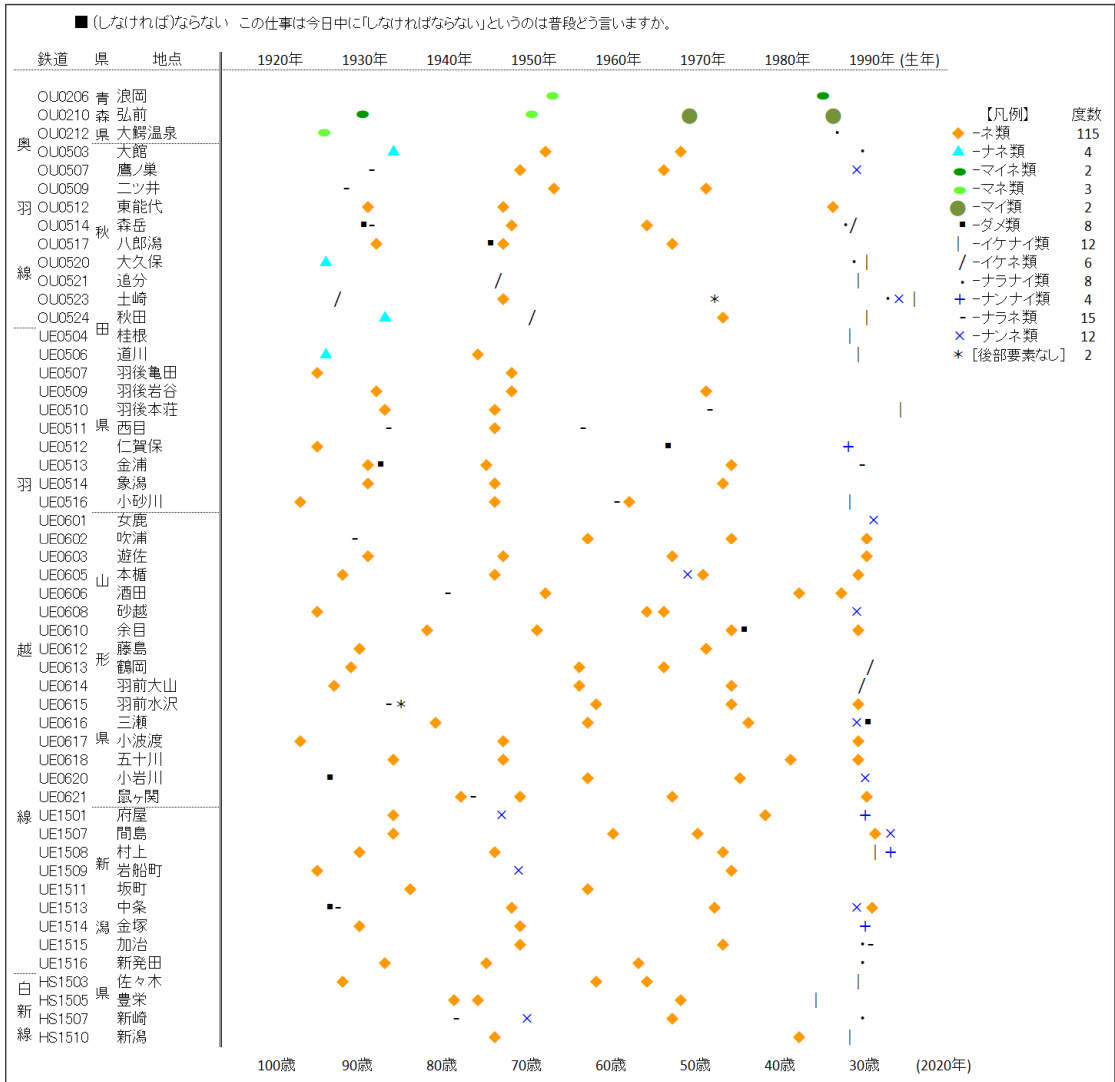


図 4-2:後部要素(-ならない)[日本海グロットグラム]

### 3.3. 東北南部の狭域グロットグラム

太平洋側と日本海側でそれぞれ、東北地方を南北に縦断するグロットグラム調査のデータを用いて、言語地図による経年比較の結果を検証した。グロットグラムに示される見かけ時間データ(世代差)も、2節にみた実時間比較にもとづく分布変動の推定をおおむね支持するものであり、また(鉄道沿線地域に限られるが)より若い世代の変容状況も一部把握することができた。

この2本のグロットグラムに加え、東北地方南部(主に宮城県と福島県)では1990年代以降、多数のグロットグラム調査が実施されている。広域の調査ではとらえきれていない形式が確認されている例もあり、以下に一部紹介しよう。

図5は福島県内で実施した水郡線沿線の調査結果である(2014年実施)。上下に前部要素と後部要素を並べて示した。上下両図を対照すると、福島県南部(東白川郡)を中心に、GAJ, FPJD および広域グロットグラムのいずれにも認められないスル-ヨーダという形式が広く確認される。この形式が当該地域に分布することが想定されたため誘導語形に設定し、調査時にはインフォーマントに使用の有無を必ず確認した<sup>16</sup>。当為表現としてのスル-ヨーダは、東日本各地に報告があり(山浦玄嗣 2000, 松田勇一・高丸圭一 2009, 松田勇一・高丸圭一 2010, 野島本泰 2010, 白岩広行 2019, 大塚貴史・菅野倫匡 2023), GAJ, FPJD, 広域グロットグラムの調査でも、同様に形式を提示して使用の有無を確認すれば東北地方の各地で回答された可能性が残る。共通語の場合と同様、当為表現は同一方言内に複数の形式が併存することが多く、調査時に回答された語形が当該地域の使用語形を網羅するとは限らない。この問題については5節で再度言及する。

図6には1998年に実施した宮城県内陸部の調査結果を示した。図5と同様、上下に前部、後部要素の結果を並べる。かつて存在した旧仙台鉄道および東北線沿線での調査である(武田拓・半沢康 1998)。前部、後部要素ともに GAJ, FPJD および TH グロットグラムにおいて宮城県内に確認される形式とおおむね共通するが、TH グロットグラムに比べ、より年齢の高い世代(1950年代生まれ世代)からナキャ-類の増加することが特徴的である。さらにこのナキャ-類に対応する後部要素を図6下段で確認すると、多く、方言形の-ナイ類(■)や-ネ類が後接して、シナキャ-ナイ, シナキャ-ネといった形式を形成していることがわかる。宮城県内に従来分布したシナクテ-ナイ, シナゲ-ネなどといった方言形の前部要素のみを共通語形に取り換えた、共通語と方言の中間的な形式とみることができる<sup>17</sup>。

<sup>16</sup> 同様の質問法を用いた福島県西部(会津地方)の調査では、スル-ヨーダはまったく確認されないことから、東北地方内部にも地域差の存在することが想定される。

<sup>17</sup> GAJでは北海道に1地点、イカナキャ-ナイが確認されている(国立国語研究所2002:p.15)。THグロットグラムでも前後部要素の対応を確認すると、八雲1932年生、1966年生、仙北町1974年生、水沢1965年生、仙台1948年生、1970年生の6名のインフォーマントがシナキャ-ナイを回答している。

半沢「東北地方における当為表現の変化」

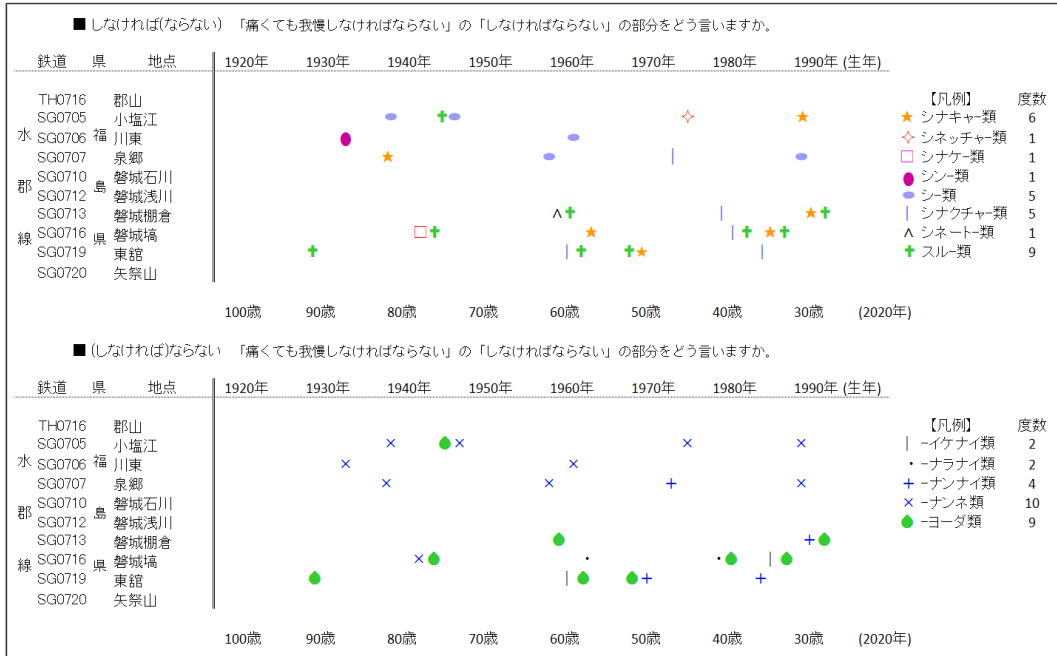


図 5:しなればならない[水郡線グロットグラム]

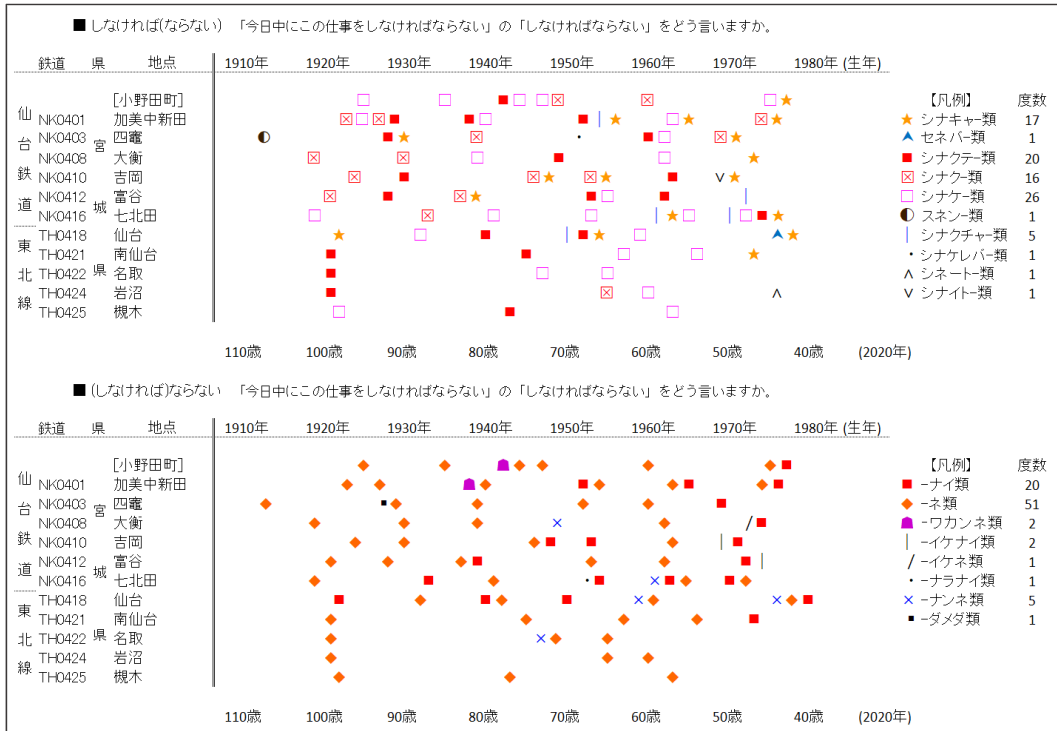


図 6:しなればならない[旧仙台鉄道・東北線グロットグラム]

このグロットグラム調査でも、図 5 水郡線調査と同様の方法で、シナキャ-ナイという形式自体をインフォーマントへ提示し、都度使用の有無を確認している。TH グロットグラムでは拾いきれなかった使用語形が把握されたもので、この中間的な形式が、比較的近年になって宮城県内に広がったことが推察される。

#### 4. 新しい中間形の発生

##### 4.1. 「気づかない方言」としてのシナキャ-ナイ

3.3 節にて、宮城県内に新しい中間形シナキャ-ナイが生じている実態を示した。この当為表現については、前節の宮城県内陸地域のグロットグラム(図 6)とほぼ同時期に実施した仙石線沿線グロットグラム調査でも語形提示方式で使用状況を把握している(図 7)。県の内陸部と同様、この形式がおおむね戦後生まれ世代に広く使用されていることが確認できる。仙石線調査ではまた、使用者が当該形式を方言と共通語のいずれと認識しているのかを知るために、同時に「テレビに出演した場合に使用するか」という質問を用意し(図 7 下段)、使用の場面差もとらえた。日常の使用場面(図 7 上段)に比し、公的な場面でも使用数に遜色がないことから、シナキャ-ナイは使用者に共通語形と誤認された「気づかない方言」とであると目される。前部は共通語形に置き換わり、さらに後部要素も連母音融合形が非融合形に回帰して、形式上は前後部とも共通語形に一致する。地域性を感じにくく、方言形ではないと判断する人が多いものと考えられる。

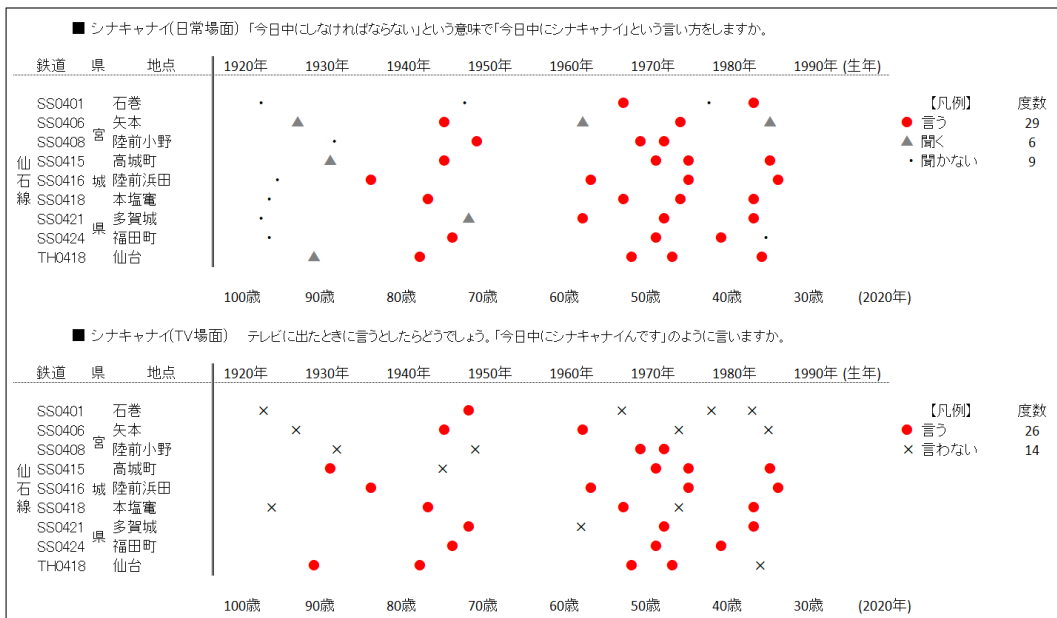


図 7: シナキャ-ナイの使用(仙石線グロットグラム)

このことは、仙台市内で行われた小規模な調査によっても裏付けられる。図 8-1、8-2 は仙台市西部の茂庭台団地で 2010 年に行われた自記式調査の結果である(安齋千秋 2011)。全インフォーマントから宮城県内出身のみを抜き出し(78 名)、世代別の使用率と方言意識(「方言」「共通語」のいずれだと思うか)をまとめた。グロットグラムの結果と同様、使用率は世代が下がるにつれて増加し、シナキャ-ナイが比較的近年に広がった形式であることを示唆する。また使用者の増加にともなってシナキャ-ナイを「共通語形」と認識する割合も増え、若年層では 60%を超える。

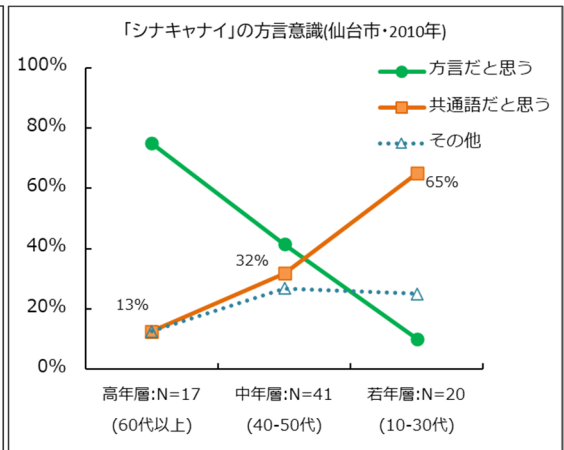
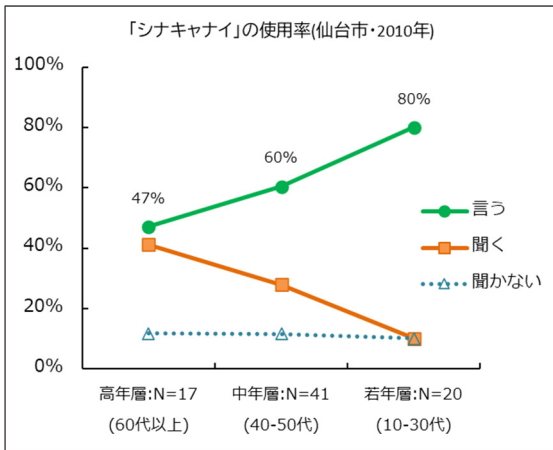


図 8-1:シナキャ-ナイの使用率  
[仙台市]

図 8-2:シナキャ-ナイの方言意識  
[仙台市]

## 4.2. 地方議会での使用実態

4.1 節に示した調査結果より、宮城県内の新しい方言形シナキャ-ナイが共通語形と誤認された「気づかない方言」である可能性が確認された。本節ではシナキャ-ナイのように、共通語の前部要素と方言の後部要素-ナイが結合した諸形式の、地方議会会議録における記載状況を検討する<sup>18</sup>。4.1 節にみたとおりの形式が「気づかない方言」であるなら、都道府県議会や市町村議会という高文体の場面においても使用されることが予測できる。会議録への出現状況を見ることで前節までの調査結果を検証することが可能となる。また、同種の形式が宮城県以外でも発生していた場合、会議録の出現状況からその地理的な分布概況が把握できるかもしれない。

用例の検索には「地方議会会議録コーパス(第 I 期)」の会議録検索システムを使用した<sup>19</sup>。2010 年開始の研究プロジェクトによって構築されたコーパスで

<sup>18</sup> 地方議会会議録を用いた「気づかない方言」の研究には二階堂整(2014)、二階堂整他(2015)、桑戸孝子・上野誠司(2018)などがある。

<sup>19</sup> [https://hkt.tl.fukuoka-u.ac.jp/gijiroku\\_elastic/](https://hkt.tl.fukuoka-u.ac.jp/gijiroku_elastic/) 使用に際しては、小樽商科大学の木村泰知先生をはじめ、プロジェクトの関係者にさまざまご高配を賜った。あらためて厚く御礼を申し上げます。

(齋藤誠他 2011, 木村泰知他 2012), 第 I 期は 425 自治体<sup>20</sup>の議会会議録(2011 年以前のものが収録されている(高丸圭一 2019)。言語研究においても多数の活用事例がある(高丸圭一 2013, 高丸圭一他 2015, 井上史雄 2016, 佐藤亜実 2016, 高丸圭一 2018)。

共通語の当為表現として使用されうる前部要素を網羅し、ナキヤ-ナイのほか、ナクチャ-ナイ、ナクテワ-ナイ、ナイト-ナイ、ナケリヤ-ナイ、ナケレバ-ナイの 6 形を検索ターゲットとした。前部要素の否定辞部分に予想外の変異形が含まれる可能性を考慮し、いずれも冒頭の「ナ」を除いた文字列を検索語として完全一致の用例を検索した。

表 1: 検索結果

検索文字列	抽出数	重複	他用法	検討数	誤記	不明	確定例
いとない	194	3	114	77	23	48	6
くてはない	68	4	1	63	14	2	47
けりやない	73	1	58	14	8	4	2
ければない	2034	101	1552	381	62	47	272
きやない	588	5	120	463	13	24	426
くちやない	82	2	5	75	6	1	68
計	3039	116	1850	1073	126	126	821

結果を表 1 に示す。抽出された全用例を目視確認し、同一の発話の重複(表 1「重複」欄)および当為表現ではないことが明らかな用例(表 1「他用法」欄)を排除した。後者は以下(1)~(6)のようなものである<sup>21</sup>。これらを除いた数が、表 1 の 5 列目「検討数」欄となる。

- (1)\*\*さんも環境についてはもう一方ならぬ熱意をお持ちの方ですから、これは私はやるっきやないと思っているんです。(東京都調布市:2009 年)
- (2)そしてめちゃくちやないいわゆる低価格で落札をし、その結果その落札した工事を下請に出すんですが、(長野県:2005 年)
- (3)やはりその中には養成を図るといふうにありますが、養成を図ってそれをほうっておくてはないと思うのです。(埼玉県坂戸市:1996 年)
- (4)何もみんな払いたいけど、払えないというふうな人やないよ、だれでも払いとないよ。(大阪府八尾市:2002 年)

<sup>20</sup> 高丸圭一(2019)には423自治体と記載されている(p.26)。

<sup>21</sup> 以下、用例の末尾に当該発話のあった地方議会の自治体と、当該会議が開催された年度を示す。都道府県議会の場合は都道府県名のみを、市町村議会の場合は都道府県名に加えて市町村名も示した(都道府県庁所在地は都道府県名省略)。また会議録中に記される個人名は伏字とする。

- (5)なけりゃないと言ってもらわないと答弁にはならない。  
(神奈川県三浦市:1998年)
- (6)制度があっても知らされなければないものと同じだということで、ぜひこのことはやっていただいて、(兵庫県西宮市:2002年)

この「検討数」の中には、さらに下記(7)~(11)に示すような、記録ミス、誤記の可能性の高い例が多数含まれていた(表1「誤記」欄)。

- (7)そうなると、どこかで返さなきゃないない。(静岡県伊東市:2003年)
- (8)1列目に座っている人じゃなくて2列目以降の人もよく考えておかなく  
ちやないならよ。(宇都宮市:2005年)
- (9)生き残れる道は何なのかというのを考えなくちやないらないわけなんです。(沖縄県石垣市:2004年)
- (10)統一化を図っていかなければないけないのかというの、余りしっくりこないと思うわけですが、いかがでしょうか。  
(東京都立川市:2004年)
- (11)それをするためには耐震工事をしなくてはないらないという附帯工事がついてきたわけですね。(京都府宇治市:2005年)

また、文脈から当為表現以外の意味に解釈することが可能な下記(12)~(15)のような例もみられる(表1「不明」欄)。いずれも「ない」が「存在しない」という実質的な意味を有しているとも、また当為表現であるようにも読める。

- (12)そうしますと、ここから12億円ここから仮にあつたとして、土地を売  
らないとない話ですから雲をつかむような話ですが、  
(福島県南相馬市:2000年)
- (13)要するに紙ひもというのはわざわざ買ってこないとないんですよ。  
(神奈川県川崎市:2010年)
- (14)あとは町内会費で出さなきゃないんだけど、そこまでやるということについてはなかなか、町内会の中の出費の問題もあるからということで、  
(青森県八戸市:1997年)
- (15)次のバスというと11時過ぎの、もう受け付けが終わるか終わらないかのバス時刻の、それに乗らなければないというようなことで、  
(埼玉県新座市:2010年)



こうした用例をさらに除き、当為表現として確定的に扱える用例のみを抜き出した。その数が表 1 右端の「確定例」欄となる。ナキャ-ナイ、ナケレバ-ナイが 3 桁のオーダーで確認され、ナイト-ナイ、ナケリャ-ナイは少数にとどまる。確定例の一部を以下に掲載する。

- (16) 地方自治体の行政も民間会社の資本主義社会がやっているんだから、効率的な投資をやらなきゃない。(札幌市:1991 年)
- (17) 現金の取り扱いが、非常に公金なるがゆえに厳正に的確に取り扱わなきゃないというのが私は基本だと思うんですよ。(宮城県:1993 年)
- (18) 市長は私にとんでもないあれが出されたと、これはもう警察にも届けなきゃないし、(新潟県五泉市:2008 年)
- (19) 河川、犀川や梓川、堰など、水害のおそれはないかななどを十分考慮していかなきゃないということであります。(長野県安曇野市:2011 年)
- (20) 何とかこれがプラス・マイナスゼロで、市民の税金を入れなくても運営ができる方向をつくっていかなきゃない。(山口県下松市:2003 年)
- (21) 今、この除雪で軒下が折れたとか、小屋のところのひだがひびが入って取りかえなければいとか、結構あるんですよ。(岩手県紫波町:2010 年)
- (22) 例えば、ガードマンを用意しなければい、職員も用意しなければい、(宮城県石巻市:2008 年)
- (23) 任命権者である新潟市長とすればやっぱり怒らなければいと思うんですよ。(新潟市:2004 年)
- (24) みんなで声をかけ合い、市民一人ひとりの認識の高揚を図らなければいと思います。(福島市:1991 年)
- (25) あなたは在任中はこんなことしまいと思うけれども、いつの日かやめなくちゃいわけだ。(宮城県:1991 年)
- (26) 場合によっては計画を場所を、会場を変更しなくちゃい事態も出てくると思うんです。(新潟県新発田市:2005 年)
- (27) それは脱税の問題と絡むのかどうか詳しく調べてみなくちゃいわけですけれども、(宮城県:2003 年)
- (28) そして滝沢村のあるべき新しいコミュニティと協働、共生というものをつくっていなくちゃいということですから、(岩手県滝沢村:2005 年)

この確定例について、都道府県別に出現数を集計した(表 2)。宮城県および隣接する岩手県に集中し、他には北海道と新潟県に一定の使用が確認される。第 1 期の「地方議会会議録コーパス」は、収載される発話数が都道府県によって大きく異なっていて用例数を単純に比較することはできないので、10 万発話あたりの出現数も算出し、表 2 右端に併せて示している。単純な度数比較と

傾向は変わらず、先の東北2県が突出、他に出現数が1を超えるのは北海道と新潟県の2道県である。

議会における全発言者の出身地が確認できるわけではなく、実際には各都道府県の数値に当該都道府県以外の出身者の発言が含まれている可能性が高い<sup>22</sup>。また(7)～(11)のような明確な誤記は除いたものの、それでもなお他のすべての記録が完全に発言通り記載されているという保証もない<sup>23</sup>。こうした問題が含まれる以上、各都道府県の数値を厳密なものとはできないが<sup>24</sup>、それでもなおこの4道県(とりわけ東北の2県)にみられる数値の偏りは、背後に地域差の存在を措定するに十分であると考えられる。

宮城県とともに当該形式の使用が突出する岩手県は、宮城県同様、当為表現の後部要素として-ネ類、-ナイ類を用いる地域であることが図2-1、2-2より確認される。他に比較的数値の高い北海道と新潟県もまた同様である。4.1節に示したように、やはり当該形式は共通語を受容する過程で本来の方言形の影響を受け、後部要素にのみ方言形式を残した「不完全な共通語形」である蓋然性が高いといえよう。これらが議会という共通語使用が想定される場で多数使用され、また修正の目をすり抜けてそのまま議事録に記載されていることからみても、宮城県内の調査結果のとおり、非共通語形であるという認識が薄いまま使用されているものと思われる。

後部要素に-ネ類、-ナイ類が分布する東北5県のうち、先の2県にのみ中間形が広まった理由は、各県の共通語化の遅速によって説明できそうである。3.1、3.2節に述べたとおり、TH、日本海グロットグラムデータの分析から、東北地方の共通語化は太平洋側の南部で先行している。この地域は1982年に東北新幹線が先行開通し、首都圏からの人の入り込みが増大した。とくにその傾向は、東北地方唯一の政令指定都市仙台市を擁する宮城県に顕著であった。このように東北地方の太平洋側の地域において、共通語が先んじて日常に浸透することにより、方言形との中間形式が生まれたのであろう。

<sup>22</sup> 地方議会議員の87%以上が当該都道府県出身者であるとする竹安栄子(2004)の調査データは存在するが、このデータはあくまでも2002年時点のもので、かつ調査に回答した議員のみについての結果である。また地方議会議録には議員以外の地方自治体職員の発言も含まれている。各議会の発話中、当該都道府県出身者の発話が占める割合を正確に見積もることは非常に難しい。

<sup>23</sup> 二階堂整他(2015)は、web上に公開されている会議録と会議映像を照合し、実際の発言にみられるフィラーや言い間違い、方言形などが修正される場合があることを指摘している(pp.301-303)。

<sup>24</sup> 松田謙次郎(2004)は、地方議会会議録の使用にあたり、その品質のばらつきに留意すべきことを指摘する(p.79)。事実、今回の検索結果でも例(7)～(11)にあげたような誤記とおぼしき例が相当数みついている。また浅川哲也(2019)は「国会会議録検索システム」や「現代日本語書き言葉均衡コーパス」においてすら、入力ミス等に起因する誤りが存在することを指摘している(p.5, p.7)。議会会議録を含む各種コーパスによる用例検索から、こうした不確実性を完全に排除することはおそらく困難であり、精度に一定の限界があることを前提として取り扱わざるを得ないと考えられる。

表 2: 地方議会における中間形の使用数(都道府県別)

	(シ)ナキヤー ナイ	(シ)ナクチャ ナイ	(シ)ナイト ナイ	(シ)ナクテワ ナイ	(シ)ナケリヤ ナイ	(シ)ナケレバ ナイ	合計	出現数/ 10万発話
北海道	45	0	6	0	0	21	72	<b>1.07</b>
青森県	1	0	0	0	0	0	1	0.23
岩手県	120	4	0	1	0	26	151	<b>8.96</b>
宮城県	231	59	0	36	0	114	440	<b>18.51</b>
秋田県	0	0	0	0	0	0	0	0.00
山形県	5	0	0	0	0	0	5	0.34
福島県	1	1	0	1	0	2	5	0.17
茨城県	1	0	0	0	0	2	3	0.21
栃木県	0	0	0	1	0	4	5	0.25
群馬県	0	0	0	0	0	1	1	0.08
埼玉県	0	0	0	4	0	6	10	0.08
千葉県	0	0	0	1	1	7	9	0.19
東京都	0	0	0	0	1	5	6	0.04
神奈川県	4	0	0	0	0	6	10	0.14
新潟県	12	4	0	0	0	5	21	<b>1.06</b>
富山県	0	0	0	1	0	1	2	0.49
石川県	0	0	0	0	0	7	7	0.54
福井県	0	0	0	0	0	3	3	0.35
山梨県	0	0	0	0	0	0	0	0.00
長野県	2	0	0	0	0	8	10	0.17
岐阜県	1	0	0	0	0	1	2	0.08
静岡県	0	0	0	0	0	6	6	0.31
愛知県	1	0	0	0	0	0	1	0.02
三重県	0	0	0	0	0	4	4	0.16
滋賀県	0	0	0	0	0	6	6	0.21
京都府	0	0	0	0	0	0	0	0.00
大阪府	0	0	0	0	0	7	7	0.06
兵庫県	0	0	0	0	0	6	6	0.07
奈良県	0	0	0	0	0	2	2	0.80
和歌山県	0	0	0	0	0	0	0	0.00
鳥取県	0	0	0	0	0	2	2	0.26
島根県	0	0	0	0	0	0	0	0.00
岡山県	0	0	0	0	0	5	5	0.20
広島県	0	0	0	0	0	1	1	0.04
山口県	2	0	0	0	0	1	3	0.14
徳島県	0	0	0	0	0	0	0	0.00
香川県	0	0	0	0	0	0	0	0.00
愛媛県	0	0	0	0	0	0	0	0.00
高知県	0	0	0	0	0	0	0	0.00
福岡県	0	0	0	1	0	0	1	0.08
佐賀県	0	0	0	0	0	2	2	0.14
長崎県	0	0	0	0	0	5	5	0.42
熊本県	0	0	0	0	0	2	2	0.15
大分県	0	0	0	0	0	0	0	0.00
宮崎県	0	0	0	1	0	0	1	0.13
鹿児島県	0	0	0	0	0	1	1	0.13
沖縄県	0	0	0	0	0	3	3	0.07
合計	426	68	6	47	2	272	821	

80年代に高速交通網が整備され、首都圏との時間的距離が短縮されたという点では、新潟県も先の東北2県に共通する。また近代以降の入植によって日本各地の方言が持ち込まれた北海道(の特に内陸部)には、早い時期から全国共通語によく似た北海道共通語が成立したことも知られる。井上史雄(2004)に掲載の、東京からの鉄道距離と共通語使用率(LAJデータ)の相関図(p.22)では、北海道が外れ値を示す。文法項目(GAJデータ)も同様であり(井上史雄2020:p.5)、北海道は中央からの「鉄道距離に比べて例外的に共通語使用率が高い」(p.4)とされている。こうした使用実態はまた「北海道では共通語(標準語)が話されている」という道民の言語意識にも反映する(田中ゆかり・前田忠彦 2012:p.130, 田中ゆかり他 2016:p.141)。

こうしてみると当該中間形が比較的多く確認された4道県には共通の条件が揃っているように見受けられる。他方、同様に-ネ類、-ナイ類を有する東北地方日本海側の地域では共通語の普及が遅れ、変容が起こらないまま、比較的近年まで本来の方言形が保持されたものと推察される<sup>25</sup>。

## 5. まとめ

東北地方における当為表現の変化について、主として戦前に生育した高年層の方言分布変動からは、共通語化の波をかぶる以前の、方言自体の変化が確認された。後部要素については、東北の広い範囲で-ナンネ類などから-ネ類への交代が確認され、また前部要素についても、各地に短縮形の存在が確認された。当為表現の文法化進行とともに、形式についても短縮化、一語化が進行したものと推察される。GAJ, NLJを概観するにこの変化は東北地方以外の他地域方言にも広く確認され、日本語全体の変化の方向性を反映するものと思われる。また一部、従来の分布域が県域に即して変容する状況も確認された。

こうした方言自体の変化はより若い世代では臙化し、地域による遅速の差を示しつつも、各地で着実に共通語化が進行している。そうした方言形衰退の過程において、方言と共通語の中間形式が宮城県、岩手県を中心に新しく広まったことを4節に示した。地方議会会議録の用例からもわかるように、シナキャ-ナイなどの中間形は「共通語スタイル」の発話においても使用されている。図7や図8-2に示される話者の使用意識も参照すると、これらの形式は真田信

<sup>25</sup> 岩崎真梨子他(2018)には、青森県八戸市および周辺地域の大学生、高校生世代でシナイト-ナイが使用されている旨の報告がある(「日常使用する」という回答は大学生78.3%、高校生58.0%)。学生・生徒のみを対象とした調査のため、それ以外の世代の使用状況は不明だが、本稿のTHグロットグラムや地方議会録のデータをふまえると、この地域には比較的近年に新しく普及した可能性が考えられる。もしそうであるとすれば、やはり共通語普及のタイムラグが影響し、宮城、岩手両県に比べて新形発生が遅れたものと解釈することができる。

治(2009)のいう「クァージ標準語」(p.47)の構成要素とみなすことができそうである。

こうした中間形の現状を把握するため、宮城県内の若年層(高校生世代)を対象に、2023年に自記式の小調査を実施した。中間形に加え、共通語形、伝統的方言形、関西方言形を示し、それぞれの使用の有無と方言意識を聞いている<sup>26</sup>。結果の一部を図9-1、9-2に示す。

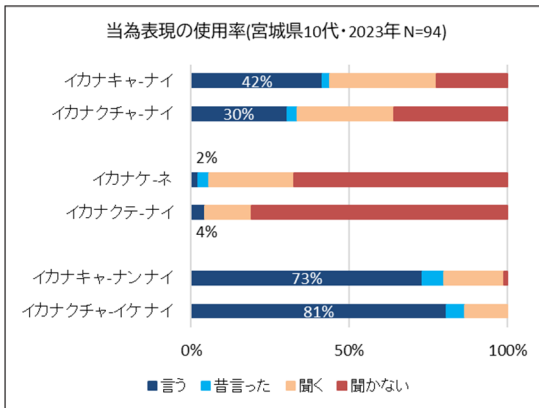


図9-1:当為表現の使用率  
[宮城県10代(2023年)]

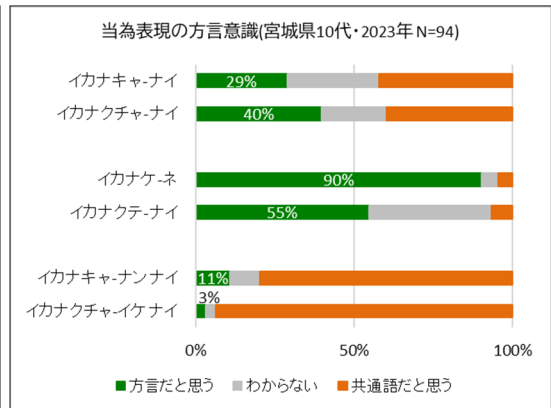


図9-2:当為表現の方言意識  
[宮城県10代(2023年)]

2000年代後半生まれ世代では、伝統的方言形(イカナケ-ネ、イカナクテ-ナイ)に比べれば多いものの、それでも中間形のイカナキャ-ナイ、イカナクチャ-ナイとも使用率は50%を割りこむ。図7や図8-1の若年層(1980~90年代生まれ世代)に比して使用者は減り<sup>27</sup>、また「方言である」という認識も30~40%程度へ上昇している。共通語形の使用率をもっとも高く、それぞれの形式の使用率の多寡が、伝統的方言形から、中間状態を経て、共通語化が完成するという変化過程を反映しているものとみなせる。

小林隆(1999)は、共通語と方言の接触による中間方言の発生は西日本でより活発であるとし、その理由を、東西方言と共通語との距離、異なりの度合いおよび両地域の共通語や方言に対する意識、志向性の違いに求める(p.79)。東北地方ではそもそもこうした中間形式が生じにくく、また仮に発生しても、共通語化の進行とともに方言形であることに気づかれ、結局は淘汰されていくので

<sup>26</sup> 調査にあたっては仙台高等専門学校の武田拓先生にご高配を賜った。ご協力いただいた生徒さんともども、あらためて厚く御礼を申し上げます。

<sup>27</sup> 注25に示すとおり、青森県でも大学生世代に比し、高校生世代では使用率が下がっている(岩崎真梨子他2018)。

あろう。中間形式が若い世代に受け入れられ、ネオ方言形として定着する西日本とは状況が大きく異なっている<sup>28</sup>。

本稿では言語地図およびグロットグラム調査データが、実際の方言分布状況を一定正確に反映していることを前提に考察を進めてきた。しかしながら 3.3 節に示したように、GAJ や NLJ および各グロットグラムに、当該地域で使用されている当為表現すべてが網羅されていない可能性があることには注意しなければならないだろう。当為表現のように、同一地域内に複数の形式が多数併存する場合、同一の話者を対象にまったく同じ調査を短期間のうちに再度繰り返すと、2 回の結果が一致せず、異なる地図が描かれてしまうことが指摘されている(半沢康 2015:p.9)。調査時の質問に対し、その時々で複数の形式のいずれかがランダムに回答されてしまうことがあるためと考えられる。

先に述べた地方議会会議録のデータと同様、言語地図データやグロットグラムデータについても、その質、信頼性(調査結果の一貫性)について精査が必要となるが、本稿ではその点の検討には至らなかった。残された課題である。

## 参考文献/References

- 秋田県教育委員会(編)(2000)『秋田のことば』秋田:無明舎出版。
- 浅川哲也(2019)「そのテキストに信頼性はあるのか—書き間違い・打ち間違い・意図的改変—」『人文学報 日本語教育学』515(07):1-11.
- 安齋千秋(2011)「仙台市方言の研究—方言意識を中心として—」福島大学卒業論文。
- 井上史雄(1985)『関東・東北方言の地理的・年齢的分布(SFグロットグラム)』東京:東京外国語大学。 <http://www.innowayf.net/>
- 井上史雄(2004)「標準語使用率と鉄道距離にみるコミュニケーションの地理的要因」『社会言語科学』7:19-29。 [https://doi.org/10.19024/jajls.7.1\\_19](https://doi.org/10.19024/jajls.7.1_19)
- 井上史雄(2016)「日本語2人称代名詞の地方議会会議録における出現」『明海日本語』21:1-16。 <https://dl.ndl.go.jp/pid/12395113>
- 井上史雄(2020)「文法形式の全国分布と鉄道距離—『方言文法全国地図』の地点別共通語化—」『明海日本語』25:1-9。
- 井上史雄(編)(2008)『日本海沿岸地域方言の地理的・年齢的分布(日本海グロットグラム)』浦安:明海大学。
- 井上史雄・永瀬治郎・沢木幹栄(1980)『最上地方新方言図集』東京:東京外国語大学。
- 井上史雄・玉井宏児・鎌水兼貴(編)(2003)『東北・北海道方言の地理的・年齢的分布(THグロットグラム)』東京:東京外国語大学。

---

<sup>28</sup> 宮城県では、方言文末詞-ツチャを-サに置き換えた確認要求表現-ダサが仙台を中心に広まり、「気づかない方言」として使用されたが(玉懸元2000:pp.78-79)、現在の若い世代では衰退した模様である。こうした中間形式以外にも「クァージ標準語」の構成要素とみなしうるジャス(ジャージ)やナゲル(捨てる)など、方言形であることに気づかれ、衰退した語形が複数確認される。

- 井上史雄・加藤和夫・中井精一・半沢康・山下暁美(編)(2011)『北陸方言の地理的・年齢的分布 (北陸グロットグラム)』浦安:明海大学. <http://www.innowayf.net/>
- 岩崎真梨子・夏坂光男・日比俊介・畑文子(2018)「八戸市の若者の「気づかない方言」と言語活動」『八戸工業大学紀要』37:21-39. <https://hi-tech.repo.nii.ac.jp/records/3833>
- 大島一郎(1959)「福島県檜枝岐村方言の性格について」『人文学報』19:37-54.
- 大塚貴史・菅野倫匡(2023)「茨城県央方言における「評価のモダリティ」を担うヨウダ」『方言の研究』9:133-160.
- 大西拓一郎(2017)「言語変化と方言分布—方言分布形成の理論と経年比較に基づく検証—」大西拓一郎(編)『空間と時間の中の方言』1-20.東京:朝倉書店.
- 大橋勝男(1975)『関東地方域方言事象分布地図 第二巻 表現法篇』東京:桜楓社.
- 菅野宏(1982)「福島県の方言」飯豊毅一・日野資純・佐藤亮一編『講座方言学4 北海道・東北地方の方言』365-397.東京:国書刊行会.
- 木村泰知・渋木英潔・高丸圭一・乙武北斗・森辰則(2012)「地方議会議録コーパスの構築とsoの利用」『2012年度第26回人工知能学会全国大会論文集』. [https://doi.org/10.11517/pjsai.JSAI2012.0\\_3B3NFC43](https://doi.org/10.11517/pjsai.JSAI2012.0_3B3NFC43)
- 櫛引祐希子(2012)「方言特有の「イキナリ」「ナゲル」「オチル」の分布状況」小林隆(編)『宮城県・岩手県三陸地方南部地域方言の研究』141-152.仙台:東北大学国語学研究室.
- 桑戸孝子・上野誠司(2018)「長崎方言話者による「イチゴが売ってある」という表現について—質問紙および地方議会議録調査から—」『長崎総合科学大学紀要』58(2):115-125.
- 国立国語研究所(2002)『方言文法全国地図解説5』東京:国立国語研究所. <http://doi.org/10.15084/00001567>
- 小林隆(1999)「方言語彙・表現法の現在」真田信治(編)『展望現代の方言』62-80.東京:白帝社.
- 齋藤誠・大城卓・菅原晃平・永井隆広・渋木英潔・木村泰知・森辰則(2011)「地方議会議録の収集とコーパスの構築」『言語処理学会第17回年次大会論文集』368-371. [https://www.anlp.jp/proceedings/annual\\_meeting/2011/pdf\\_dir/P2-21.pdf](https://www.anlp.jp/proceedings/annual_meeting/2011/pdf_dir/P2-21.pdf)
- 佐藤亜実(2016)「議録における接尾辞ラヘンの用法記述」『言語科学論集』20:91-103. <http://hdl.handle.net/10097/00065097>
- 真田信治(2009)「方言研究の新たなる出発」『言語』38(12):44-49.
- 渋谷勝己(1988)「江戸語・東京語の当為表現—後部要素イケナイの成立を中心に—」『大阪大学日本学報』7:99-118.
- 白岩広行(2019)「福島県北部方言の「ようだ」」『阪大社会言語学研究ノート』16:1-14. <https://doi.org/10.18910/73636>
- 高丸圭一(2013)「形態素N-gramを用いた地方議会議録における地域差の分析手法の検討—ひらがなで構成された文末の4-gramに着目して—」『明海日本語』18:1-10. <https://dl.ndl.go.jp/pid/10305379>
- 高丸圭一(2018)「地方議会におけるオノマトペの使用分布」小林隆(編)『感性の方言学』119-146.東京:ひつじ書房.
- 高丸圭一(2019)「地方議会議録コーパスと地方議会議録を用いた学術研究の現状」『知能と情報』32(2):25-33. [https://doi.org/10.3156/jsoft.31.2\\_25](https://doi.org/10.3156/jsoft.31.2_25)
- 高丸圭一・内田ゆず・乙武北斗・木村泰知(2015)「地方議会議録コーパスにおけるオノマトペ—出現傾向と語義の分析—」『人工知能学会論文誌』30(1):306-318. <https://doi.org/10.1527/tjsai.30.306>

- 武田拓・半沢康(1998)「仙台市・中新田町間の言語動態」加藤正信・遠藤仁(編)『宮城県中新田町方言の研究』143-188.仙台:宮城教育大学.
- 竹安栄子(2004)「地方議員のジェンダー差異—「2002年全国地方議員調査」結果の分析より—」『京都女子大学現代社会研究』7:99-118. <http://hdl.handle.net/11173/359>
- 田中章夫(1967)「江戸語・東京語における当為表現の変遷」『国語と国文学』44(4):102-114.
- 田中章夫(1969)「近代東京語の当為表現」佐伯梅友博士古稀記念国語学論集刊行会(編)『佐伯博士古稀記念国語学論集』649-673.東京:表現社.
- 田中ゆかり・前田忠彦(2012)「話者分類に基づく地域類型化の試み—全国方言意識調査データを用いた潜在クラス分析による検討—」『国立国語研究所論集』3:117-142.  
<https://doi.org/10.15084/00000493>
- 田中ゆかり・林直樹・前田忠彦・相澤正夫(2016)「1万人調査からみた最新の方言・共通語意識—「2015年全国方言意識Web調査」の報告—」『国立国語研究所論集』11:117-145.  
<https://doi.org/10.15084/00000844>
- 谷謙二(2018)『フリーGISソフトMANDARA10 パーフェクトマスター』東京:古今書院.
- 玉懸元(2000)「終助詞ツチャ, サ」小林隆(編)『宮城県仙台市方言の研究』76-89.仙台:東北大学国語学研究室.
- 二階堂整(2014)「福岡の議会会議録における方言についての一考察」『筑紫日本語研究』2013:151-157.
- 二階堂整・川瀬卓・高丸圭一・田附敏尚・松田謙次郎(2015)「地方議会会議録による方言研究—セミフォーマルと気づかない方言—」『方言の研究』1:299-324.
- 野島本泰(2010)「神奈川県座間市の方言における,「必要」を表す「ようだ」」『地球研言語記述論集』2:73-85. <https://chikyu.repo.nii.ac.jp/records/934>
- 花菌悟(1999)「条件形複合用言形式の認定」『国語学』197:90-104.  
[https://bibdb.ninjal.ac.jp/SJL/view.php?h\\_id=1971040900](https://bibdb.ninjal.ac.jp/SJL/view.php?h_id=1971040900)
- 半沢康(2001)「宮城・福島太平洋沿岸地域の方言動態—常磐線沿線グロットグラム調査の結果から—」『言文』48:1-14.
- 半沢康(2015)「言語地理学調査データの信頼性」『言文』62:1-20. <http://hdl.handle.net/10270/4317>
- 半沢康(2018a)「福島県檜枝岐方言の現状—その独自性と変容—」『日本語学』37(7):2-10.
- 半沢康(2018b)「現代における方言語彙の動態」小林隆(編)『シリーズ日本語の語彙 第8巻 方言の語彙—日本語を彩る地域語の世界—』132-147.東京:朝倉書店.
- 半沢康(2021)「『庄内浜藪』調査データの多変量解析」『福島大学人間発達文化学類紀要』33:49-64. <http://hdl.handle.net/10270/5406>
- 日高水穂(2023)「『方言文法全国地図』と方言地理学」小林隆・大西拓一郎・篠崎晃一(編)『方言地理学の視界』71-88.東京:勉誠出版.
- 松田謙次郎(2004)「言語資料としての国会会議録検索システム」『トークス』7:55-82.  
<https://shoin.repo.nii.ac.jp/records/168>
- 松田勇一・高丸圭一(2009)「栃木方言「～ヨウダ」の用法と使用実態—相手の年齢差と親疎による表現の使用差—」『茨城大学留学生センター紀要』7:65-77. <https://doi.org/10.34405/00009915>
- 松田勇一・高丸圭一(2010)「栃木方言「～ヨウダ」の待遇表現—相手の属性による敬語の使い分け—」『茨城大学留学生センター紀要』8:65-84. <https://doi.org/10.34405/00010121>
- 矢島正浩(2013)『上方・大阪語における条件表現の史的展開』東京:笠間書院.
- 山浦玄嗣(2000)『ケセン語大辞典<下巻>』秋田:無明舎出版.



鍵水兼貴(2003)「東北・北海道における共通語化の地理的・年齢的動態」『日本方言研究会第77回研究発表会発表原稿集』81-88.

湯浅彩央(2019)『近代日本語の当為表現』東京:武蔵野書院.

Inoue, Fumio and Yasushi Hanzawa (2021) Multivariate Analysis of Geography and Age in Dialect Vocabulary: Comprehensive Analysis of 250 Years of Language Change. *Dialectologia: revista electronica* 27, 97-160. <https://doi.org/10.1344/dialectologia2021.27.5>

## 出版情報

投稿受理日：2023年8月31日

Research article

## A further study of kinship terms for ‘mother’ in Vietnamese

TRẦN, Thị Hồng Hạnh

Faculty of Linguistics, USSH, Vietnam National University Hanoi

**Abstract:** This paper offers a comprehensive look at kinship terms in Vietnamese, highlighting the specific case of kinship terms for ‘mother.’ The terms for ‘mother’ are abundant in the Vietnamese kinship term system and include terms that are used nationwide as well as those that have been preserved in dialects. Among them are some cases which seem to have sufficient historical phonetic evidence to argue about their origin. Meanwhile, the origin of other terms for ‘mother’ is still under debate. Based on historical linguistic evidence supported by the interpretation of linguistics maps, this paper aims to give more insight into the origin and development of four Vietnamese kinship terms for ‘mother’: *mẹ*, *mê*, *mạ*, *má*. The findings indicate that there are terms for ‘mother’ that are Austroasiatic etymons (*mẹ*, *mạ*, *mê*), while *má* is likely to be a loanword from Chinese dialects.

**Keywords:** kinship term; mother; Vietnamese; historical linguistics; geolinguistics

### 1. Introduction

The study of Vietnamese kinship terms has indeed captivated the attention of both Vietnamese and foreign linguists for a long time. These scholars have approached these terms from several perspectives, which has contributed to a deep understanding of Vietnamese language and culture. Previous studies by linguists have delved into structural and semantic aspects as well as the use of these terms as pronouns and terms of address in daily social communication.

In terms of historical perspective, the study by Mark Alves (2017) on Chinese loanwords in Vietnamese pronouns and terms of address and reference pointed out the source of some kinship terms, which has made a noteworthy contribution to this issue. According to his findings, a dozen kinship terms were borrowed from Chinese during an early period of Sino–Vietnamese contact and that this borrowing took place over multiple periods of Chinese historical phonology: *thím* (father’s younger brother’s

---

TRẦN, Thị Hồng Hạnh. 2023. A further study of kinship terms for ‘mother’ in Vietnamese. *Studies in Geolinguistics* 3: 137–145. doi: <https://doi.org/10.5281/zenodo.8437092>

wife), *mợ* (mother's younger brother's wife), *chú* (father's younger), *cậu* (mother's younger brother), *bác* (father or mother's elder sister/brother), *chị* (elder sister), *bà* (grandmother), *ông* (grandfather), *cô* (father's younger sister), *đì* (mother's younger sister). Their original semantic features were kept while developing anaphoric functions not seen in Chinese, with a distinction between kinship terms within families versus those used in society more generally. The study also revealed that the etymological source of some of the terms is Austroasiatic: *con*, *cháu*. However, the sources of some of the terms are still unknown: *cụ/cố* (great-grandfather, very old person), *anh* (elder brother), *em* (younger sibling), *duyệt* (aunt's husband), *cha/bố/thầy* (father), *so* (great-grandparent). With regard to the term for 'mother,' Alves listed two cases, *mẹ* and *má*, but the source of these two terms was left open because as Alves noted, comparable forms of *mẹ* and *má* could be found in both Austroasiatic and Tai-Kadai languages.

In my research (Tran Thi Hong Hanh 2021), I discussed the origin of certain kinship terms in the Phú Thọ dialect (a Northern Vietnamese dialect) and indicated that the term *bằm* used to refer to 'mother' here is an Austroasiatic etymon. Based on historical phonetic correspondences, I argued that *bằm* has the etymological meaning of 'to suck, to breastfeed' and therefore, it was developed to designate a 'breastfeeding woman.' I also proposed that the process was likely the same with other terms referring to 'mother': *bu*, *vú*, *u*, *đẽ*. In another study of mine (Tran Thi Hong Hanh 2023), I also examined other kinship terms and gave my explanation of the source of *con* (child, offspring), *cháu* (grandchild), *chắt* (great-grandchild), *chít* (great-great-grandchild), *chít* (great-great-great-grandchild), some terms for 'mother' such as *me*, *mợ*, *mụ*, and some female kinship terms such as (bà) *vãi* (maternal grandmother), *ả* (sister), and *già* (aunt/ grandmother). However, my analysis differed from that of Alves in several respects. I was inclined to suggest that the source of *mẹ* (and *mạ*) was Austroasiatic, whereas *má* seemed to originate from Chinese dialects.

In reviewing the literature, I discovered that although a general consensus seems to exist regarding the source of certain kinship terms in Vietnamese, there are still differences of opinion and some weak arguments concerning this, which need further study. Thus, this paper continues the discussion of the origin of the kinship terms in Vietnamese, focusing on the origin analysis of some terms for 'mother.' By providing additional evidence on the geographical distribution of these terms, this paper is expected to contribute to identifying and clarifying their origin.

## 2. Data and discussion

There are various terms to refer to ‘mother’ among Vietnamese dialects. These terms can be divided into two groups: terms that are all open syllables starting with the first consonant /m/: *mẹ, me, mạ, mợ, mệ, mẹ, má,* and terms that do not have /m/ as first consonant: *bà, bu, u, vú, đẽ, cái, nạ.* Some of these terms specifically and primarily mean “mother”: *mẹ, me, mạ, má, bà, u,* while others have multiple meanings that refer to female relatives who are not necessarily mother. These include words such as *mụ, mệ,* and *mợ* or words that are used as nouns to refer to female breasts, verbs meaning ‘to give birth,’ or adjectives meaning ‘female,’ ‘king-sized,’ or ‘main.’ These data were collected from various sources such as dictionaries of idioms and proverbs and dialect dictionaries. Moreover, in order to draw the linguistic map, we conducted surveys and interviews with informants in different localities using questionnaires. The data related to the reconstructions of Mon-Khmer languages are all from the Mon-Khmer Etymological Dictionary (MKED) and reconstructions of Tai languages are cited from Li Fang Kuei (1977) and/or Pittayaporn (2009). Among the kinship terms for ‘mother,’ the terms *bà, bu, u, vú,* and *đẽ* were discussed in my previous studies (Trần Thị Hồng Hạnh 2021, 2023). Thus, this paper will focus on other terms.

### 2.1. The case of *mẹ, mệ, and mạ*

According to the MKED, Proto-Mon Khmer has four reconstructed forms of *mẹ*: *\*mee?, \*ma?, \*ʔmee?* and the reconstructed form *\*[ʔ]bo?*, which changed to *vợ* ‘wife’ in modern Vietnamese (Shorto 2006). To enable a comparison among the other branches of Mon-Khmer languages, the following table presents the reconstructed results for ‘mother’:

Table 1 The reconstructed forms of ‘mother’ in other Proto-Mon Khmer branches and in Mon Khmer languages (Source: MKED).

Proto West-Bahnaric	<i>*mee?</i>
Proto South-Bahnaric	<i>*me:</i>
Proto Central-Bahnaric	<i>*me:ʔ</i>
Proto Bahnaric	<i>*me:iʔ</i>
Proto Katuic	<i>*ʔamee?</i>
Proto Khmuic	<i>*maʔ</i>
Proto Vietic	<i>*me:ʔ ~ me:ʔ</i>
Khmu	<i>maʔ</i>
Bahnaric	<i>maa, meʔ</i>
Mnong	<i>ma:j</i>

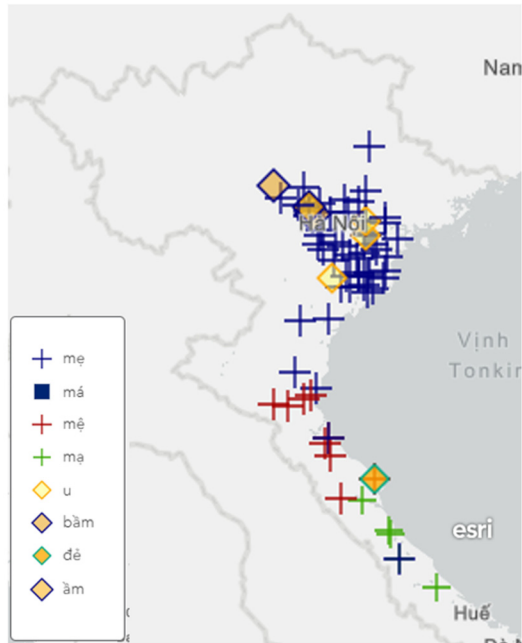
Stieng	<i>mej; me:i</i>
Old Khmer	<i>me, 'me, ame</i>
Khmer	<i>mae</i>
Chứt/Rục/Sách	<i>m<sup>ə</sup>ɛ̃</i>
Maleng	<i>mê:ʔ</i>
Pong	<i>me:</i>
Mường	<i>me:<sup>46</sup></i>

Regarding the similarity of these word forms in Vietnamese and Tai languages, in Proto Tai, the reconstructed form of mother is *\*me:<sup>B</sup>*, and in some other Tai dialects, *me:<sup>B2</sup>*, *me:<sup>B2</sup>*, and *me<sup>B2</sup>* (Pittayaporn 2009). This is probably the why Alves (2017) claimed that there was not enough evidence to confirm the origin of the term *mẹ*. Other scholars such as Ly Tung Hieu (2015) even asserted that both *mẹ* and *bố* originated from a Tai language ('Tày' language as he wrote).

However, the term *mẹ* together with two other terms *mệ* all have phonetic forms that are more related to those reconstructed forms mentioned in Table 1. Thus, if *mẹ* was identified as a Tai loanword, this implies the possibility that not only Vietic languages but also other Mon-Khmer languages borrowed the term 'mother' *mẹ/mệ* from Tai languages. These conditions are unlikely to have occurred. Moreover, it is worth noting that in the Mường language, the term to refer to 'mother' is *mế*.<sup>1</sup> With regard to the tonal and vowel correspondences, the reciprocal between /e/ in *mệ* and /ɛ/ in *mẹ*, between 'sắc' tone in *mế* and 'nặng' tone in *mệ* further support the idea that *mệ* and *mế* as well as *mệ* and *mẹ* originated from the same source and that they can all be identified as Austroasiatic terms.

Meanwhile, the term *mạ* has a clear phonetic relationship with the reconstructed form *\*maʔ* in Proto-Mon Khmer. Moreover, in terms of geographical distribution, *mạ* (as well as *mệ*) is still used in Central dialects (Map 1). As analyzed and indicated in my previous study (Trần Thị Hồng Hạnh 2023), some kinship terms in Central dialects such as *o* (aunt) and *ả* (sister) can be identified as terms of Proto-Austroasiatic origination based on historical evidence of phonetic correspondence (Table 2).

<sup>1</sup> I use this spelling according to the spelling in *Đẻ đất đẻ nước* (The Muong Epics of 'The Birth of the Earth and Water') (Đặng Văn Lung 1988). This is also the spelling whose use is quite popular in newspapers. In the Mường - Việt Dictionary (Nguyễn Văn Khang 2002), the spelling is *mế*.



Map 1 Terms for ‘mother’ in Northern and Central dialects.

Table 2 Correspondences between *o*, *â*, and Proto-Austroasiatic/Austroasiatic forms.

<i>Terms</i>	<i>Proto-Austroasiatic/Austroasiatic forms related</i>	<i>Meaning</i>
<i>o</i> ‘father’s sister’	Proto Mon-Khmer: *ʔəh	younger sibling
	Proto Mon-Khmer: *ʔ[o]ʔ	elder sibling
	Proto Vietic: *ʔə: Chứt, Malieng, Tho: ʔə: <sup>1</sup>	father’s younger sister
<i>â</i> ‘sister’	Chứt (Arem, Sách, Rục): ʔa:	father’s elder brother’s wife, father or mother’s elder sister

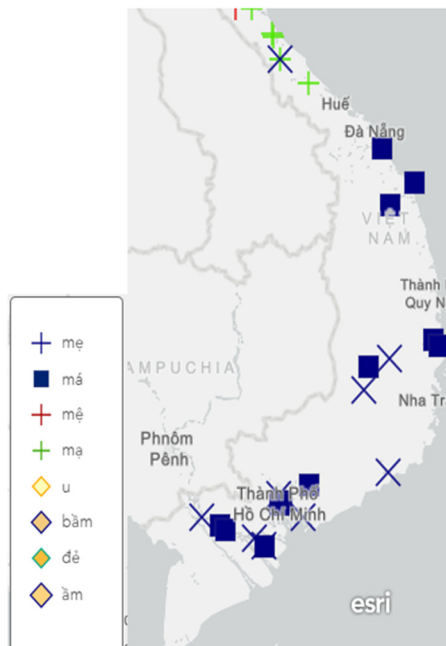
Based on the above arguments, versus the possibility that *mệ* is a Tai loanword, I was inclined to suggest that *mệ* as well as *mệ*, and *mạ* are terms of Austroasiatic origin.

## 2.2. The case of *má*

The source of *mệ* was left open by Alves, but he did note that there were comparable forms in both Austroasiatic and Tai-Kadai. Meanwhile, the source of *má* is also difficult to assume because comparable forms could be found in both Austroasiatic and Chinese dialects.

Regarding the possibility that *má* is Austroasiatic in origin, an issue that emerges from the correspondence between *má* and *mạ* is whether it can be considered to be essentially the same as the correspondence between *mệ* and *mẹ*. In other words, a possible explanation for this correspondence is that *má* and *mạ* were also originally related. In this case, *mạ* is likely to be more ancient than *má* because it is more likely that *\*maʔ* in Proto-Mon Khmer changed into *mạ* rather than *má*, and, as noted above, *mạ* is used in Central dialects that are mainly agreed to be dialects that have preserved many of the original elements of Vietnamese language. It can therefore be assumed that *má* was a variation of *mạ* as the term spread southward.

Nevertheless, the distribution of this term seems to provide further support for the hypothesis that *má* is the result of the later Chinese–Vietnamese language contact. Southern Vietnam has a number of speakers of Chinese dialects such as Cantonese, Teochew, Hakka, Hokkien, and Hainanese who have been migrating to this region since the 17th century. In Cantonese, mother 妈妈 is pronounced as /ma<sup>21</sup> ma<sup>35</sup>/. In Teochew dialect, mother is 阿嫲 /a<sup>33</sup> ma<sup>35</sup>/ and father is 阿爹 /a<sup>33</sup> tia<sup>33</sup>/. Thus, this vowel and tonal correspondence indicates the possibility that *tía*, *má* in Southern Vietnamese dialects originated from these dialects. In terms of geography, *má* in the pair of *ba má*, *tía má* (father and mother) only appear in the Southern dialect. It can therefore be assumed that the term *má* was the term that appeared along with the wave of Chinese immigrants to Vietnam.



Map 2 Geographical distribution of *má* in Southern dialects.

Another piece of evidence that can further support this explanation is the fact that in the dictionary *Đại Nam quốc âm tự vị* (*Dictionnaire Annamite* 大南國音字彙) published in 1895, the entry for ‘má’ is explained: ‘*n. Hai miếng thịt ở hai bên mặt; mẹ, (kêu theo tiếng Khách)*’ (Nôm word. two pieces of meat on either side of the face; mother (called by Hakka language)).

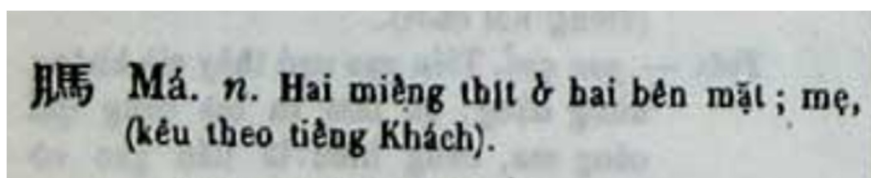


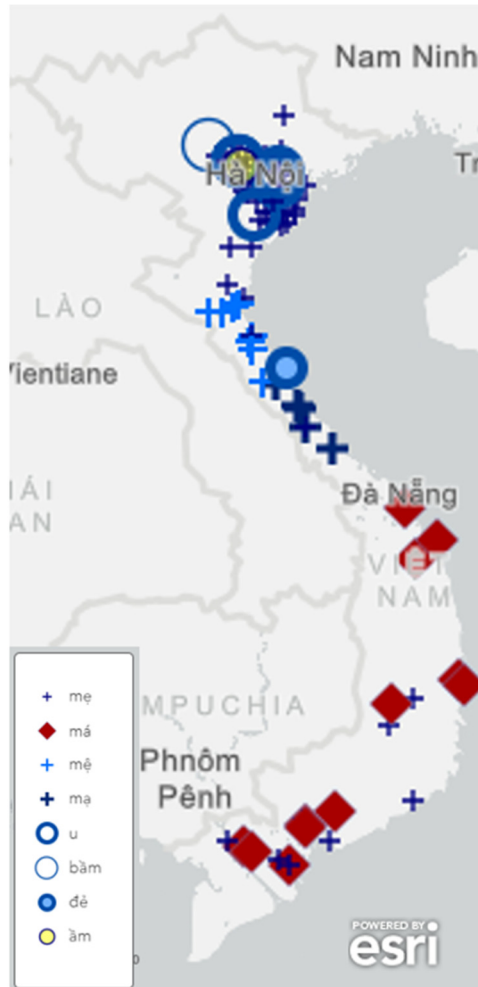
Fig. 1 The entry for ‘má’ in *Đại Nam quốc âm tự vị*.

A. de Rhodes’ dictionary (1651) and Taberd’s *Dictionarium Anamitico Latinum* (1838) contain no entry for *má* with the meaning ‘mother’. The term referring to ‘mother’ in the first literary work written in *Chữ quốc ngữ* (national romanized Vietnamese script), *Thầy Lazaro Phiền* (The story of Lazaro Phiền) (Nguyễn Trọng Quản 1887), is *mẹ*. It was not until Ho Bieu Chanh’s literary works in the 1920s that the term *má* seemed to have just begun to appear. Therefore, the term *má* can be assumed to have appeared very late, around the end of the 19th century or the beginning of the 20th century.

However, Map 3 has important implications concerning which hypothesis on the diffusion of the term *má* appears to be stronger. If *má* had changed from *mạ*, the more ancient term preserved in the Central dialects, *má* (together with *mẹ*) must have spread to the South due to the Southern wave of Vietic residents moving southward around the beginning of the 17th century. This time period coincided with the first wave of the Hoa people<sup>2</sup> migrating to Vietnam. If so, there is no explanation for the fact that *má* did not diffuse to the North but only to the South and that *má* was first found only in documents at the end of the 19th century as analyzed above. This finding, although preliminary, suggests the possibility that *má* came from Chinese dialects.

<sup>2</sup> The Hoa people are the name Vietnamese people call Chinese groups like Cantonese, Teochew, Hakka, Hokkien, Hainanese, etc. who migrated to Vietnam at different times from the 16th century and later in the late Ming and early Qing dynasties, which seems to last until the first half of the 20th century.





Map 3 Geographical distribution of *má* in Southern dialects.

### 3. Conclusion

Based on an analysis of both historical phonetic evidence and the geographical distribution of the terms *má*, *mạ*, *mệ*, *mẹ*, this paper indicated that *mẹ*, *mệ*, *mạ* can be argued to be Austroasiatic etymons (*mẹ*, *mạ*, *mệ*), while cases of *má* are likely to be loanwords from Chinese dialects. These findings may help provide us with more understanding of the Vietnamese kinship terms, which are not only rich in number but also diverse in origin. However, further studies of Vietnamese kinship terms are called for that can provide more evidence of the proposed hypotheses of how those terms originated while also taking other terms into account.

## References

- Alves, Mark (2017) Chinese Loanwords in Vietnamese Pronouns and Terms of Address and Reference. In: Lan Zhang (ed.) *Proceedings of the 29th North American Conference on Chinese Linguistics* (NACCL-29), Volume 1, 286–303. University of Memphis, Memphis, TN.
- Alves, Mark (2020) Historical ethnolinguistics notes on Proto Austroasiatic and Proto Vietic vocabulary in Vietnamese, *Journal of Southeast Asian Linguistics Society* 13.2 (2020): xiii–xlv. URI: <http://hdl.handle.net/10524/52472>
- Đặng, Văn Lung (1988) *Đẻ đất đẻ nước - Sử thi Mường* [The Muong Epics of ‘The Birth of the Earth and Water’], Nhà xuất bản Khoa học xã hội, Hà Nội.
- Hồ Biểu Chánh (2014) *Tại tôi* (the novel ‘Because of me’), Nhà xuất bản Văn hóa văn nghệ, Thành phố Hồ Chí Minh.
- Hồ Biểu Chánh (2020) *Con nhà nghèo* (the novel ‘Children in poor families’), Nhà xuất bản Văn hóa văn nghệ, Thành phố Hồ Chí Minh.
- Huỳnh, Tịnh Paulus Của (1895) *Đại Nam quốc âm tự vị (Dictionnaire Annamite 大南國音字彙)*, Khai Trí, Sài Gòn.
- Li, Fang Kuei (1977) *A Handbook of Comparative Tai*, Oceanic Linguistics Special Publications, no. 15 (1977): i–389. <http://www.jstor.org/stable/20006684>
- Lý, Tùng Hiếu (2015) *Từ xưng hô tiếng Việt - Tiếp cận từ góc nhìn dân tộc - ngôn ngữ học*, *Tạp chí Ngôn ngữ* [Vietnamese address terms – An approach from the perspective of ethnolinguistics, *Journal of Linguistics*], Số 6: 30–42.
- Nguyễn Trọng Quản (1887) *Thầy Lazaro Phiền* (The story of Lazaro Phiền), Rey et Curriot, Sài Gòn.
- Nguyễn, Văn Khang (eds.) (2002), *Từ điển Mường Việt* [The Dictionary of Mường - Việt], Nhà xuất bản Văn hóa dân tộc, Hà Nội.
- Pittayaporn, Pittayawat (2009) *The Phonology of Proto-Tai* (Doctoral dissertation), Department of Linguistics, Cornell University. URI: <https://hdl.handle.net/1813/13855>
- Rhodes, Alexandre de (1651) *Từ điển Annam – Lusitan – Latinh (Từ điển Việt - Bồ - La)* [The Dictionarium Annamiticum Lusitanum et Latinum (The Vietnamese - Portuguese - Latin dictionary)], Nhà xuất bản Khoa học xã hội, Hà Nội.
- Taberd, Jean-Louis & Pigneau de Behaine (1838) *Dictionarium Anamitico-Latinum*. J.Marshnam, Serampore, India.
- Trần, Thị Hồng Hạnh (2021) *Về hai từ bú, bằm trong thổ ngữ Phú Thọ, Kỳ yếu Hội thảo khoa học quốc tế “Nguyễn Tài Căn - Tư tưởng, tác phẩm và kỷ niệm”* [About two words bú and bằm in Phu Tho dialect, Proceedings of the International Scientific Conference “Nguyen Tai Can - Thoughts, works and memories”], Nhà xuất bản Đại học Quốc gia Hà Nội, Hà Nội, 133–144.
- Trần Thị Hồng Hạnh (2023) A contribution to the discussion on kinship terms in Vietnamese, In: Alves, Mark J., Lâm, Quang Đông, Trịnh, Cẩm Lan, Trần, Thị Hồng Hạnh, & Dương, Xuân Quang (eds.) *Researching and Applying Linguistics and Vietnamese Language Studies*, 270–289. Tokyo: Geolinguistic Society of Japan. <https://doi.org/10.5281/zenodo.8153483>
- The Mon-Khmer Etymological Dictionary (The MKED): <http://www.sealang.net/monkhmer/dictionary/>

## Publication history

Date received: 1 September 2023

Research article (Featured theme: Geolinguistic approach to Sino-Tibetan)

## Geolinguistics and Sino-Tibetan: An introduction

SHIRAI, Satoko  
The University of Tokyo

**Abstract:** Geolinguistics is expected to contribute to the diachronic study of the Sino-Tibetan language family, which typologically shows significant diversity across regions. This article introduces concepts, methods, and recent results of geolinguistic study. We also survey the differences between geolinguistics and the adjacent fields of study, including areal linguistics.\*

**Keywords:** geolinguistics; language geography; areal linguistics; linguistic map; Sino-Tibetan

### 1. Overview of the geolinguistic approach to Sino-Tibetan

The Sino-Tibetan languages show remarkable diversity and regional distribution of their typological characteristics (cf. the concept of spheres in Matisoff 1990, 1991, 2003, 2021). In this feature, “Geolinguistic approach to Sino-Tibetan,” we present the method of geolinguistics as an approach to the diversity of Sino-Tibetan languages and provide practical examples of research.

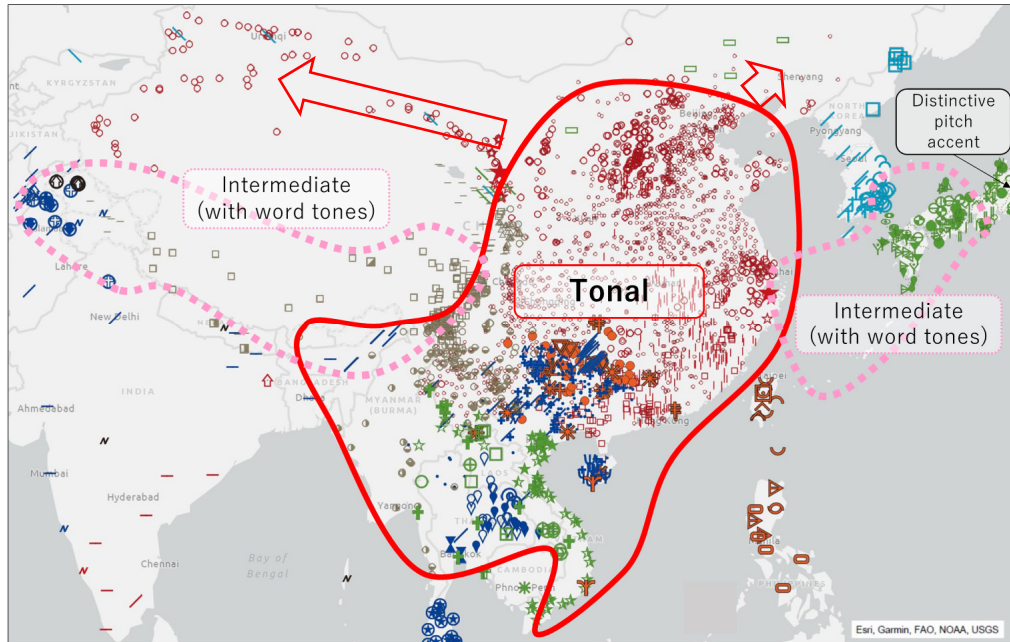
Figure 1 presents an example from Endo et al. eds. (2021: Map 7), a linguistic map produced by the geolinguistic method, to which I have added borderlines and annotations. The red circle corresponds to the boundary between the inner and outer groups defined by Iwata (2021). This figure visually illustrates a hypothesis of tone diffusion from southern China; to the east and west of the syllabic tone language region lie word tone regions, and further out are the distinctive pitch-accented regions (cf.

---

SHIRAI, Satoko. 2023. Geolinguistics and Sino-Tibetan: An introduction. *Studies in Geolinguistics* 3: 146–151. doi: <https://doi.org/10.5281/zenodo.8437101>

\* This study was presented at the pre-conference workshop of the 55<sup>th</sup> International Conference on Sino-Tibetan Languages and Linguistics (Kyoto University, 2022), entitled “Geolinguistic approach to Sino-Tibetan.” I would like to express my sincere gratitude to Prof. Mitsuaki Endo for his tremendous comments during the preparations for that workshop. However, all errors in this paper are attributed to the author. This work was supported by JSPS Grants-in-Aid for Scientific Research (Kakenhi) JP18H03577 and JP19K00543.

Hayata 1999, Matisoff 2001). The Sinitic languages fall within the tonal area, whereas the Tibeto-Burman languages are found in all the tonal, intermediate (with word tones), and atonal (outside the circles) areas. This geographical distribution suggests that, from a macro perspective, the tone systems among the Tibeto-Burman languages are innovations.



**Sinitic**

**The number of phonological tones**

- ★ 1
- ☆ 2
- 3
- 4
- | 5
- 6

**Tibeto-Burman**

- |            |               |
|------------|---------------|
| □ WT+NP    | ⊞ WT+SA+NP    |
| ○ ST+NP    | ▣ WT+YT       |
| ◐ ST+YL    | ⊞ NT/NA+ST+YL |
| — NT/NA+NP | ⊞ PA+ST+YT    |
| △ RG+YC    | ⊞ ST+SA+NP    |
| ⊞ ST+YT    | ⊞ ST+SA+YL    |
| ⊞ NT/NA+YC | ⊞ WT+RG+YC    |
| ⊞ PA+NP    |               |

Fig. 1: Tone and accent in Sino-Tibetan and adjacent languages (Excerpted from Endo et al. 2021: Map 7. Boundaries, arrows, and annotations in the map were added by the author for illustrative purposes.)<sup>1</sup>

<sup>1</sup> Abbreviations in the Tibeto-Burman Map Legend: NP: no phonation pattern, NT/NA: no tone or accent, PA: pitch accent, RG: register, SA: stress accent, ST: syllabic tone, WT: word tone, YC: phonation influenced by initial consonants, YL: distinctive phonation, YT: phonetic phonation.

We consider that a geolinguistic approach could be effective in studying the Sino-Tibetan language family. This introductory paper is followed by three related specific studies: Ebihara and Saitô (this volume) discuss the interactions between Tibetic and Mongolic languages belonging to different language families. Then, H. Suzuki, Yagi, and F. Suzuki (this volume) examine the distribution of word forms within the Sino-Tibetan language family, using examples from animal and plant terms. Finally, Iwasa (this volume) provides a unique study of Yi characters using diverse regional manuscripts, which are mutually almost incomprehensible. Each study was conducted from a different perspective, but they all demonstrate the validity of a geolinguistic approach.

## **2. Basic concepts**

### **2.1. Philosophies of linguistic geography**

Geolinguistic methods are based on those of linguistic geography (or dialect geography). This section outlines the basic concepts of linguistic geography, mainly based on Sibata (1969). This paper distinguishes geolinguistics and linguistic geography, and their differences are explained in Section 2.3.

“Every word has a history.” This epigram is attributed to Jules Gilliéron, a Swiss-French dialectologist of the 19th and early 20th centuries. Sibata (1969: 16–19) cited these words and added, “Language systems do not change at once. When dealing with individual and personal phenomena, even members of the same community do not necessarily have the same language. However, when dealing with the geographic distribution, we can find orderliness in it because it is a projection of history.” He clearly identified the purpose of geolinguistics as uncovering the history of a language (Sibata 1969: 11–12). These are also the basic assumptions of our work. More recently, Onishi (2016: 185) claimed that geolinguistic study should not be limited to historical facts, as it can reveal various aspects of dialects that can be studied in relation to geographic distribution. This can be understood as an expansion of linguistic geography.

### **2.2. Basic method of geolinguistics**

The basic procedures of geolinguistic studies are common to linguistic geographical studies as outlined by Sibata (1969: 11), as cited below:

1. Select a certain linguistic part to be examined.
2. Identify the regional variants of the linguistic form that denote it.
3. Map the regional variants.
4. Identify the geographic distribution on the map.

5. Focusing on that distribution, along with other clues, estimate diachronic change and its factors.

Sibata (1969) places particular emphasis on the fourth of these steps: finding the geographic distribution: “Simply filling in a map with regional variants does not automatically produce their geographic distribution” (Sibata 1969: 12). Therefore, it is necessary to find the classification significantly distributed for historical composition through the examination of various classifications of the target linguistic phenomena (ibid.).

### **2.3. Difference among adjacent research areas**

Geolinguistics adopts the abovementioned methods to identify the history of multiple languages using a certain linguistic feature, mainly in relation to its geographical distribution (Endo 2015). Here, we survey the differences between geolinguistics and the adjacent fields of study: linguistic geography, areal linguistics, and typological feature maps. Sibata (1969: 11–16) also made comparison with other fields, including comparative linguistics.

Linguistic geography, also referred to as dialect geography, shares common methods with geolinguistics, but its targets are slightly different: it aims to examine the history of a language (or languages in a country) based on the geographical distribution of word forms of vocabulary items. An example of this in the earliest stages is Gilliéron (1918), and Iwata ed. (2009) is an example of recent studies.

The main purpose of areal linguistics is to investigate linguistic features shared across a certain linguistic area. The historical process of their formation may be within the range of study. In this case, it seems like a geolinguistic study, but areal linguistics does not necessarily discuss detailed geographic distribution. Examples of this include Aikhenvald and Dixon eds. (2001) and Hickey ed. (2017).

Typological feature maps like those found in the *World Atlas of Language Structure* (Dryer and Haspelmath eds. 2013) only map certain typological features. Finding distributions or examining the process of formation of a distribution does not fall within the scope of their research. Hence, the research attitudes of geolinguistics and typological feature maps are very different.

## **3. Results of recent geolinguistic projects in Japan**

Our articles form part of the result of joint research projects at the Research Institute for Languages and Cultures of Asia and Africa (ILCAA), “Studies in Asian Geolinguistics” and “Studies in Asian and African Geolinguistics,” both headed by

Mitsuaki Endo. These projects were completed in March 2023 and have resulted in a number of publications. *Studies in Asian Geolinguistics* series (eight volumes), *Studies in Asian and African Geolinguistics* (two volumes), and a monograph series were published online by ILCAA and are available at <https://publication.aa-ken.jp/>. In addition, the *Linguistic Atlas of Asia* (Endo et al. eds. 2021) was published by Hituzi Syobo.

The Geolinguistic Society of Japan (GSJ) was founded in April 2019, and it continues to publish the journal *Studies in Geolinguistics* and Monograph series, including the *Linguistic Atlas of Asia and Africa* series. All publications of the GSJ are available online.

#### 4. Summary

Geolinguistics not only maps linguistic features but also examines their distribution and history. Geolinguistics can synthesize a variety of clues to identify a history that is not visible to comparative linguistics cannot. Although comparative linguistics is an extremely sophisticated research method, the history of a language can be shown by using the findings of geolinguistics in a complementary way as well. Geolinguistics is expected to play a significant role in the diachronic study of Sino-Tibetan.

#### References

- Aikhenvald, Alexandra Y. and R. M. W. Dixon (eds.) (2001) *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press.
- Dryer, Matthew S. and Martin Haspelmath (eds.) (2013) *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info>
- Ebihara, Shiho and Yoshio Saitô (this volume) Language contact between Tibetic and Mongolic, with a special focus on the Amdo Linguistic Area. <https://doi.org/10.5281/zenodo.8437105>
- Endo, Mitsuaki [遠藤光暁] (2015) Gengochizu 「言語地図」 [linguistic map] In: Yoshio Saito et al. (eds.) 『明解言語学辞典』 [*The Sanseido dictionary of linguistics*], 70–71. Tokyo: Sanseido.
- Endo, Mitsuaki, Makoto Minegishi, Satoko Shirai, Hiroyuki Suzuki, and Keita Kurabe (eds.) (2021) *Linguistic Atlas of Asia*. Tokyo: Hituzi Syobo.
- Gillieron, Jules (1918) *Généalogie des mots qui désignent l'abeille: d'après l'Atlas linguistique de la France*. Paris: Librairie ancienne Honoré Champion, éditeur; Édouard Champion.
- Hayata, Teruhiro (1999) *Onchoo-no Taiporojii* 『音調のタイポロジー』 [*The Typology of Tonal Phenomena*]. Tokyo: Taishuukan.
- Hickey, Raymond (ed.) (2017) *The Cambridge handbook of areal linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107279872>

- Iwasa, Kazue (this volume) Current issues and perspectives on Yi characters: From the past and into the future. <https://doi.org/10.5281/zenodo.8437127>
- Iwata, Ray (ed.) [岩田礼 編] (2009) 汉语方言解释地图 / Interpretative Maps of Chinese Dialects. 東京：白帝社. [Also available online since 2019: Chinese Dialect Geography (online version). <http://chinesedialectgeography.jp/>]
- Iwata, Ray (2021) Tone and accent in Asia. In: Mitsuaki Endo et al. (eds.) *Linguistic Atlas of Asia*, 232–236. Tokyo: Hituzi Syobo.
- Matisoff, James A. (1990) On megalocomparison. *Language* 66(1): 106–120. <https://doi.org/10.2307/415281>
- Matisoff, James A. (1991) Sino-Tibetan linguistics: Present state and future prospects. *Annual Review of Anthropology* 20: 469–504. <https://www.jstor.org/stable/2155809>
- Matisoff, James A. (2001) Genetic versus contact relationship: Prosodic diffusibility in South-East Asian languages. In: A. Y. Aikhenvald and R. M. W. Dixon (eds.) *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 291–327. Oxford: Oxford University Press.
- Matisoff, James A. (2003) *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley: University of California Press.
- Onishi, Takuichiro [大西拓一郎] (2016) *Kotoba no chirigaku: hoogen wa naze soko ni aru no ka* 『ことばの地理学 方言はなぜそこにあるのか』 [Geography of languages: Why is each dialect there?] Tokyo: Taishukan.
- Sibata, Takesi [柴田武] (1969) *Gengochirigaku no hoohoo* 『言語地理学の方法』 [Methods in linguistic geography]. Tokyo: Chikuma Shobo. (Translated into Chinese by 崔蒙, 《语言地理学方法》北京：商务印书馆 2018)
- Suzuki, Hiroyuki, Kenji Yagi, and Fumiki Suzuki (this volume) Geolinguistic approach to Sino-Tibetan: Lexical relationship in some animal and plant terms. <https://doi.org/10.5281/zenodo.8437113>

## Publication history

Date received: 31 August 2023



Research article (Featured theme: Geolinguistic approach to Sino-Tibetan)

## Language contact between Tibetic and Mongolic, with a special focus on the Amdo Linguistic Area

EBIHARA, Shiho

Tokyo University of Foreign Studies

SAITÔ, Yoshio

Takushoku University

**Abstract:** Tibetic and Mongolic languages have been influenced by each other on various levels throughout history. Especially in the Amdo linguistic area, since the migration of the Oirat Mongolian nomads around the 17th century, there has been close cultural contact. Linguistically, both language groups have experienced language contact in both directions. In this paper, we focus on loanwords between Tibetic and Mongolic languages by using a geolinguistic approach. In addition, the Amdo area showcases a distinct preference for words tied to pastoral culture and weaponry. By comparing the general tendency of borrowing in each language group, the characteristics of cultural contact between the Tibetan and Mongolian tribes in the Amdo linguistic area are discussed.\*

**Keywords:** loanwords; Tibetic; Mongolic; Amdo linguistic area

### 1. Introduction

This study examines language contact between Tibetic (belonging to Tibeto-Burman) and Mongolic by using a geolinguistic approach. Although the language contact between these two language groups at the macro level dates back to the ancient era, this paper focuses on micro-level contact between Tibetic and Mongolic in the Amdo linguistic area (the ‘Amdo Sprachbund’ in Slater 2019). Although language contact in grammatical forms and functions has recently attracted attention in the Amdo linguistic area (Sandman and Simon 2016 etc.), we limit the focus of this discussion only to loanwords between Tibetic and Mongolic.

---

EBIHARA, Shiho and Yoshio SAITÔ. 2023. Language contact between Tibetic and Mongolic, with a special focus on the Amdo Linguistic Area. *Studies in Geolinguistics* 3: 152–163. doi: <https://doi.org/10.5281/zenodo.8437105>

\* This work was funded by JSPS KAKENHI 18H05219, 15H03203, 20J40127, and 19K00567.

The Amdo linguistic area lies on the northeastern edge of the Tibetan plateau. In the modern administrative division, it is located in the eastern Qinghai and southern Gansu provinces. Historical causes, including political and military control, have formed the present situation of language contact among four language groups: Tibetic, Mongolic, Turkic, and Sinitic.

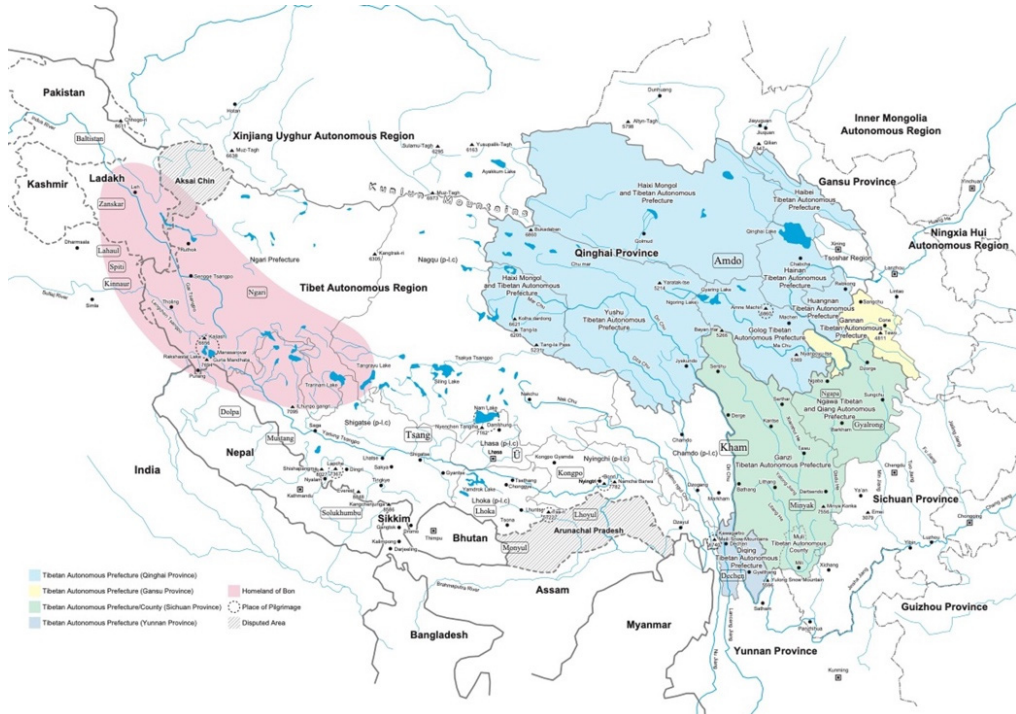


Fig. 1 Amdo Linguistic Area (Iwao & Ikeda 2021)

In the following, we examine loanwords borrowed from Tibetic to Mongolic (§2) and from Mongolic to Tibetic (§3) in turn. In each section, we first discuss the situation in Mongolian- or Tibetan-speaking areas outside Amdo. Next, we discuss the situation inside the area, and then we make a more general comparison, which will identify the characteristics of borrowing in the Amdo area.

## 2. Tibetic loanwords in Mongolic languages

Because of the long historical and cultural relationship with Tibetans, Mongols have adopted many words from Tibetan.

### 2.1. Tibetic loanwords in Mongolic languages in general

Many Tibetic words in Mongolic, either of Tibetic origin or borrowed through Tibetic, are related to Buddhism, the names of planets, days of the week, mineral substances, plants, and diseases. Examples include the following:

Written Mongolian (WT)		Written Tibetan (WT)
<i>lama</i>	‘monk, priest’	<i>bla ma</i>
<i>rabsal</i>	‘textbook of minor rites’	<i>rab gsal</i>
<i>jada</i>	‘rainmaking by use of magic spell’	<i>gyad</i>
<i>bumba</i>	‘vessel for holy objects’	<i>bum pa</i>
<i>namtar</i>	‘biography, legend’	<i>rnam thar</i>
<i>diglim</i>	‘order, system’	<i>sgrig lam</i>
<i>maday</i>	‘mistake, fault, error’	<i>ma dag</i>
<i>saya</i>	‘million’	<i>sa ya</i>
<i>bum</i>	‘hundred thousand’	<i>'bum</i>
<i>nima</i>	‘Sunday’	<i>nyi ma</i>
<i>lhayba</i>	‘mercury; Wednesday’	<i>lhag pa</i>
<i>nomin</i>	‘azure stone’	<i>mu men</i>
<i>dorjipalam</i>	‘diamond’	<i>rdo rje pha lam</i>
<i>jay</i>	‘gonorrhoea’	<i>zag</i>
<i>yariγ</i>	‘paralysis, apoplexy’	<i>ka rigs</i>
<i>badayana</i>	‘cancer of the stomach’	<i>bad kan</i>
<i>sügmil</i>	‘cardamom’	<i>sug rmel</i>
<i>buram</i>	‘cane sugar’	<i>bu ram</i>

Most of these words do not seem to have entered Mongolic through everyday contact between local Tibetan and Mongolian commoners.

### 2.2. Tibetic loanwords in Mongolic languages in Amdo

In the Amdo area, Tibetic and Mongolic people are routinely in contact with one another, and the Mongolic languages spoken in this area have words of Tibetic origin that are not known in other Mongolic languages.

As an example, consider the case of animal vocabulary. Figures 2 to 5 show maps of the geographic distribution of words for ‘wolf’ (Figures 2 and 3) and ‘bear’ (Figures 4 and 5) in Tibetic and Mongolic in general and in Amdo.

The original Mongolic word for ‘wolf’ is *čono*, and this is the term used most widely in Mongolic languages. But, in Amdo, Dongxiang has the word *dzaŋgəi*, and Monguor

uses the form *kadam/kadan*, which are related to Tibetic *hcaŋkhu* and *khadam*, respectively.

The original Mongolic word for ‘bear’ is *bābgai*, but languages in peripheral regions adopt words from their neighboring languages. In the Amdo area, Baoan uses the form *dermoŋ* for ‘bear,’ which is a copy of the Tibetic word *dred mong*.

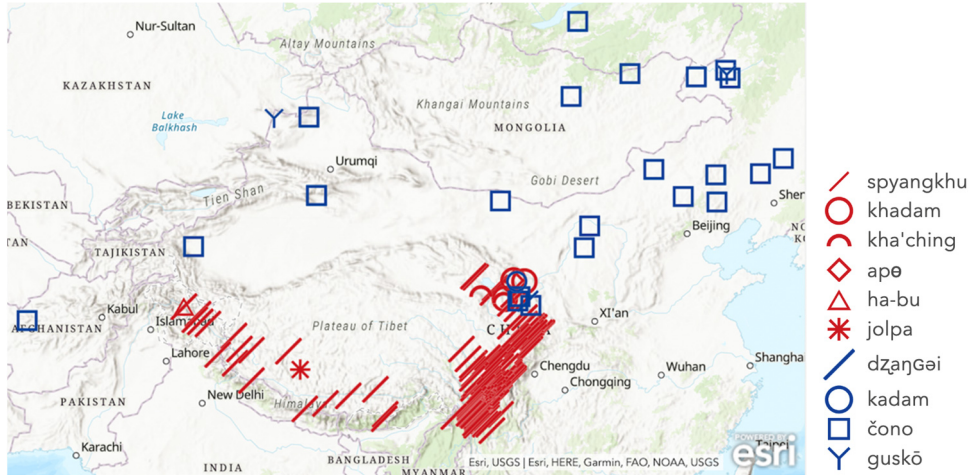


Fig. 2 ‘Wolf’ in Tibetic and Mongolic (in general)

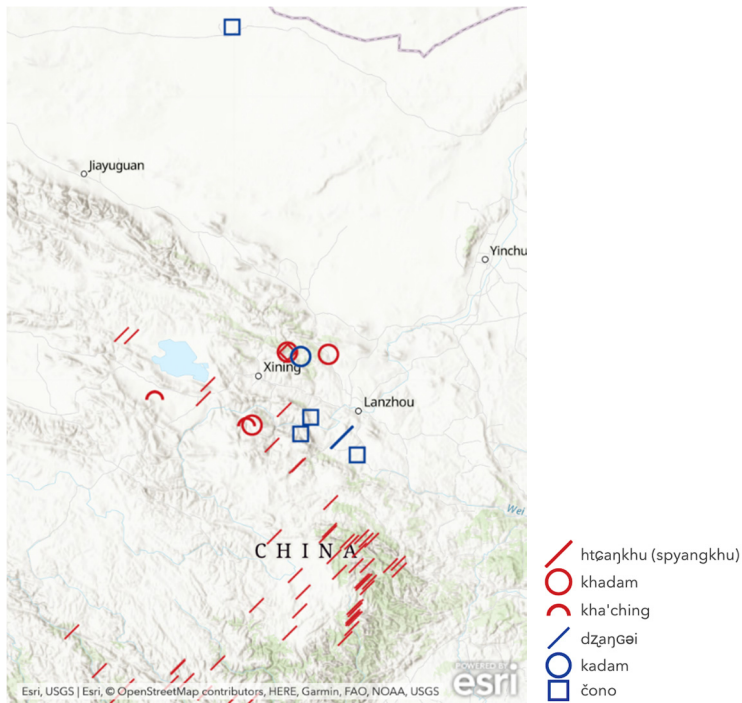


Fig. 3 ‘Wolf’ in Tibetic and Mongolic (in Amdo)

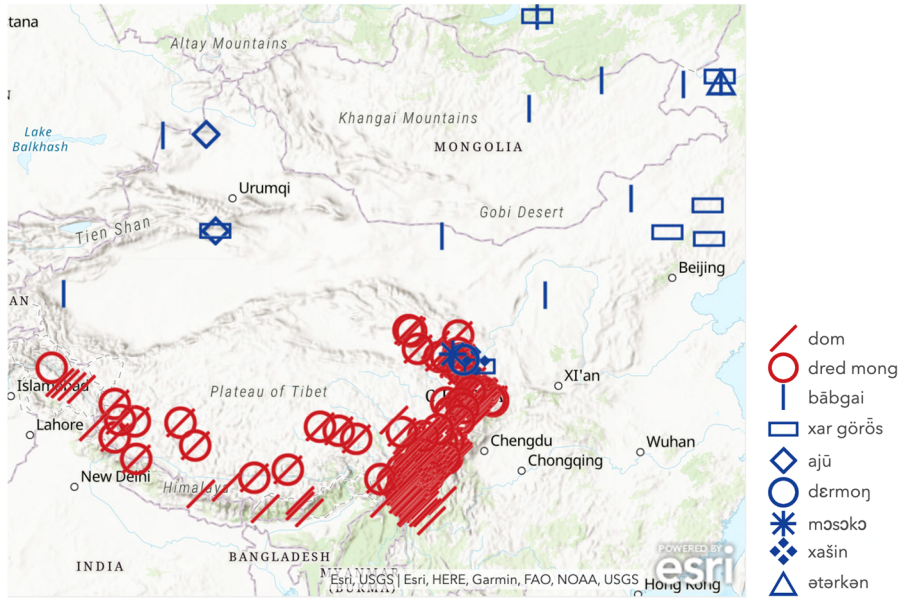


Fig. 4 'Bear' in Tibetic and Mongolic (in general)

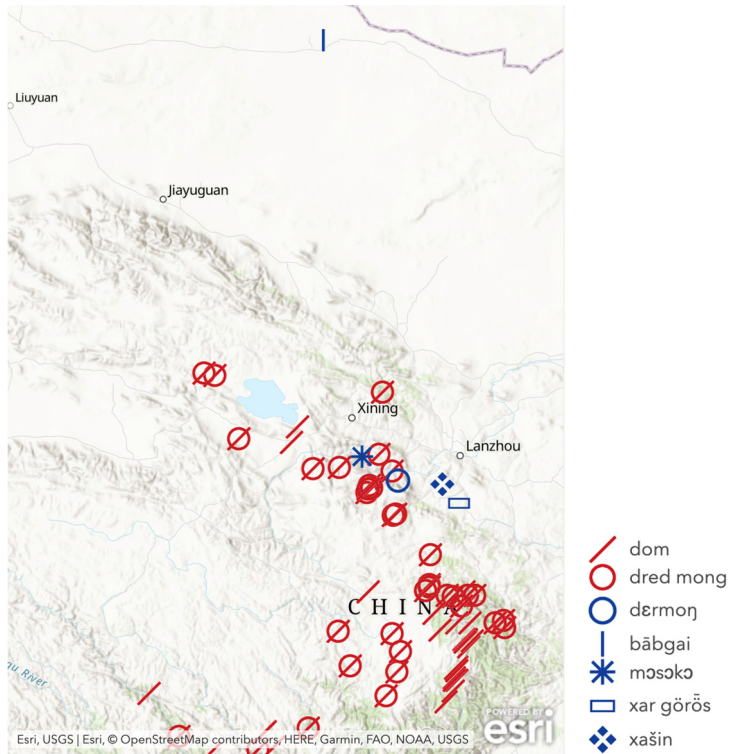


Fig. 5 'Bear' in Tibetic and Mongolic (in Amdo)

### 2.3. Discussion about Tibetic loanwords in Mongolic languages

Unlike religious and scientific terms, which are used by intellectuals and are independent of region, everyday words adopted from local neighboring languages may be found only in specific areas. The abovementioned Dongxiang, Monguor, and Baoan animal-related words of Tibetic origin are not known to the other Mongolic languages and are found only in the Amdo area, where Tibetans and Mongols are in close proximity.

## 3. Mongolic loanwords in Tibetic languages

Next, we examine the case where Tibetans have adopted words from Mongols.

### 3.1. Mongolic loanwords in Tibetic languages in general

Mongolian adopted loanwords in the sphere of religious affairs from Tibetan, whereas Tibetans adopted Mongolian words for secular affairs from the beginning of the Mongol Empire (Laufer 1916; Thubten Jigme Norbu & Takeuchi 1991). Three good examples from these previous studies of categories of Mongolian loanwords in Central Tibet are (1) administrative terms and titles, (2) foreign goods and trade goods, and (3) physicians. Examples include the following:

#### (1) administrative terms and titles

WT		WM
<i>dza sag</i>	‘Mongolian title’	<i>ḡasaḡ</i>
<i>the ci</i>	‘Mongolian title’	<i>taiḡi</i>
<i>taa la’i</i>	‘Tibetan title’	<i>dalai</i>
<i>er te ni</i>	‘treasure’	<i>erdeni</i> (< Skt. <i>ratna</i> )

#### (2) foreign goods and trade goods

WT		WM
<i>thob ci</i>	‘button’	<i>tobči</i>
<i>a cor</i>	‘towel’	<i>alčiḡur</i>
<i>’o mo su</i>	‘stocking’	<i>oyimosu</i>

#### (3) physician

WT		WM
<i>am ci</i>	‘doctor’	<i>emči</i>

### 3.2. Mongolic loanwords in Tibetic languages in the Amdo Area

Thubten Jigme Norbu and Takeuchi (1991) stated that Mongolian loanwords were not established in Amdo Tibetan, but upon close observation some loanwords can be



found, and they have a tendency to be different from loanwords in Central Tibetan. Similar to the loanwords given by Laufer (1916), terms in group (1) ('administrative terms and titles') are also found in modern Amdo Tibetan. However, as noted by Thubten Jigme Norbu and Takeuchi (1991), loanwords referring to (2) ('foreign goods', such as towels, socks, and buttons) and (3) ('physicians') are not found.

Thubten Jigme Norbu and Takeuchi (1991) mentioned that 'while these Mongolian words have been employed and are still in use in the dialect of Lhasa, these words could not have been established in the dialects of Eastern Tibet, such as the Amdo dialect, even though Amdo is the place where Tibetans and Mongolians live side by side.' However, through fieldwork and a literature survey, we have found different kinds of Mongolian loanwords used in Amdo Tibetan, namely (i) place names, (ii) *ger* (a Mongolian mobile dwelling), (iii) camel names and equipment, (iv) horse tack, (v) natural salt, and (vi) guns. Of these, (ii) to (v) can be summarized as words related to pastoral culture.

An example of a term related to pastoral culture belongs to group (iv) ('horse tack') and refers to 'a cushion under a saddle' made by using felt or grass as stuffing. This cushion is called /hom/ (WT: *hom*) in Amdo Tibetan, originally borrowed from the Mongolian word /xom/ (WM: *qom* 'a piece of felt placed under the pack on a camel'). A second example is 'natural salt'. This salt is not for human consumption, but rather to feed livestock, and it is very important for animal growth and maintaining health. This natural salt is called /hordjor/ (WT: *ho rjor*) in Amdo Tibetan and was borrowed from the Mongolian word /xuǰir/ (WM: *quǰir* 'salt marsh; soda').

We can also examine the geographic distribution of group (vi) ('guns'). The word for gun in Amdo Tibetan is pronounced /wu/ (WT: *bo'u*). The origin of this word is /pao/ ('砲') in Chinese, but according to Janhunen and Kalsang Norbu (1999: 274) it was borrowed from the Mongolic term /bū/. This interpretation is supported by the fact that the aspirated /p<sup>h</sup>/ in Chinese does not correspond to the labial-velar approximant /w/ in Amdo Tibetan.

From the perspective of Tibetic languages, seven types of the word for 'gun' were found:

(1) /wu/ type:

Tib. < Mon. < Ch.

(2) /menda/ type: /mentā/ and /ŋēnda/

Compound forms meaning 'fire arrow' (*me mda* (fire arrow) in WT).

(3) /mengko/ type

/me/ corresponds to *me* (fire) in WT.

(4) /trobda/ type

/bda/ corresponds to *mda* ' (arrow) in WT.

(5) /nda/ type

/nda/ corresponds to *mda* (arrow) in WT.

(6) /tubaq/ type

/tubaq/ is of Persian origin and might be borrowed from Turkic (/tüfek/). Tib. < Tur. < Per.

(7) /ts<sup>h</sup>u/ type

/ts<sup>h</sup>u/ is a loan from the Chinese *qiang* (‘槍’).

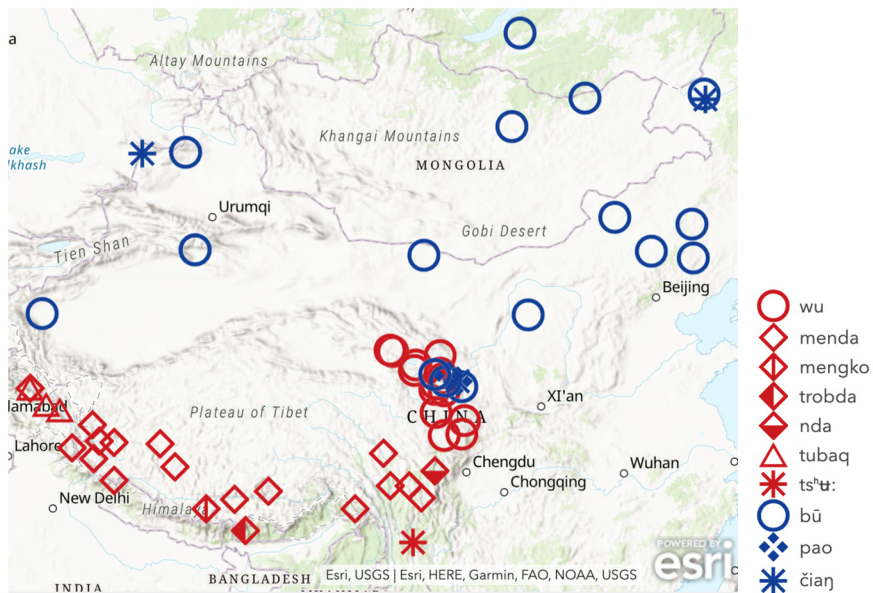


Fig. 6 ‘Gun’ in Tibetic and Mongolic (in general)



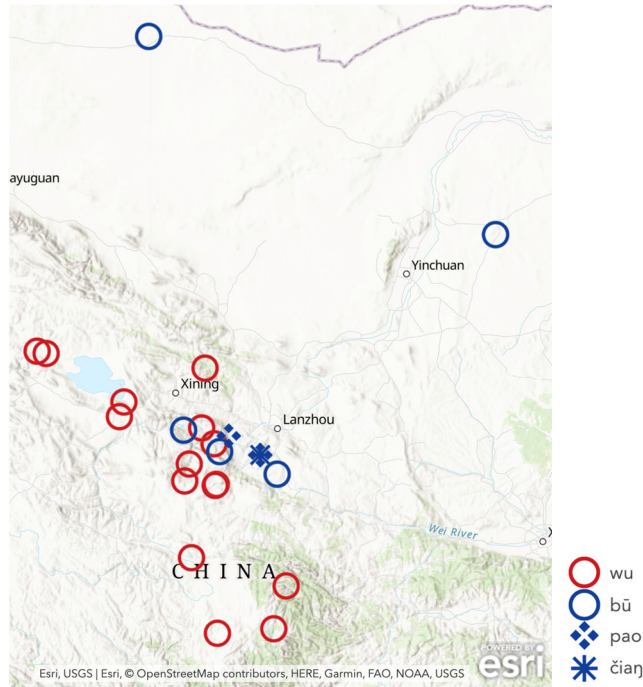


Fig. 7 'Gun' in Tibetic and Mongolic (in Amdo)

### 3.3. Discussion about Mongolic loanwords in Tibetic languages

A comparison of Mongolian loanwords in Central Tibetan and Amdo suggests that there are differences in the way Mongolian and Tibetan speakers interact in each area. In Central Tibet, words related to administration and titles, foreign goods and trade goods, and physicians were borrowed, whereas in Amdo, words related to pastoral culture and guns were found to be distinctive borrowings (Table 1). This suggests that, in Amdo, the Mongolian and Tibetan peoples shared skills related to daily life while living in close proximity to each other.

Table 1 Mongolic loanwords in Tibetic (Central and Amdo Tibetan)

Loanwords	Central Tibetan	Amdo Tibetan
(1) Administrative terms and titles	○	○
(2) Foreign goods, and trade goods	○	—
(3) Physicians	○	—
(4) Pastoral goods	—	○
(5) Gun	—	○

#### 4. Conclusions

This paper has delved into the intricate dynamics of contact between the Tibetic and Mongolic languages, with a specific focus on loanwords within the Amdo linguistic area. Throughout history, the interaction between these two language groups has manifested itself in various ways, leaving traces of cultural exchange and linguistic borrowing. The Amdo area, situated on the northeastern fringes of the Tibetan plateau, has witnessed sustained cultural contact due to historical migrations and geopolitical factors, resulting in a unique linguistic landscape.

The analysis of loanwords highlights the multifaceted nature of this contact. In the case of Tibetic loanwords in Mongolic languages, the borrowing tends to revolve around religious and scientific terminology, reflecting the cultural and intellectual interchange between the two groups. Additionally, everyday words of Tibetic origin find their way into the Mongolic languages spoken within the Amdo area, demonstrating the influence of proximity and routine interaction on vocabulary adoption.

Conversely, an examination of Mongolic loanwords in Tibetic languages, especially in Amdo, unveils a fascinating pattern. Whereas traditional loanwords in Central Tibetan encompass administrative terms, foreign goods, and more, the Amdo area showcases a distinct preference for words tied to pastoral culture and weaponry. This divergence underscores the sociocultural implications of linguistic contact, suggesting that the close coexistence of Tibetan and Mongolian peoples in Amdo has led to the sharing of practical skills relevant to their daily lives.

The geographic distribution of certain loanwords, such as the term for gun, further underscores the intricate web of borrowing. The diversity in borrowing patterns between Central Tibetan and Amdo Tibetan suggests varying modes of interaction and cultural exchange in these areas.

#### Sources

- Baochaolu [保朝鲁] et al. (eds.) (1985) *Dongbu Yuguyu Cihui* 《东部裕固语词汇》 [Vocabulary of the Eastern Yugur Language]. Huhehaote: Neimenggu Daxue Chubanshe.
- Bielmeier, Roland, Katrin Häsler, Chungda Haller, Felix Haller, Veronika Hein & Brigitte Huber (eds.) (2018) *Comparative Dictionary of Tibetan Dialects. Volume 2 Verb*. Berlin: De Gruyter Mouton. doi: <https://doi.org/10.1515/9783110554076>
- Buhe [布和] et al. (eds.) (1983) *Dongxiangyu Cihui* 《东乡语词汇》 [Vocabulary of the Dongxiang Language]. Huhehaote: Neimenggu Daxue Chubanshe.
- Chen, Naixiong [陈乃雄] et al. (eds.) (1986) *Baoanyu Cihui* 《保安语词汇》 [Vocabulary of the Baoan Language]. Huhehaote: Neimenggu Daxue Chubanshe.

- van Driem, George (1992) *The Grammar of Dzongkha*. Bhutan: RGoB, Dzongkha Development Commission.
- Ebihara, Shiho [海老原志穂] (2010) *Amudo-Tibetogo no Hatuon to Kaiwa* 《アムド・チベット語の発音と会話》 [Amdo-Tibetan pronunciation and conversation]. Fuchu: Research Institute for Languages and Cultures of Asian and Africa, Tokyo University of Foreign Studies. doi: <https://doi.org/10.15026/95120>
- Ebihara, Shiho [海老原志穂] (2019) *Amudo Tibetogo Bunpoo* 《アムド・チベット語文法》 [Amdo-Tibetan grammar]. Tokyo: Hituzi Syobo.
- Haller, Felix (2000) *Dialekt und Erzählungen von Shigatse*. Bonn: VGH Wissenschaftsverlag.
- Haller, Felix (2004) *Dialekt und Erzählungen von Themchen: Sprachwissenschaftliche Beschreibung eines Nomadendialektes aus Nord-Amdo*. Bonn: VGH Wissenschaftsverlag.
- Herrmann, Silke (1989) *Dialekt und Erzählungen von Dinri*. Bonn: VGH Wissenschaftsverlag.
- Hoshi, Izumi, Shiho Ebihara, Nantaijia, and Yusuke Bessho (eds.) (2020) *Dictionary of Tibetan Pastoralism (Tibetan-Japanese)*. Fuchu: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. URL: <https://nomadic.aa-ken.jp/search>
- Hoshi, Michiyo and Thondup Tsering (1978) *Zangskar Vocabulary: A Tibetan Dialect Spoken in Kashmir*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Hoshi, Michiyo and Izumi Hoshi [星実千代 & 星泉] (1995) *Tibetogo Goisyuu* 《チベット語語彙集》 [Vocabulary of the modern Tibetan spoken in Lhasa]. Tokyo: Research Institute for Languages and Cultures of Asian and Africa, Tokyo University of Foreign Studies.
- Hua, Kan [华侃] (2002) *Zangyu Anduo Fangyan Cihui* 《藏语安多方言词汇》 [Vocabulary of the Amdo Dialect of Tibetan]. Gansu: Gansu Minzu Chubanshe.
- Huber, Brigitte (2005) *The Tibetan Dialect of Lende (Kyirong): A Grammatical Description with Historical Annotations. (Beiträge zur tibetischen Erzählforschung, 15)*. Bonn: VGH Wissenschaftsverlag.
- Kakudō, Masayoshi [角道正佳] (2012) *Dozokugo Goishū* 《土族語語彙集》 [Vocabulary of the Monguor Language]. Osaka: Osaka University.
- Lessing, Ferdinand D. (ed.) (1960) *Mongolian-English Dictionary*. Berkeley and Los Angeles: University of California Press.
- Qu, Aitang, and Kerang Tan [瞿霭堂 & 谭克让] (1983) *Ali Zangyu* 《阿里藏语》 [Ngari Tibetan]. Beijing: Zhongguo Shehui Kexue Chubanshe.
- Quejingzhabu [确精扎布] et al. (eds.) (1998) *Weilate Fangyan Cihui* 《卫拉特方言词汇》 [Vocabulary of the Oirad Dialect]. Huhehaote: Neimenggu Daxue Chubanshe.
- Siqinchaoketu [斯钦朝克图] (1999) *Kangjiayu* 《康家语》 [Kangjia Language]. Shanghai: Shanghai Yuandong Chubanshe.
- Sprigg, Richard Keith (2002) *Balti-English English-Balti Dictionary*. Richmond: Routledge Curzon.
- Sun, Zhu [孙竹] et al. (eds.) (1990) *Mengguyuzu Yuyan Cidian* 《蒙古语族语言词典》 [Dictionary of Mongolic Languages]. Xining: Qinghai Renmin Chubanshe.
- Suzuki, Hiroyuki [鈴木博之] (2007) *Sensei Minzoku Sooroo Tibetogo Hoogen Kenkyuu: Tibetogo Hoogen Bunrui Goi Siryooisyuu* 《川西民族走廊・チベット語方言研究 チベット語方言分類語彙資料集》 [Study on Tibetan Dialects in the Ethnic Corridor of West Sichuan: Materials of Vocabulary of Tibetan Dialects Sorted by Semantic Fields]. Doctoral dissertation, Kyoto University. doi: <https://doi.org/10.14989/doctor.k12734>
- Sühbaatar, O. (1997) *Mongol khelnij khar' ügijn tol'*. Ulaanbaatar: Mongolian Academy of Sciences.

Zhoumaocao [周毛草] (2003) *Maqu Zangyu yanjiu* 《玛曲藏语研究》 [A study of rMachu Tibetan]. Beijing: Minzu Chubanshe.

## References

- Iwao, Kazushi, and Takumi Ikeda [岩尾一史・池田巧] (eds.) (2021) *Chibettogo no Rekishi to Shakai* 《チベットの歴史と社会》 [History and Society of Tibet]. Kyoto: Rinsen Shoten.
- Janhunén, Juha and Kalsang Norbu (2000) Aspects of Amdo Tibetan Segmental Phonology. *Studia Orientalia* 85: 249–276. URI: <https://journal.fi/store/article/view/45082>
- Laufer, Berthold (1916) Loan-words in Tibetan. *T'oung Pao*, Second Series, 17(4/5): 403–552. doi: <https://doi.org/10.1163/156853216X00157>
- Sandman, Erika & Camille Simon (2016) Tibetan as a “model language” in the Amdo Sprachbund: evidence from Salar and Wutun. *Journal of South Asian Languages and Linguistics* 3(1): 85–122. doi: <https://doi.org/10.1515/jsall-2016-0003>
- Slater, Keith W (2021) Introduction: Language contact in the Amdo Sprachbund. *Himalayan Linguistics* 20(3): 1–7. doi: <https://doi.org/10.5070/H920355370>
- Thubten Jigme Norbu and Tsuguhito Takeuchi (1991) Mongolian Loan-words in Tibetan and their Socio-Cultural Implications. In: Steinkellner Ernst (ed.) *Tibetan History and Language: Studies Dedicated to Uray Géza on His Seventieth Birthday*, 383–396. Wien: Universität Wien.

## Publication history

Date received: 4 September 2023

Research article (Featured theme: Geolinguistic approach to Sino-Tibetan)

## Geolinguistic approach to Sino-Tibetan: Lexical relationship in some animal and plant terms

SUZUKI, Hiroyuki  
Kyoto University

YAGI, Kenji  
Kokushikan University

SUZUKI, Fumiki  
Nanzan University

**Abstract:** This study examines the results of a geolinguistic approach to the Sino-Tibetan family by combining data described separately for Sinitic and Tibeto-Burman in the projects *Studies of Asian Geolinguistics* and *Studies of Asian and African Geolinguistics*. Four animal and plant terms were discussed: ‘dog’, ‘pig’, ‘rice plant’, and ‘wheat’. Among them, ‘dog’ and ‘rice’ exhibit a lexical relationship at the root level, whereas ‘pig’ shows different roots in the two language groups, and ‘wheat’ exhibits their loan relationship.\*

**Keywords:** Sino-Tibetan; geolinguistics; lexical distribution; language contact; morphological difference

### 1. Introduction

In light of recent research, Sinitic and Tibeto-Burman are considered distinct branches of the Sino-Tibetan single-language family. However, the ILCAA projects *Studies of Asian Geolinguistics* and *Studies of Asian and African Geolinguistics* have treated both as separate members owing to differences in history, research focus, data availability, quantity, and geographical density. This study examines the results of combining data from both language groups.

Recent studies (M. Zhang et al. 2019, Sagart et al. 2019, H. Zhang et al. 2020) have shown that Sinitic and Tibeto-Burman are closely related (Sino-Tibetan), with Sinitic

---

SUZUKI, Hiroyuki, Kenji YAGI, and Fumiki SUZUKI. 2023. Geolinguistic approach to Sino-Tibetan: Lexical relationship in some animal and plant terms. *Studies in Geolinguistics* 3: 164–179. doi: <https://doi.org/10.5281/zenodo.8437113>

\* An earlier version of the article was presented at the workshop entitled ‘Geolinguistic approach to Sino-Tibetan and beyond’ in the 55th International conference of Sino-Tibetan Languages and Linguistics. We should thank the participants who provided us with insightful comments. The work was supported by the ILCAA Joint Research Project ‘Studies in Asian and African Geolinguistics’ and JSPS KAKENHI Grant Numbers JP17H04774, JP18H00670, JP18H05219, JP19K13187, JP20K13032, JP23H00625, and JP23K00486.

being the first branch of Sino-Tibetan (Bradley 2022).<sup>1</sup> An essay combining Sinitic and Tibeto-Burman using the geolinguistic method (see Shirai, this volume) is possible because the dataset is readily available.

From a technical viewpoint, combining Sinitic and Tibeto-Burman data is not complicated. However, we should pay attention to the choice of the word form. A target may share an innovation to signal a lexical derivational relationship, unlike a combination of the datasets consisting of different language families (Ebihara and Saitô, this volume).

A newly acquired term cannot adequately explain the Sino-Tibetan relationship. For instance, the word forms for ‘horse’ are common in most Sino-Tibetan languages (Figure 1)<sup>2</sup>; however, the introduction of the ‘horse’ is assumed to be after 3500 Before Common Era (BCE), a period that is considered to be the domestication of horses in Central Asia (see Librado et al. 2021). Furthermore, the word form for ‘horse’ is regarded as a *Wanderwort* ‘migratory word’ that is attested across language families (Bjørn 2022).

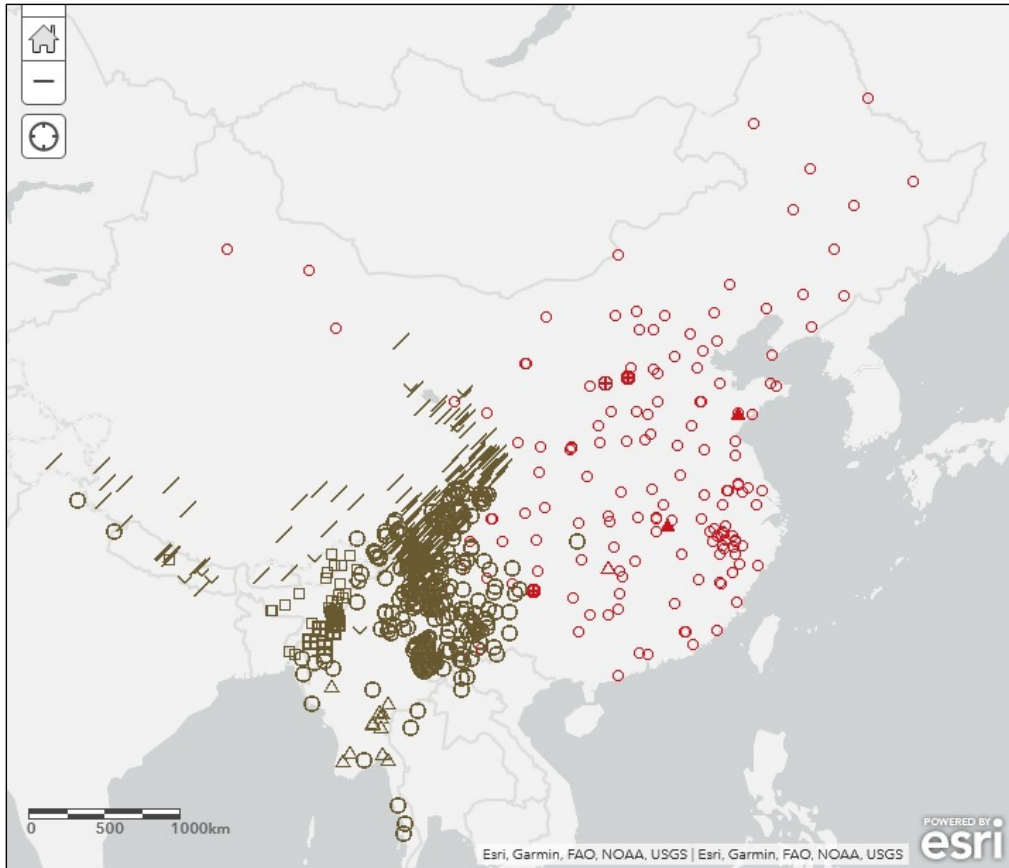
In addition to the ‘horse’ shown in Figure 1, this article considers four lexical forms, the ‘dog,’ ‘pig,’ ‘rice plant,’ and ‘wheat.’ We present maps containing both Sinitic and Tibetan-Burman data simultaneously. This study aims to demonstrate how Sino-Tibetan linguistic maps function in geolinguistic studies. The abbreviations SN (Sinitic), TB (Tibeto-Burman), and ST (Sino-Tibetan) are used in the tables and legends. The symbols on the maps are red for SN and brown for TB. For detailed descriptions and analyses of each SN and TB lexical item, please refer to Table 1. For bibliographical information on the data sources, see LAA (2021) and LAAA-III (2023).

Table 1 Previous works of each lexical item.

SN	‘horse’	Yagi (2022a)
	‘dog’	Yagi (2022b)
	‘rice plant’	Yagi and Ueya (2021)
	‘wheat’	F. Suzuki (2023)
TB	‘horse’	Ebihara et al. (2022a)
	‘dog’	Ebihara et al. (2022b)
	‘pig’	H. Suzuki (2019)
	‘rice plant’	H. Suzuki et al. (2016, 2021)
	‘wheat’	H. Suzuki et al. (2023a)

<sup>1</sup> We find more recent studies examining the Sino-Tibetan lexical relationship, for example, Jacques and d’Alpoim-Guedes (2023) for ‘*Zanthoxylum*’.

<sup>2</sup> A clear exception is observed in Tibetic and Tamangic languages, which use a word derived from Type B: *\*r-ta*. In many Tibetic languages, a form derived from Type A (*\*s-mraj*) is used for a word meaning ‘wild yak’ (Literary Tibetan [LT] ‘*brong*’).



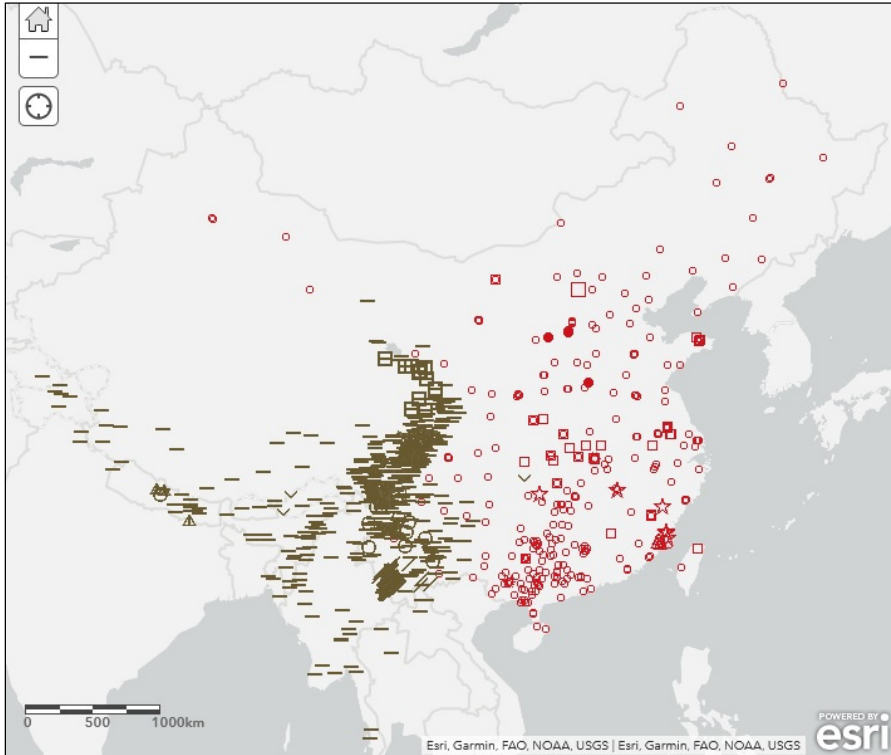
- A. type
- A-1 马 ma
  - ⊕ A-2 马儿 mar
  - A. *\*s/m-raŋ* type  
nbzəŋ, m̄zəŋ<sup>31</sup>, <sup>a</sup>brā, wə, lu, ran, zu, ɬəq, 'jy, kò, guē<sup>13</sup>, buro, byin, nyang, m̄ŋo<sup>35</sup>, man, myin<sup>H</sup>, sɔrɔ, mbo ro, mbro, vre, fire, etc.; mo<sup>21</sup>ku<sup>55</sup>lu<sup>55</sup>, mu<sup>21</sup>pa<sup>55</sup>, mɔ<sup>21</sup>pha<sup>21</sup>, mo<sup>44</sup>ŋ<sup>21</sup>, mjo<sup>44</sup>tha<sup>42</sup>, etc. (suffixed); a<sup>55</sup>m̄(y)<sup>21</sup>, a<sup>21</sup>bəŋ<sup>21</sup>, a<sup>55</sup>mo<sup>31</sup>, i<sup>35</sup>mv<sup>53</sup>, ḡum<sup>31</sup>raŋ<sup>31</sup>, gim rang, pa<sup>31</sup>xoŋ<sup>35</sup> (prefixed); a<sup>55</sup>mo<sup>31</sup>za<sup>31</sup>, etc. (affixed).
  - ∕ B. *\*r-ta* type  
l̄ta, rta, ʃta, ta, <sup>h</sup>ta, <sup>h</sup>to, te<sup>53</sup>, da<sup>1</sup>, etc.
- SN:
- ▲ A-3 马子 ma t̄ɬi
  - △ A-4 马特 ma ku
  - C. *\*ghoḍā* type  
(N)gohḍa:q, guree, gura, gum<sup>4</sup>raŋ<sup>1</sup>, k̄ierü, keru, kuri, kor ḡu-ri, etc; sə-kuu, etc. (prefixed).
  - ⊞ C-compound type  
sya-n +C: sa gol, sa kawr, sa-koi, sāl kòl, si-kwe.
  - △ D. *\*k-sre<sup>7</sup>* type  
kǎ θe<sup>3</sup>, kǎ<sup>11</sup>se<sup>45</sup>, se<sup>31</sup>, θè, ngshe, etc.
  - ∨ X. others  
lalo, cop̄i, s<sup>h</sup>əp̄ü<sup>2</sup> mok, pferi, <sup>h</sup>gō, etc.
- TB:

Fig. 1 'Horse' in Sino-Tibetan.



## 2. ‘Dog’

Figures 2 and 3 show the word forms for ‘dog’ in ST. SN and TB employ common roots of a word for ‘dog’ (Type A in SN and TB).



- A. ☆ A-1 犬k<sup>h</sup>eiŋ
- B. ○ B-1 狗kou, kɿ, kau, kai, ki, tɛiv, teiɬu  
● B-2 狗儿kəu ər, 狗儿kəur  
□ B-3 狗子kou tsɿ, ke tsæʔ, kieu tsɿ, kau a, e tsie
- C. ☆ C Others: 来福, 地羊ti iəŋ
- B-4 狗娃子kəu va tsəʔ  
△ B-5 狗团kəu (k-)jiaN
- SN:
- A. \**d-k<sup>v</sup>əy-n* type  
c<sup>h</sup>i<sup>53</sup>, ci<sup>55</sup>, tɛ<sup>h</sup>ə, khur<sup>33</sup>, tshi<sup>33</sup>, tshz<sup>11</sup>khv<sup>33</sup>,  
khi<sup>33</sup>, khui<sup>44</sup>, phur<sup>53</sup>, kui<sup>31</sup>, tefü, kei, si, ʔüy,  
ci, güy, thwī, hi, thi<sup>11</sup>, <sup>23</sup>dži, šəy, hui, i, üi,  
ʃi<sup>33</sup>, fū za, thwi, wī, gui hen, ʔüy, foi, hi,  
etc.; tɛhəru, chi<sup>55</sup>bo<sup>55</sup>, khur<sup>31</sup>ga<sup>31</sup>,  
khu<sup>33</sup>jo<sup>33</sup>, khli-tsa, khi:bu, fhurro, khira,  
khinu, etc. (suffixed); ɔ<sup>31</sup>khur<sup>31</sup>, azü,  
mɔ<sup>55</sup>khur<sup>21</sup>, etsü, lə<sup>31</sup>kha<sup>35</sup>, <sup>3</sup>jukyu, əgi<sup>31</sup>,  
atsü, etc. (prefixed)
  - △ C+A: 'no kju, nā ki
  - ∨ B. \**m-par* ∞ *pra* type  
phur<sup>3</sup>, phv<sup>3</sup>, phi<sup>3</sup>, etc.
  - C. \**na* type  
nagi, nā ki, 'nakyn, etc. (suffixed); a<sup>33</sup>no<sup>21</sup>,  
etc. (prefixed)
  - D. *hapa* type  
haba, xapa, hapa, etc.
  - ∨ X. Others  
wathi, kotsbo, a-chak, wək, boh, nyi<sup>4</sup>mur<sup>2</sup>,  
tsuʔyu, etc.
- TB:
- ◀ A+C: khe<sup>33</sup>ne<sup>55</sup>, khəna, khue<sup>53</sup>ɿ<sup>33</sup>

Fig. 2 ‘Dog’ in Sino-Tibetan.



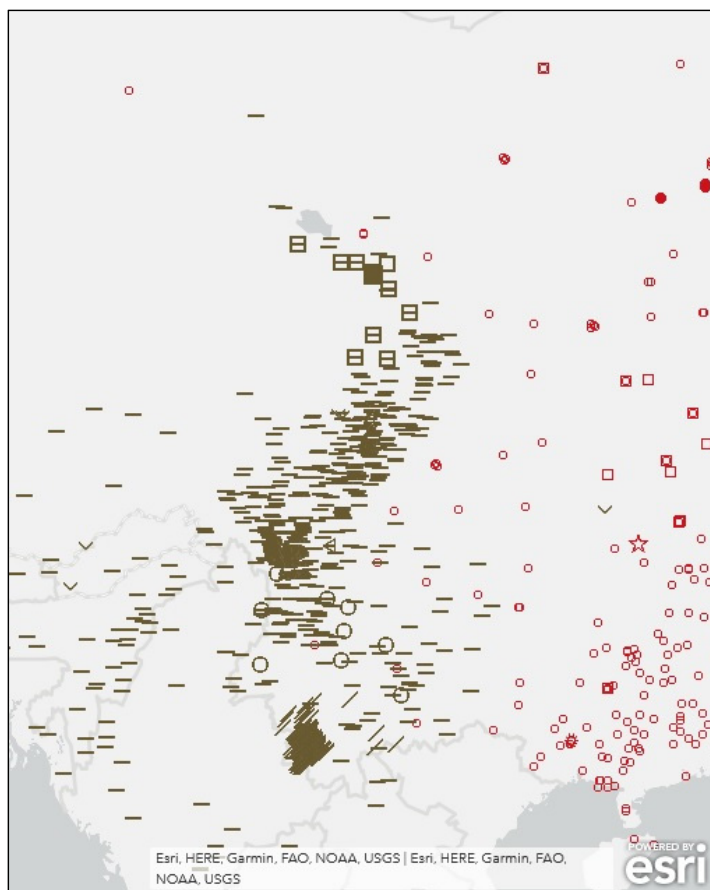


Fig. 3 ‘Dog’ in Sino-Tibetan (ST borderland enlarged).

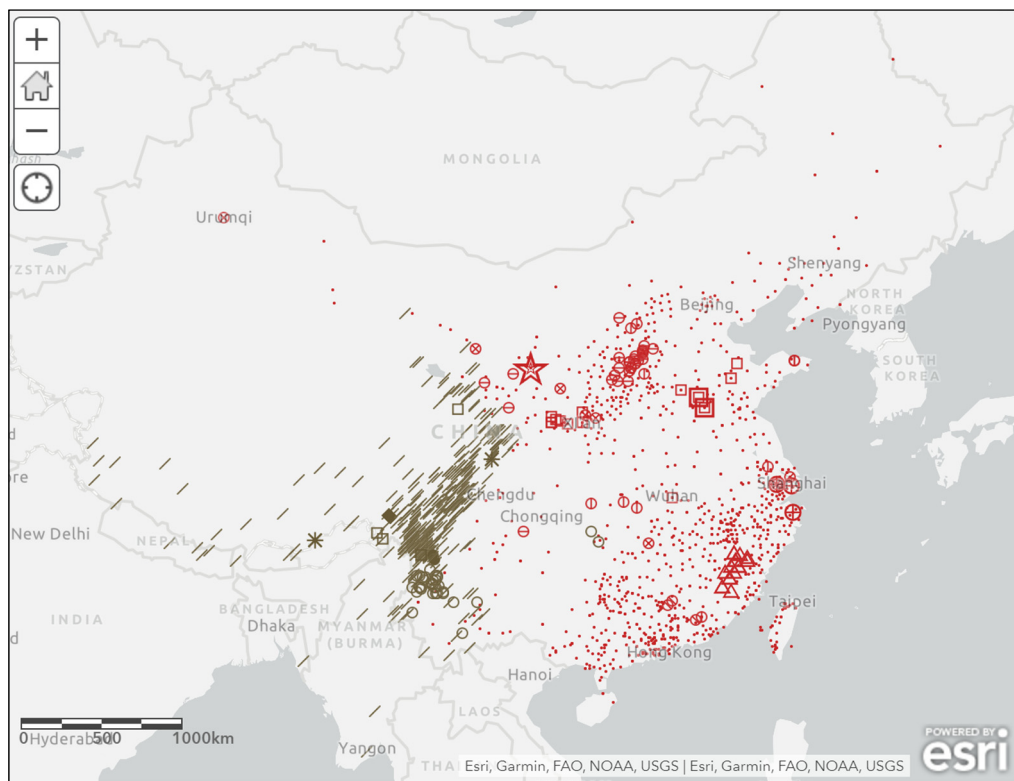
The SN *quan* 犬 (OC \*[k]<sup>whs</sup>[e][n]?) and TB \*d-k<sup>w</sup>əy-n are from the same root. Most SN languages and varieties use *gou* 狗 (OC \*Cə.k<sup>ʰ</sup>ro?) for ‘dog’, considered to have originally been borrowed from Hmong-Mien (Akitani et al. 2022).<sup>3</sup> In addition, the TB languages spoken in south-western Yunnan use other roots.

Tibetic languages spoken in north-eastern Tibetosphere (mainly Amdo Tibetan) distinguish a large herding dog (< \*d-k<sup>w</sup>əy-n) from a small long-haired pet dog by borrowing the Sinitic form *habagou* 哈巴狗 ‘Pekinese’ (Ebihara et al. 2022b). Sound forms of the same origin in the SN and TB generally follow their proto-forms (proto-Sinitic and proto-TB). For example, Tibetic follows the sound correspondence of each variety with LT *khyi* (or its ancient variant *kʷi*). The present dataset suggests that there is no influence beyond these branches.

<sup>3</sup> However, this view is still in debates.

### 3. ‘Pig’

Figures 4 and 5 show the word forms for ‘pig’ (general term) in ST.



- |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>A. <i>zhu</i> 猪 type</p> <ul style="list-style-type: none"> <li>• A-1 <i>zhu</i> 猪: tʂu, tʂy, tʂɿ, tsu, tsy, tsɿ, tsy, tsɿ, tey, tei, teie, ʃu, ʃy, ʃɿ; pʃu, ty, ti, tu, tuu, te, tiu, tiəu, tɔy, ʔdu, ly, li, etc.</li> <li>⊖ A-2 <i>zhur</i> 猪儿: tʂur, tsur, tʂyər, ʃy zɿ, tsu ə, tsu v, tsuβ əl, tsy ər, tsu ə, tsu ai, tsəf ər, tsy zɿʔ, tʂuəv, etc.</li> <li>⊕ A-3 <i>zhuzi</i> 猪子: tʂu tsɿ, tsu tsɿ, tsu ve, tɔy tə, tsu tsəʔ, tsu ləʔ, etc.</li> <li>⊗ A-4 <i>zhuli</i> 猪哩: tsu ɛ, tɔy lei, ʃu li</li> <li>⊗ A-5 <i>zhunong</i> 猪农: tsɿ nonɿ</li> </ul> <p>SN</p> <p>／ A. PTB *pʷak: pʰaʔ, haʃ, va, ve, paʃ, weʔ, etc.</p> <p>◆ B. Proto-Tani *rjek: ʔa' re.</p> <p>TB</p> <p>○ C. archaic loan from Sinitic <i>zhu</i> 猪: te, etc.</p> | <ul style="list-style-type: none"> <li>⊕ A-6-1 <i>zhulo</i> 猪罗: tsɿ lu, tsy lau, tsy lo</li> <li>⊗ A-6-2 <i>zhulolo</i> 猪唠唠: tʂu lo lo, tʂu lao lao, tsy loo loo, tɔy lo lo, tsu lo lo, etc.</li> </ul> <p>B. <i>xi</i> 豨 type</p> <ul style="list-style-type: none"> <li>△ kʰy, kʰyi, kʰye, kʰə, kʰui, kʰəu, hui, etc.</li> </ul> <p>C. <i>lolo</i> type: onomatopoeia</p> <ul style="list-style-type: none"> <li>□ C-1 <i>lolo</i> 唠唠: lo lo, lau lau, no no, etc.</li> <li>◻ C-2 <i>heilolo</i> 黑唠唠: xei lo lo</li> </ul> <p>D. others: Muslim avoid “pig” as taboo</p> <ul style="list-style-type: none"> <li>☆ D-1 <i>heizi</i> 黑子</li> <li>☆ D-2 <i>jiudiwen</i> 就地闻</li> </ul> <ul style="list-style-type: none"> <li>● D. derived from Sinitic <i>zhu</i> 猪: tʂu.</li> <li>□ E. <i>IV</i>-type: lo lo, bi la, etc.</li> <li>* X. others</li> </ul> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Fig. 4 ‘Pig’ in Sino-Tibetan.

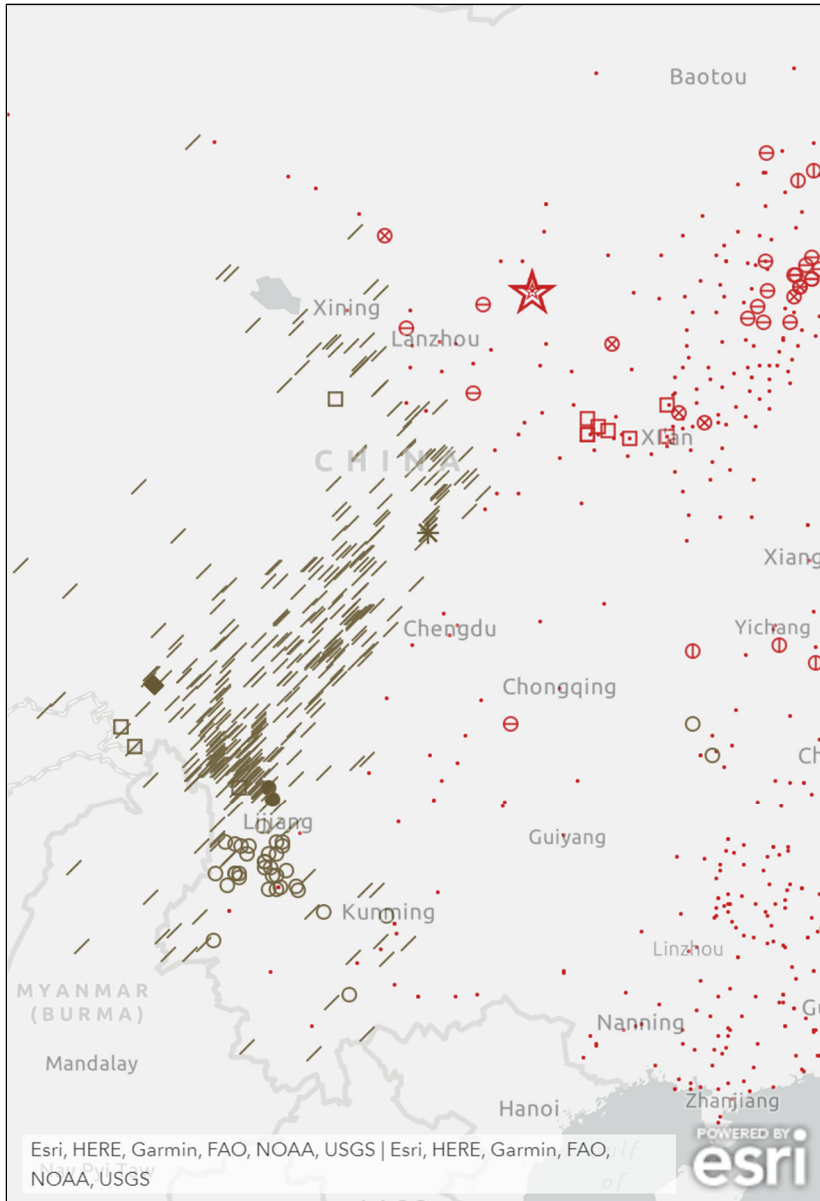


Fig. 5 'Pig' in Sino-Tibetan (ST borderland enlarged).

Although pigs are widely domesticated in the ST area,<sup>4</sup> Figure 4 shows that SN and TB employ different roots of the word for the 'pig'. The root of SN is derived from *zhu* 猪 (OC \*tra > MC trjo) with a minority derived from *xī* 豨 (OC \*q<sup>h</sup>əjʔ > MC xj+jX), whereas the principal root of TB is derived from \*p<sup>w</sup>ak (STEDT #1006; Type A) with

<sup>4</sup> The first domestication of pigs in China dates back to ca. 6000 BCE (Price and Hongo 2020). This period is before the hypothetical split of SN and TB.

two minorities; one is derived from \*rjek (STEDT #6019; Type B), whereas the other is a form including a denti-alveolar plosive or affricate initial (Type C).<sup>5</sup>

SN *zhu* was most likely the original connection for type C TB. The distribution of Type C, which is limited to the Baic (Bai) and Bisic (Tujia) languages, is closely connected to Sinitic languages from the viewpoint of language contact. SN Type C and some subtypes of Type A contain a /IV/ syllable, which can be explained as an onomatopoeia for calling pigs (e.g., Shen 2005:165). Interestingly, some TB languages use a similar form (Type E), including examples with the initial /l/ in the word form, some of which are connected to an expressive reflecting the addressing pigs, although no reports have identified the lexical origin. However, it is possible that the /IV/ syllable also functions in addressing pigs.<sup>6</sup> Nevertheless, this commonality is coincidental from a geolinguistic viewpoint.

New word forms that avoid taboos in Muslim culture were also found in SN (Type D). Several TB languages are also spoken in similar cultural environments, such as in Qinghai and Gansu; however, it has not been reported whether the word ‘pig’ was substituted by another form. Notably, Type D-1 uses the morpheme ‘black’ for ‘pig’, which is generally black in the given region. Some Tibetic languages have a similar implication to the word derived from *phag*, which is always black (H. Suzuki 2019).

#### 4. ‘Rice plant’

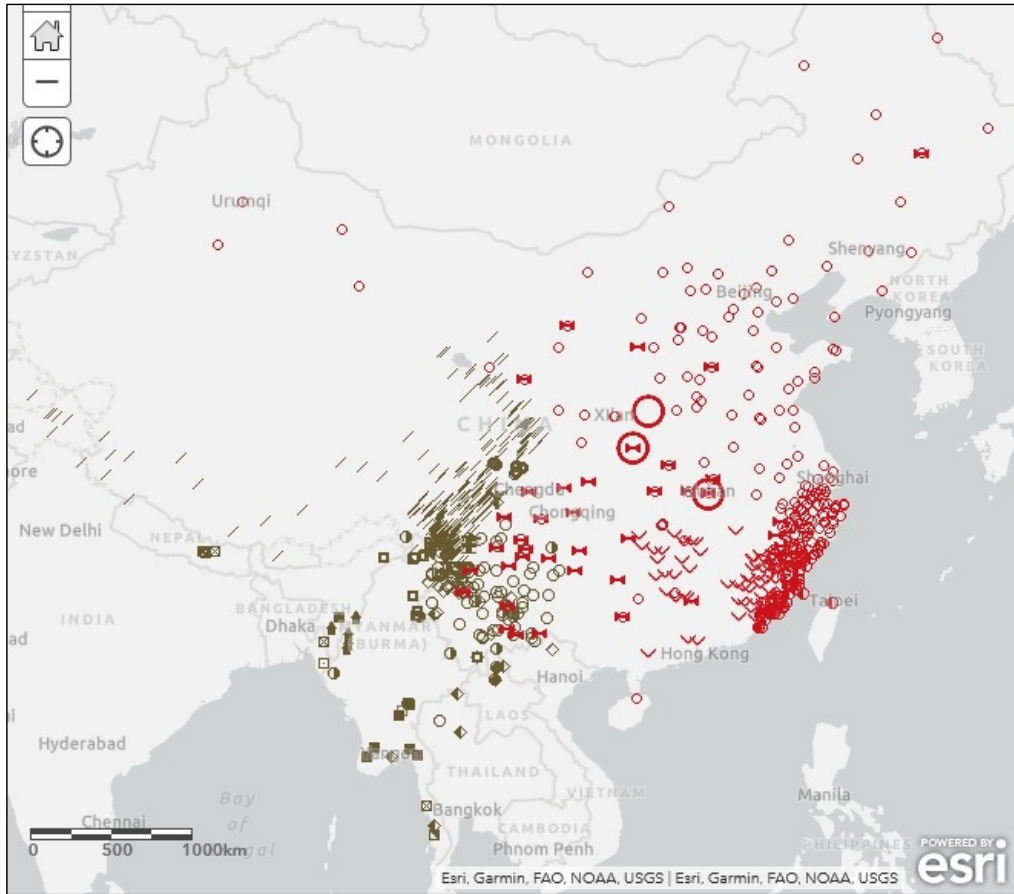
SN and TB have many roots for the rice-related semantic field, such as ‘rice plant’, ‘rice grain’, and ‘cooked rice’; for details on TB, see Suzuki et al. (2016). Figures 6 and 7 show the word forms for ‘rice plant’ in ST.

It is believed that rice cultivation originated within the current Sinitic-speaking area (Huang et al. 2012). Hence, the rice plant has played an essential role in the alimentionation of the people inhabiting the ST-speaking area. As mentioned previously, many ST languages possess multiple lexical roots for rice-related words. Due to this background, word forms for ‘rice plant’ in the languages within ST diverged greatly. This led to difficulty in identifying cognates between the SN and TB.

The root of SN is derived from *dao* 稻, whereas the root of TB is derived from various forms, such as \*b-ras (STEDT #2071), \*khrəy (STEDT #5867), and \*ma-y 𑄎 \*mey (STEDT #2445). The last root is related to SN *mi* 米 ‘rice grain’; \*khrəy is used for ‘millet’ in some TB languages (H. Suzuki et al. 2023b, c).

<sup>5</sup> Note that large lexical discrepancies are attested in specific terms for ‘pig’, such as ‘boar’, ‘sow’, and ‘piglet’, especially in the eastern Tibetosphere (H. Suzuki 2022).

<sup>6</sup> In the Tibetic context, the /IV/ syllable is connected to a noun for ‘cat’ and addressing voice to it (Qin and Suzuki 2016).



- |   |            |   |           |   |         |
|---|------------|---|-----------|---|---------|
| ○ | A1 tau/dao | ⊕ | C1 ku     | ★ | D1 sy   |
| ● | A2 tiu     | ⊙ | C2 tau ku | ⚡ | D2 tsau |
| ∇ | B vu       |   |           |   |         |
- SN
- ∕ A. PTB etymon \**b-ras*, incl. reflexes of WrT '*bras*': *mbrat*, *mdzwi*, *bras*, *bdza*ː, <sup>1</sup>*ɖah*, <sup>1</sup>*ɖe*, <sup>1</sup>*ɡu*ː, <sup>1</sup>*ji*ː, <sup>1</sup>*dze*ː, <sup>1</sup>*be*ː, etc.
  - B. PB etymon \**čan1*, incl. reflexes of WrB *chan*: <sup>33</sup>*tshu*, <sup>33</sup>*tehe*, <sup>33</sup>*tshe*, <sup>33</sup>*tshu*<sup>44</sup>, *cà*, etc.
  - ◆ B+ C. PTB etymon \**bu*: *biq*, *bi*, *biu*, *biu*, etc.
  - E+ D. PTB etymon \**ma-y* < \**mey*: DM: *mam*, *maam*, <sup>a44</sup>*me*<sup>44</sup>, <sup>me35</sup>*me*<sup>31</sup>, etc.; DA: *aŋ*, <sup>am33</sup>*am*<sup>33</sup>, <sup>aŋ31</sup>*bu*<sup>55</sup>, *a:ŋ*<sup>55</sup>, etc.
  - ▣ DA E. PTB etymon \**kuk*, incl. that of PB \**ʔgok*: *kaukF*, *kuk*<sup>31</sup>, *ko*<sup>33</sup>, *khə*<sup>33</sup>, *quo*<sup>31</sup>, *go*<sup>31</sup>, etc.
  - DM F. PB etymon \**hay*: <sup>a55</sup>*ho*<sup>21</sup>, etc.
  - L G. PTB etymon \**m-dz(y)a-k/n/t/s*: *dza*<sup>31</sup>.
  - ⊕ H. PKC etymon \**θaaj* ('millet'): *ta:ŋ*<sup>2</sup>, *tsaaj*, etc.
  - ⊙ I. \**kʰr*/<sup>1</sup>*qʰr*-type (cf. WrT *khre* 'millet'): *kʰru*<sup>235</sup>, *qhə*<sup>4</sup>, *kʰə*<sup>44</sup>, *kʰrŋ*, etc.
  - ◆ J. reflexes of WrT ᨡᨣᨠᨣ *drus ma* 'polished grain': *te*: *ma*, *te ma*, etc.
  - ▣ DM+ K. *ba*-type: *ba*, *wā*.
  - F L. *s*-type: <sup>siu</sup>*ur*<sup>55</sup>, <sup>eu</sup>*ur*<sup>55</sup>, <sup>so</sup>*u*<sup>44</sup>, <sup>ɣv</sup>*u*<sup>44</sup>, <sup>ei</sup>*u*<sup>31</sup>, <sup>ɣu</sup>*u*<sup>55</sup>, etc.
  - + G M. *do*-type: *də*<sup>35</sup>.
  - K+ N. *l*-type: <sup>li</sup>*za*<sup>53</sup>.
  - † M
  - ⊕ N
  - \* O. *ny*-type: *ny*.
  - TB

Fig. 6 'Rice plant' in Sino-Tibetan.

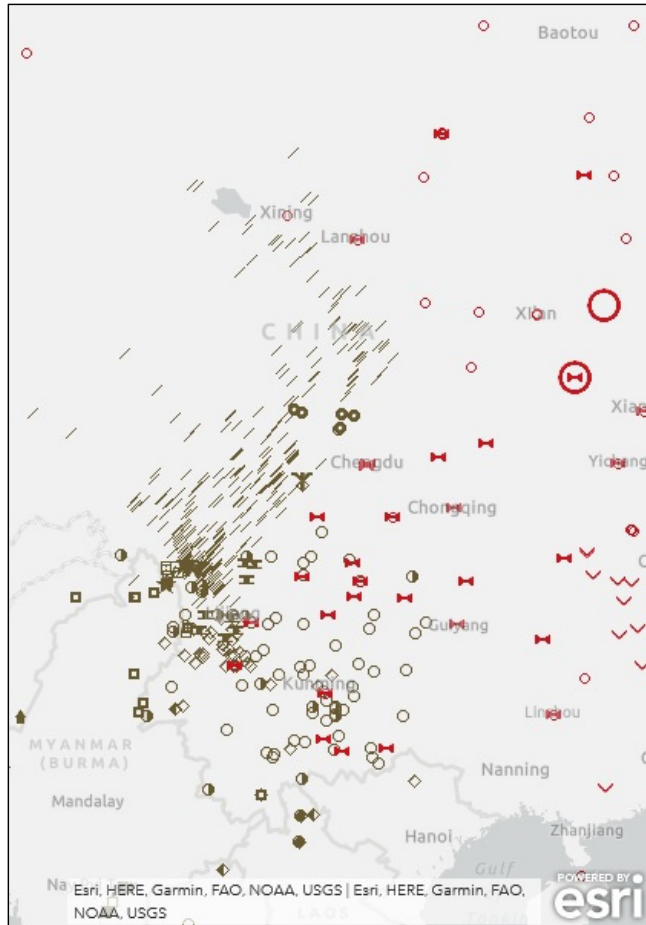


Fig. 7 'Rice plant' in Sino-Tibetan (ST borderland enlarged).

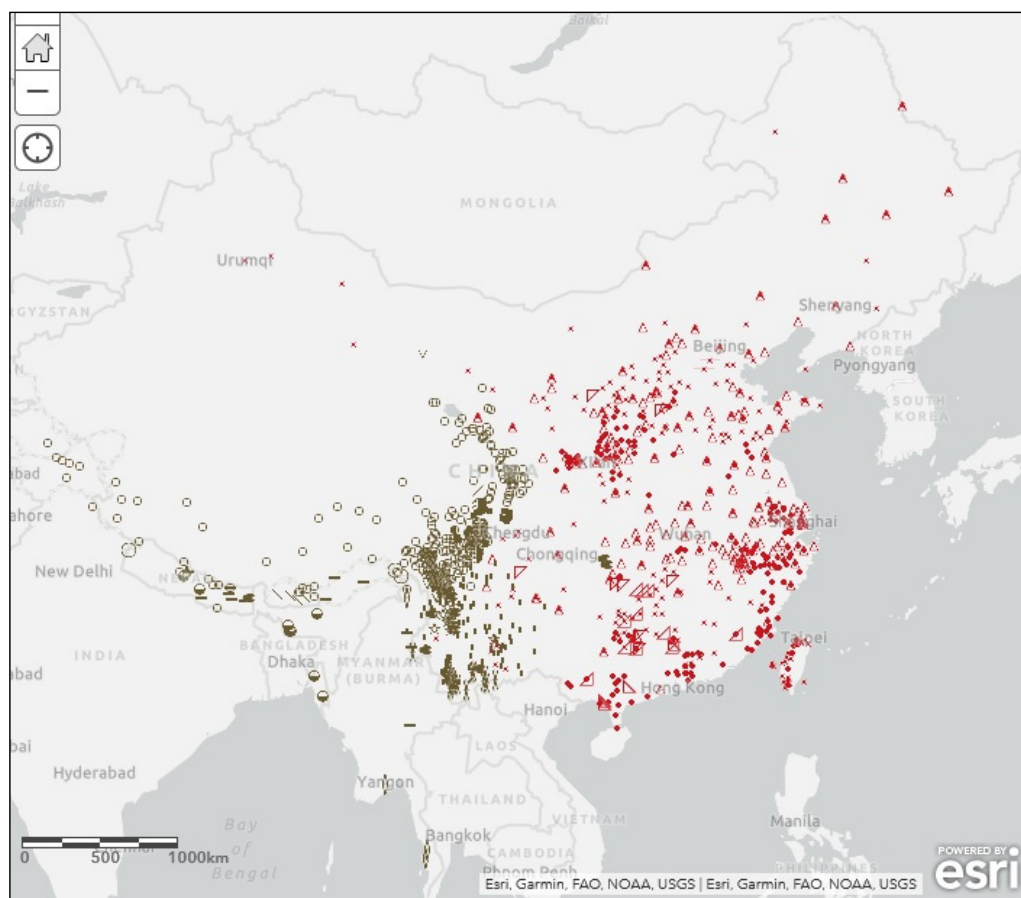
Although rice is cultivated in the ST-speaking region, some languages are also spoken in non-rice-cultivating areas, such as the Tibetan Plateau. Therefore, mismatches between roots and meanings beyond languages can occur due to different processes of lexical diffusion.<sup>7</sup> This also suggests that crop terms at the ST proto-language level have not yet been fixed for specific plants. This further implies that rice began to be cultivated after the split of ST into SN and TB, based solely on linguistic evidence.

## 5. 'Wheat'

Figures 8 and 9 show the word forms for 'wheat' in ST.

<sup>7</sup> See, for example, Suzuki and Sonam Wangmo (2016) for the expansion of the word form *'bras* in the Tibetosphere.





- |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>A. stem (+ suffix) type</p> <ul style="list-style-type: none"> <li>• A-1 <i>mai</i> 麦 (monosyllabic)</li> <li>• A-2 <i>maizi</i> 麦子</li> </ul> <p>B. modifier type</p> <ul style="list-style-type: none"> <li>△ B-1 <i>xiaomai</i> 小麦</li> </ul> <p>SN</p> <p>A. PTB *m-grwa</p> <ul style="list-style-type: none"> <li>■ A1: <sup>n</sup>qo, <sup>n</sup>qu, etc.</li> <li>○ A2: <sup>t</sup>o, <sup>k</sup>ro, <sup>co</sup>, <sup>teo</sup>, <sup>c</sup>ço, <sup>t</sup>ʂ<sup>h</sup>o<sup>12</sup>, etc.</li> <li>⊕ A3: <sup>rgak</sup>, <sup>rgo</sup>, <sup>lgap</sup>, <sup>ka lie</sup>, etc.</li> <li>● A4: <sup>lu</sup>, <sup>zi</sup>, <sup>ji</sup>, <sup>jo</sup>, etc.</li> <li>● A5: <sup>kaj</sup>, <sup>qo</sup><sup>24</sup>, <sup>vuə</sup><sup>241</sup>, etc.</li> <li>● A6: <sup>gom</sup>, <sup>goŋ</sup>, etc.</li> </ul> <p>B. Coronal initial type</p> <ul style="list-style-type: none"> <li>▮ B1: <sup>zi</sup><sup>33</sup>, <sup>zɿ</sup><sup>33</sup> <sup>ze</sup><sup>31</sup>, <sup>n</sup>dza, <sup>tso:dza</sup>, etc.</li> <li>▮ B2: <sup>sa</sup>, <sup>fua</sup><sup>44</sup>, <sup>ʂa</sup><sup>44</sup>, <sup>xa</sup><sup>33</sup>, etc.</li> </ul> <p>TB</p> <p>C. /m/-initial type</p> | <ul style="list-style-type: none"> <li>△ B-2 <i>ximai</i> 细麦</li> <li>△ B-3 <i>hemai</i> 禾麦</li> <li>△ B-4 <i>mianmai</i> 面麦</li> <li>▽ B-5 others: <i>huangmai</i> 黄麦, etc.</li> <li>— C. others: <i>xialiang</i> 夏粮, <i>suozi</i> 梭子</li> </ul> <ul style="list-style-type: none"> <li>☆ C1: <i>mu</i>, etc.</li> <li>● C2: <i>se mu</i>, etc.</li> </ul> <p>D. Semantic change type</p> <ul style="list-style-type: none"> <li>▽ D1: <sup>nē</sup>., <sup>nje</sup>., <sup>ne</sup>., etc.</li> <li>△ D2: <sup>s<sup>h</sup>o</sup> wa</li> <li>∕ D3: <sup>ɛ</sup>ŋo, <sup>ŋo</sup></li> <li>∖ D4: <sup>kə</sup> rə, <sup>ka</sup> rə, <sup>kā</sup>, <sup>kar</sup></li> </ul> <p>E. Loanword type</p> <ul style="list-style-type: none"> <li>✦ E1: Sinitic; <i>maizi</i>, <i>xiaomai</i>, <i>mu tsi</i></li> <li>✦ E2: Indic; <i>joŋ</i>, <i>jouŋ<sup>M</sup></i>, <i>zong</i></li> <li>— X. miscellaneous</li> </ul> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Fig. 8 'Wheat' in Sino-Tibetan.

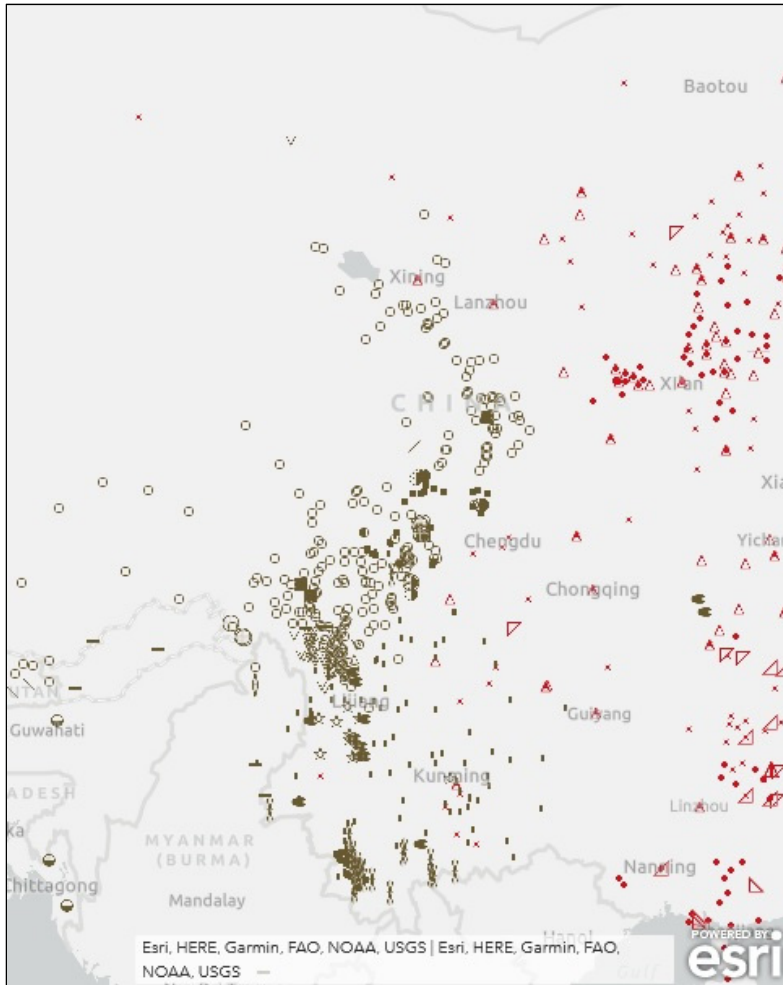


Fig. 9 ‘Wheat’ in Sino-Tibetan (ST borderland enlarged).

The word forms for ‘wheat’ generally differ in Sinitic and TB. A small number of TB varieties spoken in Sichuan and Yunnan present forms similar to their Sinitic counterparts (Types A and B for SN and Type C for TB). This similarity does not exhibit a clear relationship with proto-level forms but with a loan process in an earlier period since wheat was introduced in China around 2600 BCE (Long et al. 2018), corresponding to the period after the split into SN and TB. Although there are some lexical similarities between languages spoken in the Sinitic-TB borderland, this is predominantly due to recent lexical borrowing.

Bradley (2011:140) proposes three TB etyma: \*ša<sup>3</sup>, WrT *gro*, and \*gom/goŋ. Our analysis identified two main types: A (\*m-grwa group) and B (\*ša<sup>3</sup> group; south-eastern TB area). Type C is mainly attested within Type B territory. We relate Type C to the



Sinitic word *mai* (OC \*m-rʰək). Bradley (2011:136) states that the Sinitic form *mai* ‘appears to be a loan from Indo-European, possibly via Turkic’. Hence, if our analysis is correct, the varieties using Type C received the Sinitic lexical form after Sinitic was borrowed from IE via Turkic (i.e., via the present Xinjiang).

The Tibeto-Burman analysis does not distinguish between Sinitic loan forms (Type E1). However, the analysis of Sinitic distinguishes *maizi* 麦子 (Type A2) from *xiaomai* 小麦 (Type B1). Type B1 was observed at several locations in Sichuan, whereas Type A2 was observed in Yunnan. In Tibeto-Burman, many loanword forms are Type A2 (Hani, Jino, Lahu, and Kucong), with a minority of Type B1 (Hani). Because Type B1 is used in the northern and central regions of the Sinitic distribution area, the use of Type B1 in TB could be a recent loan through the expansion of Putonghua.

In the Tibetosphere, the distribution of Type A2 is wide, and its form corresponds to the LT *gro*. However, in the area where the wheat grows instead of the highland barley (Yunnan), the word form for ‘wheat’ is substituted by other crops’ terms, as shown in Types D1 (< LT *nas* ‘highland barley’) and D2 (< LT *so ba* ‘barley’). It appears in the Tibetic languages, particularly in varieties spoken at lower altitudes. Therefore, vegetal cultivation is likely to trigger semantic changes. Potentially, the lexical classification of crop species in Tibetic may have originated from varieties spoken at higher altitudes.

## 6. Concluding words

This article examined the lexical forms for ‘dog’, ‘pig’, ‘rice plant’, and ‘wheat’ in ST languages. The SN and TB datasets were initially compiled and analysed separately following the projects *Studies of Asian Geolinguistics* and *Studies of Asian and African Geolinguistics*. This article discusses how similarities and differences appear in these words.

The lexical variation between the SN and TB was relatively large, even for basic words. Among them, ‘dog’ and ‘rice’ exhibit the lexical relationship at the root level, whereas ‘pig’ demonstrates different roots in the two language groups, and ‘wheat’ exhibits their loan relationship. Furthermore, more roots were used differently in TB than in SN.

Early loan forms and their expansion could make the two branches divergent. More case studies are needed to analyse mutual lexical development. Additionally, semantic changes are worth considering in the field of geolinguistics.

## References

- LAA=Endo, Mitsuaki, Makoto Minegishi, Satoko Shirai, Hiroyuki Suzuki, and Keita Kurabe (eds) (2021) *Linguistic atlas of Asia*. Tokyo: Hituzi Syobo.
- LAAA-I=Suzuki, Hiroyuki, Mika Fukazawa, Akiko Yokoyama, and Mitsuaki Endo (eds) (2022) *Linguistic atlas of Asian and Africa I*. Tokyo: Geolinguistic Society of Japan. doi: <https://doi.org/10.5281/zenodo.7118188>
- LAAA-II=Suzuki, Hiroyuki, Kohei Nakazawa, and Mitsuaki Endo (eds) (2023) *Linguistic atlas of Asian and Africa II*. Tokyo: Geolinguistic Society of Japan. doi: <http://doi.org/10.5281/zenodo.7754469>
- LAAA-III=Fukushima, Chitsuko, Satoko Shirai, Mika Fukazawa, Hiroyuki Suzuki, and Mitsuaki Endo (eds) (2023) *Linguistic atlas of Asian and Africa III*. Tokyo: Geolinguistic Society of Japan.
- STEDT=The Sino-Tibetan etymological dictionary and thesaurus. Online database. <https://stedt.berkeley.edu/~stedt/cgi/rootcanal.pl>
- Akitani, Hiroyuki [秋谷裕幸], Weihui Wang [汪维辉], and Masaki Nohara [野原将挥] (2022) Shuo {gou} 说 {狗} [On {DOG}]. In *Iwata Ray Jiaoshou rongxiu jinian lunwenji bianjizhu* 《岩田礼教授荣休纪念论文集》编辑组 (ed) *Iwata Ray Jiaoshou rongxiu jinian lunwenji (shangce)* 《岩田礼教授荣休纪念论文集 上册》, 264–280. Tokyo: Geolinguistic Society of Japan. doi: <https://doi.org/10.5281/zenodo.6342364>
- Baxter, William H. and Laurent Sagart (2014) *Old Chinese: A new reconstruction*. Oxford: Oxford University Press. doi: <https://doi.org/10.1093/acprof:oso/9780199945375.001.0001>
- Bjørn, Rasmus G. (2022) Indo-European loanwords and exchange in Bronze Age Central and East Asia: Six new perspectives on prehistoric exchange in the eastern steppe zone. *Evolutionary Human Sciences* 4: e23. doi: <https://doi.org/10.1017/ehs.2022.16>
- Bradley, David (2011) Proto-Tibeto-Burman grain crops. *Rice* 4: 134–141. doi: <https://doi.org/10.1007/s12284-011-9074-y>
- Bradley, David (2022) Sinitic as the first branch of Sino-Tibetan. Conference paper presented at CIEL-13.
- Ebihara, Shiho, Kazue Iwasa, Keita Kurabe, Satoko Shirai, and Hiroyuki Suzuki (2022a) ‘Horse’ in Tibeto-Burman. *LAAA-I*, 138–140.
- Ebihara, Shiho, Kazue Iwasa, Keita Kurabe, Satoko Shirai, and Hiroyuki Suzuki (2022b) ‘Dog’ in Tibeto-Burman. *LAAA-I*, 180–182.
- Ebihara, Shiho and Yoshio Saitô (this volume) Language contact between Tibetic and Mongolic, with a special focus on the Amdo Linguistic Area. doi: <https://doi.org/10.5281/zenodo.8437105>
- Guo, Xiliang [郭锡良] (2010) *Hanzi guyin shouce (zengdingben)* 《汉字古音手册（增订本）》 [*A handbook of ancient sounds of Chinese characters (enlarged and revised version)*]. Beijing: Shangwu yinshuguan.
- Huang, Xuehui, Nori Kurata, Xinghua Wei, et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490(7421): 497–501. doi: <https://doi.org/10.1038/nature11532>
- Jacques, Guillaume and Jade d’Alpoim Guedes (2023) Sichuan peppercorn and the birth of numbing spices in East Asia. *Ethnobiology Letters* 14(1): 10–23. doi: <https://doi.org/10.14237/eb1.14.1.2023.1842>
- Karlgren, Bernhard (1957) *Grammata serica recenssa*. The Museum of Far Eastern Antiquities, Bulletin 29.
- Librado, Pablo, Naveed Khan, Antoine Fages, et al. (2021) The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* 598(7882): 634–640. doi: <https://doi.org/10.1038/s41586-021-04018-9>

- Long, Tengwen, Christian Leipe, Guiyun Jin, Mayke Wagner, Rongzhen Guo, Oskar Schröder, and Pavel E. Tarasov (2018) The early history of wheat in China from  $^{14}\text{C}$  dating and Bayesian chronological modelling. *Nature Plants* 4: 272–279. doi: <https://doi.org/10.1038/s41477-018-0141-x>
- Matisoff, James A. (2003) *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley/Los Angeles/London: University of California Press. URI: <https://escholarship.org/uc/item/19d79619>
- Matisoff, James A. (2015) *The Sino-Tibetan etymological dictionary and thesaurus*. The Regents of the University of California.
- Price, Max and Hitomi Hongo (2020) The archaeology of pig domestication: Methods, models, and case studies. *Journal of Archaeological Research* 28: 557–615. doi: <https://doi.org/10.1007/s10814-019-09142-9>
- Qin, Liying and Hiroyuki Suzuki (2016) Chasing a cat from the Mekong to the Salween: A geolinguistic description of ‘cat’ in Trung and Khams Tibetan in North-western Yunnan. *Studies in Asian Geolinguistics* I: 61–71. Online: [https://publication.aa-ken.jp/sag1\\_sun\\_2016.pdf](https://publication.aa-ken.jp/sag1_sun_2016.pdf)
- Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill, and Johann-Mattis 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: <https://doi.org/10.1073/pnas.1817972116>
- Schuessler, Axel (2007) *ABC etymological dictionary of Old Chinese*. Honolulu: University of Hawai‘i Press. URI: <https://www.jstor.org/stable/j.ctt6wr0n4>
- Shen, Xinghua [沈兴华] (2005) *Huanghe sanjiaozhou fangyan yanjiu* 《黄河三角洲方言研究》 [A study on dialects in the estuary of Yellow River]. Jinan: Qilu Shushe.
- Shirai, Satoko (this volume) Geolinguistics and Sino-Tibetan: An introduction. doi: <https://doi.org/10.5281/zenodo.8437101>
- Suzuki, Fumiki (2023) ‘Wheat’ in Sinitic. *LAAA-II*, 11–12.
- Suzuki, Hiroyuki (2019) How Tibetans classify pigs in their languages in the eastern Tibetosphere: Revisiting the pig issue through geolinguistics. In Hiroyuki Suzuki, Keita Kurabe, and Mitsuaki Endo (eds) *Papers from the Workshop “Phylogeny, Migration, and Contact of East and Southeast Asian Languages and Human Groups”*, 40–53. Fuchu: Research Institute for Languages and Cultures of Asia and Africa. URI: [https://publication.aa-ken.jp/sag\\_mono7\\_phylogeny\\_dispersion\\_contact\\_2019.pdf](https://publication.aa-ken.jp/sag_mono7_phylogeny_dispersion_contact_2019.pdf)
- Suzuki, Hiroyuki (2022) *Geolinguistics in the eastern Tibetosphere: An introduction*. Tokyo: Geolinguistic Society of Japan. doi: <https://doi.org/10.5281/zenodo.5989176>
- Suzuki, Hiroyuki, Keita Kurabe, Kazue Iwasa, Satoko Shirai, and Shiho Ebihara (2021) ‘Rice plant’ in Tibeto-Burman. *LAA*, 75–79.
- Suzuki, Hiroyuki, Keita Kurabe, Kazue Iwasa, , Satoko Shirai, Shiho Ebihara, and Ikuko Matsuse (2016) A geolinguistic analysis of the ‘rice’ category in Tibeto-Burman. *Studies in Asian Geolinguistics* 2: 37–51. URI: [https://publication.aa-ken.jp/sag2\\_rice\\_2016.pdf](https://publication.aa-ken.jp/sag2_rice_2016.pdf)
- Suzuki, Hiroyuki, Shiho Ebihara, Kazue Iwasa, Keita Kurabe, and Satoko Shirai (2023a) ‘Wheat’ in Tibeto-Burman. *LAAA-II*, 15–18.
- Suzuki, Hiroyuki, Shiho Ebihara, Kazue Iwasa, Keita Kurabe, and Satoko Shirai (2023b) ‘Foxtail millet’ in Tibeto-Burman. *LAAA-II*, 79–80.
- Suzuki, Hiroyuki, Shiho Ebihara, Kazue Iwasa, Keita Kurabe, and Satoko Shirai (2023c) ‘Barnyard millet’ in Tibeto-Burman. *LAAA-II*, 99.
- Suzuki, Hiroyuki and Sonam Wangmo (2016) Cultural contexts of the expansion of a Tibetan word ‘bras’ ‘rice’ in the easternmost Tibetosphere. In Mitsuaki Endo (ed.) *Papers from the Third International Conference on Asian Geolinguistics*, 72–79. Fuchu: Research Institute for Languages and Cultures of Asian and Africa. URI: [https://publication.aa-ken.jp/papers\\_3IC\\_Asian\\_geolinguistics\\_2016.pdf](https://publication.aa-ken.jp/papers_3IC_Asian_geolinguistics_2016.pdf)

Yagi, Kenji (2022a) ‘Horse’ in Sinitic. *LAAA-I*, 134.

Yagi, Kenji (2022b) ‘Dog’ in Sinitic. *LAAA-I*, 175–176.

Yagi, Kenji and Takashi Ueya (2021) ‘Rice plant’ in Sinitic. *LAA*, 68–69.

Zhang, Hanzhi, Ting Ji, Mark Pagel, and Ruth Mace (2020) Dated phylogeny suggests early Neolithic origin of Sino-Tibetan languages. *Scientific Reports* 10(20792): 1–8. doi: <https://doi.org/10.1038/s41598-020-77404-4>

Zhang, Menghan, Shi Yan, Wuyun Pan, and Li Jin (2019) Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569(7754): 112–115. doi: <https://doi.org/10.1038/s41586-019-1153-z>

## Publication history

Date received: 21 August 2023

Research article (Featured theme: Geolinguistic approach to Sino-Tibetan)

## Current issues and perspectives on Yi characters: From the past and into the future

IWASA, Kazue  
Nagoya University of Foreign Studies

**Abstract:** This paper aims first to show current issues in the study of the Yi characters, then to demonstrate the high affinity between geolinguistics and the study of the Yi characters as well as the efficaciousness of a geolinguistic approach when many issues can be solved by mapping the Yi characters.

Yiyu (彝語) is the language of the Yi ethnic group living in the southwestern part of China, and the northern parts of Vietnam and Laos. Yi people possess their own script and to date have written a great number of manuscripts so far. Due to the huge dialectal differences, and various changes throughout history, there has come to be an enormous disparity among the Yi characters in their shape, phonetic values and meaning from region to region. As a result, these characters are mutually almost incomprehensible among the dialects.

Therefore, in the Yi-character study, there exist such issues as clarifying the change patterns in the Yi script and detecting possible routes of its spread, and these puzzles have attracted the interest of many scholars. The author hence applied the geolinguistic approach to tackling them by mapping as many Yi characters of all regions as possible. As a result of these efforts, certain features and tendencies have been found in the Yi-character change patterns, and the chronological order in several graphemes.

However, at the same time, this attempt also suggests the necessity of analysing the Yi characters in a much narrower area to elucidate details of their change process and the directions of spread more precisely and accurately. Accordingly, in this paper, the Yi characters written in *Hua-Yi Yiyu (Chinese-Foreign Glossaries)* are geolinguistically analysed, because the places of the data sources in these glossaries are located in a relatively narrow area where Sichuan, Guizhou and Yunnan intersect. This study is preliminary but effective in solving some of the current issues of the Yi-character study.\*

**Keywords:** the Yi script; Yi-character maps; geolinguistic approach; *Hua-Yi Yiyu*

---

IWASA, Kazue. 2023. Current issues and perspectives on Yi characters: From the past and into the future. *Studies in Geolinguistics* 3: 180–198. doi: <https://doi.org/10.5281/zenodo.8437127>

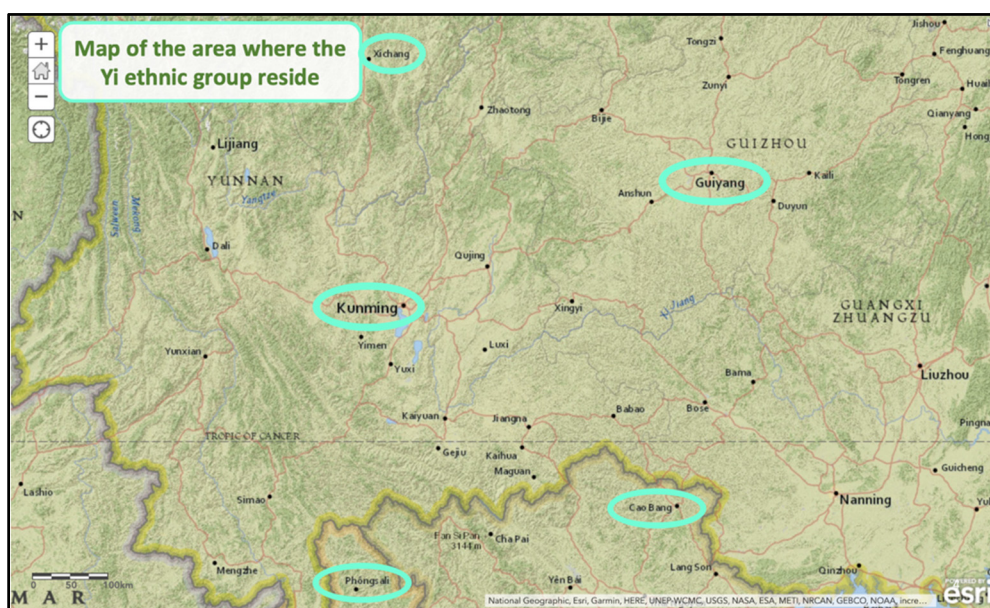
\* Here I express my deep gratitude to Professor Mitsuaki Endo, who encouraged me to explore the study of geolinguistics and gave me precious information about it. I also thank Mr Hiroyuki Suzuki, who gave me many useful pieces of advice.

## 1. Introduction

First, I will provide you with a brief introduction to Yiyu (彝語), the language of the Yi ethnic group, and the Yi script.

### 1.1. Yiyu and the Yi script

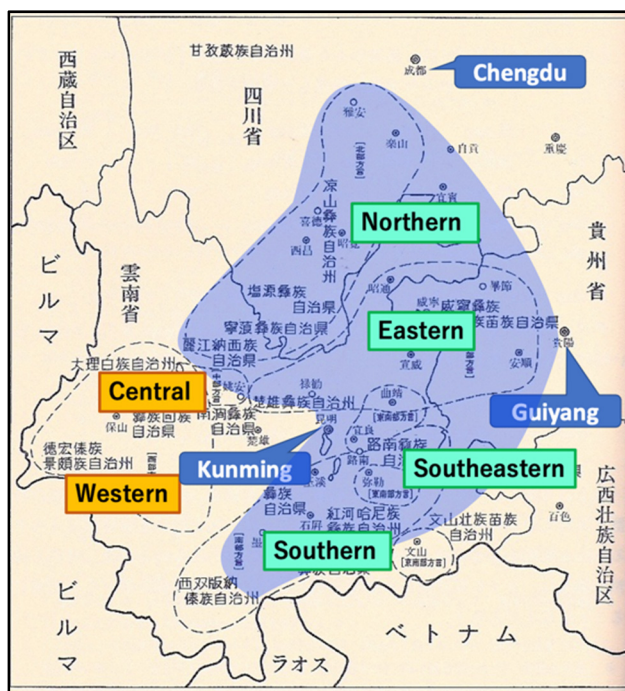
Yiyu is the language of the Yi ethnic group. It belongs to the Lolo-Burmese language group of the Tibeto-Burman language family. It is spoken in the southwestern part of China, in areas included in the provinces of Sichuan, Guizhou, Yunnan, and Guangxi, as well as northern areas of Vietnam and Laos. Map 1 shows approximately the entire area where the Yi ethnic group live.



Map 1 The area where the Yi ethnic group reside.

According to the official classification of languages in China, there are six Yi dialects: Northern, Eastern, Southern, Southeastern, Western, and Central. Four of them – i.e., the Northern, Eastern, Southern, and Southeastern dialects – possess the Yi script and manuscripts written in it. No script has so far been reported as being used among the Yi people in Vietnam and Laos.





Map 2 Distribution of the six Yi dialects in China and the approximate area with the Yi script<sup>1</sup>.

Map 2 indicates the distribution of the Yi dialects in China. The area shaded in blue demonstrates the dialects possessing the Yi script and manuscripts.

Yi characters are basically all syllabic. However, they probably used to be logographic. It is said that Yi characters were created around the time of the Yuan or Ming dynasties.

*Bimos*, religious leaders among the Yi people, have dealt with the script and manuscripts exclusively within their clan throughout history, basically from father to son. Yi characters have been used within *Bimos* to keep secrets like religious scriptures. Therefore, from the point of view of the *Bimos*, the characters were not a tool for communication but were used instead for the prevention of any leakage of their clan's heritage. In such circumstances, the *Bimos* quite freely used a homophonous character or a character with very similar phonetic value, and they often ignored such differences in tones or voicedness in consonants in their writing. They also sometimes freely added certain alternations to characters as long as they could understand what was written. In other words, the Yi characters can be said to be very personal.

<sup>1</sup> The original map was retrieved from Nishida (1979: 182).

As a result of such high individuation, exclusiveness, broad spread of the script, and possible lexical changes, the characters in each area or village came to show great diversity in their shapes, phonetic values, and meaning.

### 1.2. Current issues in the study of Yi characters

There are several important current issues in the study of the Yi characters. As mentioned above, the Yi characters show great diversification in shapes and phonetic values (presented in Chart 1), due to the regional differences or processes or degrees of phonological or semantic changes.

Consequently, it is not easy to determine the chronological order among the dialects and to trace the route of their propagation. Also, elucidating the patterns of changes in the Yi characters is a knotty problem.

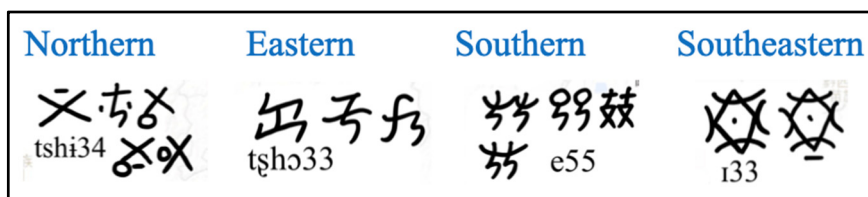


Chart 1 The Yi characters meaning ‘this’ in four dialects.

### 1.3. Compatibility between geolinguistics and Yi-character study

As explained above, the critical issues in the Yi-character study begin with the history of Yi characters, as follows:

- Possible route(s) by which Yi characters would have spread
- Patterns in which Yi characters have changed

Contemplating these issues, I struck upon a novel idea that drawing Yi-character maps and analysing them geolinguistically might help to solve these hard issues. Here is the reason why I chose this path.

These issues are related to the history of the Yi characters. Then, geolinguistics is one of the methods in clarifying the history of a language. Geolinguistics aims at revealing the direction of language changes<sup>2</sup>. Geolinguistic principles and aims seem to match very well with sorting out the current issues in Yi characters.

<sup>2</sup> Sibata (1969: 11) mentions as follows: 「……言語地理学は言語史の方法の一つである。したがって、言語地理学の目的は言語の歴史を明らかにすることにある。……」 (“Geolinguistics is one of the methods in studying language history. Therefore, geolinguistics aims at elucidating the history of a language...”: translated by the author.)



## 2. Geolinguistic approach to Yi-character study

In this section, I will trace how the geolinguistic approach can be applied to Yi-character study and what it has revealed to us so far.

### 2.1. Main research conducted by the author

One of the studies of Yi characters in which the geolinguistic approach was applied is Iwasa (2018a).

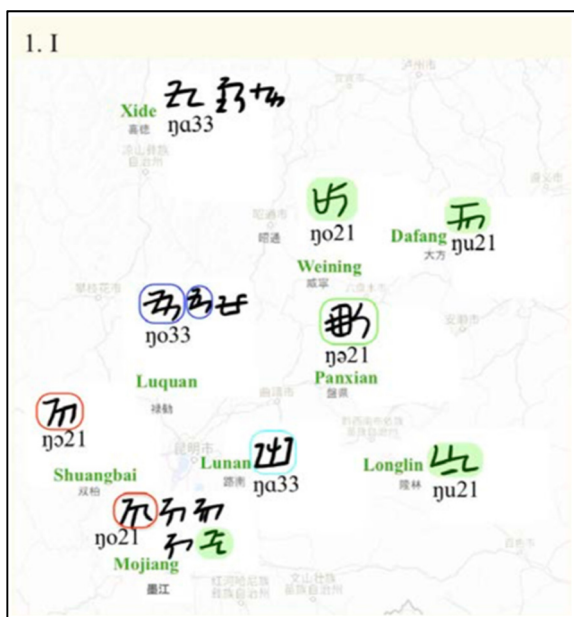
First, a map was drawn by picking up from all resources I have with as many Yi characters as possible. See Map 3 as an example.



Map 3 Yi-character map for the meaning of '1'.

Second, I pick up a key character, sometimes a few key characters. These key characters are determined mainly from the perspective of the character shape, as the logographic nature of the Yi characters is respected.

Third, I observe the whole distribution of the characters and attempt to find how and why the characters show such a spread. At the same time, I also search if there is a substitution of homonyms or any other features. (See Map 4 below, for example.)



Map 4 A result of geolinguistic analysis of Map 3 (Iwasa, 2018a: 15).

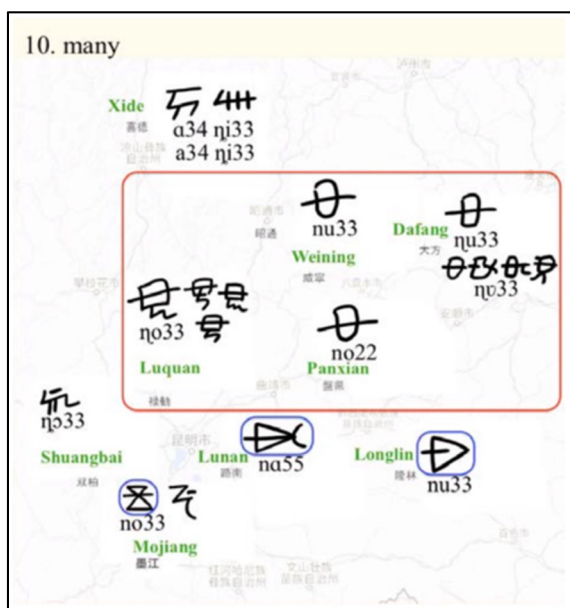
In Map 4, the characters regarded as belonging to the same group are indicated by the same colour, namely red, blue, or green. Several research results will be presented in the next section.

## 2.2. What the study has yielded to us so far

In this section, I will introduce three main research results (chiefly from Iwasa, 2018a).

### 2.2.1. Substitution of homonyms

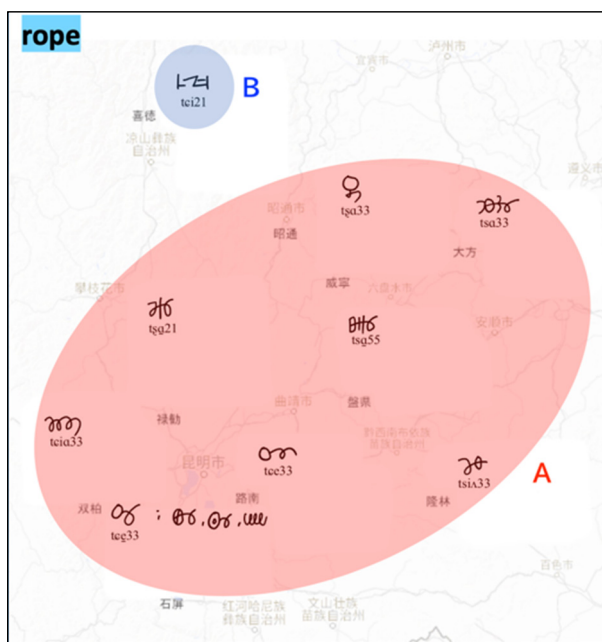
Here is an example of the substitution of homonyms, or phonetic loans. In Map 5, you can see the characters in the red frame and those in the blue circles. Both show formal and phonetic resemblances. These characters originally meant ‘arrow’, but it seems that they began to be used for expressing ‘many’ because of their phonetic similarity. (See Iwasa, 2018a for more details.)



Map 5 Yi-character map for the meaning of 'many' (Iwasa, 2018a: 10).

### 2.2.2. AB distribution

There are several maps showing AB distribution. One of them is shown below, Map 6. In this map, some characters in the red shade (group A) can be in the same continuum, whereas the character in the blue shade (group B) might also be included. It is very hard to determine which group (A or B) is older just judging from the map. In this case, more investigation into the continuum is needed to figure out their feasible metamorphoses.



Map 6 Yi-character map for the meaning of ‘rope’ (Iwasa, 2018a: 62)<sup>3</sup>.

### 2.2.3. ABA distribution

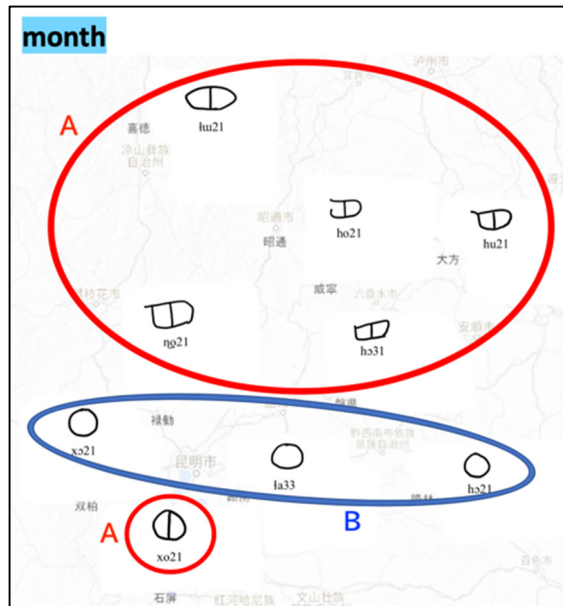
Next, we have an example of ABA distribution. In Map 7, you can see two red circles. Let’s call them A. You can also see a blue circle, which we can call B.

ABA distribution is thought to imply that the characters in A could be anticipated to be relatively older than the characters in the blue circle, B area, although it is still difficult to affirm it immediately. What we can find from this map is at least the probability that in the B area a new change could have happened, because it would be improbable that the same innovation could coincidentally happen in two remote locations, namely the unconnected A areas.

<sup>3</sup> Some amendments have been made to the original map for clarity by the author.



Map 7 Yi-character map for the meaning of ‘to stand’ (Iwasa, 2018a: 49)<sup>4</sup>.

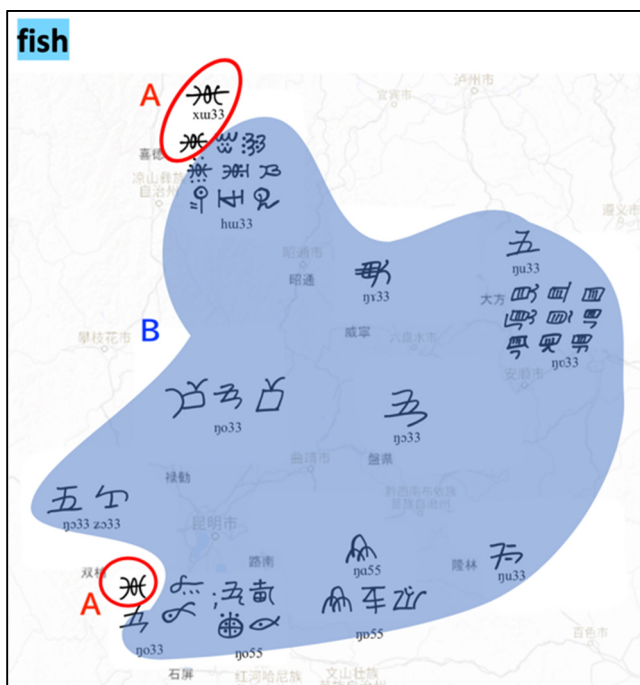


Map 8 Yi-character map for the meaning of ‘month’<sup>5</sup>.

<sup>4</sup> Some amendments have been made to the original map for clarity by the author.

<sup>5</sup> This map was drawn for this paper by the author.

Looking at Map 8, in the B area simplification seems to have happened, although the characters in both A and B areas show close similarity.



Map 9 Yi-character map for the meaning of ‘fish’ (Iwasa, 2018a: 24)<sup>6</sup>.

In Map 9, although area B is heterogeneous and contains various types of characters, it may indicate in this area many innovative changes could have happened. On the other hand, in both areas A, identical characters are found and intervened by area B, while these two areas are very distant from each other.

### 3. What the geolinguistic approach to Yi-character study can show us and what it indicates about future research needs

In the previous section, I demonstrated several findings facilitated by the geolinguistic approach to the important issues in the Yi-character study. This section will present the potential of geolinguistic analysis and future tasks in this domain.

<sup>6</sup> Some amendments have been made to the original map for clarity by the author.

### 3.1. Urgent tasks in the Yi-character study

My previous studies on the Yi characters have proven that a geolinguistic approach is effective and practical in clarifying their change patterns and possible routes of their spread.

However, to elucidate more accurately the change patterns in the character shapes and possible spread routes of the script, it is urgently necessary to analyse the data of the Yi characters in a much narrower area. For this purpose, studying and mapping the data in *Hua-Yi Yiyu* (華夷譯語) is a very desirable attempt<sup>7</sup>.

*Hua-Yi Yiyu* is a series of Chinese–Foreign Glossary books compiled during the Ming and Qing Dynasties and contains five books on the vocabularies of distinctive Yi dialects and scripts. These five dialects are thought to have been located in a relatively narrow area where Sichuan, Guizhou, and Yunnan intersect. The locations of three of them are identified, whereas the rest are still open to question.

*Hua-Yi Yiyu* also gives us the phonetic information of the Yi characters by the phonetic transcriptions written in Chinese characters. It helps to detect the two unknown places mentioned above and to observe phonological changes that happened in all locations throughout history, even though it contains many discrepancies between the Yi characters and their phonetic transcriptions<sup>8</sup>.

In the next section, I will introduce my current geolinguistic analysis of the Yi characters in *Hua-Yi Yiyu*.

### 3.2. Current challenge: Geolinguistic exploration of the Yi characters in *Hua-Yi Yiyu*

Here I will show several results from my preliminary geolinguistic exploration into the Yi characters written in *Hua-Yi Yiyu*.

#### 3.2.1. Mapping the Yi characters written in *Hua-Yi Yiyu* and comparing them with the existing Yi-character maps

The materials used for mapping the Yi characters from *Hua-Yi Yiyu* are as follows<sup>9</sup>:

- ① *Gugong Book, the third version, Yongning shu Shuiliao Luoluo Yiyu* (故宮本第三種 永寧屬水潦猺獮譯語): The basic data are based on Nishida (1979).
- ② *Gugong Book, the second version, Jianchang shu Shamaliangshan ge Luoluo Yiyu* (故宮本第二種 建昌屬罵梁山各猺獮譯語): The basic data are based on Matsukawa and Nie (2015).

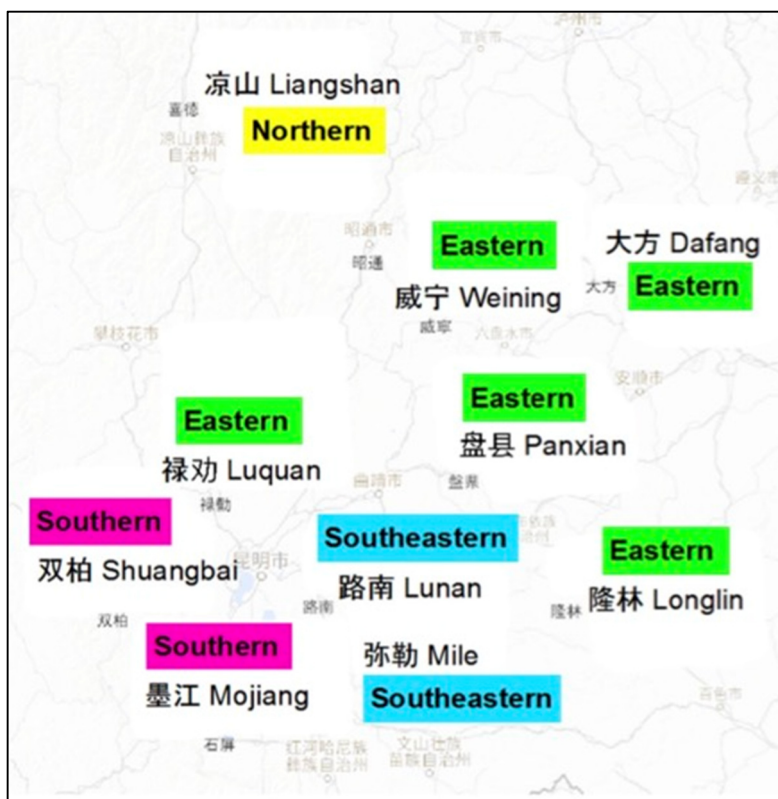
<sup>7</sup> Based on the present study, the author has developed the data and findings then written a paper (Iwasa 2023) with the entire data on the Yi languages recorded in *Hua-Yi Yiyu*.

<sup>8</sup> These errors can have several causes. I will mention them in detail on another occasion due to the limited space available in this paper.

<sup>9</sup> When I started this research, the whole reproduction of the five books of *Luoluo Yiyu* was not available. That is why I referred only to these three data sources. In my paper, Iwasa (2023), all five versions are analysed.

- ③ *Gugong Book, the first version, Dongchuan-fu shu Luoluo-Yiyu* (故宫本第一種 東川府屬猥獮譯語): Information provided as a courtesy.

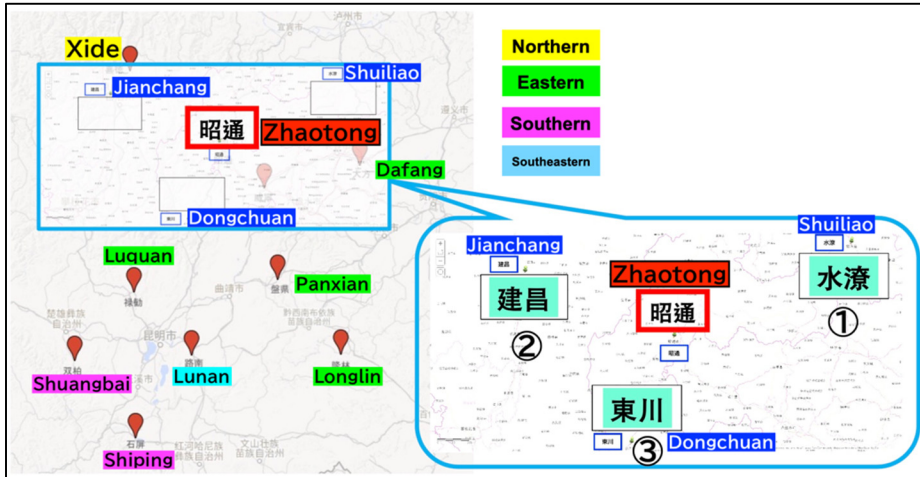
I have collected the data of the regions indicated in the map below and drawn the Yi-character maps prior to this exploration. The main data sources for these existing maps are Yi-Chinese dictionaries. See the References for more details.



Map 10 The locations of the existing Yi-character maps.

The locations of the data from *Hua-Yi Yiyu* are shown below in Map 11. The numbers on Map 11 correspond to which of the versions of *Hua-Yi Yiyu*, ①, ② or ③, mentioned earlier at the beginning of this section, was the source. These locations on Map 11 are illustrated on the right side of Map 12 to 16.

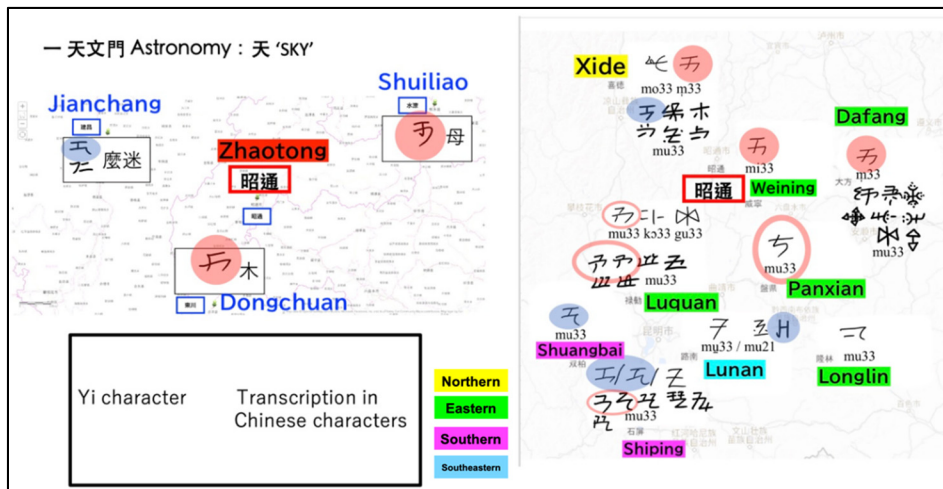




Map 11 The locations of the data from *Hua-Yi Yiyu*.

### 3.2.1.1. ‘SKY’

Here are maps for ‘SKY (天)’. The left map shows the Yi characters from *Hua-Yi Yiyu*, while the right one shows the Yi characters in the whole area. The right map was drawn previously by the author. I have indicated each grapheme by using a distinctive colour – red and blue – that we can observe if the characters from *Hua-Yi Yiyu* correspond to those on the right map. We can also examine the phonetic values by comparing the pronunciations of the modern Yi dialects on the right map and the phonetic transcriptions in *Hua-Yi Yiyu* on the left – for example, ‘麼迷’, ‘木’, and ‘母’.



Map 12 The Yi characters indicating ‘SKY’.

The characters indicated by both red and blue seem to be in the same continuum, but focusing more minutely on their shape, they can be divided into two groups, namely, the red group and the blue group. On both maps, the characters shaded in red bear a strong resemblance to each other, while those in blue also do. Therefore, it can be said that the characters recorded in *Hua-Yi Yiyu* reflect the regional characteristics very well.

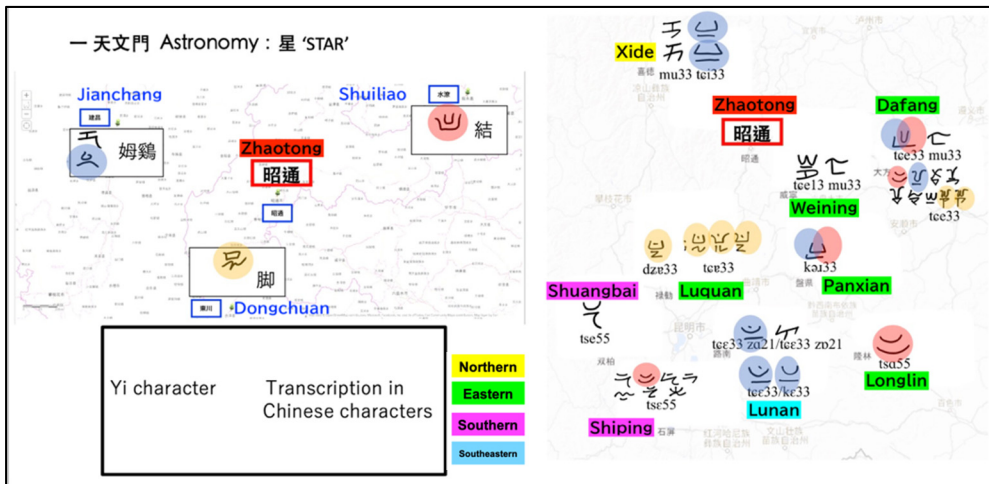
Additionally, the transcriptions of Chinese characters in *Hua-Yi Yiyu* also seem to correspond very well to the pronunciations in the modern dialects. For example, on the left map, the transcription ‘麽迷’ in Jianchang – a locale very near Xide on the right map – might be pronounced like [mɤ mi] in disregard of the tones, if its pinyin ‘me mi’ is taken into consideration. This phonetic value [mɤ mi] is similar enough to [mo<sup>33</sup> m<sup>33</sup>] in Xide on the right map.

### 3.2.1.2. ‘STAR’

As you can see on the left map of Map 13, there are three graphemes in *Hua-Yi Yiyu*. The Yi characters on the right map correspond well with them formally and geographically as well.

We also find that the graphemes shaded in red, and the ones in blue seem to have a strong relation to each other. Although they should be treated as allographs of one particular grapheme, they are regarded as distinctive because the shape of any parts and/or the number of strokes obviously appear to be different.

Then, the grapheme shaded in yellow may have been derived from the same grapheme as the rest. For example, this grapheme may have been simplified through time and changed towards the graphemes as shaded in red and blue. In particular, the characters shaded in both red and blue seem to show a certain state of transition phase from the yellow-shaded ones.



Map 13 The Yi characters indicating 'STAR'.

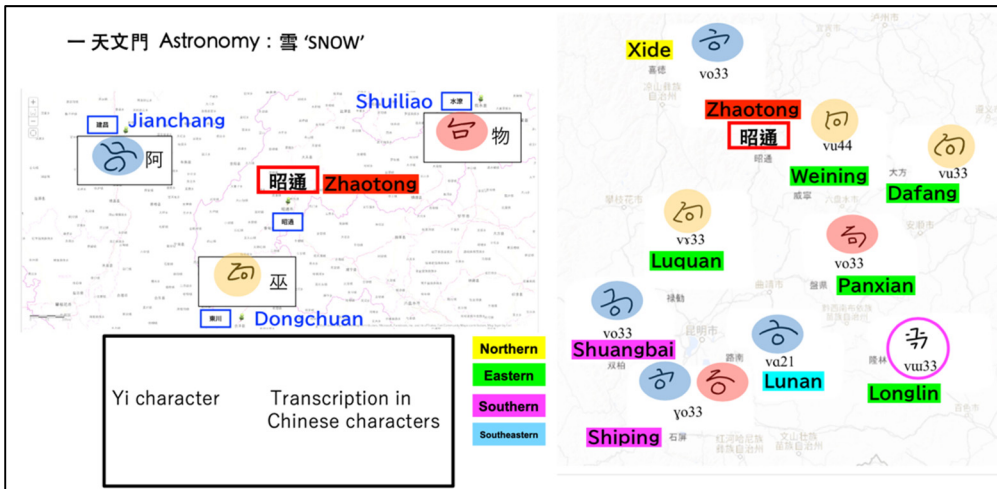
### 3.2.1.3. 'SNOW'

Map 14 gives us more concrete information about the change patterns of the Yi characters throughout history. All characters on both maps should probably be interrelated and illustrate their process of the formal transition. To clarify each phase of the transition, I will set three graphemes and observe their spread.

The character of Shuiliao on the left map, which is shaded in red, is similar to those of Panxian and Shiping on the right map, which are also marked by red shades.

Then, the characters shaded in yellow seem to intervene with the red ones. The ones in yellow shade may be more innovative than those in red.

The characters shaded in blue are found in the most peripheral areas, they hence might have once spread across a vast area.

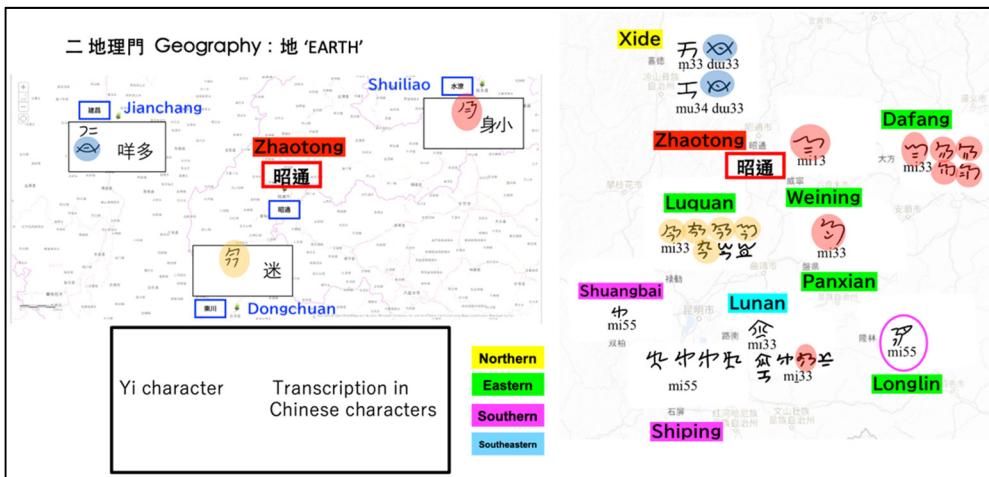


Map 14 The Yi characters indicating 'SNOW'.

### 3.2.1.4. 'EARTH'

On Map 15, each grapheme of the left map corresponds well to the distribution of the characters on the right map. Whereas the word forms around Jianchang are different from the others, the rest show a high resemblance to one another.

Except for the character in the pink circle around Longlin, the characters spread from Zhaotong to the southernmost area around Shiping, both shaded and unshaded, appear to be interrelated judging from the curious similarity in their forms and pronunciation.

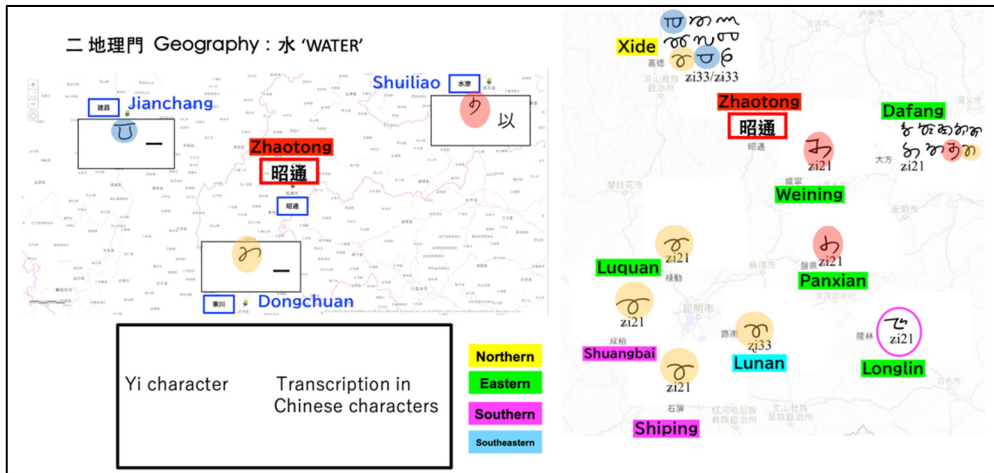


Map 15 The Yi characters indicating 'EARTH'.

### 3.2.1.5. ‘WATER’

Map 16 shows very well the correspondence of the distribution of each grapheme. The grapheme shaded in yellow is the most widely spread on the right map. The one shaded in red can be treated as an allograph of it.

Very interestingly, the character in the pink circle at Longlin demonstrates a very different form from the others. This needs close and thorough investigation.



Map 16 The Yi characters indicating ‘WATER’.

### 3.3. Final remarks and future perspectives in the Yi-character study

To sum up, from the maps, we have found that the Yi characters written in *Hua-Yi Yiyu* demonstrate a high level of resemblance to those on the existing Yi-character maps according to the region and that the Yi characters in Longlin may bear different features from those in other regions.

Hence, in the next phase of this study, we need to examine the process of the metamorphoses of the Yi characters in *Hua-Yi Yiyu* by comparing them to the existing Yi-character maps. We also need to analyse the transcriptions by Chinese characters and reconstruct the phonetic values of the Yi characters in *Hua-Yi Yiyu*. Needless to say, it is undoubtedly necessary to refer to all five versions of *Hua-Yi Yiyu* in that case. Because two of the five locations of *Hua-Yi Yiyu* are unknown, this study will contribute to locating them in space.

This paper demonstrates that a geolinguistic approach has proved to be efficacious for the historical study of the Yi characters. I hope that geolinguistic analysis of all the Yi characters written in *Hua-Yi Yiyu* will lead us to the elucidation of their change patterns and possible spread routes as soon as is feasible.

## References

- Huang, Jianming [黃建明] (1993) *Yizu Guji Wenxian Gaiyao* 《彝族古籍文獻概要》 [Outline of the ancient manuscripts of Yi people]. Kunming: Yunnan Minzu Chubanshe.
- Iwasa, Kazue (2017) Mapping the Yi characters – an attempt to explore a route of their historical change and origin –. *Studies in Asian Geolinguistics, Monograph series no.2, Proceedings of the Workshop “Geolinguistic Method and Southeast Asian Linguistics”*, 59–67. Fuchu: ILCAA, Tokyo University of Foreign Studies. [https://publication.aa-ken.jp/proceedings\\_geolinguistic\\_SouthEastAsian\\_2017.pdf](https://publication.aa-ken.jp/proceedings_geolinguistic_SouthEastAsian_2017.pdf)
- Iwasa, Kazue (2018a) *Remarks on Maps of the Yi Script Based on the Swadesh 100 Wordlist. Studies in Asian Geolinguistics, Monograph Series No. 5*. ILCAA 2015–2017 Joint Research Project Studies in Asian Geolinguistics Research Institute for Languages and Cultures of Asia and Africa Tokyo University of Foreign Studies. [https://publication.aa-ken.jp/sag\\_mono5\\_yi\\_2018.pdf](https://publication.aa-ken.jp/sag_mono5_yi_2018.pdf)
- Iwasa, Kazue (2018b) “Current studies and future perspectives on the Yi manuscripts preserved in Europe” at Conference on Silk Road Studies, held in Göttingen in 2018.
- Iwasa, Kazue (2019) Dili Yuyanxue Yuanli de Yingyong – Yi Yiwen Weili- 「地理語言學原理的應用 – 以彝文為例-」 [An application of geolinguistic principles - The Yi script as a case study -], *Collected Papers on Eastern Asian Geolinguistics. Studies in Asian Geolinguistics, Monograph Series* 《亞洲地理語言學研究》單刊 第六 東部亞洲地理語言學論文集 No. 6, 131–144. ILCAA 2015–2017 Joint Research Project Studies in Asian Geolinguistics Research Institute for Languages and Cultures of Asia and Africa Tokyo University of Foreign Studies. [https://publication.aaken.jp/sag\\_mono6\\_eastern\\_asian\\_2019.pdf](https://publication.aaken.jp/sag_mono6_eastern_asian_2019.pdf)
- Iwasa, Kazue (2023) Geolinguistic Analysis of the Yi Script written in *Hua-Yi Yiyu*. Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, Hiroyuki Suzuki, and Mitsuaki Endo (eds.) *Proceedings of the fifth International Conference on Asian Geolinguistics*, 172–202. doi: <https://doi.org/10.5281/zenodo.8374673>
- Matsukawa, Takashi [松川節] and Nie, Hongyin [聶鴻音] (2013) Ōtani Daigaku Shozō *Roro Yakugo Jutsuryaku* 「大谷大学所蔵『僮僮譯語』述略」 [The Summary of Lolo Vocabulary preserved at Otani University]. *Kaiyakugo (Seibankan Yakugo Yonshu Roro Yakugo Isshu Eiin to Kenkyu* 『『華夷譯語（西番館譯語四種 僮僮譯語一種）』影印と研究』 [The Photographic Reproduction and study on *Huayi Yiyu* (Four Tibetan Vocabularies and One Lolo Vocabulary)] (Ōtani Daigaku Bunken Kenkyū Sōsho 1), 23–29. Kyoto: Shokado.
- Nishida, Tatsuo (1979) *Roro Yakugo no Kenkyu: Rorogo no Kōzō to Keitō* 『僮僮譯語の研究：僮僮語の構造と系統』 [A study of the Lolo-Chinese vocabulary Lolo i-yu: The structure and lineage of Shui-liao Lolo]. (Ka-i Yakugo kenkyu sosho 6) Kyoto: Shokado.
- Sibata, Takeshi (1969) *The methods of linguistic geography (Gengo Chirigaku no Houhou)*. Tokyo: Chikuma Shobo.
- Wen, You [聞宥] (1940) Luoluo Yiyu Kao 「僮僮譯語攷」 [On Lo Lo I Yu—Lolo-Chinese Vocabularies]. *Studia Serica, Journal of The Chinese Cultural Studies Research Institute*, 77–97.
- [Dictionaries]
- Bai, Xianyun (ed) [白顯雲主編] (1995) *Yi-Han Zidian* 《彝漢字典》 [Yi-Chinese Character Dictionary]. Yunnan Minzu Chubanshe.
- Dianchuanqiangui Yiwen Xiezuozu (ed.) [滇川黔桂彝文協作組編] (2004) *Dianchuanqiangui Yiwenziji* 《滇川黔桂彝文字集》 [Collection of Yi characters of Yunnan, Sichuan, Guizhou, and Guangxi]. Guizhou Minzu Chubanshe. Sichuan Chuban Jituan. Sichuan Minzu Chubanshe. Yunnan Minzu Chubanshe.

- Li, Zenghua (ed.) [李增華主編] (2014) *Guyiwen Changyong Zidian* 《古彝文常用字典》 [*Dictionary of Commonly Used Old Yi Characters*]. Yunnan Minzu Chubanshe.
- Pu, Zhangkai (ed.) [普璋開主編] (2005) *Diannan Yiwen Zidian* 《滇南彝文字典》 [*Dictionary of Diannan Yi Characters*]. Yunnan Minzu Chubanshe.
- Wu, Zili [武自立], Ji Jiafa [紀嘉發] (2011) *Han Yi Jianming Cidian* 《漢彝簡明詞典》 [*Chinese-Yi Concise Dictionary*]. Sichuan Minzu Chubanshe.
- Yunnansheng Lunan Yizu Zizhixian Wenshi Yanjiushi (ed.) [雲南省路南彝族自治縣文史研究室編] (1984) *Yi Han Jianming Cidian* 《彝漢簡明詞典》 [*Yi-Chinese Concise Dictionary*]. Yunnan Minzu Chubanshe.
- Zhang, Jinzhi [張晉智] (2014) *Jianming Yi Han Zidian* 《簡明彝漢字典（祿勸版）》 [*Concise Yi-Chinese Dictionary (Luquan version)*]. Yunnan Minzu Chubanshe.
- Zhongyang Minyuan Yizu Lishi Wenxian Ban (ed.) [中央民院彝族歷史文獻班主編] (1984) *Dianchuanqiangui Yi Han Jiben Cihui Duizhao Cidian* 《滇川黔桂彝漢基本詞彙對照辭典》 [*Yi-Chinese Comparative Dictionary of Basic Vocabulary in Yunnan, Sichuan, Guizhou, and Guangxi*]. Zhongyang Minyuan Yizu Lishi Wenxian Bianyishi.

## Publication history

Date received: 6 September 2023

## 彙報 2022年度

### Memoranda and personalia 2022

第4回大会は東京外国語大学で開催準備していたが、引き続きコロナのためオンライン開催となった。一般発表は2件応募があり、いずれも採択した。招待講演としてオランダの Fryske Akademy の Wilbert Heeringa 氏の“Introduction to Dialectometry”があった。これについては開催校の多大の努力の末実ったものであり、深甚なる謝意を表したい。今回は参加費 1000 円を一般参加者から徴収し、学生・海外オンライン接続者は無料とした。

2022年9月には学会誌『地理言語学研究』第2号と Monograph Series No.3 として Hiroyuki SUZUKI, Mika FUKAZAWA, Akiko YOKOYAMA, and Mitsuaki ENDO eds. *Linguistic Atlas of Asia and Africa, Volume 1* を pdf 版公開し、翌月に紙版も Amazon より販売開始した。2022年10月4日に学会誌に投稿論文が到着し、編集委員会・世話人会で協議した結果、同日付で受理した扱いで第3号への掲載を決定した。

大会発表者と学会誌投稿者が非常に少なくなっているため、次年度からは6月初旬前後に大会を開催して、その口頭発表を夏休み中に論文化して学会誌に投稿できるよう締切を8月末に伸ばした。また、予稿集ではなく要旨集に改め、口頭発表の敷居を低くした。これからはまた大会発表と学会誌寄稿が増えるよう願っている。

世話人の改選時期にあたり、第三期は

総務	岩田 礼・白井聡子
集会	清水政明・鈴木史己
編集	斎藤純男・鈴木博之
ウェブ	倉部慶太・清宮貴雅
会計監査	木部暢子・中澤光平

の顔ぶれとなる。いずれも筆頭の者は 2023-24 年度の 2 年任期で、各パートのコーディネーターを務める。岩田礼氏は第一期に 2 年の任期を務め、第三期では代表相当の総務に 2 年任期で入る。後の者は 2023-26 年度の 4 年任期となる。2023~2024 年度の編集委員会は斎藤純男(コーディネーター)・鈴木博之・岩田礼・白井聡子からなる。

講座についてはコロナで開催実績がないまま 4 年間が経過してしまった。学会 HP に「本学会の趣旨に合う内容でしたら地理言語学プロパーに限らず関連分野も含めてどなたでも企画・実行できます。当面は細かな要項や企画申請時期は定めず、実績が増えていってから徐々に開催方法を固めていきます。企画



のある方は大会募集のアドレスに任意の書式の企画書を通例 1 ヶ月以上の余裕を持って随時お送りください。」とアナウンスすることとした。

また、モノグラフ募集要項の新版を HP に挙げた。修正事項は「2022 年度版」を削除し、末尾の「当該年度の」という文言を「本学会ホームページに出ている最新の」に修正したのみである。

学会 HP の末尾には連絡先として青山学院大学の研究室が表示してあり、学会連絡先表示が ISSN 取得の必須条件となっているが、次期代表は退職後のため研究室を表示することができず、2023-24 年度は便宜的にそのまま表示することとした。

2022 年度の会計報告は以下の通り（会計監査・世話人会確認済）：

収入の部	30,408 円	2021 年度繰越金
	12,000 円	2022 年度大会参加費 1000 円×12 名
計	42,408 円	
支出の部	5,568 円	学会 HP サーバー使用料
	5,478 円	学会誌紙版公開手数料（MyISBN/デザインエッグ）
	31,362 円	次年度繰越金
計	42,408 円	