

Received 24 January 2020.

Accepted 20 October 2020.

DOI: 10.1344/Dialectologia2021.27.5

**MULTIVARIATE ANALYSIS OF GEOGRAPHY
AND AGE IN DIALECT VOCABULARY
— COMPREHENSIVE ANALYSIS OF 250 YEARS
OF LANGUAGE CHANGE —**

Fumio INOUE¹ & Yasushi HANZAWA²

Tokyo University of Foreign Studies¹ / Fukushima University² *

innowayf@nifty.com / yhanzawa@educ.fukushima-u.ac.jp

Abstract

In this paper, we analyze long-term changes in dialect vocabulary. By analyzing the age difference of 140 years and the regional difference of 27 survey points, we will examine historical change and geographical spread over the 250 years since the compilation of a dialect glossary. First, we present the distribution of eight representative words according to the traditional method of linguistic geography. After that, we present comprehensive maps and graphs of the entire 406 words in order to grasp the general pattern of regional and age differences. Furthermore, multivariate analysis (MCA) is applied and the meanings of the axes produced by it are analyzed for the dimension of person (respondent) and the dimension of word (vocabulary). In the dimension of person, the magnitude of the age difference was confirmed in the data as a whole. Before WWII there was a large regional difference, but now the age difference has become larger. It was possible to demonstrate the double concentric distribution, that is, a regional concentric distribution under the national concentric distribution. In the dimension of word, the process of standardization and the regression of antiquated forms were confirmed. From pre-modern times to the present, most of the words show a process of disappearance, but some dialectal forms continue to

* Tokyo University of Foreign Studies 3-11-1, Asahicho, Fuchu, Tokyo 183-8534 Japan; Fukushima University, 1, Kanayagawa, Fukushima-city 960-1296 Japan.

increase even today. The words that match the colloquial and slang in Tokyo are expected to survive. In conclusion, this study will contribute to the expansion and development of dialectometry.

Keywords

regional difference, age difference, standardization, multivariate analysis, multiple correspondence analysis, dialect vocabulary

ANÁLISIS MULTIVARIANTE DE GEOGRAFÍA Y EDAD EN EL LÉXICO DIALECTAL — ANÁLISIS INTEGRAL OF 250 AÑOS DE CAMBIO LINGÜÍSTICO —

Resumen

En este artículo, analizamos los cambios producidos a largo plazo en el léxico dialectal. Al analizar la diferencia de edad de 140 años y la diferencia regional de 27 puntos de la encuesta, examinamos el cambio histórico y la distribución geográfica durante los 250 años que han transcurrido desde la compilación de un glosario de dialectos. En primer lugar, presentamos la distribución de ocho palabras representativas según el método tradicional de la geografía lingüística. A continuación, presentamos mapas y gráficos completos de un total de 406 palabras para comprender el patrón general de diferencias regionales y de edad. Además, se aplica el análisis multivariante (MCA) y se analizan los significados de los ejes producidos por él para la dimensión de la persona (encuestado) y la dimensión de la palabra (léxico). En la dimensión de la persona, el valor de la diferencia de edad se ha confirmado en el conjunto de los datos. Antes de la Segunda Guerra Mundial había una gran diferencia regional, pero ahora la diferencia de edad se ha incrementado. Se ha podido demostrar la distribución concéntrica doble, es decir, una distribución concéntrica regional bajo la distribución concéntrica nacional. En la dimensión de la palabra, se ha confirmado el proceso de estandarización y de regresión de formas anticuadas. Desde la época premoderna hasta la actualidad, la mayoría de las palabras muestran un proceso de desaparición, pero algunas formas dialectales continúan aumentando incluso en la actualidad. Se espera que sobrevivan las palabras que coinciden con la variedad coloquial y el argot de Tokio. En resumen, este estudio contribuye a la expansión y desarrollo de la dialectometría.

Palabras clave

diferencia regional, diferencia de edad, estandarización, análisis multivariante, análisis de correspondencia múltiple, léxico dialectal

1. Research history of dialect vocabulary

1.0 Regional and age differences in modern Japan

In this study, we try to comprehensively grasp the dialect situation based on a dialect glossary compiled 250 years ago. By dividing the data into the dimension of person (geography and age) and the dimension of word, we were able to grasp the dynamics of dialects over 250 years. We made comprehensive maps and summary graphs, and compared them with individual dialect maps to grasp the entire distribution of all 406 words in the Shonai district, Japan. The age difference was large in the past, and now the regional difference has become larger. Renovation in the central cities and relic forms in the surrounding area were observed. In the following, cross-references will be used for easier understanding.

1.1 Research history of language change

Research on language change in progress has become popular since the proposal of the variation theory of sociolinguistics. It was pointed out that not only age difference (*apparent time*) but also difference in survey time (*real time*) is important (Labov 1972, Chambers & Trudgill 1980), and memory of the past (*memory time*) can be utilized for investigating language usage (Long 2007, Inoue 2017.5). Follow-up studies of past surveys have been widely executed (Chambers et al. 2002, Sankoff 2006). Decline in vocabulary has always occurred in history throughout the world, and the accumulation of research supporting this is enormous. They show the trend of convergence, while the trend of divergence (*new dialect*), the opposite concept, is hard to ascertain through research (Inoue 1986.6, 1993.2, 2010.7, 2010.12, 2016.8).

Since large-scale surveys were conducted in Japan shortly after the war, various real-time research results were produced in the form of follow-up surveys. Research that takes into account time differences and regional differences is also extensive. There are regional-scale studies based on repeated surveys of linguistic geography, and national-

scale studies based on LAJ (*Linguistic Atlas of Japan*) (Onishi ed. 2017). Regional differences are still noticeable among the elderly in the 21st century. However, according to a nationwide survey of junior high school students in the late 20th century, most of the regional differences have already diminished, and nearly 90% of the young respondents have the ability to use standard language (Inoue 2011.1, Yarimizu 2017).

The retreat of dialects and the reduction of regional differences have been mainly examined in dialect maps and age-based graphs. The *glottogram*, which shows geography and age in one graph, was developed in Japan and applied to various geographic areas (Inoue 2017.8, 2019.12, Tsuzome 2012), and is spreading to Asia (Li 2014, Ang 2019, Huang 2019). See Inoue (2001.2), Yarimizu (2007, 2017), and Hanzawa (2018) for the history of research using glottograms and computational dialectology. Hanzawa (2017, 2020) repeated glottogram surveys in the same geographical area, and confirmed not only the later younger generations adopting new dialect, but also the middle age generation (who were the younger generation in the former survey) adopting the new dialect later in life. It is a lifelong change, and it demonstrates *late adoption* of new dialect by combining apparent time and real time.

According to the result of application of the multivariate analysis to the glottogram survey, age difference and regional difference appeared as two dimensions. In a certain size (distance), regional differences work larger than age differences (Inoue 2001.2, Yamashita & Hanzawa 2010). However, since the end of the 20th century, the standard language has spread among the younger generation, and the age difference has become remarkable as a whole.

The *glottogram* is an effective technique to see the competition of regional differences and age differences. The limit of the glottogram was that it is usually applied to linear regions. However, in order to overcome this limit, there was an attempt at a three-dimensional glottogram (Inoue 2003.7). There was also an attempt to show two parallel glottograms side by side (Honda 2005). There is also a technique to display a large number of complicated glottograms vertically and horizontally (Tsuzome 2012, Inoue 2017.8, Hanzawa 2020). An attempt to create a 3D graph by combining the 2D geographical surface and age axis was also presented (Huang 2019), and the technique of rotating the Figure to make it look three-dimensional was applied (Hidaka 2019). The

present paper is another attempt to overcome the weakness of the glottogram by applying it to the two-dimensional surface of the Shonai district and by representing a three-dimensional *age pillar map* to add the age difference.

1.2 *From individual word history to comprehensive study*

In the study of traditional linguistic vocabulary, researchers were interested in individual phenomena. Attempts to apply quantitative methods to analyze large amounts of data appeared at the last quarter of the 20th century (Heeringa 2004, Heeringa & Nerbonne 2001, Jeszenszky et al. 2019). These are called *dialectometry*. The vocabulary as an aggregate was processed (Goebel 2010, Nerbonne 2009, Nerbonne et al. 2005, Yarimizu et al. 2004, Aurrekoetxea et al. 2017). This trend has increased the momentum of research aimed at regularity, lawfulness, generalization and universality. This paper is one stage of an attempt to further develop the quantitative dialectical research.

The study of vocabulary (words) is often executed individually. The expression “Every word has its own history” is famous in dialectology. This expression does not belong to Jules Gilliéron, the founder of linguistic geography. According to Bloomfield (1933: 328, 520, 533), the source is Karl Jaberg (1908) *Sprachgeographie*. Presumably it also has its own source in “Jedes Wort hat seine Geschichte” at the beginning of the preface of Jakob Grimm (1819) *Deutsche Grammatik I*.

As a preceding study of this paper. Matsuda (1978) investigated remnant rates among old and young generations of glossaries compiled before the modern age in the Kyushu area, and ascertained the decline of many dialectal words. We can compare 3 time points, one being the time of compilation. The remnant rate of dialect vocabulary seemed to be decreasing at a constant rate. The summary graph is shown in Inoue (1985.2).



Figure 1. Survey area: Shonai District

1.3 Survey method of “Hamaogi” data

The starting point of the data is “Hamaogi”, a dialect glossary of Tsuruoka compiled in 1767 (transliterated in Mitsuya 1932). Regarding the remaining vocabulary of “Hamaogi”, the first survey was conducted in 1950 for 3 generations (Kokuritsu Kokugo Kenkyujo 1953). The second survey was conducted on four generations in 2018. There are 27 survey points. Figure 1 shows a map of the survey area of the Shonai district. The mark ● is for localities near railway stations, ▲ for localities in eastern hilly mountains, and ■ for other localities. All the dialect distribution maps were published as DASH (Inoue and Hanzawa 2019).

The following “Hamaogi” data is for 7 generations each about 20 years apart for a maximum difference of 140 years by birthyear. This data enables us to continuously study vocabulary changes over a long period of time. In this paper, we deal with all the words of all members and apply multivariate analysis to clarify the internal structure of the whole data.

1.4 Results of MCA: From regional to age differences

This study started from a research question of which is larger, the regional difference within one district or the 140-year age difference. The geographical size of the Shonai district of Yamagata Prefecture in the Tohoku region of Japan is 80km north-south, 50km east-west, and the birthyear difference of respondents is 140 years. In conclusion, age is the largest factor within the background (non-linguistic determinants) of standard language use (Inoue & Hanzawa 2021.2). The regional differences have diminished and changes have rapidly progressed. The pre-war generation was dominated by regional differences, but in the post-war era, standard language rapidly infiltrated and the age gap increased. The ultimate solution was obtained by applying multivariate analysis to the whole data. As a result of MCA (Multiple Corresponding Analysis) of all the data, the age difference was confirmed to be larger, appearing in axis 1. Then regional differences appeared as axis 2. On axis 2, the remnant rates along the north and south ends and the eastern mountains are conspicuous. In other words, a regional concentric distribution was captured. It was found that there is a large regional difference in the north-south direction on axes 3 and 4. The results of the multivariate analysis teach us much more than this. We tried to let the data talk, and grasped the overall picture of the whole results.

1.5 Significance of this research

This paper has the following features compared to conventional research.

1. The long-term age difference of 140 years was dealt with within the geographical distribution of one district.

2. Since we dealt with the words found in a dialect glossary from about 250 years ago, we can continuously observe the transformation of the dialect from pre-modern times.

3. Multivariate analysis was applied to comprehensively analyze all words of all respondents. Previous studies have succeeded in analyzing regional differences as aggregates. However, the results were rarely applied to the explanation of maps of individual dialect (forms).

As a result, the following phenomena were observed.

4. By comparing the results of the multivariate analysis with individual dialect maps, many Figures were properly positioned with each other. It turned out that dialect forms that match slang and colloquial language in Tokyo can survive.

5. The apparent time of age difference does not straightforwardly reflect language changes. There is *late adoption* which looks similar to *age grading* wherein new words are adopted after adulthood (Inoue 2013.4, Inoue and Yamashita 2014).

6. As a result of applying multivariate analysis, it was found that the age difference of 140 years worked greatly. People who were born during WWII are the boundaries of a kind of language shift. Before WWII there was a large regional difference, but the age difference has become larger at present.

7. There is a large north-south difference in terms of regional differences, and many words are retained at both ends of the region. It shows a regional concentric diffusion from the central city.

8. Dialect classification and dialect division has been attempted by applying multivariate analysis like cluster analysis. However, in the case of the “Hamaogi” vocabulary of Shonai district the result did not show clear geographical cleavage. It may be due to the fact that this district forms one cultural unit. It may also be because the

analysis is based on lexicon. Phonology and grammar show distinct regional differences in the Shonai district (Inoue 2000.2).

9. MCA analyzes both dimensions of a matrix (of person and word). In analyzing the word dimension, characteristics of words were found to influence the destiny of those words. Coincidence with words in Tokyo is crucial and also usage in neighboring areas is influential (words distributed in small areas tend to disappear earlier).

2. Regional and age differences: eight representative words

2.0 “Age Pillar Map” of “Hamaogi” data

The order of description of this paper is from simpler to more complex to understand. In this section, *eight representative words* will be described individually according to the traditional dialectological procedure. We extracted four major types of patterns according to the results of multivariate analysis of overall distribution pattern. Two Figures for each word, a total of eight Figures will be shown. Also, the depth of commentary is deeper than most of the traditional dialect maps. In the past, discussion about relative year of change was based on the geographical distribution at the time of the survey. Unlike those surveys, this is a follow-up survey of words used in 1767, so the description has historical depth. We can see the usage in Tsuruoka (or Sakata) 250 years ago. In other words, it also gives a clue as to the absolute year of change (Inoue 2004.7, 2010.7). It is a traditional dialect study method to explain the distribution of individual words by showing them on maps. In this paper, we also refer to the analysis from §3 below and add cross references.

Pillars at 27 survey points were set up on the map of the Shonai district to display the dimension reflecting the 7 generations. We later learned that the same visualization technique had been applied in Germany (Mang & Wollin 2010), and we renamed it to “**age pillar map**”. Figures 2 to 9 show the eight *representative* “age pillar maps”. ■ is

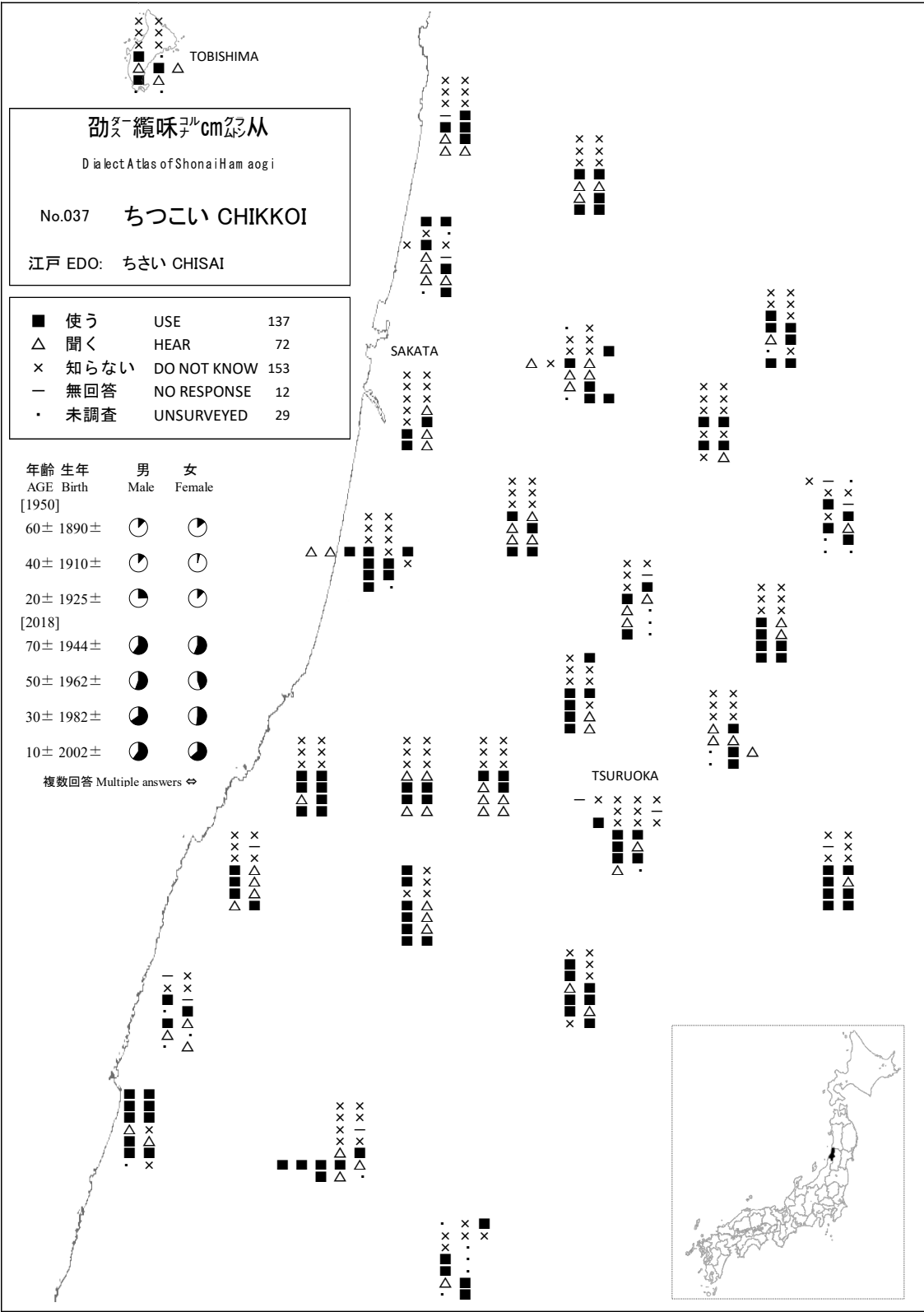
“use”, Δ “hear”.× “do not know”. The location of the survey area in Japan is shown in the inset at the bottom right of the map.

2.1 Revival of colloquial “a chikkoi” (age difference, axis 1)

The first example shows an impressive exceptional distribution. Figure 2 shows age-area distribution of the dialectal form *chikkoi* (standard *chiisai*) ‘small’. Compared to the first survey of the upper three generations of the pillar, the mark ■ (use) appears more frequently in the second survey of the lower four generations. Over the last 70 years or so, the colloquial (slang) expression *chikkoi* has become popular in Shonai. It is not standard Japanese, but colloquial Tokyo speech (Inoue 2000.2, Inoue & Hanzawa in press). Contrary to common sense, this Hamaogi form shows an increase in use (or revival of word before the modern age).

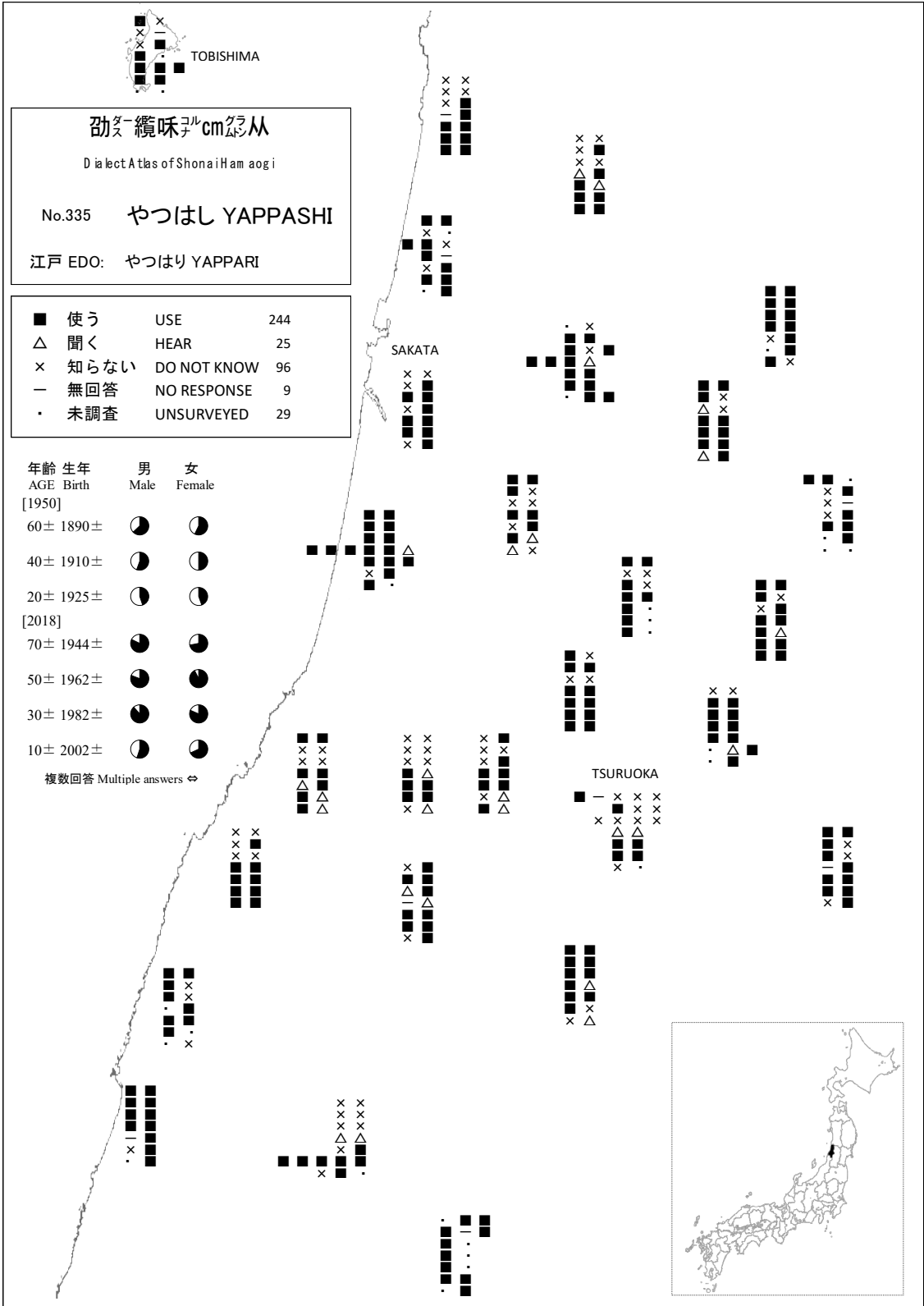
2.2 Spread of colloquial “b yappashi” (age difference, axis 1)

In Figure 3 *yappashi* (*yahari*) ‘after all’, more usage is shown in the lower four generations of the second survey than in the upper three generations of the first survey. Colloquial *yappashi* has spread to various parts of Japan since the early 16c, and flowed back to Tokyo (Inoue 1998.1). Similar changes have progressed in the Shonai district. (These two words and about 20 words distributed near in Figure 19 are exceptional because they are still increasing despite being words from 250 years ago).



『庄内浜藪』方言地図 (国立国語研究所 NLRI 1950; 井上・半沢 Inoue & Hanzawa 2018)

Figure 2. a *chikkoi* ‘small’



『庄内浜藪』方言地図 (国立国語研究所 NLRI 1950; 井上・半沢 Inoue & Hanzawa 2018)

Figure 3. b *yappashi* ‘after all’



Figure 4. c *waniru* 'be shy'

2.3 Remnant of the old “c waniru” (regional difference, axis 2)

Figure 4 shows *waniru* (*hanikamu*) ‘to be shy’. The “Hamaogi” form was maintained in the vicinity of the southern tip and in the eastern mountain hills. This is a representative regional concentric distribution showing remnants in peripheral areas. But among the present younger generation, it has decreased in the whole area. This is similar to Figures 5, 6, 7 (d, e, f) in terms of showing regional concentric distribution.

2.4 Sakata word “d hocha” in the periphery (regional difference, axis 2)

Figure 5 *hochu* (*hocho*) ‘kitchen knife’ is annotated as a Sakata word in “Hamaogi”. According to Figure 5, it is no longer used in Sakata itself (except the youngest student among the survey respondents, who admitted usage), and is scattered mainly in the areas of eastern and southern peripheries.

2.5 Sakata word “e yami” in the periphery (regional difference, axes 3 and 4)

Yami (*tsuri*) ‘fishing’ is also annotated as a Sakata word in “Hamaogi”. In Figure 6, many elderly people use it in the north, but users are scattered in the periphery area in the south, showing a regional concentric distribution. Before modern times, Sakata's *yami* had already shown a periphery distribution as an old form. After 250 years, it disappears from the vicinity of Sakata but remains in the elderly generation around Shonai district. It shows a residual, periphery distribution in the north and south. The distribution pattern is similar to Figure 4, *waniru*. About 10 or so words show similar distribution. We can calculate approximate speed of diffusion to be less than 1km per year (Inoue 2003.7, 2010.7).

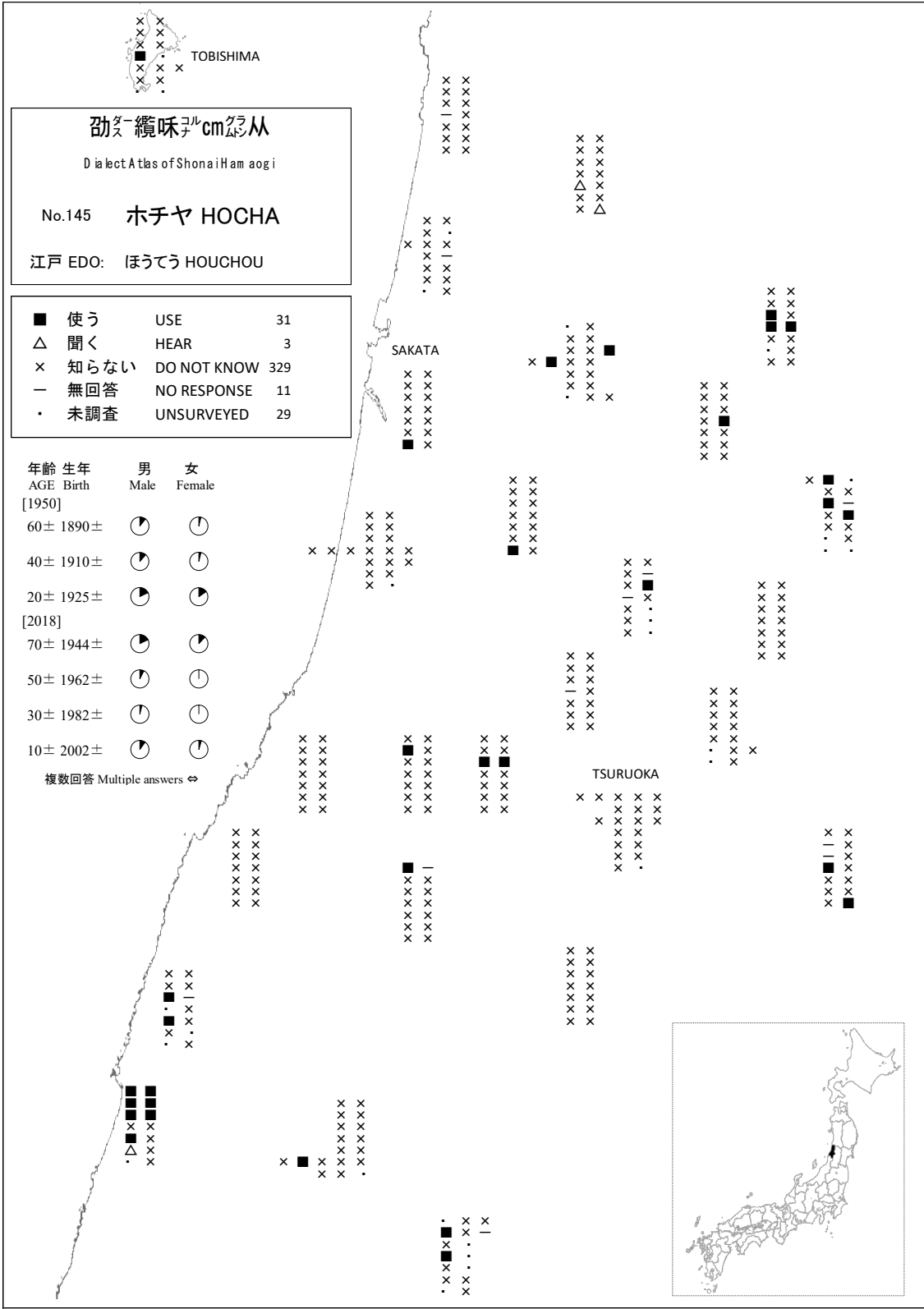
2.6 Periphery residual of “f tenokoppa” (regional difference, axes 3 and 4)

Figure 7 *tenokoppa* (*tenokou*) ‘back of the hand’ shows a marked decrease. *tenokoppa* was found mainly among respondents of the first survey in the northern part

and southern periphery of this area (Inoue 2001.2, 2009), but mostly disappeared in the second survey.

2.7 Decline of “g danma” near Tsuruoka (regional difference, axes 3 and 4)

Figure 8 *danma* (*otedama*) ‘beanbags, jackstones’ are found mainly in the southern half of Shonai and also among the elderly. It once spread near Tsuruoka, but then faded among the respondents of the second survey due to pressure from the standard language *otedama*. (About 20 or 30 words show similar distribution according to Figure 19).



『庄内浜藪』方言地図 (国立国語研究所 NLRI 1950; 井上・半沢 Inoue & Hanzawa 2018)

Figure 5. d *hocha* 'kitchen knife'

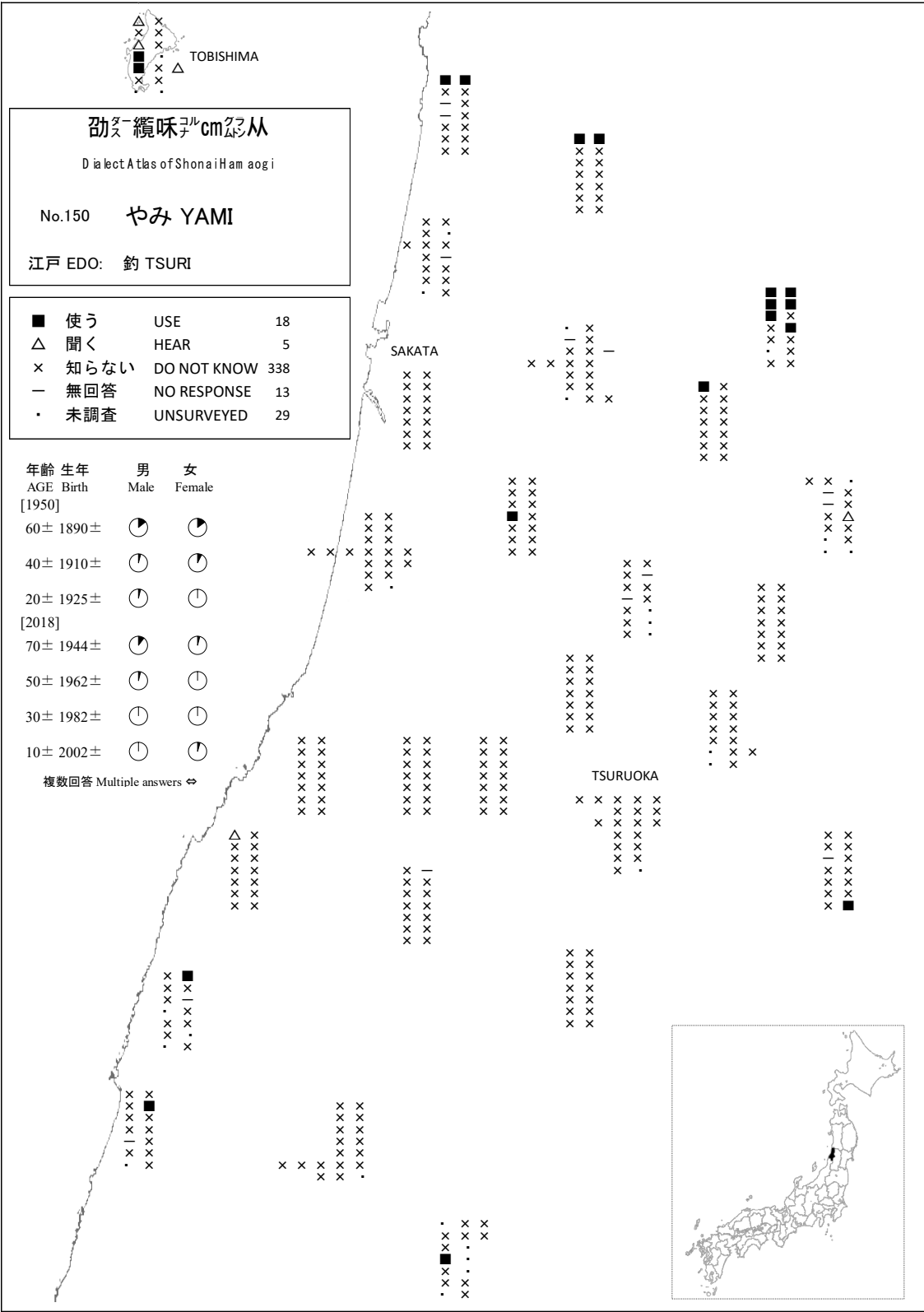
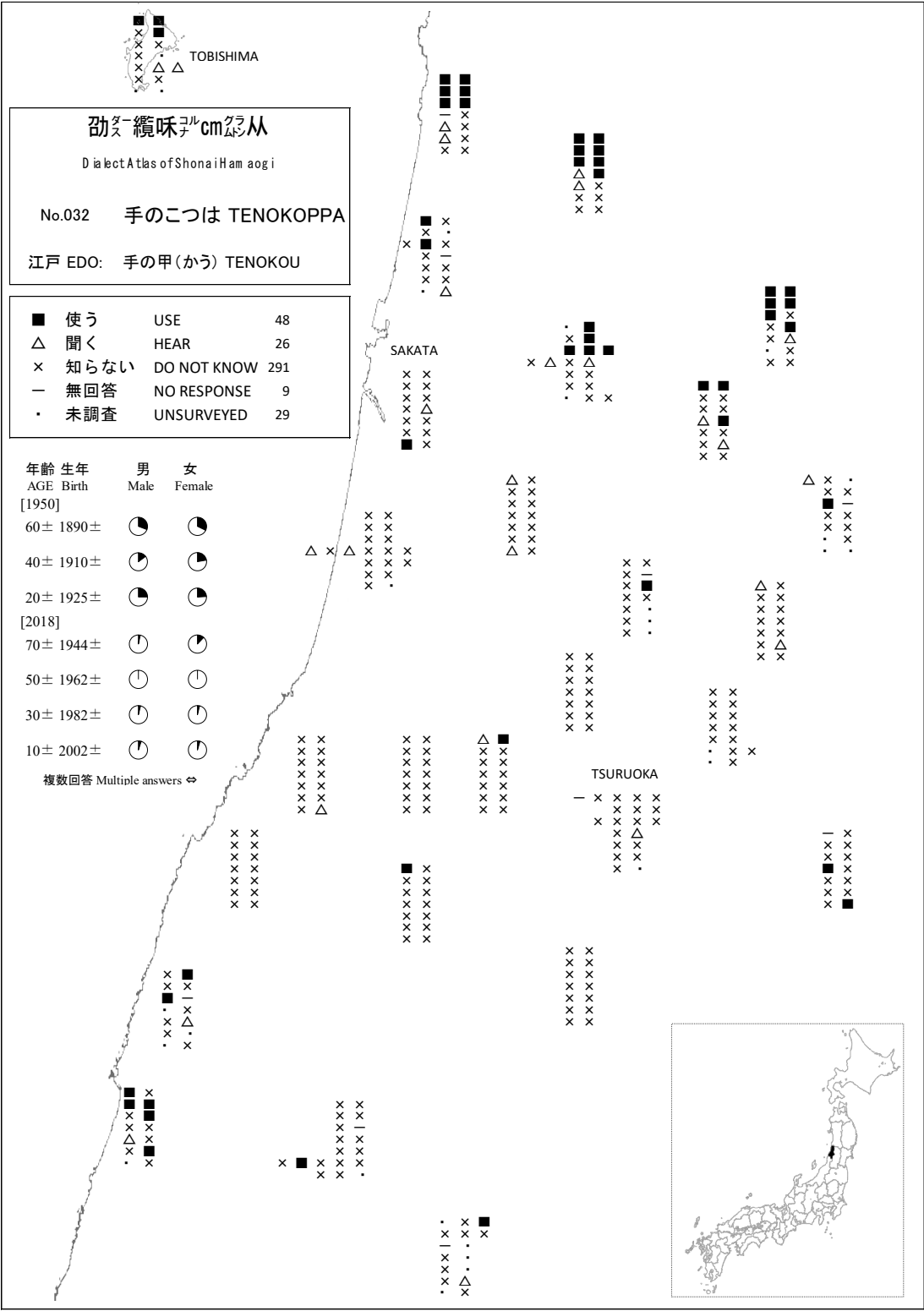


Figure 6. e *yami* 'fishing'



『庄内浜藪』方言地図 (国立国語研究所 NLR1 1950; 井上・半沢 Inoue & Hanzawa 2018)

Figure 7. f *tenokoppa* 'back of the hand'

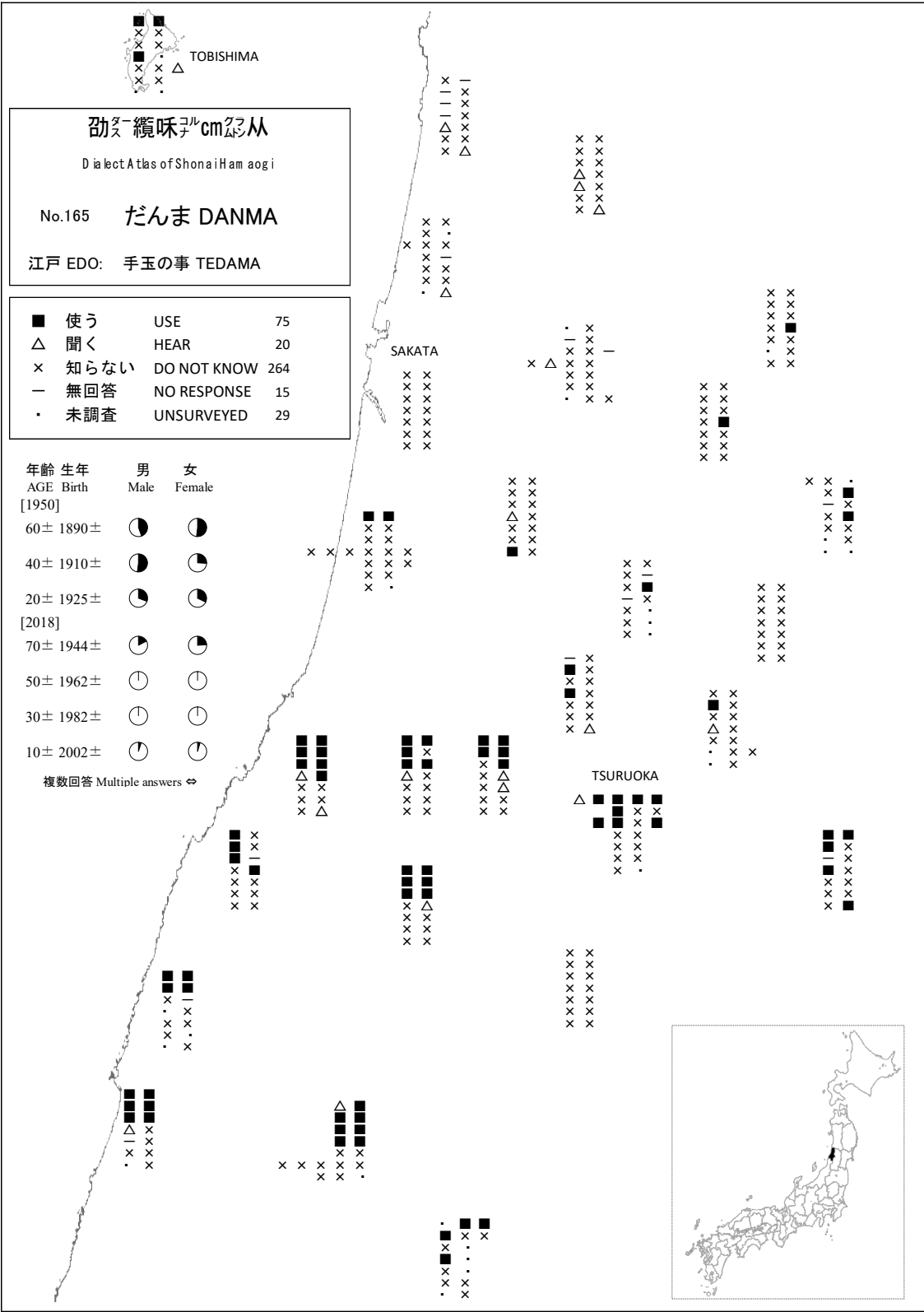
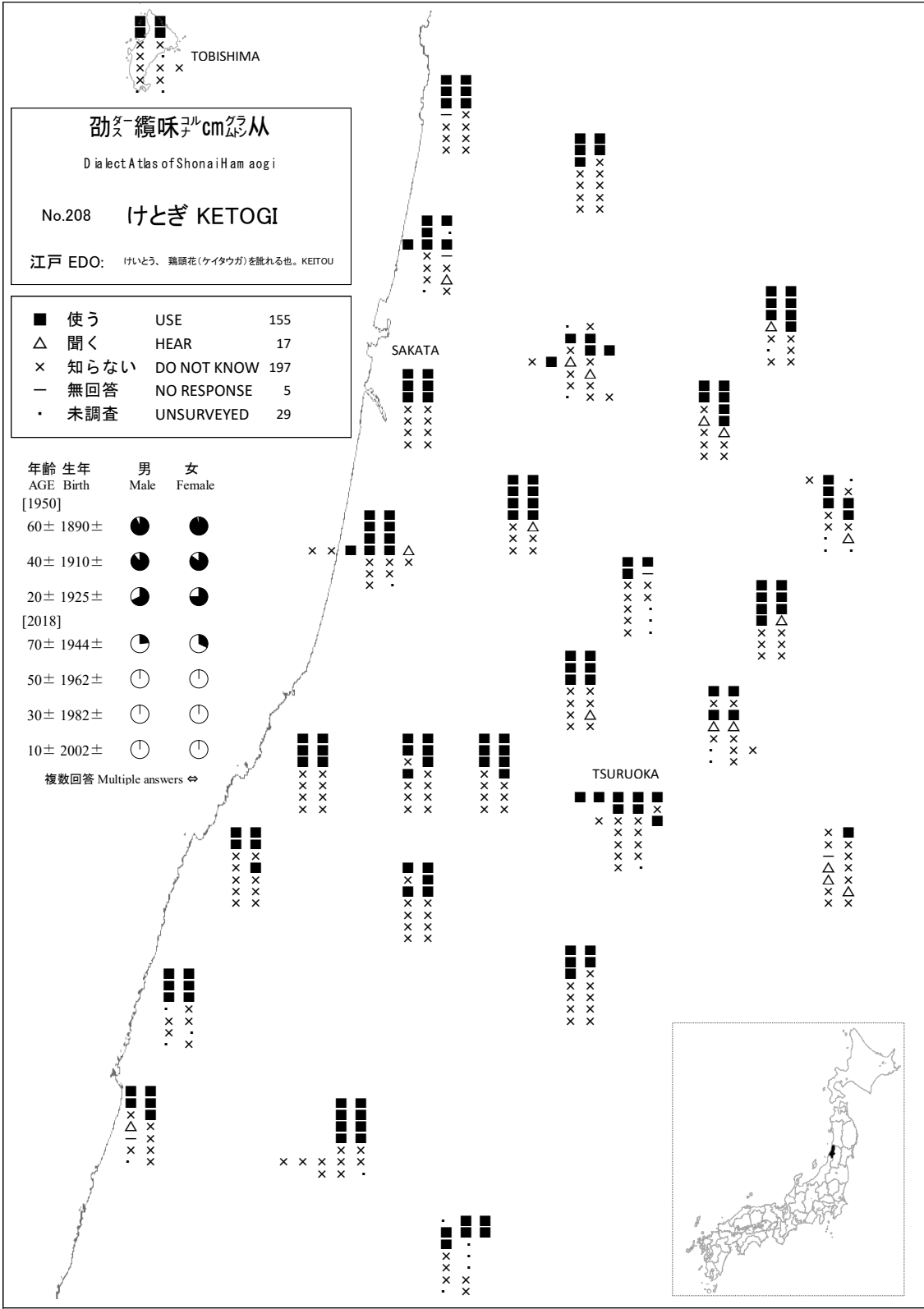


Figure 8. g *danma* 'beanbags, jackstones'



『庄内浜藪』方言地図 (国立国語研究所 NLRI 1950; 井上・半沢 Inoue & Hanzawa 2018)

Figure 9. h *ketogi* 'cockscomb'

2.8 Rapid standardization of “*h ketogi*” (age difference, axis 1)

The age difference of *ketogi* (*keitou*) ‘cockscomb’ in Figure 9 is remarkable. Almost all generations answered “use” in the first survey, but in the second survey, only a part of the elderly people use it near Sakata and Tsuruoka. It is not inherited by the generations below it. It is one representative of about 300 words (according to Figure 19) with large age difference without showing any clear geographical tendency.

3. Regional and age differences: entire “Hamaogi” vocabulary

3.1 Remnant rate of all 406 vocabulary items

In traditional dialect geography, individual words were described as shown above. In the last quarter of 20th century, dialectometric research that numerically grasps and analyzes the words as a whole became popular. We will perform this procedure in §3.

Figure 10 shows the combined average remnant rate or usage rate of all 406 items in an “age pillar dialect map”. On the map, 27 pillars each show the age difference in remnant rate. Responses of “use” were set to 1, while “hear” were set to 0.5, and “don't know” to 0, and the remnant rate was subdivided in 20 steps in 5% increments. Increments exceeding 80% are represented by solid black symbols while those falling short of 79% are represented by non-solid symbols with more white. Below the legend on the left of the figure, 14 pie charts show the average remnant rate by gender and seven age groups. The black part of the pie is the Hamaogi remnant rate. Symbols are whitish for the younger generation, that is, the remnant rate is low. This shows that after the generation born around 1944 (the elderly in the second survey), the decline suddenly progressed.

The blacker the symbol of each person at each survey point, the greater the remnant rate of Hamaogi vocabulary. As a whole in Figure 10, the age difference is conspicuous, the symbols are darker in the upper generations, and lighter in the modern

younger generation at the lower end of pillars. Age differences are observed at all points, but at each point, the order is not orderly and individual deviation can be detected.

Geographically, there are many dark-colored individuals at the southern end. Basically, the difference between north and south is large, but the remnant rate is also large in the north and along the east mountains in addition to the southern end. Although some individual differences can be seen in this figure, the map of the average value for each survey point shows a high remnant rate in the old-fashioned remote areas (Hanzawa 2021), and shows a regional-scale concentric distribution.

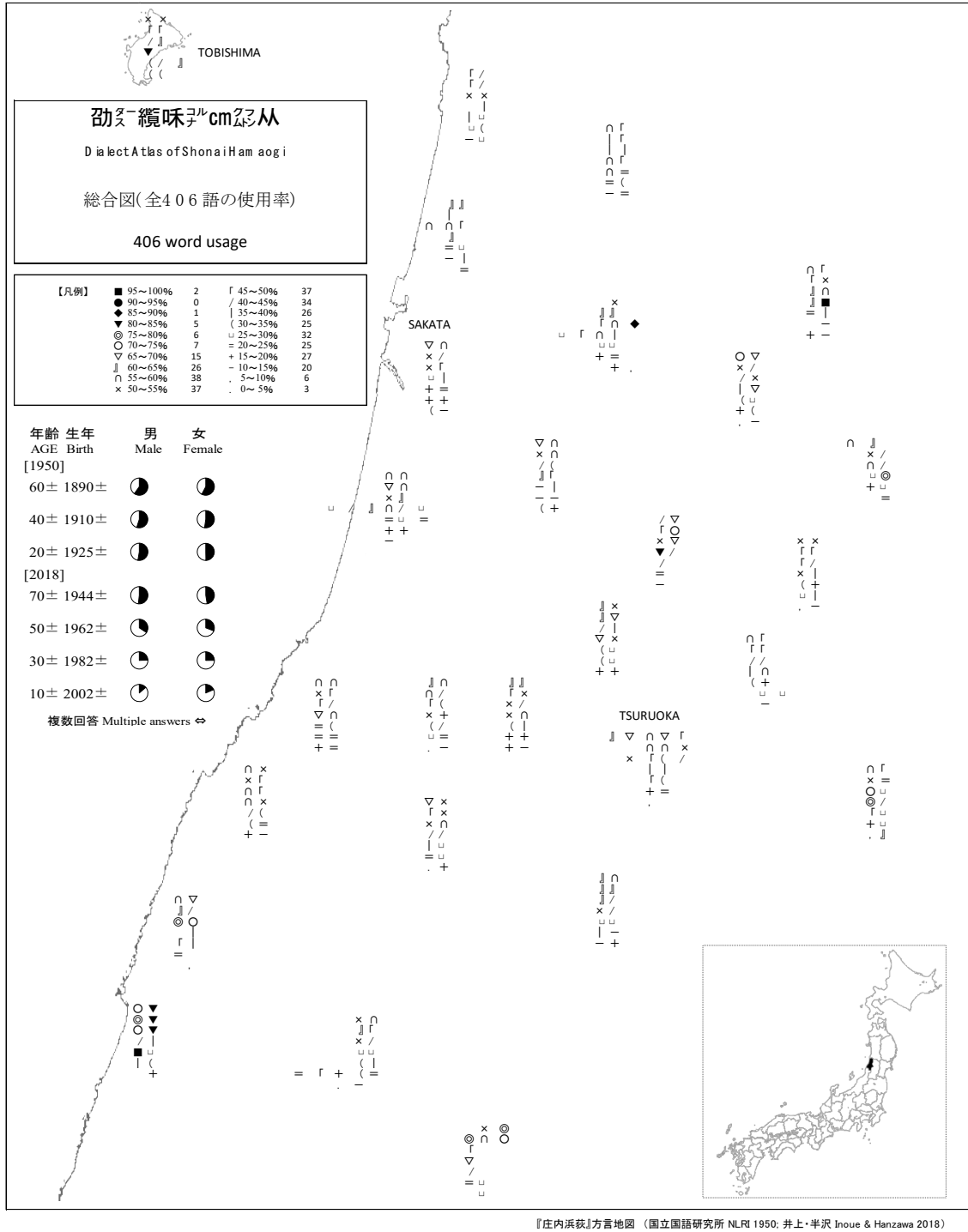


Figure 10. Geographical and age distribution of remnant rate

The “age pillar maps” are effective for showing regional differences and age differences in detail in one figure. Before we made the comprehensive map in Figure 10 with the remnant rates, we had browsed over 400 maps and although we noticed some

distribution patterns, we could not grasp the major tendency. It can be seen from Figure 10 that the age difference is large. This can be further summarized in Figure 11 by taking up eight representative words (*a* to *h*) instead of all words.

The regional differences within the Shonai district are not distinctive. Fourteen words of Sakata dialect were recorded in “Hamaogi”, but the regional differences within Shonai were not retained in the first and second surveys. As for the approximately 400 words of “Hamaogi”, a clear dialect border within the region cannot be drawn.

3.2 Age group remnant rate of eight words

The data above are composed of a matrix of dimension of words and dimension of person. The two dimensions will be summarized separately in the following sections. First, the dimension of person will be analyzed. Figure 11 shows the average remnant rate of a set of eight representative words divided into 7 age groups. The thick dotted lines *a* and *b* show an exceptional increase in the younger generation. Other words are declining in the younger generation. The dotted lines *c* and *d* and the thick lines *e* and *f* are at the bottom, which had already declined in the first survey, have decreased further in the second survey. The solid lines *g* and *h* show rapid decrease.

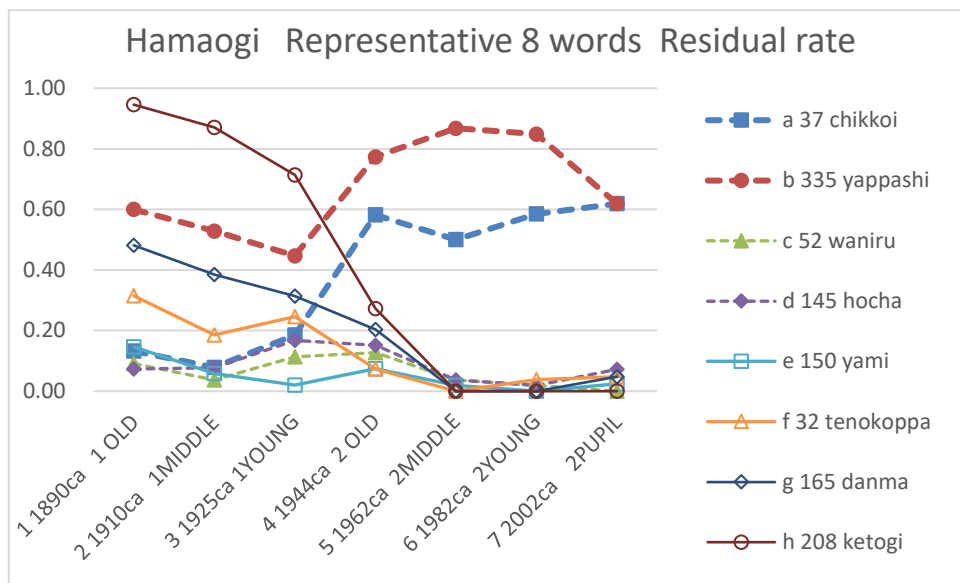


Figure 11. Age differences: eight representative words

In the first survey on the left, the differences among the eight words were large, but in the second survey on the right, the two extremes are conspicuous. In particular, there are only small differences between words in the generations born after 1962. It was possible to summarize the age difference in remnant rate across eight age pillar maps with this one graph.

3.3 Scattergram of all respondents by age

Regarding the dimension of person, the geographical summary was shown in the comprehensive map of Figure 10 and by generation in Figure 11. Furthermore, the values of all respondents will be shown in relation to age more precisely. Figure 12 shows respondents' age and remnant rate in a scattergram. The horizontal axis faithfully represents the birthyear of each respondent. The vertical axis represents the remnant values of 406 words for each respondent. Although it is falling to the right as a whole, it can be seen that there is a large dispersion (individual difference) in the middle generations.

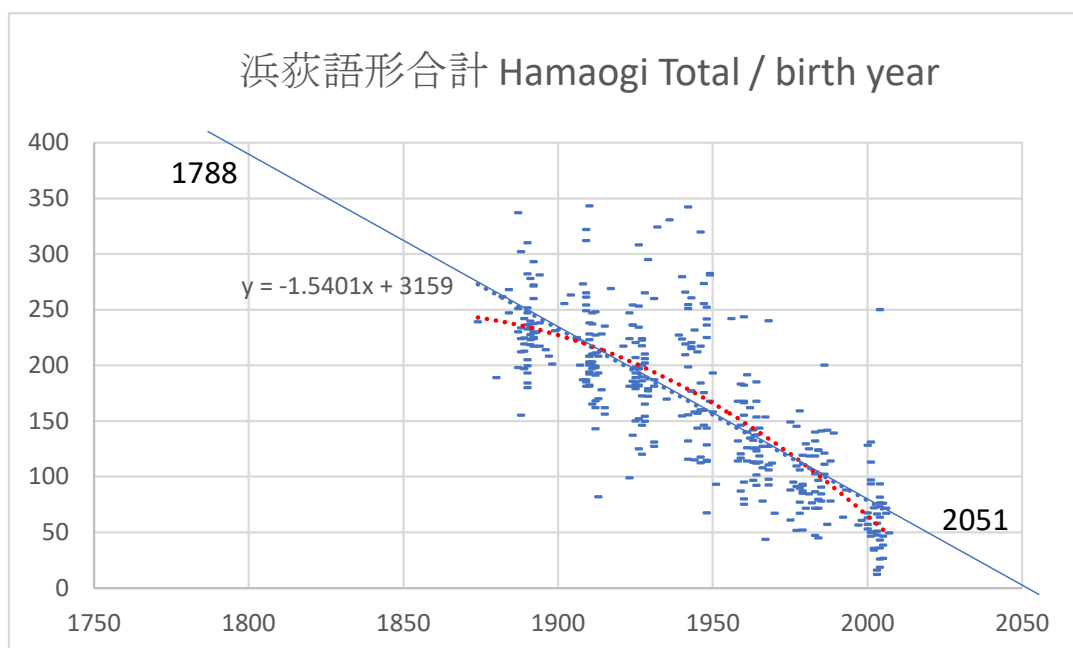


Figure 12. Age and remnant rate

The approximate straight line can be extended back and forth, and mathematical formulas can be calculated. 100% corresponds to 406 words in 1788, a little later than 1767 when “Hamaogi” was compiled. It will become 0% in 2051, which is about 30 years from now. One generation from now, the “Hamaogi” vocabulary may disappear. However, some words will survive as described later (§5.1). The linguistic change is not linear but forms an S-shaped curve (Aitchison 1991, Sankoff 2006, Inoue 2020.2), and it is possible that the rate of 0% will be delayed to the 22nd century, especially because some Tokyo colloquial words are included in the data.

4. MCA of “Hamaogi”: Geographical and age differences

As mentioned above, we started from observation of individual vocabulary items and proceeded to comprehensive consideration. The simple tabulations were shown by arithmetic calculation, and the overall tendency of the whole was grasped. For further investigation to grasp the internal structure of the data as a whole, application of multivariate analysis is appropriate. In §4, the internal differences are analyzed by applying the multivariate analysis based on the overall trends above. A two-dimensional analysis of the word dimension and the person dimension is possible, and in §4, we will focus on the dimension of person (region, age). The relationship between the dimension of person and the dimension of word, the focus of §5, is two sides of a coin in the multivariate analysis.

4.1 MCA and geographical result

MCA, or Multiple Correspondence Analysis was applied to the data. MCA, also referred to simply as “Correspondence Analysis” has been applied to nominal (as opposed to continuous numerical) variable data in the European research. MCA is equivalent to Hayashi 3 (Hayashi’s Quantificational Theory Type 3) which had been developed independently in Japan after the War (Hayashi 2004) and had been used in Japanese dialect research for a long time (Inoue 1996.10, 2001.2, Yarimizu et al. 2004, Ueda &

Perea 2014). MCA and Hayashi 3 are similar in principle to the classic Factor Analysis which is applied to numerical data. These techniques are similar in essence with VARBRUL developed in the US (Gorman 2009, Singler 2001).

MCA was applied to the data of 372 respondents, out of the 374 whose data was obtained from the two surveys in 1950 and 2018, except for 2 who had many missing data. The answers for all 406 words were recoded into binary data of “use or not”, and a 372 x 406 data matrix was created. The “R” “corresp” function was used for the analysis. Below, we will proceed with the analysis based on the results of the four axes produced. The correlation coefficients of the four axes are shown in Table 1. The values of the axes 1 and 2 are large, and these two explain more than two thirds of the set of dialect forms. For each axis, a word score and a respondent score are given. Below, the scattergram shows the word and respondent values from axis 1 to axis 4.

By applying MCA, we were able to capture the tendency of distribution and change of dialects, and summarize complex patterns. Based on this, it is appropriate to consider axes 1 and 2, produced by MCA, in combination, and then the axes 3 and 4, in order to show the age and geographical distribution patterns. Below, we will compare the results with the patterns observed in the eight representative words introduced above. The conclusion of the meaning, i.e. the tendency, of the four axes is shown in Table 1 for easier understanding.

	Correlation coefficient	Function	Tendency	Representative words + / -
axis 1	0.397	Age difference	Increase in youth	a b / c d
axis 2	0.276	Regional difference	Concentric distribution	c d / h
axis 3	0.180	Regional difference	North-Center contrast	e f / g
axis 4	0.174	Regional difference	North-South contrast	e f / g

Figure 2 a. *chikkoi* ‘small’; Figure 3 b. *yappashi* ‘after all’; Figure 4 c. *waniru* ‘be shy’

Figure 5 d. *hochu* ‘kitchen knife’; Figure 6 e. *yami* ‘fishing’; Figure 7 f. *tenokoppa* ‘back of the hand’

Figure 8 g. *danma* ‘beanbags, jackstones’; Figure 9 h. *ketogi* ‘cockscorn’

Table 1. Meaning of the four axes

4.2 Individual MCA values: Overview of age differences

In MCA, the numerical value of the axis is given for each sample (respondent). In this section, we will look at age and regional differences. First the differences by age of the sample scores for each of the axes 1 to 4 are shown in the line graph of Figure 13. It shows the age differences in remnant rate over 7 generations and nearly 140 years at intervals of about 20 years.

In Figure 13, the bold line **axis 1** shows a large age difference, and words that are increasing in the younger generation are sorted out (negative values are words that are common in the elderly). The words reflected in this axis are “Hamaogi” forms that have been in existence for 250 years. This most important axis shows a good correspondence with the 7 generations, which means that there is a large age difference in the whole data. (In Figure 14 below, there are few regional differences.) The age difference of **axis 2** (the thin line) is small, becoming slightly larger in the youngest generation. (In Figure 15 regional differences are noticeable.) The age difference of the broken line **axis 3** is large, but it shows the opposite direction to axis 1 and takes a large negative value in the young generation. Most of the positive values are in the elderly in the second survey. (In Figure 16 the north has a positive value and the southern half has a negative value.) The age difference on the dotted line **axis 4** is rather large, lower in the middle and higher in the younger generation. (In Figure 17 the northern half is positive and the southern half is negative.) Axes 3 and 4 are interpreted to show the “Hamaogi” words that remain in the north.

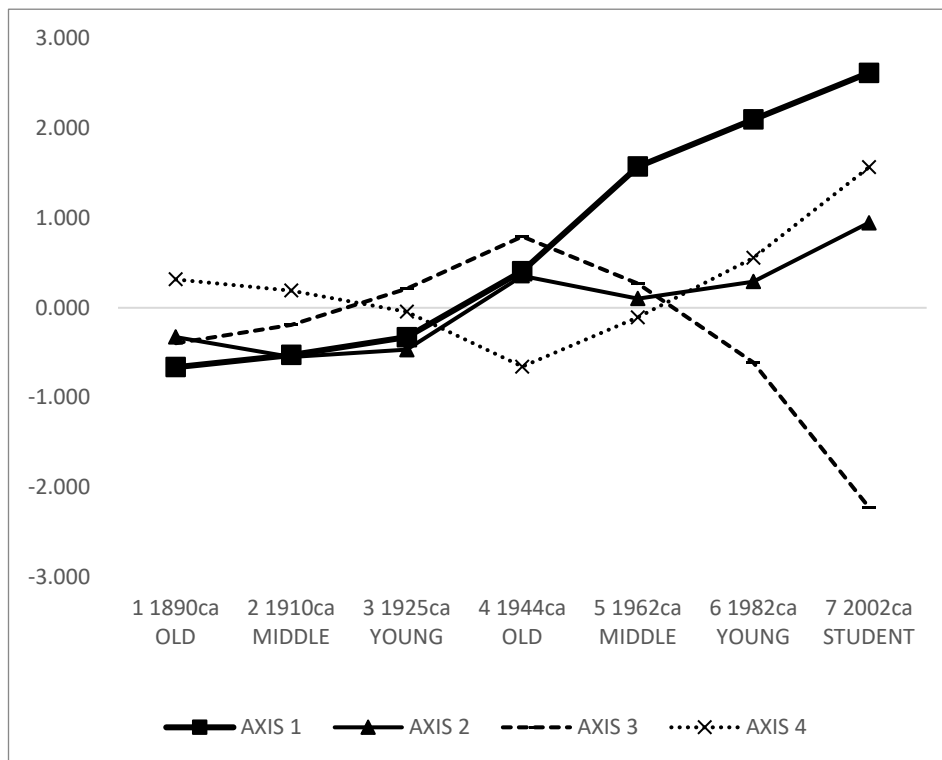


Figure 13. MCA person score: axes 1 to 4

In summary, data with intervals nearly 140 years show a large age difference. Due to the research based on the dialect glossary compiled before modern times, the age difference between the late 19th century birth and the 21st century birth was conspicuous; the spread of modern Japanese and the standard language after the war worked strongly. In the conventional glottogram surveys, words with geographical difference within the region were often selected, so the regional difference was noticeable. The surveyed words in “Hamaogi” are based on a dialect glossary focusing on the difference between former Edo (Tokyo) and Shonai, so the vocabulary was selected based on a large-scale, colorless, objective and general criterion, different from traditional small-scale geographical research. This valuable information shows steady language change over 250 years, or a quarter millennium.

4.3 Axis 1 Sakata vs. Tsuruoka

Above, we have seen a large-scale trend in the line graphs by age group. Below, regional differences among the respondents will be analyzed. The resulting values of the multivariate analysis are displayed on the map as the average value for all the 27 survey points. The axes 1 to 4 are shown in separate maps. Figure 14 shows the geographical distribution of the **axis 1**. All are positive values, and Sakata is the largest, Tsuruoka the smallest. (According to Figure 13, axis 1 showed a clear age difference.) This axis shows that usage of “Hamaogi” words is still increasing among young people. Commercial port city Sakata leads the trend, and administrative old castle city Tsuruoka is the last in using colloquial forms. Returning to Figures 2 and 3, the number of young users is increasing in all areas. (According to Figure 19 below, the effect of “a *chikko*” and “b *yappashi*” is great.)

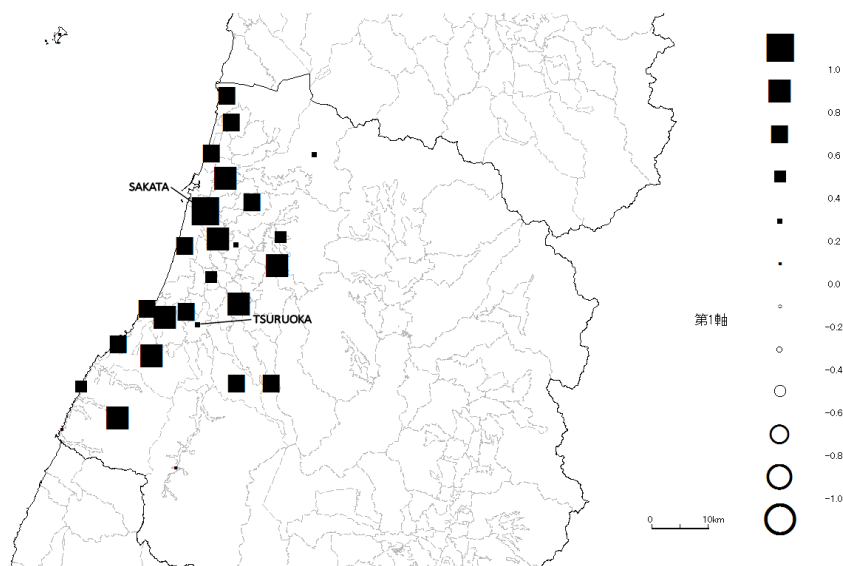


Figure 14. MCA person score axis 1

4.4 Axis 2 Tsuruoka centered concentric circle

As for the axis 2 in Figure 15, large positive values are found in and around the southern part of the district, and eastern mountains also show positive values. The northern and central plains show negative values. It is a regional concentric distribution

centered on Tsuruoka, and old words remain in the surrounding mountainous areas (it was flat in the age graph in Figure 13). (In Figure 19 described later, “c waniru, d hochu” was separated in axis 2.) Returning to Figures 4 and 5, the regional concentric distribution around Tsuruoka is shown in the north and in the southern tip.

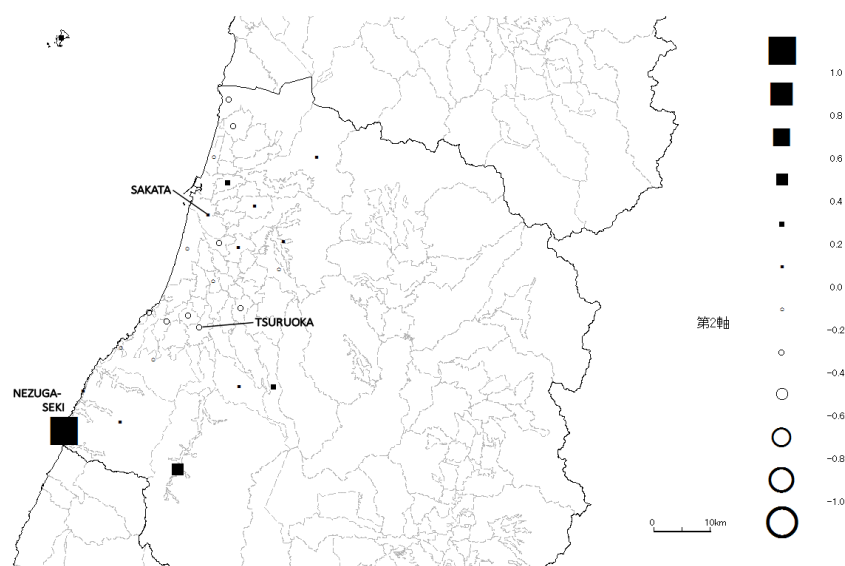


Figure 15. MCA person score axis 2

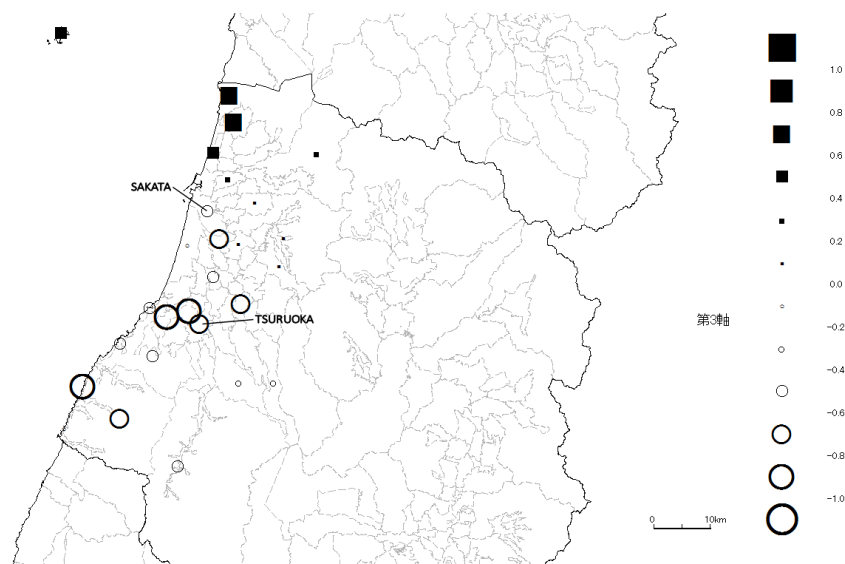


Figure 16. MCA person score axis 3

4.5 Axis 3 North-Center contrast

In Figure 16, axis 3 shows clear contrast between the north and the rest. The northern part shows a positive value, indicating the “Hamaogi” words that remain in the northern part. Large negative values are observed near Tsuruoka, so this axis should be interpreted as showing a contrast between the north and the center. (In the age graph in Figure 13, the peak was in middle age). “e *yami*”, “f *tenokoppa*” is representative. (Both are plotted upper left in Figure 19 below, and upper right in Figure 20 which picked up the exceptional distribution pattern.) In Figures 6 and 7, they remained in the north and south, showing a regional concentric distribution. It is a continuous distribution with “c, d” in Figure 15.

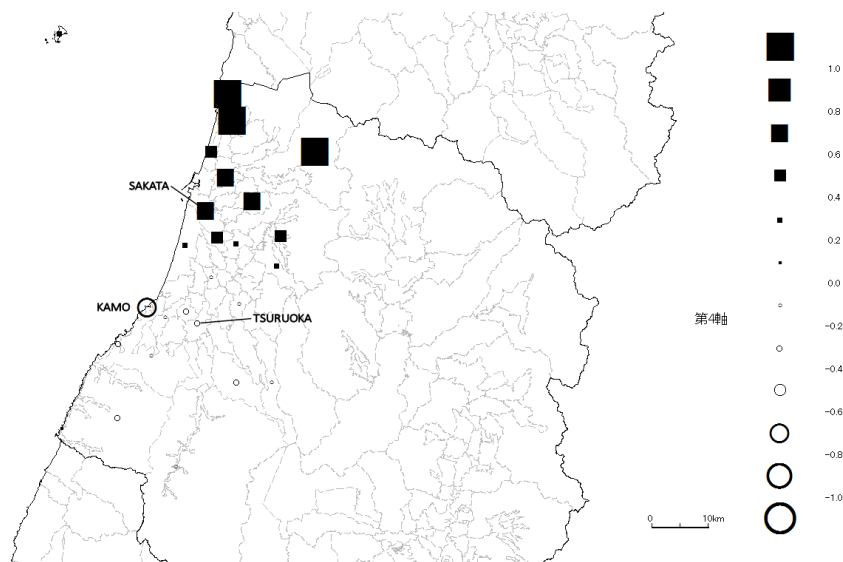


Figure 17. MCA person score axis 4

4.6 Axis 4 North-South contrast

As shown in Figure 17, axis 4 also shows north-south contrast. The northern part has a positive value, the southern part a negative value. Small port town Kamo stands out on the negative side. Kamo was an important town for providing fish to Tsuruoka. The north-south difference is so clear-cut that we may draw an isogloss between Sakata area vs. Tsuruoka area. The border line is near Mogami River (Figure 1), which divides the

Shonai district into *kawa-kita* 'river north' and *kawa-minami* 'river south' in lay people's perception of dialects, but they do not coincide exactly. It is rather near the border line between two commercial areas of competitor cities Sakata and Tsuruoka. The border line is almost the same distance from the two cities. Thus, axis 4 can be interpreted as showing a north-south contrast. (In the age graph of Figure 13, it decreases in middle age). Axis 4 picked up the exceptional patterns of axes 1 and 2. (According to Figure 20, "e yami", "f *tenokoppa*" was representative again on the positive side. On the negative side, "g *danma*" is found near Tsuruoka.)

4.7 Summary of the 4 axes: A North-South contrast

In the above, the values of axes 1 to 4 for each individual in the multivariate analysis were graphed by age group and mapped for each survey point. The appropriateness of interpretation was confirmed by comparing with the axis value for each word and by comparing with the individual "age pillar maps" shown in §2. As shown in Table 1, axis 1 shows age differences, and axes 2, 3 and 4 show regional differences. In summary, the main geographical trends of the Shonai district can be concisely approximated by a north-south contrast and especially by distance from the cultural center, Tsuruoka. We will simplify the 2-dimensional earth surface into one dimension using distance from Tsuruoka as the criterion in Figure 21.

5. MCA of "Hamaogi": Difference by word

5.1 Word difference and eight medical conditions

The dimension of words will be dealt with here. As could be seen from the eight representative words, there are differences among the trends each word follows. From the above analysis, it was found that the age difference is large in the data as a whole, followed by the regional difference. We will here analyze the age differences. By focusing

on age differences (and abstracting or ignoring the regional differences), each word can be displayed by a single line. The 420 words of the questionnaire (406 words, if the words from Sakata were not separated) were arranged in the order of the remnant rate and divided into eight levels. In parallel with the vitality of language (or pathology of endangered languages), they were metaphorically named as follows: 1 dying, 2 critical, 3 serious, 4 bedridden, 5 absolute rest, 6 unstable, 7 stable, 8 healthy. 50 words were allotted to each category except 1 dying, which was allotted 70 words.

The eight graphs are shown from Figure 18-1 to Figure 18-8. The overall pattern can be grasped, though the dialect label in the legend are hard to read. The dialect labels with ■ are words which are used as colloquial speech in Tokyo, and those with ● (also shown by red lines) are *eight representative words* discussed above. The vertical axis of each graph shows the usage rate, the upper end being 100%. The horizontal axis shows age, the left is the elderly in the first survey, and the right is young, junior high school students in the second survey.

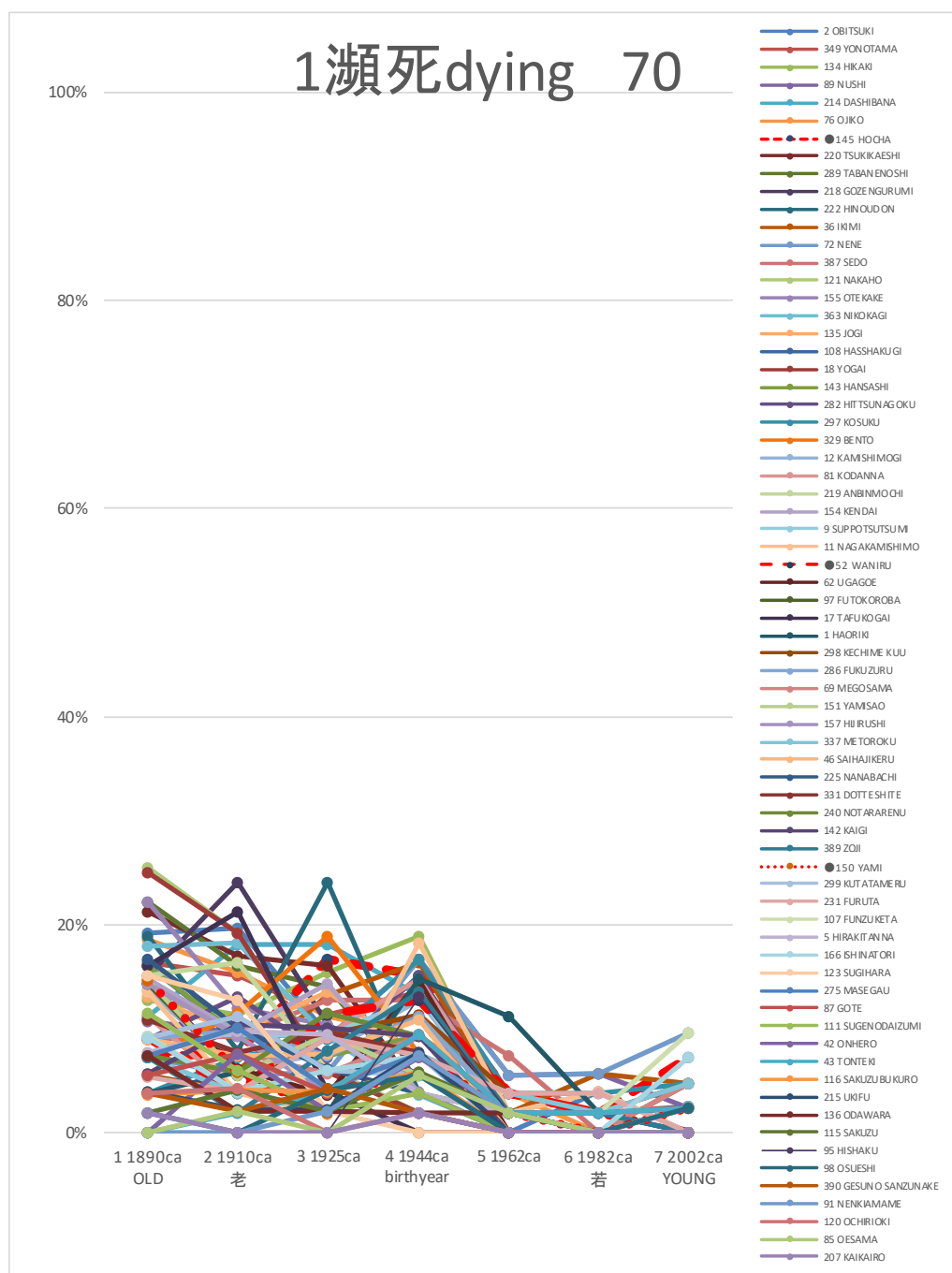


Figure 18-1. Generation and eight conditions: dying

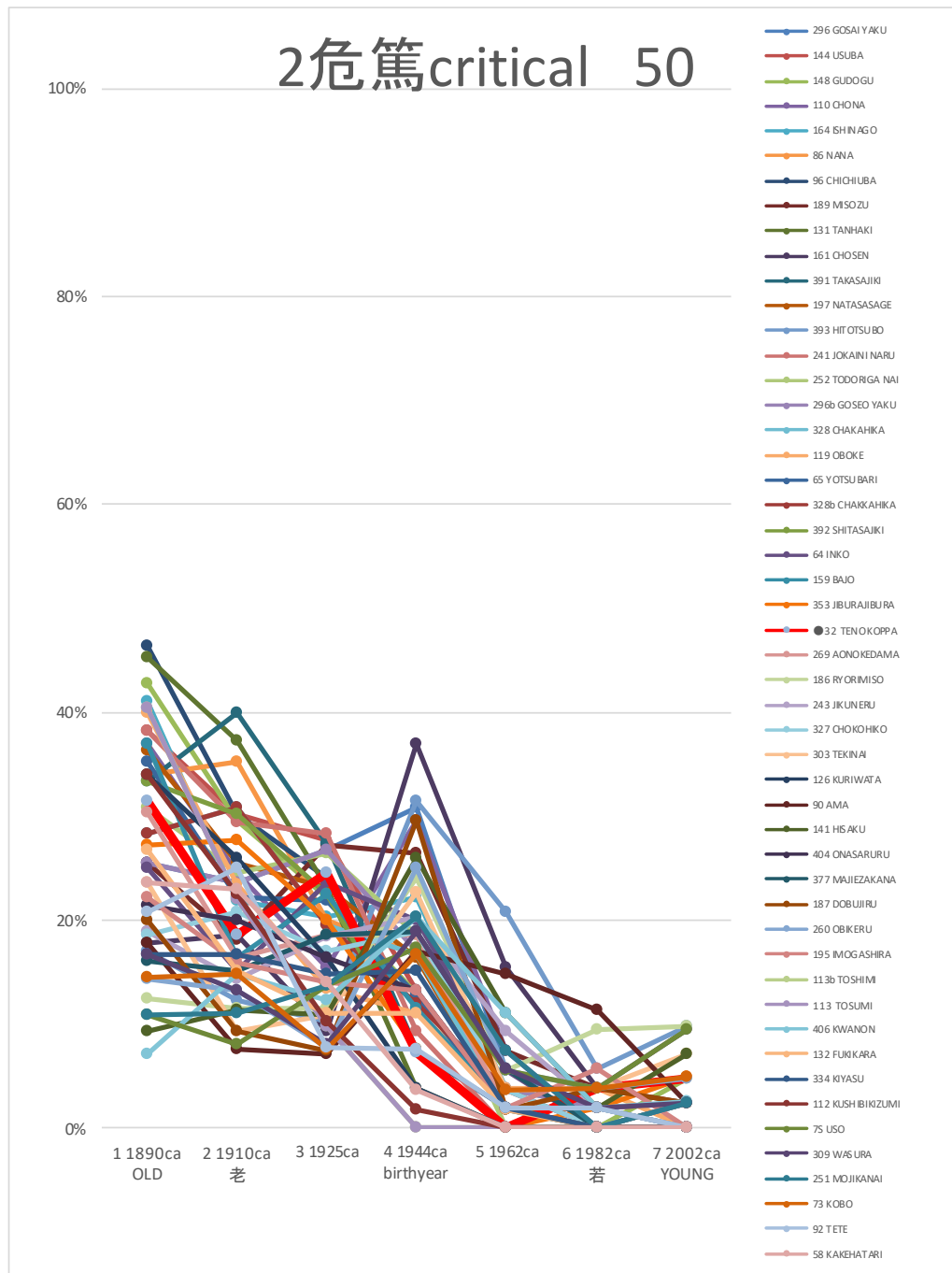


Figure 18-2. Generation and eight conditions: critical

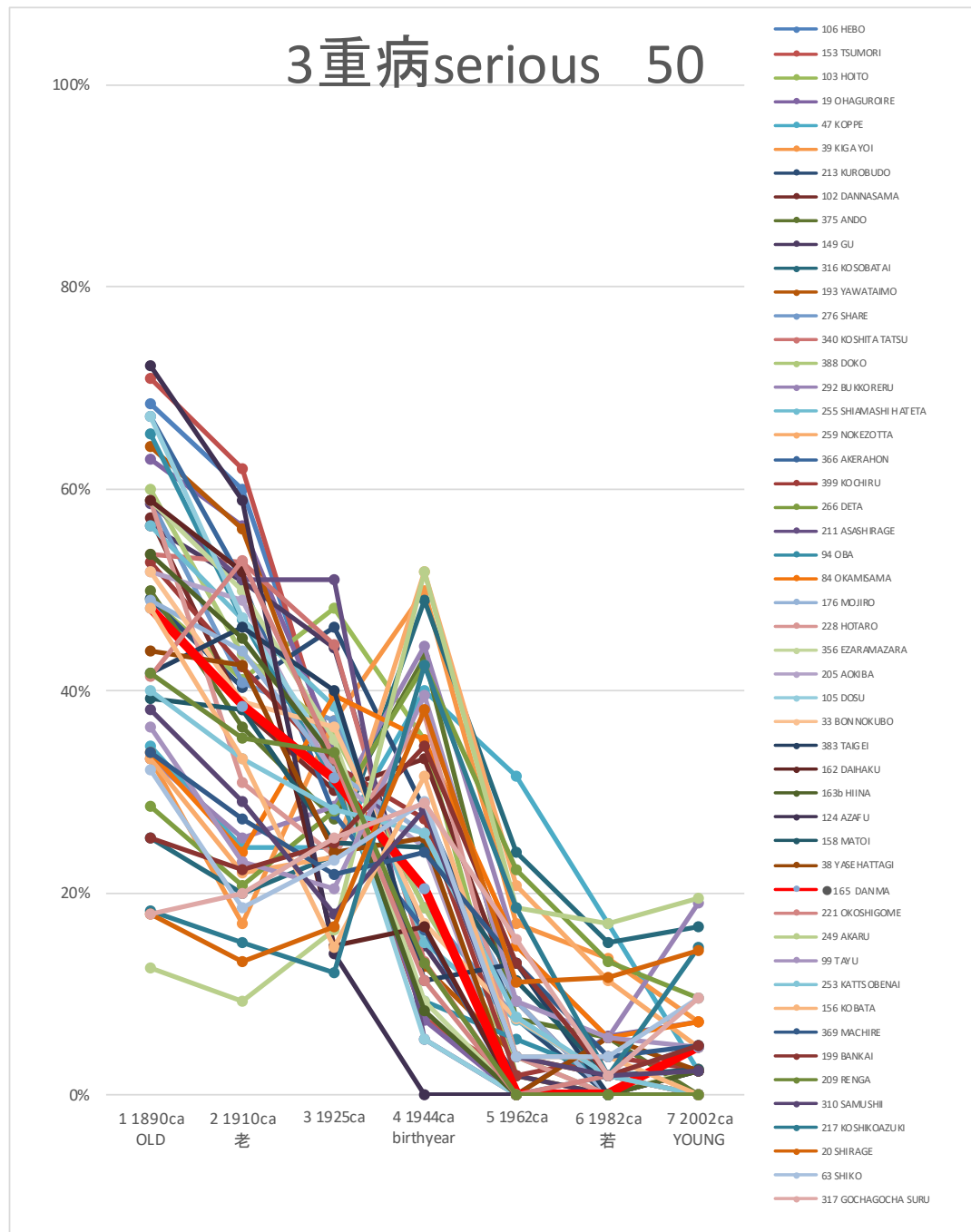


Figure 18-3. Generation and eight conditions: serious

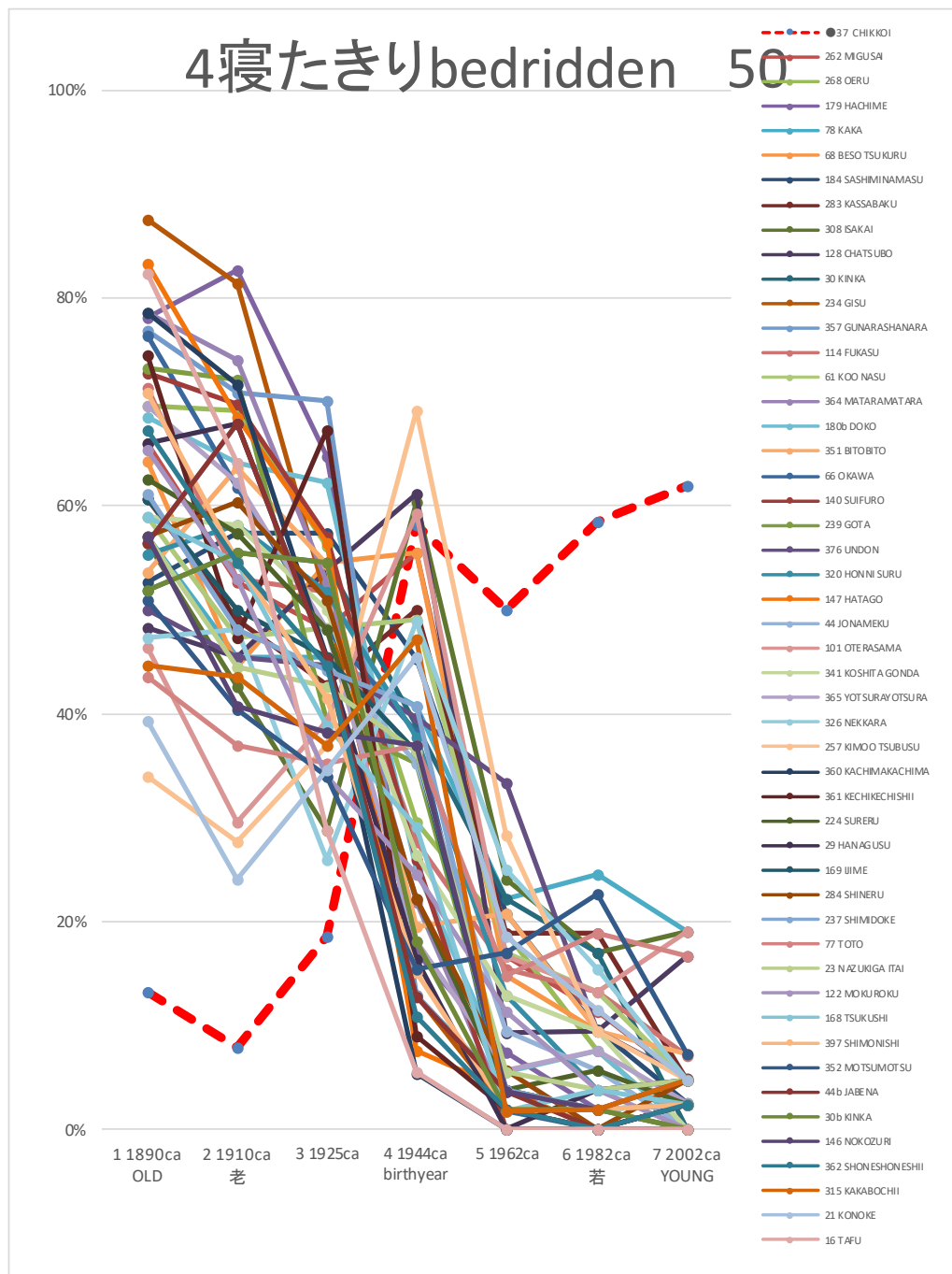


Figure 18-4. Generation and eight conditions: bedridden

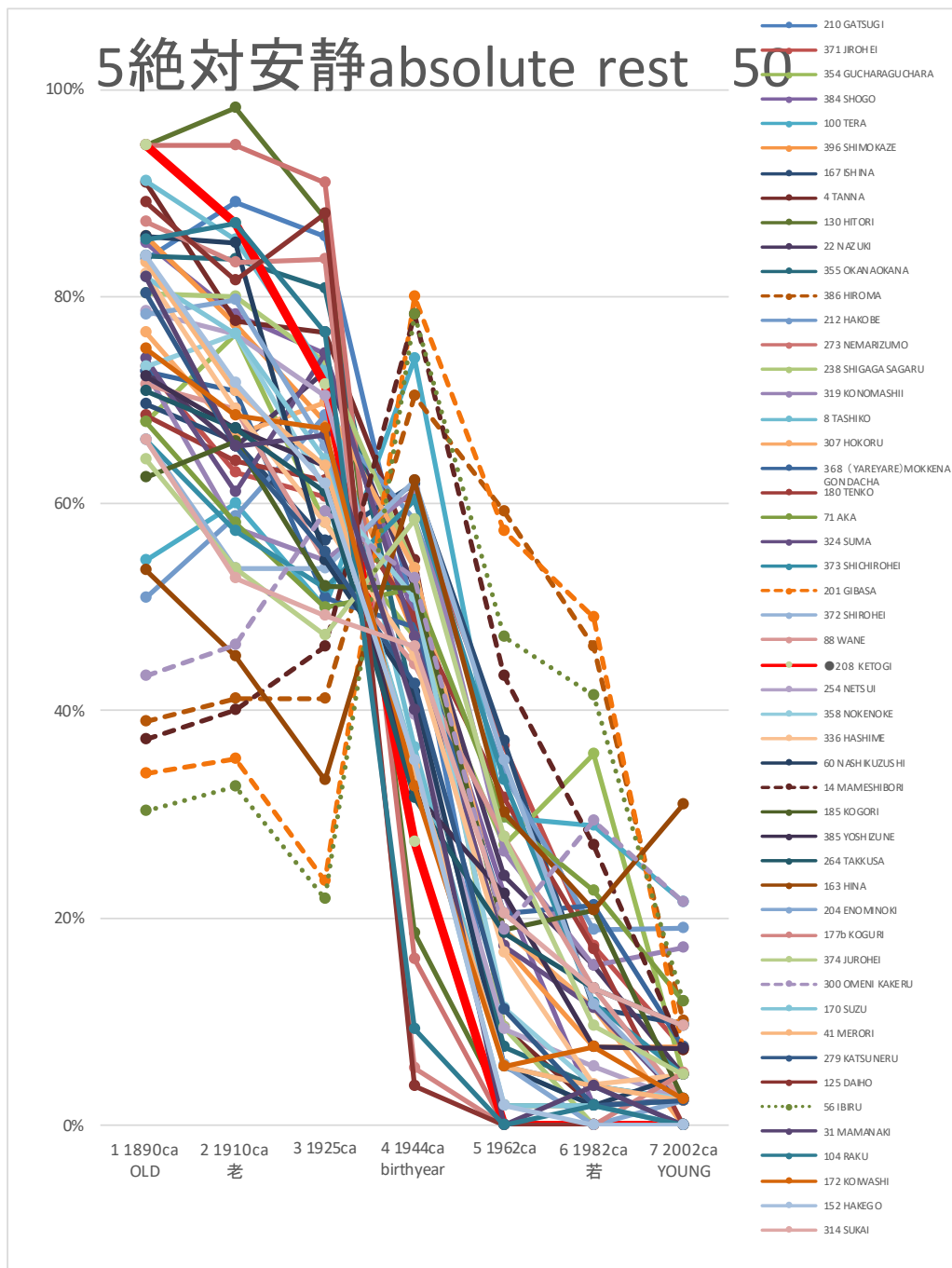


Figure 18-5. Generation and eight conditions: absolute rest

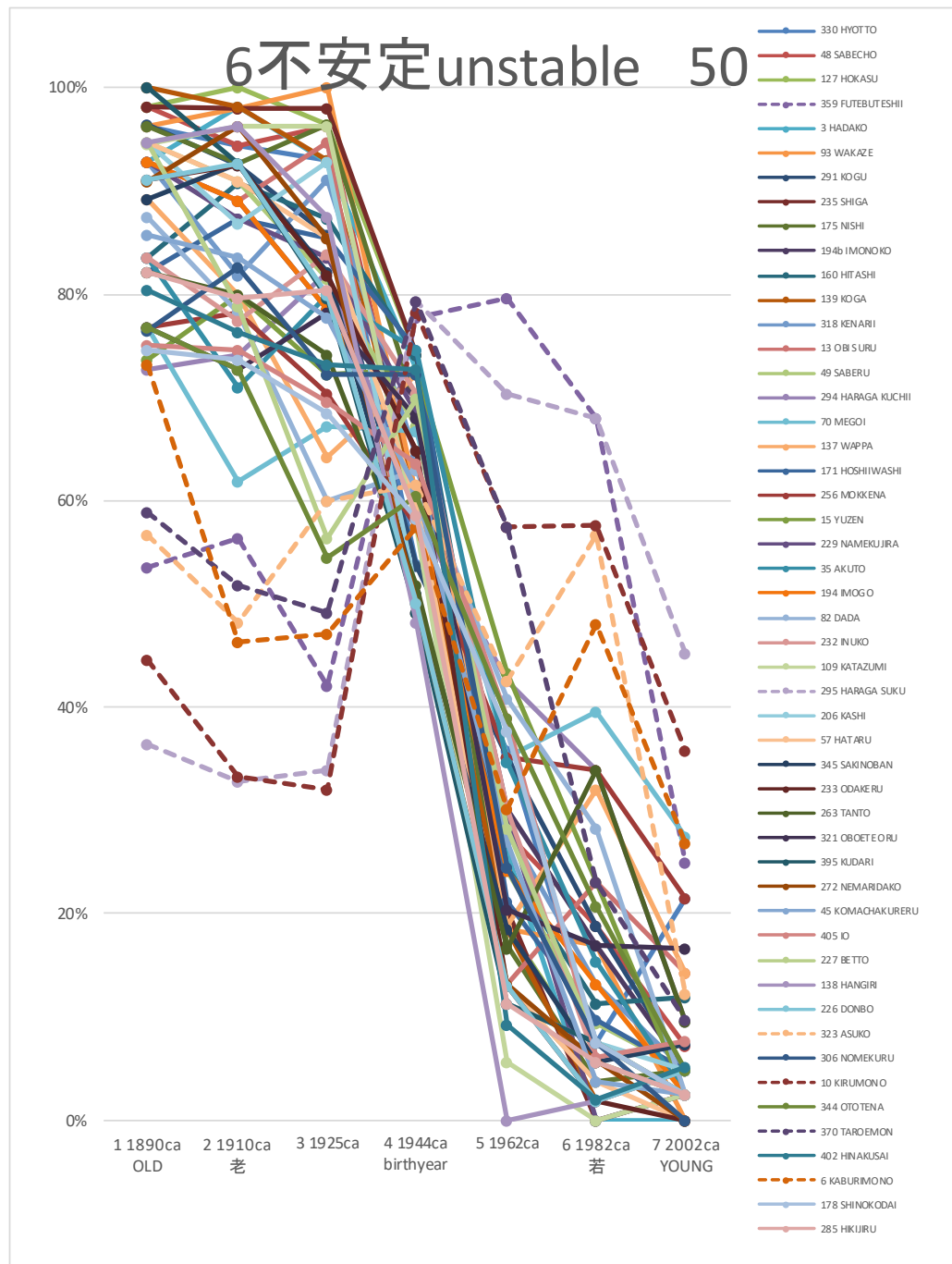


Figure 18-6. Generation and eight conditions: unstable

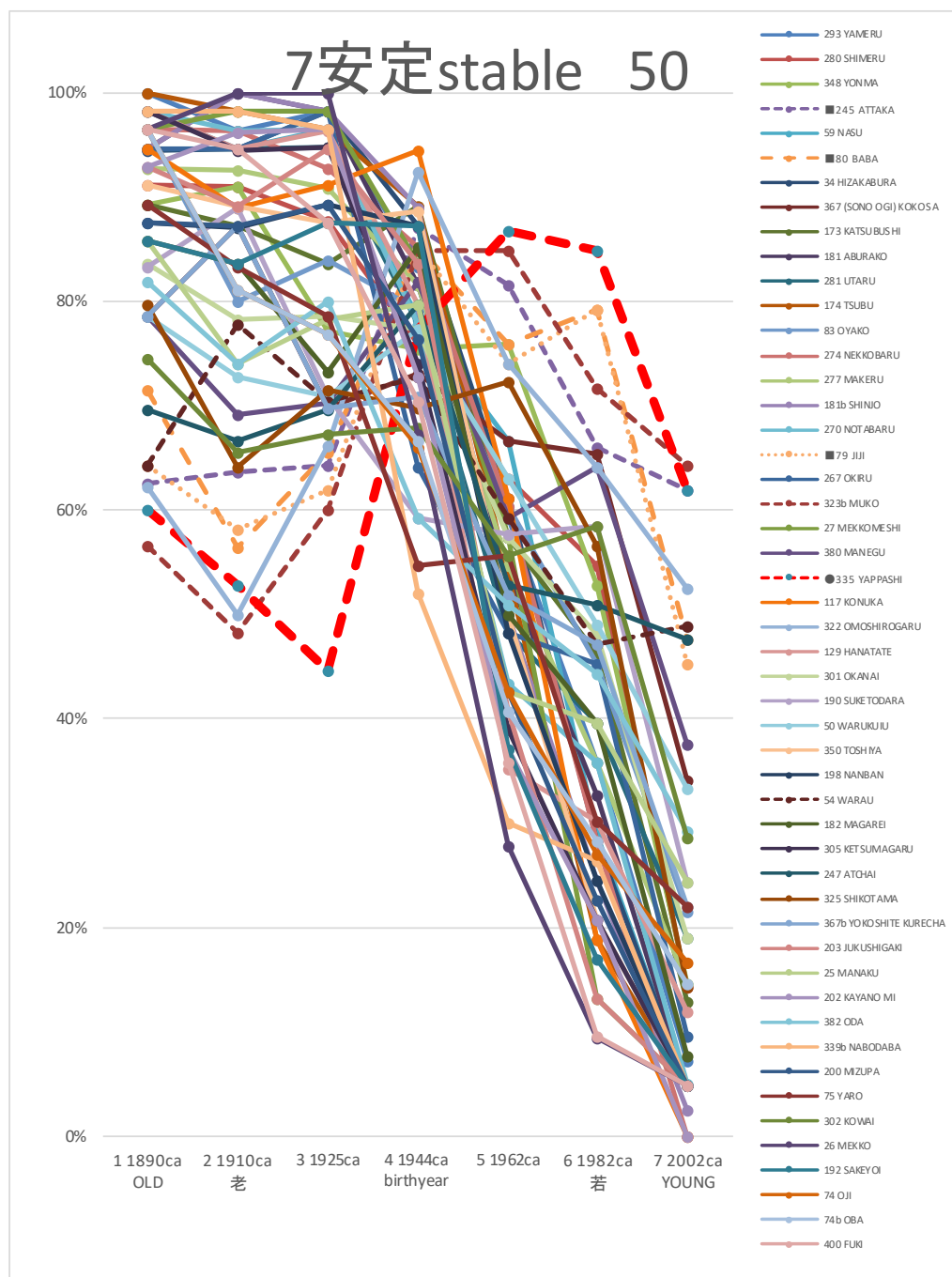


Figure 18-7. Generation and eight conditions: stable

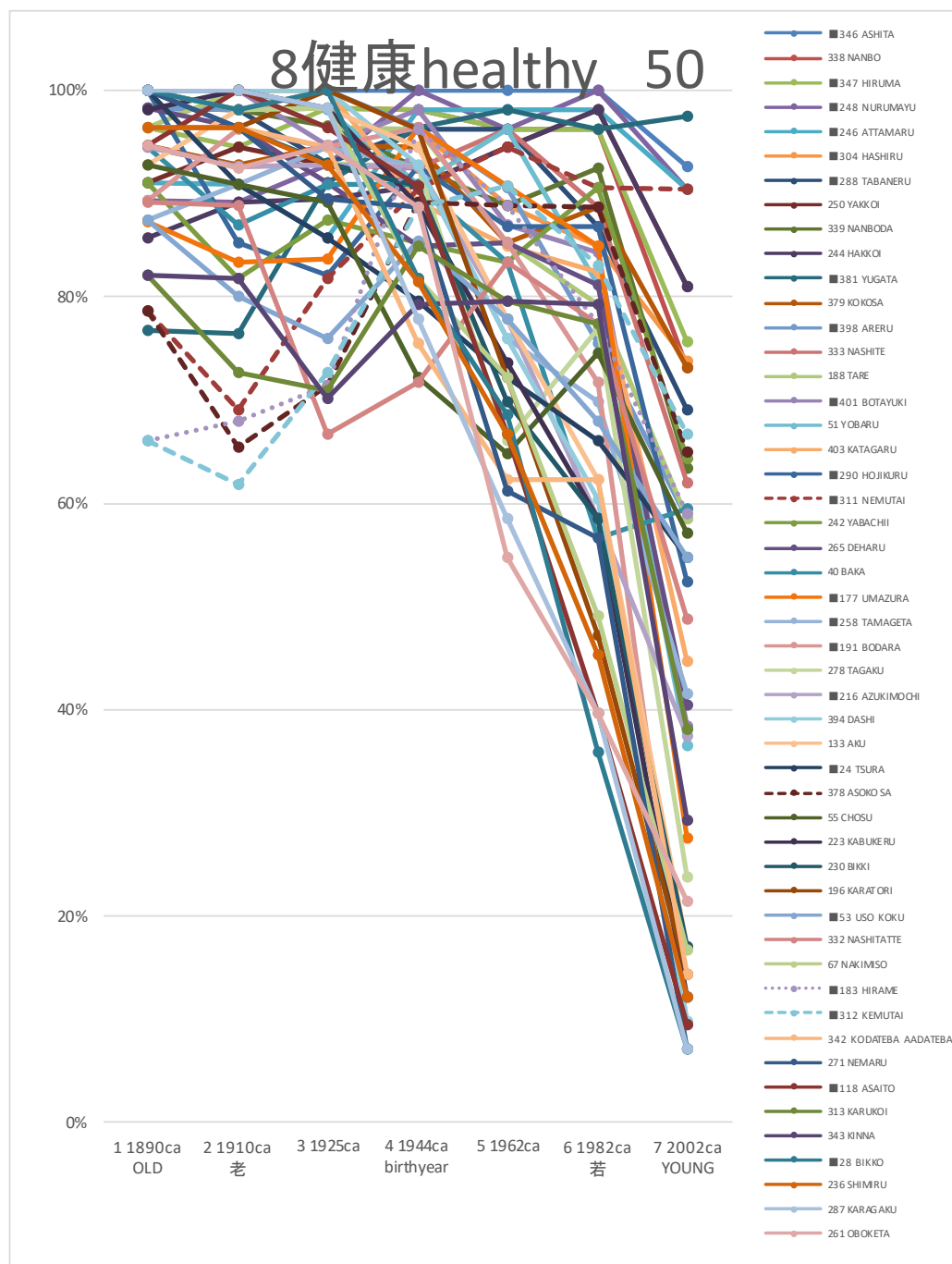


Figure 18-8. Generation and eight conditions: healthy

A continuous pattern is observed when seen in the order from 8 to 1. This order partly reproduces the historical process of decrease of vocabulary. The patterns of the graphs of eight medical conditions almost match internally within each graph and show a downward-sloping curve. A rapid decline in dialect over 140 years is displayed, but there are words showing exceptional (increasing) curves in the middle. In the graph of “1

dying", "Hamaogi" forms are hardly used even among the elderly respondents in the first survey born around 1890. In **"2 critical"** and **"3 serious"**, "Hamaogi" forms are disappearing, forming bundles. In **"4 bedridden"** "Hamaogi" forms are already on the decline in the three generations of the first survey. One word *chikkoi* 'small' is increasing exceptionally. In **"5 absolute rest"**, **"6 unstable"**, rapid decline in the middle generation is more conspicuous, but there are miscellaneous movements and there are some words that indicate increase. In **"7 stable"** most of the words show decline after the generation of birthyear around 1962. *Yappashi* 'after all' and some others show exceptional increase. In **"8 healthy"** many words were used vigorously in the first survey, but the lines show a sudden rapid decline in the fourth generations of the second survey, showing rapid standardization among the youth. Here, many words with ■ can be found.

Looking at the 7 generations within the eight graphs, the three generations of the first survey on the left half show small generation differences, and the four generations of the second survey on the right half show large generation differences. The rankings of the first and second surveys changed significantly. There were several patterns, such as words that were rapidly lost in elderly people in the second survey and words that were recently lost in junior high school students in the second survey. This corresponds with the diversity of the age difference curves in Figure 13 above.

This type of observation was possible by making use of simple arithmetic calculation and several studies have been done in the past.

5.2 MCA axis 1: age difference

In the following, the dimension of words from the MCA will be analyzed. In order to grasp the meaning of the axes, the axis value for each word is handled. Axes 1 and 2 are combined and plotted in Figure 19. In analyzing the words in order to examine the meaning of the axes, the distribution in each dialect map is the ultimate clue, and the summary line graphs of age in Figure 18-1 to Figure 18-8 are also convenient clues. The eight representative dialect maps in §2 are useful as representatives of words following the same trend. The alphabet labels of the eight representative words shown in Figures 2 to 9 (*a* to *h*) are added. We can also make use of the relationship with the result of person

(region, age) in §4. We put marks for eight medical conditions (or vitality) (§5.1) in the center positions of the plotted 406 word.

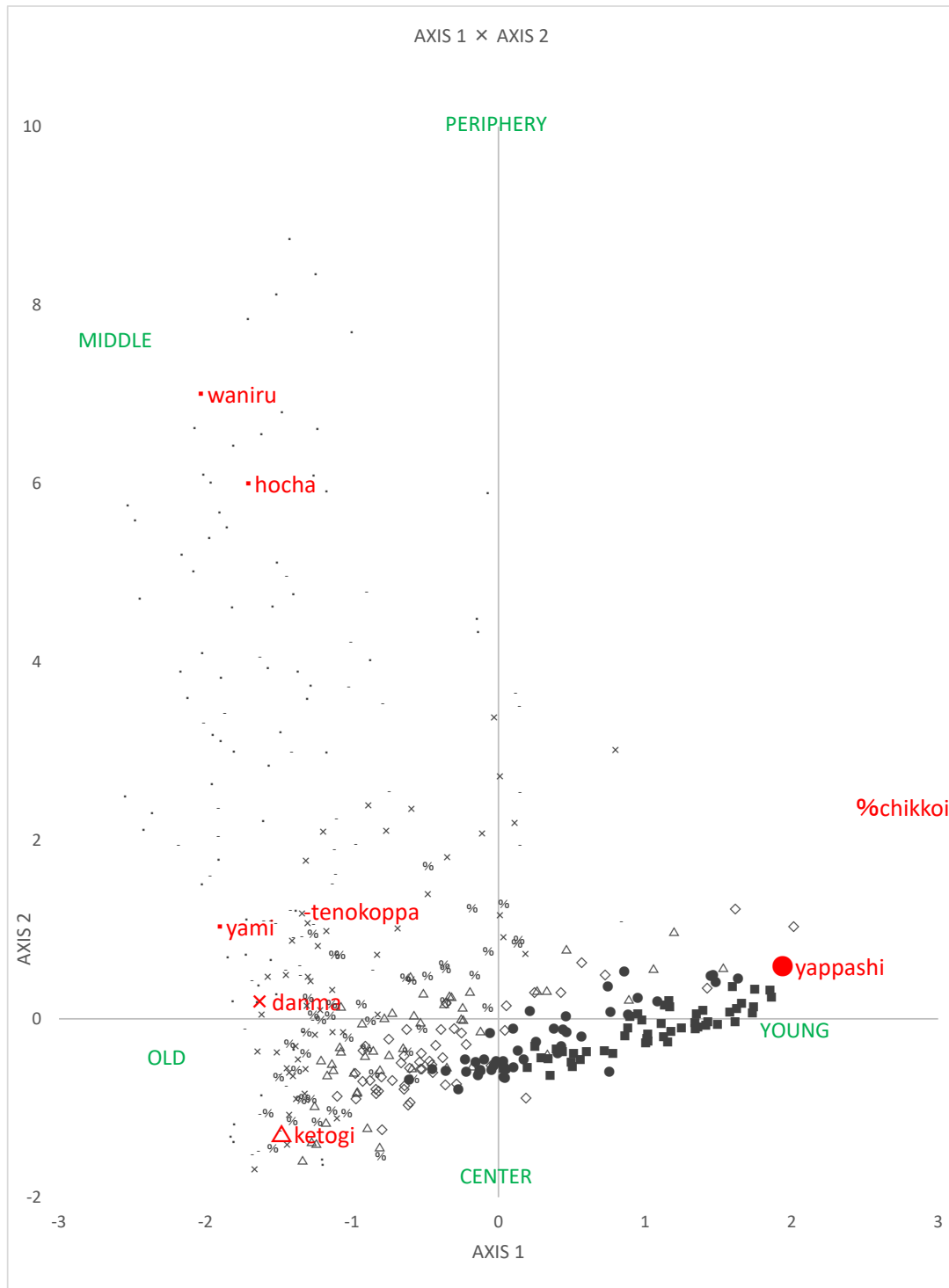


Figure 19. MCA axes 1 & 2 north-south difference

■8 healthy, ●7 stable, ◇6 unstable, Δ5 absolute rest,
 %4 bedridden, ×3 serious, -2 critical, • 1 dying

Figure 19 shows a horseshoe shape as a whole. From the experience of applying multivariate analysis, main general forms are plotted in the top part of the horseshoe close to the origin, and various distributions of the minority usage are scattered in the remote deviated parts. The results of multivariate analysis become usually dispersed in the case of a small number of responses.

As a whole, the combination of axes 1 and 2 corresponds with the eight medical conditions shown by marks. Healthy words with dark marks are plotted in the lower right, and words of slim marks with serious medical conditions are scattered in the upper left. The horizontal **axis 1** has a weak relationship with the medical condition, but healthy majority are plotted on the right (that is, words used by the younger generation), and weak minority are plotted on the left (used by the elderly). The vertical **axis 2** arranges the majority healthy forms below and the minority weak forms above.

Let us take a closer look. “8 ■ healthy” and “7 ● stable” are plotted in a line at the bottom right. Figure 18-1 and Figure 18-2 suggest that their age peaks are various. “6 ◇ unstable” and “5 △ absolute rest” are plotted linearly from right to lower left. Figure 18-3 and Figure 18-4 suggest that there may be peaks somewhere between old and young depending on the word. The “4 % bedridden” are plotted in a strip near the origin. “3 × serious” are plotted within a small area above the origin. “2 - critical” and “1 . dying” are scattered over a wide area on the upper left.

Looking at the distribution of the eight representative words in Figure 19 shown by large letters, “a *chikko*” and “b *yappashi*” are plotted on the right (axis 1 plus side) and “c *waniru*” and “d *hochu*” are plotted in the upper left corner (axis 2 plus side). At the bottom left (minus side of axis 1 and minus side of axis 2), “h *ketogi*” is plotted. “f *tenokoppa*”, “g *danma*” and “e *yami*” are plotted on the left-hand side. From this, it can be inferred that horizontal **axis 1** is related to the age difference and to the spread among the younger generation. This is consistent with the observation in §4 that axis 1 reflects age differences. In Figure 19, the right “a *chikko*” and “b *yappashi*” are the words prevalent among younger people, and the left “c *waniru*” and “d *hochu*” are the words used by elderly people. The order of remnant rate described in §5.1 is similar to the

arrangement from the positive side to the negative side of axis 1 from MCA. Axis 1 showed the difference in age, but the words that are often found in the younger age group are naturally higher in the older age group because this is the follow-up survey of the old dialect glossary. Therefore, **axis 1** can be interpreted as representing age peaks of remnant rate (from the words that are common to all age groups to the words that remain only in the elderly).

When the maps of about 400 words were arranged in the order of values of axis 1, they showed not only age differences but also a historical process of diffusion. Approximately 400 words were arranged in the order of the MCA's axis 1, and movement in a set direction was observed. They showed a continuous movement like a movie or a video, and showed what the array of words in Figure 19 represents.

(1) The old process of “Hamaogi” forms spreading from Tsuruoka to the whole of the Shonai district,

(2) then the process of spreading of other word forms,

(3) the process of “Hamaogi” forms remaining in distant peripheries,

(4) the process of increasing age difference due to rapid standardization, and remaining among the elderly.

In short, we were able to present the process of the decline of old dialect and the development of standard language in a reconstructed form. When the maps of about 400 words were arranged in the order of remnant rate, continuous movement was observed. They exhibited how multivariate analysis is processed. It can be presented on slides for oral presentations at academic conferences. Since Kumagai (1993), many videos have been presented.

In the word dimension, a few notable exceptional words such as “a *chikko*” and “b *yapashi*” were found that continue to increase in modern times. If we look at the 50 words that are still healthy in Figure 18-8, we can see the words that will continue to be used in the future. Some of the “Hamaogi” forms survive. The key is their matching with the colloquial speech and slang of Tokyo (Inoue 2000.2, 2020.2), as shown by ■ in Figures 18-6 to 18-8. On the other hand, words related to pre-modern life and culture had already disappeared in the elderly of the first survey as shown in Figure 18-1. Also, compared to the nationwide dialect distribution, words with a narrow geographic

distribution only in the Shonai district declined quickly (Inoue, & Hanzawa in review). Words limited to Yamagata Prefecture are also disappearing. However, the words that are widely distributed in the Tohoku region and eastern Japan have survived. To see the fate of each word, it is necessary not only to look at the survey results of Shonai district, but also to position them in the distribution of Japan as a whole (Hinskens 1998).

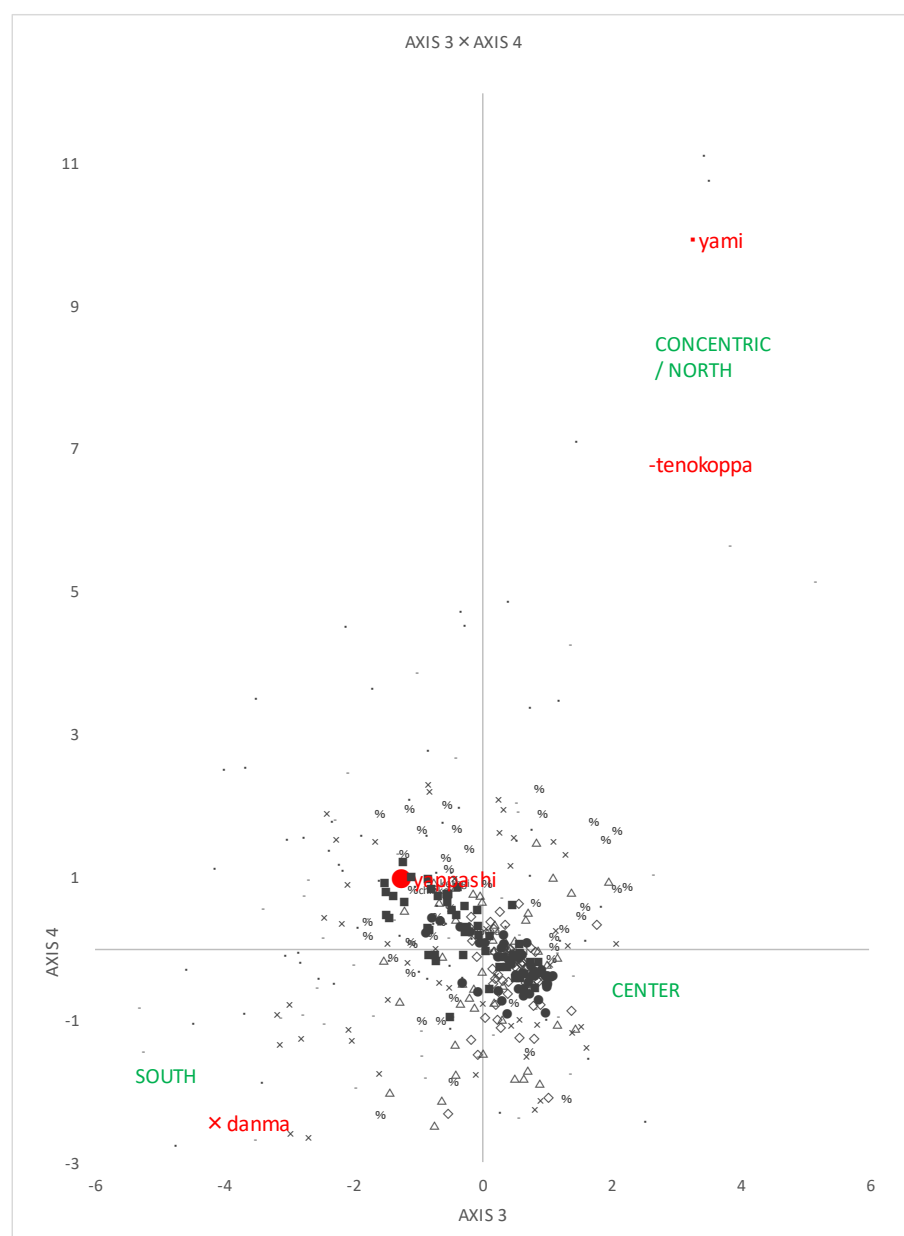


Figure 20. MCA axes 3 & 4 north-south difference

■8 healthy, ●7 stable, ◇6 unstable, △5 absolute rest,
%4 bedridden, ×3 serious, • 1 dying

5.3 MCA axis 2: regional distribution

On the other hand, vertical **axis 2** of Figure 19 is considered to be the district internal classification of dialects by geographical distribution. The marks for the eight medical conditions show a close relationship, with dying and critical words taking large positive values, and healthy stable words taking negative values. Thus, axis 2 can be interpreted to show an inverse relation to vitality or vigorousness of “Hamaogi” forms.

“c waniru” and “d hocha” are plotted in the upper left, taking large positive values and indicating their scattered remnant distribution in the north. “e yami”, “f tenokoppa” and “g danma” are plotted in the left half, and also show scattered distribution mainly in the north. It can be inferred that the words showing regional concentric distribution have been gathered here. Returning to Figures 4 and 5, concentric distribution is shown, as we saw in §4. (They show large values on the axes 3 and 4 in Figure 20.) In contrast, “h ketogi” which is plotted on the top of the horseshoe distribution in lower left showed prominent age differences and almost no geographical differences.

5.4 MCA axes 3 and 4: North-South difference

Next, we will consider axes 3 and 4. Reviewing §4, axes 3 and 4 show a pattern in which old words remain in the north due to a contrast between the north and the rest. In Figure 20, the horizontal axis is axis 3, and the vertical axis is axis 4. Most words are heavily concentrated around the origin. Words with large absolute values (with special, various distribution) are scattered in all directions. In the upper right corner (both the axes 3 and 4 are plus), there are “e yami” and “f tenokoppa”. They were distributed to the north (Figures 6 and 7). In the lower left corner (both axes 3 and 4 are minus), there is “g danma” which is used in south (in Figure 8). The two axes both show the north-south difference. (“f tenokoppa” does not take large values in Figure 19 below, and are buried in many words in the lower left. It was extracted as a sub-class on the axes 3 and 4.)

Axes 3 and 4 characterize words which did not show specific tendencies in axes 1 and 2 of Figure 19. Many words concentrated in the middle of Figure 20 do not show a

characteristic tendency. Of the remaining five words, “a *chikko*”, “b *yappashi*” and “h *ketogi*” are largely concentrated at the upper left of the origin (in Figures 2, 3 and 9 they spread to the whole area). “c *waniru*” and “d *hocha*” are mixed among many in the upper right corner of the origin (in the north in Figures 4 and 5). Since “g *danma*” is a special distribution pattern, a large negative value was shown for axes 3 and 4 and plotted in the left bottom. Other words plotted in the lower left of Figure 20 are used near Tsuruoka and show a similar geographical distribution pattern. In Figure 20, “h *ketogi*” was buried among many, near the origin. The fact that there are many words around this in Figures 19 and 20 indicates that there are many words having similar distribution. The size of the age difference and the rapid development of standard language throughout Shonai are suggested. As a result of applying the MCA in §4, it was found that the age difference was large.

We will examine the relationship with the eight medical conditions (or vitality) using the markers in Figure 20. As a whole, healthy words were plotted near the origin and words with severe medical conditions were plotted in peripheries as concentric circles. “8 ■ healthy”, “7 ● stable” lump near the origin. “6 ◇ unstable”, “5 △ absolute rest” are plotted around them. “4 % bedridden” is mostly in the upper right. “3 × serious” is mostly in the lower left. “2 - critical” and “1 . dying” are scattered over a wide area. This is because the number of answers is small. The **axes 3 and 4** subdivide the words distributed at limited areas, and the upper part is the north and the lower part is the south.

5.5 Summary of word forms

Above, we have considered the characteristics of each word. Based on the results of MCA, we picked up eight words that exhibit representative geographical and age distribution. It was confirmed that eight words concisely represent the entire tendency of 406 words. The words plotted nearby in Figures 19 and 20 show similar geographical and age distributions in the age pillar maps. It was also possible to comprehensively position the entire 406 words in Figures 19 and 20. It was thus found that the arrangement of

words by the values given by MCA faithfully correspond to the geographical and age distribution in the dialect map.

In conclusion, the meaning of the axes (Table 1) can be summarized by contrasting with the individual age pillar maps above. **Axis 1** corresponds to age difference. The positive side reflects the colloquial or slang forms of Tokyo, often found among the younger generation. The negative side is a pattern displayed by the declining words remaining only among the elderly. **Axis 2** corresponds to regional differences, showing concentric distribution within the Shonai district. This is a pattern of “Hamaogi” forms being pushed away by other forms and kept to places where traffic is inconvenient. It can be interpreted that this geographical decline process has been a major trend in dialects since pre-modern times. **Axes 3 and 4** are for further detailed analysis of small regional differences between the north and south.

5.6 Age difference

The ultimate conclusion of MCA is the large age difference. The age difference of 140 years worked and whole data showed a rapid decline curve. In the past, regional differences were conspicuous within the Shonai district. At present, the difference between generations stands out as the district internal difference in the Shonai district. The “Hamaogi” words are rapidly lost. The large age difference is not only due to the regional (small) scale of Shonai district. In Japan as a whole, there is a large age difference of over 100 years from the LAJ informants (born at the end of the 19th century) to current children who were born more recently (Inoue 2020.2, 2020.12). It can be said that language standardization in Japan is just a short distance from completion.

5.7 Possibility of regional concentric distribution

Looking at the distribution pattern of the entire Shonai district, newer forms spread from Tsuruoka and Sakata, diffused to the north and south, and the words 250 years ago remain in the surrounding peripheral areas. As a result of examining the MCA results by the geographical distribution of individual words, it was possible to interpret that a

double concentric distribution at the regional level was established under the nationwide concentric distribution (Inoue 2000.2). If one of the north and south of the regional distribution is missing, the north-south difference results, so both (regional concentric distribution and north-south contrast) are continuous. It is the same as Japan as a whole.

The regional distribution of Shonai can be interpreted metaphorically to be a microcosm of the Japanese archipelago. The administrative city Tsuruoka is Tokyo and the commercial city Sakata is Osaka (Tobishima island is Okinawa). Just as in LAJ, there was a contrast between Tokyo and Kyoto-Osaka (Inoue 2010.12), and in Shonai there was a contrast between Tsuruoka and Sakata.

Instead of a single decline of the “Hamaogi” forms, there were four changes. One is vanishing without alternatives: e.g. words for pre-modern tools and customs (Inoue 2013.12), the other is replacement by alternative forms, most of which are standard language forms (Inoue 2008.5), the third is new dialect forms and the last is old dialect forms kept because of match with Tokyo colloquial. Before the overwhelming power of the standard language, regional dialect forms became powerless, but the new dialect forms were born even after modern times (Inoue 1993.2, 2016.8).

6. Discussion: from geography to generation

6.1 Make data talk about internal structure

In this paper, we proceeded from the consideration of individual dialect maps to synthesis by multivariate analysis. All words of all respondents were subjected to multivariate analysis to clarify the internal structure of the data: the dimension of person and the dimension of word. 250 years have passed since the compilation of “Hamaogi”, and we tried to grasp the age difference in dialect usage spanning 140 years and the regional difference in Shonai district as a whole. The age difference of 140 years is larger than the regional difference of about 80 km north and south distance of the Shonai district. Application of multivariate analysis showed that the general pattern changed

from regional difference to age difference due to standardization. As for the regional differences, diffusion of dialect from the central city and its remaining in the surrounding areas were confirmed.

6.2 *From regional to age differences in simplified glottogram*

Walking distances from Tsuruoka to 27 survey points in premodern feudal ages were calculated on the basis of an old map in Feudal ages (1861) making use of “curvimètre”, and 27 survey points were arranged by the order of distance, thus constituting a “simplified glottogram”. These simplified glottograms were found effective in showing overall trends of dialect differences, though geographical locations cannot be shown faithfully. Examples for the eight representative words in §2 are used in Inoue & Hanzawa (in press).

The general trend in the data discussed in this paper is presented in a simplified glottogram of Figure 21. The horizontal axis represents age and the vertical axis represents geography. The basic data is a little different from Figure 10. Responses of “use” were set to 1, while “hear” and “don't know” were set to 0, and the remnant rate was subdivided in 10 steps in 10% increments. Increments exceeding 50% are represented by solid black symbols while those falling short of 49% are represented by non-solid symbols with more white. The survey points are arranged from north to south according to walking distance on the basis of an old map. This figure shows local concentric distribution. In the 19th century, the old Hamaogi forms remained in the south and north ends of the research area, showing a difference from the center, through which new phenomena had been diffusing. In the 20th century, especially after WWII, the remnant rates became smaller and the regional differences diminished. Instead, generational difference became conspicuous. Regional differences will further diminish in the 21st century, mainly because of the rapid spread of standard language.

As for the dimension of person, there was a rapid standardization in the generation born in the 1940s. This time-point is similar to the result of four surveys of language standardization in Tsuruoka (Yoneda 1997, Inoue 2010.1, 2020.12). However, the standardization of pitch accent was delayed (Yokoyama et al. 2018). The standardization

of grammar has also been delayed, and the spread of new dialect was observed. Regarding the time of change based on age difference, it is necessary to consider a wider area which includes Shonai. (Note that factors such as the subtle differences among different types of Hamaogi forms and the spread of new dialect forms are omitted from Figure 21.)

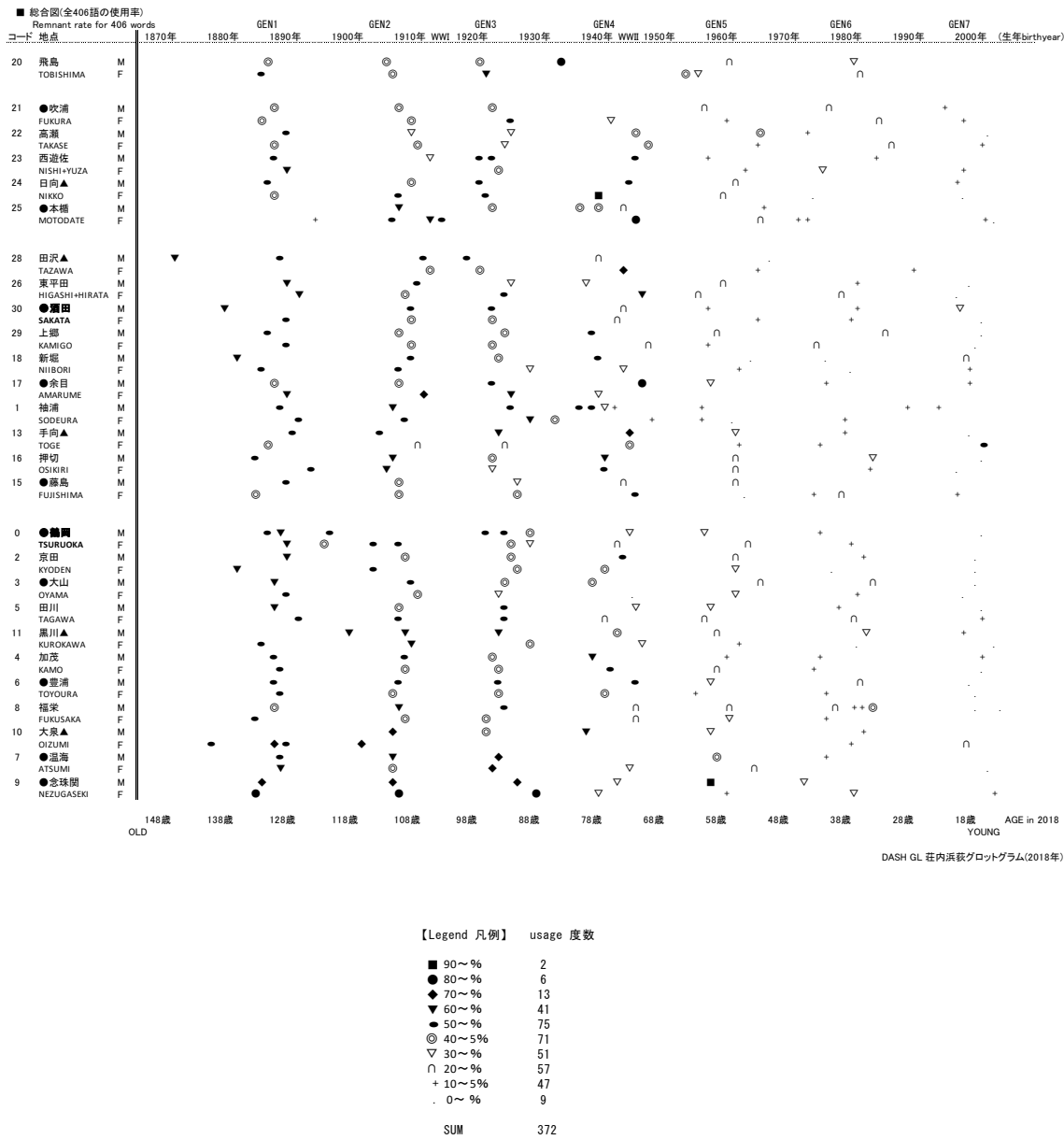


Figure 21. Simplified Glottogram of remnant rate

The basic tendency of the regional difference in the Shonai dialect is the north-south contrast observed in axes 2, 3 and 4. However, since the difference is a continuum and has no fault lines, it is almost impossible to set dialect boundaries based on vocabulary. The geographical and age distribution of each word indicates a continuum. From a different perspective, the many Figures as a whole showed the overall tendency of change and diffusion. The change progresses gradually, including the random process, depending on the word and depending on the person.

Let us review the above results from a different viewpoint. The 140-year generation gap, together with the 1950 three-generation survey and the 2018 four-generation survey, was analyzed. The result of multivariate analysis gave a different value from the usual glottogram of the linear region with age difference (time depth) of about 60 years. Since the geographical spread was also investigated, the “age pillar map” can be interpreted as a 3D (three-dimensional) glottogram (or multi-layered dialect map).

From an extralinguistic point of view, factors such as interaction with the central city, transportation routes, industrial structure, and the movement of inhabitants are important. Even for residents at the same location, many more factors should be taken into account, including age, gender, occupation, geographical/social activity, social network, status attribution etc. Furthermore, as personal psychological factors in interaction, the scale of accommodation, identity and belief also influence dialect use. However, it is difficult to extract the functions of other factors in the population with large geographical and age variations. In order to prove without a doubt, objective analysis with strict control of other conditions is necessary. The Tsuruoka Survey (Yoneda 1997, Yokoyama et al. 2018) is the most suitable for this research, and it is necessary to carry out related research.

6.3 *Time-space model*

The major trend in the data discussed in this paper is concisely visualized in Figure 22. The horizontal axis represents time and the vertical axis geography. This *time-space model* or *age-area model* can be applied to nationwide spread of standard language in

Japan, and also in many nations in the world. It can also be applied to the spread of English language all over the world.

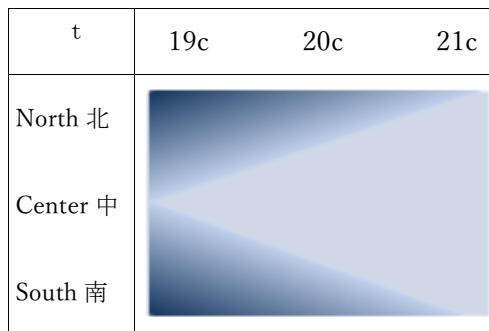


Figure 22. Time-space model of dialect change

6.4 Dimension of vocabulary

With regard to the dimension of vocabulary, more than 400 words were analyzed at once. Thanks to the application of multivariate analysis for all survey items, it was possible to process all the words as groups rather than treating them separately. There was a big difference according to age, but we also found words that young people continue to use today and words that are increasing now, rather than simply having a simple continuous decline between young and old. Words that match the colloquial speech and slang of Tokyo will survive.

If we look at the process of linguistic change for each word, the semantic field, the time of disappearance of the real object, the frequency of use, the scene (or *domain*) of use, size of distribution area etc. will affect their use. Adding information such as semantic field and national distribution to the words in the survey data can reveal general principles applicable to other regions (Inoue & Hanzawa 2021.4).

6.5 Continuum of language change

In this paper, the continuum of the Shonai dialect is clarified. Clear boundaries or fault lines were not found for regional differences (Figures 14-17) or age differences (Figure 12). It can be considered that the change progressed and diffused at an almost constant rate for 250 years in the entire Shonai district. Even when the multivariate analysis was applied, no significant difference was sorted from inside the data. However, historically, there are periods of rapid change and periods of gradual change, and it is commonly accepted that there are areas with large dialect differences and areas with small dialect differences.

Let us review the data with that perspective. If we assume that there is a certain law or rule, the exception become visible. In Figure 18, the graph was divided into eight medical conditions. Of these, 50 items of eight medical conditions showed a process in which Hamaogi forms were rapidly lost in the younger generation of the second survey. The Hamaogi forms that kept life until recently are no longer used by the youngest population. A *diglossia* should be postulated for this area. In the past, the “High” standard language and the “Low” dialect were used properly according to domains. Now, the process of becoming monolingual of the “High” standard language is observed among young people. In other words, the decline of dialect and the spread of standard language do not always progress at a constant rate. Depending on the time, the words of a group can change all at once in a group in a certain area. Of the 250 years since the compilation of Hamaogi, there must have been a rapid disappearance of vocabulary due to lifestyle changes since the modernization of 1868. The 140-year data in this paper changed significantly in response to changes in lifestyle during the economic growth period of the 1970s. Especially, the spread of television had great significance. The change in pitch accent in Nagano Prefecture is related to the spread of television. The major changes in the generation born in the 1940s, in wartime (Figure 12), are worthy of reassessment. However, we do not think standardization happened all over the country at the same time. Until recently, the northern part of the Tohoku region was said to be a “dialect mainstream society”. It is also believed that the standard language gradually diffused from the vicinity of Tokyo according to the railway distance (Inoue 2010.7).

Regarding vocabulary, there was a regional difference corresponding to the railway distance in LAJ (informants born in the 1890s), but the regional difference almost disappeared nationwide in junior high school students (born in the 1980s) (Inoue 2020.2). At some point there was a dialect decline process. A large-area glottogram survey will serve as a clue. We hope new data by age in each region will be collected. Slow diffusion is observed according to the distance (Inoue 2003.7), i.e. the double concentric circle theory holds (Inoue 2016.8. p.87). The “cascade model” from city to medium city applies (Labov 2002) and Christaller's central place theory of geography is also applicable.

6.6 *Future of dialect and dialectology*

The analysis by age group reveals the dialect usage patterns of modern youth. On the basis of these dialect words, we can predict the future. New dialect forms have been created in Tsuruoka and Sakata, which are also influenced by the prefectural capital Yamagata City. This new dialect can be a regional *dialect levelling* (Hinskens 1998, Kerswill 2003). Moreover, the close relationship between dialect, geography and society was deduced above. This “Hamaogi” survey enhances the status of Tsuruoka for fixed-point observation both by expanding the survey area and by historically extending to 250 years in the past since the compilation of the dialect glossary. Research on social dialects, dialect geography, and glottograms addressing the Shonai district has been accumulated.

The linear glottogram had the disadvantage that the distribution and diffusion of the two-dimensional earth surface cannot be captured. This study tried to overcome this, and examined the entire Shonai district as a surface to add to the age difference. This analysis has made it possible to elucidate the mechanism of the decline of dialect vocabulary and the progress of new dialect and standard language.

This survey did not take into account situational use or bidialectalism. However, the “Hamaogi” survey questionnaire also has another column for the actual usage of words, and entries of multiple word usage is observed. We will soon analyze these entries. This will allow consideration to be given to difference by domains. Regarding modern bilingualism and diglossia, the language shift among immigrants usually occurs from the

1st to the 3rd generation. Emotional return to the ancestral language among the 3rd generation is also reported. An attempt to divide the historical status of Japanese dialects into “eradication, description, and entertainment” (Inoue 1997.10, 2011.1) can be regarded in parallel with the language shift from the 1st to the 3rd generations. There is a dialectal revival movement. The concrete examples are the economic utilization of dialects as shown by dialect landscape and dialect souvenirs (Inoue 2012.1). These examples are also observed in the Shonai dialect.

By faithfully representing individual answers in “age pillar maps”, we were able to study the relationship between words and humans, and the underlying psychology, geography, society, and economy. Dialect geography was denounced in the past as atomistic because of its limited viewpoints. It was rarely positioned within the broad perspective of the humanities and social sciences. In order to approach the general theory, it is necessary to comprehensively deal with a large amount of data and grasp the big overall trends. One such method is to simplify complicated phenomena and find rules, tendencies and principles. The multivariate analysis adopted in this paper is one realization of this.

Acknowledgments

In the Hamaogi survey, we asked the local board of education, community center, and other organizations to find the respondents. The respondents kindly filled out the survey form. We would like to express our deepest gratitude. This research was supported by the following scientific research funds. JSPS Grant-in-Aid for Scientific Research, Basic Research (B) 16H03420, Fumio Inoue “Sociolinguistic Comprehensive Research on Regional Differences of Public Language”.

References

* English papers by F. Inoue before 2008 are accessible through the internet.

https://dictionary.sanseido-publ.co.jp/affil/person/inoue_fumio/index_eng.html

<http://innowayf.net/index-e.html>

<http://urayasu.meikai.ac.jp/japanese/meikainihongo/18ex/default.htm>

AITCHISON, Jean (1991)² *Language change: progress or decay?*, Cambridge: Cambridge University Press.

ANG, Uijin (2019) *Studies on Social Language Geography of Taiwan I II*, Taipei: Avant-Garde Publisher.

AURREKOETXEA, Gotzon et al. (2017) "Principal components and clustering on lexical distances from standard Basque to local varieties", *Journal of Research Design and Statistics in Linguistics and Communication Science*, vol 3 (1), 5-22.

BLOOMFIELD, Leonard (1933) *Language*, New York: Holt.

BRITAIN, David & Jenny CHESHIRE (eds.) (2003) *Social Dialectology in Honour of Peter Trudgill*, Amsterdam/Philadelphia: John Benjamins.

www.ling.upenn.edu/~wlabov/Papers/PCM.html

CHAMBERS, J. K., Peter TRUDGILL & Natalie SCHILLING-ESTES (2002) *The Handbook of Language Variation and Change*, Oxford: Basil Blackwell.

CHAMBERS, J. K. & P. TRUDGILL (1980) *Dialectology*, Cambridge: Cambridge University Press.

GOEBL, Hans (2010) "Dialectometry: Theoretical prerequisites, practical problems, and concrete applications (mainly with examples drawn from the *Atlas Linguistique de la France*, 1902-1910)", *Dialectologia*, special issue, I, 63-77.

<http://www.edicions.ub.edu/revistes/dialectologiasp2010/>

GORMAN, Kyle (2009) "On VARBRUL – Or, The Spirit of '74", University of Pennsylvania.

<http://ling.auf.net>

GRIMM, Jakob (1819) *Deutsche Grammatik I*, Gütersloh: Bertelsmann.

HANZAWA, Yasushi (2017) "Gurottoguramu chosa deta no jitsujikan hikaku (Real-time comparison of glottogram survey data)", in Takuichiro Onishi (ed.), *Dialect in Space and Time*, Tokyo: Asakura Shoten, 283-303.

HANZAWA, Yasushi (2018) "Gendai ni okeru hogen goi no dotai (Dynamics of dialect in modern

- times)", in Takashi Kobayashi (ed.), *Vocabulary of Dialects Tokyo*: Asakura Shoten, 132-147.
- HANZAWA, Yasushi (2020) "Fukushima-ken ni okeru r datsuraku gensho no denpa to henka (Diffusion and change of r deletion in Fukushima prefecture)", *Kokugogaku kenkyu*, 59, 259-273.
- HANZAWA, Yasushi (2021) "Shonai Hamaogi chosa deta no taehenryo kaiseki" (Multivariate analysis of research data of "Shonai Hamaogi"), *Fukushima daigaku ningen hattatsu bunka gakurui kiyo*, 33, 49-64.
- HAYASHI, Chikio (2004) "*Hayashi Chikio chosakushu 3 shitsu o hakaru suryoka riron (Collection of Hayashi Chikio, vol.3, Theory of Quantification Measuring Quality)*", Tokyo: Bensei Shuppan.
- HEERINGA, Wilbert (2004) *Measuring Dialect Pronunciation Differences using Levenshtein Distance*, Groningen Dissertations in Linguistics, 46.
- HEERINGA, Wilbert & John NERBONNE (2001) "Dialect areas and dialect continua", *Language Variation and Change*, 13(03), 375-400.
- HIDAKA, Chiemi (2019) "3D gengo chizu o mochiita chugoku kososho joshu hogen no shakai-gengo-chirigakuteki kenkyu (Social linguistic geographical study using 3D linguistic maps in the Xuzhou dialect of Jiangsu Province, China)", *Nihon chirigengo gakkai yokoshu (Proceedings of the Geolinguistic Society of Japan)*, 1, 72-76.
- HINSKENS, Frans (1998) "Dialect levelling: A two-dimensional process", *Folia Linguistica Historica*, 32, 1/2, 35-51.
- HONDA, Masashito (2005) "Heiko suru gurotoguramu to heimenbunpuzu ni yoru gengo shinnyu no rittaiteki haaku (Three-dimensional grasping of language invasion by parallel glottograms and plain distribution maps", *Iwaki Meisei Daigaku Daigakuin Jinbunkagaku Kenkyuka Kiyo (Bulletin of Graduate School of Humanities, Iwaki Meisei University)*, 3, 51-62.
- HUANG, He (2019) "Exploring the diffusion models of lexis by glottogram: A pilot study in Yixing Wu dialects", *Nihon chirigengo gakkai yokoshu (Proceedings of the Geolinguistic Society of Japan)*, 1, 113-118.
- INOUE, Fumio (1985.2) *Atarashii Nihongo: "shin hogen" no bunpu to henka (New Japanese — distribution and change of "new dialect")*, Tokyo: Meijishoin.
- INOUE, Fumio (1986.6) "Sociopsychological characteristics of users of 'New dialect forms'", *Journal of Pragmatics*, 10 (3), 327-345.
- <http://www.tufs.ac.jp/ts/personal/inouef/inoue-e-1986-06b.pdf>
- INOUE, Fumio (1993.2) "The significance of new dialects", *Dialectologia et Geolinguistica*, 1, 3-27.
- INOUE, Fumio (1996.3, 1996.10) "Computational dialectology (1) (2)", *Area and Culture Studies*, 52, 67-102; 53, 115-134.

- INOUE, Fumio (1997.10) "Market value of languages in Japan", *Japanese Linguistics*, 2, 40-61.
- INOUE, Fumio (1998.1) *Nihongo Uotchingu (Japanese Language Watching)*, Tokyo: Iwanami shoten.
- INOUE, Fumio (2000.2) *Tohoku Hogen no Hensen (Transitions of the Tohoku Dialect)*, Tokyo: Akiyama Shoten.
- INOUE, Fumio (2001.2) *Keiryoteki Hogen Kukaku (Quantitative Dialect Classification)*, Tokyo: Meiji Shoin.
- INOUE, Fumio (2003.7) *Nihongo wa Nensoku 1 kiro de Ugoku (Japanese Moves at a Speed of 1km per year)*, Tokyo: Kodansha.
- INOUE, Fumio (2004.7) "Multivariate analysis, geographical gravity centers and the history of the standard Japanese forms", *Area and Culture Studies*, 68, 15-36.
<http://repository.tufts.ac.jp/bitstream/10108/23707/1/acs068002.pdf>
- INOUE, Fumio (2008.5) "Geographical distance center and multivariate analysis of the standard Japanese". *Dialectologia*, 1, 65-81.
<http://www.publicacions.ub.es/revistes/dialectologia1/>
- INOUE, Fumio (2009) *Shonai Hogen Chizu (Linguistic Atlas of Shonai Hamaogi)*
<http://www.urayasu.meikai.ac.jp/japanese/inoue/LASD>
- INOUE, Fumio (2010.1) "S-shaped curve of phonological standardization — six surveys in Tsuruoka and Yamazoe areas —", *Slavia Centralis*, III/1, 101-112.
- INOUE, Fumio (2010.7) "Real and apparent time clues to the speed of dialect diffusion", *Dialectologia*, 5, 45-64
<http://www.edicions.ub.edu/revistes/dialectologia5/>
- INOUE, Fumio (2010.12) "Gravity model of diffusion for Tokyo new dialect forms", *Leeds Working Papers in Linguistics and Phonetics*, 15, 92-100.
<http://www.leeds.ac.uk/linguistics/WPL/WPL15.html>
- INOUE, Fumio (2011.1) "Standardization and de-standardization processes in spoken Japanese", in Patrick Heinrich & Christian Galan (eds.), *Language life in Japan: Transformation and prospects*, Abingdon and New York: Routledge, 109-123.
- INOUE, Fumio (2012.1) "Improvements in the sociolinguistic status of dialects as observed through linguistic landscapes". *Dialectologia*, 8, 85-132.
<http://www.edicions.ub.edu/revistes/dialectologia8/>
- INOUE, Fumio (2013.4) "A contemporary history of Okazaki honorifics: Democratization and 'te-itadaku'", *Working Papers from NWAV Asia-Pacific*, 2, 1-9.

<http://pj.ninjal.ac.jp/socioling/nwavap02/Inoue-NWAVAP2-2013.pdf>

INOUE, Fumio (2013.12) "Dialect lexicography and the standard language — Words for snow and suburban Tokyo dialect —", *Dialectologia*, special issue, IV, 93-121

<http://www.publicacions.ub.edu/revistes/dialectologiaSP2013/>

INOUE, Fumio (2016.8) "A century of language change in progress — New dialect in Tsuruoka—". *Dialectologia*, 17, 71-89.

<http://www.publicacions.ub.edu/revistes/dialectologia17/>

INOUE, Fumio (2017.5) "Observation of linguistic change in progress through real time comparison of glottogram data", Vugar Sultanzade et al. (eds.), *VIII. Congress of International Society for Dialectology and Geolinguistics*, Famagusta: Eastern Mediterranean University Press, 31-43.

INOUE, Fumio (2017.8) "Age-area distribution of linguistic change in progress observed in glottograms", in Eveline Wandl-Vogt & Amelie Dorn (eds), *Dialekt / Dialect 2.0 Langfassungen / Long papers*, Wien: Praesens Verlag, 174-194.

INOUE, Fumio (2019.12) "Common developments of Japanese and Dutch-Flemish dialectology — Computational dialectology in Japan —", *Dialectologia*, special issue, 8, 35-64.

<http://www.edicions.ub.edu/revistes/dialectologiasp2019/>

INOUE, Fumio (2020.2) "Standard language distribution in LAJ and railway distance — Geographical and historical interpretation of limestone cave model —", *Dialectologia*, 24, 111-156.

<http://www.edicions.ub.edu/revistes/dialectologia24/>

INOUE, Fumio (2020.12) "Dialect vocabulary changes over 140 years — Standardization and new dialect forms observed in Hamaogi glossary —" *Dialectologia et Geolinguistica*, 28, 105-116.

INOUE, Fumio & Akemi YAMASHITA (2014) "Change in the use of beautifying "o-" and late adoption: A historical interpretation of data one decade apart", *Bamberg Studies in English Linguistics*, 57, 315-326.

INOUE, Fumio & Hanzawa YASUSHI (2019) *Dialect Atlas of Shonai Hamaogi (DASH)*, CD private edition.

INOUE, Fumio & Hanzawa YASUSHI (2021.4) "Hogen no chiikisa kara neneisa e --- Shonai Hamaogi tsuiseki chosa no taju taiou bunseki (From geographical differences to age differences of dialect — Multiple correspondence analysis of Shonai Hamaogi)", *Gengo Kenkyu*.

INOUE, Fumio & Hanzawa YASUSHI (2021.2) "Hogen goi no chiikisa to nenreisa — Shonai Hamaogi chosa 2 kai no taju taiou bunseki (Regional and age differences of dialect vocabulary — Multiple correspondence analysis of two surveys of Shonai Hamaogi —)", *Chiiki Sozo*, 32(2), 5-24.

INOUE, Fumio & Hanzawa YASUSHI (in review) Hogen suitai no goironteki katei — Shonai Hamaogi

no 250 nen — (Lexicological process of dialect decline — 250 years of Shonai Hamaogi —),
Nihongo no kenkyu, 17(1), 1-18.

JABERG, Karl (1908) *Sprachgeographie*. Aarau (p.6 p.22).

<https://archive.org/details/sprachgeographi-00reesgoog/page/n37>

JESZENSZKY, Péter, Yoshinobu HIKOSAKA, Satoshi IMAMURA & Keiji YANO (2019) Japanese lexical variation explained by spatial contact patterns, *International Journal of Geo-Information* 8(9).

<https://doi.org/10.3390/ijgi8090400>

KERSWILL, Paul (2003) "Dialect levelling and geographical diffusion in British English", in David Britain & Jenny Cheshire (eds.), *Social Dialectology in Honour of Peter Trudgill*, Amsterdam/Philadelphia: John Benjamins, 223-245.

KOKURITSU KOKUGO KENKYUJO (1953) *Chiiki shakai no gengo seikatsu — Tsuruoka ni okeru jittaishosa —* (Language Survey in Tsuruoka City, Yamagata Pref.), Tokyo: Shuei Shuppan.

KUMAGAI, Yasuo (1993) "Suryoka riron 3 rui ni yoru hogen kukaku to nettowaku ho (Dialect division based on quantification theory III and network method)", *Nihon hogen kenkyukai happyo genkoshu 57* (The 57th Annual Meeting of the Japanese Dialect Society), 27-34.

LABOV, William (1972) *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press.

LABOV, William (2002) "Pursuing the cascade model". Published online: 29 August 2003.

www.ling.upenn.edu/~wlabov/Papers/PCM.html

LI, Jung-min (2014) *Diliyuyanxue de shijian* (The Practice of Geo-linguistics), Taichung: ICJJ.

LONG, Daniel (2007) *English on the Bonin (Ogasawara) Islands*. Publication of the American Dialect Society Nr. 91, Durham: Duke University Press.

MANG, Alexander & Markus WOLLIN (2010) "Sprachraum and sociodemographic variables", in Alfred Lameli et al., *Language and Space*, Berlin: Mouton de Gruyter, 607-626, maps 3101-3106.

MATSUDA, Masayoshi (1978) *Ko-hogensho no Tsuiseki Kenkyu* (Follow-up Study of Old Dialect Dictionaries), Tokyo: Meiji Shoin.

MITSUYA, Shigematsu (1930) *Shonaigo oyobi Goshaku* (Shonai Language and Commentaries), Tokyo: Toko Shoin.

NERBONNE, John (2009) "Data-driven dialectology", *Language and Linguistics Compass*, 3/1, 175-198.

NERBONNE, John, Ilse VAN GEMERT & Wilbert HEERINGA (2005) "A dialectometric view of linguistic 'gravity'". Available at <<https://www.let.rug.nl/nerbonne/papers/gravity2004.pdf>>

ONISHI, Takuichiro (ed.) (2017) *Kukan to jikan no nakano hogen* (Dialect in space and time), Tokyo:

Asakura Shoten.

- SANKOFF, Gillian (2006) "Age: Apparent time and real time", *Elsevier Encyclopedia of Language and Linguistics, Second Edition*. pages?
- SINGLER, John Victor (2001) "Why you can't do a VARBRUL study of quotatives and what such a study can show us", *University of Pennsylvania Working Papers in Linguistics*:
<https://repository.upenn.edu/cgi/viewcontent.cgi?article=1484&context=pwpl>
- TSUZOME, Naoya (2012) "Gurottoguramu de miru hogen no dotai (dynamics of dialect seen in glottograms)", *Nihongogaku*, 31(13), 66-75.
- UEDA, Hiroto & Maria-Pilar PEREA (2014) "The degree of union resulting from reaction points expressed in a diatopic table. An application to a Catalan verb morphology database". *Dialectologia et Geolinguistica*, 22, 16-38.
- YAMASHITA, Akemi & Yasushi HANZAWA (2010) "The language variation and diffusion in Japan — research through Glottogram—", in B. Heselwood & C. Upton (eds.) *Proceedings of Methods XIII*, Bern: Peter Lang, 310-318.
- YARIMIZU, Kanetaka (2007) "Hogen bunpo zenkoku chizu ni okeru kyotsugoka no jokyo (The situation of common language in GAJ)", *Nihongogaku*, 26 (11), 112-119
- YARIMIZU, Kanetaka (2017) "Gendai nihongo no kyotsugo katei (Common language process of contemporary Japanese)", in Onishi, Takuichiro (ed.), *Kukan to jikan no nakano hogen (Dialect in space and time)*. Tokyo: Asakura Shoten, 304-322.
- YARIMIZU, Kanetaka, Kawaguchi YUJI & Ichikawa MASANORI (2004) "Multivariate Analysis in Dialectology — A case study of the standardization in the environs of Paris—", *Linguistic Informatics*, 3, Tokyo University of Foreign Studies, 99-119.
- YOKOYAMA, Shoichi, Seiju SUGITO, Kazuyuki SATO, Masato YONEDA, Tadahiko MAEDA & Takahito Abe (ed.) (2018) *Shakai Gengogaku no Genryu o Ou (Pursuing the source of sociolinguistics)*, Tokyo: Hituzi Syobo
- YONEDA, Masato (1997) "Survey of standardisation in Tsuruoka, Japan: Comparison of results from three surveys conducted at 20-year intervals", *Nihongo Kagaku [Japanese Linguistics]*, 2, 24-39.