Dialectologia 24 *(2020), 111-156.*
ISSN: 2013-2247

Received 3 April 2019.

Accepted 24 June 2019.

STANDARD LANGUAGE DISTRIBUTION IN LINGUISTIC ATLAS OF JAPAN (LAJ) AND LIMESTONE CAVE MODEL: GEOGRAPHICAL AND HISTORICAL INTERPRETATION OF RAILWAY DISTANCES

Fumio INOUE

Tokyo University of Foreign Studies *

innowayf@nifty.com

Abstract

This paper presents a novel technique of showing the railway distance on a real map. Standard Japanese words were grouped into several categories to explain their historical distribution patterns. The influence of the old capital, Kyoto, was significant in the historical development of Japanese, against popular belief that Tokyo speech is the base of the standard language. It has already been shown that the railway distance is effective for explanation of regional differences in dialect. By taking the railway distances from Tokyo and Kyoto as a clue, and by plotting usage values laterally left-to-right on the graph allowed straight-forward visual interpretation of the East-West distribution throughout the Japanese Archipelago. In this paper, we report on the further results of application of two attempts. (1) We used past data on railway distance, comparing the railway distance data in 1980 and in 1910. The railway distance in 1910 showed closer correlation with the standard form usage rate than that of 1980. The correlation coefficient of the railway distance from Kyoto was low. However, it suggested a lot about the correspondence with the actual geographical distribution pattern. (2) The display method was improved. We superimposed a line graph reflecting the standard form usage rate and railway distance on a map of Japan. By visualizing this relationship, it became possible to intuitively understand the correspondence of standardization with traffic conditions and so on. We were also able to interpret the results from (1) above, and we discussed the theoretical possibilities related to railway distance. This analysis made it possible to see two peaks, one at

* Tokyo University of Foreign Languages, 3-11-1, Asahi-cho, Fuchu-shi, Tokyo, 183-8534 Japan.

Kyoto and one at Tokyo, for the propagation rate of standard Japanese on the basis of the railway distance. We also explored the possibility of applying the "limestone cave model" to other phenomena. We examine the possibility that geographical location can be used in interpreting the historical diffusion process.

Keywords

Standard language, Linguistic Atlas of Japan, Limestone cave model, Concentric propagation theory, Railway distance

DISTRIBUCIÓN ESTÁNDAR DE LA LENGUA EN EL ATLAS LINGÜÍSTICO DE JAPÓN (LAJ) Y EL MODELO *LIMESTONE CAVE*: INTERPRETACIÓN GEOGRÁFICA E HISTÓRICA DE LAS DISTANCIAS FERROVIARIAS

Resumen

Este artículo presenta una técnica novedosa de mostrar la distancia del ferrocarril en un mapa real. Las palabras japonesas estándar se agruparon en varias categorías para explicar sus patrones históricos de distribución. La influencia de la antigua capital, Kioto, fue significativa en el desarrollo histórico del japonés, contra la creencia popular de que el habla de Tokio es la base de la lengua estándar. Ya se ha demostrado que la distancia del ferrocarril es efectiva para explicar las diferencias regionales en el dialecto. La toma de las distancias ferroviarias de Tokio y Kioto como una pista, y el trazo de los valores de uso lateralmente de izquierda a derecha en el gráfico permitieron una interpretación visual directa de la distribución este-oeste en todo el archipiélago japonés. En este artículo, damos cuenta de los resultados adicionales de la aplicación de dos tentativas. (1) Utilizamos datos anteriores sobre la distancia ferroviaria, comparando los datos de la distancia ferroviaria en 1980 y en 1910. La distancia ferroviaria en 1910 mostró una correlación más cercana con la tasa del uso de las formas estándar que la de 1980. El coeficiente de correlación de la distancia ferroviaria desde Kioto era bajo. Sin embargo, sugirió mucho sobre la correspondencia con el patrón de distribución geográfica real. (2) Se mejoró el método de visualización. Superpusimos un gráfico lineal que refleja la tasa del uso de las formas estándar y la distancia del ferrocarril en un mapa de Japón. Al visualizar esta relación, fue posible comprender intuitivamente la correspondencia de la estandarización con las condiciones del tráfico, etc. También pudimos interpretar los resultados de (1), y discutimos las posibilidades teóricas relacionadas con la distancia ferroviaria. Este análisis hizo posible ver dos picos, uno en Kioto y otro en Tokio, para la tasa de propagación del japonés estándar en función de la distancia del ferrocarril. También exploramos la posibilidad de aplicar el "modelo *Limestone*" a otros fenómenos. Examinamos la posibilidad de que la ubicación geográfica se pueda utilizar para interpretar el proceso de difusión histórica.

Dialectologia 24 *(2020), 111-156.*
ISSN: 2013-2247

Palabras clave

lengua estándar, Atlas lingüístico de Japón, modelo *Limestone cave*, teoría de la propagación concéntrica, distancia ferroviaria

1. Railway distance as an indicator of dialect differences

1.1 Quantificational dialect classification and diachrony

Dialect research stems from an interest in regional differences. There is a long-time controversy as to whether dialects form a continuum or are separated by boundaries. Languages in modern societies (national languages) are often divided into several dialects. For example, German, French and Spanish can be partitioned in several dialects by lines. On the other hand, computational methods have allowed a more objective analysis of the dialect differences within languages than in the past, which has resulted in a growing emphasis on continuity. In fact, a chain of mutual intelligibility across dialects which belong to different national and standard languages has been suggested (Chambers & Trudgill 1980). There are also attempts to measure the similarity among dialects and languages (Gooskens & Heeringa 2004). Since the end of the 20th century, a trend has arisen to treat what were considered dialects by conventional standards as independent languages, as in Spain and Yugoslavia. Thus, the borderline between a language and a dialect is becoming vague again. In addition to analyzing language itself, some studies have focused on the regulatory factors and formation processes behind dialect differences.

In Japanese also, the elucidation of regional linguistic differences was the starting point and has been an enduring task for dialectology. The evidence used for classifying dialects has evolved from the isogloss to bundles of the isoglosses, and furthermore, the measurement of the degree of commonality (or degree of difference) between dialects is now used as well. Additionally, arguments for the need for comprehensive consideration of many phenomena increased. Along with the popularization of computers, various

quantificational methods, including multivariate analysis, were applied to analysis of differences between localities. Dialects were grasped as a continuum in factor analysis. This data was interpreted through the framework of diffusion theory and propagation studies, and dialectology theory progressed towards “concentric propagation theory”.

Japanese dialectology showed strong interest in differences between East and West Japan. Behind this interest are the cultural exchange between Kanto of the east and the Kinki Regions of the west and the perception of cultural conflicts. While phonology and grammar are interesting from a linguistic perspective, they are less susceptible to influences from outside the language such as culture, geography, economy and so on due to pressure from the language system. Vocabulary, in contrast, is useful for finding basic mechanisms stemming from geographical differences because it involves many phenomena and is highly diverse, reflecting the flow of culture. Linguistic and cultural phenomena are mutually influenced, and traffic is a particularly significant explanatory extra-linguistic factor.

This article addresses the words studied in LAJ “Linguistic Atlas of Japan” (National Language Research Institute 1966-1974). The field surveys of LAJ were executed from 1957 to 1964. The informants were male and born before 1903. Publication was from 1966 to 1974. The "Kasai data" is the average value of standard Japanese word forms in LAJ by prefecture. Individual differences among the informants are not retrievable, but as an advantage of the average value, regional differences have been beautifully represented. Diachronic interpretation can be derived from the synchronic distribution of these standard words. The standard Japanese words will be analyzed in this paper, but the survey is based on a questionnaire about the vocabulary used daily as the colloquial, local language. Whether or not it corresponds to the standard word is coincidental and should be regarded as an external factor on the regional dialect. For example, the result of aggregation varies depending on which form of synonyms *osoroshi·*/ *kowai* ‘scaring’ or *sakana* / *uo* ‘fish’ is taken as the standardized word.

In section 4 of this paper, we will present four classifications of linguistic diffusion according to Ancient / Modern and East / West. Among those four, we can observe influence from Kyoto in three. This diverges from the common view that the words of modern Tokyo were adopted as the standard language. Continuity was observed in the

formation between Ancient and Modern dialects. Various past changes and dissemination show continuity with the modern "new dialect" Inoue 2010.1, 2010.12, 2017). This idea that the past and the present are regarded as continuous is in agreement with the uniformitarian principle of geology as discussed by Labov (1972).

1.2 Research history of railway distance

It has long been recognized that geographical distance is useful in explaining regional differences in dialect. The origin of this method is referred to as the "dialect concentric propagation theory". Various methods have been proposed. The simplest one is straight-line distance, but other types of distance such as traffic, road distance (Kumagai 2013, 2016), railway distance, sea route, etc. have also been considered. Railway distance was adopted in Inoue (2004).

Comprehensive analysis of dialect differences developed greatly thanks to dialectometry, developed by Goebl (1982), Nerbonne et al. (2005), Heeringa (2004), etc. The use of railway distance is one development of this method. As quantificational studies on dialect distribution, the relationship between road distance, railway distance, traveling time and the like, there are works in Switzerland (Jeszenszky et al. 2017) and Sweden and the Netherlands (Gooskens 2005). Also in Japan, Inoue (2006, 2008), Yarimizu (2009) applied and confirmed the validity of railway distance. Kumagai (2013, 2016, etc.) studied the correspondence between dialect and road networks.

As quantificational studies of LAJ, Inoue (2001) can be cited. Kumagai (2013, 2016) is a dynamic presentation of the data. For GAJ "Grammar Atlas of Japanese dialects" (National Institute for Japanese Language and Linguistics, 1989-2006), it was confirmed that the umbrella model of Inoue (2001) is applicable also to grammatical phenomena. Yarimizu et al. (2004) also applied multivariate analysis MDS (multidimensional scaling method) to the standard French word usage around Paris and found that the geographical position could be represented based on these data. Onishi (2019) rejects the simple dialect concentric propagation theory based on the new nationwide data FPJD

"Formation Process of Japanese Dialects" and Onishi (2016) NLAJ "New Linguistic Atlas of Japan" (Onishi 2019).

1.3 Limestone Cave Model

In this paper, we propose the "limestone cave model" for geographical propagation of standard Japanese.[1] Figure 1 shows a map of Japan (Inoue 2007) created by Yarimizu which divides the Kasai data on standard form usage rate into 20 or more intervals between 62% (Tokyo) and 3% (Okinawa). It shows nearly concentric circles from Tokyo as the center.

See §6.4 for a list of region names used in this paper with the prefecture labels used in the graphs and charts.

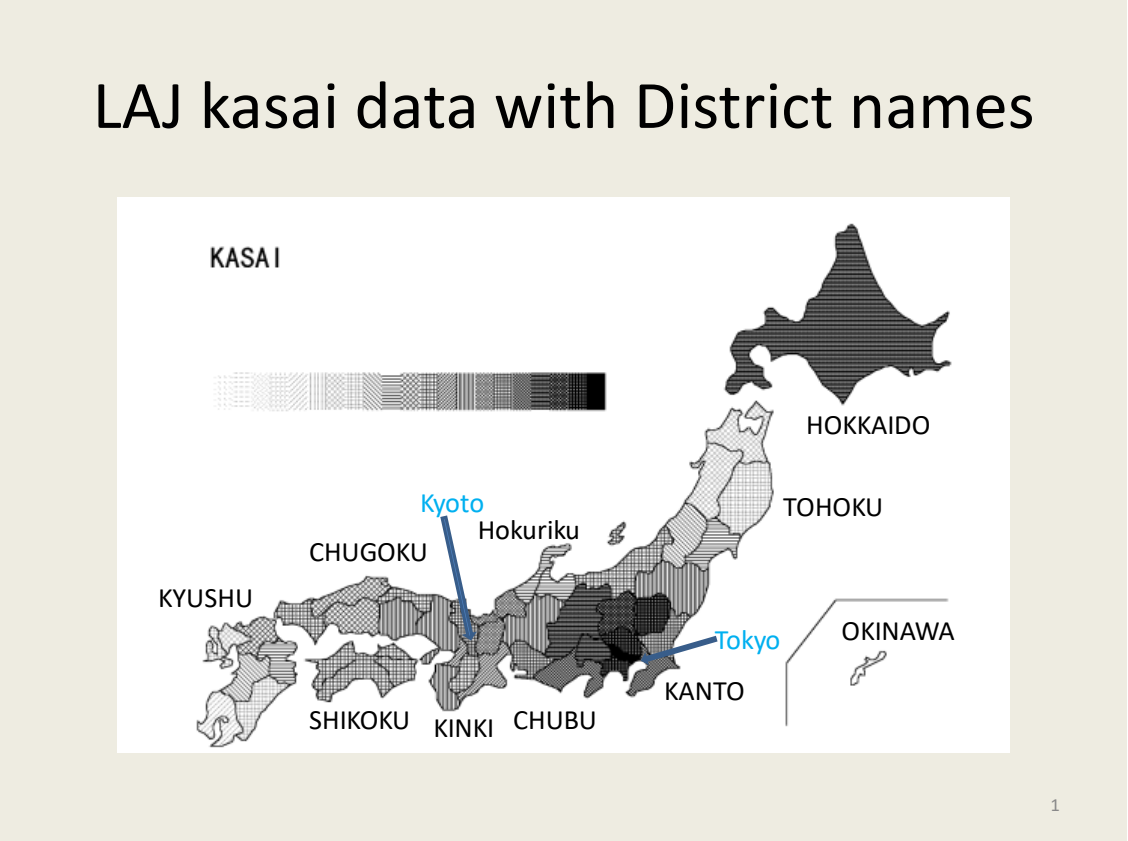

Figure 1. Distribution map of the 82 standard forms from LAJ with Region names.

[1] The limestone cave model is a revised version of the umbrella model (Inoue 2010.7, 2010.12) and double umbrella model (Inoue 2017).

Standard form usage rate can be displayed as a continuum by using a bar graph. The distance on the map can be expressed as a one-dimensional progressive numerical value by using the railway distance. In Figure 2, both are displayed in a scattergram. The vertical axis is the standard word usage rate. The numerical values are shown inverted, with 0 at the top. The horizontal axis shows the railway distance from Kyoto to the station of each prefectural capital.

This has the following advantages. Standard form usage was expressed as continuous numerical values, as if hanging from the ceiling of the cave. The two-dimensional ground surface is expressed in one dimension by the horizontal axis, so that it can be read that the standard form usage rate is higher in the prefecture closer to Tokyo (T●K), and the second peak is in the vicinity of Kyoto (K▼T). Furthermore, when eastern Tohoku and western Kyushu are viewed from Kyoto, they are almost equidistant, and the standard form usage rate is similar, making symmetry with a midpoint at the center of Kyoto. In other words, we can metaphorically compare the standard form usage rate with two stalactites (hanging limestone), one in the center of Kyoto and another in the center of Tokyo. Note that Okinawa and Hokkaido are exception, Okinawa as the former dominion of the Ryukyu Kingdom and Hokkaido as the former frontier where Ainu people had been living.

The two stalactites can be represented by connecting the tip of each prefecture according to the railway line. For instance, it can be observed that the low usage rate in the Hokuriku region connects to the low rate in Niigata NIG, Yamagata YMG and Akita AKI by following the Japan Sea railway, which connects Osaka OSK and Aomori AOM. From Section 3 onwards, we will use the graphs of this method.

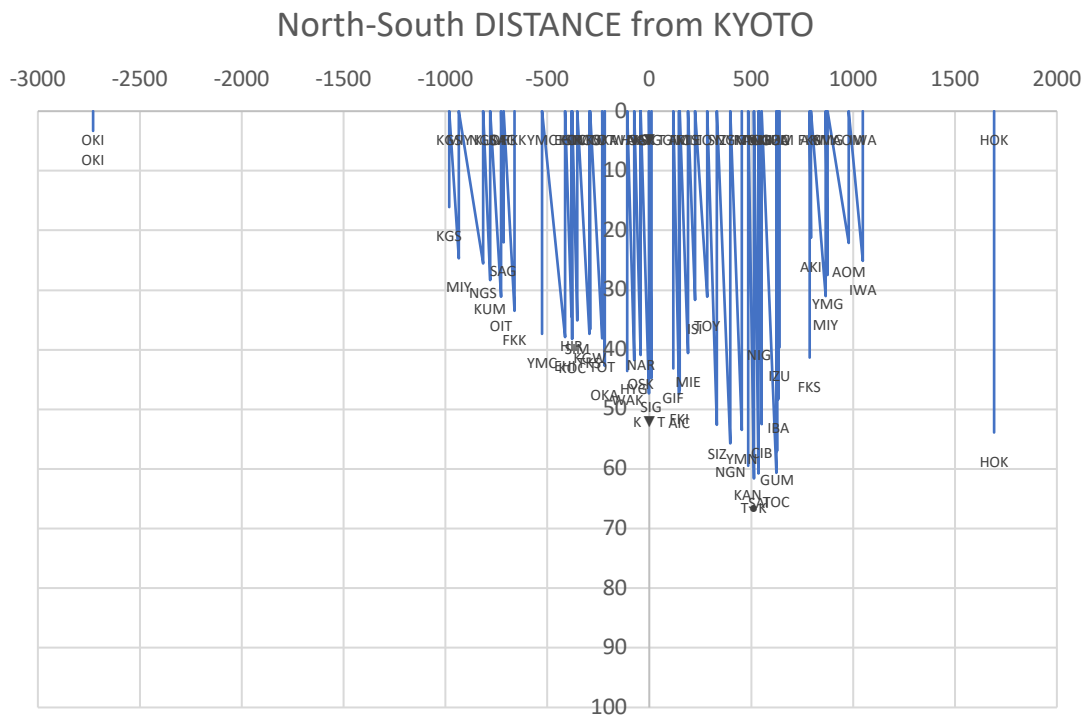

Figure 2. LAJ standard form by railway distance from Kyoto.

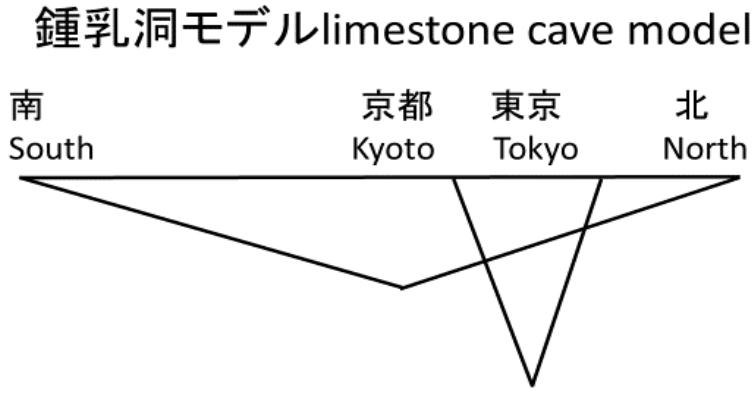

Figure 3. Limestone Cave Model of standardization

Figure 2 can be graphically simplified as Figure 3. Jumping ahead of the conclusion, Figure 3 represents a "limestone cave model" on standard form dissemination, allowing a metaphorical comparison between the geographical propagation process of standard language and the development of stalactites. In a

limestone cave, when water drips from the ceiling it produces a stalactite. The distribution pattern can be explained by viewing both the Kyoto and Tokyo peaks. In the past, water dripped from the ceiling of the Kyoto region of the "limestone cave" but in the early Modern times water dripped from the Edo=Tokyo region. This corresponds to the distribution map by prefecture of LAJ standard form usage rate in Figure 1. Figure 2 visualizes the correspondence with the graph (displayed later) of the railway distance divided to the East-West.

1.4 Advantages of railway distance as an indicator

Railway distance is a convenient indicator. The routes never change significantly so stable distance values are always available. It also reflects past transportation networks. Railway network construction often follows former road networks in order to maximize profitability, reflecting people's long-term interaction through history. The modern bullet train Shinkansen network replicates the development of the railway network since the beginning of Modern times. Many of the railway networks in Modern times trace the main roads of the Edo period. Going back further, it has also copied the official roads in the ancient Heian dynasty. The Meiji Era road network (shown in LAJ Vol. 4) utilized by Kumagai (2013, 2016) overlaps considerably with the railway network of the same period.

The railway network at the end of the Meiji Era, one which we deal with in this paper, reflects the traffic situation of an older period. However, the railway network does not usually directly reflect maritime shipping routes that competed with land ways. The results in this paper are indicative of the importance of the shipping route over the Seto Inland Sea (between Chugoku and Shikoku) in addition to the railway.

Railway distance also has an advantage as a technique of visualization. By using the railway distance from the cultural centers, it is possible to represent the two-dimensional ground surface on one dimension, still allowing distinguishing between each prefecture. Inoue (2001) applied various multivariate analyses to the distribution of LAJ standard

forms.[2] One of them was to calculate the geographical gravity center (Inoue 2004) and express the Japanese archipelago on one dimension, i.e. longitude only. However, longitude does not allow us to observe differences according to latitude, which is significant for areas with a long north-south latitude such as between Tokyo and Tohoku. Another method, railway distance does not have this kind of drawback. The usage rate can be shown on the other dimension of the scattergram. In addition, Inoue (2004, 2010.2) displayed the railway distance, the standard form usage rate, and the first year of appearance in one figure, using a three-dimensional (3D) graph, discussing their close relationship with the findings of this paper and the Japanese language history.

1.5 Characteristics of 1910 railway distance data

The railway distance used in previous research is based on the JNR (Japan National Railway) timetable of the 1980s.[3] It accurately reflected the railway at the time of the completion of the investigation of LAJ. However, if the railway distance at the time the informant was grown is available, we can verify which time period most closely corresponds to the usage data. Data showing the railway distance at the beginning of 20th century was found in Appendix at the end of the book of "*Hakubunkan Diary for Taisho Year 2*" (published in 1912 = Meiji 45 = Taisho 1) (Miyaji et al. 2013: 787).

Figure 4 presents a railway route map in 1906, slightly before this time (Kubota 2005). It shows the situation immediately before the establishment of the state-owned railway law in 1906. It corresponds to the stage when the railway network was owned by the state after the experience of poor military transport during the Sino-Japanese war (1894) and the Russo-Japanese War (1904). For eastern Japan, all the prefectural capitals and Tokyo were connected by railway. In western Japan, although the line in Chugoku was connected, the construction of the railway network of Shikoku and Kyushu was later.

[2] The 82 words of Kasai data are listed in Inoue (2001).

[3] Publication was in Oct. 1981. However, we use 1980 for the sake of simplicity. The railway distances between the prefectural capitals have not changed significantly from the time of the LAJ field survey (from 1957 to 1964). Construction of railway was active until the 1930s, that is, when the average age of the LAJ informants was in the 30s. Since then, growth of the railway was gradual until the 1980s except deficit local lines.

The differences with the railway distance between this time and 1980 are found mainly in western Japan.

Figure 4. Railway network in 1906 (Kubota 2005).

1.6 Development of railway network

The rail route from Tokyo to Hokuriku and East Kyushu is indirect. This is also true of the main roads (Kumagai 2013, 2016). In Shikoku and Eastern Kyushu, the Seto Inland Sea was used, so passenger ships developed. The advent of the steamship improved safety and allowed scheduled, regular service, which is not affected by the

weather. As another means of transport in the Meiji Era, there were also rickshaws and horse-drawn carriages to connect to parts which did not have rail.

2. Explanatory power of 1910 distance and 1980 distance

2.1 Correlation coefficients of 1910 /1980 distances of Kyoto / Tokyo

Two types of railway distance, 1910 and 1980, are used below. Based on the correlation coefficient, we see which time period's distance better corresponds with the LAJ standard form usage rate. We will first examine the overall image by comparing correlation coefficients between 1910 and 1980 as shown in the 9 graphs listed in Table 1. *x* and *y* indicate maps which are not exhibited in this paper. Bold letters mark the largest numerical values.

		Correlation coefficient
Fig. 5	Kyoto 1910 railway distance (Mainland)	0.1307
Fig. 5x	Kyoto 1910 railway distance (nationwide)	0.1185
Fig. 6	Kyoto 1980 railway distance (Mainland)	0.1937
Fig. 6x	Kyoto 1980 railway distance (nationwide)	0.2215
Fig. 7	**Tokyo 1910 railway distance (Mainland)**	**0.5773**
Fig. 7x	Tokyo 1910 railway distance (nationwide)	0.5646
Fig. 7y	**Tokyo 1910 railway walking distance (Mainland)**	**0.5886**
Fig. 8	Tokyo 1980 railway distance (Mainland)	0.5616
Fig. 8x	Tokyo 1980 railway distance (nationwide)	0.5501

Table 1. Formula of correlation coefficient.

According to Table 1, the largest correlation coefficient is 0.5886 in Figure 7y. However, this is a numerical value of "Tokyo 1910 railway walking distance (Mainland)" which factors in the time taken to walk through the areas without rail, so it is a different

class from the others. The next highest numerical value is 0.5773 in Figure 7 “Tokyo 1910 train distance (mainland)”, and the next is 0.5646 in Figure 7x “Tokyo 1910 railway distance (nationwide)”. The corresponding numerical values of 1980 are a little smaller at 0.5501 and 0.5616. Therefore, it seems that in terms of explaining the whole result of the LAJ survey conducted in 1957-1964, the 1910 railway distance, which is the informant's growth period, is more effective than the railway distance in 1980 at the time after the publication (1966-1974). The average birthdate of the LAJ informants is 1894.2, so, in 1910 they were 16 years old on average. Although past research has already shown that railway distance provides a good explanation for the distribution of language forms, this endeavor has improved the accuracy.

Above are the correlation coefficients of the railway distance from Tokyo. Looking at the railway distance from Kyoto in Table 1 in Figure 5, Figure 5x, Figure 6 and Figure 6x, they are low numerical values from 0.11 to 0.22. As we can see from later graphs, the distance from Kyoto shows two large stalactites, and its relationship with the usage rate of standard language forms is complicated. Here we can see the effectiveness of the railway distance from Tokyo, so, further analysis of railway distance from Kyoto seems unnecessary. The fact that it is significant will be shown in the third section through a different method.

2.2 Four types of railway distance:1910 / 1980 and Tokyo / Kyoto

In the following scattergrams between railway distance and usage rate of standard language forms, approximate straight lines are added, and the formulas are shown. In order to make Tokyo and Kyoto stand out, the prefecture label was changed to T●K, K▼T instead of TOK, KYT in some graphs.

2.3 Kyoto 1910 railway distance

Let us look at the numerical values from Kyoto Station in 1910. Figure 5 shows the railway distance from Kyoto in 1910 for mainland prefectures. Although Hokkaido and Okinawa were omitted due to their being exceptional, the pattern is still complicated. The overall pattern shows a downward trend to the right, and two tendencies can be observed: a gentle curve from Kyoto to the far end of the Japanese archipelago and a steep mountain around Tokyo which is 600km from Kyoto. Kanto is plotted on the upper side, and Tohoku and Kyushu are each plotted on the lower righthand side. The correlation coefficient is 0.13, which is low, and the approximate straight line and the equation shows that the rate is 46% near Kyoto and reduces 0.1% each 100 km. In the nationwide data in Figure 5x, which is omitted here, the new frontier Hokkaido is positioned at the upper righthand side, and the numerical value grew thanks to the spread of the Hokkaido common language. Okinawa is situated in the lower righthand side with the largest distance and the lowest standard language forms usage.

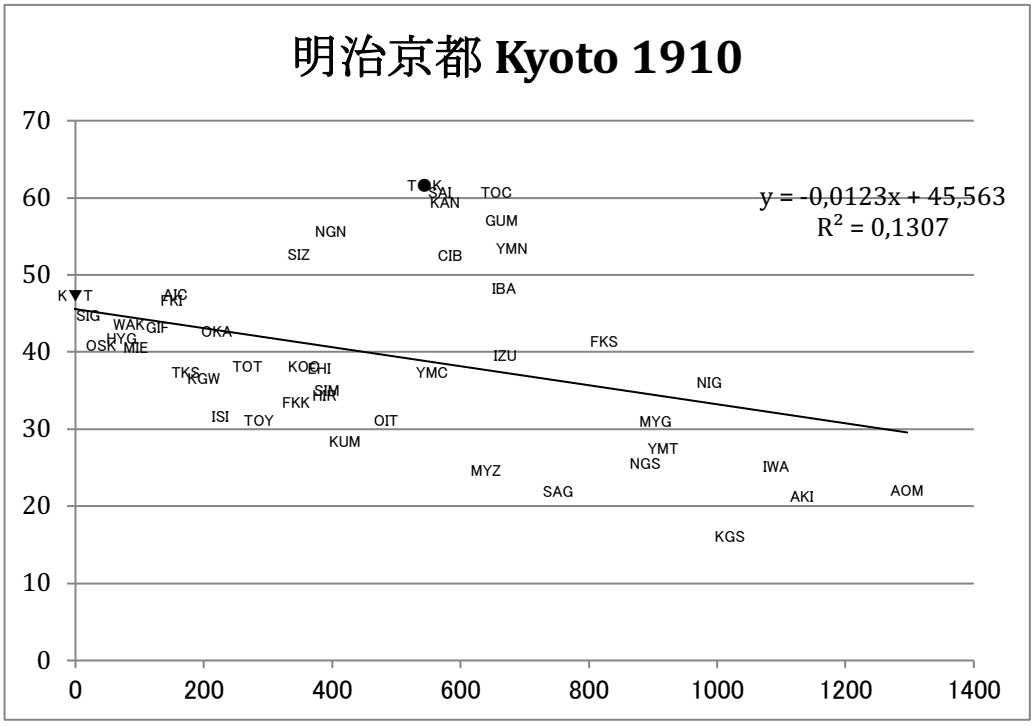

Figure 5. Kyoto 1910 railway distance.

Dialectologia 24 *(2020), 111-156.*
ISSN: 2013-2247

2.4 Kyoto 1980 railway distance

Figure 6 shows the data of the 1980 railway distance from Kyoto in the mainland (Hokkaido and Okinawa are omitted). The numerical values near Tokyo are large and those in western Japan are small. The whole pattern is similar to Figure 5 and it is difficult to find differences. As mentioned above, the correlation coefficient is low for the railway distance from Kyoto, and this gives an impression that Kyoto distance is not helpful, but graphs divided into East-West, which will be described in section 3, show the nationwide trends. This allows a more complete interpretation of the history of Japanese dialects.

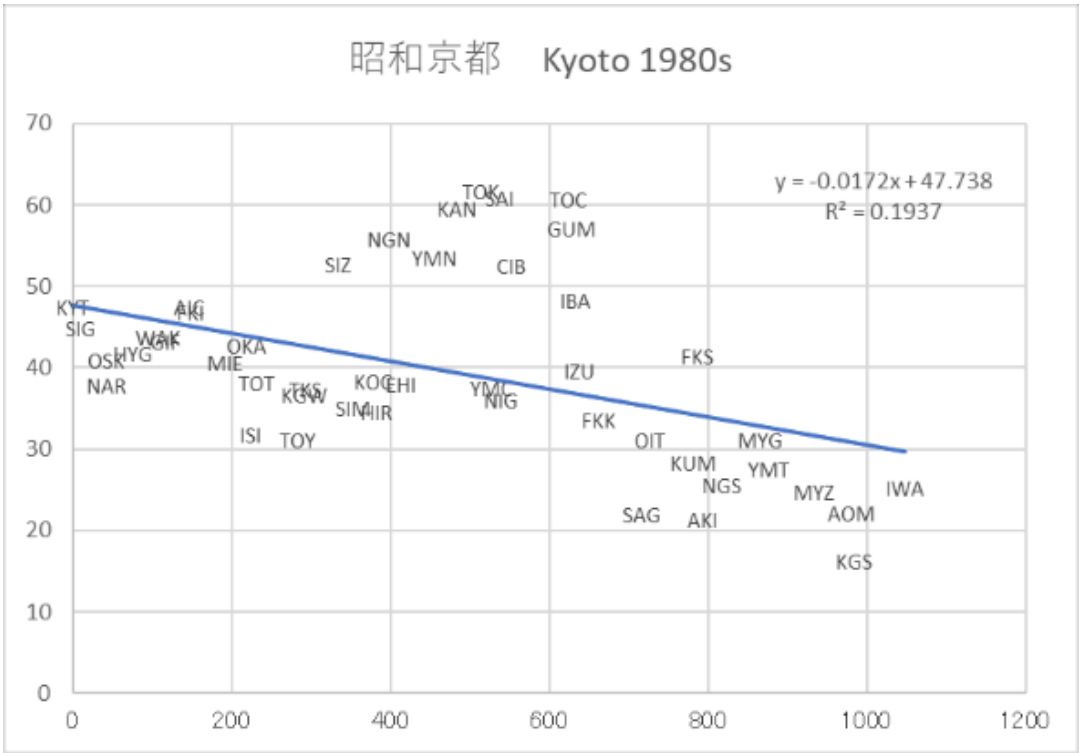

Figure 6. Kyoto 1980 railway distance.

2.5 Tokyo 1910 railway distance

Let us next view the railway distance from Tokyo. In Figure 7, we see the data in 1910. Okinawa prefecture and Hokkaido were excluded in order to focus on similar prefectures. It shows a single downward slope as a whole. As the distance from Tokyo increases, the usage rate of standardized language forms decreases. However, Tohoku

and Hokuriku are in the lower left, below the approximate line, indicating that distance may not be a good indicator for standard form usage. The correlation coefficient is 0.5773, which is the highest in the Figure for Tokyo. Approximation straight lines and mathematical formula show a trend of 52.8% near Tokyo and 2.26% reduction per 100 km distance. The % attenuation per distance is larger than other Figures.

Let us focus on the prefectures near Tokyo here. Prefectures in Kanto are plotted above the approximate line. The existence of the Five Main Roads in the feudal Edo period seems to be an extralinguistic factor in this. The prefectures within the circles, through which the Five Main Roads pass, show higher standard form usage rates. In contrast, within the Kanto Region, Ibaraki IBA and Chiba CIB, through which the Five Main Roads do not pass, have low standard form usage rates.

In Figure 7x (omitted here), which contains Okinawa and Hokkaido, the correlation coefficient is 0.56. The formula of the approximate straight line shows a trend of 52.6% near Tokyo and 2.15% reduction per 100 km distance. In Kagoshima, 1500 kilometers away, the trend is less than 20%, so the prediction holds.

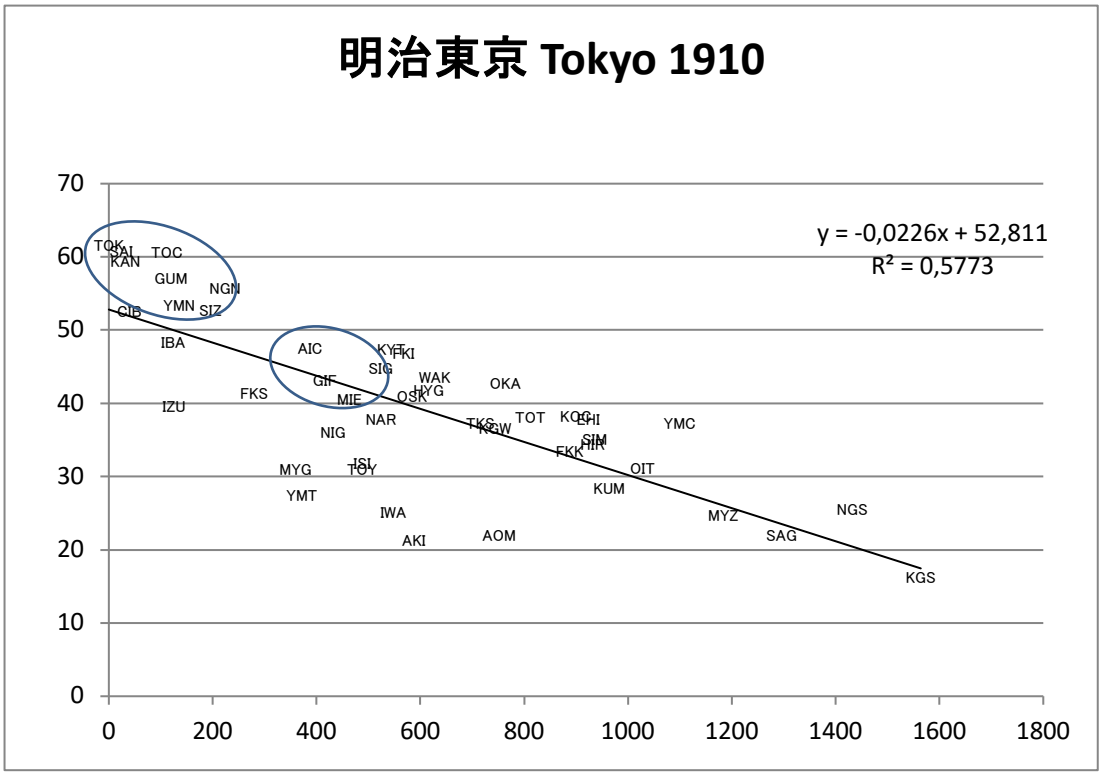

Figure 7. Tokyo 1910 railway distance.

2.6 Tokyo 1980 railway distance

Next in Figure 8, let us look at the relationship of Tokyo 1980 railway distance. Observe the linear tendency of the mainland as a whole that falls to the right, with the standard form usage decreasing as the distance from Tokyo increases. The correlation coefficient is 0.5616, which is lower than that of Figure 7 which portrays the 1910 data. For explaining the results of LAJ, the railway distance of the informants' growth period is more effective than the railway distance at the time of publication.

In all the graphs above, the clumps at Kanto and Eastern Chubu Region are conspicuous. As for the quantificational tendency based on the railway distance, the 1910 data corresponds to the standard word usage rate better than that of 1980. This trend has its roots in the traffic situation during the feudal Edo period. In the same way, Tohoku and Hokuriku Regions are equidistant from Tokyo and are separated under the approximate line, in contrast with the Kinki Region where communication was more frequent on the Pacific Ocean side. This shows that traffic and main roads before the establishment of the railway had a significant influence on the standard usage rate in the LAJ. This point will be discussed in section 4 when discussing the relationship of geography with the first appearance of words in Figures 13 to 16.

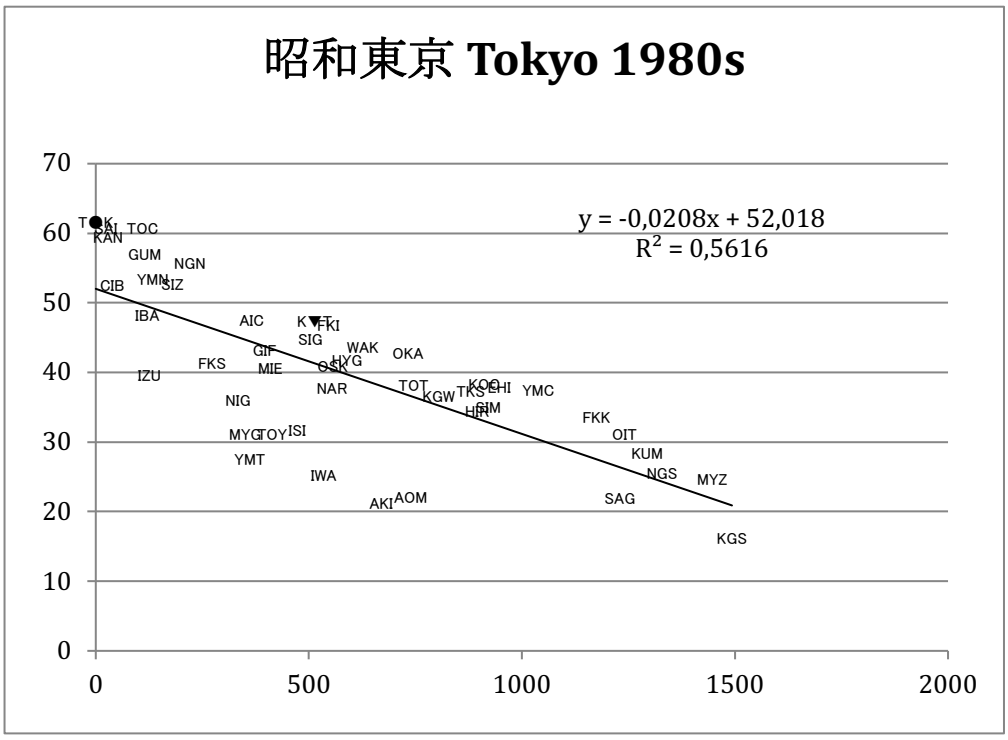

Figure 8. Tokyo 1980 railway distance.

According to Jeszenszky et al. (2017), travel time in the past corresponds well with the current dialect distribution. The required travel time in 1885 was analyzed for our data (graph omitted here). The correlation coefficient was 0.59, which was the highest. However, the reason should be interpreted as being that the railway range in this year was limited, which corresponds to the area where the standardized forms are used frequently. This was when the railway network around Tokyo reached the whole Kanto Region. For other regions, the travel time was calculated as 5 times the time that would be required by railway.

3. Railway distance fits with limestone cave model

3.0 Mapping by East-West division and axis reversal

Above, we looked at the correspondence between words and distances by scattergrams. Below, we will divide the railway distance into East-West according to geographical locations in order to make it easier to understand the East-West collation with the map. Also, we invert the numerical value on the vertical axis,[4] so it shows the rate of non-use of standard forms, while the upward direction shows the dialect usage ratio. Imagine the drops of the hanging stalactite when interpreting them.

Further, line graphs based on the railway distance are superimposed on the Map of Japan. The horizontal and vertical scales of the numerical graph were adjusted so that they would match the corresponding geographical position of Japan. The angle of the Map of Japan was also tilted to increase correspondence. In the figures below, if the position of a prefecture is plotted above the actual geographical position, the standard form usage rate was lower than expected (i.e. they use more dialect forms). Examples include prefectures along the Japan Seaside and Kyushu.

We next compare the one-dimensional railway distance data with the two-dimensional map. This technique reveals that the regional differences in standard form usage rate correspond to their relative geographical position. In Figure 1, we divided the usage rate into 20 or more gradations, visualizing finer differences as compared to the several gradations used in previous research, but the change in usage rate was still discontinuous. In contrast, the line graphs in Figures 9-12 show the continuous aspect of the usage rate change. The correspondence with the map is also shown.

Let us overview the four figures. If we align any of the four line graphs with the geographical position of Japan as a whole, we get roughly two lines, one going north of Tokyo and one going west of Tokyo. One lines up almost exactly to the map at Aomori in

[4] This technique of reversing the axis value was presented in Inoue (2019).

Tohoku, and another line extends to the west along the Seto Inland Sea. This data does not take geographical direction into account, so Okinawa is plotted in the East China Sea.

3.1 Kyoto 1910 distance

Let us consider the Kyoto railway distance in 1910. Figure 9 is the East-West division of Figure 5. As a whole it can be approximated by two lines: one gentle line from the west of Tokyo and a sharp line east of Tokyo. There are small discrepancies between the locations of the prefectures according to the line graph and the map. It is difficult to find anything like the two stalactites which will be shown in Figure 10.

For East Japan, the Kanto Region prefectures are grouped together in the vicinity of Tokyo. This means that the language usage is closer to that of Tokyo than is predicted by geographical position. Also, due to a lack of railway along the Japan Sea, there was no direct route from Kyoto to Akita AKI and Aomori AOM, instead stopping in Tokyo first. They cluster with the other prefectures of Tohoku Region. Since Niigata NIG prefecture was reached via Nagano NGN prefecture, there was also a long distance, so it was plotted near Miyagi MYG prefecture. Hokuriku was close to Kyoto via Shiga SIG prefecture. Ishikawa ISI and Toyama TOY were almost in agreement with their actual geographical positions. They are plotted in the Japan Sea, which evinces a low standard form usage rate. Western Japan lined up almost along Chugoku Region. Moreover, it kept about the same angle east of Tokyo. Kyushu and Okinawa are extended and plotted near Korea. That is, even if the geographical position goes southward, railway distance is still the controlling factor. In other words, it seems that standard forms were accepted at about the same rate in Tokyo and in western Japan.

Dialectologia 24 *(2020), 111-156.*
ISSN: 2013-2247

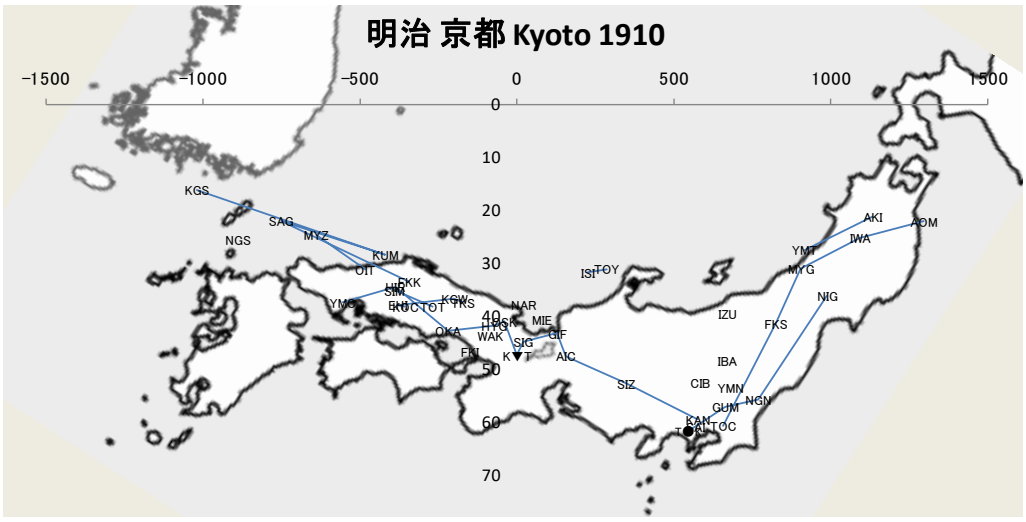

Figure 9. Kyoto 1910 railway distance and LAJ.

3.2 Kyoto 1980 distance

In Figure 10, the distance from Kyoto station in 1980 is used. It corresponds to the East-West division of Figure 6. Since, unlike the 1910 data, the railways along the Japan Sea have been connected between Kyoto and Aomori AOM, there were two railway routes for east Japan. Two stalactites can be observed in this graph. One consists of a line from the center of Kyoto heading towards Kyushu and another gentle line along Japan Sea heading east toward Aomori. The other stalactite is the lines from the center of Tokyo.

In East Japan, the Tohoku Region calls for attention. The Pacific Ocean side and the Japan Sea side of Tohoku Region are separated, and the prefectures' positions on the line graph almost agree with their geographical positions. Since prefectures in Tohoku along the Japan Sea are connected with the Hokuriku Region by railway, the slope from Kyoto forms a symmetrical line similar to the west of Kyoto, constituting one stalactite. The Kanto and Chubu Regions cluster together, constituting a separate stalactite. This is because the numerical value of each prefecture is close to the numerical value of Tokyo owing to the rapid diffusion of standardized forms from Tokyo.

Regarding Western Japan, all the prefectures of Chugoku and Shikoku Regions overlap with the main railway line, thanks to the connection of the railway. In addition, Kyushu is plotted above the other regions, showing less usage of standard forms.

Both Chugoku and Shikoku are plotted almost on one line to the west from Kyoto, but the east prefectures are clearly divided between the Pacific Ocean side and the Japan Sea side. Seen from the center of Kyoto, the standard form usage rate of the Japan Sea side (from Hokuriku to Eastern Japan) is low (i.e. there is greater dialect use), which agrees with the geographical position on the Map of Japan.

From this map, we can see two historical processes in the spread of standard forms. (1) The railways to the west and along the Japan Sea form a stalactite and show the propagation from the Kyoto center in ancient times. (2) The other stalactite shows propagation from Tokyo to the Kanto and Chubu Regions in modern times. Thus, there are two historical stages to the spread of standard Japanese, (1) from Kyoto in ancient days, (2) from Tokyo, over a moderate range, in modern times (Inoue 2010.2). The East Japan and West Japan factors were already taken out by application of multivariate analysis. Further cluster analysis showed that 82 words were separated into East and West clusters according to which side of Japan showed a greater usage rate (Inoue 2001, 2008). This served as the basis for an interpretation of the geographical distribution through history.

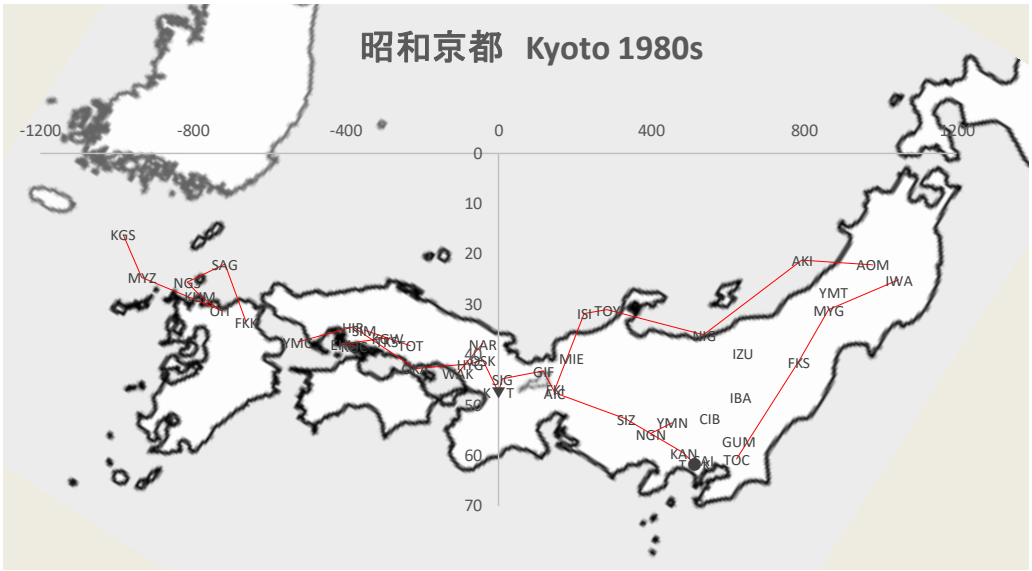

Figure 10. Kyoto 1980 railway distance and LAJ.

Dialectologia 24 *(2020), 111-156.*
ISSN: 2013-2247

3.3 Tokyo 1910 distance

Next, we shall look at the railway distance data from Tokyo. Figure 11 shows the graph by distance from Tokyo in 1910 divided into the east and west of Tokyo. This figure is the East-West expansion of Figure 7. Tohoku Region, which deviated from the overall trend, lined up from center to right, corresponding to lateral geographical position on the Map of Japan. This made the difference easier to interpret. This can be compared with Fig 12 below to observe the effect of time period.

Looking at the whole picture, some of the prefecture's position on the line graph is slightly displaced as compared with the map. In east Japan, the angle of the line is different for Kanto than for Tohoku. Tohoku Region is located above the railway distance from Tokyo. In other words, Tohoku has maintained more dialectal forms. The steep angle between Tokyo and Fukushima FKS means that the change in standard forms usage per distance is large. This is likely due to the presence or absence of a traffic route by the Five Main Roads in Feudal Japan. The Kanto Region and nearby Yamanashi YMN, excluding the Izu Islands in the Pacific Ocean, constitute one group because their positions are close and the standard form usage rate high. Hokuriku (FKI, ISI, TOY) was plotted above Kyoto in Figure 11, due to railway route from Tokyo being indirect, going around Shiga SIG prefecture in those days. From the viewpoint of Tokyo, the standard word usage rate shows a symmetrical correspondence with Tohoku. It can be interpreted that the usage rate of standard forms in Hokuriku is low because it was far from Tokyo by railway at that time.

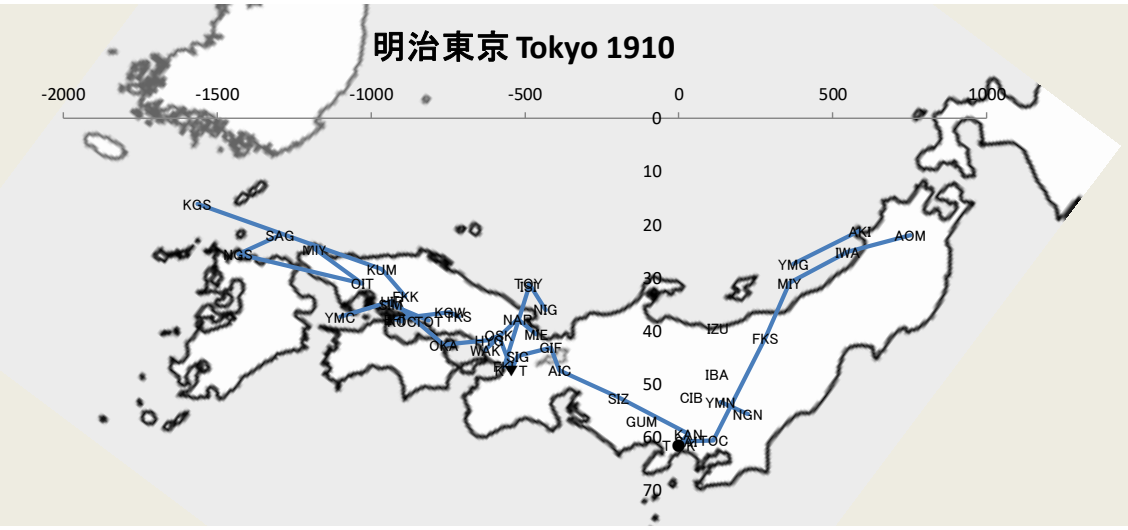

Figure 11. Tokyo 1910 railway distance and LAJ.

When looking at West Japan, the prefectures form nearly a straight line from Tokyo to Kagoshima KGS prefecture at a similar angle. Kyushu is plotted in Chugoku Region because the sea route in the Meiji Era was a shorter distance than railway.

3.4 Tokyo 1980 distance

Figure 12 shows the distance from Tokyo in 1980. The data corresponds to the East / West division of Figure 8. For East Japan, the main railway network had been completed in the early 20th century, so railway distance is not different from Figure 11 of 1910. In Tohoku, the Japan Sea side (Yamagata YMT and Akita AKI) is plotted near the Pacific Ocean side (Aomori AOM, Iwate IWA and Miyagi MYG). If not for the graph of distance from Kyoto in Figure 10, researchers would likely have tried to explain usage rate with a single stalactite.

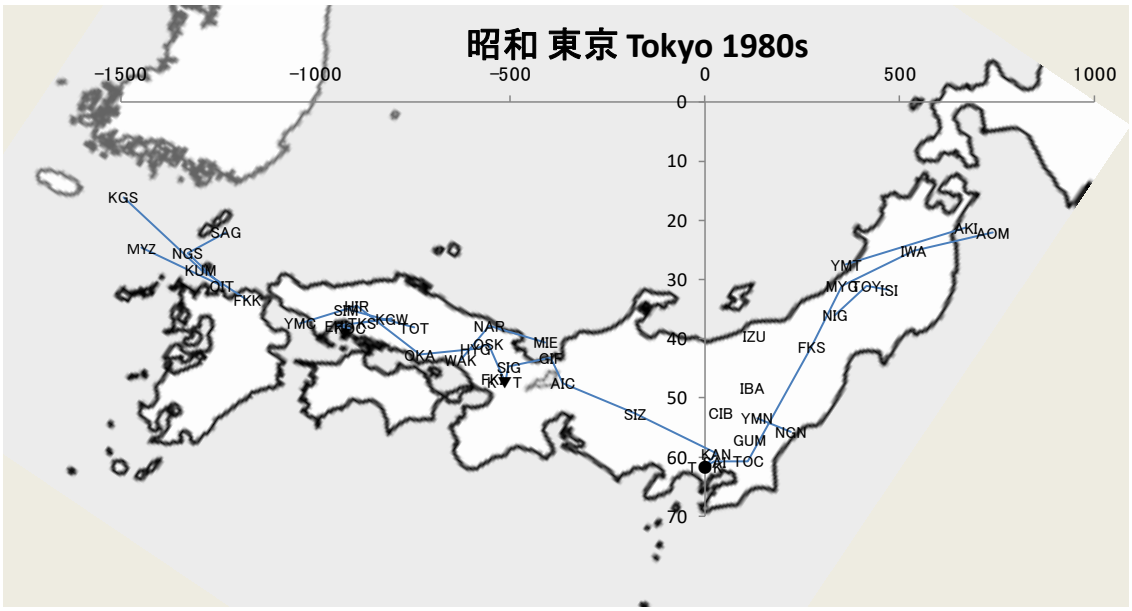

Figure 12. Tokyo 1980 railway distance and LAJ.

The Kanto Region prefectures cluster around Tokyo. The sharp line from the Kanto Region to Tohoku indicating the difference between the areas stands out. Since Niigata NIG got a short route via a new tunnel, its position on the line is similar to Tohoku Region. Since Hokuriku (TOY, ISI, FKI) got a short route via Nagano NGN prefecture, its position is now in the Japan Sea on the right side. From the point of Tokyo, Hokuriku is similarly far as Tohoku Region and shows a common tendency. It seems that the Hokuriku here shows a stage where it had not yet been absorbed in the sudden wave of standardization from Tokyo.

Regarding West Japan, from Tokyo to Kyushu, the angle is gentle compared to the steep angle east of Tokyo. It shows a gentle angle which follows along Chugoku Region, and Kyushu is extended further westward at a similar angle. In other words, even if the geographical position is southward, the controlling factor in the rate of standardization is railway distance.

3.5 Evaluation on effect of overlaying maps

As described above, thanks to the effect of reversing the vertical axis (standard form usage rate) of the scattergram, the geographical position became easier to grasp allowing clearer interpretation. Visualization has allowed a deeper interpretation. By dividing further in the east and west, overlapping with the actual map, it was shown more concretely that the railway distance from Tokyo and Kyoto was the controlling factor in the difference in language use, and that there are slight differences due to the traffic conditions of each prefecture. This indicates the effectiveness of displaying the map in the background.

Kyoto 1980 in Figure 10 has shown a distinct pattern from the other three maps. The eastward lines were separated by the railway along the Japan Sea and Pacific Ocean. New words propagated from Kyoto to the East and to the West as a central language, and in Modern times new words diffused as standard words from Edo=Tokyo to East Japan. This process can be observed if the data is viewed as it corresponds to the map. Although, as mentioned earlier, the Table 1 correlation coefficient of standardization with railway distance was less significant for Kyoto in 1910 than for Tokyo, the information that could be read from Kyoto distance is rich when plotted on the Map of Japan as in Figure 10.

By 1980, the railway network had developed nationwide. Indirect routes have almost disappeared. In other words, we can reproduce quite a lot of the Edo period road. This differs from the railway construction stage of only the main route in 1910. We can follow a railway route which is similar to walking routes in the past. Thanks to that, we can even reproduce the propagation pattern starting from Kyoto in the feudal Edo period.

In the four figures above, the one-dimensional numerical values of railway distance and of standard word usage rate were combined in two dimensions and a pattern suitable for the physical shape of the Japanese archipelago was obtained. This means that the propagation of standard forms was influenced by geographical distance from Tokyo (and Kyoto). It shows that change in language phenomena is dominated by geographical factors in a rule-governed manner when spreading and propagating. The “geographical proximity effect”, which is the fundamental principle of geography, was reflected. The same data was subjected to multivariate analysis without geographical factors, and East

and West factors were extracted (Inoue 2008, 2010.2). Similar to that, the two-dimensional ground surface was simplified to one-dimensional railway distance and it showed a close relationship to standard form usage rate.

So far, we have compared the railway distances of 1910 and 1980 and analyzed the standard word usage rate. Even from the correlation coefficient, the superior explanatory power of the railway distance in 1910, when the LAJ informants were in the process of growing, was clear. The influence of the Five Main Roads in the Edo period was also considered. In a slender nation like the Japanese archipelago, the railway distance accurately reflects the usage of dialect (standard language) and past geographical propagation. Using railway distance, we rendered the two-dimensional geographical position as one dimension, and then restored the geographical dimension with a map of Japan to make a more concrete analysis possible.

3.6 Temporary conclusions: two stalactites

Above, we represented the Kyoto-centered “dialect concentric propagation theory” as a graph rather than a map. Two-dimensional figures of Schmidt's Wellentheorie (wave theory) in historical linguistics and Thuenen’s isolierte Staat (isolated country) in geography were rendered as one dimension by using the distance from the center, allowing a comparison of the 1910 and 1980 data showing only one-dimension. Using the correlation coefficient, we were able to compare the appropriateness of grouping all prefectures with respect to standard word usage rate and railway distance. The correlation coefficient allowed us to notice slight differences which are not observable by eye. The use of the average value of LAJ’s many items was effective for grasping the whole of Japan. The trend we saw was of many words, not individual words. The Law of Large Numbers works. We discovered geographical trends in communication and the effects of people's interaction on language usage. Differences according to time, region and transportation methods were observed.

The standard form usage rate of LAJ matched well with the limestone cave model (Figure 3). The two stalactites were reflected in the graph from Kyoto 1980 in Figure 10,

which reflected the completion of the railway along the Japan Sea. The rapid growth of Edo=Tokyo was also visualized by the limestone cave model, indicating the model of dialect differentiation and propagation process occurring over more than one thousand years. The limestone cave model is a modern version of the "dialect concentric propagation theory" and was established because Edo=Tokyo appeared as the center of culture and transmission. The strength of the explanatory power of the railway distance with two centers of Kyoto and Tokyo was thus exhibited.

The geographical distribution of the LAJ standard form usage rate corresponds to the traffic routes during the adolescence of the LAJ informants and can be interpreted as reflecting the traffic route available prior to the invention of railway. The areas through which the Five Main Roads of the Edo period pass have a high standard form usage rate. The high usage rate around Kanto Region reflects the frequent traffic before modernization. In contrast, the indirect route from Tokyo to Hokuriku Region, and the fact that railways did not reach western Kyushu (due to competition with Setouchi boats) worked as hindrances.

The two stalactites were clearly visualized by railway distance taking 1980 Kyoto as the center. It is also recognizable in the Figure of 1910, but it is not as clear as in 1980 because the railway along the Japan Sea had not been constructed. The two stalactites are consistent with the results of multivariate analysis of LAJ vocabulary data (Inoue 2008). It was a model obtained by a technique which does not factor in geography, only the structure inherent to the linguistic data. The result of multivariate analysis can be simplified, and complex structures can be simply reproduced by arithmetic calculation of the data. Even switching the linguistic data to *Grammar Atlas of Japanese Dialects* (GAJ), gave almost the same results (Yarimizu et al. 2004, Yarimizu 2009). Proceeding progress will be discussed in another paper.

3.7 Reading history from geography

The distribution pattern based on railway distance dealt with in this paper was divided into (1) a gentle-sloping stalactite centering on Kyoto and (2) a steep-sloping stalactite centering on Tokyo, and these were interpreted historically as (1) gradual

propagation of various words in the past and (2) a large number of words propagating from Tokyo in Modern times. However, given the clear correspondence with the map as we saw along the Japan Sea in Figure 10, a doubt arises. A geographical interpretation may be all that is required, rendering historical interpretation excessive. Below, to confirm the validity of the historical interpretation, we will observe the structure of the LAJ data itself making use of year of the first appearance in text. In section 3, the Kasai data were analyzed in relation to four different conditions. In section 4, it will be divided into four groups.

4. Historical interpretation of the two stalactites of dialect propagation

4.1 Ancient / Modern and East / West clusters by railway distance from Kyoto

Let us examine the limestone cave model incorporating historical information. We will use the railway distance from Kyoto, by which two stalactites were clearly observable. The 82 words of the Kasai data will be first divided into east and west. The east-west difference of the LAJ data was pointed out early on in Japanese dialectology. It was quantitatively ascertained by standard form usage data using the result of cluster analysis. Inoue (2008) has divided the LAJ Kasai data into an East cluster and a West cluster.

As a second factor, the Kasai data will be divided into Ancient[5] and Modern, taking into account the year of first appearance of the words in literature. The boundary is the 14th century because there was no word appearing first in the 14th century in the Kasai data. These four categories, "Ancient / Modern and East / West" were shown on maps and were analyzed in three-dimensional graphs by combining detailed numerical values of the first year, average railway distance, and the LAJ usage rate (Inoue 2010.2).

[5] The term “ancient” is used here because it was adopted in Inoue (2004). It includes centuries before 8c when there are scarce evidences of written Japanese.

We will here verify these relationships using diagrams of the railway distance from Kyoto. Below, four figures categorized according to "Ancient / Modern, East / West" elements are displayed in a form which corresponds to the limestone cave model.

This visualization had a significant effect. As shown in the following four Figures, there are big differences in pattern between the West and East clusters. In the West cluster shown in Figures 13 and 14, there is a gentle slope in the vicinity of Kyoto, and the two figures largely coincide along the Japan Sea. Compared with this, in the East cluster in Figures 15 and 16, there is a steep slope in the vicinity of Tokyo, indicating that there was a large amount of propagation originating around Tokyo. The difference between Ancient and Modern is not large, but historical interpretation of it is possible.

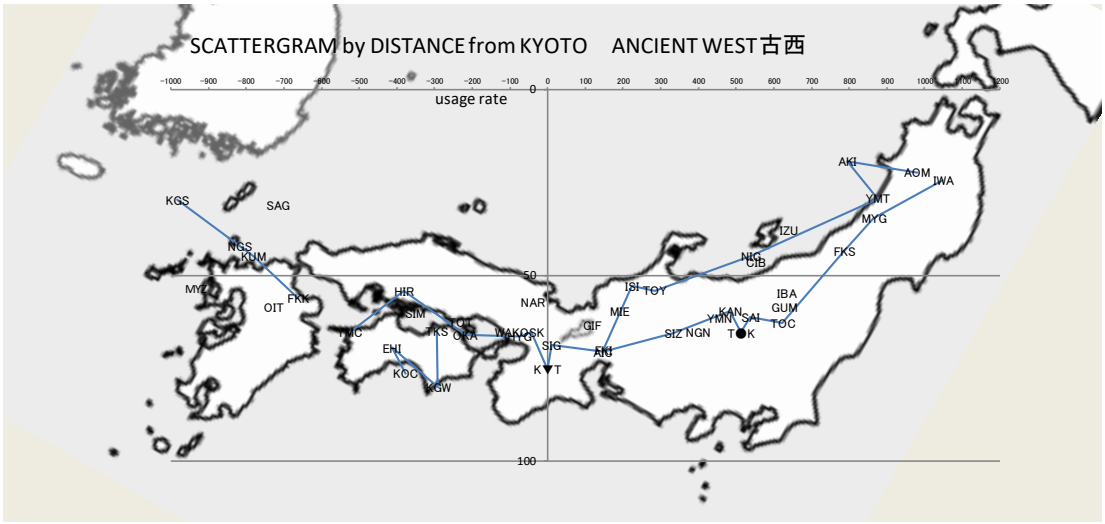

Figure 13. Ancient West cluster by Kyoto railway distance.

Let us look at the four figures according to historical order of propagation. Figure 13 Ancient West shows data for 19 words in the West cluster which first appeared prior to the 13th century. It resembles a gentle-sloping stalactite spreading to east and west Japan from the center of Kyoto. However, the peak is not Kyoto but Shikoku. This is a result of the survival of the old words in the periphery of Kyoto, which happened because newer words spread to the immediate vicinity of Kyoto later on, corresponding to the “Kinki hollow distribution” which will be described later. In the Ancient days the propagation pattern centered mainly on Kyoto. Approximately the same usage rates are

obtained at both ends of the Japanese mainland at about 1000 km from Kyoto, looking nearly symmetrical. In East Japan, there is a difference between the Japan Sea side and the Pacific Ocean side, where the Japan Sea side has a lower usage rate. This may be partly because the official large and middle roads in ancient Heian dynasty did not pass through the Japan Sea side. In Eastern Japan, the stalactite near Tokyo is inconspicuous., showing the influence of Tokyo had not reached the West cluster.

Figure 14 shows data for 20 words in the West cluster which first appeared sometime after the 15th century. Like Figure 13, it resembles a gentle-sloping stalactite spreading over east and west Japan. Kyoto itself is the peak this time. The lines look symmetrical from Kyoto to Kyushu and to Tohoku. The west area through Chugoku has a high usage rate according to the distance from Kyoto. In contrast, the line to East Japan has a steep slope showing the low usage rate on the Japan Sea side. Tokyo shows less influence from Kyoto in Modern times. In Modern times, Kyoto's influence continues in West Japan but seems to have waivered in East Japan.

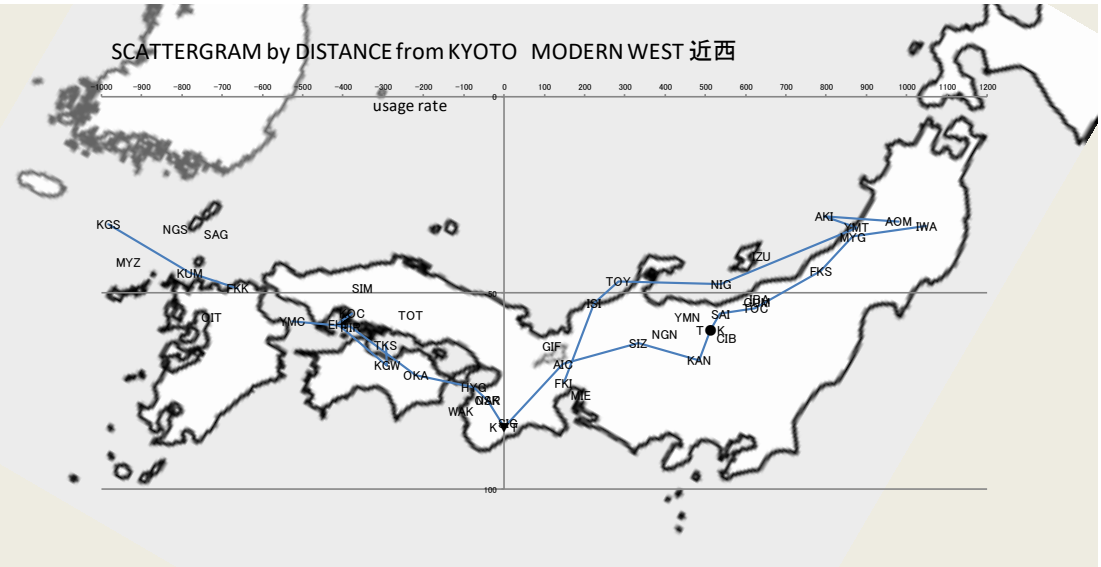

Figure 14. Modern West cluster by Kyoto railway distance.

Figure 15 shows data for a small group of 15 words in the East cluster that first appeared prior to the 13th century. The first appearance in documents is Ancient, but newer forms of them seem to have appeared near Kyoto. Since older forms continued to be used near Tokyo until Modern times, they revived as a standard word and showed a

rise in their stylistic class in literary language. For that reason, they are widely used in the Kanto and Kyushu Regions. The large stalactite hanging in eastern Japan is conspicuous. The standard form usage rate in north Kanto Region was higher than Tokyo itself, being plotted in the Pacific Ocean. In the west, Kyushu surpasses the Kinki Region and hangs low. As shown in Figure 13, propagation of the first words in ancient times centered on Kyoto. There are great differences between the Japan Sea side and the Pacific Ocean side of East Japan, wherein only propagation to the Pacific Ocean flourished. The revival pattern in Figure 15 is an exception.

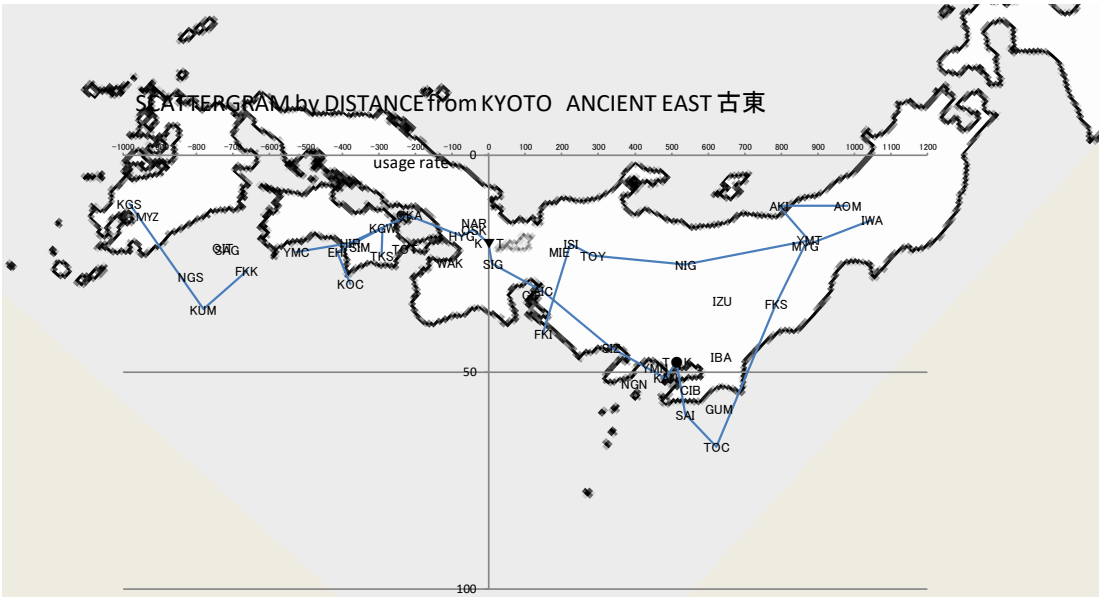

Figure 15. Ancient East cluster by Kyoto railway distance.

Figure 16 shows the data of 28 words in the East cluster which first appeared sometime after the 15th century. They represent a majority in the data, occupying more than one-third of the 82 forms. Here for the first time, Tokyo is the peak of the slope. There are many words which were born in Edo=Tokyo and became popular in Modern times. Propagation is noticeable in the area through which the Five Main Roads of the Edo Period pass. This causes differences to appear between the Japan Sea side and the Pacific Ocean side of eastern Japan. For West Japan, the slope of the railway distance is not noticeable, showing that influence from Tokyo in Modern times was weak in West Japan. In contrast, the stalactites show a sharp decline at the northern end of Tohoku

Region, showing a high standard form usage rate. The same trend can be confirmed in the Figure 5 Tokyo cluster graph in Inoue (2004). Note that the vertical axis of the graph has been condensed to match the relative positions of Tokyo and Kyoto on the map of Japan.

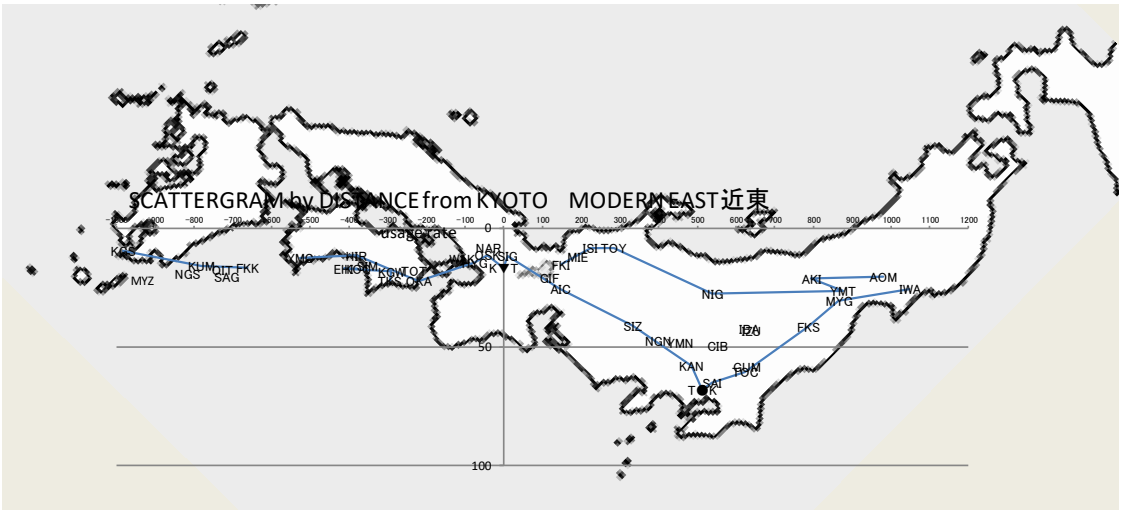

Figure 16. Modern East cluster by Kyoto railway distance.

4.2 Diachronic interpretation of four synchronic clusters

Taken together, the differences between the East-West clusters are large, and the differences between Ancient and Modern are small. In the first half of this article we observed the gradual propagation from Kyoto and the sudden propagation from Tokyo using the line graph pattern. To affirm its validity, we compared the results with historical information. The data is arranged in historical order of propagation as summarized in Table 2.

Ancient West	Fig. 13 Propagation from Kyoto to the whole country
Modern West	Fig. 14 Propagation from Kyoto to West Japan
Ancient East	Fig. 15 Revival of ancient words, Kinki hollow distribution
Modern East	Fig. 16 Propagation from Edo=Tokyo to Kanto

Table 2. Historical propagation process of Ancient-Modern East-West.

Both the Ancient and Modern West clusters reflect the propagation from Kyoto. Besides that, the Ancient East cluster showed the remnants (and revival) of the former propagation originating in Kyoto. Newer words were later born and spread from Kyoto, but they were not adopted as standard forms, instead treated as colloquial or slang. Accordingly, they can be named "Kinki hollow distribution". In the end, only the Modern East cluster of Figure 16 shows the power of new city Edo=Tokyo. When we observe Ancient words, appearing before the 13th century, both the East and West cluster show the power of Kyoto. It is largely taken for granted that standard Japanese is based on the speech of the uptown middle class of modern Tokyo. Given that, it is unexpected that the Kasai forms should be classified into almost the same number when divided into an East cluster (15 + 28 = 43) and a West cluster (19 + 20 = 39). Thus, the fact that the flow from Kyoto can be seen also in the basement of Ancient East gives an even more unexpected impression. This shows that Kunio Yanagita's focus on Kyoto's influence in *Kagyu Ko (Reflections on Snail)* (1943) is highly insightful.

There were no big differences regarding the years of the first appearance (between Ancient and Modern graphs), but a fine difference is suggestive. The positions of the peaks did not correspond to Kyoto or Tokyo in two Figures of Ancient. Figure 13 shows the peak in Shikoku and indicates the survival of the ancient words in the periphery. Meanwhile, Figure 15 Ancient East has its peak in the periphery of Kanto Region, and also indicates the survival of the ancient word.

The difference between the East and West clusters was large, and we saw two stalactites at the centers of propagation. When connecting the railway route along the Japan Sea side of eastern Japan, old-fashioned areas gather together and show bilaterally symmetrical patterns similar to the Chugoku Region in western Japan. This shows that interpreting the graph of Kyoto 1980 in Figure 10 as indicating gradual propagation from Kyoto along the Japan Sea side was not a mistake. The route on the eastern Japan Sea side corresponds to the West Japan cluster, and the East Japan Pacific Ocean side route reflects the East Japan cluster. For Western Japan, the cluster of West Japan and East Japan are almost identical, indicating overlapping patterns. In other words, using geography as a basis for interpreting history was not unreasonable. In the limestone cave model, the gentle-sloping stalactite hanging down from Kyoto was interpreted as Ancient

movement, and the steep-sloping stalactite hanging down from Tokyo was interpreted as Modern movement. This is consistent with the year of first appearance of forms in Ancient / Modern.

The limestone cave model shown in Figure 3 was typically applied to Kyoto 1980 in Figure 10 and the modern eastern cluster in Figure 16. In other words, the separation of two stalactites from Kyoto and Tokyo appeared in a stage when the power of modern Tokyo became big enough.

In an analysis of data from junior high school students, Inoue (2001, 2004) found that the usage rate of standard words became nearly 90%. Based on this data and the past dialect eradication campaign and the current treatment of dialect as entertainment, he proposed three dialect types (eradication, description and entertainment) classified according to their historical sociolinguistic status. Yarimizu (2009) has further subdivided historical types. Regarding the Edo period, we should consider shifting of the cultural power from the west to east (from Kyoto to Edo=Tokyo), but it is beyond the scope of this paper. By analyzing the synchronic state, it was possible to restore or reconstruct the state in the past. In other words, the synchronic state reflects the diachronic state. Also, in the words of Grootaers, “geography reflects history” and “space reflects time” (Inoue 2011.7).

4.3 Geographical interpretation of four synchronic clusters

The Kyoto-centered stalactite is a combination of ancient and modern Western clusters. In other words, it consists of new words spread out from Kyoto and older words cast out by the newly expanded words from Kyoto. For this complicated situation, the slope became gentle. The hanging stalactite around Tokyo is steep. It consists of many words with large usage rate over a short distance, distinct from the gentle slope over the center of Kyoto. The propagation speed of the standard forms is different between the Ancient Kyoto center and the Modern Tokyo center, and the difference in slope of these lines reflects the standard propagation speed per year, or average over 1km/y (Inoue 2003, 2010.7, 2016).

Apart from the above, cluster analysis of the Kasai data has given results of 4 clusters, 5 clusters and 6 clusters (including unpublished findings). In all cases, the Northern Alps mountains between Gifu GIF and Nagano NGN prefectures in Chubu Region appear to serve as an east-west boundary. This is different from the boundary of pitch accent and phonology. However, the dialect boundary line along the Pacific Ocean side, Aichi AIC and Shizuoka SIZ cannot be separated by a line, as they are either the same or similar.

5. Railway distance, language and society

5.1 Effectiveness of railway distance and limestone cave model

Using the railway distance, we have proposed a model incorporating two flows of propagation. The graphs which served as evidence were corroborated by the geographical position (proximity effect) by overlapping the graphs with the Map of Japan. Moreover, it showed the effectiveness of the limestone cave model based on the hypothesis that the pattern reflects historical propagation. Next, let us consider non-linguistic phenomena where the Japan Sea side tends to differ from the Pacific Ocean side. Examining the relation between linguistics and non-linguistic data will allow an evaluation of the extent to which standard form usage is influenced by economic factors.

5.2 Standard language usage rate and prefectural income

First, we will examine economic phenomena which seem to form the basis of many social and cultural phenomena. The prefectural income and the LAJ standard language usage rate show a close correlation (Inoue 2007).

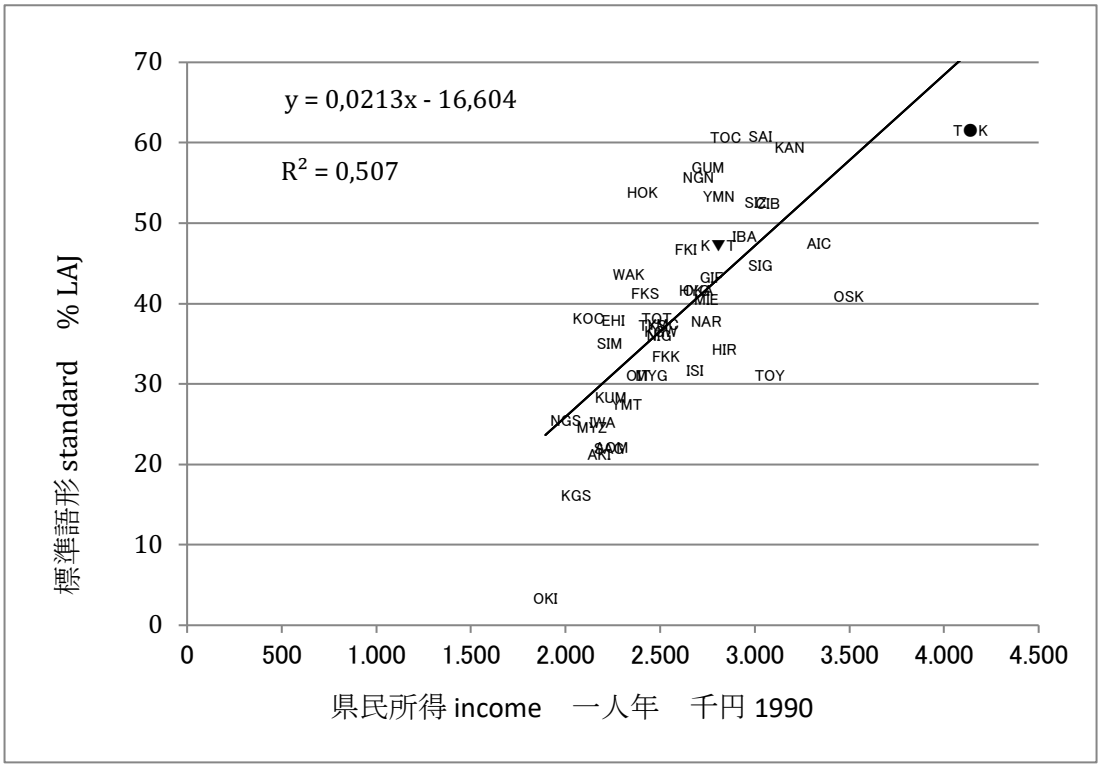

Figure 17. Standard form usage rate in LAJ and prefectural income.

Figure 17 is a scattergram of the LAJ standard form usage rate and prefectural income (Inoue 2011.12). There is a positive correlation. The correlation coefficient is 0.51, which is high for a human societal phenomenon (in comparison with the numerical values in Table 1). Prefectural income near Tokyo is overwhelmingly high, as is the standard form usage rate. Conversely, remote Tohoku and Kyushu, etc. have a low prefectural income and a lower standard form usage rate. There are exceptional prefectures such as Osaka and Aichi (Nagoya city) with high income and low use of the standard forms, but they are not disparate enough to hurt the trend given by the approximate steep straight line in the graph. The standard word usage rate and the prefectural income show close correlation. The two stalactites were explained by the historic propagation of the forces of the two cultural centers of Kyoto and Tokyo, but they can also be explained as a pattern dependent on the third factor of income level.

In other words, the controlling factor in the low usage rate of standardized words along the Japan Sea is railway distance. It also shows parallelism with low prefectural income. Also, in the nationwide distribution of dialect image, the intellectual image of

Tohoku and Hokuriku is negative (though emotional images are positive and rich) (Inoue 2007, 2011.12). The images of rural regions and the people who live there can neither be said to be good (Preston ed. 1999). This is fueled by a generally-held stereotype causing discrimination concerning regional income level, standard of living, and degree of civilization.

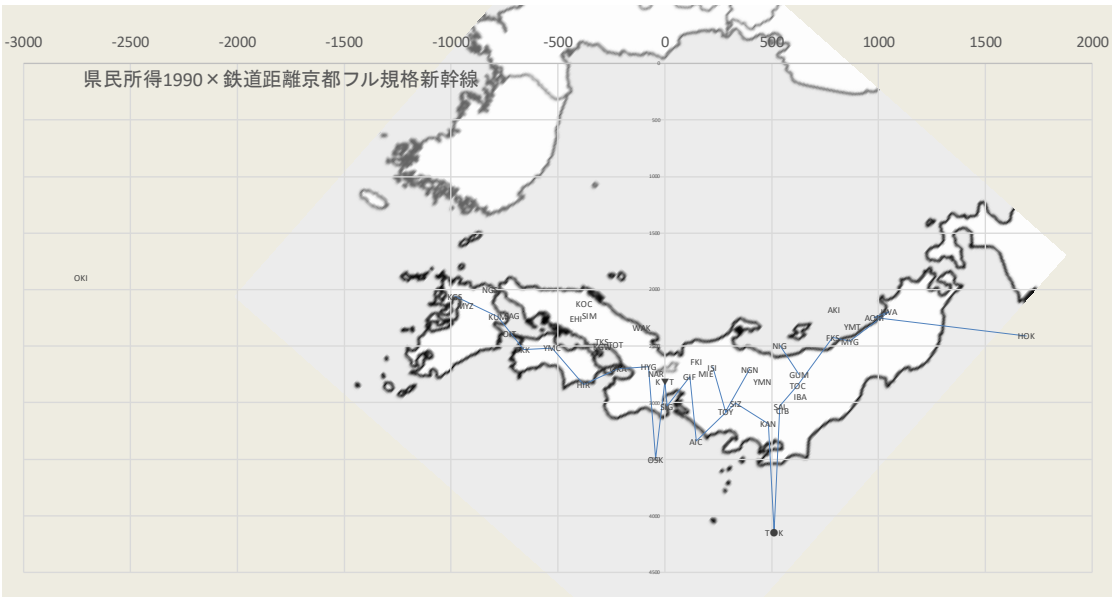

Figure 18. Prefectural income and railway distance.

In Figure 18, the prefectural income is shown in the same way as the previous scattergrams on maps. The vertical axis is prefectural income, but since it was reversed, the low-income prefectures are plotted on top. The horizontal axis divides the railway distance from Kyoto to the East-West and the link connects the Shinkansen bullet train routes as of 2018. Prefectures not on the Shinkansen route are not connected with a line. Rendering the geographical location on one dimension allowed us to show income as a continuous value on another dimension.

Two major trends can be read. (1) The prominence of prefectures with big cities. Tokyo, Aichi and Osaka are plotted in the lower Pacific Ocean. (2) On the Pacific Ocean side, income is large, and they are connected by Shinkansen. On the other hand, the Japan Sea side and Shikoku, etc. have low income and are plotted without the Shinkansen line. These prefectures show a gentle slope, and their income is largely inversely

proportional to the railway distance from Tokyo. The East and West show different patterns. In the west of Kyoto, the line heads to the upper left toward Kyushu. In the east of Kyoto, the lines are divided into two: The upper right line heading for Tohoku and the line going down towards Tokyo. The Hokuriku prefectures of Fukui, Ishikawa, and Niigata can be seen as relay points with the Tohoku Region. This interpretation is made possible by the two-stalactite standard form hypothesis. Normally we would be distracted by the great prefectural income of Tokyo, Aichi and Osaka.

The pattern in Figure 18 is a complicated version of the limestone cave model shown in Figure 3. The distribution pattern resembles Figure 12. However, in Figure 18, the overall correspondence with the railway distance (i.e. income to distance) collapses. In other words, railway distance is a bigger factor in standardization than income.

5.3 Railway distance and prefectural income

Below, we explain the explanatory power of railway distance. Looking at a simple scattergram of railway distance and prefectural income (not shown here), it shows inverse correlation. The farther from Tokyo the lower the prefectural income. The correlation coefficient is 0.42, which is not so high for a social phenomenon. It shows a tendency for the standard word usage rate, railway distance and prefectural income to be similar. The correlation coefficients are as follows.

Standard word usage rate and railway distance	0.55 - 0.58
Standard word usage rate and prefectural income	0.51
Railway distance and prefectural income	0.42

In other words, the standard word usage rate and the railway distance are the most relevant among the three. From a global viewpoint, language is dominated by economy (Inoue 1997, 2011.12). Regulation by geographical distance is also great. The development of the railway network is dominated by the economy. Taking this into

account, this paper can be regarded as a preliminary work in a field of economic linguistics.

5.4 Standard language usage rate, railway distance and income

We have considered the relationship between language (standard language usage rate), geography (railway distance) and economy (prefectural income). In conclusion, geography is a larger factor than economy in language standardization. There are other related, similar phenomena. The Sapir-Whorf hypothesis is hard to be applied. The influence of the standard form usage rate on the other phenomena is not large, but the reverse is true. The influence on standard language usage rate from railway distance and income is a product of population concentration caused by urbanization and of the resulting diversity of residents. The mechanism can be explained as follows. People from various places are drawn to the city by promise of higher income. If the population grows denser, the railway density in that area also increases. The traffic routes are improved and the potential for economic development increases. As communication becomes frequent, common language is demanded, and standard form usage rate grows. Standardized language is based on literary or written language and urban areas are home to many highly-educated white-collar workers who use such language. All of these phenomena are intertwined to form complicated causal relations. In short, the urbanization demands common language for exchanges.

One task for future research is to gather all of these factors together and analyze the causal relationships by multivariate analysis. In terms of language, GAJ grammatical phenomena (Yarimizu 2009), LAJ lexical phenomena (Kumagai 2016, Kumagai ed. 2013), new dialect and dialect image (Inoue 2007) should be incorporated. On the extralinguistic side, many factors related to traffic and human exchange (passengers, transport volume, number of express trains etc.) would be necessary. Population size, population density, population shift, industrial structure, trade volume and hierarchical social structure should also be taken into account. Also, incorporating data from the past would allow deeper insights.

Differences between the Pacific Ocean side and the Japan Sea side are apparent in many of the above non-linguistic figures. However, unlike the standard language usage rate, it is difficult to find two stalactites, i.e. to apply the limestone cave models to them. Figure 12 showed a typical distribution, and interpreting two historical flows of culture based on two distinct peaks was not unreasonable. Language is a reflection of many cultural phenomena and, since individuals' adoption of language is free of charge, the process of propagation can be observed as a typical phenomenon.

6. Conclusion: Evaluation of limestone cave model

6.1 Effectiveness of limestone cave model

In this paper, we applied a new representation method using existing numerical data and tried to achieve deeper insight from geographical and historical points of view. By improving the visualization method, the depth and width of the analysis increased. The standard form usage rate and railway distance were not just applied to scattergrams, but were further visualized by plotting left-to-right, East-West to show how they correspond with the Japan archipelago. Further, by overlapping the line graphs with the Map of Japan, the correspondence between the trend and geographical position became more concrete. Comparison of the propagation of the standard forms with hanging stalactites allowed an intuitive understanding of its correspondence with the railway route. We were able to derive the propagation from former capital Kyoto and from modern capital Edo=Tokyo from the synchronic distribution data of the LAJ. In sum, we succeeded in understanding the history of Japanese language as it is linked with the formation process of dialect distribution.

Although the railway distance from Kyoto did not appear useful when comparing the correlation coefficients of simple scattergrams, in analyzing east/west differences of Japanese archipelago, distance from Kyoto turned out to be more useful in explaining the mechanism of diffusion. The 82 words of the Kasai data were then divided into four

groups and compared with the map of Kyoto distance in 1980. The historical process of standardization was clearly explained. The fourth groups of words, that is, the Modern Tokyo cluster, was found to show the powerful standardization spreading from Tokyo. On the other hand, the other three groups showed the long-persisting, complicated influences of Kyoto in various ways. The limestone cave model proposed here can also be applied to other languages such as English in England (Inoue 1996) and Italian.

6.2 The future of the limestone cave model

Although we were not able to discuss in this paper, we also drew limestone cave model graphs for individual forms. In order to examine the word history, all of the 82 Figures of the Kasai data were examined for the year of first appearance, and words which were remaining dialect forms, or the survival of the ancient forms were also noted. The presentation will however be given at other opportunities.

In addition, based on the correspondence with the year of first appearance, we were able to calculate the propagation rate of individual words. There is a possibility that the Modern standard forms spread faster than the basic 1 kilometer per year (Inoue 2003). For LAJ, data for each locality (or informant) including the dialect forms is being prepared. The propagation of dialectal words is likely different in speed. It is also possible to calculate for GAJ grammatical phenomenon, so this will be a task for the future.

6.3 The Future of railway distance

Railway distance is related to socio-economic (extra-linguistic) phenomena. Numerical charts and graphs of numerous social phenomena and cultural phenomena are currently being published, and mapping of these phenomena is also prosperous. As dialect researchers, we are distracted by the East-West difference and we are looking for the distribution of the Kyoto-centered peripheral zones, but the number of instances is small. The majority is the Tokyo-centered distribution. Some of them also show the contrast between the Pacific Ocean side and the Japan Sea side. In other words, the

correlation of the railway distance is higher than the straight distance for many of the phenomena.

This paper has contributed to uncovering the relationship between the railway distance and linguistic phenomena. Traffic indicates human exchange as do dialects. Further comparison of how traffic routes relate with modern dialectal distribution and standard form usage rate in the past and in the present seems to be a worthwhile endeavor. Dialect distribution also depends on the systematicity of language itself. Another factor is human contact. Uncovering the process of the formation of modern dialect distribution requires an examination of extralinguistic factors as well. Railway distance has proved to be an appropriate indicator for making observations about traffic over a long period in Modern times.

Prefecture abbreviation list

Prefecture abbreviations in this paper are classified by Regions as follows:
HOK Hokkaido. *TOHOKU*: AOM Aomori, IWA Iwate, MYG Miyagi, AKI Akita, YMT Yamagata, FKS Fukushima. *KANTO*: IBA Ibaraki, TOC Tochigi, GUM Gunma, SAI Saitama, CIB Chiba, TOK T●K Tokyo, IZU Izu Islands (Tokyo island department), KAN Kanagawa, *HOKURIKU*: NIG Niigata, TOY Toyama, ISI Ishikawa, FKI Fukui. *CHUBU*: YMN Yamanashi, NGN Nagano, GIF Gifu, SIZ Shizuoka, AIC Aichi. *KINKI***:** MIE Mie, SIG Shiga, KYT K▼T Kyoto, OSK Osaka, HYG Hyogo, NAR Nara, WAK Wakayama. *CHUGOKU***:** TOT Tottori, SIM Shimane, OKA Okayama, HIR Hiroshima, YMC Yamaguchi. *SHIKOKU***:** TKS Tokushima, KGW Kagawa, EHI Ehime, KOC Kochi. *KYUSHU***:** FKK Fukuoka, SAG Saga, NGS Nagasaki, KUM Kumamoto, OIT Oita, MYZ Miyazaki, KGS Kagoshima. *OKINAWA***:** OKI Okinawa.

References

* English papers by F. Inoue before 2008 are accessible through the internet.
http://dictionary.sanseido-publ.co.jp/affil/person/inoue_fumio/doc/

http://www.urayasu.meikai.ac.jp/japanese/meikainihongo/18ex/achievements.xls

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

GOEBL, Hans (1982) *Dialectometrie: Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialectgeographie,* Oesterreichische Akademie der Wissenschaften.

GOOSKENS, Charlotte (2005) "Travel time as a predictor of linguistic distance", *Dialectologia et Geolinguistica*, 13, 38-62.

GOOSKENS, Charlotte & Wilbert HEERINGA (2004) "The position of Frisian in the Germanic language area", in Dicky Gilbert, Maartje Schreuder, & Nienke Knevel (eds.), *On the boundaries of phonology and phonetics*, Groningen: Klankleergroep, Faculty of Arts, University of Groningen.

HEERINGA, Wilbert (2004) *Measuring Dialect Pronunciation Differences using Levenshtein Distance,* Groningen Dissertations in Linguistics, 46.

INOUE, Fumio (1996) "Subjective dialect division of Great Britain", *American Speech* 71.3, 118-135; also in Preston (1999).

INOUE, Fumio (1997) "Market value of languages in Japan", *Japanese Linguistics*, 2, 40-61.

INOUE, Fumio (2001) *Keiryoteki Hogen Kukaku (Quantitative Dialect Classification)*, Tokyo: Meiji Shoin.

INOUE, Fumio (2003） *Nihongowa Nensoku 1 kilo de Ugoku* (*Japanese language moves 1 km per year),* Tokyo: Kodansha.

INOUE, Fumio (2004) "Multivariate analysis, geographical gravity centers and the history of the standard Japanese forms", *Area and Culture Studies*, 68, 15-36.

INOUE, Fumio (2006) "Geographical distance center and rate of diffusion of Standard Japanese", in Agris Timuška (ed.), *Proceedings of the 4th International Congress of Dialectologists and Geolinguists*, Riga: Latvijas Universitāte, 239-247.

INOUE, Fumio (2007) *Kawaru Hogen Ugoku Hyojungo (Changing dialects and moving standard)*, Tokyo: Chikuma Shinsho.

INOUE, Fumio (2008) "Geographical distance center and multivariate analysis of the standard Japanese", *Dialectologia,* 1, 65-81.
<http://www.publicacions.ub.edu/revistes/dialectologia1/>

INOUE, Fumio (2010.2) “Sprachraum and infrastructure: Abstracting geographical space via railway distance”, in Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and Space: Language Mapping*, Berlin/New York: De Gruyter Mouton, 542-561, Maps 2801-2806.

INOUE, Fumio (2010.7) “Real and apparent time clues to the speed of dialect diffusion”, *Dialectologia*, 5, 5-64.
< http://www.publicacions.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 (2011.7) “First dialectologists Willem A. Grootaers”, *Dialectologia: revista electronica*, 7, 57-64.
<http://www.publicacions.ub.edu/revistes/dialectologia7/dialectologists.asp>

INOUE, Fumio (2011.12) *Keizai Gengogaku Ronko* (*Papers on Econolinguistics)*, Tokyo: Meiji Shoin.

INOUE, Fumio (2016) “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) “Age-area distribution of linguistic change in progress observed in glottograms”, *Dialekt / Dialect 2.0 Langfassungen / Long papers,* Wien: Praesens Verlag, 174-194.

INOUE, Fumio (2019) “Common development of Japanese and Dutch dialectology - Computational dialectology in Japan”, *9th International Congress of Dialectology and Geolinguistics,* Vilnius, Lithuania.

JESZENSZKY, Peter, Philipp STOECKLE, Elvira GLASER & Robert WEIBEL (2017) “Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German”, *Journal of Linguistic Geography*, 5-2, 86-108.

KUBOTA, Hiroshi (2005) *Nihon no Tetsudoshi Seminar,* Tokyo: Grand Prix Shuppan.

KUMAGAI, Yasuo (2013) “Development of a way to visualize and observe linguistic similarities on a linguistic atlas”, *Working Papers from NWAV Asia-Pacific*, 2.
<https://pj.ninjal.ac.jp/socioling/nwavap02/Kumagai-NWAVAP2-2013.pdf>

KUMAGAI, Yasuo (ed.) (2013) *Daikibo Hogen Data no Takakuteki Bunseki (Analyzing Large-scale Dialectal Survey Data from Multiple Perspectives)*, Tokyo: NINJAL.
<file:///C:/Users/owner/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempState/Downloads/review040101%20(1).pdf>

KUMAGAI, Yasuo (2016) "Developing linguistic atlas of Japan database and advancing analysis of geographical distributions of dialects", in Marie-Helene Cote, Remco Knooihuizen & John Nerbonne (eds.), *The Future of Dialects,* Berlin: Language Science Press, 333-362. <file:///C:/Users/owner/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempSt>

LABOV, William (1972) "Some principles of linguistic methodology", *Language in Society* 1(1), 97-120.

MIYAJI, Masato et al. (2013) *Meiji Jidaishi Daijiten,* 4, Tokyo: Yoshikawa Kobunkan.

NERBONNE, John, Ilse VAN GEMERT & Wilbert HEERINGA (2005). *A Dialectometric View of Linguistic "Gravity"*, Alfa-informatica, University of Grooningen. <http://odur.let.rug.nl/~nerbonne/papers/gravity2004.pdf -html>

NINJAL (National Institute for Japanese Language and Linguistics) (1989-2006), *Grammar Atlas of Japanese Dialects* (GAJ), Okurasho Insatsukyoku. <https://www2.ninjal.ac.jp/hogen/dp/gaj-pdf/gaj-pdf.index.html>

NLRI (National Language Research Institute) (1966-1974) *Linguistic Atlas of Japan*, vols. 1-6, Okurasho Insatsukyoku. <https://mmsrv.ninjal.ac.jp/laj_map>

ONISHI, Takuichiro (ed.) (2016) *Shin Nihon Hogen Chizu (New Linguistic Atlas of Japan)*, Tokyo: Asakura Shoten.

ONISHI, Takuichiro (2019) "On the relationship the degrees of correspondence of dialects and distances", *Languages,* 4(2), 1-15.

PRESTON, Dennis R. (ed.) (1999) *Handbook of Perceptual Dialectology*, vol. 1, Amsterdam / Philadelphia: John Benjamins.

YARIMIZU, Kanetaka (2009) *Kyotsugoka Katei no Keiryoteki Bunseki (Quantitative Analysis of Standardization Process)*, Doctor's Thesis, Tokyo University of Foreign studies. <http://repository.tufs.ac.jp/handle/10108/51462>

YARIMIZU, Kanetaka, Yuji KAWAGUCHI & Masanori ICHIKAWA (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.

YANAGITA, Kunio (1943) *Kagyu Ko* (Reflections on Snail), Tokyo: Sogensha.