多変量解析と地理的重心法と標準語形の歴史

井 上 史 雄

Multivariate Analysis, Geographical Gravity Centers and the History of the Standard Japanese Forms

Fumio INOUE

0. Two kinds of simplification

In this paper two kinds of simplification were attempted to represent geographical distribution patterns. The first makes use of the gravity center method, that is, the representation of a two-dimensional distribution pattern by one point. The second plots geographical locations on a one-dimensional line, an east-west line. This second simplification is further simplified by making use of railway distances from cultural centers. The overall pattern of simplification will first be presented in order to show the flow of reasoning in this paper.

0. Usually one linguistic map is drawn up for one item or one word, showing various word-forms by symbols or phonetic alphabets. This is the ordinary custom in linguistic atlases.

1. Sometimes only one word-form is represented on one map. If this is done for a word with many dialectal expressions or with many word-forms, it can be seen as a waste of resources. But the merit is that each word-form is shown separately and can be compared purely from the standpoint of distribution patterns (with semantically unrelated terms). A sample of such a map is KENKEN 'to hop on one leg'[i] from the LAJ in **Fig 1**. It would be laborious if this technique is applied to all the dialect distribution data of an atlas. For this paper, however, this technique was applied to the 82 standard Japanese forms and a computer was utilized.

2. The idea of representing individual word-forms (standard Japanese) separately on individual maps leads to the idea of simplifying the geographical distribution pattern to one point on a map using of the "geographical gravity center method"(Inoue 1996.3, 1996.10).

3. This attempt to use the geographical gravity center method in Japan inspired me to make another simplification by representing the real geographical location only in one dimension. Because the Japanese dialects are spoken along the long archipelago spreading from southwest to northeast, localities can be shown along the coastline of the Pacific Ocean. The Japan Islands can be represented in a one-dimensional line, as seen from an airplane or a space shuttle flying over the Pacific Ocean (Inoue 1996.10).
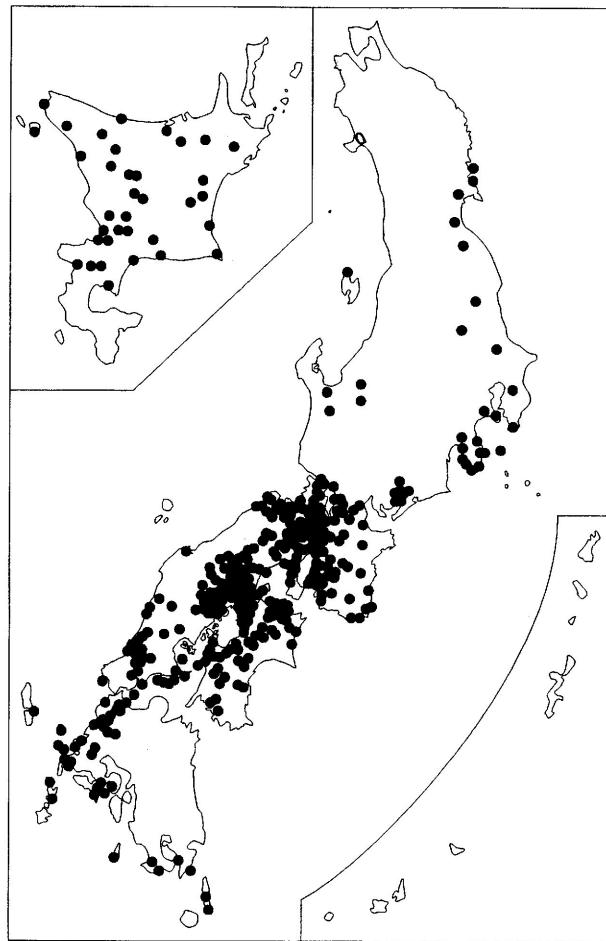
Fig 1. KENKEN 'to hop on one leg' from LAJ

4. During the gravity center method analysis, the power of influence of the old capital Kyoto and the modern capital Tokyo became the main topic of analysis. This lead to the idea of making use of the geographical distance from Kyoto or Tokyo using calculations based on railway distance. Prefectures situated in the east, west, south and north can be treated in the same way by this technique, and can be plotted in the same direction on a new kind of graph. This technique corresponds to the idea of folding the Japanese Islands into two at the axis of Kyoto or Tokyo. This procedure can be said to represent a circular model of the "Isoliert Staat," a system defined by geographer Thünen in a one-dimensional graph and further in a one direction graph; or in linguistics, to represent the figure of Wellentheorie (wave-theory) by Johannes Schmidt in one direction.

5. As the geographical locations are shown in one dimension by the technique above, another dimension on a sheet of paper can be utilized for another factor. In this paper the first appearance of the word form in historical documents will be considered. Recently I have also analyzed the ratio of standardization nationwide for each prefecture and a very interesting theoretical tendency was discovered (Inoue, in print). However in this paper I will only consider the relationship between geographical distribution patterns and the first appearance of a word in historical documentation.

Now I will first quickly review what I presented at the Joensuu conference in 2002 (§1-3), and then I will proceed to my new technique of simplification of geographical distribution (§ 4-7).

1. Data and Maps

The recent diffusion of computers has enabled studies of computational dialectology. Several multivariate techniques have been applied to various data in many countries. I myself have analyzed a matrix of standard Japanese forms and believed I had successfully classified and divided Japanese dialects (Inoue 1996.3, 1996.10). However, I have recently discovered that the same data can be interpreted otherwise. That is, various distribution patterns of standard Japanese forms can be interpreted as showing continuous stages of dissemination.

The relation between the history and the geographical distribution of the standard Japanese forms will be discussed in this paper. The century or year in which a form was first documented and geographical distribution on the basis of the dialect gravity center method will be compared. The shift of the main cultural center of Japan from the west to the east (Kawaguchi and Inoue 2002) is reflected in the graphs obtained from factor analysis and cluster analysis, and in the geographical distribution patterns of the standard Japanese forms. Forms that disseminated from modern Tokyo are considered to be the "literary" standard.

Thus we will consider the geographical distribution patterns of the standard Japanese forms together with the century in which each form was first documented, and look at how historical aspects relate to the geographical tendencies on the maps.

The most reliable data on the dialectal distribution of Japanese is recorded in the "Linguistic Atlas of Japan" or LAJ (NLRI 1966-1974). There is also a database which shows the general tendency of standardization of Japanese. It is computational data compiled by Ms. Kasai and is referred to as the "Kasai data" (Inoue and Kasai 1989).

Fig 2 shows the basic structure of the data matrix of the Kasai data (Data for first year of appearance of word-forms and railway distance of prefectures were later added). For each of the 82 standard Japanese forms the average degree of usage was calculated for each of the 48 prefecture areas. One map was constructed for each of the 82 forms selected from the *Linguistic Atlas of Japan*.

WORD-FORM	PREFECTURES									First Year of
	Hokkaido	Aomori	····	Tokyo	····	Kyoto	Hyogo	····	Okinawa	Appearance
MABUSHII dazzling	28.5	6.8	····	88.9	····	11.1	11.2	····	0.0	1550
KOGEKUSAI smelling burnt	95.2	58.1	····	100.0	····	97.2	87.3	····	0.0	1601
NASU eggplant	15.4	0.0	····	100.0	····	18.8	18.2	····	0.0	1483
TSUYU rainy season	39.8	0.0	····	22.2	····	100.0	98.6	····	0.0	1401
····	····	····	····	····	····	····	····	····	····	····
Railway Distance	1179	739	····	0	····	514	590	····	3242	····

Fig 2. Data matrix of the Kasai data

Fig 3 TSURARA (icicle) is one of the typical distributions of a standard Japanese form that originated in western Japan.

More information was added to all the standard Japanese forms of the Kasai data. For this paper information on usage derived from historical documents was utilized. The first appearance of the forms was determined on the basis of "Nihon Kokugo Daijiten" (Great Dictionary of the Japanese Language) (2nd Edition), the largest Japanese dictionary at present, which cites historical examples.

2. Factor Analysis and Cluster Analysis

2.1. Factor Analysis

In order to show the basic patterns of distribution of the Japanese standard forms, the results of factor analysis of the Kasai data will first be introduced. By applying factor analysis to the matrix of LAJ Kasai data, a clear-cut difference between western and eastern Japanese dialects was discerned (Inoue and Kasai 1989). This can be shown in the two dimensions of the data. Firstly, factor score, or the dimension of prefectures, showed that the first factor is western and that the second factor is eastern. Secondly, the factor load, or the dimension of the 82 individual standard Japanese forms, showed that about half of the forms are characterized as
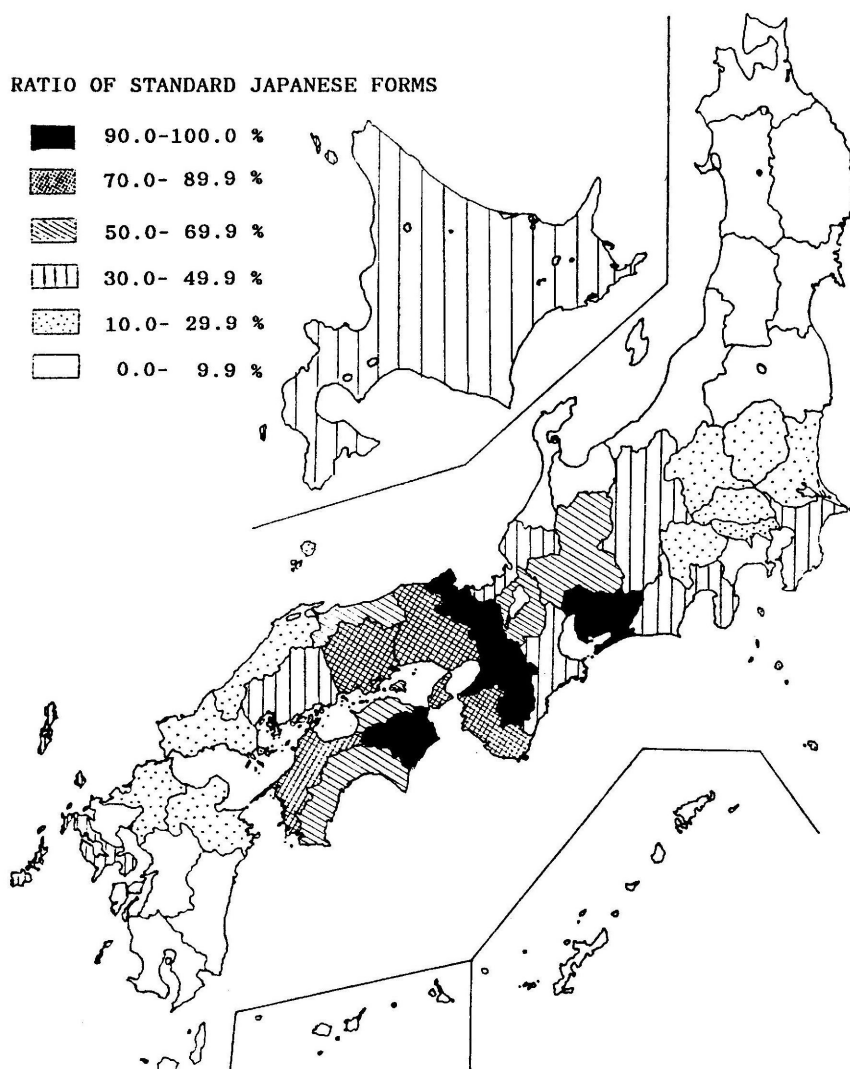
west-centered forms.

RATIO OF STANDARD JAPANESE FORMS

- ■ 90.0-100.0 %
- ▨ 70.0- 89.9 %
- ▥ 50.0- 69.9 %
- ⊞ 30.0- 49.9 %
- ▦ 10.0- 29.9 %
- □ 0.0- 9.9 %

Fig 3. TSURARA (icicle), a standard Japanese form that originated in western Japan

Fig 4 shows the graph of factor load for the 82 individual standard Japanese forms. From the above results of the LAJ Kasai data, two typical groups of the standard Japanese forms could be distinguished, that is, western standard forms and eastern standard forms. Comparison of this graph with individual distribution maps has shown that the four quadrants of Fig 4 correspond each with eastern forms, western forms, nationwide forms and sporadic literary forms, as will be discussed later. Fig 4 also shows the century of the first appearance of

a from in historical documents; see also Fig 5 and Fig 6.

The merit of factor analysis is that it can distinguish between densely distributing forms and sparsely distributing forms. This can be observed again later when we look at the results of the gravity center analysis.
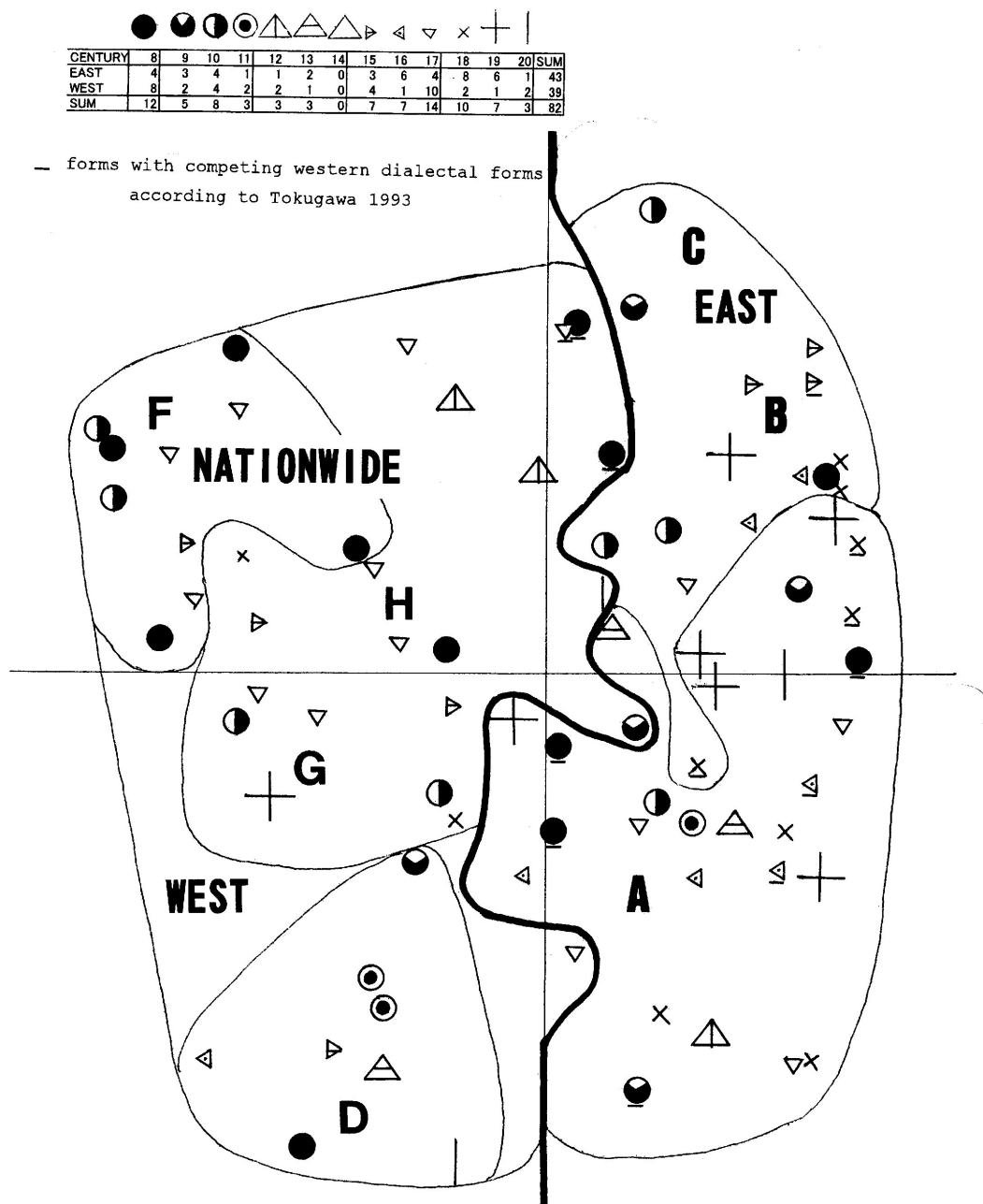
CENTURY	8	9	10	11	12	13	14	15	16	17	18	19	20	SUM
EAST	4	3	4	1	1	2	0	3	6	4	8	6	1	43
WEST	8	2	4	2	2	1	0	4	1	10	2	1	2	39
SUM	12	5	8	3	3	3	0	7	7	14	10	7	3	82

— forms with competing western dialectal forms according to Tokugawa 1993

Fig 4. Factor load for the 82 standard Japanese forms

2.2. Cluster Analysis

In Fig 4, the results of the cluster analysis of the forms are also entered by thick and thin lines which encircle the forms. The results of factor analysis and cluster analysis coincided fairly well. All the forms are first divided into the western cluster and the eastern cluster. These clusters were then further subdivided into several clusters, from A to C and D to H.

2.3. Actual Distribution Maps

The above results of the factor scores can be summarized to reflect the east-west differentiation of Japanese dialects. As has been pointed out earlier, factor 1 is the western factor, and factor 2, the eastern factor. In order to grasp the results of factor analysis more concretely, the graph of Fig 4 was enlarged, and 82 reduced maps of the Kasai data were pasted on it. This very large graph was again reduced in **Fig 5**. The small maps on the upper-left are darker, that is, the standard forms are used more heavily throughout Japan, and the maps on the lower-right are pale, that is, the standard forms are used only sparsely. The maps on the upper-right are dark in eastern Japan; the maps on the lower-left are dark in western Japan.

This technique of arranging all of the 82 maps on the basis of factor analysis has shown that the data matrix of the Kasai data clearly reflects the geographical patterns of distribution of the standard Japanese forms.

3. Gravity Center

Now we proceed to look at gravity center technique. A gravity center is the point on a map that shows the center of geographical distribution (Inoue 1996.3, 10). The point of gravity of a map is calculated so that if a piece of string was attached to a point of a stiff paper map, the map would be suspended horizontally in mid air.

The formula for calculating gravity centers is as follows:

$$\text{Gravity center (x)} = \frac{w^1 \times x^1 + w^2 \times x^2 + w^3 \times x^3 + \ldots\ldots\ldots\ldots w^i \times x^u}{w^1 + w^2 + w^3 + \ldots\ldots\ldots\ldots w^i}$$

$$\text{Gravity center (y)} = \frac{w^1 \times y^1 + w^2 \times y^2 + w^3 \times y^3 + \ldots\ldots\ldots\ldots w^u \times y^u}{w^1 + w^2 + w^3 + \ldots\ldots\ldots\ldots w^u}$$
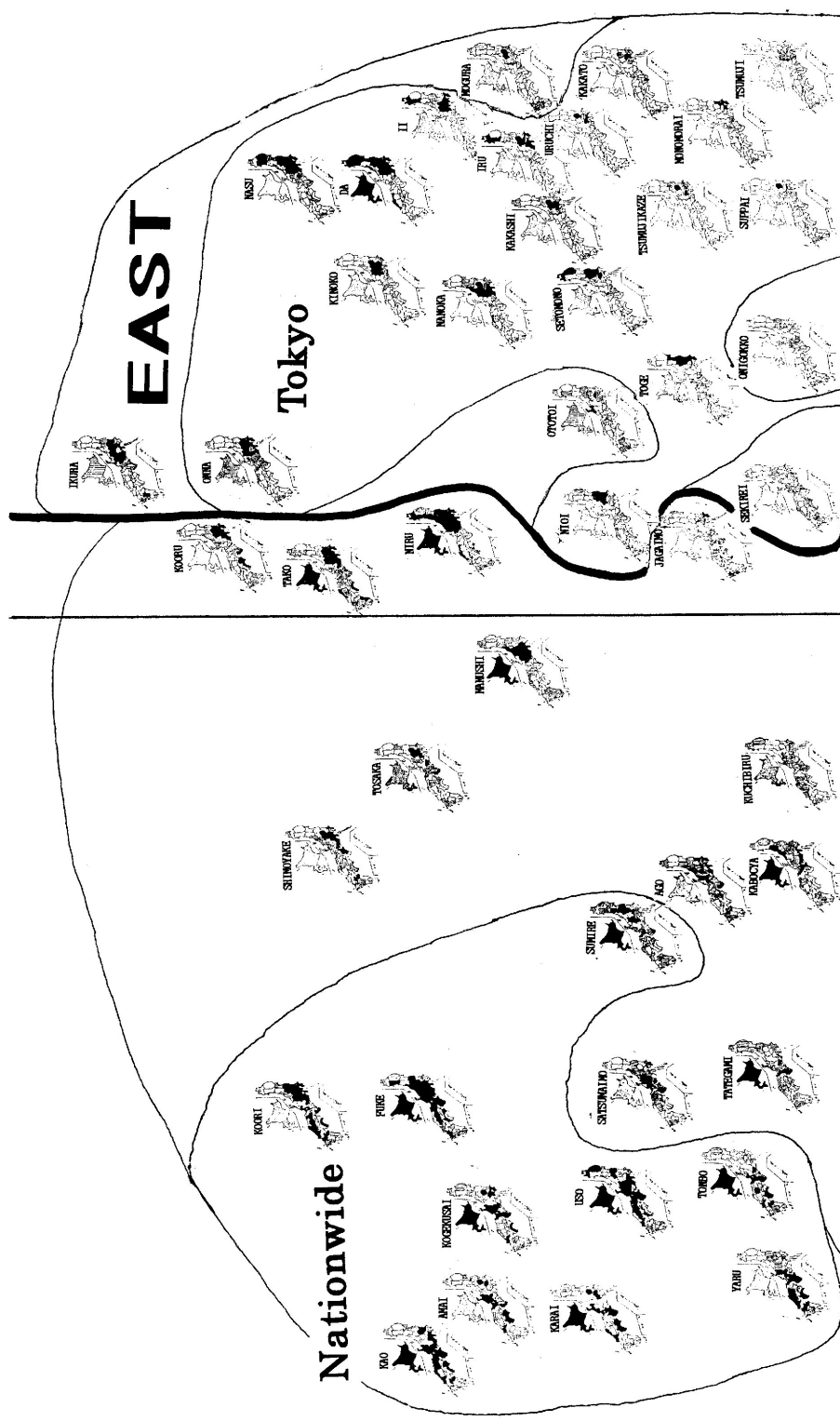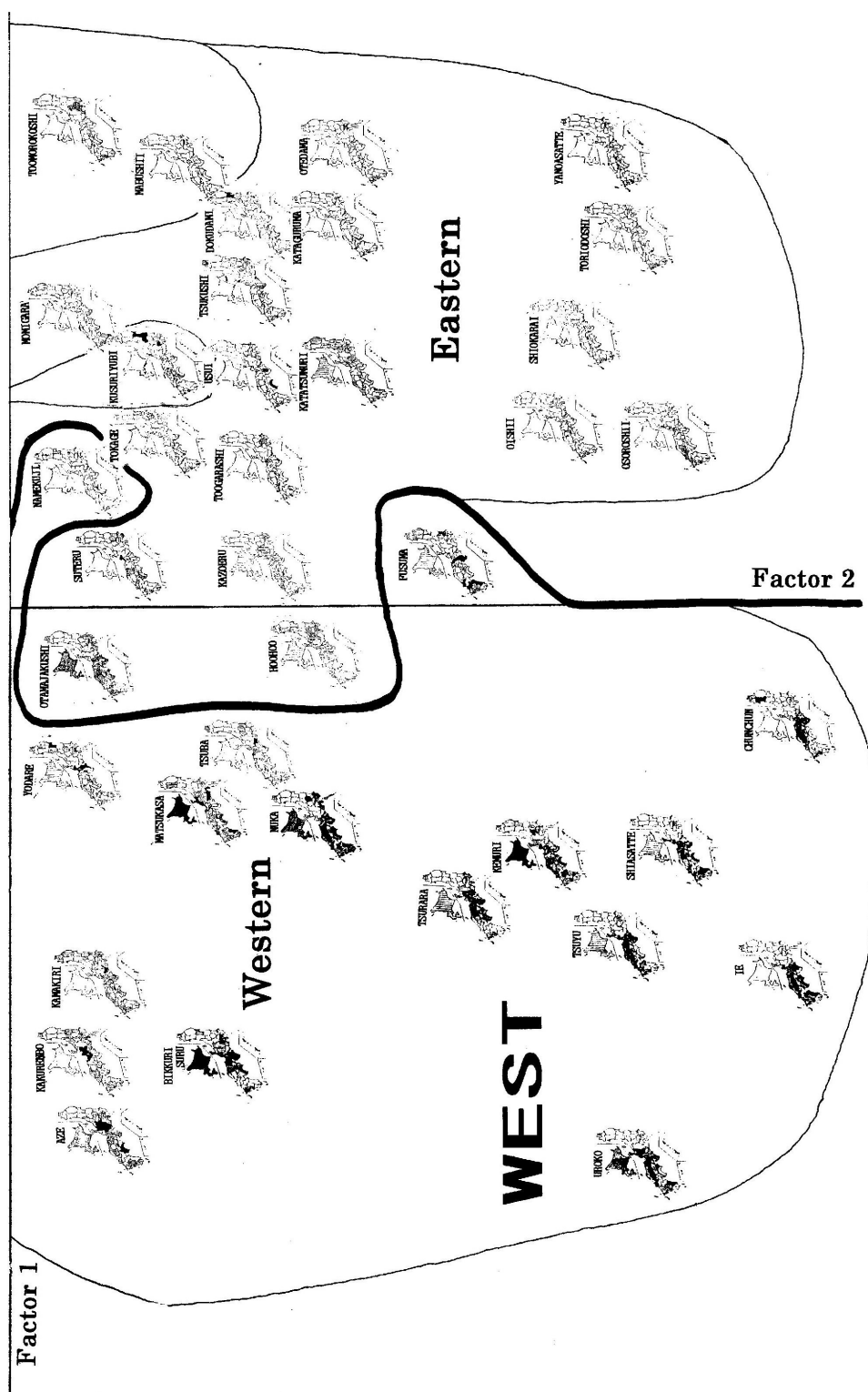
Fig 5. Reduced maps arranged by factor load

where w^1 to w^n mean "weight" (average ratio of standard forms), and x^1 to x^n and y^1 to y^n refer to the positions of prefectures in two dimensions (directions).

For the Kasai data, the positions of the capital cities of the prefectures were utilized because of a lack of better indexes.

Fig 6 shows the gravity centers of the 82 standard Japanese forms. Most of the forms are distributed in Central Japan between Kyoto and Tokyo. Fig 6 also shows the century of the first appearance of a form in historical documents together with the gravity center data. The forms which appeared early are shown by larger black, round symbols; and those which appeared in the modern ages are shown by lines (plus signs and bars). Darker forms that appeared earlier are distributed all over the areas from the west to the east.

The older standard forms seem to have experienced various histories during nationwide propagation and in the struggle with newly appearing dialectal forms. In contrast, later forms, which were first recorded after the 18th or 19th centuries, were distributed either near Tokyo or in the central area of gravity centers. They seem to have a shorter and simpler history of nationwide propagation.

4. Gravity Center and History

Now I will proceed to the portion that I added for this paper.

Fig 7 is a scattergram of the 82 standard Japanese forms. It shows geographical gravity centers on the x-axis, the horizontal dimension of the graph. Only the east-west locations of the 82 standard forms are represented. The y-axis, or the vertical scale, shows the first century of appearance in historical documentation of each of the 82 standard Japanese forms. By adding cluster analysis information to this scattergram, encircling with solid lines, three kinds of information can be examined in one figure; that is, geographical distribution, historical background and cluster groupings.

The horizontal locations of the 82 forms naturally correspond to the gravity center map in Fig 6. However the continuous relation of first year of appearance is more realistically shown in Fig 7. The continuous values of the first year of appearance can be shown more faithfully in this graph; while in the map of Fig 6, centuries are shown discretely in the form of symbols. The fact that there were no words which appeared for the first time in the 14th century is revealed clearly. The lack of documentation for this period is partly due to the turmoil of war in the feudal middle ages in Japan.
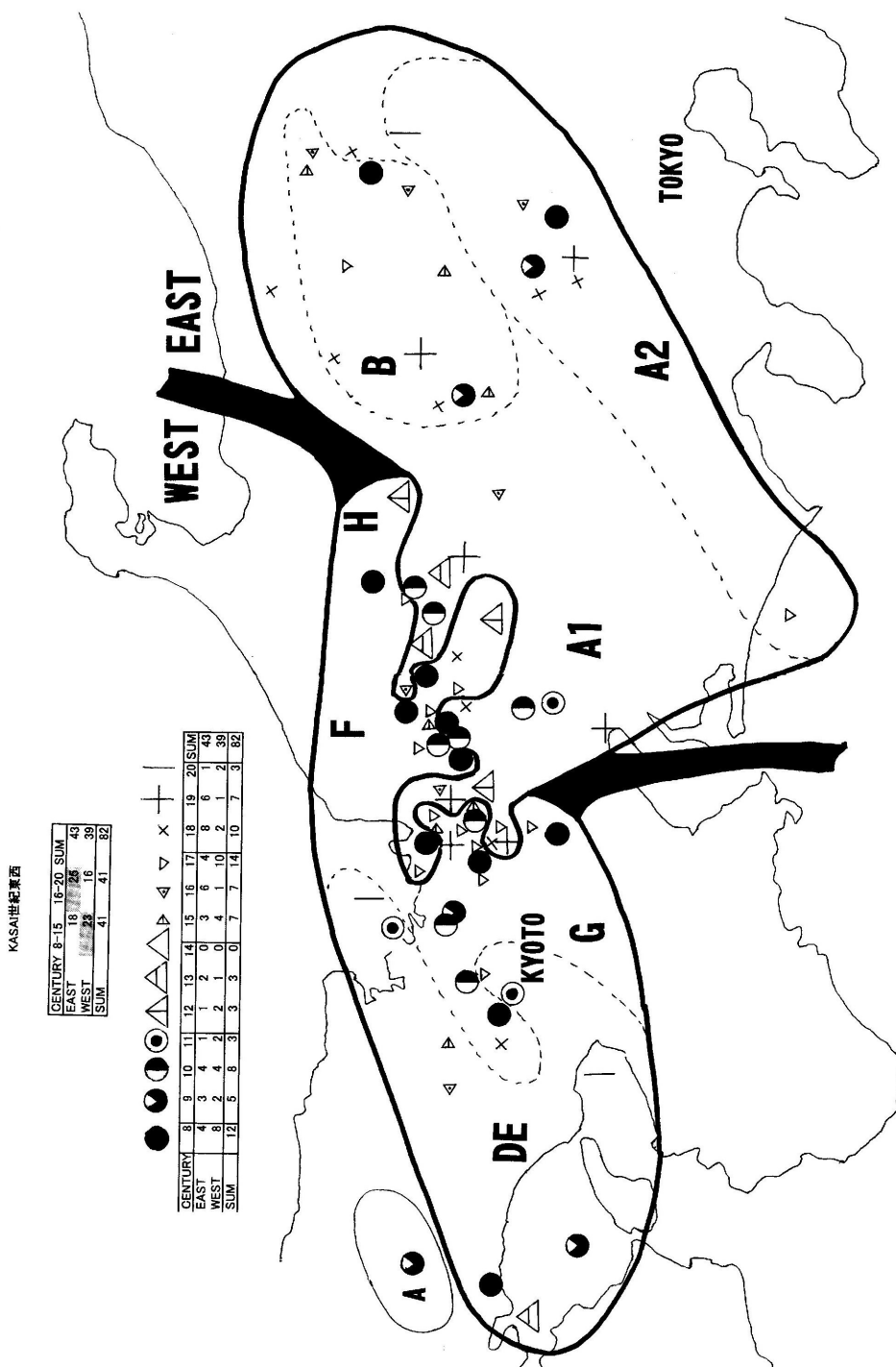
Fig 6. Gravity centers of the 82 standard Japanese forms

The important division into east and west clusters does not correspond to the first year of appearance at all. Rather, the year of appearance crosses at right angles with east-west clusters, as if denying any relation between geography and the history of words. The east-west clusters better correspond to geographical locations, but even then the correspondence is not clear and reveals many exceptions. The overall pattern seems to show that the 82 forms can be analyzed into 5 or 6 main groups. The 4 forms in the lower-right-hand side are exceptional, being old forms which appeared very early in Japanese history and are distributed in eastern Japan. This exceptional group will be discussed later again. The other 4 forms in the lower-left-hand side are also exceptional. These forms also appeared early in history but were expelled from the Kyoto area to the west by newer forms, which however did not become standard and continued to be used as western dialectal forms.

If these 8 forms are ignored as exceptions, the remaining forms show that forms that appeared before the 13th century are concentrated around central Japan near Kyoto, and that more recent forms appearing after the 15th century expanded from central Japan to the east in the direction of Tokyo (Kawaguchi and Inoue 2002). This seems to show that recent standard forms were first adopted in Tokyo and disseminated in many directions near Tokyo. An approximation line was calculated and entered to show the general tendency. This straight line shows the tendency that the forms plotted near Tokyo appeared more recently, while that the forms plotted west appeared earlier in history.

Fig 7 is misleading in the sense that the average differences of standardization in northern Japan are not represented properly. The general map of standardization showing the cumulative average ratio of usage of the 82 standard forms was omitted in this paper (see Kawaguchi and Inoue 2002). There is a great difference in standardization between the Tokyo area and northern Japan. But the Tokyo area and the northern area are calculated similarly on the basis of the west-east dimension of the gravity center method.

In order to represent this difference properly in one dimension, calculation of geographical locations in a diagonal direction on the map would be useful. However, even by this new technique, the general difference of standardization between Tokyo and the countryside to the north-west along the coast of the Japan Sea would not be properly represented. Although the Japan archipelago may be seen as linear from southwest to northeast, it cannot be grasped as a pure line without width.
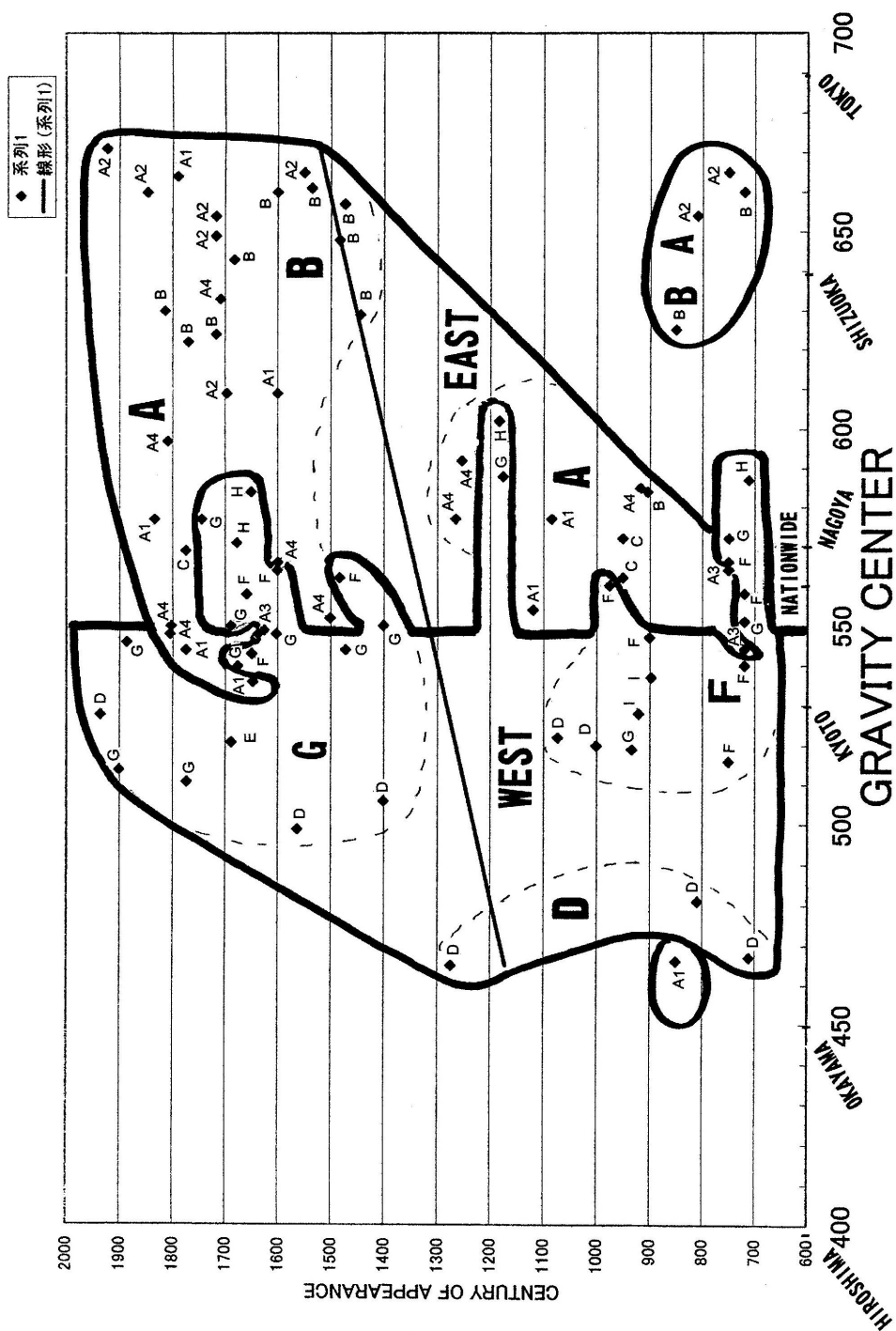
Fig 7. Scattergram by geographical gravity centers and by the year of appearance

5. Railway Distance Gravity Centers

5.1. Methods

In order to evade the problem of width of the Japan archipelago, a new technique of calculating the geographical locations of the 48 prefecture areas was adopted. It is the use of railway distance from cultural centers to the 48 prefecture centers. Distances were calculated both from the old cultural center Kyoto and from the modern cultural center of Tokyo. Railway distances were calculated along the shortest and fastest railway network in operation on the basis of official timetable of Japan National Railway in 1980's. When there was no railway, for example, between islands, the distance by marine transport was adopted. The geographical distances to administrative central cities were calculated, because there was no other reliable technique for calculating the distance from Tokyo or Kyoto to any given prefecture. There have been no great changes to the railway networks over the past fifty years. Although the rapid train networks of the "Shinkansen" have dramatically changed travel time, there have been no changes to the basic networks. According to a rule of transportation geography, new traffic networks repeat old traffic networks. The Japanese railway network can be interpreted as a reappearance of the old highway networks of earlier centuries. Thus, a railway network is a representation of the traditional communication network of residents.

Railway distance gravity centers were calculated in the same way as geographical gravity centers. The formula of calculation is as follows:

Railway distance gravity center = (rate of usage of each area) multiplied by (railway distance) divided by (the sum of the rate of usage of each area)

$$\text{Distance center (d)} = \frac{w^1 \times d^1 + w^2 \times d^2 + w^3 \times d^3 + \ldots\ldots\ldots\ldots w^n \times d^n}{w^1 + w^2 + w^3 + \ldots\ldots\ldots\ldots\ldots w^n}$$

where w^1 to w^n mean "weight" (average ratio of standard forms), and d^1 to d^n refer to the distances of prefectures from the center.

By using this kind of calculation, the degree of standardization of any prefecture in any direction can be taken into consideration, and also the real traffic distance can be taken into consideration. This is important in explaining the prefectures which lie outside of the main railway networks, especially the islands in western Japan. There is no space for explaining the general results, but the correspondence of the cumulative ratios of standardization of the 82

forms and the railway distances of the 48 prefecture areas was remarkably clear (Inoue 2004, in print).

5.2. Results

In **Fig 8**, the horizontal axis shows the railway distance centers from Tokyo. The left-hand end of the graph corresponds to the location of Tokyo. Kyoto and some northern cities correspond to the distance of 530km from Tokyo. If a hypothetical form with a nationwide distribution with 100% usage in all the prefectures (like many basic words with no dialectal differences, that is, words usually neglected in dialect geographical research) is plotted on this graph, it would be situated 662.0km away from Tokyo.

The vertical axis shows the first year of appearance of the form in historical documents. The result of the cluster analysis is added by solid and dotted lines encircling east and west clusters. The approximation line again shows the tendency that the forms plotted near Tokyo appeared recently, and that the forms plotted farther from Tokyo appeared early in history.

Fig 8 looks neater than Fig 7. East-west cluster differences are more clearly shown in this graph. Standard forms belonging to the eastern cluster are naturally plotted near Tokyo. The forms lying nearer (than 400km) are divided into two main groups, one group of four exceptional forms appearing very early in history, and the other forms appearing after the 16th century. The four exceptional forms which appeared early and are distributed in the east are the forms that were later expelled from the Kyoto area by newer forms. These four forms ('whirl of wind', 'whirl of hair', 'woman' and 'to be') became standard Japanese, perhaps because these continued to be used in written documents as stylistically higher (orthodox or literary) forms.

The remaining forms of the eastern cluster group show an inclination to be distributed from the upper-left to the lower-right, indicating that recent forms first cited in the 19th and 20th centuries are distributed mainly near Tokyo. This further indicates that the forms disseminate farther from Tokyo as time passes. Recent forms will follow the same pattern as forms that appeared in the 16th, 17th and 18th centuries. The newly adopted standard forms seem to disseminate nationwide from Tokyo in four or five centuries. In contrast, forms belonging to the western cluster are mostly distributed near Kyoto and near the theoretical center of Japan in the case of the nationwide distribution, that is, between 500km and 700km from Tokyo.

The forms of the western cluster also seem to be divided into two groups according to the year of first appearance. Masses of forms of the western cluster are observed both in the 8th and
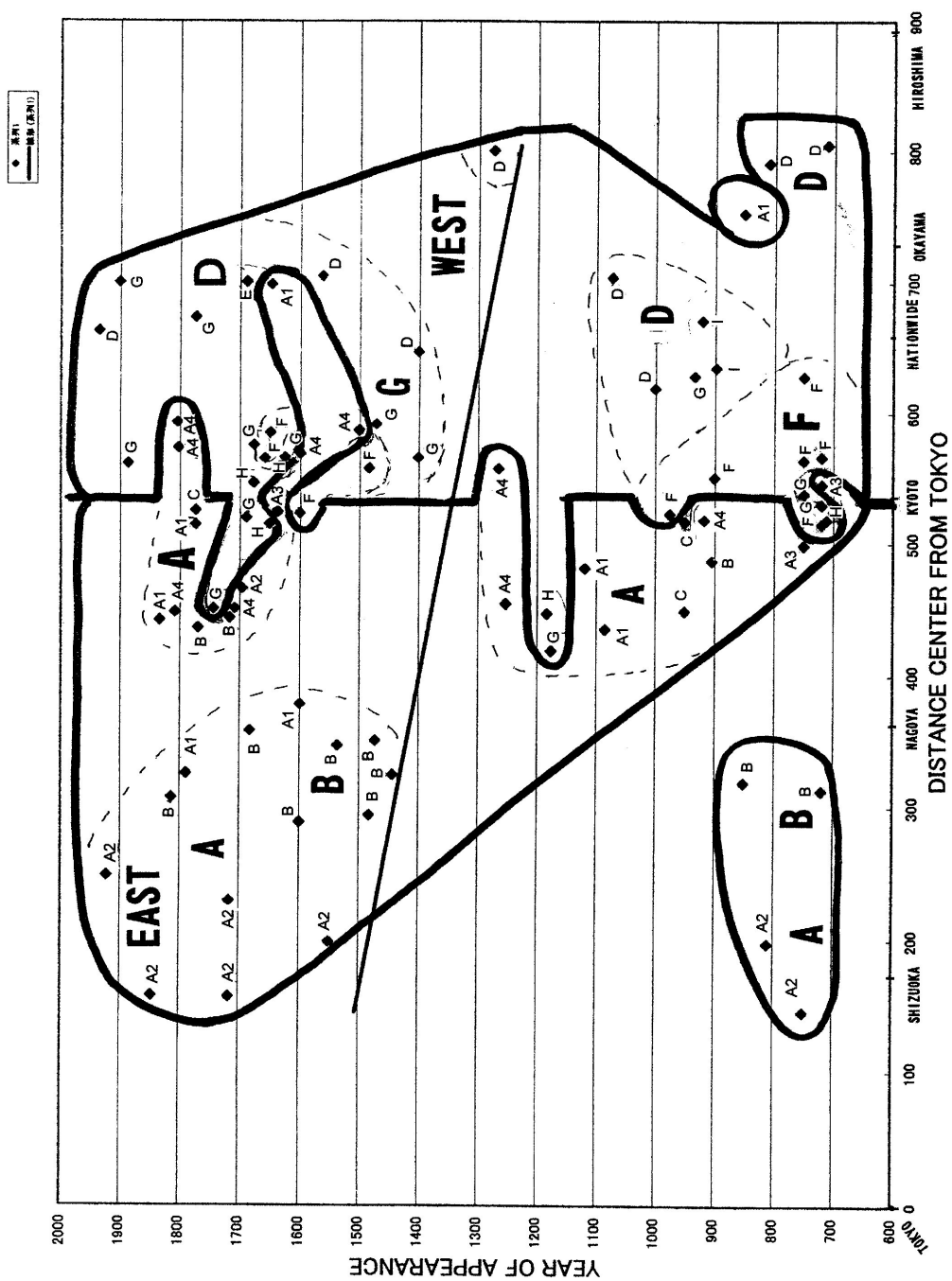
Fig 8. Scattergram by railway distance centers from Tokyo and by the year of appearance

17th century, around 530km or at about the same distance as Kyoto. In returning to the original LAJ maps (or maps arranged in Fig 5), most of them were found to be forms which were not expelled from the Kyoto area by newer forms. This implies that the forms appearing in similar ages behave similarly if taken as a whole. It also suggests that seemingly semantically unrelated forms show continuous stages of geographical dissemination if taken as a whole.

The forms plotted further than 600km are mostly those which have confronting colloquial expressions in eastern Japan. Some of the forms appearing earlier seem to be expelled farther from Tokyo as time passes, as can be explained by the general velocity of dissemination.

Forms in the F cluster have been interpreted as showing "nationwide" distribution as in Fig 4 and Fig 5. Similarly, F cluster (or nationwide cluster) forms appearing before the 10th century achieved "nationwide" distribution. The F cluster forms appearing after the 16th century were also able to achieve nationwide distribution. It seems that four or five centuries seem to be enough for a form to spread to the whole of Japan (Tokugawa 1993, Inoue 2003.7).

In summary, the railway distance gravity center method showed quite successfully the relationship with the first year of appearance of the standard forms. Tokyo was clearly shown as the modern center of dissemination of standard forms, while Kyoto seems to be the older center of dissemination.

Distance centers from Kyoto will be skipped in this paper. I will only include the last summary of the explanation.

"Thus, although Fig h seems to show another clear distribution pattern, this is not due to the power of Kyoto itself, but due to the recent powerful dissemination from Tokyo. Recent waves from Tokyo were again observed in the distance center graph."

6. Theoretical implications of the Distance Center Method

Representation in one dimension by a gravity center graph can be effectively applied to countries with a long geographical configuration, for example Japan, Chili, Italy and Portugal. However, the further simplified representation on the basis of railway distance from a cultural center could be applied to any country or to any area, if the center of linguistic diffusion is clear. The distance could be measured in various ways other than railway distance. Road distance, direct distance, and distance on the basis of time, fare and frequency of travel could also be utilized.

This idea of simplifying geographical distribution into one dimension and one direction

is just like "an egg of Columbus". It is not a great invention but only a newer idea of representation. In natural science it is said that the truth must be simple and applicable to many phenomena. This reasoning can be applied to this distance center technique. When matters are simplified, they can be applied to more related phenomena. As the relationship between geographical locations of distance centers and other factors like the first year of appearances can be represented in continuous values, correlation coefficients can be calculated, and furthermore, the number of years necessary for nationwide dissemination can be calculated. This will lead to the study of diffusion velocity of dialectal forms.

The technique of "Dialektometrie" has recently been applied to many dialect atlases in Europe, like Italy, France, Switzerland and so forth by Professor Goebl (1982, 1993) and others (Heeringa 2004). The difficulty with this technique is the representation of continuous scales of differences among many localities. It is customary to present many maps with continuous scales of differences by discrete colors or shades. However if a center can be selected a priori like Paris in France, using the distance center graph method, the degree of differences can be shown graphically as continuous scales. The distance from the center can be shown on the horizontal axis, and the similarity value (or difference value) can be shown on the vertical axis. All the localities investigated can then be plotted on the graph. Thus, the relationship between distance and dialectal differences can be analyzed and can be compared with different languages or with different dialect atlases.

If one does not like to make use of a center given from outside unrelated to linguistic reasoning, a typical central location could be decided upon based on some kind of calculation, and this could be used to determine the center of linguistic dispersion. In contrast, even a graph with a starting point from a far-away location is theoretically possible. It would show the trend of dissemination in a reverse way.

Theoretically speaking, representation in two dimensions by the "distance gravity center method" can be applied to any kind of geographical distribution. For this study, only the standard forms were shown in the "distance center graph". But any other phenomena can be shown. The merit of a "distance center graph" is that many phenomena or a phenomenon from many localities can be shown individually at one time. Thus the distribution of the many word-forms of one word can be summarily shown in one graph. It can show the historical relationships of word-forms more concisely and more reliably. Theoretically, it would be possible to include all the word-forms of all the investigated words of a linguistic atlas. This

graph would show that seemingly disparate word-forms without any semantic similarity are plotted near each other, suggesting some historical reasons for the similar geographical distribution.

In some quantificational techniques making use of computers, individual differences of investigated words are sometimes ignored, and only overall differences are considered. However, this type of observation is sometimes misleading. In my own study of another linguistic phenomenon, the study of new dialect forms in present-day Japanese dialects, trends of new dialect forms born in the countryside and adopted by young people in Tokyo have been amply observed. This tells us that all forms do not propagate from the cultural center to the periphery. However, minute investigation of the phenomena tells us that there is another process of stylistic upward dissemination from dialectal, colloquial style to standard, literary style. Many forms can be adopted in colloquial style by young people in Tokyo, and as time passes, these forms become to be used by people in older age groups (or young people themselves become older), and then start to be used in written sentences and in books. The same procedure of stylistic upward shift must have been necessary also in the past. This process seems to explain why new forms adopted later in Kyoto did not become standard, as was pointed out earlier. Thus, linguistic changes in the past and linguistic change in progress at present are closely connected, and studies of both phenomena can help to understand the real picture of the linguistic change (Labov 1966).

This technique can again be applied to any kind of geographical distribution other than dialectal phenomena, like diffusion of fashion, average income, ratio of older people, and so on. This idea of distance center can also be applied to phenomena larger than domestic distribution. For example, the evaluation of Japanese language is higher in nearby countries in Asia and also other Pacific Rim countries like Australia and the United States. This can be explained by the idea of "geographical neighboring (proximity) effect". Data can be graphically represented by the distance center graph, and if shipping charges for freight, for example, are used to measure the distance between Japan and the countries in question, a clear correlation may appear graphically.

The idea of calculating and representing railway distance centers for geographical distribution seems to be quite new in Japanese dialectology. However, some other scholars may have applied the same technique many years ago in some unrelated areas of study. I would be grateful if I could obtain information on any similar attempts in the western world.

7. Discussion: Continuity of distribution

All in all, this study emphasized continuity rather than division of dialectal phenomena. The direction of diffusion is similar for most of the standard Japanese forms.

Language is used as a means of communication, so that no sudden overall change is possible. To secure effectiveness of communication, small changes occur incessantly. The hint of this view of continuum is the sociolinguistic view of dialects, that dialects reflect linguistic changes in progress. By arranging dialect distribution maps to show the order of changes, an overall perception of the various maps becomes possible.

General tendencies that can be applied to various phenomena have been treated in this paper. My belief is that with the application of multivariate analysis, even very complicated data can be represented concisely using simple arithmetic calculations, and thus complicated reality can be resolved by simple rules or laws.

Linguistic geography was once admonished for its atomism, paying too much attention to the distribution of individual words. Though this analysis is based on data acquired in the far eastern part of the Eurasian Continent, it is presumed that the same processes may also be ascertained in the other parts of the world. If similar tendencies are found in other languages, some universal rule-governed nature of dialects may be attainable. Dialectology has tended to pay too much attention to domestic matters, and an international, interlingual point of view has been lacking. By paying more attention to dialects of foreign languages and by grasping more theoretical tendencies, dialectology can develop in a new direction in the new 21st century[ii].

This is a revised and enlarged version of the presentation at the Symposium of "Measuring lexical variation and change" held in Leuven in October 2002[iii].

Bibliography

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

Goebl, Hans (1993) "Dialectometry: A short overview of the principles and practice of quantitative classification of linguistic atlas data" in Koehler et al (eds.) *Contributions to Quantitative Linguistics*, (Kluwer)

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

Inoue, Fumio and Hisako Kasai (1989), Dialect classification by standard Japanese forms, in Mizutani (S.) ed. *Japanese Quantitative Linguistics* (Altmann (G.), ed. *Quantitative Linguistics* Vol. 39 (Bochum,) p.220-235.

Inoue, Fumio (1993), The significance of new dialects, *Dialectologia et Geolinguistica* (SIDG) 1 p.3-27.

Inoue, Fumio (1996.3) (1996.10), Computational dialectology (1) (2), *Area and Culture Studies* 52 p.67-102, 53 p. 115-134.

Inoue, Fumio (2002.10) "Multivariate Analysis, Geographical Gravity Center and the History of the Standard Japanese" Leuven Symposium

Inoue, Fumio (2003.7) "*Nihongo wa nensoku 1 kiro de ugoku (Japanese language moves 1 km per year)*" (Kodansha Gendai Shinsho)

Inoue, Fumio (2004) "Hyojungo shiyoritsu to tetsudokyori ni miru komyunikeshon no chiriteki yoin (geographical factors of communication observed between standard usage and railway distance)" *Shakai gengo kagaku* 7-1

Inoue, Fumio (in print) "Geographical Distance Center and Multivariate Analysis of the Standard Japanese" *Proceedings for ICL 17 Prague* (CD-ROM)

Inoue, Fumio (forthcoming) "Geographical Distance Center and Rate of Diffusion of Standard Japanese" *Proceedings of ICDG 4* Riga

Kawaguchi, Yuji and Fumio Inoue (2002) "Dialectology in Japan in historical perspectives" (PART II Historical characteristics and geographical distribution of standard Japanese forms), *Revue Belge de Philologie et d'Histoire* 80.

Labov, William (1966). *The Social Stratification of English in New York City* (CAL)

Nihon Daijiten Kankokai (2000-2002), *Nihon Kokugo Daijiten (Great Dictionary of the Japanese Language)* Second edition (Shogakukan)

NLRI (National Language Research Institute) (1966-1974), *Linguistic Atlas of Japan Vols. 1-6* (Okurasho Insatsukyoku)

Tokugawa, Munemasa (1993) *Hogen-chirigaku no Tenkai (Expansion of Dialect Geography)* (Hitsuji Shobo)

---

[i] LAJ consists of 2400 localities. More than hundreds of word-forms were reported for the item KATAASITOBI 'to hop on one leg'. One of the forms KENKEN is exceptional in being not standard but spreading from the western half of Japan to eastern Japan. This trend is still continuing among younger generation in 21st century. KENKEN is one of typical examples of new dialectal forms (Inoue 1993, 2003.7).

[ii] I am very grateful to Mr. Stefan Grondelaers for inviting me to the international conference in Leuven and to all the people who have made my visit here possible. I was first invited to present the same paper as the one I presented at the "METHODS in dialectology" conference in Joensuu, Finland, in August 2002. However Mr. Stefan Grondelaers very sensitively added that he found one of my old papers written many years ago to be interesting. This comment reminded me that my former technique of simplification of geographical distribution had not been fully applied to the data which I presented in Joensuu. Thanks to Mr. Grondelears' suggestion I was able to develop a new technique while preparing for this conference.

[iii] The following is the message given at the end of the symposium.

It is an honor to have this valuable opportunity to report to you about the recent progress of computational dialectology in Asia. When I received the letter of invitation to this conference, I thought that it was almost my duty to participate, because this country, and especially this university, is connected to my scholastic life in two ways. When I decided to major in linguistic geography as a young student, I was introduced to the Reverend Willem Grootaers, the son of Professor Grootaers who worked at this university. Reverend Willem Grootaers later taught me the basic ideas of linguistic geography in his special lectures at a university in Japan. The other scholar whose work greatly influenced me was Dr. Sever Pop who also worked at this university. Through his voluminous book of "La dialectologie", I learned how studies of linguistic geography had developed in various parts of the world. At the same time, I learned that some studies of linguistic geography had been accused of atomism, that is, of having a tendency to treat dialect maps one by one, or individually. As a result, throughout my life I have been trying to grasp whole trends of geographical distribution at a glance, by processing many words at the same time, and making use of computational techniques.

多変量解析と地理的重心法と標準語形の歴史

井　上　史　雄

　この論文では，従来分析を進めてきた日本語標準語形の地理的分布データについて，新たな地理的情報を加え，二つの単純化技法を適用した．一つは地理的位置を鉄道距離によって表現したことである．もう一つは2次元的な地理的情報を点で表わして，日本列島を東西の1次元で示したことである．これにより，紙面上の余った次元に別現象を表示できる．ここに標準語形の文献初出年を示した．

　本論文ではいわゆる「河西データ」を用いた．『日本言語地図』(LAJ) の一部項目の数値データであり，これまで様々な統計手法を用いて分析してきた．今回は鉄道距離という非言語的，つまりは広義の社会言語学的情報を加え，県ごとの標準語使用率との関連を分析した．東京の影響力の大きさをみることにより，標準語形使用に古都京都が基本的な力を及ぼしていたことと，その後東京からの伝播が大きく作用したことが，読み取れた．初出年と地理的分布は直接きれいな対応関係をみせない．一方近世初出の語が全国的に普及・伝播している例もあることが分かった．平均しての年速は1キロ程度と見られる．

　この考察をふまえて，方言地理学，計量方言学の理論的問題にも言及する．