COMMON DEVELOPMENTS OF JAPANESE AND DUTCH-FLEMISH DIALECTOLOGY: COMPUTATIONAL DIALECTOLOGY IN JAPAN

Fumio INOUE, Professor Emeritus
Tokyo University of Foreign Studies *
innowayf@nifty.com

Abstract

In this paper, quantitative method of dialectology developed in Dutch-Flemish and Japanese dialectology will be discussed. Three main topics will be treated here. (1) Glottogram was conceived of in Japan on the basis of Itoigawa survey which was a collaborative work of Dr Sibata and Dr Grootaers. Because of this technique it became possible to ascertain that “new dialect” is born and spreading even now. Age difference or apparent time is a key concept of linguistic change. (2) Perceptual dialectology has the roots in Dutch dialectology and was imported in Japan through Grootaers and Sibata. It developed into psycholinguistic analysis of dialect image. (3) Dialectal differences can be studied in order to classify dialects into divisions or areas. They can also be studied as a continuum, reflecting dynamics of linguistic change. Railway distance is a convenient basis to explain dialect differences. Commonalities and differences in Dutch-Flemish and Japanese dialectology will be discussed at the end.

Keywords
linguistic geography, computational dialectology, Glottogram, Takesi Sibata, Willem Grootaers

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

En este artículo, se discutirá el método cuantitativo de dialectología desarrollado en las dialectologías flamenco-neerlandesa y japonesa. Se tratarán tres temas principales. (1) El glotograma fue concebido en Japón sobre la base de la encuesta de Itoigawa, que fue un trabajo colaborativo entre los doctores Sibata y el Grootaers. Gracias a esta técnica, fue posible determinar que el "nuevo dialecto" nace y se extiende incluso ahora. La diferencia de edad o el tiempo aparente es un concepto clave del cambio lingüístico. (2) La dialectología perceptual tiene sus raíces en la dialectología neerlandesa y fue importada al Japón a través de Grootaers y Sibata. Se convirtió en análisis psicolingüístico de la imagen del dialecto. (3) Las diferencias dialectales se pueden estudiar para clasificar los dialectos en divisiones o áreas. También se pueden estudiar como un continuo, que refleja la dinámica del cambio lingüístico. La distancia ferroviaria es una base conveniente para explicar las diferencias dialectales. Los puntos en común y las diferencias entre las dialectologías neerlandesa-flamenca y japonesa se discutirán al final.

 PALABRAS CLAVE

geografía lingüística, dialectología computacional, glotograma, Takesi Sibata, Willem Grootaers

1. Itoigawa Survey, Glottogram and New Dialect

1.1 Linguistic Atlas of Itoigawa

Linguistic Atlas of Itoigawa was a collaborative work by Takesi Sibata, Willem Grootaers, Munemasa Tokugawa and Yoshio Mase. Fieldworks were conducted in 1957, 1959 and 1961 at all the localities in the survey area. Techniques developed in Dutch dialectology were transmitted through Grootaers (Figure 2) to Sibata (Figure 1). Supplementary surveys were added as further development: surveys of all residents at several localities and of junior high school students.

1 This workshop was held in 2018 in commemoration of 100th anniversary of birth of Dr Takesi Sibata [takeji jibata] and on the occasion of 100 years of restored Lithuania.

2 2019 is 20 years after the return of Reverend Willem Grootaers [wilm éro:ta:rs].
Three volumes of *Linguistic Atlas of Itoigawa* (Sibata 1988, 1990, 1995, Grootaers 1988, 1990) are shown in Figure 3. The size and volumes can be guessed with a tape measure put on the right extending more than one meter. Examples of the maps are
shown on the left. Dr. Sibata dedicated years to manually draw distribution maps using rubber stamps. But his daughter Dr Mie Inokuchi thought it was time to adopt computer technology to draw maps. All the maps were redrawn by the computer. ³

1.2 Glottogram

As the first topic of connection between Dutch and Japanese dialectology Glottogram will be introduced. This technique originated in the survey for Linguistic Atlas of Itoigawa.

As a development of dialect distribution map, Grootaers and Sibata planned a new survey of all residents in some localities in Itoigawa area (Sibata 1998). Figure 4 shows a change from [ŋ] to [ɡ]. Propagation of [ɡ] from the mountainous area to the seaside area is clearly shown. Age differences were conspicuous as shown on the lower half of this map. Three localities were later added along the Hayakawa valley for the pronunciation change of nasal [ŋ]. These supplementary surveys also showed clear age differences. The information of apparent time was not consistent with deduction from geographical distributions (for lexical items but seemingly contradictory for pronunciation). Supplementary surveys of three localities along the Hayakawa valley on the righthand side was impressive and persuasive, showing diffusion from mountainous upstream area to downstream area. This is the beginning of the technique of surveying age differences in many localities which developed later into the glottogram (Inoue 2017a, 2017b).

³ Raw data of Linguistic Atlas of Itoigawa was also included and made public.
• Itoigawa survey
  – 1957
  – 1959
  – 1961

• All-resident survey
• 3 + 3 localities

Figure 4. Incipient glottogram in Itoigawa.

Figure 5 is a typical “glottogram”. This is a survey of young and old informants at all localities along the Hayakawa valley (Tokugawa 1993). The typical new dialect presented in this glottogram is “Sty (in the eye)”. Informants are plotted along the vertical axis, younger people over the elderly like geologic stratum. A dialectal form MEBOITO is used upstream and by older informants, and we can deduce it must be older. A dialectal form MEBOITA is used downstream and by younger informants, and we can deduce it must be a new dialect form. The technique of glottogram is powerful and effective to observe the linguistic change in progress. In usual studies of dialect geography, the comparison of maps of different age groups was necessary in order to see age differences of an area. By utilizing glottogram age differences can be easily observed at a glance. However, the glottogram technique has only been applied in Taiwan by Prof. Li and in China by Dr. Huang He and some other scholars, but not applied in Europe yet.
1.3 New dialect

The concept of “new dialect” was born in Japan thanks to the stimulus by Grootaers and Sibata. “New dialect” is defined with three conditions (Inoue 1993, 2016): It is (1) a non-standard linguistic form, (2) used more among younger people, (3) in informal situations. In other words, it is a typical linguistic change in progress, and a change from below. It is also anti-standardization. New dialect can be ascertained by observing differences by age groups or apparent time. New dialect forms are reported in all the prefectures of Japan, and a dictionary of new dialects is published and is being revised in my computer anytime when reports of new dialect are found. Tokyo itself is not the only vanguard of new dialect. New dialect forms that spread from the countryside to Tokyo were detected (Inoue 2010c, 2012).
New dialect forms were analyzed using a multivariate analysis “Hayashi 3” which is similar to “correspondence analysis”. Since then, areas of glottogram survey extended to almost all the parts of Japan. GLAPS program developed by Professor Ogino (1980) was useful in drawing glottograms. Some more computational programs were developed by Fukushima (2000, 2014) and other scholars. Once survey data are computerized to draw maps or to make graphs, multivariate analyses can easily be applied to grasp the overall patterns of linguistic changes in progress. Age and area differences are analyzed, and multivariate analyses are applied. An example will be shown next.

Correspondence analysis was applied to new dialect adopted in Tokyo, and a resultant graph is shown in Figure 6. The first component on the horizontal axis is Tokyo factor shown by green. The second component on the vertical axis is Osaka-Kyoto factor shown by yellow. These two cultural centers seem to hold their own
prestige and power of propagation even now, dividing Japan in east and west. Computational techniques provided a new general view of Japanese dialects here again.

2. Perceptual Dialectology and Multivariate Analysis

2.1 Similarity and difference of dialect perception

Figure 7. Dutch subjective dialect boundary showing similarity, Weijnen (1946).

As the second topic of the relation of Dutch-Flemish and Japanese dialectology, perceptual dialectology will be discussed. The discussion on whether similarity or difference should be surveyed and plotted will be introduced, and the development in the direction of dialect image will follow.

The research idea of perceptual dialectology was conceived and realized by a Dutch dialectologist Weijnen and imported from Dutch dialectology by Grootaers and
Sibata, who conducted research in Itoigawa as a part of Linguistic Atlas of Itoigawa survey. Perceptual dialectology was later advocated by Dr Preston, and this classic research question is again attracting attentions in many countries of the world, as many presentations of SIDG congress in Lithuania symbolize. Preston’s introductory book and handbooks (Preston 1989, 1999) are appropriate and impartial in paying attention to both Dutch-Flemish and Japanese early studies.

There was a small difference in the technique at the beginning of research. In the early Dutch approach (Weijnen 1946) informants’ consciousness of dialect similarity is shown by little arrows, and borders are shown by bold lines (Preston ed. 1999) as shown in Figure 7.

![Figure 7](image)

Figure 8. Japanese attempt: difference map, Itoigawa.

On the Japanese side, at Nechi valley in the Itoigawa research in Figure 8, respondents indicated the language of which nearby village was 1) not different, 2) a little different, 3) quite different, or 4) mostly incomprehensible. The Dutch maps, therefore, appear to be ones of similarity and the Japanese maps ones of difference. There were many discussions so far as to what the influencing factors of dialect
perception are, linguistic or non-linguistic in the first place. Both factors were found to work as far as the researches so far have discovered.

2.2 Multivariate analysis of dialect image

By utilizing multivariate analysis, the perceptual dialectology took off from earth surface and developed in the field of psycholinguistics. A technique of SD method or Semantic Differentials was adopted for characterizing speech regions by dialect image. By applying multi-dimensional analysis of Hayashi 3, intellectual and emotional components were separated (Inoue 1996) (shown later in Figure 10). This suggests that dialect perception is mainly controlled by non-linguistic factors. Two factors of intellectual and emotional correspond to “correct” and “pleasant” in Preston’s analysis (Preston 1989).

Figure 9. Multivariate analysis: dialect image.
This technique was also applied in Great Britain (Inoue 1996). As shown in Figure 9, two components were “intellectual” (or standardness) and “emotional”. A description of a British linguist on “English Accents” (Honey 1989) is shown on the righthand side. He classified accents first into acrolect, mesolect and basilect, and then added hyperlect, paralect and so on. His analysis beautifully corresponds with responses of British university students. American and Australian English were included as an attempt at incorporating World Englishes. “Emotional” components approximately correspond with Honey’s description of media exposure of dialects and world Englishes.

2.3 Dialect image and dialect souvenirs

The dichotomy of factors in intellectual and emotional seems to be universal. However, this dichotomy is not so widespread in European analysis of perceptual dialectology as far as I noticed in SIDG congress. It can be shown on a map of Japan (Figure 10). Dark areas are dialects which are felt emotionally plus or positive. They are distributed in three separate areas, northern, central and southern Japan.

---

4 This data was acquired at the same occasion with the survey of perceptual dialectology which Dr Chris Montgomery kindly referred to in the plenary session of SIDG Congress.
This seemingly mysterious distribution pattern reminds us of the distribution pattern of various dialect souvenirs sold in Japan (Figure 11). As the graph on the upper righthand side of Figure 10 shows, the large numbers of dialect souvenirs are distributed in a similar pattern in northern, central and southern Japan. As shown by three arrows which connect the graph of dialect souvenirs with the map of Japan, they correspond with the areas where dialects are emotionally positively evaluated (Inoue 1995).

In the areas of emotionally positive dialect image, dialect souvenirs are sold. This is the mechanism connected to positive emotional feeling of these three areas. Dialect souvenirs prosper in these three areas as a symbol of love for local life. Thus, the study of dialect image has successfully explained people’s behavior and economical utilization of dialects.
Figure 11. Dialect souvenirs (Kyushu, Kinki and Tohoku).

Japan as a double caldera volcano

Figure 12. Japan as a double caldera volcano.
This mysterious distribution pattern can be logically explained if we take the figurative image of a double caldera volcano in Figure 12. Northern and southern Japan have been the peripheries of Japanese culture for a long time. Because they were far from cultural centers, they were full of dialect characteristics. Kyoto-Osaka area was the cultural center of Japan in olden days; however, after the modern era a new city Edo, which later became the capital Tokyo, was constructed. Tokyo developed and dialect differences of Kyoto Osaka area became famous or salient among residents of Tokyo as a cultural rival. Consciousness of dialect differences against Tokyo became conspicuous in these three areas, leading to love or positive image for their own dialects.

3. Dialect Classification and Dialectometry

3.1 Dialect classification of Japan

The third topic is the dialect classification which was the basic starting point and essence of dialectology. The contrasting ideas of dialect division (classification) and continuity will be discussed, and utilization of railway distance will also be examined.
Figure 13. Dialect divisions of Japan.

Figure 14. Linguistic Atlas of Itoigawa: isoglosses.
Figure 13 is a compilation of maps of dialect divisions by various dialectologists. Some dialectologists stressed differences and others stressed continuities between dialects. As the red arrow of this map indicates, the survey area of Linguistic Atlas of Itoigawa is just on the border of east and west division of Japanese dialects.

Figure 15. Dialect classification by cluster analysis of vocabulary.

In spite of the fact of east-west differences, Sibata and Grootaers did not pay much attention to dialect division or classification within their survey area. As shown in Figure 14, they drew several maps showing bundles of isoglosses in the surveyed area, and sometimes counted number of words showing similar distribution patterns. However, they seemed to be more interested in dynamic changes of dialect than static division of dialect. They considered linguistic geography as a means to reconstruct the
history of dialect. To add to this, it was troublesome and time-consuming for them to manually count phenomena. Computerization of the whole data by Dr Mie Inokuchi was just in time. An original dialectometric technique was applied to Itoigawa data in Kumagai (1993).

For computational dialectology in Japan further to develop, a new kind of nationwide data was necessary. Dialect geography has long been accused of being atomistic and individualistic, paying attention to individual words without looking at the general overall patterns. Aggregate data can be utilized to study dialect differences or dialect classification and dialect division. On the basis of raw data showing usage rate of many words, dialect differences can be analyzed from an objective point of view. Numerical distribution data of Standard Japanese in LAJ (Linguistic Atlas of Japan) was made public. So-called Kasai data was made of LAJ 82 standard Japanese words. Cluster analysis was successfully applied to this data. The resulting dendrogram was redrawn in the form of a map in Figure 15 (Inoue et al. 1989). This classification coincides with previous attempts of dialect division in Japan as shown in Figure 13. Hokuriku region was separated from Tokyo area in the dichotomy of east and west Japan. Explanation of continuity is also possible.
Returning to the already existing raw data, a map was drawn by Dr Yarimizu which faithfully shows continuity of degrees of standardization in about 20 steps of shades (Inoue 2010a). This map in Figure 16 reveals that Hokuriku region in central Japan is low in usage of Standard Japanese words if compared with direct distance from Tokyo. This can be better explained if railway distance is taken into consideration. Hokuriku region is far and inconvenient in railway because of detours. There are natural barriers of steep Japan Alps mountains between Tokyo and Hokuriku.

3.2 Railway distance

In order to explain the basic mechanism of dialect differences in LAJ, railway distance was taken into consideration. Railway distance is reliable and convenient because it is often a repetition of old road system and it does not change so often.
Figure 17 is a scattergram of railway distance and ratio of LAJ standard forms (Inoue 2004, 2006, 2008) overlapped on a map of Japan. Horizontal axis shows railway distances from old Capital Kyoto to all the prefectural capitals. Vertical axis shows percentage of non-standard, dialectal forms (upside-down presentation of standard forms) in LAJ. This can be metaphorically conceived of as a limestone cave\(^5\) where drops of water come from the ceiling (or sky) above Kyoto and later above Edo-Tokyo. Two stalactites (hanging limestone) can be detected. One forms gentle slopes around Kyoto, and the other forms steep slopes around Tokyo. Hokuriku region is solely controlled by old influences from Kyoto. The distribution of prefectures is similar to actual geographical locations of the prefectures. By overlapping the map of Japan, beautiful coincidence with actual geographical locations appeared. Usage rates of standard Japanese forms are found proportional to railway distances. Railway distances thus represent old and new diffusions from the two capitals of Japan (Inoue 2010b).

\(^5\) Limestone cave model is a revised version of the (double) umbrella model (Inoue 2010b: 60, 2010c: 99, 2017b: 189) which is still applicable for ordinary representation of language standardization and new dialect.
The same technique was applied for grammatical phenomena of GAJ or *Grammar Atlas of Japanese Dialects* (NINJAL 1989 - 2006). This time not the average usage rate of prefectures but usage rate of each locality (informant) was calculated, and also railway distance plus walking distance to informants’ household were measured. Overlapping with the map of Japan was again tried in Figure 18. Horizontal axis shows the railway distances from the Capital Tokyo, and vertical axis shows percentage of non-standard, dialectal forms (upside-down presentation of standard forms) in GAJ. Labels show prefecture names. This time again two metaphorical stalactites (hanging limestone) were ascertained. One peak with gentle stalactite around Kyoto was detected, the other steep stalactite around Tokyo also appeared. In this case of grammatical phenomena too, beautiful coincidence appeared by overlapping with the map of Japan, though the actual locations of westernmost Kyushu island show a small disagreement because north-south difference cannot be shown faithfully by this railway distance technique.
Thus, basic geographical distributions of lexical phenomena in LAJ and grammatical phenomena in GAJ are common. These results show the effectiveness and explanatory power of the railway distance. In other words, dialect distribution is a reflection of human activities.

3.3 Levenshtein distance

Utilization of railway distance appeared independently in Dutch dialectological research, perhaps by coincidence. Gooskens (2005) discusses dialect differences in southern Norway utilizing Levenshtein distance, which will be introduced next. Railway drastically changed transportation system of Norway as the map on the left in Figure 19 shows. Travelling time by railway was found useful to explain the dialect differences.

Figure 19. Travelling time and Levenshtein distance in Norway.

Accumulation of isoglosses or bundle of isoglosses were the basic ideas in the past to observe dialectal differences. Many dialectologists now accept that borders of dialects are difficult to identify, and they are beginning to show the continuity of neighboring dialects more faithfully. However, in this field too computational
technology and multivariate analysis have offered a great progress. Methodological influence was great when Levenshtein distance was applied to Dutch dialects by Dr Heeringa (Heeringa 2004, Heeringa & Nerbonne 2001, Nerbonne et al. 2005). The technique of calculating Levenshtein distance is shown on the left-hand side in Figure 20. And the result is shown as a colorful map of the Netherlands showing (gradual) gradation and continuity⁶.

![Figure 20. Levenshtein distance by Heeringa.](image)

Levenshtein distance was introduced in Japan and was successfully applied to GAJ data by Dr Yarimizu (2009). He applied cluster analysis for his data of Levenshtein distance. This colored map in Figure 21 brought deeper reading or interpretation of geographical distributions. In western Japan waves of diffusion from the old capital Kyoto were clearly observed. In eastern Japan waves of diffusion from the new capital Tokyo were shown. The border between east and west does not lie along steep Japan Alps.

Application of Levenshtein distance to Japanese dialects was recently executed by a Dutch scholar Huisman (2018). The resulting map showed similar pattern of Kyoto-centered propagation though the data is different (East-west border is not clear). By the way all the maps and raw data of LAJ and GAJ are available on the internet (Kumagai 2013, 2016) by the efforts of Prof. Sawaki and Prof. Onishi (2018).

⁶ By the way professor Chambers gave him a nickname of “mapman”.

©Universitat de Barcelona
4. Interlingual dialectology and Lendword (export word)

4.1 Interlingual Dialectology

As the fourth common topic, interlingual dialectology can be discussed. This is an interest on neighboring languages. It forms the field of “interlingual dialectology” advocated by Weijnen (1978). There are signs of appearance in Japan as Sibata had discussed dialectal forms imported from Dutch language (Inoue 1990). Similar studies are executed in many language areas as presentations on borrowings of this SIDG congress show. This research field symbolizes that we are now free from narrow
nationalistic vision of dialectology in the 20th century. I hope this kind of study with broad and international view will flourish in the future. Relation of dialect and neighboring languages is an interesting topic, and this field will develop more in this age of internationalization (Weijnen 1978, Gooskens & Heeringa 2012, Inoue 2015). In this periodical, Mr Sawaki will treat the theme from a different point of view.

4.2 Lendword (export word)

The study of loanword (import word) in borrowings is possible within a single accepting language and leads to general theory, but the lendword (export word) requires a multilingual contrastive study, and it was almost impossible to obtain uniform data in the past. In the 21st century, it became possible with the Internet search “Google trends”, as shown by Inoue (2007.2, 2013.3, 2013.12). In addition to Japanese, examples of English, Spanish, Chinese, and Korean have been dealt with, but there seems to be very few followers. This is a promising area in the future. In the latter half of the 20th century, there is a possibility that general rules of economic linguistics will be established outside of the Japanese language (Inoue 2015).

5. Commonalities and Differences of Dutch-Flemish and Japanese Dialectology

5.1 Value of multivariate analysis

Commonalities and differences between Dutch-Flemish and Japanese dialectology will be discussed at the end. Common characteristics is utilization of computer technology. Dialectometry or computational dialectology developed in both areas. Application of multivariate analysis prospered in both countries. In all the three topics above multivariate analyses provided general overall tendencies. In most of the international congresses of dialectology new progresses have been reported every time. Levenshtein distance was first applied by Dutch scholars and soon it was adopted in many countries including Japan. Railway distance was taken into consideration in
explaining dialect differences in Japan, The Netherlands, Switzerland and other countries.

The value of multivariate analyses should be emphasized. The general pattern on the basis of aggregate data can be presented by multivariate analysis. Multivariate analysis is effective for further and deeper understanding, for heuristic study of many variables, like gender, profession, education, or geographically urban/rural, north/south, sea/inland, etc.

By attaining the vantage point of the computer, dialectologists can acquire a wider general view which can lead to broad and rule-governed universal phenomena.

5.2 Differences between Dutch-Flemish and Japanese dialectology

There are also differences between Dutch-Flemish and Japanese dialectology. Glottogram survey was recently applied in Taiwan and mainland China. However, no study of glottogram was reported in the western studies. The Glottogram technique is especially useful to observe linguistic change in progress in time and space, because it can show time dimension and space dimension in one graph.

Birth and spread of new dialect forms are often reported in Japanese dialectology. Age is a cue for apparent time change. Interest on age differences is conspicuous in Japan in the form of glottogram.

5.3 Power of English

It can be said in conclusion that seeds brought from the West by Grootaers (Inoue 2011, 2014) was raised as a tree by Sibata (Sibata 1998, Fukushima 2012), and then new seeds in Japan have grown as more trees, as the Japanese participants of SIDG congress symbolize.7

---

7 Workshop: Bridging East and West: Grootaers, Sibata and Beyond. 9th Congress of the International Society for Dialectology and Geolinguistics Vilnius, Lithuania on July 27(Fri), 2018.
At the end of this paper I must express what Professor Sibata often complained. He said that his works were very rarely cited or referred to in western studies. It must have been because his works were mostly written in Japanese though parts of his works are now available in a book (Sibata 1998) translated in English as shown in Figure 22. He urged younger scholars to present and write papers in English. There is a saying “Publish or perish”. It should be revised as follows: “Publish in English or perish in trash”.

References

* English papers by Fumio Inoue before 2002 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


GOOSKENS, Charlotte (2005) “Travel time as a predictor of linguistic distance”, Dialectologia et


INOUE, Fumio (2010a) “Sprachraum and infrastructure: abstracting geographical space via railway distance”, Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), Language and

Inoue, Fumio (2010b) “Real and apparent time clues to the speed of dialect diffusion”, Dialectologia, 5, 45-64.  

<http://www.leeds.ac.uk/linguistics/WPL/WPL15.html>.

< http://www.publicacions.ub.edu/revistes/dialectologia7/dialectologists.asp>

INOUUE, Fumio (2012) “Improvements in the sociolinguistic status of dialects as observed through linguistic landscapes”, Dialectologia, 8, 85-132.  
<http://www.publicacions.ub.edu/revistes/dialectologia8/>

<http://innowayf.net/download/Works_in_Dialectology_by_Reverend_Grootaers.pdf#search='willem+grootaers'>


Bochum, 220-235.


NERBONNE, John, Ilse VAN GEMERT, and Wilbert HEERINGA (2005) *A dialectometric view of linguistic “gravity”*, Alfa-informatica, University of Groningen


