UNITED NATIONS

Group of Experts on Geographical Names

Fifteenth Session
Geneva, 11-19 November 1991

WORKING PAPER

No.36

Agenda item 5

REPORTS OF THE DIVISIONS

Report on Standardization of Geographical Names in Japan

(Geographical Survey Institute, Japan)

Report on Standardization of Geographical Names in Japan (Item No.5 of the Provisional Agenda for the 15th Session of UN Group of Experts on Geographical Names)

Geographical Survey Institute of Japan

Standardization of geographical names and related activities in Japan since the 14th Session of UN Group of Experts on Geographical Names are as follows:

1. Standardization of Geographical Names

(1) Domestic Names

There is no single central agency responsible for the collection, registration and standardization of geographical names in Japan. Several agencies and organizations have been sharing the responsibility in their respective fields of speciality, and have been cooperating to achieve the consistent standardization of geographical names.

Efforts of standardization have mainly been made on natural feature names and place names. Administrative division names are determined by laws and regulations, therefore they are already standardized.

The agencies engaged in mapping, namely, the Geographical Survey Institute and the Hydrographic Department of Maritime Safety Agency, have been collecting geographical names for the purpose of annotating their maps and charts. In order to coordinate the use of geographical names by the two agencies, the Joint Committee on the Standardization of Geographical Names was established in 1960 and has been convened once or twice a year.

The committee completed in 1978 the standardization of about six thousand (6,000) broad natural feature names to be used on 1:500,000 scale maps and charts. And review work of the standardized names, which includes delineation of the area indicated by each geographical name, has been continued since then.

The committee has been trying to standardize the names to be placed on the 1:25,000 scale maps and similar scale charts since 1978. Nineteen thousand (19,000) out of total about four hundred thousand (400,000) geographical names were standardized by the time of the last conference of the committee held in 1989. With this slow progress rate, namely almost half a percent (0.5%) of total number per year, it will take two hundred (200) years to finish the work. To make a breakthrough in this work it is necessary to apply some automatic processing means.

For reference, there are three cases for determining a standard name. The first case is that the name is found only either in the maps of the Geographical Survey Institute or in the

charts of the Hydrographic Department, the second is that the same name is found in both maps and charts, and the third is that different names are found in maps and charts for the same geographical entity. In the former two cases, whereas the found name becomes standard immediately, most parts of standardization works are simple but laborious confirmation works, which can be well managed by the application of computer technologies.

In the third case, the standardization is difficult because of the following two reasons. One reason is that each agency adopts geographical names from different sources. The Geographical Survey Institute adopts geographical names from the Geographical Names Descriptions reported by the local governments, while the Hydrographic Department adopts geographical names from the Geographical Names Cards which are the result of its own survey. The other reason is that each agency has a different attitude to change the usages of geographical names. Regarding to the geographical names on the 1:25,000 scale topographical maps, the Geographical Survey Institute has a policy which is to follow exactly present-day local usages, which are change in time, while for charts the Hydrographic Department tends to keep the usages once adopted for charts to avoid confusions on maritime safety administration. Consequently, both the agencies can not agree to standardize some geographical names of the third case.

(2) Names of Undersea Features

The names of undersea features adopted by the GEBCO Subcommittee on Geographical Names of Ocean Bottom Features and approved by IHO are being used as the standards with no exception in Japan.

Undersea features discovered and observed by Japanese agencies have officially been named by the Hydrographic Department after deliberation by the Assembly on Geographical Names of Oceans and Bottom Features, the members of which are agencies associated with ocean and undersea observation.

Since the 14th Session, one hundred and sixty (160) names have been adopted and the total number of the adopted names is nine hundred and thirty (930) at the present time.

(3) Antarctic Geographical Names

The naming of natural features and important places discovered and observed by the Japanese Antarctic Research Expedition is promoted by the National Institute of Polar Research.

The director of the Institute, seeking counsel from the Antarctic Place-Names Committee of Japan, makes out a draft of the new names. On the basis of the draft, the new names are adopted

at the general convention of the Headquarters of the Japanese Antarctic Research Expedition.

The names of two hundred and seventy-six (276) places have been adopted by the time of the 14th Session, and fourteen (14) new names have been added since then.

2. Romanization of Geographical Names

Two systems of Roman spelling of the Japanese language, which are called the "Kunrei Siki" system and the "Syusei Hebon Siki" system (the modified Hepburn system), have been advertised and wide spread for a long time.

In order to achieve unification the Government of Japan issued in 1954 the notification of the cabinet which meant the "Kunrei Siki" system should always be used while the "Syusei Hebon Siki" system might be tolerated when the immediate change of the system was difficult.

The geographical names placed on the Japanese official maps and charts prepared for international use issued by the Geographical Survey Institute and the Hydrographic Department have been accordingly spelled with the "Kunrei Siki" system.

The "Syusei Hebon Siki" system, however, has been yet widely used in passport names, road signs and other indications for example. Considering the historical background and the present situation, it appears to be quite difficult immediately to abandon one system.

For reference, it must be noted that only fourteen (14) out of a hundred (100) basic sounds are differently spelled with the two systems, and that the conversion rule from one system to the other is very simple.

3. Construction of Geographical Names Data Base

A geographical names data base of Japan was constructed by the Geographical Survey Institute in 1991. This data base consists of one hundred and twenty thousand (120,000) geographical names placed on the 1:200,000 scale regional maps. The data base is expected to be an efficient tool to make the list of the geographical names which is to be compared with the corresponding list prepared independently for charts. Background, purposes, contents, construction procedure and use of the data base are described below:

(1) Background

Geographical names, which are important means to designate geographical entities, are widely used in social activities from ancient ages. For example, it is in the eight century that the

old Government of Japan ordered the local governments to compile the "Hudoki", which is an old gazetteer and topographic description.

In addition, the Japanese language is written in three character systems such as ideogram "Kanji" system, phonogram "Hirakana" system and another phonogram "Katakana" system. Official Japanese documents are written with a combination of "Kanji" and "Hirakana" characters, while "Katakana" characters are used only for imported words and proper nouns such as some personal, commodity and geographical names. Since "Kanji" characters usually have more than two pronunciations and most of geographical names include "Kanji" characters in it's official spelling, it is difficult even for native speakers to pronounce or to read written geographical names correctly and to write down pronounced geographical names correctly.

By these reasons, many geographical names data bases including gazetteers have been constructed in Japan. In these existing data bases, the location of the entity indicated by the geographical name is shown simply by the administrative division name or the rough geographical coordinates. Furthermore, official data bases, which have usually authoritative information about spellings and pronunciations, contain names of administrative divisions only, while in private company's data bases, which include other feature names such as the names of natural features and transport facilities, information about spelling and pronunciation of geographical names is not necessarily authoritative. In other words, there is no sufficient data base for the standardization of geographical names.

(2) Purposes of the data base

From the Joint Committee's activities for more than ten (10) years, it became clear that the standardization of the geographical names placed on the 1:25,000 scale topographic maps takes unfeasibly long time by manual means. The most troublesome part of the work is to make two respective lists of geographical names on maps and charts and to compare them. The progress of computer technology makes it possible with less cost to exempt the above mentioned job by means of data base.

Under this situation, the Geographical Survey Institute decided to set up a geographical names data base in order to use it as a tool for standardization of geographical names. In addition there is another purpose of the data base to eliminate duplication in time and money spent by government agencies, private companies and institutions to organize similar data bases for specific needs, in particular, to automate the processes of lettering the maps and to provide an indexing system of regions, areas and places for establishing geographical information sys-

tems.

To achieve these purposes, the new data base should cover entire Japan and all kinds of features, and it should contain precise information on such items as official spelling, pronunciation, location, etc. of the geographical names.

Hence the following characteristics are indispensable for primary sources of the geographical names in the new data base.

a. The sources cover all over Japan in the same accuracy;
b. The sources contain all classes of features;
c. The sources contain the geographical names which were adopted according to a consistent criterion;
d. Correct pronunciations and spellings can be obtained from the sources; and
e. Precise positions can be measured from the sources.

To meet these requirements, the most suitable sources are the 1:25,000 scale topographic maps, which cover all over Japan with approximately four thousand (4,000) sheets and contain approximately four hundred thousand (400,000) geographical names. However the 1:200,000 scale regional maps, which cover all over Japan with one hundred and thirty (130) sheets and contain one hundred and twenty thousand (120,000) geographical names, was selected as sources of the new data base, because of budgetary limitations and of the committee's urgent need for the data base.

(3) Contents of the data base

The information items of the data base are as follows:

a. Map sheet name

The sheet names of both a 1:200,000 scale regional map and a 1:50,000 scale topographic map are recorded. The sheet name of the 1:50,000 scale topographic map indicates where the geographical name is found in the 1:200,000 scale regional map.

b. Identity number

The sequential number of the geographical name is recorded.

c. Feature classification

Features are categorized into ten (10) classes such as administrative division, summit, lake, river, sea area, beach, island, broad natural feature, transportation facility and others. Some classes are divided into subclasses. The feature class of the geographical name is recorded. The size and spacing of characters can be known by the feature class because this classification is consistent to the lettering rules of the 1:200,000 scale regional maps.

d. Administrative division code

The code number of the administrative division which includes the area delineated or pointed by the geographical name is recorded. The code number is decided by the Ministry of Home Af-

e. Spelling

The official spelling, which usually includes ideogram "Kanji" characters, is recorded.

f. Pronunciation

The pronunciation derived from the Geographical Names Description reported by the local government is recorded.

g. Number of the characters

The total number of characters in the official spelling of the geographical name is recorded.

h. Position of the geographical name on the 1:200,000 scale regional map

The lower left corner point coordinates of both the first and the last characters of the spell string are recorded. The spacing can be calculated from this data and the number of characters. The spacing of characters of the geographical names of subdivided administrative division or railway station is fixed according to the lettering rules. Therefore only the corner point of the first character is recorded for this kind of name.

i. Relation between character orientation and spell string direction

The character orientation with respect to the spell string direction, parallel or cross, is distinguished and recorded.

j. Mispronunciations

Possible mispronunciations of the geographical name is recorded. Therefore the geographical name can be found even if correct pronunciation is unknown.

k. Data source

A principal source material of the geographical name in the data base, namely the sheet name of the 1:200,000 scale regional map which contain the geographical name, is recorded.

(4) Procedure of construction of the data base

There are four main stages in preparing the data base, namely, collection, listing, digitization, and compilation.

a. Collection

Primary sources such as the 1:200,000 scale regional maps and the Geographical Names Descriptions are collected. Then all geographical names placed on the source maps are marked for digitization and numbered for listing.

b. Listing

The geographical names are listed with the attribute information such as official spelling, pronunciation, feature class and so on obtained from the sources.

c. Digitization

Necessary points are digitized with accuracy of one tenth second (0.1") in longitude and latitude.

d. Compilation

The listed attribute information and digitized coordinates are combined into a data base, which is stored on thirty (30) megabyte hard disk.

(5) Use of the data base

The main use of the data base is, of course, the stadardization of geographical names. Other uses are as follows;

a. To publish a gazetteer of natural features

Using the data base, a gazetteer of natural features was compiled as a printing product. Twenty thousand (20,000) natural features are recorded in the gazetteer with their official spellings, pronunciations, standard mesh code numbers, feature classes and sheet numbers of the respective 1:50,000 scale topographic maps. Besides a data base of the same contents stored on floppy disk is available for a personal computer. The gazetteer will be used in government and private organizations for multiple purposes.

b. To establish a reference system of aerial photographs

A system to search the principal points and flight lines of all aerial photographs taken by the Geographical Survey Institute was constructed using the data base.

c. To assist construction of a geographical information system

The Geographical Survey Institute is constructing a geographical information system which utilizes the data base as a reference tool.

d. To study on automatic making of names plate for map printing

One of characteristics of the new data base is to contain precise and various map annotation data such as position, type size and styles of characters, relation between character orientation and spell string direction and others. It is expected to create a names plate automatically or to place the names on a draft of small scale map by using computer technology. The use of the data base for this purpose will be studied.

e. To make scientific study on geographical names

Distribution of specific geographical names, preferred "Kanji" characters used for spelling geographical names and so on are easily studied by using the data base. Analyzing these results, origins of geographical names, influence of natural and social conditions on geographical names and effect of people's wishes on geographical names can be studied.

For example, the Japanese language has three types of suffixes which mean small river and are generic part of geographical names. The first type is "-tani" or "-dani", the second type is "-sawa" or "-zawa", and the third is "-nai". The distribution of geographical names with each type of suffix is shown in Fig.1. The former two distributions well reflect the regional structure of Japanese culture, while the last one coincides with the area

of non-rice culture called "Ainu" race culture till the 11th century. Though the fact itself is already well known in Japan by qualitative analysis, the further study becomes possible by quantitative analysis using the data base.

Fig.2 shows the density of annotation on the 1:200,000 scale regional maps. From Fig. 2, it can be recognized that the coastal area has more geographical names than the inland area, and that the southwestern part of Japan has more geographical names than the northeastern part of Japan. Some useful information will be derived from the figure for the planning survey and mapping works in Japan.

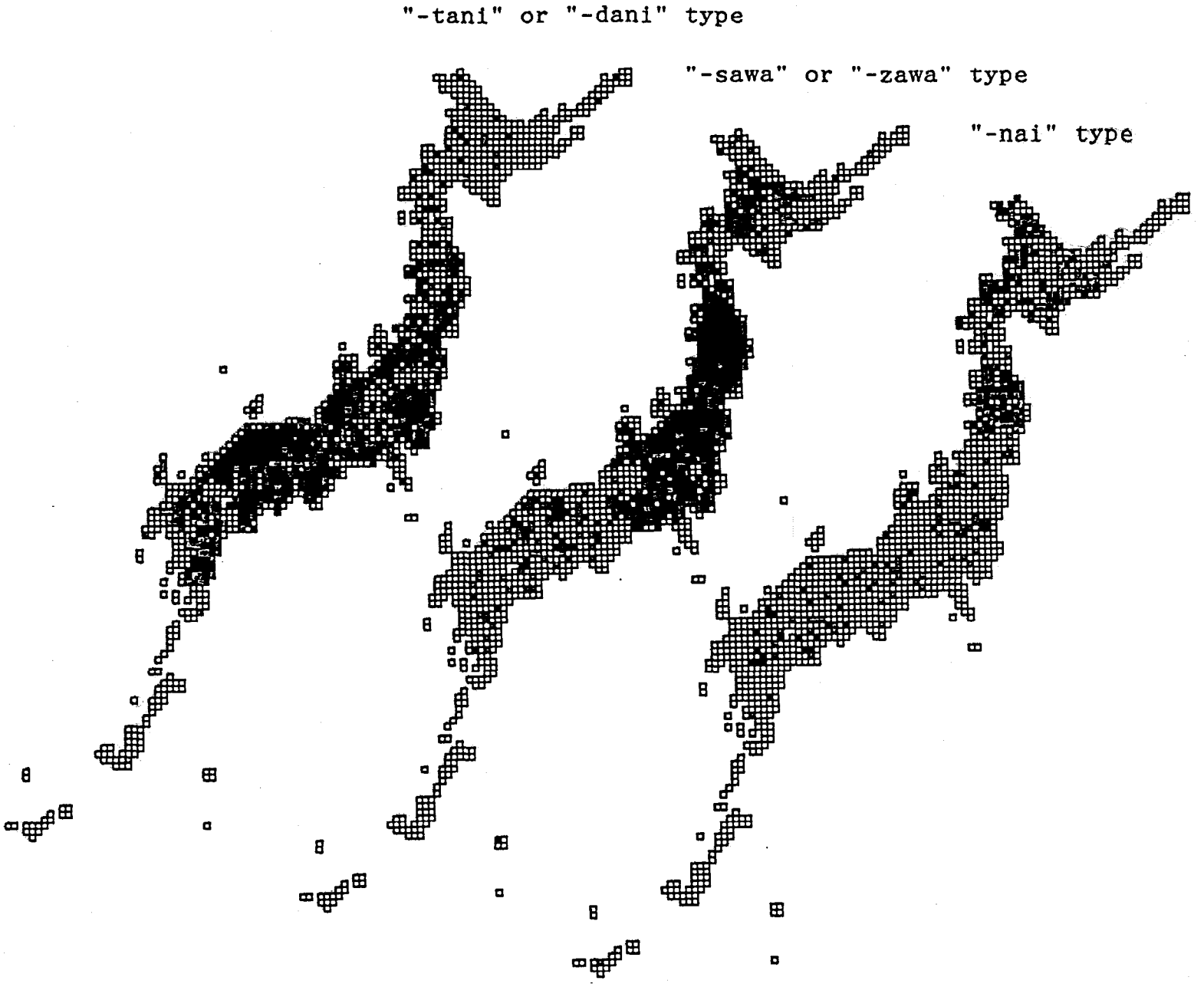

Fig.1 Distribution of the geographical names with three types of suffixes, which mean small rivers

The cells with more than three geographical names are black, the cells with one or two geographical names are gray and the cells with no geographical name are blank.

The southwest ends of each distributions are relatively sharp and the northeast ends are loose. This fact seems to suggest the direction of cultural transition is from southwest to northeast.

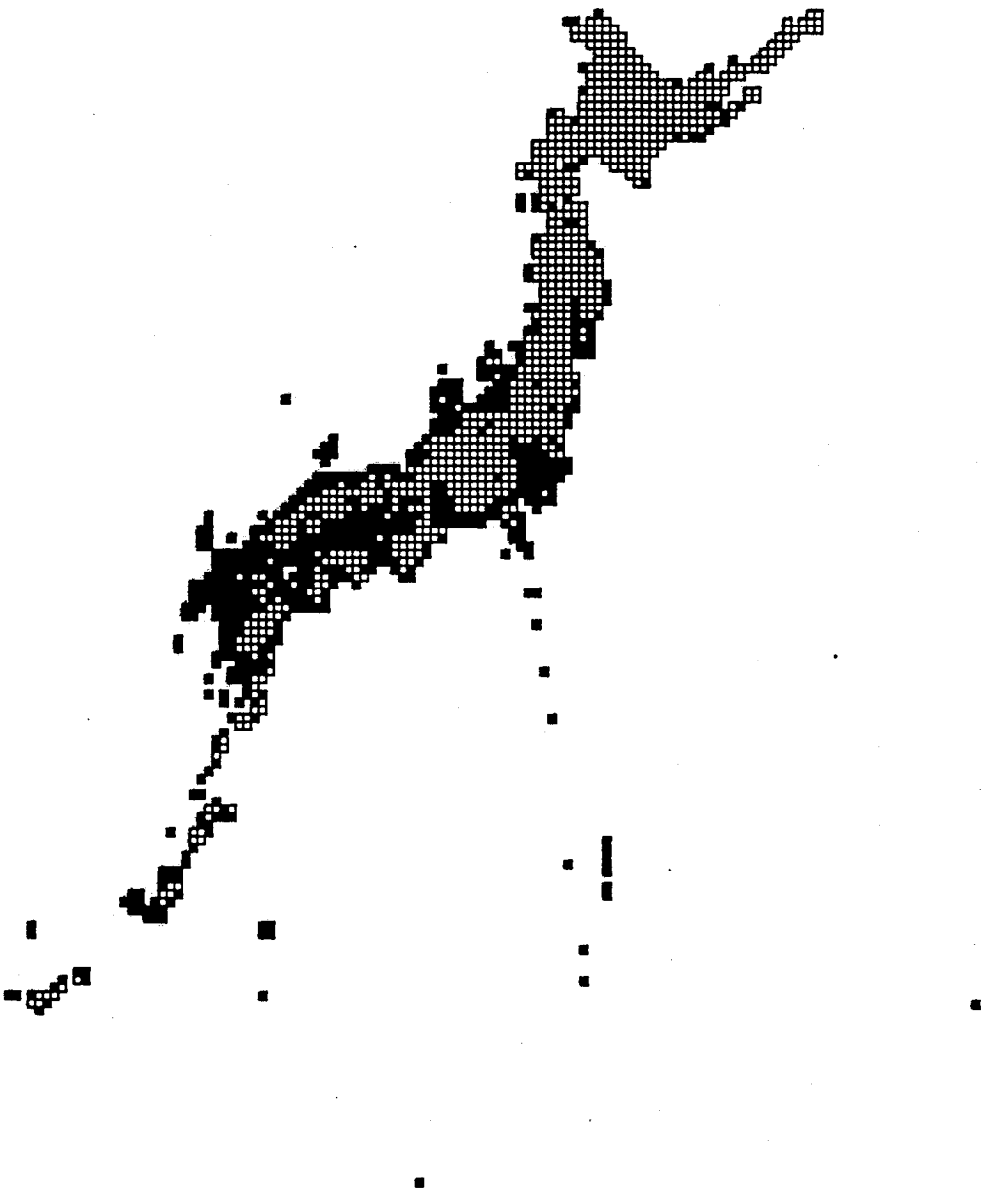

Fig.2 Distribution of annotation density

The cells in which annotation density is more than 0.32/km^2 are black, while the other cells are blank.