ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN – POLONIA
VOL.XLIX, 23 SECTIO B 1994

Zakład Kartografii
Wydziału Biologii i Nauk o Ziemi UMCS

Mieczysław SIRKO

APPLICATION OF BINARY SYSTEM TO OBJECTIVIZE SETTLEMENT GENERALIZATION PROCESS ON MAPS

Próba zastosowania systemu biernego jako środka prowadzącego do aktywizacji procesu generalizacji osiedli na mapach

The problem of generalization of map contents was and still is one of the basic goals of cartography, which despite ample literature on the subject, has not been fully solved. It has been dealt with both in monographs (F. Töpfer 1979) and in comprehensive studies (M. Sirko 1988) as well as by men who have greatly contributed to the development of cartography (E. Sydov 1866, M. Eckert 1921, 1925, E. Imhof 1923/24, 1956, K. A. Salischev 1939, E. Raisz 1938, A. Robinson 1960, L. Ratajski 1973, V. J. Suchov 1947, 1951, 1953, W. Pawlak 1971, 1972, W. Grygorenko 1973, 1979, 1981).

The main direction of present-day studies on cartographic generalization is the tendency to objectivize this process. It is the tendency to express, in a mathematically formal way, the basic principles that must be followed in working out map contents.

One of the main map elements that still poses difficulties in map drafting is settlement.. This is an element whose mode of representation on the map has been seemingly solved but none of the existing methods has been generally accepted.

One of the conditions of a correct generalization of settlements is to classify them according to how important or essential they are from the standpoint of the intended generalization. It was assumed that every settlement can be characterized by a number of features that will permit to objectively assess its value (importance), which is essential for the process of generalization. Cartographic references provide numerous examples of such features adopted by various authors (A. V. Borodin 1976, O. M. Dixon 1967, W. Keller 1970, V. V. Ivanov 1964).

In order to better and fully examine the set of the studied localities, both in the cross-section of its elements and in the cross-section of the distinguished variables (features), a system of classification was to be applied. Therefore a research procedure had to be developed that would be formally correct. A system of classification is correct when it fulfills the conditions of adequacy and disjunction. The condition of adequacy stipulates that the sum of subsets is identical with the set, while the condition of disjunction indicates that particular subsets do not contain elements in common. (Cf. D. L. Armand 1980, J. J. Parysek 1982.)

Consequently, a correct classification can affect the value of a study with regard to its adequacy. The probability of obtaining objective results is greater when the variables (features) characterize the objects of investigation (i.e. localities) as fully as possible.

The adopted investigative procedure fundamentally affects the obtained results: as early as the stage of feature selection the problem of credibility and adequacy of further investigations is to a large extent resolved.

Fully aware of the significance of this stage of the study we decided that in the conducted investigations variables will have to be taken into account that permit a relatively complete characteristics of settlements with respect to their geographical features.

Table 1

No	Feature	Number of distinguished classes
1	Number of inhabitants in a settlement	10
2	Administrative importance of a settlement	5
3	Level of settlement urbanization	5
4	Economic importance of a settlement	5
5	Road transport access to a settlement	7
6	Railway transport access to a settlement	7
7	Historical importance of a settlement	5
8	Tourist value of a settlement	5
9	Settlement location with respect to river network	5
Total		54

The selected features were divided into 54 classes, which permitted to unequivocally determine the characteristics of the studied settlements*.

* Division into classes and justification of their distinction have been discussed in an earlier study (M. Sirko 1988).

Each of the adopted features classifies the whole set of settlements. At the same time particular subsets are disjunctive: they do not contain elements in common. The grouping in Table I meets thereby the two basic conditions of correct classification: adequacy and disjunction. The objective of feature selection was that the adopted classification criteria should comprehensively characterize all settlements, in particular the small ones that are largely the main object and problem of cartographic generalization.

As a consequence of adopting definite features it is necessary to determine their rank. Out of the nine variables only two, the number of settlement inhabitants and economic importance, are quantitive, the other seven being qualitative features. Some of the adopted features could obviously be quantitive under certain assumptions: for example road or railway transport access or settlement location with respect to river network. A possible solution for the former variable would be to adopt the degree of intensity of traffic flows or the number of roads linking a given locality (A. Gotz 1976). However, those ways of solving the problem of quantitive characteristics of the settlement's transport value are not fully satisfactory. The degree of intensity of transport flows, measured with the amount of carried goods, is not identical with the significance of particular type of transport connection. The other indicator, the number of roads linking a settelment, is also inadequate for their importance.

In the case of settlement location with respect to river network, the quantitive index could be the volume measure of river flow or the degree of river utilization by industries and public utilities. The adoption of these measures is very difficult because there are no data on which to base calculations.

Assuming that it would be possible to obtain values characterizing the above two variables, there is still a problem of how to determine the values for the remaining five features, which are exclusively qualitative whereas they seem to play a significant role in defining the rank of a settlement: for example historical importance, administrative rank, tourist value or the level of urbanization. That is why the quantitative representation of features for which numerical data could be obtained was abandoned, all criteria being treated as equally important. That accounts for the necessity to adopt a different, so far not applied system that would permit to uniformly characterize each studied variable through quantitative indicators.

Every feature was divided into several classes whose order and arrangement were not accidental. It was established in such a way that each subsequent class of localities should have an inferior rank with respect to settlements assigned to a higher class, from the stanpoint of the studied variable. That way the system of the division of variables reflects the rank of a locality and the hierarchy inside the division into classes has been preserved. The adopted manner of ordering groups (classes) within particular variables is extremely important since it ensures the possibility of obtaining indicative values for each of them.

The next stage of the study was description of particular features in the binary system, where two values are employed: zero and one (0-1). In view of the fact that a particular settlement can be assigned exclusively to one of the classes of the studied variable, the class to which it is assigned receives the value of one (1) while all the others zero (0). We can illustrate this with the following example. The category of the number of settlement inhabitants was divided into ten classes (Table 1). Thus, if the studied settlement is assigned to the highest class (over 50,000 inhabitants), then its binary notation will have the form:

10,000,000

and the corresponding number will be 2^9 or 512. The minimum value will go to settlements assigned to the lowest class (below 100 inhabitants) and will equal 2^0 or zero (0).

By analogy, we can examine another variable, for example settlement location with respect to river network, whose feature was divided into 5 classes. Assuming that the studied settlement belongs to the highest group (class), i.e. it is situated on the first – order river, its notation in the binary system will have the form:

10,000

Its corresponding number equals 2^4 or 16, the minimum value being, as with the settlement size, 2^0 or zero (0).

In view of the fact that the studied features are divided into different numbers of classes, the presented mode favours variables with a greater number of groups. We do not know, however, if the features described with a greater number of classes are more or less significant with settlement generalization than variables divided into a smaller number of classes. That is why the values of the obtained indicators should be made comparable regardless of the number of adopted classes. The obtained data should be therefore standardized. This can be done in such a way that the obtained results will range from zero to one (0-1). For that purpose, the obtained binary value for a given locality should be devided by the maximum value that a given settlement would have if it was assigned to the highest class of the studied variable.

This can be described by the following formula:

$$\frac{k}{k_{max}}$$

0 – when a locality is assigned to the lowest class

1 – when a locality is assigned to the highest class

In other words, if while examining the importance of a given settlement according to a particular variable, we find that the settlement should be assigned to the lowest class, then the "k" value will be zero (0) and thereby the indicator value will also equal zero. This conforms with the binary notation because it corresponds to 2^0 or zero. If, however, the locality is assigned to the highest class, then the "k" value will equal the "k_{max}" value or the indicator value will be one (1).

Let us illustrate this mode of reasoning with an example. The criterion of the number of settlement inhabitants was divided into 10 classes. Each has a corresponding binary notation. Starting from the lowest to the highest class it will have the form:

class no.	notation	numbers
1	1	2^0
2	10	2^1
3	100	2^2
4	1,000	2^3

5	10,000	2^4
6	100,000	2^5
7	1,000,000	2^6
8	10,000,000	2^7
9	100,000,000	2^8
10	1,000,000,000	2^9

The maximum value in this notation will thus have the form:

1,000,000,000

This corresponds to 2^9 or 512. If we find with a given settlement that its size assigns it to class nine (25.000 to 50.000 inhabitants, $2^8 = 256$) then the indicator will have the value:

$$\frac{k}{k_{max}} = \frac{156}{512} = 0,5$$

By analogy let us examine the feature of settlement location with respect to river network. This variable was divided into five classes, the binary notation having the following form:

class no.	notation	numbers
1	1	2^0
2	10	2^1
3	100	2^2
4	1,000	2^3
5	10,000	2^4

Maximum value will go to localities situated on the first-order rivers (class 5) and will have a binary notation 10,000, which corresponds to 2^4 or 16. If a given locality is situated on the second-order river (class 4), the notation will have the form 1,000 and the number 2^3 or 8. The indicator value for this locality will be:

$$\frac{k}{k_{max}} = \frac{8}{16} = 0,5$$

The results are thus comparable in both cases. Before we proceed, however, we should examine the question whether the adopted procedure of indicator calculation is correct. Or, whether the binary system applied to calculate indicator values, adopted to investigate qualitative features, does not contain some methodological faults.

Let us return to the example in question. The number of classes for particular features varies, ranging from five with settlement location with respect to river network, to ten with the number of settlement inhabitants. We should therefore find out what values indicators will have for particular classes in examining the two cited variables:

settlement location with respect to river network (5 classes)	number of inhabitants in a settlement (10 classes)
1,000	1,000
0.500	0.500
0.250	0.250
0.125	0.125

0.0625
0.0312
0.0156
0.0078
0.0039
0.0000

The foregoing figures show that indicator values for the variable with a smaller number of classes as compared with the variable with a higher number of classes are not proportionate. In each case, values were obtained that decreased geometrically, the variable with five classes having indicators grouped in the upper end of the value scale of the other variable with ten classes. With this mode of calculating numerical indicators based on qualitative features we would undeniably have total equivalence of particular features if the number of classes in each studied feature was identical, and the importance of variables for the problem of settlement selection had the same value (weight). This assumption is theoretically possible to implement. This can be done by adding several classes to be divided for the features with the lowest number of them, or by reducing the number of classes for the features with the highest number. As, however, we do not know the actual importance of the variables, this procedure would artificially increase information not substantiated by its real value, or it would decrease the amount of significant information. Consequently, the obtained results might contain a large error margin.

The five-degree scale needs therefore to be remodelled in such a way that the distribution of indicator values should be preserved according to geometric progression while at the same time they should be distributed along a ten-degree scale. For that purpose I assumed that the middle value on the five-degree scale corresponding to the middle of the ten--degree scale is a geometric mean of two midddle values on the degree scale. The remaining indicators were calculated according to the definition of geometric progresion, assuming that the last value equalled zero. As a result we obtained the following distribution of indicator values on the five-degree scale versus the ten-degree scale.

settlement location with respect to river network (5 classes)	number of inhabitants in a settlement (10 classes)
1,000	1,000
	0.500
	0.250
0.210	0.125
	0.0625
0.0441	0.0312
0.0093	0.0156
	0.0078
	0.0039
0.0000	0.0000

I used the analogous method to calculate indicator values for the features that were divided into seven classes (road and railway transport access to a settlement).

The foregoing methodological assumptions concerning settlement selection on the map were then tested in the course of investigations conducted in the Zamość voivodeship.

Fig. 1. Zamość Voievodeship – settlement shown by signature (without generalization)
Województwo zamojskie – osiedle przedstawione za pomocą oznaczeń (bez uogólnienia)

When we started to collect reference materials that would permit to characterize every settlement with respect to the adopted nine variables, I understood that their completeness and reliability would determine whether the obtained results would be free from errors. For, if the map on which we base our investigations contains a number of inaccuracies concerning for example locality selection, then this error will be duplicated in a new cartographic study. Following that assumption, we sought to select such reference sources where the picture of settelment is complete, without a quantitive generalization. If a reference source was found incomplete with respect to the adopted criteria, it was necessary to supplement it with indispensable elements.

The cartographic source that served to collect the basic data for generalization of settlement network were map sheets in the scale of 1:100,000 made in the GUGiK-80 projection and the administrative map of Poland in the scale of 1:300,000, the Zamość voivodeship.

Maps in the scale of 1:100,000 served to determine the road and railway transport access to localities and to establish their situation with respect to river network. The administrative map, due to settlement representation with cartographic symbols, was used to determine the position of every locality. It was found, however, that the 1:300,000 map did not give the complete picture of settlements. Many localities, due to quantitive generalization, were not represented on the map. Following the principle that the absence of a settlement symbol on the 1:300,000 map does not mean that it is of lesser importance than localities represented on the map, we decided to complete the picture of settlement to the same

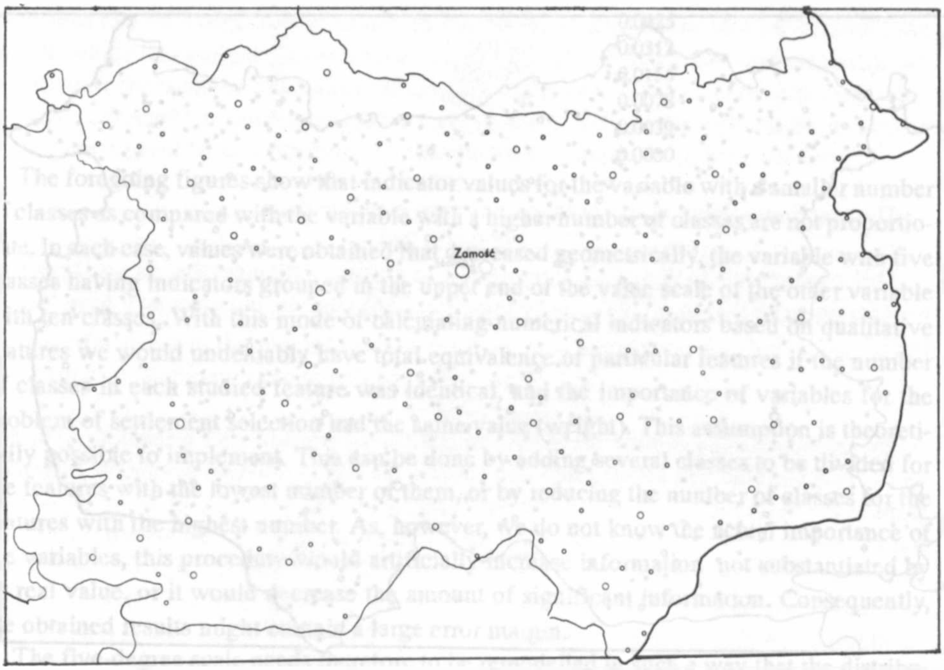

Fig. 2. Zamość Voivodeship – settlement shown by signature (settlements classified on map in scale 1:1,000,000)

Województwo zamojskie – osiedle przedstawione za pomocą oznaczeń (osiedle klasyfikowano na mapie w skali 1:1 000 000)

degree as on the 1:100,000 map. For that purpose, the focal point was established for every locality absent on the administrative map and subsequently projected on to the map in the scale of 1:300,000. The administrative map complemented in the described manner gives practically the same picture of settlements as on the map in the scale 1:100,000, being shown with cartographic symbols. (Fig. 1). Out of the total of 1,024 settlements in the Zamość voivodeship, this procedure covered 186 localities or 18.2%.

It was subsequently necessary to determine its class for each individual settlement from the standpoint of the particular features. Or, it was necessary to assign each of the 1,024 settlements to a definite class by assessing it for its administrative, economic or tourist importance, or the level of urbanization, etc.: that is for each of the nine adopted features. This is equivalent to determining indicator values for the particular features. As a result, I obtained a collection of data that characterize all settlemens in a detailed way.

Having calculated the values of all the individual indicators w_i that characterize a given settlement in relation to all variables, and subsequently added them up, I obtained a new, total indicator w_s that expresses, comprehensively, the importance of every settlement in relation to other settlements in the studied area.

The calculations of total indicators characterizing all the 1,024 settlements in the studied area differentiated with respect to the obtained numerical values. As each of the individual indicators w_i determines the importance of a settlement in the set of all settlements

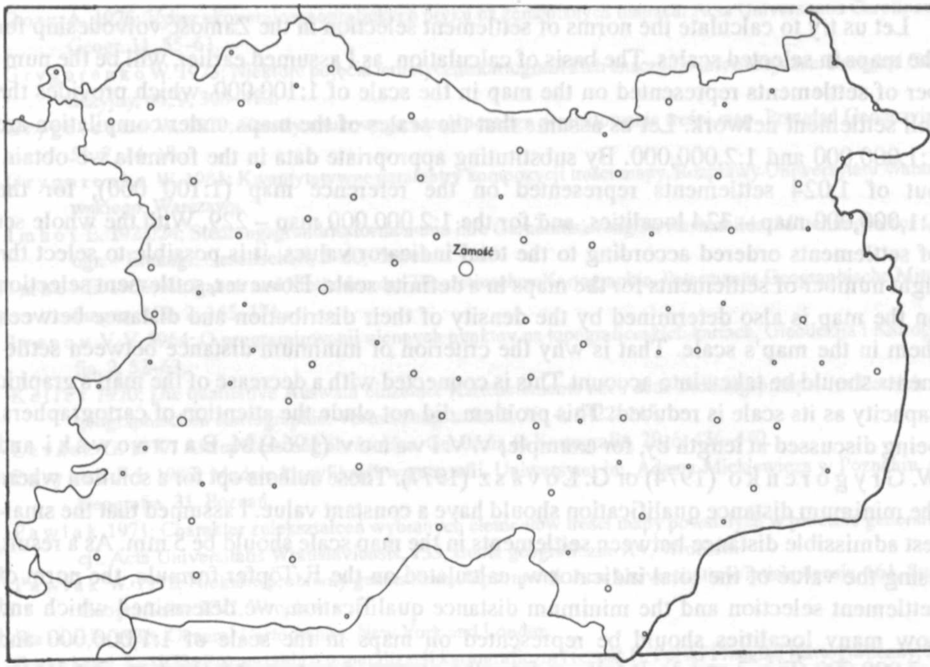

Fig. 3. Zamość Voievodeship – settlement shown by signature (settlement classified on map in scale 1:2,000,000)
Województwo zamojskie – osiedle przedstawione za pomocą oznaczeń (mapa w skali 1:2 000 000)

with respect to one feature, I assumed that the total indicator would also determine the importance of the settlement but more comprehensively, that is with respect to all the nine features adopted for the investigation of settlements.

The whole set of localities, ordered according to the decreasing value of w_s does not, however, answer what number of them should be represented on a given map. This depends, among others, on the scale of a map: the smaller the scale, the smaller its area. It can be said that there is a dependence between the map scale and the degree of filling it with content. Therefore, while determining the number of settlements that are to be projected onto the map, it is necessary to use a method that takes into account the scale of the map. Such a method is the F. Töpfer (1979) square root formula. It assumes that the number of objects (in this case settlements) on the map under compilation equals the product of the number of objects on the reference map multiplied by the square root of the quotient of the denominator of the reference map scale and the denominator of the scale of the map under compilation.

$$n_b = n_a \sqrt{\frac{M_a}{M_b}}$$

where: n_b – number of objects on the map under compilation, n_a – number of objects on the source map, M_a – denominator of the source map scale), M_b – denominator of the scale of the map under compilation.

Let us try to calculate the norms of settlement selection in the Zamość voivodeship for the maps in selected scales. The basis of calculation, as I assumed earlier, will be the number of settlements represented on the map in the scale of 1:100,000, which provides the full settlement network. Let us assume that the scales of the maps under compilation are 1:1,000,000 and 1:2,000,000. By substituting appropriate data in the formula we obtain, out of 1,024 settlements represented on the reference map (1:100 000), for the 1:1,000,000 map – 324 localities, and for the 1:2,000,000 map – 229. With the whole set of settlements ordered according to the total indicator values, it is possible to select the right number of settlements for the maps in a definite scale. However, settlement selection on the map is also determined by the density of their distribution and distance between them in the map's scale. That is why the criterion of minimum distance between settlements should be taken into account.This is connected with a decrease of the map's graphic capacity as its scale is reduced. This problem did not elude the attention of cartographers, being discussed at length by, for example, V. V. Ivanov (1964) M. Baranowski and W. Grygorenko (1974) or G. Lovász (1977). Those authors opt for a solution where the minimum distance qualification should have a constant value. I assumed that the smallest admissible distance between settlements in the map scale should be 5 mm. As a result, using the value of the total indicator w_s calculated on the F. Töpfer formula, the norm of settlement selection and the minimum distance qualification, we determined which and how many localities should be represented on maps in the scale of 1:1,000,000 and 1:2,000,000 (Fig. 2 and 3). It follows from the presented picture that there are 206 localities on the 1:1,000,000 map (Fig. 2) and 94 on the 1:2,000,000 map (Fig. 3). In comparison with the calculations obtained only with the Töpfer formula, 324 and 229 respectively, the number of settlements to be plotted on the map has considerably decreased. We can therefore say that the effect of the minimum distance coefficient is significant, its role becoming greater as the map scale decreases. This prompts an assertion that completion of map content changes with the change of its scale. The foregoing discussion can also lead to a conclusion that during the generalization process there occurs a certain flattening out of differences in the density of map content elements between high and low intensity areas. The process of generalization leads to the levelling out of differences in object density (i. e. settlements) between large-number areas and the rest of the map.

REFERENCES

Armand D. L. 1980; Nauka o krajobrazie. Podstawy teorii i metody logiczno matematyczne. PWN, Warszawa.

Baranowski M., Grygorenko W. 1974; Próba obiektywnego doboru osiedli na mapach z zastosowaniem maszyny cyfrowej. Polski Przegląd Kartograficzny, 6, 4, 149–158.

Borodin A. V. 1976; Voprosy gienmeralizacji kartograficeskogo izobraženija pri avtematiceskom sezdanii kart. Gieodezija i Kartografia, 7, 57–64.

Eckert M. 1921; Die Kartenwissenmschaft. Forschungen und Grumndlagen zu einer Kartographie als Wissenmschaft. Erster Band, Berlin und Leipzig.

Eckert M. 1925; Die Kartenwissenschjaft. Forschungen und Grundlagen zu einer Kartographie als Wissenschaft. Zweiter Band, Berlin und Leipzig.

G o t z A. 1976; Vyber ekonomicegeografickych pryku na zemepisnych mapach. Acta Universitatis Carolinae – Geogr. 11, 55–61.

G r y g o r e n k o W. 1973; Niektóre pojęcia teoretyczne kartografii i ich znaczenie informatyczne. Przegląd Geodezyjny, 44, 9, 398–400.

G r y g o r e n k o W. 1979; Zasady naukowego i technicznego projektowania treści map. Przegląd Geodezyjny, 51, 7, 14–18.

G r y g o r e n k o W. 1981; Kwantytatywne parametry kompozycji treści mapy. Rozprawy Uniwersytetu Warszawskiego, Warszawa.

I m h o f E. 1923/24; Siedlungsgrundrissformen und ihre Gerneralisierung im Kartenbilde. Mitteilungen der Geogr. – Ethnogr. Gesellschaft, 5–60, Zurich.

I m h o f E. 1956; Aufgaben und Methoden der Theoretiuschen Kartographie. Petermanns Geographische Mitteilungen, 100, 2, 165–171.

I v a n o v V. V. 1964; O programirovanii olennych punktov na topograficeskich kartach. Gieodezija i Kartografia, 2, 52–63.

K e l l e r 1970; Die qualitative Auswahl einzelner Kartenelemente nach dem Ordnungsprizip – im Bereich der geographischen Kartographie. Vermessungstechnik, 18, 4, 1422–145.

L e v á s z G. 1977; Atclepülésuség ábrázolása. Geodézia és Kartografia, 29, 6, 436–440.

P a r y s e k J. J. 1982; Modele klasyfikacji w geografii. Uniwersytet im. Adama Mickiewicza w Poznaniu, ser. Geografia, 31, Poznań.

P a w l a k 1971; Charakter zniekształceń wybranych elementów treści mapy powstałych w procesie generalizacji. Acta Universitatis Wratislaviensis, 133, studia geograficzne XV, Wrocław.

P a w l a k W. 1972; Niektóre problemy generalizacji hipsometrii. Acta Universitatis Wratislaviensis, 161, Studia Geograficzne XVI, Wrocław.

R a i s z E. 1938; General Cartography. New York and London.

R a t a j s k i L. 1973; Rozważania o generalizacji kartograficznej (część I). Polski Przegląd Kartograficzny, 5, 2, 49–55.

R a t a j s k i L.. 1973; Rozważania o generalizacji kartograficznej (część II). Polski Przegląd Kartograficzny, 5, 3, 103–110.

R o b i n s o n A. 1960; Elements of Cartography. 2nd edition, New York – London.

S a l i š č e v K. A. 1939; Osnovy kartoviedienija. Obscaja cast, Moskwa.

S i r k o M. 1988; Teoretyczne i metodyczne aspekty obiektywizacji doboru osiedli na mapach (Studium na przykładzie Polski Środkowo-Wschodniej). Uniwersytet Marii Curie Skłodowskiej, Lublin.

S u c h o v V. J. 1947; Izobraženije nasielennych punktov SSSR na topograficeskich kartach. Trudy CNIIGAAiK, 48, Geodezizdat, Moskwa.

S u c h o v V. J. 1951; Norma otbora nasielennych punktov dla mielkomastabnych obcegeograficeskich kart. Geodezizdat, Moskwa.

S u c h o v V. J. 1953; O tieoretgiceskich esnpvachj katograficeskoj gieneralizacji. Sbornik GUGik, vyp. 4, Moskwa.

S y d o w E. 1866; Drei Kartonklippen. Geographisches Jahrbuch, 348–361.

T ö p f e r F. 1979; Kartographische Generalisierung. VEB Hermann Haack Geogr.-Kart. Anstalt, Gotha/Leipzig.

STRESZCZENIE

Głównym kierunkiem współcześnie prowadzonych badań nad generalizacją kartograficzną jest dążenie do wyrażania w sposób matematycznie sformalizowany podstawowych zasad, którymi należy się kierować przy redagowaniu treści map.

Do głównych elementów map, stwarzających kartografom wiele trudności przy ich redagowaniu, należy osadnictwo. Jednym z warunków przeprowadzenia poprawnej generalizacji osiedli jest pogrupowanie ich według stopnia ważności. Każde osiedle można scharakteryzować przy pomocy wielu cech pozwalających na obiektywną ocenę jego wartości (ważności), co ma istotne znaczenie w procesie generalizacji. Prawdopodobieństwo uzyskania obiektywnych wyników jest większe wówczas, kiedy zmienne (cechy) charakteryzują w sposób mo-

żliwie pełny obiekty będące przedmiotem badania, w tym przypadku miejscowości. Dobrane 9 cech podzielono na 54 klasy, które pozwoliły na jednoznaczne określenie charakterystyki badanych osiedli. Konsekwencją przyjęcia określonych cech jest konieczność ustalenia ich rangi.

Można tego dokonać przy zastosowaniu systemu binarnego. Konkretne osiedle może być zaliczone do jednej z klas cechy badanej, a klasa, do której zostanie ono zaliczone, otrzyma wartość jeden (1), pozostałe wszystkie inne zero (0). W niniejszej pracy cechy podzielone są na różną liczbę klas. W związku z tym należało doprowadzić do tego, aby wartości wskaźników mogłyby być porównywalne. Dokonano tego poprzez rozkład wskaźników zgodnie z postępem geometrycznym. Z kolei należało ustalić dla każdego osiedla indywidualnie jego klasę z punktu widzenia poszczególnych badanych cech. W wyniku powstał zbiór danych charakteryzujących wszystkie osiedla. Jednakże nie otrzymano odpowiedzi, jaka liczba osiedli powinna się znaleźć na konkretnej mapie w określonej skali. Odpowiedź na tak postawione pytanie daje wzór pierwiastka kwadratowego F. Töpfera. Innym elementem, który należało uwzględnić, to kryterium minimalnej odległości między osiedlami.

W rezultacie ze względu na trzy wymienione wyżej elementy (wskaźnik wynikający z zastosowania systemu binarnego, wzór Töpfera i kryterium minimalnej odległości określono, które i ile osiedli powinno się znaleźć na mapach w określonej skali.