

SPATIAL CONFIGURATION OF TYPES OF PHONOLOGICAL SYSTEMS OF CENTRAL AND SOUTH-EUROPEAN LANGUAGES

M.I.Lekomtseva

Institute of Slavic and Balkan Studies of Academy of Sciences of USSR, Moscow

ABSTRACT

The paper presents an areal classification of the phonological systems of 18 languages of Central and South-East Europe. By means of cluster analysis of phonological specifications of these languages two areal types are obtained: Balto-Balkan type, presented by two areas, and the Central type which is situated between these areas.

After the discovery of phonological features N.Trubetzkoy and R.Jakobson were led to conclude that the distribution of phonological features in different languages was not random. The areal distribution of the phonological features resembled the isoglosses of lexemes studied in linguistic geography. At the 1st International Congress of Linguists N.Trubetzkoy outlined the task of studying the areal configurations of linguistic features /9/. R.Jakobson presented the first example of a language area (the Euro-Asian Sprachbund or linguistic area) modelled on the basis of compact territorial distribution of languages of different degrees of genealogical relation. The phonological systems of the corresponding languages were characterized by the two following dominant features: palatalization of consonants and monotony of vowels /4/. Areal language studies show that the phonological system of a language derives not only from its genealogical relations, but is likewise influenced by the neighbouring languages /2, 3, 5, 6, 7/. The development of the phonological system of a language is further conditioned by the language situation in the given region and the spatial structure of the linguistic area.

While the task of finding out the distribution of phonological features is all over clear, the choice of the guidelines for setting up a basis for defining the areal type of phonological systems (i.e. the domain of connectedness of a linguistic area) presents serious difficulties /3, 6/. E.g. though the phonological system of Irish is characterized by the two abovementioned dominant fea-

tures: the opposition of palatalized/non-palatalized consonants and monotony of vowels, the Irish language can hardly be recognized as belonging to the Euro-Asian language area. Two coincidences of features are not enough to decide whether the language belongs to a definite type /against: 3/.

The Balkan language area (like other currently defined language areas) does not hold such features that might be regarded as necessary and sufficient in the traditional sense. It is frequently noted in areal typology that every feature defined as specific for a linguistic area can also be found in a language beyond the area /3, 7/. On the other hand, some languages belonging to the area may lack a feature defined as specific. Essentially it is a high degree of similarity characterizes the languages belonging to the same domain of connectedness of a linguistic area /1, 2, 3, 6, 7/. The setting up of abstract ideal types to serve as a basis for a quantitative evaluation of real language systems /3/ seems to be of little value: the problem becomes one of constructing ideal types to be used for a further description of the language systems of a given area. A high degree of similarity among the languages of a given area however is frequently due to non-exclusive intersecting features, that do not fall into clear-cut patterns.

This paper presents an attempt of suggesting areal types of phonological systems on the basis of the languages of the Baltic-Balkan areal.

The phonological systems are viewed with regard to their inherent features, i.e. the features of monotony - polytony go beyond the scope of our study. Phoneme identification was accomplished according to the distinctive features of Chomsky-Halle and their amendment in Halle-Stevens. The syntagmatic aspect of the phonological systems was not taken into account.

Areal studies naturally fall into the domain of dialectology. N.Trubetzkoy regarded them as a continuation of dialectal studies /9/. The language material used and the way we have set our task inclines us, however, in the present preliminary stage of analysis to consider the area as represented by lan-

guages in standard form. Further tasks include a description of dialects as global phonological systems, each with a given set of features and setting up areal types of dialectal phonological systems of genealogically related and unrelated languages.

The areal types are set up on the basis of the phonological systems of the following languages belonging to the Balto-Balkan linguistic area: Latvian, Lithuanian, Upper-Lusatian, Polish, Belorussian, Russian, Ukrainian, Slovak, Czech, Hungarian, Romanian, Bulgarian, Turkish, Greek, Albanian, Macedonian, Serbo-Croatian and Slovenian languages.

A transparent procedure of construction, controlled objectivity on all stages and unambiguity of results are essential for setting up areal types of phonological features.

Due to the abovementioned peculiarities of areal language relations the traditional Aristotelean classification is believed to be of little value. In the present case a quantitative approach seems preferable. Of all techniques of cluster analysis, employed for the purpose of obtaining groups of objects characterized by a maximum degree of similarity, we have considered the Linker algorithm as best suited for our needs /8/.

The Linker algorithm gives hierarchically arranged object clusters and determines the relative degree of similarity by which the objects are clustered. Generally the algorithm guarantees a local rather than a global optimum. In our case if the objects can be enumerated, the algorithm guarantees the global optimum as well /8/. Consequently, if the data matrix has inherent structure cluster analysis will succeed in identifying it. The result of the Linker algorithm is unambiguous only if the degrees of similarity (or distance) among the objects are different from each other. If, however, there are identical degrees of similarity between two or more pairs of objects (as may frequently be the case in areal studies) the result of clustering becomes ambiguous. To eliminate this disadvantage of the Linker algorithm we have introduced a subalgorithm to be applied in the case of identical degrees of similarity (or distance). The subalgorithm assigns a clustering value to each claimant depending on the next step of the main algorithm, i.e. the joining of a given claimant with all the other clusters is preferred if this results in the maximum sum of the mean degrees of similarity (or correspondingly the least sum of the mean degrees of distance).

The Linker algorithm can be applied to the matrix of distance (or similarity) between given languages if metric space has been determined. The condition of metricity is met by applying the formula defining distance /8/ between the phonological systems of the given languages. I.e.

$$d = 1 - \frac{\alpha + \delta}{1}$$

where

$$\alpha = \sum_{k=1}^1 \min(x_k, y_k)$$

$$\beta = \sum_{k=1}^1 x_k - \alpha$$

$$\gamma = \sum_{k=1}^1 y_k - \alpha$$

$$\delta = 1 - (\alpha + \beta + \gamma)$$

1 - the number of positions (features) chosen to represent the given language. Else: α - the number of positions (features) where both languages have a positive values (1); δ - the number of positions (features) where both languages have a negative value (0).

The data matrix of the distance between the given languages is filled according to the abovementioned formula.

According to the algorithm the least distance between the languages is selected and the corresponding languages clustered. Next the mean distance from the obtained cluster to the rest of the languages is calculated. The languages showing minimum distance once more undergo clustering. In the case of several identical distances we introduce the subalgorithm. The routine is run until all the languages have been clustered.

The algorithm can be presented in the form of a dendrogram mapping the sequence of element and group clustering and showing the minimal distances at which the clusterings take place.

The languages considered present the following picture of language groups marked by an increasing degree of similarity in paradigmatic phonology (see Fig. 1).

Fig. 1 shows the maximum similarity (minimum distance) between the phonological systems of Romanian and Turkish (similarity, here, is viewed paradigmatically, - f.e. the introduction of syntagmatic features would naturally change the position of both Romanian and Turkish). Likewise the maximum degree of similarity characterizes Latvian and Lithuanian, Czech and Slovak, Serbo-Croatian and Slovenian, and Bulgarian and Macedonian.

The next cluster is formed by Czech, Slovak, Serbo-Croatian and Slovenian, while the following step adds Hungarian to the cluster. Last the group including Latvian and Lithuanian is added.

The Bulgarian-Macedonian cluster is joined by Greek, and next - by Albanian. It should be noted, however, that the distance between the initial language cluster (Bulgarian-Macedonian) and the Greek is much greater than the distance between the languages of the formerly mentioned language group. Both language clusters are united into one which is globally opposed to another cluster formed by consequent joining of Uk-

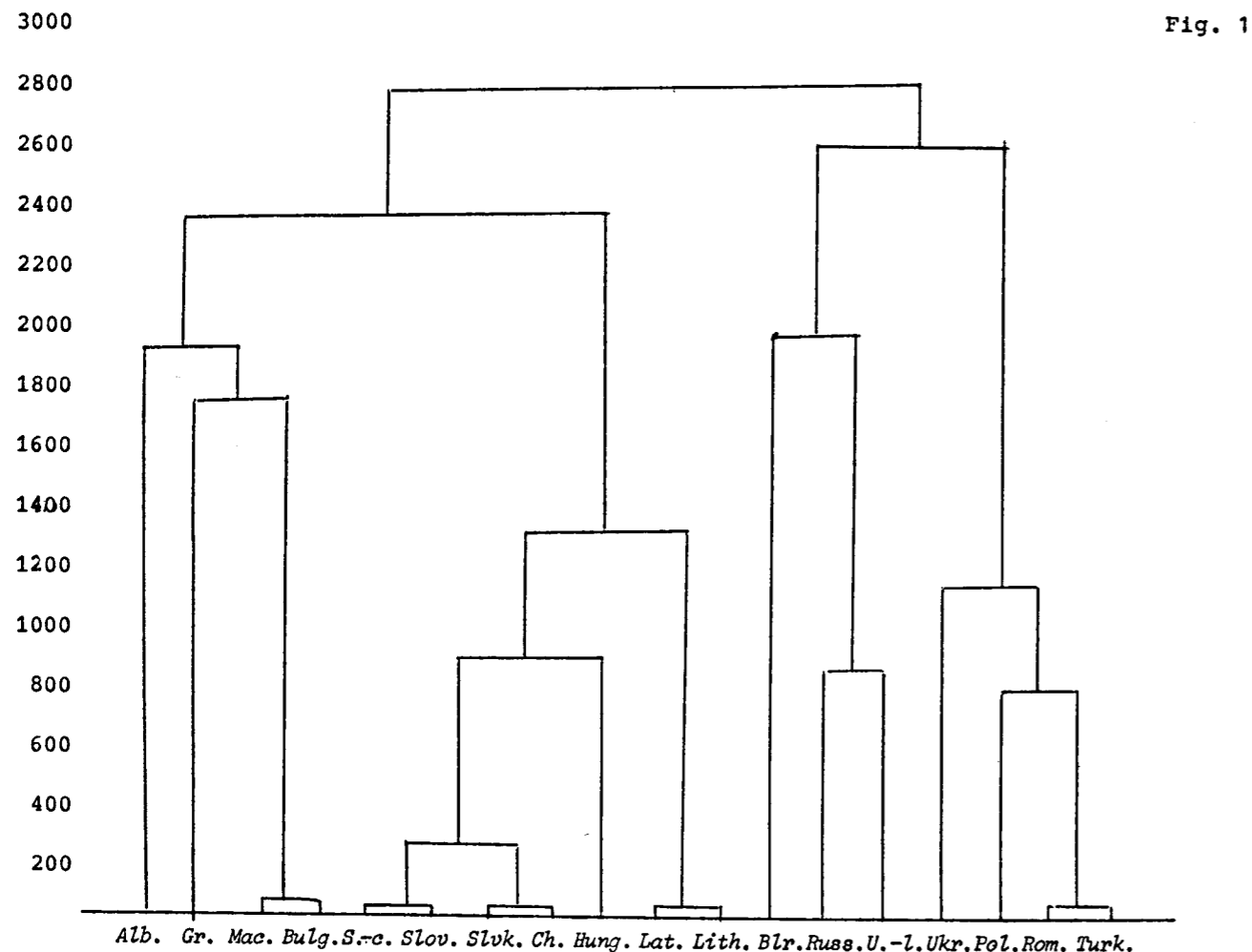


Fig. 1

ranian (Polish (Romanian Turkish)) and (Russian-Upper Lusatian) Belorussian.

It should be noted that while the clustered languages demonstrate a high degree of similarity, none of them is characterized by one or several of specific features. One and the same feature could appear as almost specific or almost non-specific depending on the concrete constellation of features in the given language and its neighbours. Thus the opposition of palatalized vs. non-palatalized consonants was noted to be specific not only to the languages of the first cluster, but also to mark the phonological systems of the languages belonging to the 2nd group (Lithuanian and Bulgarian). This, however, did not interfere with their clustering correspondingly with the Latvian language and the Macedonian language on the basis of whole sets of features.

The dendrogram showing the clustering of the phonological systems depending on the degree of their similarity can be represented on the map as a structure of the corresponding linguistic territory. Thus the relative similarity of languages can be shown as an altitude of the corresponding territories and marked by different hatching. In the

present case we have established two areas of similar phonological system or two areal types: Balto-Carpathian-Balkan area represented by two territories that are separated by a compact Central area of phonological systems including languages from Upper Lusatian to Russian.

The features of syntagmatic phonology as well as features of other language levels should be likewise included in the data matrix. The interpretation of areal types thus obtained will yield the best results when an increased degree of approximation of the global characteristics of the language does not change the output of the algorithm.

CONCLUSIONS

1. From the point of view of paradigmatic phonology the Balto-Balkan region is a single relatively connected linguistic area with two language areas: a marginal Balto-Balkan area proper and a central compact area, separating the two territories of the Balto-Balkan language area.

2. A quantitative analysis of the degree of similarity characterizing the phonological systems of a given area helps to establish the systems of maximum similarity, the hierarchy of their clustering and the optimum of the borderlines between the obtained areal types.

3. The areal types obtained as a result of the cluster analysis described in the present paper show a high degree of overall similarity, while they do not necessarily contain specific features. As specific features characterizing areal types tend to be absent, the procedure of establishing the types is believed to be rather objective.

REFERENCES

- /1/ L.Campbell, T.Kaufman and T.Smith-Stark. Meso-America as a linguistic area. - *Language*, v. 62 N 3 (1986), 530-570.
 /2/ M.B.Emeneau. *Language and linguistic area*. Stanford, 1980.

/3/ H.Haarmann. *Aspekte der Arealtypologie. Die Problematik der europäischen Sprachbünde*. Tübingen, 1976.

/4/ R.Jakobson. К характеристике евразийского языкового союза. Paris, 1931. Reprint: R.Jakobson (1971). *Selected Writings. I. Phonological Studies*. The Hague-Paris, 144-201.

/5/ R.Jakobson. Über die phonologischen Sprachbünde. - *TCLP, IV* (1931), 234-240. Reprint: R.Jakobson (1971), 137-143.

/6/ R.Kattein. Zur Definition des Begriffs Sprachbund. - *Sprachwissenschaft*, B. 11 H. 3 (1986), 276-287.

/7/ C.Masica. *Defining a linguistic area: South Asia*. Chicago, 1976.

/8/ H.Späth. *Cluster-Analysen-Algorithmen zur Objektklassifizierung und Datenreduktion*. München Wien 1975.

/9/ N.S.Trubetzkoy. *Phonologie und Sprachgeographie*. - *TCLP, IV* (1931), 228-234.