

On the Nostratic Language Family Hypothesis

The present subject, one of the basic characteristics of which is that it probes into an almost endless and distant past, brings to mind the following words of Thomas Mann: „... mit unserer Forschungsangelegentlichkeit treibt das Unerforschliche eine Art von foppenden Spiel: es bietet ihr Scheinhalte und Wegesziele, hinter denen, wenn sie erreicht sind, neue Vergangenheitsstrecken sich auftun...”

And it is an actual fact! An investigation of the ancient Uralic Age characteristics, then the search for pre-Uralic contacts, that is, the creation of an Indo-Uralic, Uralic-Altaiic, and Uralic-Jukagir theory, as well as the creation of theories involving other language families, seem to be nothing more than transitory stages in the mirror of the Nostratic language family theory.

The designation „Nostratic” was first used by HOLGER PEDERSEN. In 1903 he enriched comparative linguistics with an interesting idea. According to this it is conceivable that more language families form one single large language community. To this linguistic community, which was called Nostratic by him, belong the following languages or language families: Uralic, Indo-European, Altaiic, Hamito-Semitic, Basque, the Caucasian languages, the dead languages of Asia Minor and its neighboring territories, as well as Eskimo, Jukagir, and the dead languages of the Bering Straits (Türkische Lautgesetze ZDMG 57: 535—61).

It was only in the sixties, after a rather long silence, that any attempt was made towards producing some sort of scientific foundation for such a type of macro-language family theory, if I may put it so. The long pause was justified; moreover, it can even be said that it was worthy of praise, because without a more thorough knowledge concerning the members of the macro-language family, the notion of a Nostratic language family would have become degraded to only an irresponsibly playful idea. Then, however, when V. M. ILLICH-SVITYCH and A. V. DOLGOPOLSKIJ settled down to work in about the middle of the sixties—to use the words of B. COLLINDER—„das fast unabsehbare nostratische Feld zu pflügen” (Hat des Uralische Verwandte? 171), it could already be argued that given the knowledge of the newer results, an investigation of a language relation in a wider perspective appears to be much less likely to be labeled illusory.

There are, however, viewpoints on the basis of which the question may well be raised as to whether it wasn't a premature undertaking to collate several language families. Above all, the theory which is at issue causes one to think in the direction—as W. VEENKER suggests (UAJb. 41: 367—8)—that

ILLICH-SVITYCH and DOLGOPOLSKIJ were not entirely of the same opinion as to which language families belong the Nostratic language community (DOLGOPOLSKIJ earlier called it Borealis and then Europo-Siberian). According to ILLICH-SVITYCH, Hamito-Semitic, Kartvelian, Indo-European, Uralic, Dravidian, and the Altaic language families belong to the Nostratic language community. DOLGOPOLSKIJ at first included Chukchi-Kamchadal, while he later left it out, but he left Dravidian completely out of consideration. In my opinion, such exclusions and inclusions respectively are quite arbitrary at this initial stage of research. For what is actually the question here? The question is that once as far back as Paleolithic times the languages of the Eurasian continent could have once been members of a related linguistic community in a certain territory. From the followers of the hypothesis we could expect as a logical starting-point the a-priori argument that there was such a relationship of such an early age between all the languages, or rather, all the language families of Eurasia. Further research will decide, in the course of bringing to light the distance or proximity of the relationships between the language families, which language families should be excluded from the so-called Nostratic linguistic community.

Raising some doubt regarding the hypothesis is also the circumstance that the internal relationships of some of the language families have by no means made clear that as far as the Nostratic hypothesis is concerned, there would not be quite enough to be done in micro-comparative linguistics. As an example, B. COLLINDER mentions the internal reconstruction problems of the Hamitic languages (*Antiquitates Indogermanicae* 1974: 370), but we may well suppose that within the Altaic language family, whose mere existence is furthermore subject to debate, the problem of comparing and genetically relating is not of an insignificant nature. Moreover, the relationships between the two major branches of the Uralic language family, Finno-Ugric and Samoyedic, is subject to refinement. Concerning this, we may cite GYULA DÉCSY (*UJb.* 41: 375—80), according to whom Finno-Ugric and Samoyedic should for practical purposes be separated. He proposes that only through this division can Samoyedic linguistics be created as an independent discipline which is without doubt necessary for hypothesizing a Samoyedic proto-language, just as through the help of the same for a more perfect reconstruction of the Uralic proto-language.

After all of this, we can return to the question of whether it is not too premature to compare more language families or even perhaps two language families. I believe and I wish to emphasize that it is not too early but just the right time. But why? Here I am resorting to a very simple commonplace. In any case it would be to the advantage of a Uralic or Indo-European linguist if he could be fully aware of how far his research in his own specialization can be assimilated in future investigations of a wider scope. The deductive utilization of a seemingly well theory, which appears to be well-established and correct, can, as compared to this, advance inquires of an inductive nature.

DOLGOPOLSKIJ regards it possible that the Nostratic theory should already be utilized in this way. It has as its starting point the fact that the effectiveness of language comparison can be increased if many languages are compared all at once. The possibility of two morphemes corresponding by chance, as generally known, is more probable if we compare only two languages than if we do the same with more languages. On the basis of this, the method

of comparing all of these language families collectively can strengthen the correspondences between the Uralic and Altaic, the Uralic and Indo-European, the Indo-European and Hamito-Semitic, etc. language families, and it can also significantly reduce the number of chance correspondences.

DOLGOPOLSKIJ clearly saw that in this collective comparison the traditional methods of comparative linguistics could not be applied. He therefore called upon the help of strict laws based upon the exact statistical data of the theory of probability. First of all he chose from among a group of meanings those whose designations in the majority of languages show the least amount of instability. He was able to filter out fifteen such meanings. After that he calculated what was the probability of the chance correspondences of the morphemes carrying these meanings. It turned out that among the fifteen morpheme types, thirteen could not be explained as chance correspondences! It cannot even be a question of borrowing. According to the statistical investigation, out of two hundred European, Asian, and African languages, words having such meanings are borrowed from one language to another only in the rarest of cases. DOLGOPOLSKIJ came finally to the conclusion that the Indo-European, Hamito-Semitic, Uralic and Altaic, Chukchi and Kartvelian language families were probably in an ancient genetic relationship with one another; since in finding the origin of the word material under investigation we have to exclude the possibility of chance correspondences and borrowings, we have only the possibility of a genetic relationship to consider. Says DOLGOPOLSKIJ! (VJa. 1964/2: 53—63)

But I wonder whether we can actually think of only that. ANDRÁS RÓNA-TAS, in investigating the problems concerning the relationships between the Altaic languages, lists all of those factors which can bring about correspondences or agreements between languages. After considering chance and typological correspondences as well as correspondences based upon convergence, he returns to those he calls correspondences resulting from historical relations. Here he distinguishes four types: genetic relationship, areal contacts, mutual influence of a third language, as well as certain historical contacts which are manifest in borrowings (VJa. 1974/2:31). DOLGOPOLSKIJ then, after excluding the chance possibility as well as the possibility of borrowing, exclusively regards a genetic relationship as a possible explanation for the correspondences, and does not regard the correspondences based upon areal contacts.

PÉTER HAJDÚ, in a paper given at the Finno-Ugric Congress held in Tallinn (Uralistik und areale Sprachforschung), taking the notion of Sprachbund as his starting point, treats in some detail the possibilities that stem from a state which is characteristic of a group of neighboring languages in a relatively uniform or at least contiguous area and which we can recognize through the help of a group of linguistic similarities. HAJDÚ establishes that the areal contact between languages existing together for a long time and between non-related languages can in a surprising manner bring about corporeal similarities in the grammatical morpheme structure. As an example, the instrumental suffixes of Selkup and Ket (Selkup *-sä*, Ket *-as*, *-äs*) can be mentioned. Although the Selkup suffix presumably comes from an *-s* adjectival derivational suffix, there is also unexpectedly in Ket an *-s*, *-sa* adjectival derivational suffix. The Eastern Ostyak and Chukchi instrumental suffix likewise shows a striking resemblance between one another. After this,

HAJDÚ poses the question: can these examples mean that the areal language relationship can bring about not only typological correspondences but genetic corporeal correspondences as well? The answer: „Vielleicht dürfte man diese Frage bejahend beantworten.“

According to HAJDÚ, through this areal contact, that is actually an application of a more modern version of BUBRICH's contact theory, we can perhaps find a suitable explanation for the pre-Uralic genetic-like correspondences. If that is to say the areal relations during the historical period could have brought about genetic-like correspondences, then we must with this phase consider with increased effort the Paleolithic Age when human society from an economic and cultural point of view was rather unified; that is, the language creation role of the areal contacts must have also been stronger.

Of course, if we also accept areal contact as an explanatory principle, we cannot leave out of consideration the possibility that areal contact could just as well have come about between related languages some time after their separation and could form a system of secondary correspondences. But one of the reasons for the Altaic problem is that the Altaic nations after the 13th Century were in contact with one another. The existence of a contact of this kind must be recognized as being in favor of their genetic relationship. This contact being a fact, we can suppose that there had been contacts between them earlier as well so that the correspondences thought to be ancient are perhaps only secondary. The latter assumption, however, does not always exclude the possible ancient correspondences, that is, the possibility that the Altaic language family had really existed at one time.

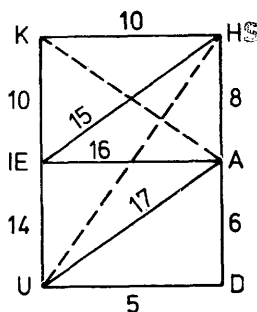
The correspondences in that case are of three types: 1. ancient inheritance, 2. secondary correspondences stemming from the later areal contact of languages probably related genetically, 3. correspondences stemming from the areal contact of non-related languages. These three types of explanatory principles are similarly valid if based upon the Nostratic area. RÓNA-TAS, in working towards a solution of the Altaic question, proposes a detailed examination of the historical relationships of the periods from the beginning of the II millenium B.C. to the beginning of the II millenium A.D. (op. cit. 45). This test could actually bring about a solution, because the relatively recent past must be investigated. But I wonder whether we can think of a similar solution, if we must make a decision regarding the language contacts during the Paleolithic and Mezolithic Ages. Even if we can do so, our method must be different in the case of such a distant past.

In the following I will introduce the method that according to me might be of perhaps more help to us towards finding a solution.

On the basis of ILLICH-SVITYCH's 245 Nostratic, called so by him, lexical items (Опыт сравнения ностратических языков. Moscow), I have made some statistical analyses. I mention in advance that the specialists dealing with the various language families will without doubt have to check the reliability of the etymologies, in particular when we have at our disposal Nostratic word material of more than 600 vocabulary items as estimated by DOLGOPOLSKIJ. At the time he included the Uralic equivalents, ILLICH-SVITYCH relied mostly upon B. COLLINDER's dictionary as well as the first three volumes of the Finnish etymological dictionary. He however used word agreements not occurring in these works from older and possibly obsolete findings. In Uralic linguistics etymological research is continually carried on. Being in possession

of the complete material of the SKES, as well as the MSzFE and the Uralic Etymological Dictionary, a thorough screening of the Nostratic vocabulary items from a Uralic point of view will be possible. With all of this I wish to express that now as I am utilizing the 245 etymologies that already exist, I am doing it with the knowledge that I can only carry out a methodological experiment on them in so much as I am not and cannot be convinced of their reliability.

In the language families I selected related variations appearing in the individual etymologies. What was particularly noticeable at first glance was that disyllabic word agreements, which contained correspondences from only two language families, were in the great majority (120). 78 etymologies connect three language families, 32 connect four, 10 connect five, and only 5 connect all six. It turns out from the disyllabic related types that on the basis of these, a close connection can be assumed first between U-IE-A, second between IE-A-HS, third between IE-K-HS and fourth between U-A-D. We can demonstrate this as follows in a diagram :

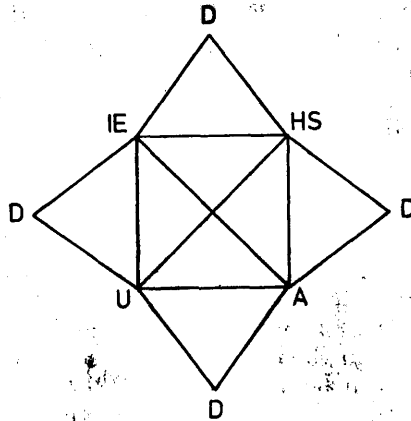


(The numbers signify the numerical ratio of the individual related variations.

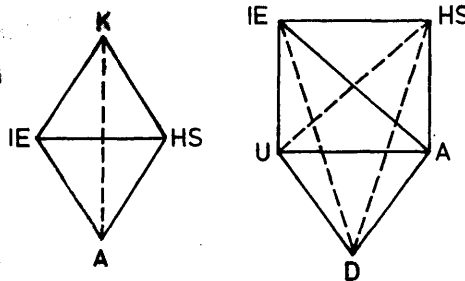
Abbreviations :

- HS — Hamito-Semitic
- K — Kartvelian
- IE — Indo-European
- U — Uralic
- D — Dravidian
- A — Altaic)

Considering these connections, I found in insignificantly small numbers etymologies containing HS-D (1), K-U (3), K-D (1) and IE-D (3) pairs. Agreements between A and K (6) as well as between HS and U (5) show a relatively stronger connection, which is nevertheless indicated by only a broken line. Through the help of the diagram, I wanted to demonstrate that outside of possibly the group IE-U-A-HS, which constitutes a more solid block, occurs K and D; but the last two cannot hardly be connected with one another. Etymologies, such as these, which contain K or D reflexes — this can also be seen from the numbers — occur, on the other hand, in a smaller quantity than those in which the other language families play a role. In addition, K appears in the 245 etymologies only 65 times, D 92 times; but on the other hand U appears 118 times, HS 125 times, A 138 times, and IE 144 times. This would support the assumption, formed on the basis of the preceding, that particularly K and perhaps D are connected to the others by possibly weak threads. After all of this and while thinking of the fortuitousness of the disyllabic comparisons mentioned repeatedly by ДОЛГОПОСКИУ, I examined the tri-syllabic word agreements. I took into consideration only those connection variations whose numerical ratio was not smaller than 5. I naturally set this limit arbitrarily. The largest number found was 9. My tests resulted in the following diagram :



From the disyllabic word agreements it was disclosed that D exhibits a close relationship with U and A. Also in the majority of threefold etymologies, as can be seen in the diagram, the D reflex appears only if the etymologies have at the same time an A or a U reflex. We however have a type of relationship whose reflexes are IE-D-HS. On the basis of the disyllabic etymologies, D cannot be connected with either one of the other two language families. We must now, however, make an alteration in the picture. Of course, we cannot in any case decide — alongside such small order of magnitude — from among the two language families, which one D is closer to. We can barely establish that D might have been in contact with the HS-IE group. After all of this, we can determine the relationships between the language families in two diagrams :



The question of whether the relationships shown in the two diagrams reflect a temporal level of difference or only a geographical localization during a certain period may now be logically presented. It can also of course be imagined that they reflect both at the same time. In other words, the position of K, which excludes both U and D, can be explained on the one hand in that K sooner or later joined the IE-HS-A stock-group; but on the other hand it can be explained that geographically the previously mentioned stock-group occupied, at one time, a central place; while D-U, or rather K, settled down in peripheral areas further apart from one another.

I repeatedly emphasize that I did not carry out all of these calculations with the hope that some kind of definite conclusion would be disclosed. I was on the one hand aware that ILLICH-SVITYCH's word material was small and on the other, as I have already mentioned, that a thorough re-examination would be required. My goal was only that the same kind of approximating method could, to a greater extent, be proposed when a larger and more reliable collection of words would be at our disposal.

What the calculations carried out in the present material can tell us is only that we can observe in the relationships of the six Nostratic language families certain temporal or spatial, or rather temporal and spatial shifts. This point is sufficient to arouse our curiosity as to how the Nostratic area was divided and what kind of events took place there.

In conclusion I can say that the Nostratic theory is an interesting hypothesis. Instead of the concept „language family”, however, we should in the present case be satisfied with the concept „area”. Future research would achieve significant results if through the help of the calculus of probability, or rather through the help of certain statistical methods similar to those shown above, or other non-traditional means of paleo-linguistics, the movements that took place in the Nostratic area and the genetic type of relations that came into being there would, to a certain degree, become clear. If it could be shown which language families were in a more distant relationship with one another and which were in a closer relationship, then perhaps the internal language teachings of the language families in closer contact and which correlate with one another could also be utilized. Attempts in such a direction have already taken place (ILLICH-SVITYCH: *Etim.* 1966: 304 pp.; VFUJa. 4 (1967): 95 pp.; DOLGOPOLSKIJ: *Etim.* 1964: 259 pp., 1968: 237 pp., 1970: 356 pp. etc.), but because of their prematurity, their results are not convincing. We can await help of a different kind from a picture drawn more clearly of the Nostratic area. We can perhaps more exactly designate the individual language families, ancient areas of their diffusion, and also be able to determine the possible direction of their migrations.

All of this, however, belongs to future tasks. Researchers involved with the individual language families, and thus Uralic linguists, must still wait for the Nostratic hypothesis to help them in their respective areas and in the research they are carrying out. I will close with BO WICKMAN's conviction according to which in such research, that is in an investigation of the relationships between language families, temperament, or rather optimism and scepticism, will always play a large role (UAJb. 41:310).

ÉVA KORENCHY

