

## METHODOLOGY OF LONG-RANGE COMPARISON

### 1. Is there any need for long-range comparison?

The question is not uncommon. Some people (even prominent specialists in particular language families) think that they can learn nothing from the outside world and are quite content with what is available.

However, there are two main reasons which, in my opinion, justify the existence of this branch of linguistics:

a) We need to have some classification of the world's languages. The traditional classification (which lists several hundred linguistic families) is a perpetual challenge for comparative linguists. Are there any genetic links between at least some of the world's major linguistic families? If not, how did this extremely strange situation arise? As far as I know, nothing of the kind exists in other disciplines dealing with *Homo Sapiens*, e. g., in the biological sciences.

b) Comparative linguistics is at this time one of the very few branches of science which can supply information about the preliterate history of man. There have been several attempts to combine linguistic data with archeological and genetic evidence, some of which have given very promising results. Surely, if we could extend linguistic evidence to dates earlier than the 4th-5th millennia B.C., this could be very useful for the whole field of human history.

### 2. Comparison and reconstruction

The method used by the best long-range comparative linguists was not invented especially for this kind of research. It is the same traditional comparative method which has been used in linguistics for nearly two centuries.

There exists, however, a difference - not a methodological, but rather a strategic one: traditional comparative linguistics relies basically on written and spoken languages, whereas the basic material for long-range comparison is reconstructions.

Of course, the idea of reconstruction is a legitimate part of the traditional comparative method. However, in very many cases, when languages are closely related, genetic classification and different kinds of comparative research are quite possible without any reconstruction. One does not really need a reconstruction to arrive at the idea that, e.g., Slavic languages are related to each other genetically. No reconstruction was needed in the initial stages, when the idea of the Indo-European family was born. In these cases, reconstruction may either be absent altogether (there still exists a large number of commonly accepted linguistic families with no available proto-language reconstruction), or it may be there just as a means of explaining the similarities and correspondences between languages.

For long-range comparison, reconstruction is absolutely vital. One often hears from critically-minded people that if two languages exist separately for a time span of more than 5-6 thousand years, they may lose all traces of similarity and any comparison becomes impossible. They forget, however, that one may deal not with modern languages, but with reconstructed intermediate stages which - for all practical and theoretical reasons - must have been closer to each other than their modern descendants. A few examples:

Modern Chinese numerals *èr* 'two', *wǔ* 'five' and *bā* 'eight' are totally unlike Modern Burmese numerals *ne*, *ṇa* and *hrać*. However, if we compare reconstructed Old Chinese *\*nij-s* 'two', *\*ṇā?* 'five' and *\*prēt* 'eight' with reconstructed Tibeto-Burman *\*g-nis* 'two', *\*ṇa?* 'five' and *\*p-riat* 'eight', we get a fairly good idea

of the languages' relationship.

Russian *слышать* 'to hear' (or Old Indian *çru-* id., or English *loud*) are certainly not similar to Korean *kwi* 'ear', or Turkish *kulak* id., or Evenki *ūl-ta-* 'to be heard, resound'. But the reconstructed Proto-Indo-European *\*kleu-* 'to hear' is much closer phonetically to reconstructed Proto-Altaic *\*k`ujla* 'ear, to hear'.

Chechen *dog* 'heart' (or Agul *jurk* id., Circassian *g<sup>w</sup>ə* 'heart, breast') do not resemble Chinese *yi* 'breast' (or Burmese *raL* id.). If we know, however, that the Caucasian forms go back to Proto-North-Caucasian *\*jerkwi*, and the Sino-Tibetan forms - to Proto-Sino-Tibetan *\*ʔrək / \*ʔrəL*, the comparison becomes much more plausible.

### 3. Statistical methods

Statistics is not widely used in traditional comparative linguistics. However, it is an important tool for long-range comparison for several reasons:

a) Statistical methods are good for verifying hypotheses about linguistic relationship. Since in many cases long-range genetic links are not superficially obvious, statistical testing is useful for distinguishing between genuine relationships and look-alikes or massive borrowings.

b) Subgrouping in comparative linguistics is usually done using the criterion of shared innovations. In practice, this criterion works best on morphological data. Since morphological reconstruction of the macrofamilies is basically in an initial stage, there is an urgent need for some substitute.

It can be shown that the lexicostatistical method of classifying languages can be applied both to "short-range" and long-range comparison. Since the results obtained in the classification of closely related languages generally correlate rather well with traditional subgroupings, one can assume that the results of long-range classification are also plausible.

c) The application of statistic methods to linguistic dating (also known as glottochronology) has been widely criticized.

While doing "short-range" comparison, one can generally dismiss it and guess the approximate dates of divergence using other evidence (oldest written records, sometimes archeological data). It is, however, the only method which can be applied to distant relationships and, therefore, seems to be worth re-examining.

#### **4. Computer methods.**

A researcher dealing with long-range comparison has to process a huge amount of linguistic data, which grows exponentially once any new linguistic family is being added. Modern computer technology allows one to deal with this flow of data more efficiently, although there still are very few computer applications designed for comparative linguistics. It is possible to use computers for storing large comparative databases, for processing data of related languages (even for establishing phonetic correspondences), and, of course, for performing all kinds of statistical calculations.