PERSPECTIVES OF COMPARATIVE STUDIES ON THE MANDE AND WEST ATLANTIC LANGUAGE GROUPS:
AN APPROACH TO THE QUANTITATIVE COMPARATIVE LINGUISTICS[1]

Konstantin Pozdnyakov

Statistical methods in comparative linguistics have been widely applied in only one peripheral field, namely the glotto-chronology and lexico-statistics which are intended to establish absolute or relative chronology of language divergence. In the meantime, statistics can be applied in all the principal fields of comparative linguistics. In certain relations, statistical data become the decisive factor. This paper deals with principal branches of quantitative comparative linguistics dealt with during my work in the field of Mande and West Atlantic historical grammar.

The principal starting point for the following reasonings is very simple: any qualitative divergence between cognate languages evoked by any innovation process results in a respective quantitative divergence. For example, palatalization of dental stops before front vowels in one of a couple of related languages results automatically in an increased frequency of palatals and a decline in the frequency of dentals in the vocabulary. As a

[1] This electronic version of Konstantin Pozdnyakov's article has undergone some extra editing, so it has some differences from the paper version published in Mandenkan 22. However, the modifications concern only formulations of ideas but not their content. – *Editors.*

result, the statistical distribution of phonemes reflects all basic divergence processes in cognate languages (with few exceptions). Therefore, any case of quantitative difference in any pair of related languages is to be interpreted by comparative historical linguistics.

Taking this thesis as a starting point, we obtain two alternative sets of facts to explain a phenomena, instead of only one. Any contradiction between qualitative and quantitative evidence should be considered as an indication of a shorcoming in the analysis. These contradictions are of great importance in those frequent situations, when traditional methods of comparative linguistics do not provide criteria for the choice between two or more reconstructions which seem to be equally probable. For the solution of some problems of linguistic reconstruction, there are no other methods than statistical ones. They manifest their advantages especially when we deal with «young» branches of comparative linguistics handling dozens of poorly described languages.

In this paper, I am going to try to summirize the results achieved during a ten-year project in the field of historical comparative interpretation of phoneme frequency divergences in the vocabularies of cognate languages. It is possible today to pick out at least seven basic problems of comparative linguistics, in which the use of statistical data yields positive results. These are:

1. Discovery of regular phonemic correspondences in related languages at the most preliminary stage of analysis preceding compiling of a comparative vocabulary.
2. Discovery of archaic fossilized morphemes in lexical stems, such as archaic noun class markers regarded today as integral parts of roots.

3. Internal reconstruction of the phonemic proto-system of an isolated language when «classical» comparative methods cannot be applied. Internal reconstruction is also very efficient when dealing with closely related languages.

4. Search for the genesis of an isolated language.

5. Verification of results of a comparative analysis, including evaluation of each segment of a comparative etymological dictionary and detection of incorrect segments; comparative evaluation of the level of correctness of different etymological dictionaries of one proto-language.

6. Detection of the most archaic languages in a group of related languages.

7. Genetic classification based on statistical analysis of etymological data.

Regular phonemic correspondences. The principal quantitative characteristic of a phoneme is its frequency, in a text and in the vocabulary. What is important for comparative studies it is the frequency in the vocabulary. How can we explain a considerable difference between frequencies of a phoneme in related languages? For example, what can be the significance of the fact that 12% of words in a Bisa dictionary begin with phoneme, *h*- and in a closely related Mande language, Lebir, initial *h*- is absent (frequency 0%)? Two answers are possible:

1) In Lebir, **h*- yeilded another phoneme, for example **h* > *k*. Therefore, the frequency of *h*- dropped to 0%, and the frequency of *k*- rose;

2) *h*- in Bisa reflects another proto-phoneme, for example **f* > *h*. Therefore, the frequency of *h*- rose, and that of *f*- dropped. In both cases

we have a phonemic change in one of the cognate languages which results in a regular correspondence between two different phonemes. Hence the inverse conclusion can be proposed: if there is any regular correspondence between different phonemes, such as *f- : h-*, the frequency of *f-* in the first language will be above the average, and the frequency of *h-* will be below average.

The following example makes it clear. Suppose a proto-language had 42 stems with **f-* and 10 stems with **h-*. Assume the first daughter language underwent very little change, retaining 40 stems *f-* and 10 *f-*. In the second language **f- > h-*, resulting in 52 stems with *h-* (42 + 10). Making our example more natural, suppose the second language to be heavily influenced by English or French and to have 42 stems substituted by loan words. So the second language retains 10 stems with *h-* as well as the first one.

Now we can try to find out the correspondence *f-* (lang.1) : *h-* (lang.2) which has appeared due to the phonemic change **f- > h* in the second language:

		f	h	∑
Stem number:	language 1	40 stems	10 stems	50 stems
	language 2	0 stems	10 stems	10 stems
Stem frequency:	language 1	80%	20%	
	language 2	0%	100%	

Frequency deviation:

language 1 + 40% (80% - 40%) - 40% (20% - 60%)

language 2 - 40% (0% - 40%) + 40% (100% - 60%)

Deviations from the average frequency of *f-* and *h-* allow to find out the correspondence *f-* (lang. 1) : *h-* (lang. 2). It is possible to show that a complicating of the conditions of the problem yields the same result: any correspondence between two different phonemes brings forth two pairs of frequency deviations from the average with opposite distribution of plus and minus signs.

We can take three real examples from three principal branches of the Mande group. Let us look at the table of frequncy deviations from the average for Kpelle and Mende (South-Western group). In two rows in the bottom, the deviations in per cents are given:

Phoneme (C-)	mb	b	nd	l	nj	y	ng	w	kp	kpw	s	h
Average (%)	2	5	4	5	1	6	5	4	6	3	2	8
Kpelle	-2	+2	-4	+4	-1	+3	-5	+2	-3	+3	-2	+2
Mende	+2	-2	+4	-4	+1	-3	+5	-2	+3	-3	+2	-2

Frequencies of other consonants (stops, nasals and *f-*, not included in the chart) are almost equal in the two languages. Taking into account regular consonant alternations in

both languages, it is not difficult to postulate regular correspondences as follows:

Kpelle	b	l	w	y	y	kpw	h	h
Mende	mb/b	nd/l	ng/w	ng/y	nj/y	kp	s	h

Apart from this, the elementary statistics shows correspondences *m - m, n - n, ny - ny, p - p, f - hw, t - t, k - k*. Therefore, without compairing any pair of real words we get nearly all the regular correspondences which could be found out through comparative linguistic analysis.

Let us compare now the frequencies of initial phonemes in two languages of the Eastern Mande group, Guro and Yaure. In both languages, the frequencies of sonorants (nasal as well as oral) are the same. On the contrary, the frequencies of all non-sonorant consonants differ:

Phoneme (C-)	b	p	v	f	d	t	z	s	g	k	gy	ky
Average (%)	10	8	4	9	5	10	3	11	4	7	2	3
Guro	+3	-3	+2	-2	+3	-4	+3	-2	+2	-3	+2	0
Yaure	-3	+3	-2	+2	-3	+4	-3	+2	-2	+3	-2	0

It is not necessary to compile a comparative dictionary in order to discover regular correspondences between these two languages. It is quite evident that initial voiced consonants of Guro correspond to their voiceless counterparts in Yaure.

At last, consider an example from Nothern Mande. It will be more complicated with three languages involved, i.e. Bozo, Soninke and Susu. The most important frequencies deviations are given below:

Phoneme (C-)	p	f	k	x	s	j	y	gb	g
Average (%)	4	6	14	7	14	3	5	1	5
Bozo	+5	-4	+5	-6	+4	0	-1	-1	-1
Soninke	-3	+2	-5	+3	-1	+4	-1	-1	+2
Susu	-3	+2	0	+3	-3	-3	+2	+1	-2

The table shows the same distribution of deviations for voiceless labials and velars: the stops in Bozo *(p-, k-)* correspond to the fricatives in Soninke and Susu *(f-, x-)*. The distribution shows also a high probability of the following correspondences: Bozo *s-* : Soninke *j-* : Susu *y-* and the correspondence of Susu *gb-* : Soninke *g-* (in Bozo it might be *k-*).

The method illustrated above has its limits. Theoretically, we may single out situations when it does not work. However, some experience obtained during research work in Mande and West Atlantic comparative linguistics persuades that it is possible to find out large majority of phonemic correspondences before any comparison of real words.

Revealing of rudimentary morphemes and of morpho-phonemic alternations in a proto-language.

Reconstruction of West Atlantic languages seems to be much more

complicated than that of Mande. The West Atlantic phonemic reconstruction could be regarded as exemplar for its inaccessibility. First, some consonants which are included in today's noun roots, represent archaic prefixes of noun classes merged with lexical stems. Before these fossilized class markers are detected, a reliable comparative study of initial consonants is impossible. Besides, there is a powerful morpho-phonemic fator that influences initial consonants: historical phonemic change has been often determined by synchronous consonant alternations in the proto-language. Evidently, phonemic frequency distribution in present-day languages reflects both these factors. It means that statistical data gives a real support for morphemic and morpho-phonemic reconstructions.

Let us compare phoneme frequencies in nouns and in verbs. If no specific factor is involved, we can expect them to be equal. In the Mande languages, this assumption is true for the majority of initial consonants. However, there are some interesting exceptions. For example, in the Bamana nouns, frequency of pre-nasals is higher than in the verbs. It becomes clear if we take into account that in Bamana, nouns with initial prenasals retain archaic noun class prefix **N-*. Word-initial prenasals in nouns have a morphemic component, and it results in the frequencies distribution.

In the West Atlantic languages, frequencies of identical phonemes in nouns and verbs differ very often. Let us analyse a typical example.

In the Noon language (Cangin subgroup), the initial *p-* is attested in 12% of nouns and only 5% of verbs. Therefore,

we can conclude that many Noon nouns retain traces of the **pV-* prefix which results in rising of the frequency of initial *p-* in nouns. It is easy to confirm this conclusion by linguistic data.

There exists another way to detect archaic prefixes. Let us compare the frequencies of initial *p-* in two distantly related languages, namely Noon (Northern Atlantic branch) and Sherbro (Southern Atlantic branch, Mel group):

	noun	verb
Noon	12 %	5 %
Sherbro	7 %	9 %

In Noon, the frequency of *p-* in nouns exceeds that in Sherbro. In the meantime, the correlation of its frequencies in verbs is inverse. It is impossible to explain it by the existence of an ordinary phonetic correspondence. It seems more expedient to search an explanation in the field of morphemic processes and to suppose that nouns in Noon have traces of a prefix containing *p-*.

It should be remarked that both independent statistical procedures lead us to the same conclusion. Therefore, it must have some linguistic grounds behind it. Suffice it to give here only one example. It is known that the Sherbro language belongs to the Mel group, named by David Dalby according to the noun root *mel-* for «tongue», as opposed to *lem-/rem-* in the Northern branch of West

Atlantic. In Noon, the noun for «tongue» is *perem.* Therefore it is possible to single out an archaic prefix **pV-* in this noun: *pe-rem*, the more so as in many Atlantic languages the cognate nouns have the same class marker *pV-/fV-* which is attested in present-day noun class systems: Manjak (dial.) *p-remt*, Mankaɲ *p-diim,* Pepel *p-remte*, Diola (dial.) *fu-rim, hu-rim*, Jaad *pe-deema*, Biafada *bu-deema*, Nyun *bu-lemuc*, Limba *fi-liŋ*.

These two complimentary procedures allow to discover rudimentary class prefixes in the majority of the West Atlantic languages, such as *b-* (Ndut, Kisi, Sherbro, Temne, Limba), *k-* (Noon, Mankañ ?, Kisi, Sherbro, Gola ?), *p-, f-* (Ndut, Noon, Kisi), *r-* (Gola, Limba), *j-, ll-*(Diola), *l-* (Mankañ ?), *m-* (Noon), *d-, g-, gb-* (Gola ?) and some others. There is no space here to examine these questions in detail. Some problems of noun class reconstruction in the West Atlantic and the Mel languages are dealt with in my fothcoming book.

It is important to note that a situation when frequency of a phoneme in nouns exceeds its frequency in verbs does not always reflect the retaining of an archaic morpheme. Let us take a look at another fragment of distribution of frequencies in Noon:

	mb-	*nd-*	*nj-*	*ng-*	*w-*	*l-*	*y-*
Noun (%)	5	4	2	4	2	8	2
Verb (%)	1	2	2	1	5	12	3

We can notice systematic characteristics in both series of initial phonemes: the summarized frequency of prenasals

in nouns (15%) is higher than in verbs (6%). On the contrary, in the series of sonorants we have 12% for nouns and 20% for verbs. Such frequency distribution points out to a diachronical morpho-phonemic process and allows to reconstruct some consonant alternations in the proto-language.

It is interesting for this purpose to confront frequency distributions in Fula and Serer. For the majority of consonants the distributions are the same in both languages. Nasals (i.e. non-alternative consonants) have the same frequency in nouns and and verbs in both Fula and Serer. Voiceless stops (except *t*-, the only non-alternative voiceless consonant) and voiced prenasals (the «strong» alternation grade) are more often met with in nouns than in verbs. On the contrary, the frequency of voiceless fricatives and oral sonorants (the «weak» alternative grade) is much higher in verbs than in nouns. In both Fula and Serer, the frequency of preglottals is also higher in verbs, although they are not included today in any alternative series. It shows with certainty that pre-glottals had a weak alternation grade in the proto-language. However, there is one series where the frequency distribution is always different in two languages: in Fula, the summarized frequency of voiced stops is much higher in nouns (26%) than in verbs (only 9%), in Serer we find an inverse situation (18% for nouns and 25% for verbs). According to a linguistic reconstruction (Pozdnyakov, 1987), Fula retains main features of the proto-language alternation system, while in Serer this proto-system has been changed: **nj* III/*j* II/*y* I > Serer *nj* III/*c* II/*j* I. In other words, voiced stops assumed in Serer a weak

alternation grade (grade I), and in the series of voiceless consonants, the same happened to the fricatives. Therefore, it is this process that has been reflected in the frequency distribution in Serer: frequency of voiced stops rose in verbs. Therefore, phonemic frequencies distribution reflects not only diachronical phonetic processeses, but also morpho-phonemic and morphemic ones. It must be regarded as an important support for the reconstruction.

Internal reconstruction. Some ways to outline the main features of the phonological proto-system without resorting to the traditional comparison of lexical items have been elaborated. It is important to note that some valid results in this field can be achieved even even before we compare either qualitative data or quantitative ones in different cognate languages. Very simple statistical methods allow to reconstruct a diachronical phonemic processes on the basis of only one language, without referring to comparative data. It would be naive to pretend that any research work in the field of comparative linguistics would be possible without comparison of languages. What we do compare is not statistical data of different languages, but phonemic frequencies in different positions in one single language.

Let us compare the frequencies of initial and intervocalic phonemes in one of the Mande languages, e.g. in Susu. Before one arrives to any sort of conclusions, one must be sure of reliability of the primary data. It is difficult to say a priori, whether a small glossary of less than 1000 items can adequately reflect essential quantitative correlations in the lexical stock of a

language. In order to examine this point we can take two different glossaries of one language compiled by different authors of two distant epochs. The frequency of each phoneme in the dictionary of J.-B. Raimbault published in 1885 (939 lexical items) and in that of M. Friedländer published a century later (1647 lexical stems) is given below (the first figure indicates the percentage of a phoneme in the initial position, the second one, after a slash, represents the percentage of the phoneme in the intervocalic position):

Dictionary of M.Friedländer					Dictionary of J.-B.Raimbault				
p	t	-	k	-	p	t	-	k	-
1/1	9/6		13/3		1/0	10/5		16/2	
f	-	s	x	-	f	-	s	x	-
8/2		11/5	10/6		9/3		12/3	11/6	
b	d	-	g	gb	b	d	-	g	gb
11/4	8/1		4/5	2/1	12/6	7/1		3/5	2/1
mp	nt	nc	nk	-	-	nt	ns	nk	-
0/1	0/2	0/2	0/3			0/2	0/2	0/5	
mb	nd	-	ng	-	mb	nd	-	ng	-
0/4	0/5		0/3		0/3	0/4		0/3	
m	n	ny	-	-	m	n	ny	-	-
4/4	3/6	3/5			5/4	5/7	2/3		
-	r	-	-	-	-	r	-	-	-
	0/14					0/17			

Comparison of frequencies in two different dictionaries reveals their striking resemblance. The correlation between initial and intervocalic frequencies of the consonants is identical

in the ancient and in the modern glossaries which differ in the amount of data represented. It is typical not only for these two lexical stocks and not only for the Susu language. My statistical data based on various dictionaries of Mande, West Atlantic and Bantu languages allows to insist on the fact that a glossary of 400 items can adequately reflect the real frequency distribution of phonemes in the vocabulary.

Let us compare the phoneme frequencies in initial and intervocalic positions as they are represented in one of the glossaries, namely that of M. Friedländer. First of all, we can notice that they differ very much, the distinctions being of systematic character. All oral consonants (except liquids *l, r*) are considerably more frequent in initial position than in intervocalic. All prenasals and probably nasals (at least, *n*) manifest the opposite distribution of frequencies. As for the grouping by local series, the dentals are opposite to all other consonants. There is the only set in which the summarized frequency of intervocalic consonants (48%) is higher than their summarized frequency in the initial position (24%). Now we can reasonably put a question: why is it so? The experience of comparative linguistics shows that alternative manifestations of a phoneme depend on the phonetic context: a difference of phonetic environment results in different of historical processes of phonetic changes for alternative manifestation. It is clear that phonemic frequencies in alternative positions reflect it automatically.

Therefore, we receive an efficient procedure to reconstruct some historical processes without recurring to external comparison. Afterwards, bearing in mind these quantitative indications, we can carry out a traditional comparative study of lexical stocks. It is important to stress the following point: conclusions drawn on the Susu data almost coincide with conclusions based on the statistical analysis of each other closely related language, i.e. Bozo and Soninke.

In order to compare these data, let us observe in more detail the frequencies of phonemes in initial and intervocalic positions (the first and the second figures accordingly):

Bozo				Soninke				
p	t	c	k	p	t	c	k	-
9/2	11/6	1/0	19/1	1/0	9/6	1/0	9/1	
f	-	s	x	f	-	s	x	-
1/2		17/7	1/0	8/3		13/4	9/7	
b	d	j	g	b	d	j	g	-
9/9	6/1	3/0	5/14	9/5	6/3	7/2	7/6	
-	l	y	w	-	l	y	w	-
	1/3	3/1	3/0		3/5	4/4	4/2	
mp	nt	nc	-	mp	nt	-	nk	nq
0/1	0/1	0/1		0/1	0/2		0/1	0/1
-	-	-	-	pp	tt	cc	kk	qq
				0/2	0/2	0/1	0/2	0/1
mp	nd	-	ng	mb	nd	nj	ng	-
0/3	0/5		0/5	0/3	0/2	0/1	0/3	
m	n	ny	ŋ	m	n	ny	ŋ	-
4/10	3/10	2/2	1/1	7/5	2/6	2/2	1/2	
-	r	-	-	-	r	-	-	-
	0/18				0/18			
-	-	-	-	-	ll	-	-	-
					0/4			

For each of the three languages we can consider that prenasals as well as geminates not found in the initial position should not be regarded as phonemes: the nasal component assimilated in localization series to a subsequent consonant can be attributed to an initial syllable as its final element. Therefore, if we consider the summarized frequencies of oral consonants and their prenasal variants (e.g. *b* + *mb*, *p* + *mp*, etc.), it will be noticed that the difference between initial and intervocalic consonants considerably decreases or even vanishes. It is clearly seen for *t, d, b* in particular. Being summed up, the frequencies will manifest predominance of intervocalic */g/* in the Susu language (*g* + *ng*: 4/8%), which is far more evident in Bozo (*g* + *ng*: 5/19%). Due to the Bozo data it is clear that the given correlation of */g/* is compensated by the correlation of */k/* (Bozo, *k* + *nk*: 18/1%). Therefore we can reconstruct the historical process **-k-* > *-g-* in the intervocalic position. For Bozo, the process **-p-* > *-b-* is also not to be excluded. If it is so, it is worth summing up the frequencies of all

occlusives. The results will be as follows:

	p, mp, pp, b, mb	t, nt, tt, d, nd	c, nc, cc, j, nj	k, nk, kk, g, ng	q, nq, qq	gb
Susu	13/10	17/14	0/2	17/14	-	2/1
Bozo	18/15	17/13	4/1	23/20	-	-
Soninke	10/11	15/15	8/4	16/13	0/2	-

It is evident that there is some equalization of phoneme frequencies in both positions. It correlates with two processes under reconstruction, namely: 1) sonorization of voiceless stops in intervocalic position; 2) prenasalization of intervocalic stops (or, on the contrary: disappearing of initial prenasals). Besides, a great exceeding of intervocalic dentals

and dental sonants in particular (total frequency for *l* + *r* + *n* is always impressive: Susu 7/34, Bozo 4/31, Soninke 5/28) indicates a historical process of dentalization in the intervocalic position.

Without further detailed discussion of this concrete reconstruction, let us consider some general points. First of all, it should be noted in this relation that the statistical data for such an analysis is far more interesting in West Atlantic than in Mande studies. In the West Atlantic languages, where words have final consonants, frequencies of phonemes in three different positions can be compared.

In a short paper, only a very simple way of using the phoneme frequencies in different positions for comparative linguistics can be displayed. Let us examine briefly the principle of another method which seems to be more productive. I have tried to outlook this way of internal reconstruction basing on the data for various West Atlantic and Mande languades. It has been testified afterwards by students and graduates of the Leningrad University for such languages as Songhai (VictorVoloshin), Mandinka (Olga Zavyalova), Hausa (Michael Kilks) and some others. Examples will be given to illustrate this principle.

In the Songhai dictionary by J.-M. Ducroz, according to

Victor Voloshin's calculations, there are 1663 lexical stems. Among them, 65 (i.e. 4%) begin with initial consonant */j/*. As for */i/* in the first syllable, it is attested in 14% of all the Songhai stems. Let us assume that there is no correlation between */j/* and */i/*, i.e. the presence of one phoneme does not provoke appearance of the other. In this case, it would be natural to expect that 0.6% of all the stems in the dictionary (0,04 x 0,14 = 0,006), i.e. 9 stems, will have *ji-* as the first syllable.

In the meantime, we find in the dictionary 32 stems with initial *ji-*, i.e. three times more than expected, and it is a very clear indication of a correlation between */j/* and */i/*. Either of two historical processes can be presumed: a) palatalization of the consonant before the front vowel; b) change of the localization series of the vowel (i.e. **u* > *i*) under the influence of the palatal consonant.

This assumption turns into an established fact when we notice the same correlation between */j/* and */e/*, another front vowel, as well as correlations between other palatal consonants */c, s, y, ny/* with front vowels */i, e/*. At the same time, the inverse kind of correlation is observed between the velar consonants and the back vowels. It is not easy sometimes to determine which of two processes theoretically probable is to be reconstructed (in our example, change of initial consonant or of the subsequent vowel), but some additional clues allow us to find a solution to this problem too.

In addition to the statistical data concerning the first syllable, we can bring to light all correlations in the second syllable, as well, such as correlations between the phonemes of the first and the second syllables (either consonants or vowels). While combining all these results, we obtain a trustworthy key to internal reconstruction of historical processes. On this way we also obtain some results in the field of historical morphology. So, in Songhai appears a phonetically groundless correlation between */m/* and */a/* in the final syllable. It can be interpreted as an indication of the fact that Songhai lexical stems include archaic fossilized suffix *-ma*.

This kind of procedure is of great importance for genetically isolated languages which are unaccessible to the standard comparative reconstruction. It is just the case of Songhai, taking into account numerous objections against including Songhai into the J. Greenberg‘s Nilo-Saharic macrofamily.

We are confronted today with attempts to breathe new life into the well-known hypothesis of the Mande-Songhai macrofamily, for instance (R.Nicolaï, 1989, H.Mukarovsky etc.). Some statistical arguments to this approach can be added.

There is a striking resemblance of Songhai and Manding phoneme frequencies in the initial as well as in the intervocalic positions. Comparison of frequencies in Songhai and in Mandinka presents a good illustration. In the table given below,

the first figure reflects the frequency of initial phoneme (per cent), the second figure is that for intervocalic consonant:

Mandinka				Songhai			
p 2/3	t 9/12	c 3/1	k 17/11	-	t 7/9	c 4/1	k 14/8
f 9/4	-	s 12/8	h 2/0	f 7/4	-	s 9/8	h 8/1
b 9/8	d 6/4	j 7/3	-	b 10/11	d 9/8	j 4/3	g 8/10
-	-	-	-	-	-	z 5/2	-
-	l 3/12	y 3/4	w 4/2	-	l 4/9	y 2/0	w 3/1
m 6/5	n 3/8	ny 4/3	ŋ 2/0	m 5/5	n 1/6	ny 1/1	-
-	r 0/13	-	-	-	r 0/13	-	-

It is evident that the frequencies in both languages coincide even in details. Does it mean that Songhai is more closely related to Mandinka than to other Mande languages? To gain an understanding of this question, it should be comprehended to what period we refer when dealing with when an internal reconstruction?

Let us assume: comparing the initial and the intervocalic frequencies in Bozo, we have arrived to a conclusion that in some historical period, the voiceless stops of this language became voiced in the intervocalic position. But we have no reason to belive that in the proto-language of the Bozo-Soninke subgroup, the frequency rates of voiceless stops

in the initial and intervocalic positions was different from that in the modern language. It is not impossible that this phonetical transformation took place in the Proto-Mande or (why not?) even in the Proto-Niger-Congo. All we affirm is the fact that this historical process took place at a certain moment. When comparing related languages, we cannot reconstruct a process which is older than their proto-language. To the contrary, internal reconstruction provides no way to date a process, but it allows to reconstruct the most archaic phonetic changes. Therefore, in some cases, statistical methods become basic for linguistic reconstruction.

According to the given thesis, the following methodological thesis can be formulated: any quantitative reconstruction cannot be regarded as satisfactory if it does not explain differences in phonemic frequencies in alternative positions as well as statistical correlations of phonemes in the lexical stock.

In guise of conclusion, let us examine another aspect of internal reconstruction on the example of the Wolof language data. Wolof is a language where nominal class markers are authonomous morphemes that follow nouns. From the very beginning, the attention of scholars was attracted to a remarkable peculiarity: in Wolof, initial consonant of a noun stem coincides or correlates in some basic features (location or articulation type) with the initial consonant of the nominal class marker (*mb*- ... $\leftrightarrow$ class M, *nd*-... $\leftrightarrow$ class L,

k-... ↔ class G, etc.). To explain this feature, Ida Ward and some other authors advanced the «alliteration theory». According to this theory, initial phoneme of a noun stem determines, on phonetic grounds, the choice of the class marker.

If semantics is taken into account, this theory can hardly be accepted. It is evident that noun stems are included into one nominal class or another according to their meaning, cf.: *dund* G "life", *dund* B "food".

There is another theory that means to explain resemblances of initial consonants in stems and in noun class markers through postulating a confix structure of noun class markers: initial consonant in Wolof noun stem is interpreted as a trace of archaic prefix. In some cases this explanation seems to be valid, especially in the nouns of the CVCVC structure.

There is however a tendency in West Atlantic studies to interprete initial consonants in nearly all nouns of the CVC structure as traces of noun prefixes (cf. Morice Delafosse's interpretation: *j-aan* J «snake», *b-at* B «neck», *g-en* G «tail», and so on).

A corollary of this assumption the *VC structure for noun stems in a West Atlantic language, which is highly doubtful. Another objection to this assumption is that consonants of the stem and the morpheme do not always coincide. More often, they correlate in some basic features. For example, class B includes nouns which have not only *b-*, but also *p-* as their initial consonants. Class L includes nouns beginning with *nd-, nj-, c-* and contains very few nouns beginning with *l-*. To explain these well-known

facts, many fanciful theories were advanced. Among them, the theory of Abbé Boilat, who believed that indigent ears could not distinguish clearly between such sounds as *b* and *p*, or *b* and *mb*.

If we perceive a correlation between stem-initial consonants and the consonants of noun class markers in Wolof, it means that such nouns outnumber certain rate. Therefore, it is easy to determine this rate for each noun class. If we calculate the frequency of each Wolof consonant in the word-initial position, we will find out that in a Wolof dictionary (Lexique, 1976-1981) containing 1688 noun stems, 4% have initial *mb*-. At the same time, 14% of nouns belong to the noun class M. If there is no correlation between the *mb*-stems and the marker M, we can expect to find approximately ten *mb*-nouns belonging to the class M (0,04 x 0,14 x 1688 = 10). Meanwhile, there are 58 such nouns, six times more than expected.

Here is an example of the contrary: the summarized frequency of initial labials is 24%, and the frequency of the L-class nouns is 5%. If there is no correlation between these elements, we would expect about 20 nouns marked with both these features (0,24 x 0,05 x 1668 = 20). In reality, there are none.

All the correlations between consonants of stems and class markers examined, the following assumption can be made: initial consonants of the majority of noun stems in Wolof were modified through two simultaneous processes:

1) Change of articulation type of consonant according to the grade of alternation required by the noun class. For instance, the noun class W require a «weak» grade of initial consonants: a fricative for a voiceless, a sonorous for voiced series. A "strong" grade is required the noun class M: stops in the voiceless series, prenasals in the voiced series of initial consonants.

2) Assimilation of initial consonants by the consonant of class marker according to the feature of articulation type. E.g., initial consonants of nouns belonging to the class M become labial, those of the nouns belonging to the class L become dental, and so on.

These two parallel processes result in Wolof in transformations of initial consonants that can be illustrated by the following example: **gVmb* «pumpkin, calabash» (cf. Serer *o gamb ol* «calabash», Fula dial. *gummb-al* «pumpkin») > Wolof **damb* L (assimilation by articulation point) > Wolof *ndad* L (fortis alternation grade of voiced initials, required by the class L).

That is the way to proceed from an analysis of sound frequencies in the Wolof lexicon to the comparison of Wolof with other West Atlantic languages.

Verification of results of comparative analysis. It is comparative etymological dictionary that summarizes results of comparative work. When the comparative studies of a language branch reach an advanced stage, such a dictionary may include about a thousand of reconstructed stems and the reflexes of each proto-language stem in the daughter languages.

In a comparative dictionary, the summarized number of reflexes in a daughter language is always inferiour to the summarized number of recons-

tructed stems. It is only but natural, because of various borrowings and innovations which substitute regular reflexes in each language. Let us assume that there are 50% proto-stems which have been substituted in a given language without leaving any trace in the modern vocabulary. Assume also that the proto-language had 1000 stems, and among them 100 stems had **k-* (10%) and 20 stems had **d-* (2%). How many reflexes of **k-* and of **d-* could be expected in the daughter language? Theoretically speaking, among the remaining 500 reflexes there must be about 50 stems (10%) whose initial phoneme(s) go(es) back to **k-*, and about 10 stems (2%) whose initial phoneme(s) go(es) back to **d-*. In other words, we can expect that the substitution of regular reflexes should not change the frequencies of the proto-phonemes; otherwise we should admit that, by some unknown reason, stems beginning with one phoneme (e.g., **k-*) were substituted more readily than those beginning with other phonemes. In other words, despite of innovations and borrowings, the frequency of reflexes of one proto-language phoneme is constant in any present daughter language, being equal to the frequency of the phoneme in the proto-language.

For example, if in a proto-language, 10% of all stems had initial **k-*, reflex of each tenth stem in any daughter language must begin with a phoneme which goes back to **k-*. The more inaccurate is the dictionary, the more numerous are significant deviations from the rate. This assumption turns to be a starting point of the method for verification of etymological dictionaries.

I tested this method while compiling a preliminary etymological dictionary of Mande based on the data of 30 Mande languages and dialects (Pozdnyakov, 1978). The first stage of the analysis gave 27 series of regular phonetic correspondences. It presupposes the reconstruction of 27 phonemes in the Proto-Mande language (some of them seem to be variants of one phoneme in alternative phonetic contexts, so the real number of phonemes in the Proto-Mande must be inferior). Each series of correspondences was supported by a set of lexical cognates. The number of reliable comparative sets represented in all the branches of the Mande languages totalled 334, and reflexes of 480 other sets were of more limited spread. From the basis of the 814 sets of cognates, the frequency of each initial phoneme in the Proto-Mande language has been calculated. According to the assumption, frequency of proto-phoneme must be equal to the frequency of its reflections in each daughter language. It means that deviations from the proto-language frequencies can be regarded as an evidence of weak points in the comparative analysis.

In Proto-Mande, 9% of stems had initial **s-* and 5% had initial **s_1-*. In mordern Gban, the frequency of stems going back to **s*-words (**s* > *s* in Gban) is 4%, and that of stems going back to **s_1*-words (**s_1* > *z* in Gban) is 9%. It means, in the first case (the reflection of **s-* in Gban), that I failed to find sufficient number of reflexes in Gban (and it is a good reason to examine whether **s-* results in some other consonants in Gban). In the second case (reflection of **s_1*) there are too many reflexes in Gban, and it should be verified, whether some of these stems in Gban do not go back to some other proto-phoneme. In such a situation it looks very likely that some Gban stems

beginning in *z*-, which are now regarded as reflexes of *$*s_1$*-, should be reinterpreted as reflexes of the initial **s*- of the Proto-Mande. In any case, further elaboration of phonemic correspondence system should be correlated with the weak points of the etymological analysis discovered through statistical methods. It is possible therefore to establish the series of correspondences or languages which were examined less accurately than others. It is also possible to calculate the percentage of words treated in a wrong way, which gives us a possibility of a quantitative evaluation of any etymological dictionary. For example, my preliminary version of the Mande comparative glossary has the percentage of errors of about 10% (that means, not less than 10%).

It was established that the etymological analysis had been more accurate for better documented languages, such as Bamana, Maninka, Mende, Susu, Kpelle, Bisa. On the contrary, such languages as Don, Ble, Vai, Busa, were much poorer in data, and it was very interesting to find out that they had the highest rate of errors. Correlation between the amount of data and the accuracy of the analysis is quite clear, which corroborates the plausibility if the applied method.

This method was tested on several etymological dictionaries. The results of the verifications were broadly discussed with specialists. It brings us to a general conclusion: no system of regular correspondences can be regarded as reliable and complete if the frequency of a proto-phoneme reflex is not equal in each daughter language, and if it deviates seriously from the frequency of the

phoneme in the proto-language.

Genetic relationship. Absolute dating of divergence of related languages is the most advanced and, at the same time, the most doubtful field of lexico-statistics. I share the skeptical attitide expressed by many specialists in what concerns absolute dating of language divergence. At the same time, I would like consider in more detail a far more reliable sphere of lexico-statistics, which is the relative dating of language bifurcations, i.e.the genetic classification of languages. There are several well-known methods based on comparison of similar lexical entries in 100- or 200-word lists. I would like to propose a quantitative procedure of genetic classifying on the basis of an etymological dictionary.

When dealing with historical divergence of languages, we assume that there was a point which can be regarded as the beginning of independent existence of two daughter languages. Therefore, if we regard language bifurcation as the beginning of independent event, we can use the statistical formula to calculate the number of stems in the proto-language of two descendants just before the beginning of their bifurcation:

$$\text{Number of proto-stems before the bifurcation} = \frac{\text{reflexes in L1 x reflexes in L2}}{\text{common reflexes in the L pair 1-2,}}$$

where L is one of cognate languages.

It is easy to calculate the total number of reflexes in each cognate language as well as the number of

common reflexes going back to the same proto-stems, for each pair of languages, on the basis of the etymological dictionary. If we arrange each pair of languages according to the number of proto-stems they share, we get the most probable succession of historical bifurcations in the language family. The more is the number of common proto-stems, the earlier began the period of independent existence of two languages.

Considering various statistical approaches to genetic classification, I have come to the conclusion that this one is the most adequate. Applying this method of calculation results in a new version of the genetic classification of Mande. It differs from others in two principal points: 1) Bozo and Soninke languages are the most closely related to Susu; 2) the Nothern branch is closer to the Eastern one, while the South-Western Mande branch should be regarded as more remoted one.

There is a whole range of statistical methods for comparative linguistic studies which are not dealt with in this paper. Eventually, it is not so important, which particular method of calculation is applied. In the guise of conclusion, I would like to highline the following idea: cross-control of the data in two correlated areas, namely qualitative and quantitative comparative studies, creates new ground for comparative studies, thus narrowing the range of choice among the possible comparative solutions.

REFERENCES

Abbé Boilat. *Grammaire de la langue woloffe.* Paris, 1858.

Delafosse M. Les classes nominales en wolof. // *Festchrift Meinhof. Schprachwissenschaftlische und andere Studien.* Hamburg, 1927.

Ducroz J.-M., Charles M.-C. *Lexique soŋey (songay)–français: parler kaado du Goroul.* Paris, 1978.

(Friedländer M.) *Lexique susu–français.* Conakry (n.d.)

Lexique wolof–français. CLAD, 42, Vol. 1-4. Dakar, 1976-1981.

Nikolaï R. Songhay et mandé. // *Mandenkan,* 18, 1989, p. 69-80.

Raimbault J.-B. *Dictionnaire Français–Soso et Soso–Français.* Rio-Pongo 1885.

Pozdnyakov K. *The Mande Languages: Results of comparative-historical analysis.* Ph.D. Dissertation. Moscow, 1978 (in Russian).

Pozdnyakov K. Development of the Consonant Alternation Systems in the Mande and West Atlantic Languages. – In: *African Historical Linguistics*, Moscow, 1978 (in Russian).