

**B. R. KARIMOV, SH. SH. MUTALOV**

# **AVERAGED LANGUAGES: AN ATTEMPT TO SOLVE THE WORLD LANGUAGE PROBLEM**

The creation of averaged languages compiled by a special method of averaging vocabularies and grammar rules and norms of cognate languages has been proposed. The method has been illustrated using examples of the vocabularies of the Turkic, Slavonic, Romance, and Iranian languages. The method is valid for compiling averaged Indian, German and so on languages. Political, legal, social, cultural, and linguistic problems of averaged languages have been investigated. The summary of the book is given in Russian, German, and French at the end of the book.

Es ist vorgeschlagen, eine Durchschnittssprache durch bildung eines durchschnittlichen lexikalischen Bestandes, der durchschnittlichen grammatischen Regeln und der durchschnittlichen Sprachnormen der verwandten Sprachen zu schaffen. Diese Methode ist mit Hilfe der Beispiele aus der Lexik der türkischen, slawischen, romanischen, und iranischen Sprachen illustriert. Die öhnliche Methode kann für die Bildung folgender Durchschnittssprachen gebraucht werden: durchschnittsindische, durchschnittsgermanische u.a. Die politisch-rechtlichen, sozialkulturellen und linguistischen Probleme werden dabei untersucht. Zu dem Buch gibt es eine Zusammenfassung auf Russisch, Deutsch und Französisch.

Il est à proposer de créer des langues moyennes par la voie de moyennage des fond léxique, des regles linguistiques. Cette méthode est illustré par l'emploi des donnés du léxique des langues turques, slaves, romanes, et iranienn. En même temps on propose de créer les autres langues moyennes: indienn-moyenne, germanique-moyenne et d'autres. On étudie les problèmes politiques, sociaux, culturels et linguistiques de la langue moyenne. A la fin de l'ouvrage on donne son resumé en russe, allemand et en français.

Предлагается создать усредненные языки методом усреднения лексических фондов, языковых правил и норм родственных языков. Метод иллюстрируется с использованием примеров из лексики тюркских, славянских, романских и иранских языков. Аналогично предлагается создать и ряд других усредненных языков: усредненных индийских, германских и так далее языков. Рассмотрены политико-правовые, социокультурные и лингвистические проблемы усредненных языков. Резюме брошюры дано на русском, немецком и французском языках.

## **Preface**

Many things have changed in the world since first edition of the booklet in 1992 but the world language situation has not, or has changed little. The value of English seems to have increased due to (a) fast expansion of Internet, and (b) stepping back of one of the rivals of English, i.e. Russian. Some experts talk about Chinese stepping forward as a world language. Despite Chinese is not number one language in the area of scientific and technological information exchange it is such as the number of native speakers and people of studying and speaking it as a second language is concerned. Thus, the world language problem understood in the terms of the people's language rights is still acute. And, philosophy of the averaged languages is given more emphasis in second edition.

We decided to add three more Turkic languages – Kyrghyz, Turkmen, and Uigur - to the previously involved five ones. It makes the model more illustrative and balanced in the terms of representation of the three language groups in the Turkic language family. Number of native speakers of the languages was updated for the second edition. Published reference books were used as the sources of population and language speakers' information sources for the first edition. This kind of sources is not available now in this country because of transitional period difficulties. Fortunately, solution came due to the Internet in the of Joshua Project web site ([www.joshuaproject.net](http://www.joshuaproject.net)). Comparison of Joshua Project population and languages' speakers figures with that available from censuses information for some individual countries showed good accuracy of the former. The Joshua Project figures are more detailed and accurate than that of CIA World Fact Book ones. Joshua Project data comes from quite a number of sources. The creators of the web site rely upon census information when it is available. These sources are sometimes supplemented by individuals who are in position to know current information. Besides, we are developing a model for the averaged languages, and thus found the accuracy of the figures satisfactory.

As population of the countries and number of native speakers of the languages vary in the time so do some coefficients in the model. The second edition provided an opportunity to study coefficients' behaviour in time. The coefficients change a little but their ratios almost do not change. Resulting lists of words from the first edition and the second, and intermediate model with five Turkic language dated 2005 have been compared. The list is quite sustainable. This analysis is given in Appendices.

## **AVERAGED LANGUAGES: AN ATTEMPT TO SOLVE THE WORLD LANGUAGE PROBLEM**

Despite some peculiarities of nation formation process the existing nations have developed of related tribes and all-nation language, literary language of a nation, has developed on the base of tribe dialects as a result of their interference. From the point of view of genesis it seems quite natural that germs of nations were among related tribes because related tribes inhabited neighbouring territories and had easy communications. Besides, the closer were their languages the closer was the communication. Thus, kins, tribes, and nations should be considered as historical stages of a concept that develops alongside with neighbouring kins, tribes, and nations and keeps communication with them. B. Karimov offered this approach and named the notion ethnation and ethnic neighbourhood was named oikumene of an ethnation and the approach was named oikumenic one. Historically, an ethnation develops together with its oikumene and passes stages of kindred, tribe and nation. In prospective oikumene of every ethnation enlarges up to whole mankind<sup>1</sup>.

The mankind is sure to be a single whole. The developing stage when related nations form communities of higher level than nation seems to be inevitable. But we must keep in mind that there may be wrong ways to consolidate nations like the attempts to form single Soviet and Yugoslavian and similar peoples. What was wrong with them? Those attempts denied the value of ethnic identity the attempts were to form homogeneous people of ethnically heterogeneous population, that was not an attempt to form a community of higher level, but that to form a larger

---

<sup>1</sup> For more detailed information please see: Каримов Б.П. К характеристике сущности нации и межнациональных отношений // Общественные науки в Узбекистане. 1991. N.9. С.27-33 (Karimov B.R. K kharakteristike sushchnosti natsii i mezhnatsional'nikh otnosheniy. In: Obshchestveniyе nauki v Uzbekistane. 1991. № 9, p. 27-33)

nation only. So, pursuing the policy of consolidating nations we must take into consideration that ethnic identity is value common to all mankind.

A number of large movements to consolidate close nation was known in the world history. For instance, Pan-Slavonism, Pan-Turkism, Pan-Mongolism, Pan-Iranism, Pan-Hindism were such movements. In spite their specific features all these movements are identical in their essence and form generic notion. We named the notion ethnic linguistic panism, the shortened form is ethnolinguopanism.

Ethnolinguopanism is an application of the philosophical conception of potential unity to a certain ethnic group. The closeness within a certain group of peoples and comprehension that the men of these peoples are able to understand each other to some extent even if they know only their own languages is a base of any form (specific kind) of ethnolinguopanism. The similarity of the communication process between above mentioned peoples and between the people speaking dialects of a language is the epistemological basis for the respective panism come to existence. The sphere of each ethnolinguopanism is completed with specific features such as affinity of culture, religion, literature, origin, historical fortune and so on of the ethnoses whose language closeness was based upon when respective ethnolinguopanism was being grounded.

In our opinion, ethnolinguopanisms form the following types:

1. Linguocultural ethnolinguopanism, i.e. ideas to bring together the languages and alphabets of nations belonging to an ethnolinguopanism, to create an auxiliary common language for them, and the ideas to enlarge common features of culture, to strengthen mutual communications and assistance;

2. Political ethnolinguopanism, i.e. aims to create federative, confederative state, commonwealth or unitary, indivisible state of these nations;

3. Extreme ethnolinguopanism, it means the ideas on necessity of introducing a single language, to form a large nation from close but still individual nations.

As at present the unity of mankind is necessary for its survival, it would be reasonable to search consciously for the ways and means to form and maintain such a unity. Uniting of close nations into federations, confederations or unitary states according to their will is one of the ways.

The language problem arises in multiethnic states. There must be a language serving as a common means of communication, so called mediating language, in such states. There are many mediating languages in the world. We can consider English as the world mediating language, and there are many regions in the world with their own mediating languages. For instance Swahili is such a language in Eastern Africa, Hindi in Hindustan, Spanish and Portuguese in Latin America and so on. But a national language as mediating one has at least two unacceptable features.

First, from the viewpoint of justice and spiritual life, the native speakers of other languages would feel themselves deprived from their rights to use their native languages. Second, from economic point of view all peoples but the native speakers of a mediating language would fall behind for years due to the necessity to study mediating language along with their mother languages to get the world information. Thus we conclude that none of the national languages does for mediating language.

Latin and planned languages such as Esperanto will not suit too as the former is very poor in terminology concerning modern science, technology and production and the latter are close to the languages of Romance stock, i.e. it is not as neutral as it would be desired.

A regional mediating language must be a language that is equally close to the people's languages of the region where it is supposed to be a mediating language. The world mediating language must meet similar requirements. In our opinion, regional mediating languages should be constructed on the basis of group of close languages according to a special method providing its neutrality as well as optimum degree of its intelligibility to as many people as possible among native speakers of these languages.

We can specify democratic averaged ethnolingvopanism that is characterized by certain peculiarities of culture and language policy. Let us dwell a bit on some aspects of related problems.

Within the framework of the concept of ethnoligvopanism two main approaches to the language policy are developed aimed at formation of: 1) averaged language, and 2) a common language for communication of linguistically kin ethnations and ethnic groups.

On the other hand, acceleration of the social processes, emergence of the humankind global problems system brings the necessity to move to noosphere, to simulation and control of information and language processes, in particular. The mediating language creation method must conform and imitatively simulate objective mechanism of *koine* formation process in the course of communication of the related languages native speakers. It means the more is number (1) of the related languages where and (2) the individual native speakers of these languages with whom certain language form (or norm) is observed the larger the probability for this very form (or norm) to be included into the *koine* forms or norms system.

The above is not merely theoretical speculation. Mechanism, or method, of mediating language creation is important because when persons that consider related languages as their vernacular realize their ethnic affiliation the closeness degree of the related languages, realization of the persons' languages by them as independent languages or dialects of one language play important, though not decisive, role. In the systems of dialects of various languages in the world the dialects closeness measure that results in realization of affiliation with the same language by bearers of a dialect varies essentially from one language to another. In order to promote elimination of the above inequality and bring mathematics into the field a method of quantitative assessment of synchronic closeness measure of dialects and related languages has been developed<sup>2</sup>. Next, a kind of inversion of the above method, a method of creating potential common communication, or mediating, language for groups of kin ethnations has been offered.

The essence of the offered method is to define a common fund for groups of dialects or languages by means of a special mathematic procedure, and to codify the common fund as a standard, i.e. as units of corresponding levels of the language being compiled. The procedure is as follows: units of language levels are connected to vectors. As a rule, the level units have versions: different meanings of a word, positional variants of phonemes and morphemes, different ways of expressing syntactic relations. A vector space is associated with the unit system of a level and this space is constructed in such a way that when the main meaning of a unit is bound with magnitude 1, other variants are correlated with numbers less than 1 according to the formula:

$$X_{n,\alpha}^i = 1 - \frac{i-1}{2s}, \quad (1)$$

where “n” stands for the number of units in the list of units; “α” stands for a concrete language; “i” stands for the variant number of the level unit with number “n”; “s” is the number of variants

---

<sup>2</sup> See: Каримов Б.Р., Муталов Ш.Ш. К вопросу о языковой политике в многоязычных развивающихся странах // Тезисы докладов республиканской научно-теоретической конференции молодых ученых и специалистов по общественным наукам. Ташкент, 2-3 марта 1982 г., секция 1. Ташкент, 1982. С.94-95 (Karimov B.R., Mutalov Sh.Sh. K voprosu o yazykovoy politike v mnogoyazychnikh razvivayushchikhsia stranakh. In: Tezisy dokladov respublikanskoy nauchno-teoreticheskoy konferentsii molodykh uchenykh i spetsialistov po obshchestvennim naukam. Tashkent, 2-3 marta 1982 g., sektsiya 1. Tashkent, 1982, s. 94-95). Каримов Б.Р., Муталов Ш.Ш. О количественной оценке синхронической близости родственных диалектов и языков // Тюркское языкознание. Материалы международной конференции. Ташкент, 1985. (Karimov B.R., Mutalov Sh.Sh. O kolichestvennoy otsenke sinkhronicheskoy blizosti rodstvennykh diolektov i yazikov. In: Tiurkskoye yazikoznaniye. Materialy mezhdinardnoy konferentsii. Tashkent, 1985) (Abstracts of the Conference were published in 1980)

of the level units with number “n”. These numbers form a set of the vector components<sup>3</sup>. We named the method averaging method and the languages built by the method averaged languages.

We describe the method of constructing an averaged language using examples of the vocabularies of eight Turkic, five Slavonic, and five Romance languages. As far as the level structure exists in all languages this method can be applied to any group of languages without fundamental changes no matter what group of languages it might be.

For the vocabulary it is necessary to select an upper part of a certain length from the frequency dictionary of one of the languages. This language serves as a basic one<sup>4</sup>. The meanings of polysemantic words that are marked in the dictionary entry of the basic language by figures are considered as separate lexemes. The chosen words in the list are to be numerated. For every lexeme of ordered list the equivalents should be found in the languages or dialects the averaged language must be constructed on.

To illustrate the idea of the vector space we can take the Uzbek and the Turkish equivalents corresponding to the word "говорить" — to speak.<sup>5</sup> Let's take the Uzbek word "гапирмоқ" (to speak) in its main meaning. There are synonyms for the word in Uzbek: "сўзламоқ", "айтмоқ", "сўйламоқ (high style). By means of the Russian word "говорить" (main meaning) we find the Turkish equivalents of the Uzbek word "гапирмоқ":

UZBEK	TURKISH
гапирмоқ	konuşmak
сўзламоқ	bahsetmek

Evidently, the number of translation equivalents of the word "гапирмоқ" is smaller than the number of synonyms with the meaning "гапирмоқ". Thus we have to take into consideration Perfect Aspect of the Verb and find equivalents of Perfect pair of the word "говорить". This is the word "сказать" (to say), which corresponds to the Uzbek word "демоқ" (to say) (Table 1).

**Table 1**

### UZBEK AND TURKISH LEXEMES WITH THE MEANING “TO SPEAK”

UZBEK	TURKISH
гапирмоқ	konuşmak
сўзламоқ	bahsetmek
тилдан фойдаланмоқ	söylemek
айтмоқ	demek
изҳор қилмоқ	

One can see that these groups of equivalents are closer in their number to the group of synonyms with the meaning 'to speak'. We shall illustrate extraction of the units from general vocabulary fund using examples from 8 Turkic languages with large number of native speakers<sup>6</sup>,

<sup>3</sup> Каримов Б.Р., Муталов Ш.Ш. О количественной оценке синхронической близости родственных языков и диалектов // Тюркское языкознание. Материалы III Всесоюзной тюркологической конференции. Ташкент, 1985. С. 126-129. (Karimov B.R., Mutalov Sh.Sh. O kolichestvennoy otsenke sinkhronicheskoy blizosti rodstvennykh diolektov i yazikov. In: Tiurkskoye yazikoznaniye. Materialy mezhdinarnodnoy konferentsii. Tashkent, 1985 s. 126-129)

<sup>4</sup> For many languages bilingual dictionaries are not worked up. In such cases it is possible to choose non-related languages as the basic ones, if there are basic – corresponding language dictionaries. For example, in the group of the Slavonic languages we can choose any Slavonic language as the basic, in case of Turkic languages it might be Russian, we have chosen Russian as the basic language.

<sup>5</sup> We have chosen Russian as the basis language because of availability of respective dictionaries.

<sup>6</sup> Please refer to: Каримов Б.Р., Муталов Ш.Ш. Ўртаюрк тили. Тошкент, 1992 (Karimov B.R., Mutalov Sh.Sh. Ortaturk tili. Toshkent, 1992)

5 Slavonic languages using Cyrillic alphabet (as far as Russian belongs to the group, lexeme "ГОВОРИТЬ" is in its main lexical meaning with its synonyms), 5 Romance, and Iranian languages. The equivalents of the basic lexeme are represented in the Tables 2, 3, 4, and 5.

Then the Tables 2, 3, 4, and 5 should be transformed into the Tables 6, 7, 8, and 9 where equivalent words of the same root from different languages are placed in one line.

For the words of the same root the function  $F_n^j(x)$  has been introduced by the formula

$$F_n^j(x) = \sum_{\alpha=1}^A \left( 1 + \frac{K_\alpha}{\bar{K}} \right) X_{n,\alpha}^j, \quad (2)$$

where A stands for total number of the languages under averaging,  $K_\alpha$  is number of native speakers of the language  $\alpha$ ,  $\bar{K}$  is arithmetic mean of the native speakers of languages from the group of related languages that is calculated as quotient of total number of native speakers of all the languages in the group divided by the number of languages in the group, j stands for ordinal number of cognate roots in Tables 6, 7, 8, and 9.

**Table 2**

**SYNONYMS OF THE UZBEK WORD "ГАПИРМОҚ"  
WITH THEIR EQUIVALENTS IN SEVEN TURKIC LANGUAGES**

AZERB $\alpha = 1$	$X_{n,1}^j$	KAZAKH $\alpha = 2$	$X_{n,2}^j$	KYRGHYZ $\alpha = 3$	$X_{n,3}^j$	TATAR $\alpha = 4$	$X_{n,4}^j$	TURK $\alpha = 5$	$X_{n,5}^j$	TURKMEN $\alpha = 6$	$X_{n,6}^j$	UIGUR $\alpha = 7$	$X_{n,7}^j$	UZBEK $\alpha = 8$	$X_{n,8}^j$
1 danışmaq	1.00	сөйлеу	1.00	сүлөө	1.00	сөйләшү	1.00	konuşmak	1.00	геплемек	1.00	сөзимек	1.00	гапирмоқ	1.00
2 demek	0.84	деу	0.84	айтуу	0.90	әйтү	0.84	bahsetmek	0.88	сөзлемек	0.90	гәп қилмақ	0.93	сўзламоқ	0.92
3 söylemek	0.67	айту	0.67	дем	0.80	әйтөп бирү	0.67	söylemek	0.75	айтмақ	0.80	өйтмақ	0.86	тилдан	
4 -	-	-	-	далил болуу	0.70	-	-	demek	0.63	гуруң этмек	0.70	демәк	0.79	фойдаланмоқ	0.84
5 -	-	-	-	билдирүү	0.60	-	-	-	-	гөркезмек	0.60	ипадә қилмақ	0.71	айтмоқ	0.76
6 -	-	-	-			-	-	-	-			делиллимәк	0.64	демөк	0.68
												билгүзмәк	0.57	изхор қилмоқ	0.60



**Table 3**

EQUIVALENTS OF THE LEXSEME "ГОВОРИТЬ" (to speak) IN FIVE SLAVONIC LANGUAGES

Russian $\alpha = 1$	$X_{n,1}^j$	Ukrainian $\alpha = 2$	$X_{n,2}^j$	Belorus. $\alpha = 3$	$X_{n,3}^j$	Serb-Croat $\alpha = 4$	$X_{n,4}^j$	Bulgarian $\alpha = 5$	$X_{n,5}^j$
1 говорить	1.00	говорити	1.00	гавариць	1.00	говорити	1.00	говоря	1.00
2 разговаривать	0.75	балакати	0.82	казаць	0.75	разговаривати	0.75	казвам	0.82
3 -	0.00	мовляти	0.64	-	0.00	-	0.00	разговорям	0.64

**Table 4**

EQUIVALENTS OF THE LEXEME "ГОВОРИТЬ" (to speak) IN FIVE ROMANCE LANGUAGES

Spanish $\alpha = 1$	$X_{n,1}^j$	Italian $\alpha = 2$	$X_{n,2}^j$	Portug. $\alpha = 3$	$X_{n,3}^j$	Roumanian $\alpha = 4$	$X_{n,4}^j$	French $\alpha = 5$	$X_{n,5}^j$
1 hablar	1.00	parlare	1.00	dizer	1.00	a vorbi	1.00	dire	1.00
2 decir	0.82	dire	0.88	falar	0.75	a dovedi	0.75	parler	0.82
3 mostrar	0.64	(di)mostrare	0.75	-	0.00	-	0.00	montrer	0.64
4 -	0.00	paleşare	0.62	-	0.00	-	0.00	-	0.00

**Table 5**

EQUIVALENTS OF THE LEXEME "ГОВОРИТЬ" (to speak) IN FIVE IRANIAN LANGUAGES

Dari <sup>7</sup> $\alpha = 1$	$X_{n,1}^j$	Kurd <sup>8</sup> $\alpha = 2$	$X_{n,2}^j$	Persian <sup>9</sup> $\alpha = 3$	$X_{n,3}^j$	Pashto <sup>10</sup> $\alpha = 4$	$X_{n,4}^j$	Tajik <sup>11</sup> $\alpha = 5$	$X_{n,5}^j$
1 гофтан	1,00	gotin	1,00	ҳарф задан	1,00	вайэл	1,00	гап задан	1,00
2 ҳарф задан	0,90	gılı kirin	0,90	гофтан	0,90	хабэри кавэл	0,88	ҳарф задан	0,93
3 гап задан	0,80	sohbet kirin	0,80	такаллоом кардан	0,80	гэредэл	0,75	гуфтан	0,86
4 соҳбат кардан	0,70	qize kirin	0,70	хекаят кардан	0,70	хабэри-атэри кавэл	0,62	сухан рондан	0,78
5 мокали-ма кардан	0,60	şedeti kirin	0,60	гофтогу кардан	0,60	-	0,00	гувоҳи додан	0,71
6 -	0,00	-	0,00	-	0,00	-	0,00	шаҳодат додан	0,64
7 -	0,00	-	0,00	-	0,00	-	0,00	далолат кардан	0,57

The formula (2) may be transformed to more general shape:

<sup>7</sup> Кедайтене Е.И., Островский Б.Я., Митрохина В.И. Учебный русско-дари словарь. Москва, 1983; Лебедев К.А., Яцевич Л.С., Конаревский М.А. Русско-пушту-дари словарь. Москва, 1983.

<sup>8</sup> Русско-курдский словарь / Фаризов И.О. – составитель. Москва, 1957

<sup>9</sup> Восканян Г.А. Русско-персидский словарь. Москва, 2005.

<sup>10</sup> Зудин П.Б. Русско-афганский словарь. Москва, 1955; Лебедев К.А., Яцевич Л.С., Конаревский М.А. Русско-пушту-дари словарь. Москва, 1983.

<sup>11</sup> Русско-таджикский словарь / М.С.Асимов – ред. Москва, 1985.

$$F_n^j(\frac{\rho}{k}) = \sum_{\alpha=1}^A \left( 1 + \gamma \frac{K_\alpha}{K} \right) \left( 1 - \frac{i-1}{\beta s} \right). \quad (4)$$

Here  $\beta$  and  $\gamma$  are constants. Even more general formulae can be dealt with. Choosing a formula means adopting different models of averaging.

The next step of the procedure is to order the roots according to the value of this function. In our case we have ordered the roots according to the function  $F_n^j(\frac{\rho}{k})$  defined by the formula (2).

Table 6

## ETYMOLOGICALLY ORDERED EQUIVALENTS OF THE LEXEME “TO SPEAK” IN EIGHT TURKIC LANGUAGES

AZERB $\alpha = 1$	$X_{n,1}^j$	KAZAKH $\alpha = 2$	$X_{n,2}^j$	KYRGHYZ $\alpha = 3$	$X_{n,3}^j$	TATAR $\alpha = 4$	$X_{n,4}^j$	TURK $\alpha = 5$	$X_{n,5}^j$	TURKMEN $\alpha = 6$	$X_{n,6}^j$	UIGUR $\alpha = 7$	$X_{n,7}^j$	UZBEK $\alpha = 8$	$X_{n,8}^j$
1 danışmak	1.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00
2 demek	0.84	деу	0.84	дем	0.80	-	0.00	demek	0.63	-	0.00	демәк	0.79	демоқ	0.68
3 söylemek	0.67	сөйлеу	1.00	сүлөө	1.00	сөйләшү	1.00	söylemek	0.75	сөзлемек	0.90	сөзлимек	1.00	сўзламоқ	0.92
4 -	0.00	айту	0.67	айтуу	0.90	әйтү	0.84	-	0.00	айтмак	0.80	ейтмак	0.86	айтмоқ	0.76
5 -	0.00	-	0.00	-	0.00	әйтеп бирү	0.67	-	0.00	-	0.00	-	0.00	-	0.00
6 -	0.00	-	0.00	далил болуу	0.70	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00
7 -	0.00	-	0.00	билдирүү	0.60	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00
8 -	0.00	-	0.00	-	0.00	-	0.00	bahsetmek	0.88	-	0.00	-	0.00	-	0.00
9 -	0.00	-	0.00	-	0.00	-	0.00	konusmak	1.00	-	0.00	-	0.00	-	0.00
10 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	гуруң әтмек	0.70	-	0.00	-	0.00
11 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	гөркезмек	0.60	-	0.00	-	0.00
12 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	геплемек	1.00	-	0.00	-	0.00
13 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	гәп қилмақ	0.93	-	0.00
14 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	ипадә қилмақ	0.71	-	0.00
15 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	делиллимәк	0.64	-	0.00
16 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	билгүзмәк	0.57	-	0.00
17 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	тилдан	
18 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	фойдаланмоқ	0.84
19 -	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	изхор қилмоқ	0.60
														гапирмоқ	1.00

**Table 7**

ETYMOLOGICALLY ORDERED EQUIVALENTS OF THE LEXSEME "TO SPEAK" IN  
FIVE SLAVONIC LANGUAGES

Russian $\alpha = 1$	$X_{n,1}^j$	Ukrainian $\alpha = 2$	$X_{n,2}^j$	Belorus. $\alpha = 3$	$X_{n,3}^j$	Serb-Croat $\alpha = 4$	$X_{n,4}^j$	Bulgarian $\alpha = 5$	$X_{n,5}^j$
1 говорить	1.00	говорити	1.00	гавариць	1.00	говорити	1.00	говоря	1.00
2 разговаривать	0.75	-	0.00	-	0.00	разговаривати	0.75	разговорям	0.64
3 -	0.00	балакати	0.82	-	0.00	-	0.00	-	0.00
4 -	0.00	мовляти	0.64	-	0.00	-	0.00	-	0.00
5 -	0.00	-	0.00	казаць	0.75	-	0.00	казвам	0.82

**Table 8**

ETYMOLOGICALLY ORDERED EQUIVALENTS OF THE LEXSEME "ГОВОРИТЬ" (to  
speak) IN FIVE ROMANCE LANGUAGES

Spanish $\alpha = 1$	$X_{n,1}^j$	Italian $\alpha = 2$	$X_{n,2}^j$	Portug. $\alpha = 3$	$X_{n,3}^j$	Romanian $\alpha = 4$	$X_{n,4}^j$	French $\alpha = 5$	$X_{n,5}^j$
1 hablar	1.00	-	0.00	-	0.00	-	0.00	-	0.00
2 decir	0.82	dire	0.88	dizer	1.00	-	0.00	dire	1.00
3 -	0.00	parlare	1.00	-	0.00	-	0.00	parler	0.82
4 mostrar	0.64	(di)mostrare	0.75	-	0.00	-	0.00	montrer	0.64
5 -	0.00	palesare	0.62	-	0.00	-	0.00	-	0.00
6 -	0.00	-	0.00	falar	0.75	-	0.00	-	0.00
7 -	0.00	-	0.00	-	0.00	a vorbi	1.00	-	0.00
8 -	0.00	-	0.00	-	0.00	a dovedi	0.75	-	0.00

Table 9

ETYMOLOGICALLY ORDERED EQUIVALENTS OF THE LEXSEME “ГОВОРИТЬ” (to speak) IN FIVE IRANIAN LANGUAGES

Dari $\alpha = 1$	$X_{n,1}^j$	Kurd $\alpha = 2$	$X_{n,2}^j$	Persian $\alpha = 3$	$X_{n,3}^j$	Pushtu $\alpha = 4$	$X_{n,4}^j$	Tajik $\alpha = 5$	$X_{n,5}^j$
1 гофтан	1,00	gotin	1,00	гофтан	0,90	-	0,00	гуфтан	0,86
2 -	0,00	gili kirin	0,90	-	0,00	-	0,00	-	0,00
3 харф задан	0,90	-	0,00	харф задан	1,00	-	0,00	харф задан	0,93
4 гап задан	0,80	-	-	-	0,00	-	0,00	гап задан	1,00
5 соҳбат кардан	0,70	sohbet kirin	0,80	-	0,00	-	0,00	-	0,00
6 мокалима кардан	0,60	-	0,00	-	0,00	-	0,00	-	0,00
7 -	-	-	-	-	-	-	-	-	-
8 -	0,00	qize kirin	0,70	-	0,00	-	0,00	-	0,00
9 -	0,00	şedeti kirin	0,60	-	0,00	-	-	шаҳодат додан	0,64
10 -	0,00	-	0,00	такаллом кардан	0,80	-	0,00	-	0,00
11 -	0,00	-	0,00	некаят кардан	0,70	-	0,00	-	0,00
12 -	0,00	-	0,00	гофтогу кардан	0,60	-	0,00	-	0,00
13 -	0,00	-	0,00	-	0,00	вайэл	1,00	-	0,00
14 -	0,00	-	0,00	-	0,00	хабэри кавэл	0,88	-	0,00
15 -	0,00	-	0,00	-	0,00	гэредэл	0,75	-	0,00
16 -	0,00	-	0,00	-	0,00	хабэри-атэри кавэл	0,62	-	0,00
17 -	0,00	-	0,00	-	0,00	-	0,00	сухан рондан гувоһи додан	0,78
18 -	0,00	-	0,00	-	0,00	-	0,00	далолат кардан	0,71
	0,00	-	0,00	-	0,00	-	0,00	-	0,57

The role of the coefficient  $1 + \frac{K_\alpha}{K}$  is taking into consideration the number of native speakers of languages in the group which the languages is being averaged for. The number  $K_\alpha$  of peoples who speak the languages in question and the values of  $1 + \frac{K_\alpha}{K}$  are given in the Table 8.

French, Spanish, and Portuguese are spoken and spread beyond the borders of the countries where they have been formed and are considered native (first) languages of many indigenous ethnic groups in Latin America. As we offer a model, we include the number of people speaking these languages as the first language plus number of people speaking these languages as the second in the countries where these languages are official languages with weight 0.5 into the Table 8.

**Table 10**

NUMBER OF 8 TURKIC PEOPLES, 5 SLAVONIC PEOPLES WHO USE CYRILLIC,  
AND SPEAKERS OF ROMANCE AND IRANIAN LANGUAGES AND VALUES OF  $1 + \frac{K_\alpha}{K}$

People or speakers of	Number	$1 + \frac{K_\alpha}{K}$
Azerbaijan	24,438,300	2.56
Kazakh	12,058,160	1.77
Kyrghyz	3,904,000	1.23
Tatar	7,205,690	1.46
Turk	55,920,550	4.58
Turkmen	7,380,000	1.47
Uigur	10,913,000	1.7
Uzbek	27,567,840	2.77
<hr/>		
Russian	137,837,215	4.33
Ukrainian	37,649,030	1.91
Byelorussian	9,916,001	1.24
Serbs and Croats	14,023,880	1.34
Bulgarian	7,394,160	1.18
<hr/>		
Spanish	675,899,070	3.1
Italian	32,833,900	1.15
Portuguese	201,112,815	1.92
Romanian	19,737,890	1.09
French	161,470,470	1.74
<hr/>		
Dari	27 856 500	1,921
Kurd	29 418 000	1,973
Persian	39 782 278	2,315
Pushtu	47 991 000	2,586
Tajik	6 762 000	1,224

Equivalent words corresponding to each lexeme of the basic selection are placed in the Table 9 in accordance with the decreasing order of the value of the  $F_n^j(\frac{P}{K})$  function.

The roots that are placed first in the list given are recommended to be included into the vocabulary of the averaged language as original lexical units. The other units may form the word-stock of synonyms to enrich the averaged language vocabulary when needed.

**Table 11**

LEXMESES ORDERED ACCORDING TO THE VALUES OF  $F_n^j(\mathcal{P})$  FUNCTION

Turkic		Slavonic		Romance		Iranian	
Root	$F_n^j(\mathcal{P})$	Root	$F_n^j(\mathcal{P})$	Root	$F_n^j(\mathcal{P})$	Root	$F_n^j(\mathcal{P})$
1 söyle/mek	13.96	говор/гавар	10	des/diz/dir	8.03	гофтан/gotin	7,031
2 de/mek	10.63	разговар/	5.22	mostr/montr	4.6	харф задан	5,182
3 айт/у	8.28	разговор	1.57	habl	4.1	соҳбат қардан	2,923
4 konuşmak	4.58	балак	1.22	parl	2.58	гап задан	2,761
5 bahsetmek	4.03	мовл	1.76	falar	1.44	вайёл	2,586
6 гапирмоқ	2.77	каз/казв		vorbi	1.09	хабэри кавэл	2,276
7 danış/mak	2.56			dovedi	0.82	гэредэл	1,94
8 тилдан	2.33			palesare	0.71	такаллоом қардан	1,852
фойдаланмоқ							
9 гəп қилмак	1.58					gılı kirin	1,776
10 гəплемек	1.47					хекаят қардан	1,620
11 ипадə қилмак	1.21					хабэри-атэри	1,603
						кавэл	
12 изҳор қилмоқ	1.66					гофтогу қардан	1,389
13 делиллимəк	1.1					qize kirin	1,381
14 гурун этмек	1.03					мокалима қардан	1,153
						şedeti	
15 билгүзмəк	1.02					kirin/шаҳодат	0,967
16 əйтəп бирү	0.98					додан	
						сухан рондан	0,954
17 билдирүү	0.97					гувохи додан	0,869
18 далил болуу	0.89					далолат қардан	0,698
19 гəркезмек	0.88						

The method described concerns to synchrony, but one can easily transform it to diachrony, i.e. to take into consideration the historical process of language development. For this purpose the historical period of language development should be divided into series of consequent periods of equal length and synchronic formulae of averaging should be applied to m-th period of time. Developers of an averaged language should find necessary language data for this averaging. For this purpose, the texts concerning to certain m-th period of time may be used. Then the following formula taking into account both synchrony and diachrony should be applied.

$$F_n^j(\mathcal{P}, \beta, \gamma) = \sum_m \sum_{\alpha_m} \left( 1 + \gamma \frac{K_{\alpha_m}}{\bar{K}_m} \right) X_{n, \alpha_m}^j = \sum_m \sum_{\alpha_m} \left( 1 + \gamma \frac{K_{\alpha_m}}{\bar{K}_m} \right) \left( 1 - \frac{j-1}{\beta s_m} \right). \quad (5)$$

Here “m” is order number of consequent time periods,  $\alpha_m$  stands for the language  $\alpha$  in its m-th period of development,  $K_{\alpha_m}$ ,  $\bar{K}_m$ ,  $X_{n, \alpha_m}^j$ ,  $\alpha_m$ ,  $s_m$  are the expressions of factors described in formulae (1),(2),(3), and (4) and are valid for m-th historical period of the language  $\alpha$ , j, N,  $\alpha, \beta$  and  $\gamma$  are exactly as previous ones. To compile averaged language according to the general formula (5) is very difficult problem. It is possible to replace the way by approximate one. Let's suppose that there were consequent relatively stable periods in language development. Let's suppose as well that we dispose of dictionaries of each such period. So we can average the language of each period according to the formulae (1), (2), and (4). Then the results of averaging for these historical periods should be multiplied with the relative length of supposed stable periods and the products should be summed up. After that we can make certain conclusions. Following formula for the calculations is desirable:

$$F_n^j(\mathcal{P}, \beta, \gamma) = \sum_m \sum_{\alpha_m} \left( 1 + \gamma \frac{K_{\alpha_m}}{\bar{K}_m} \right) \left( 1 - \frac{j-1}{\beta s_m} \right) \frac{\Delta t_m}{\Delta t}. \quad (6)$$

The meanings of new notations are:  $\Delta t$  is shortest of the periods of stable development of language. This period can be used as a measure unit,  $\Delta t_m$  — duration of  $m$ -th period of relatively stable development of a language. According to this method it is possible to use existent vocabularies of the languages.

All the work may be accomplished by means of computers. The advantages of the suggested solution of world language problem are as follows:

1. as averaging will be done on the basis of ethnic languages (dialects), native speakers of those languages (or dialects), averaged language will be built on, are sure to understand it to some extent without special learning;

2. belonging to none ethnic community the averaged language gives no privilege to either of them, so it will not promote national discord based on the language policy;

3. the averaged language eliminates some arbitrariness in choosing one of the local languages as official state language as well as interethnic conflicts;

4. the averaged language eliminates the introducing of languages of former colonizers as the only state official language, weakens the dependence on the former parent state in the spheres of culture and education.

5. many nations speaking the same language are divided by state borders. Thus an averaged language constructed by the offered method might play the role of macromediator.

The problems of ethnic consolidation of various dialectal groups are mainly connected with the degree of their propinquity or closeness. The degree of closeness when the native speakers of dialects feel themselves speaking the same language is very different, i.e. in some cases the native speakers do not feel themselves relating to one language although there is an objective linguistic closeness of dialects not realized by the speakers.

One more advantage of averaged languages may be brought to the light here – averaged language may be a rescue for endangered languages many of which are cognate minor languages or dialects.

The aforementioned method to define the degree of synchronic closeness of dialects by means of calculating the distance between the dialects (languages)<sup>12</sup> utilizes the vector space. Measured closeness of dialects (languages) when known by people will certainly influence the course of ethnic processes because the degree of such affinity has though not decisive but a great significance for the peoples who speak one of those languages as a vernacular one. Emphasizing the linguistic closeness can promote the persons' understanding of their belonging to the more wide human community. Finding out of the degree of the closeness of dialects and languages within the group may give to the persons the objective base for deep understanding of their language identity and ethnic identity as well, and promote the process of ethnic consolidation this way.

The averaged language under discussion must be worked over by a representative group of experts in the fields of each language that is in the process of averaging (or we may call the process neutralization). It is necessary to form the group of experts according to the native speakers will. Every language should be represented by equal number of experts to guarantee the equality in processes of voting. Membership of each delegation must guarantee the balanced, complex and systematic representation of different aspects of studying and usage of the languages.

The group of experts must be transformed to special research institute under the aegis of international organizations or special foundations.

The institute will to organize:

- the advanced study of languages and dialects, and of language situation in the area of group of languages in question with particular attention to sociolinguistic, psycholinguistic, and ethnolinguistic aspects of investigation;

- the discussion and establishment of the criteria that guarantee the optimal neutrality of the averaged language being constructed in a group of related languages. But the finishing the

---

<sup>12</sup> See footnote 2.



work should not be delayed due to ascribing absolute value of neutrality criterion and linguistic perfection of the averaged language. In the present conditions the offered method is a possibility.

It is necessary to submit the projects for discussion by native speakers of every language that form the basis of the averaged language and take into consideration all remarks and proposals for improving it. The improved project must be published and after discussion and approving by the highest state body can be introduced for usage to spread information with the aim to develop value orientations and positions regarding studying and usage of the averaged language:

- at first stage to introduce the averaged language as an auxiliary language of interethnic communication between native speakers of languages on the basis of which the averaged language had been constructed,

- to enlarge the sphere of usage of the averaged language through the mass media;

- to start to render humanitarian information, especially the treasures of world literature to the averaged language, but one must not aim at making the averaged language as rich as English or other world languages;

- to introduce the averaged language into education. Each people may define the degree of possession of the averaged language. The reasonable degree is that that guarantees the communication on the problems of common significance and professional communication;

- the averaged language being intended for interethnic communications, must not deprive any of basic languages of its spheres of use. In the case of depriving the balance the situation of neutrality will be destroyed.

The empirical reasons provided to justify the above method of an averaged language development can be generalized. Averaged language for a group of genealogically kin languages is a complex object and its systemic modeling as a totality is required. The process of an averaged language developing is a complex process as well, and that is why systemic analysis and synthesis of organization and control and management of the averaged language development, spread, and usage process. One of necessary conditions for achieving averaged languages modeling systematicness is that the measure of involving material from each language of the group under averaging is of the same qualitative and quantitative order. This aspect of the averaged languages is reflected in the development process structure of such languages expediency requirement to achieve equal, including equality of rights, participation of linguists' groups representing each of kin languages.

Let us to consider three options organization of one of the averaged languages development: 1) in a form of Institute of Averaged Languages with invitation of linguist scholars representing each of cognate languages of respective group to location of the Institute; 2) in a form of an association of separately working group of linguists, each in their own country or region, built of one linguist from each kin language with establishment of a coordinating centre and a central institute synthesizing the results; 3) geographically "dissociated" body of linguists – experts in each language working wherever in the world and doing averaging related work employing Internet technologies. The first option may be called concentrated work form, while the second one may be called diffuse form, and the third may be called absolutely diffuse form.

The combinations of the work forms are possible, too: 3) in a form of association of a number of the regional linguists groups where each one represents several geographically close located kin languages with central institute in place; 4) in a form of an association of a number of linguists groups each of which represents separate national literature language and a number of regional linguists groups where each one represents several closely located kin languages, with central institute. At all diffuse and combined versions establishment of a computer network is reasonable, and to equip the central institute with a specialized computer (a server), or the system of automated work places, to work on language averaging may prove to be reasonable. Work mood with utilization of Internet technologies will require, of course, to undertake security measures, to establish levels of access authority, employment of electronic signature, back-up bases to recover in a case of hackers' attack. The server may be set up at any place but it is

expedient to have it at the central institute for the system administration reasons. The combined versions allow actively utilise positive aspects and mitigate negative ones both of concentrated and diffuse forms.

In relation to the development of averaged neutral languages for related languages a question emerges on perspective of the development of the communication between native speakers of non-related languages in the worldwide, global communication among all peoples.

We assume that the world auxiliary language may be created the same way as the averaged language for a group of related languages and it will be the next step on the way of neutralization of languages. The creation of a world auxiliary language is important for solving the problem of information barriers as well as overcome information colonialism, and necessity to achieve equality of the peoples in condition when the so-called club of world languages exists, and looking forward to establishing the new information order. This urgent global problem can be solved by international group of experts under the UN aegis. In the creation of the artificial world language, it is advisable to take into consideration the facts of all the languages, it is neutrality that is the main feature of the averaged language.

In a case if the world community of nations agrees to create the world language which is to substitute for all national languages in farther future, it is possible to organize the work under the aegis of the UN but it must not oppose the work of creation of the world auxiliary international language, because the latter is more necessary. The world language must be fundamentally elaborated with taking into consideration all nuances and details of all the world languages. Now the transition of the humankind to one language is necessary because it would accelerate the development of human society. But the accelerated development cannot guarantee the survival of the accelerating community. From the point of view of self-conservation, reliability of existence and stability of the humankind it is expedient for it to exist in a form of community of ethnosocial organisms with equal rights, and which realizes the communication between them by means of the world auxiliary language of international communication.

The ideas to establish averaged languages for the groups of kin languages lead to the necessity to consider problems of their teaching in connection with teaching one of the other languages as the first language. For instance, in the case of Turkic languages and Language *Ortaturk* an opinion is put forward sometimes that simultaneous teaching of vernacular Turkic language and the averaged Turkic language will result in waste of strength and energy of the Turkic peoples and extinction of the attention to the Turkic languages of minor peoples. This opinion may be challenged, and the consideration will be true for other language groups and their averaged languages.

The language *Ortaturk* exemplifies averaged languages and we may discuss the above issues talking about this language. The language *Ortaturk* will be closely kin language for the majority of the Turkic languages. It will be capable of studying in addition to study of a native Turkic language. Each Turkic nation will be able to choose the forms, methods, and level of language *Ortaturk* study herself proceeding from her national and state interests. The language *Ortaturk* will be a voluntarily chosen language of inter-ethnic communication, accumulation of information of all-Turkic and world significance in the community of Turkic nations.

Each national Turkic language will have the status of a state language and will develop full scale in its nation-state, in its ethnic territory.

The language *Ortaturk*, when functioning, will benefit strengthening attention of the Turkic languages native speakers (especially those minor by number of bearers) to their own languages giving basis for hope and belief in usefulness and prospects of their native languages for access to the world culture. Democratic nature of the development and functioning procedures of the language *Ortaturk*, their compliance with shared human norms of language relations leads to voluntary learning of the language by majority of Turkic nations and Turkic speaking persons along with their native Turkic languages. Usage of the language *Ortaturk* results not in wasting of vital forces but in saving, optimization of usage of those both with each Turkic nation and their whole system.

For meeting the understanding and communication goals the averaged Turkic language *Ortaturk* will not require especial learning for many Turkic peoples since it will be very close to them. Bearers of a number of other Turkic languages will need some additional learning, first of all in the aspect of assimilation of the changed words usage frequency and grammatical forms difference.

For the purpose of practical introduction of the language *Ortaturk* an enthusiast group will be needed to be taught to high level of mastering norms system of this language. Members of the group could specialize on implementation of a number of functions in the media: announcers of the all-Turkic TV and radio in the language *Ortaturk*; editors of newspapers, magazines and books published in this language; teachers of the language *Ortaturk* teaching through radio, TV, computer information systems (distant e-learning). Frequently repeated auditory and visual perception of information transmitted in the language *Ortaturk* via media will result in passive mastering the language by the majority of the Turkic languages bearers and enhancement of the mutual intelligibility level of the bearers of different Turkic languages even while they will actively use only their native Turkic languages. In a while a significant part of the bearers of Turkic languages who often happen to be in the situation of inter-Turkic language communication will develop skills of some transformation of their respective idiolects in the direction of an array of basic norms of the language *Ortaturk* in the respective language situations.

The Uzbek people has additional development possibilities because it belongs to the Turkic group of peoples and languages that have big internal self-development potential. This condition opens wide prospects for independent development of each Turkic nation and its language, including the Uzbek nation and its language. The Uzbek people and many other Turkic peoples have reached independence and their languages have acquired the status of state languages. The application sphere of these languages cardinaly widened. They become languages of vocational, secondary, and higher education, languages of science, culture, the whole economic and social and political life.

In this respect the systems of scientific terms are being created in many Turkic languages but, unfortunately, without coordination with each other. In order to overcome this trend leading to loss of mutual intelligibility of the kin languages establishment of coordinating terminological system in the area of scientific terms is proposed.

The method of establishment of coordinating terms system is based on a mathematical procedure of choosing an optimal term for each scientific expression language from the point of view of the term existence criteria for the notion: 1) in majority of the Turkic languages; 2) with majority of Turkic speaking persons; 3) within the longest period on time; 4) the most semantically close. These four criteria are taken in various correlation combinations among them with help of certain system of coefficients in mathematical formulae applied for development of averaged languages. In the interpretation of these formulae, the notion of “term of coordinating terms system” should be used instead of notion “the word of the averaged language”.

The basic stages of this terms system development are: 1) selection of the most frequently used and important scientific notions, compilation of word-list in the basic language; 2) development of applications for computer aided processing of Turkic language terms stock and coordinating terms system; 3) compilation of computer stock and terms used for expression of these notions in Turkic languages; 4) data processing and obtaining results.

The following dictionaries are supposed results:

1. Russian - English - Uzbek - Azeri - Kazakh - Kyrgyz – Turkmen – Turkish – Tatar – Bashkir – Uigur – Karakalpak – Chuvash – Yakut – other Turkic languages – coordination terms system terms;

2. analogous Uzbek - Russian - English - Azeri – etc dictionaries (the first language rotates).

These dictionaries having been published could be used in the area of Turkic peoples terms systems development. By analogy, coordination terms systems for other groups of kin languages

can be developed, too. In perspective, this will foster mutual cooperation of their scientific, intellectual potentials at solution of complex problems facing them and the world community as a whole.

Closeness of spiritual world, and that of folklore, in particular, plays an important role in strengthening the cooperation between the peoples.

Outstanding epic pieces of the Turkic peoples, each having unique artistic and spiritual originality, have number of aspects common for epics of all the Turkic peoples, too, which seem expedient to be studied.

Eposes of the Turkic peoples, as well as the eposes of many other peoples of the world, most often exist in several versions. There is a probability that acceptance of one version as a main one and its introduction into teaching and education process nation-wide may be understood as ignoring other versions. But learning by teaching all versions may weaken each epos' potential as means spiritual education. With this regard we think expedient to create averaged version for each of the epos.

Usage of the following method is expedient in the averaging process. Plot of each version is divided into a number of similar elements and parts. From among these elements those components that are present in all or the majority of the versions are included into averaged version. The averaged version having been compiled epos narrator retells the epos in improvisation style in accordance with averaged plot, and thus averaged epos version is formed.

In the process of such averaging utilization of the mathematical method employed at averaging kin languages is expedient<sup>13</sup>.

Development of the averaged language *Ortaturk* that fulfils the functions of voluntarily recognized language of the neutral language of inter-national communication and language of gathering of information of all-Turkic and all-humankind significance would promote development of mutual understanding and equal cooperation between the Turkic peoples, social and economic, political and legal, information and communication, professional qualification, spiritual and intellectual development of persons and social groups both of Turkic world and the whole humankind.

A number of outstanding pieces of the Turkic peoples are a spiritual, cultural heritage of all-Turkic and all-humankind scale. They contain enormous potential of shared humanistic spiritual values. Their humanism is especially boldly manifested in the images expressing struggle against cruelty in relations between people. Such eternal philosophical world outlook problems as correlation of good and evil, meaning of human life, connection between a human and native nature, inter-connection between death and immortality, transient and eternal in human life, and a number of other problems are given their specific humanistic solutions in spiritual and intellectual, artistic and figurative context of a number of the Turkic peoples eposes.

The Turkic folklore artistic and figurative world contains many images that personify their fraternity, calling to friendship, fidelity to the ancestors' memory, cooperation and mutual assistance for the benefit of humankind progress.

If averaged plot versions of eposes (and other folklore pieces) widely spread with the Turkic peoples and these plots are expressed by means of the language *Ortatutk* averaged Turkic folklore in the language *Ortaturk* will shape. In analogy, other averaged languages and averaged folklores can be created for other language groups: averaged Romance, averaged Germanic, averaged Indian, averaged Slavonic, averaged Iranian, etc. This will foster mutual spiritual enrichment of the peoples both within each genetically cognate languages, kin group of peoples, and the whole community of peoples within the framework of the whole mankind.

---

<sup>13</sup> See: Каримов Б.Р., Муталов Ш.Ш. Уртатурк тили. Тошкент: Мехнат, 1992 (Karimov B.R., Mutalov Sh.Sh, Ortaturk tili. Toshkent: Mehnat, 1992); Они же. Проблемы создания среднетюркского языка ортатурк // Ўртатурк тилини яратиш муаммолари. Тошкент, 1993 (the same authors. Problemy sozdaniya srednetiurkского yazyka. In: Ortaturk tilini yaratish muammolari. Toshkent, 1993); Karimov B.R., Moutaloff Sh.Sh. Averaged Languages: An Attempt to Solve the World Language Problem. Tashkent, 1993.

Let us consider the second approach to the language policy aimed at creation of all-Turkic language. This approach is more traditionalistic one, and, as we see, can secure, mainly, mutual acquaintance between close kin mutually intelligible languages. In this approach, it is difficult to take into consideration more distant Turkic languages such as Chuvash, Yakut, Khakass, Tuvinian, Altai, Dolgan. And it is difficult to take into account the languages of small Turkic peoples in accordance with the equality status. The approach bears on the presumption as if of difference between the modern Turkic languages has a dialectal nature, of possibility for them to gradually merge in a single literature all-Turkic language. We think such a possibility existed as a real one for the majority of Turkic ethnic groups in the beginning of 20<sup>th</sup> century but in the course of the century the majority of Turkic ethnic groups grew into nations and by the beginning of 21<sup>st</sup> century this possibility turned into abstract, formal one. It is not possible to ordinarily shape the all-Turkic literature language relating to other Turkic language as to dialects by democratic way. The language norms of the Turkic languages have diverged so much that it is difficult now to imagine emergence of figures of genius in a group of Turkic cultures whose idiolects would be so close and capable of being merged by usual, traditional way, and spiritual and cultural values and norms created by the latter in their versions of Turkic languages would eclipse for the latter by language values created by the former, and voluntarily accepted by the latter. The real possibility by the end of 20<sup>th</sup> century was and in the beginning of 21<sup>st</sup> century is to develop the language *Ortaturk* by the way of equal averaging of the Turkic languages with help of mathematic, computer aided methods and their preservation of them as the state languages in their respective national Turkic states.

In the all-Turkic language creating concept, it is presupposed that it will oust all other Turkic languages in final account and remain the only one supplanting all the Turkic languages.

The first approach, i.e. creation of *Ortaturk* leads to equality and friendly relations while the second one, i.e. creation of all-Turkic language leads to violation of equal rights and the right of a nation for self-determination and independent development of the people, its language and culture, and, in the final account, the second way leads to complication and aggravation on international relations. In this respect the first and the second way possess differences of principle in the aspect of the consequences for development of the inter-national relations between modern Turkic nations, shaping and development of the Turkic civilization.

The great figures of the Turkic peoples struggled for preservation of fraternity, cooperation, and mutual assistance of the Turkic peoples.

The lives and creative work of Mahmud Koshghari, Khodja Ahmad Yassawi, Alisher Navoi, Muhammad Fizuly. Ismail Ghaspirali, Abay Kunanbayev, Abdulla Tukay, Ziya Giok Alp, Mahmudkhodja Behbudi, Abdurauf Fitrat, Abdulla Kadiri, Zaki Validi Togan, Mustafa Kemal Atatürk and many other our great thinkers that sought renaissance and development of culture, science, arts, and education in Turkic language and considered that the basis for prosperity of the Turkic peoples lays in their unity, cooperation, and mutual assistance. They thought strive of the Turkic peoples for strengthening the language and cultural unity to be rightful. They struggled for independence and prosperity of Turkistan, the deserved place independent Turkistan had to take in the world system of political, economic, and national and cultural relations. Their thoughts, expressed in the past, keep on having immediate relation with the modern fundamental social transformations in the world.

National self-consciousness and national languages of many Turkic peoples have formed, and thus the option of direct merge into one nation with one language is not feasible in the modern conditions though it was possible in the beginning of the 20<sup>th</sup> century. We propose slightly different way of resolution of these problems in the modern conditions. We think expedient development of the averaged Turkic *koine* type language *Ortaturk* by means of utilization of dialect and kin languages averaging method.

Attempts to create a common language for the communication of the Turkic peoples are undertaken by a number of research groups. In particular, in Turkey this work is carried out under the name “The Problem of Developing Language “Ortak Turk Dili”, i.e. a common Turkic

language. Each group has its specific approach to the solution of the problem. Difference of the approach where development of the language *Ortaturk* is substantiated from other approaches is that this approach proceeds from equal rights of the Turkic languages, and an attempt is made to employ a mathematical averaging method securing consideration of all Turkic languages.

Some critics of the language *Ortaturk* development concept attach absolute value to the element of artificiality present at development of the language. They talk of impossibility to create and use of the language *Ortaturk* because, they say, it will have the fate as having no prospects, as they think, as Esperanto.

Let us to consider this stand and arguments in its favour. First of all, let us say a word on the fate of Esperanto.

Near 5 million people<sup>14</sup> use the language Esperanto in many countries of the world. About 100 reviews and magazines are published in Esperanto on various branches of science, culture, education, and production. Scientific literature and literature of art exist in Esperanto. And this literature is not only translations but original literature of art both prose and poetry. The language Esperanto did not realize in the function of a world language only giving way to the languages from the “club of world languages”: English, Russian, Spanish, French, Chinese, Arabic, German, Portuguese, etc.

Esperanto gives way to about a hundred of other national languages as the level of the language development is concerned, but it is more developed as compared with the rest almost three thousand languages of the world peoples. Esperanto is more largely present in the world wide web than, for instance, Uzbek with its estimated 27 plus million native speakers. Popular Wikipedia had about 100 thousand articles in Esperanto while it featured less than 6 thousand articles in Uzbek by the middle of 2007, and more than in many other languages with native speakers over 5 million. That is why we can talk about prospects of the language Esperanto with regard of its ability to compete with natural national languages. And Esperanto’s defeat in its dispute with the national languages from the “the world languages club” is due to their former chauvinistic, imperialistic domination that allowed them forcefully to gain a foothold in enormous space of the ethnic territories of other peoples and among huge massifs of people with different national, confessional, world outlook, racial, and language identity.

In the future, information and language colonialism will be eliminated and new democratic information order will firmly establish itself in international and inter-state relations securing equal cooperation in the world between the nations and their equality in rights and peaceful and tolerant co-existence. In such a situation the role of artificial languages is to increase. The role of averaged languages that combine features of artificiality and naturalness will be especially important, as we believe.

Let us to consider the correlation between artificiality and naturalness in averaged languages.

Impetuous development of the science and technology, informatization and algorithmization of many aspects of the social life is a fundamental peculiarity of the present epoch. Within the framework of these processes, artificial and algorithm languages are widely used. An array of new problems emerge in epistemology such as, for instance, correlation of natural and artificial, algorithmic languages in cognition, algorithms of cognitive process, correlation of computer and dialectical logic, role of human-computer dialogue intellectual systems in the cognition process, social determination of cognition and present day informatization processes. Problems of the Republic of Uzbekistan’s, or more largely, Central Asia’s, all new independent states’, optimal integration into world informational, scientific, cognitive conceptual, apprehensive space and time.

Essential trait of a human being and its activities is that they are mediated by cognition process, thinking, and creation. At the same time, there is one more essential trait of humans, and indissolubly correlating with the aforementioned, it is human communicative nature, human

---

<sup>14</sup> This is estimation, and estimations vary largely depending on understanding of ‘uzage’ and ‘knowledge’ of the language.

intercourse in a *socium* by means of a language. For a long time, natural, national, or “world” language has been understood in this respect. However, this issue needs more careful study due to the processes of creation of artificial languages and definition of their relation with natural ones.

Mathematic simulation applied to development of the averaged languages system allows mimic algorithmic systemic simulation of the objective *koine* formation process in the course of kin languages bearers’ communication. Certain concord of natural and artificial languages takes place in this method.

The following peculiarities are features that bring averaged languages close to the natural languages:

1. there are no words, morphemes, phonemes and other linguistic units and norms that would be invented subjectively, with voluntarism; norms that are the most frequently used among the norms of natural languages are chosen according to four criteria of existence: a) with the majority of national languages; b) with the majority of individuals; c) during the longest period of time; d) with the most semantic closeness to the semantics of the norm sought.

2. The basis of an averaged language exists potentially as a historically shaped group of genealogically cognate languages like the basis of a “natural” national language exists in the form of its dialects. A national language is entitled to exclude its own dialects from the main basic spheres of the social life, and averaged language is not entitled to do so with regard of national languages. The national languages must be state languages in their respective national states.

The following peculiarities are traits that bring averaged languages close to the artificial ones:

1. An averaged language is not shaped completely historically, it is shaped by means of strong modern norm setting human interference;

2. It is artificially shaped unique combination of natural components, that is why an averaged language has the feature of emergence<sup>15</sup> and in is reasonable to take this feature into consideration;

3. Artificial synthesis of natural components is carried out by artificial means through utilization of mathematical simulation and computer aided data processing although these models are aimed at mimic simulation of the languages internal development processes with consideration of their socio-linguistic functioning;

4. In the moment of averaged language creation, historically shaped group of its bearers<sup>16</sup> does not exist actually although it exists in potential;

5. Averaged languages like other artificial languages spreads by means of purposely artificially established subsystem of the media and at the same time it bears on cognateness of the averaged language with natural languages of the basic part of its audience;

6. Averaged languages do not have many exceptions from the rules in their structure that emerged in natural languages both in the course of multilateral external distorting influences by other languages and in the shape of substratum languages and earlier stages relicts. Due to that norm of averaged languages are characterized by larger degree of universality, idealization, smoothness as compared to the norms of natural languages and with this respect they are closer to the norms of artificial languages;

7. The averaged language will be “supernaturally” rich because it will enrich in a certain way on the account of all national cognate languages and for this reason in accordance with the law of necessary diversity<sup>17</sup> both the averaged language and the system of national languages

---

<sup>15</sup> See: Ashby W. Ross. An Introduction to Cybernetics. L. NY: Methu & Co LTD, 1964, p.p.110-111; 127-128.

<sup>16</sup> Normally, the term would be “native speaker” but no one is ‘native speaker’ of an averaged language neither now no at introduction of one in the beginning. So, we preferred the term ‘bearer’ which is calque from Russian “носитель”

<sup>17</sup> See: Ashby W. Ross. An Introduction to Cybernetics. L. NY: Methu & Co LTD, 1964, p.p.214-219, 253-258.

from which it is shaped may act reciprocally in the function both of governing system and governed one in the course of their interacting development;

8. Averaged languages and the system of natural languages which they are formed from as internally differentiated self-organizing systems will be characterized by high level of homeostasis;

9. In the course of averaging system algorithmization, models monotony principle is applied because averaging at different language levels is carried out according to one and the same mathematical model. And principle of solution choice freedom is satisfied as well because averaging at phoneme level keeps freedom of choice at morpheme level, and so on for each consecutive level.

The intellectual, spiritual potential of human communities and the whole Humankind depends to the great extent on their information potentials. The information potential depends on the system of information exchange in the society, on the structure of information space and time. Multilingualism of Humankind, being inevitable, historically formed, and necessary for its sustainable self-preservation at the same time complicates solution of the most acute global world problems faced by the world civilization. For securing unity of Humankind preservation of its diversity, it is expedient to create the system of averaged languages for the groups of cognate languages, and to average them in their diversity based on Nostratic (Borean) concept, concept of the language universalia, and statistical methods of averaging of the language phenomena. The world international communication and world information accumulation auxiliary language created this way would promote solution of many global problems and mutual spiritual enrichment of all the peoples of Humankind.



# УСРЕДНЕННЫЕ ЯЗЫКИ: ПОПЫТКА РЕШЕНИЯ МИРОВОЙ ЯЗЫКОВОЙ ПРОБЛЕМЫ

## (РЕЗЮМЕ)

В эпоху формирования единого человечества и интенсивного национально-этнического развития народов, когда возникли новые возможности для их сближения в областях экономики, политики, культуры, науки, образования, языка, духовности и во многих других сферах, предлагается создать систему нормированных усредненных языков, которые способствовали бы данному процессу сближения.

Предлагаемый нами метод создания усредненных языков для диалектов и групп родственных языков (тюркских, иранских, славянских, романских, индийских, и др.) основан на математической процедуре усреднения на каждом из уровней языка (фонемном, морфемном, лексическом, синтаксическом) посредством формул (1), (2) и (3) (для синхронии) и (4), (5) (для диахронии), в которых используются следующие обозначения:  $X_{n,1}^j$  - компоненты вектора в векторном пространстве соответствующего уровня;  $n$  - номер единицы уровня в списке единиц уровня;  $\alpha$  - номер языка в списке усредняемых языков;  $j$  - номер варианта единицы уровня, имеющей номер  $n$ ;  $s$  - общее число таких вариантов для одного языка;  $F_{n,\alpha}^j(k)$ ,  $F_n^j(k, K\alpha)$ ,  $F_n^j(k, K\alpha, \beta, \gamma)$  - функции, вводимые для оценки степени распространенности и понятности  $i$ -того варианта  $n$ -ной единицы уровня;  $K\alpha$  - численность носителей языка  $\alpha$ ;  $\bar{K}$  - средняя численность носителей в группе усредняемых языков;  $\beta$  и  $\gamma$  - параметры, служащие для обобщения формул (1) и (2);  $m$  - номер периода в последовательности периодов времени, на которые разбито время развития усредняемой группы языков; вышеуказанные обозначения при добавлении к ним индекса  $m$  обозначают эти величины, взятые в соответствующий период времени;  $\Delta t_m$  - продолжительность этого периода времени;  $\Delta t$  - продолжительность наименьшего из таких периодов.

Метод построения усредненного языка мы описываем, пользуясь примерами из лексики тюркских, славянских, иранских и романских языков. Для лексики следует выбрать определенной длины верхнюю часть частотного словаря одного из языков - базисного. Список пронумеровывается. Для каждой лексемы упорядоченного списка находятся эквиваленты в языках или диалектах, на базе которых строится усредненный язык. Эквиваленты базисной лексемы представлены в таблицах 2, 3 и 4, которые затем преобразуются в таблицы 5, 6 и 7, где в одной строке располагаются однокоренные слова - эквиваленты из разных языков. Назначение коэффициента  $1 + \frac{K_\alpha}{\bar{K}}$  - учет относительной численности носителей языка из группы языков, для которой строится усредненный язык. Слова-эквиваленты, соответствующие каждой лексеме базисной выборки, располагаются в порядке убывания значений функции  $F_{n,\alpha}^j(k)$ . Слова, расположенные в указанном порядке, приведены в таблице 9. Корни, первые в списках, включаются в словарь усредненного языка в качестве основных значений слов. Остальные слова составляют своеобразный запас синонимов для обогащения лексики усредненного языка.

Преимущества предлагаемого решения языковых проблем заключаются в том, что:

1. будучи построено на основе языков этносов, усредненный язык в той или иной мере понятен носителям тех языков (диалектов), из которых он построен;
2. не являясь языком какой-либо этнической общности, усредненный язык не дает преимущества одной из них и, следовательно, не будет способствовать усилению розни по поводу языковой политики;

3. устраняет произвол, имеющийся при выборе одного из местных языков в качестве общерегионального, и вытекающие из такого выбора межэтнические трения;

4. усредненный язык будет способствовать равенству и равноправию наций и их языков (диалектных групп и их диалектов), а также установлению справедливого информационного порядка в мире.

По соглашению всех родственных народов, языки которых были объединены в усредненном языке, для усредненного языка может быть принят один алфавит или он может функционировать в нескольких алфавитах одновременно. Для создания усредненных языков необходимо организовать международный институт с филиалами в научных центрах соответствующих языков. Усредненные языки предлагается использовать как языки межнационального общения родственных народов, а также накопления информации регионального и мирового значения. При этом национальные родственные языки будут оставаться государственными языками соответствующих государств и далее свободно развиваться. Это будет служить обеспечению равноправия народов и их языков, что очень существенно в укреплении доверия и взаимопомощи между ними в их многоаспектном сближении. Демократические формы этнолингвопанализмов (пантюркизм, пангерманизм и др.) не противоречат международным нормам прав наций и человека и являются общечеловеческой ценностью. Предлагается обобщить вышеуказанный метод и создать усредненный язык всемирного международного общения.

## MITTELSPRACHENBILDUNG ALS EINER DER VERSUCHE WELTSPRACHENPROBLEM ZU LÖSEN

### (ZUSAMMENFASSUNG)

Im Zeitalter der Gestaltung von der einheitlichen Menschheit und der intensiven national-ethnischen Entwicklung von Völkern entstehen neue Möglichkeiten ihrer Annäherung auf den Gebieten der Wirtschaft, Politik, Kultur, Wissenschaft, Ausbildung, Sprache und Geistigkeit, sowie in anderen Branchen. Hiermit wird es vorgeschlagen, ein System der normierten Mittelsprachen zu bilden, das diesem Annäherungsverfahren dienen könnte.

Das von uns vorgeschlagene Mittelsprachenbildungsverfahren von Dialekten und Gruppen der verwandten Sprachen (türkische, iranische, slawische, romanische, indische, u.a.) beruht auf dem mathematischen Mittelwertbildungsaktion durch die Formel (1),(2), (3) für die Synchronie und (4), (5) für die Diachronie für jeder der Sprachebenen (Phonem-, Morphem-, Lexik-, Syntax). In den Formeln werden folgende Bezeichnungen gebraucht:  $X_{n,\alpha}^j$  ist eine Vektorkomponente im Vektorraum,  $n$  – Nummer der Ebenenmaßeinheit in der Liste der Niveauseinheiten,  $\alpha$  – Nummer der Sprache in der Mittelsprachenliste,  $j$  – Nummer der Variante von der Ebenenmaßeinheit, die die Nummer  $n$  hat,  $s$  - die Gesamtzahl von solchen Varianten für eine Sprache,  $F_{n,\alpha}^j(x)$ ,  $F_{n,\alpha}^j(x, K_\alpha)$ ,  $F_{n,\alpha}^j(x, \beta, \gamma)$  Funktionen, eingeführt für die Schätzung der Zahl und Begreiflichkeit von der  $i$ -Variante für eine  $n$ -Ebenenmaßeinheit,  $K_\alpha$  – Zahl der Sprachträger der  $\alpha$ -Sprache,  $\bar{K}$  – Mittelzahl der Sprachträger in der Gruppe von der Mittelsprachen,  $\beta$  und  $\gamma$  - Bestimmungsgrößen für die Verallgemeinerung der Formel (1) und (2),  $m$  – Nummer der Zeitperioden in der Reihe von Zeitperioden, in die die Zeitentwicklung einer Mittelsprachengruppe geteilt ist. Ist Index  $m$  zu den o.g. Bezeichnungen zugefügt, so bezeichnen sie diese Größe in einem bestimmten Zeitraum,  $\Delta t_m$  – Zeitdauer dieser Zeitperiode,  $\Delta t$  - Dauer der kleinsten von diesen Perioden.

Wir beschreiben die Methode der Mittelsprachenbildung, Beispiele aus der Lexik von Türkischen, Slavischen, Iranischen, Romanischen Sprachen benutzend. Für die Lexik ist eine bestimmte Länge des oberen Teils von Häufigkeitswörterbuch einer der Sprachen zu wählen, u.n. der Grundsprache. Diese Liste sei zu nummerieren. Für jedes Lexem der o.g.

durchnummerierten Liste sind Äquivalente in den Sprachen oder Dialekten zu finden, auf deren Grund die Mittelsprache zu bilden ist. Die Grundlexemäquivalente sind in den Tabellen 2, 3 und 4 dargestellt. Dann sind sie in die Tabellen 5, 6, und 7 transformiert, in denen die Einwurzelwörter-Äquivalente aus verschiedenen Sprachen in einer Zeile geordnet sind. Die Zweckbestimmung des Koeffizients  $1 + \frac{K_{\alpha}}{K}$  gilt als Registrierung der relativen Sprachträgerzahl

aus der Gruppe von den Sprachen, für die die Mittelsprache gebildet wird. Die Wörter-Äquivalente, die jedem Lexem der Grundabnahme entsprechen, werden auf dem Abnehmen-Weg der  $F_{n,\alpha}^j(\bar{k})$  Funktionbedeutung geordnet. Die nach solcher Reihenfolge geordneten Wörter sind in der Tabelle 9 angeführt. Die Wurzel, die als ersten in den Listen stehen, werden als Grundbedeutungswörter ins Wörterbuch eingetragen. Die anderen Wörter bilden eine eigenartige Synonymenreserve, um die Lexik der Mittelsprache zu bereichern.

Die Vorzüge der vorgeschlagenen Lösung von Sprachproblemen bestehen in Folgendem:

1. Die Mittelsprache, gebaut auf dem Grund der Ethniesprachen, wird soweit möglich den Sprachträgern verstehen sein, von denen Sprachen (Dialekten), sie gebildet ist;
2. Die Mittelsprache als eine Sprache von keiner ethnischen Gemeinschaft wird kein Privileg einer von denen geben und auf solche Weise wird sie den sprachlichen Zwist beilegen helfen;
3. Die Mittelsprache behebt die Willkür, die bei der Auswahl einer der Lokalsprachen als allgemein regionale vorlegen können, sowie die bei solcher Auswahl entstehenden zwischenethnischen Konflikte;
4. Die Mittelsprache fördert, Gleichheit und Gleichberechtigung von Nationen und ihren Sprachen (Dialekten) zu erreichen, sowie die richtige Informationsordnung in der Welt festzustellen.

Laut der Zustimmung aller verwandten Völker, derer Sprachen in einer Mittelsprache vereinigt waren, kann ein Alphabet für eine Mittelsprache akzeptiert werden. Sie kann auch zu gleicher Zeit in einigen Alphabeten fungieren. Für die Bildung der Mittelsprachen scheint es notwendig zu sein, ein Internationales Institut mit Filialen in den wissenschaftlichen Zentren der entsprechenden Sprachen zu gründen. Die Mittelsprachen sind als Sprachen für die zwischennationalen Kommunikation der verwandten Völker zu benutzen, sowie für die Akkumulation der Information von der regionalen und Weltbedeutung. Dabei werden die verwandten Nationalsprachen als Staatssprachen der entsprechenden Staaten bleiben und sich weiter frei entwickeln. Das läßt die Gleichberechtigung der Völker und ihrer Sprachen sichern, was ihr Vertrauen und gegenseitige Hilfe bei der aspektenreichen Annäherung wesentlich festigen hilft. Die demokratischen Formen der Ethnolinguopanismen (Pantürkismus, Pangermanismus u.a.) stehen zu den internationalen Rechtsnormen der Nationen und Menschen im Widerspruch nicht und gelten als die allgemeinmenschlichen Werte. Es ist vorgeschlagen die o.g. Methode auszuwerten und eine Mittelsprache für die Internationale Weltkommunikation zu bilden.

## **Des langues moyennes : une tentative de résolution du problème mondial de langue.**

### **Résumé**

l'époque de la globalisation et du développement intense des nations et des ethnies apparaissent de nouvelles possibilités de rapprochement économique, politique, scientifique, culturel et spirituel. Nous proposons de créer un système de langues moyennes qui puisse contribuer à ce processus de rapprochement.

Notre système de création de telles langues est valable pour les dialectes et les langues soeurs à l'intérieur de chaque groupe de langue : (turc, persan, indien, slave et autres) est basé sur la procédure mathématique visant à rendre moyen chaque niveau de langue : phonologique, morphologique, lexicale, syntaxique. Il s'agit de l'application des formules (1), (2), (3) opérant au plan synchronique et (4), (5) -au plan syntaxique . On emploie les désignations suivantes :  $X_{n,1}^j$  - composant du vecteur dans l'espace vectoriel du niveau concerné ;  $n$ -le numéro de l'unité dans la liste des unités ;  $a$ -le numéro de la langue dans la liste des langues moyennes ;  $j$ -numéro de la variante de l'unité du niveau ayant le numéro  $n$  ;  $s$ -nombre total des variantes pour une langue ;  $F_{n,\alpha}^j(x)$ ,  $F_n^j(x, K\alpha)$ ,  $F_n^j(x, K\alpha, \beta, \gamma)$  - fonctions introduites pour l'évaluation de la fréquence de l'emploi et de compréhensibilité de la variante  $i$  de l'unité  $n$  ;  $K\alpha$  - nombre de gens parlant la langue  $a$  ;  $\bar{K}$  -le nombre moyen de gens parlant la langue du groupe de ces nouvelle langue communs ;  $\beta$  et  $\gamma$ -les paramètres servant à généraliser les formules (1) et (2) ;  $m$ - le numéro de la période dans la succession des période de laquelle est divisé le temps du développement du groupe des langues rendues moyennes ; les désignations ci-dessus indiquées avec l'adjonction de l'indice  $m$ , signifient ces quantités prises en une période définie ;  $\Delta t_m$  -durée de cette période de temps ;  $\Delta t$  - la plus courte durée de ces périodes.

Nous décrivons la méthode de création de la langue moyenne en utilisant le lexique des langues turques, slaves et romanes. Quant au lexique, il est préférable de choisir une partie les mots les plus fréquents dans le dictionnaire de fréquence d'une des langues de base. La liste est numérotée .Pour chaque lexème appartenant à la liste réglée, il existe des équivalents dans les langues ou les dialectes donnés à la base desquels on construit la langue moyenne. Les équivalents des lexèmes de base sont présentés sur les tables 2, 3 et 4 qui se transforment ensuite en tables 5, 6 et 7, où les mots de même racine étant des équivalents des différentes langues sont disposés en ligne. La fixation du coefficient  $1 + \frac{K_\alpha}{K}$  s'opere en calculant du nombre relatif des locuteurs natifs

appartenant à un groupe de langues pour lequel est construite la langue moyenne. Les mots équivalent à chaque lexème du choix de base sont placés en ordre de diminution de la signification de la fonction  $F_{n,\alpha}^j(x)$ . Les mots disposés dans l'ordre indiqué sont donnés dans la table 9. Les racines insérées dans le dictionnaire de la langue moyenne présentent les acceptions essentielles et sont placés au début des listes. D'autres mots constituent une réserve particulière des synonymes servant de base d'enrichissement du lexique de la langue moyenne.

Les avantages de notre approche de la résolution du problème linguistique sont les suivants:

1. en se basant sur les langues soeurs, la langue moyenne peut être comprise, à un certain degré, par les locuteurs natifs appartenant au même groupe de langues;
2. n'étant pas la langue maternelle d'un groupe ethnique quelconque, elle ne peut pas être le privilège d'un groupe et par conséquent ne peut mener à l'accentuation de la politique en matière linguistique;
3. le projet proposé élimine le choix arbitraire d'une des langues soeurs comme dominante dans la région ainsi que les conflits entre les ethnies;
4. la langue moyenne va contribuer à l'égalité des nations et de leurs langues et à l'établissement d'une information objective dans le monde.

Selon la convention collective des peuples de langues proche à l'intérieur de laquelle on va établir une langue moyenne e on pourrait adopter l'alphabet unique pour cette communauté ou bien la langue moyenne pourrait utiliser l'alphabet appartenant à chaque nation de ce groupe. Pour la création des langues moyennes il est important d'établir un organisme international ayant ses filiales dans les centres linguistiques des langues concernées. Nous proposons d'utiliser la langue moyenne dans un but communicatif pour les peuples frères .Elle peut servir aussi de banque de données contenant des informations d'une importance régionale ou mondiale. Entre autres les langues d'état pourront librement se développer. La réalisation de notre projet va assurer l'égalité entre les peuples et leurs langues maternelles et va aussi contribuer à l'instauration des rapports de confiance et d'entre - aide. Les formes démocratiques d'ethnolinguapanisme (panturkisme,

panpersanisme, panslavisme et d'autres) ne sont pas en contradiction avec les droits internationaux des Nations et de l'Homme , elles présentent les valeurs propres à l'humanité. Il est à proposer de généraliser la méthode en question et de créer une langue auxiliaire moyenne pour les besoins de la communication entre les peuples.

## ANNEX

On the occasion of presentation on Averaged Languages made at Summer University Indiana University Bloomington the coefficient  $1 + \frac{K_a}{K}$  has been calculated with population estimated by the middle of 2000. Together with calculation of the coefficient for the second edition of the paper, this gave an opportunity to study its behaviour in time. The results of calculations are presented in the table below.

**Table 1**

COMPARISON OF  $1 + \frac{K_a}{K}$  COEFFICIENT IN THREE LANGUAGE GROUPS FOR  
DIFFERENT PERIODS

Ethnic group/language	Number (m)			$1 + \frac{K_a}{K}$		
	1992	2000	2005	1992	2000	2005
Azeri	16.3	23.26	24.44	1.8	1.97	1.96
Kazakh	9.53	9.93	12.06	1.6	1.41	1.47
Tatar	6.89	5.91	7.21	1.38	1.25	1.28
Turk	50.33	55.83	55.92	3.46	3.33	3.19
Uzbek	19.44	24.88	27.57	1.95	2.04	2.08
Russian	163.05	-	137.84	4.32	-	4.33
Ukrainian	47.47	-	37.65	1.97	-	1.91
Byelorussian	10.63	-	9.92	1.22	-	1.24
Serbs and Croats	14.84	-	14.02	1.32	-	1.34
Bulgarian	9.37	-	7.39	1.2	-	1.18
Spanish	274.06	-	675.9	3.3	-	4.1
Italian	67.34	-	32.84	1.54	-	1.15
Portuguese	169.6	-	201.11	2.42	-	1.92
Romanian	21.18	-	19.74	1.60	-	1.09
French	63.18	-	161.47	1.53	-	1.74

In the group of Turkic languages the coefficient for two has showed slight increase (8-4%) and that for three of them decreased a bit (3-13%) in the period of 1992 to 2000. As for the period 2000 to 2005 the coefficient increased insignificantly in three cases and decreased in two cases. The value for Turkish decreased and that for Uzbek increased for the both periods.

The same picture is observed in the group of Slavonic languages – three values increased and two of them decreased. More change is observed in the group of Romance languages, and this accounts for the change of principle in counting speakers of the language.

Insignificant change of the coefficient values allows us to conclude that the averaged languages will be fairly stable, i.e. there will be no need for frequent re-averaging due to possible change of speakers' number. Comparison of the averaging results for the notion "to speak" in the three groups by 1992, 2005, and 2007 (Table 2) makes this conclusion even more solid.

**Table 2**

COMPARISON OF AVERAGING OF THE NOTION “TO SPEAK” IN THREE LANGUAGE GROUPS BY 1992, 2005, AND 2007<sup>18</sup>

Turkic			Slavonic		Romance	
1992	2005	2007	1992	2005	1992	2005
söyle/mek	söyle/mek	söyle/mek	говор-	говор/гава	des/diz/dir	des/diz/dir
de/mek	de/mek	de/mek	разговор-	р	mostr/mon	mostr/mont
konuş/mak	konuş/mak	айт/у		разговар/	tr	r
айт/у	айт/у	konuşmak	каз(в)-	разговор	habl	habl
bahset/mek	bahset/mek	bahsetmek	балак-	балак	parl	parl
гапир/моқ	гапир/моқ	гапирмоқ	мовл-	мовл	fal	falar
daniş/mak	daniş/mak	daniş/mak		каз/казв	vorb	vorbi
тилдан	тилдан	тилдан			doved	dovedi
фойдалан/моқ	фойдалан/мо	фойдаланмоқ				palesare
изҳор қил/моқ	қ	гәп қилмақ				
әйтеп бирү	изҳор	геплемек				
	қил/моқ	ипадә қилмақ				
	әйтеп бирү	изҳор				
		қилмоқ				
		делиллимәк				
		гурун этмек				
		билгүзмәк				
		әйтеп бирү				
		далил болуу				
		гөркезмек				
		билдирүү				

In the group of Turkic languages, the order of the equivalents of the notion «говорить» (to speak) ordered according to the function  $F_n^j(\frac{p}{k})$  values has not changed at all. In the group of Slavonic languages the change is observed in the third equivalent and lower. Given the proposal that a equivalent with the highest weight is to be included into the averaged vocabulary as the main word the change may be considered insignificant. The there are no changes in the group of Romance languages as well. Appearance of the word «palesare» in the results of 2005 is seeming change. The word has been dropped out from etimologically ordered Table 7 for the Romance languages due to technical causes. This word is the notion «говорить» (to speak)’s equivalent in Romanian, and it is forth, last one in order. There is no etimological matches to the word in four other Romance languages. Thus, the word «palesare» was the least weighed equivalent in the results of 1992, too, and takes the last line in the Table 7.

<sup>18</sup> Iranian languages were not considered at that time.

The authors invite all those who are interested to collaborate in study of all aspects related to the averaged languages.

Contact information:

Bakhtiyor Karimov  
Doctor of Science (Philosophy), Professor, Academician of International Academy of  
Informatization

Tel.: 998-71-2464585

e-mail: [karimov.bahtiyor@yahoo.com](mailto:karimov.bahtiyor@yahoo.com)

Mailing address: 10 Eshon Bobokhon St.  
Tashkent 100169  
Uzbekistan

Shoahmad Mutalov  
Furbright Scholar

Tel.: 998-71-2491908

e-mail: [shahahmad.mutal@gmail.com](mailto:shahahmad.mutal@gmail.com)

Mailing address: 44 1<sup>st</sup> Pass Tinchlik St.  
Tashkent 100169  
Uzbekistan