

DOI 10.1515/linpo-2015-0001

## ARTICLES

**Prefixation Ability Index (PAI)  
as a powerful typological tool of historical linguistics****Alexander Akulov**

Independent scholar, Saint Petersburg, alexander\_akulov@gengo-chan.com

**Abstract:** Alexander Akulov. *Prefixation Ability Index (PAI) as a powerful typological tool of historical linguistics*. The Poznań Society for the Advancement of Arts and Sciences, PL ISSN 0079-4740, pp. 7-24

At present there are many odd hypotheses about the genetic affiliation of certain languages. Most such hypotheses are invented without any serious examination of the structural differences between the languages being compared. The PAI method was inspired by ideas of A. P. Volodin, who noticed that there were two types of languages, one type has prefixation and the other does not. Actually, there is no sharp divide between the two types, it is more precise to use a coefficient (i.e. PAI) rather than simply ask “does a language allow prefixation?” The PAI theory supposed there was correlation between values of PAI of genetically related languages. Tests of PAI on the material of well assembled stocks prove that such correlations exist. Being applied to Ainu and to languages that are possibly supposed to be related to it. The PAI shows that Ainu that is not related to either Altaic or Nivkh, while a search for relatives of Ainu to the south shows potential. Also PAI can be useful in the case of other unsettled questions of language affiliation in North America, New Guinea, Southeast Asia, Africa and other places.

**Keywords:** use of typology in comparative linguistics, Prefixation Ability Index, Ainu, Altai stock, Nivkh, Ainu and Austroasiatic, Ainu and Austronesian, isolated languages of Siberia and Far East

**1. Introduction to the problem**

As my main sphere of interest is the Ainu language I am also interested in possible genetic relationships of Ainu. However, all the works on the subject that I have examined (Mudrak 2013; Murayama 1993; Patrie 1982; Vovin 1993 et al.) give the impression of having been written by people who have not actually learned the basics of linguistics well, i.e.: they do not appear to realize that language is first of all structure and not a simple heap of lexemes. By taking a randomly selected set of Ainu lexemes and some look-alike lexemes from some other language and comparing them authors of such works state that Ainu and X language are related (or seem to be related) and almost no attention is paid to structures. (Due to its unsettled genetic position Ainu is one of most popular objects of perfunctory and amateurish “hypotheses”.) The oddest fact is that using almost the same methods different scholars have come to completely different conclusions: Murayama Shichirō states Ainu is related to Austronesian; Alexander Vovin states it is related to Mon-Khmer; Hattori Shirō and James Patrie

state Ainu is related to Altaic; and Oleg Mudrak believes Ainu is related to the Nivkh, Itelmen and Chukchi-Koryak languages and so on. Some characteristic examples are shown below:

Vovin in his “A Reconstruction of Proto-Ainu” makes a reconstruction of Ainu, then compares it with different stocks of Southeast Asia and finally states that Ainu seems to be a distant relative of Proto-Mon-Khmer (Proto-Austro-Asiatic in his terms).

“Items on the left are from PA<sup>1</sup> and those on the right are from PAA:<sup>2</sup>

\**paar*=*aC*=*sE* – \**par* ‘to fly’

\**prAA* – \**pāŋ* ‘mouth’

\**ra* ‘down’ – \**k=ram* ‘under’, ‘beneath’, ‘below’

\**rAm* ‘soul’, ‘heart’, ‘mind’ – \**k=lom* ‘heart’

\**rayonti* – \**rayōŋ* ‘rainbow’ (Vovin 1993: 169)

Followers of the megalocomparison<sup>3</sup> approach always say that they compare lexis, find regular phonetic correlations, and those correlations indicate the relationships between languages. It is possible that I am very unusual, but I do not see any regularity in these examples. In one example \**p* of PA correlates with PAA \**p* in another example PA \**pr* correlates with PAA \**p*; in one example \**r* of PA correlates with PAA \**r*, in other examples with PAA \**kr* and \**kl*. If these are called ‘regular correlations’ then it is possible to prove relationships between any two randomly chosen languages” (Nonno 2015: 44).

Hattori proves the Altaic origin of Ainu through the following examples:

“Ainu: *kur*: *kur* – ‘shadow’, *niskur* – ‘cloud’, (*nis* – ‘sky’, *kur* – ‘darkness’),

*kunne* – ‘black’ (*kur* ‘darkness’ + *ne* ‘to be’), *ekurok* – ‘dark’;

Japanese: *kur*: *kurashi* – ‘dark’, *kuru* – ‘darkness’, *kuro* – ‘black’, *kumo* – ‘cloud’;

Korean: *kurum* – ‘cloud’, *kurim* – ‘soot’, *kerim* – ‘soot’, *kurimcha* – ‘shadow’;

Tungusic: *kurunük* – ‘soot’, *komnomo* – ‘black’;

Mongol: *kara* – ‘black’, *küreng* – ‘brown’;

Turkish: *kurim* – ‘soot’, *kara* – ‘black’ (Majewicz 1984: 89)

How it can be that using almost the same methodology they have come to such radically different conclusions? Due to this I have been considering this so called trend of megalocomparison with a high degree of skepticism; it seems there is no accurate methodology in it except for the idea: “the artist sees so”.

Using that method I can, for instance, prove that Japanese is related to Cantonese: *boku*, the Japanese personal pronoun ‘I’ used by males is a borrowing from Southern Chinese *buk* ‘servant’, ‘I’ as well as *o-taku* ‘your family’, ‘your house’ or ‘your husband’ that originated from Southern Chinese form *zaak* ‘house’ or, for instance, *san* ‘three’ that originates from Southern Chinese *sam* and so on. If there were no other languages of the so-called Buyeo stock and no other languages of Chinese origin remained there would be no way to single those words out as borrowings since they have same regular and wide use as words of Japanese origin.

<sup>1</sup> Proto-Ainu.

<sup>2</sup> Proto-Austro-Asiatic.

<sup>3</sup> Megalocomparison is a term created by James Matisoff (Matisoff 1990), it is used to denote attempts to find distant genetic relationship base on the comparison of lexis.

In the case of Japanese and Cantonese, their history and the history of the Buyeo and Chinese stocks is rather well known and there is much firm evidence that Japanese is not related to the Chinese stock. But what can be done when dealing with much more isolated languages/stocks whose history is almost unknown is less clear.

It was Swadesh who warned that comparisons of vocabularies alone cannot be taken as proof of the genetic relationship between languages and some other methods should be used, i.e.: the analysis of structures. Swadesh's method is the estimation of the approximate time of divergence of languages which have been already proven to be related. However his warning is forgotten and it can be seen that lexicostatistics is a widely spread method as a completely relevant method of comparative linguistics. Of course, it is much easier to take a dictionary compiled by someone else and chose some "look alike" words in it than to study grammar and structures; that fact explains the widespread nature of such methodology.

Structure is something like bottle while lexicon is the liquid inside it. Wine, water, gasoline or even sand can be put into the bottle but the bottle always retains the form of a bottle. Because of this should the most attention should be paid to structures in linguistic investigations and especially when dealing with questions of genetic relationships. To those who think structures are not important element and lexicon is much more important the following example can be given:

*Bā he no rūto wo gugutte, purinto shita.* 'Having googled the route to the bar I printed it'.

What makes this phrase Japanese? Is it the "Japanese" words *bā* 'bar', *rūto* 'route', *purinto* 'print' or, may be, verb *guguru* 'to google'? One can probably say that this example is very unusual since it was made without so called 'basic lexicon'. However, it should be noted that the basic lexicon is not of chief importance since it should be kept in mind that it was Swadesh himself who warned against the use of lexis as base for establishing genetic relations. Also, as was shown above (in the case of Japanese and Cantonese), borrowings can be even found even in the 'basic lexicon'.

Moreover, it should be noted that there is actually no such thing as 'basic lexicon' since every local set of languages has its own basic lexicon and even items of that lexicon are actually determined by certain cultural factors (Hoijer 1956).

From what has been stated so far, it can be seen that comparative linguistics has not yet elaborated a real scientific precise methodology since existing methodologies allow different scholars to make completely different contradictory conclusions. Also there are no truly objective methods of testing results being offered.

I see the root of the problem in the fact that completely insufficient attention is paid to structural issues. What is saddest is that even opponents of megalocomparison have also based their research on the comparison of lexemes: even when they consider certain morphemes they actually make the same comparison of words and no attention is paid to the fact that any morpheme is a system of three components: meaning, position and material exponent and that meaning and position are of same or probably of higher importance than the material part. Typology is considered as a rather useless glass bead game to satisfy interest while comparative linguistics has actually little relationship to typology and seems to make its statements with help of some magic spells.

As far as language is first of all system of distributions, and not a pile of lexemes, structural issues should be the basis for any linguistic questions and especially questions of possible

genetic relationships. Methodology of comparative linguistics should be based on structural issues only while methodology based on the comparison of lexes should be abandoned since it has been shown that it can lead to completely different results that are determined by an author's wishes only and in most cases it gives birth to weird and perfunctory "hypotheses" that are not seriously discussed and proven but simply presented to the public for acceptance.

Among structural features one of the most important is the linear model of word forms. The linear model of word forms is a very conservative parameter and is not influenced by contact (Volodin 1997c). For this reason, parameters based on the linear model of word form can be very useful tools when speak of possible genetic relationships between certain languages. One such parameter, namely the Prefixation Index, is presented in this article.

## 2. Prefixation Ability Index hypothesis

### 2.1. Background of the Prefixation Ability Index hypothesis

A. P. Volodin points out that Paleosiberian languages<sup>4</sup> can be subdivided into two sets by the parameter of prefixation and its absence or presence: one group has prefixation and one does not. The first group consists of Ainu, Yeniseian and the Chukchi-Kamchatka languages; the second group consists of Eskimo-Aleut languages and the Yukaghir language (Volodin 1997c: 9).

After Volodin, the first set is conventionally named the "set following the American linear model of word forms"<sup>5</sup>, its linear model of word forms originally was represented as:

(p) + (r) + R + (s);

the second set is conventionally named "set following the Altaic linear model of word forms"<sup>6</sup>, its linear model of word form structure originally was thought to be as follows: (r) + R + (s); p – prefix, s – suffix, R – main root, r – incorporated root; brackets mean that the element can be absent or that in a particular word form can be represented by more than one item. These are linear model of two types according to Volodin.

Volodin supposed that languages belonging to the same set had certain structural similarities. He also supposed that typological similarities could probably hint at possible routes of ethnic migrations and therefore possible genetic relationships.

### 2.2. From Volodin's ideas to the Prefixation Ability Index hypothesis

Being aware of Volodin's idea about two types of linear models of word forms, for quite a long time I thought that there was a pretty strict division between languages that

<sup>4</sup> The term "Paleosiberian languages" was created to designate isolated languages of Siberia and East Asia; it does not refer to a hypothetical stock but rather a set of genetically unrelated languages assembled by their geographic location.

<sup>5</sup> The set was named "American" since such a model of the word form is widely spread in the Native languages of the Americas.

<sup>6</sup> The set was named "Altaic" since such a linear model of word forms is widely spread in the Altaic languages.

have prefixation and those that do not. For instance, I seriously thought that Japanese had no prefixes and tried to consider all prefixes of Japanese as variations of certain roots, i.e. as components of composites; until one day I finally realized that so called “variations” of roots actually could never be placed in nuclear position and so they all should be considered as true prefixes, so a strict dichotomy does not exist and I had to elaborate a new theory.

As far as any language actually has some capacity for prefixation so there is no strict border between “languages with prefixation” and “languages without prefixation” and ideas of subdividing all existing languages into two sets that have no intersection should be give up.

Hence thereupon, linear model of word forms have the following structures:

(P) + (R) + r + (s) – American linear model of word forms

(p) + (r) + r + (S) – Altaic linear model of word forms

In this particular case, capital letters mean that this parameter is well developed in languages of that type. Thereby, there is no principal structural difference between languages of the “American” type or “Altaic” types, the difference is one of degree of manifestation of certain parameters and so, in order for our conclusion to not be speculative, we should speak of the degree of prefixation producing ability / prefixation ability degree / prefixation ability index, i.e. of certain measure of prefixation.

I suppose that each language has its own ability to produce prefixation and that this ability does not change seriously during all stages of its history.

Also I suppose that prefixation ability manifests itself in any circumstances, i.e., by any means: by means of original morphemes existing in a certain language or by borrowed morphemes. If a language has a certain prefixation ability it will be shown.

That is why I do not make difference between original and loaned affixes.

Also for current consideration is not of principal importance whether this or that affix is derivative or relative: if we take into account relative affixes only then, for instance, Japanese is a language without prefixes.

That is why we should define prefixes not by their derivative or relative meanings but by their positions inside the word form: a prefix is any morpheme that can be placed only left from the nuclear position and can never be placed in nuclear and between this morpheme and nuclear root no meaningful morpheme with its clitics can occur (i.e. between the nuclear root and prefix no meaningful morpheme with its auxiliary morphemes can be placed). It should be noted that there are no so called semi-prefixes. If a morpheme can be placed in nuclear position it is a meaningful morpheme and any combinations with it should be considered to be compounds.

As was said above: it is supposed that each language has its own prefixation ability index that remains about the same during the entire history of the language, hence, it can also be supposed that languages that are thought to be related should have quite similar values of Prefixation Ability Index.

### **2.3. Prefixation Ability Index (PAI) calculation algorithm**

How can the Prefixation Ability Index (hereinafter PAI) be measured?

The value of the PAI correlates with the proportion of prefixes among the affixes of a language.

Hence, in order to estimate the proportion/percentage of prefixes of a certain language we should do the following:

- 1) Count the total number of prefixes;
- 2) Count the total number of affixes;
- 3) Calculate the ratio of the total number of prefixes to the total number of affixes.

Why is it important to count total number of prefixes and then calculate the ratio to the total number of affixes but not to estimate PAI by frequency of prefix forms in a random text? A certain language can have quite high value of PAI but in a particular text word forms with prefixes can occur at a low frequency. Our task is to estimate the proportion of prefixes in grammar but not the proportion of prefixes found in a random text. The prefixes/World index<sup>7</sup> estimated by Greenberg was the estimation of prefix frequency in a text. Of course, that index can give some general notion of the prefixation ability of a language, though it is extremely rough and inaccurate since in a randomly chosen text there can be very few words with prefixes: the longer a text is the more precise the conclusion is but the chance of error in such calculations still remains very high; while in counting all existing affixes of a certain language, the potential for error is extremely low and even if occasionally some affixes are missed it does not seriously influence the results. Moreover it should be noted that despite Greenberg's groundbreaking work on typology he did not actually use those results in his research; he was an adept of megalocomparison and made his conclusions based on "mass comparison" of lexes rather than on structural correlations; his typology was a "glass bead game" and was separate from his actual field of studies.

To those who think, that it is impossible to estimate number of morphemes since living languages always change, it should be noted that living languages do not create new morphemes every day, especially grammatical morphemes. The fact that in learning a language it is possible to use descriptions of its grammar written decades previously is the best proof that grammar exists at a very conservative level of any language. Hence, it is possible to estimate the total number of affixes of a living language as far as a description where all stable forms are represented can be obtained.

And there is no need to concern oneself with what might exist in a certain language in the future, i.e. the current stage of a living language should be considered and not possible future stages since they simply do not yet exist.

As for possibility of counting, it should be noted that even the set of words is a countable finite set while the set of morphemes and especially grammatical morphemes is a finite set and its cardinality is obviously much less than the set of words.

In order to show the possibility of counting and to explain the whole procedure some examples will be given:

### **PAI of Hawaiian**

(Krupa 1979 was used to count the PAI of Hawaiian)

#### **Prefixes:**

1. *hā-* – modifies the meaning of a root: *'ula-* 'red'/'brown'; *hā'ula'ula-* 'reddish'/'pink';
2. *ho'o-/hō-* – causative;
3. *kī-* – modifies the meaning of a root: *ohu* 'fog', *kī'ohu'ohu* 'foggy place';

<sup>7</sup> Greenberg 1960.

4. *mā-* – modifies the meaning of a root: *ewa* ‘fickle’, *māewa* ‘fluctuating’;
5. *pā-* – forms distributive values in numerals: *lima* ‘five’, *pālima* ‘by five’;
6. *pe/pē-* – forms the comparative meaning of directional pronouns: *nei* ‘that’, *pēnei* ‘like that’;
7. *pō-* – modifies the meaning of a root: *maika’i* ‘to be fine’, *pōmaika’i* ‘happy’;
8. *u-* – verb pluralizer: *haele* ‘to go (sg.)’, *uhaele* ‘to go (pl.)’;
9. *ū-* – modifies the meaning of a root: *kele* ‘watery’ / ‘cloudy’, *ūkele* ‘muddy’

**Suffixes:**

- 1 *-a/-hia/-lia/-mia/-nia/-na* – passive voice marker
- 2 *-na* – nominalizer

Hence, the total number of prefixes of Hawaiian is 9; the total number of affixes is 11. The PAI of Hawaiian is 9/11 that is about 0.82.

**PAI of Maori**

(The PAI has been calculated after Krupa 1967)

**Prefixes:**

(In the list of Maori affixes all prefixes except *whaka-* were marked by Krupa as stem modifiers with different meanings.)

1. *a-*
2. *ha-*
3. *hi-*
4. *ho-*
5. *hu-*
6. *i-*
7. *ka-*
8. *ki-*
9. *ko-*
10. *ku-*
11. *ma-*
12. *mo-*
13. *nga-*
14. *o-*
15. *pa-*
16. *pi-*
17. *po-*
18. *pu-*
19. *ra-*
20. *ri-*
21. *ro-*
22. *ta*
23. *ti-*
24. *to-*
25. *tu-*
26. *u-*

- 27. *wha-*
- 28. *whaka-* – causative
- 29. *whe-*
- 30. *whi-*

**Suffixes:**

- 1. *-a* – stem forming suffix;
- 2. *-hi/-ki/-mi/-i* – stem forming suffix;
- 3. *-hia/-kia/-mia/-ngia/-ria/-tia/-whia/-ia/-na/-nga/-ina/-a* – passive voice marker;
- 4. *-hanga/-kanga/-manga/-ranga/-tanga/-anga/-nga* – nominalizer.

The total number of Maori prefixes is 30; total number of affixes 34.

Hence, the PAI of Maori is 30/34 that is about 0.88.

Hawaiian and Maori were used here as material for illustrations as they have very compact and easy estimated grammars and due to this case are very convenient material for experiments and illustrations.

It would be nice to demonstrate all details of the procedure for counting the PAI of all the languages mentioned below, but since the format of the current article does not allow for this, I have shown just some examples of the procedure.

### 3. PAI testing: from hypothesis toward theory

In this chapter the results of testing of the PAI hypothesis i.e.: there is a correlation between values of the PAI of genetically related languages will be examined.

In order to test PAI hypothesis I paid attention to some languages of well known and firmly established stocks i.e. those stocks which existence is undoubtful: Austronesian, Indo-European and Afroasiatic.

#### 3.1. Austronesian stock

##### 3.1.1. Polynesian group

###### 3.1.1.1. Eastern Polynesian Subgroup

Hawaiian 0.82 (The PAI has been calculated after Krupa 1979)

Maori 0.88 (The PAI has been calculated after Krupa 1967)

Tahitian 0.66 (The PAI has been calculated after Arakin 1981)

###### 3.1.1.2. Samoan-Tokelauan subgroup

Samoan 0.5 (The PAI has been calculated after Arakin 1973)

###### 3.1.1.3. Tongic subgroup

Niuean 0.8 (The PAI has been calculated after Polinskaya 1995)

Tongan 0.78  $\approx$  0.8 (The PAI has been calculated after Fell 1918)

#### 3.1.2. Philippine group

##### 3.1.2.1. South Mindanao subgroup

T'boli 0.72 (The PAI has been calculated after Porter 1977)



**3.1.2.2 Northern Luzon subgroup**

Pangasinan 0.6 (The PAI has been calculated after Rayner 1923)

**3.1.3. Malayo-Sumbawan group****3.1.3.1. Malay subgroup**

Indonesian 0.53 (The PAI has been calculated after Ogloblin 2008)

**3.1.3.2. Chamic subgroup**

Cham 0.6 (The PAI has been calculated after Aymonier 1889)

**3.1.4. Formosan group**

Bunun 0.8 (The PAI has been calculated after De Busser 2009)

**3.1.5. PDI of Eastern Barito group**

Malagasy 0.74 (The PAI has been calculated after Arakin 1963)

**3.2. Indo-European****3.2.1. German group**

Dutch  $0.49 \approx 0.5$  (after list of Dutch affixes)

German  $0.51 \approx 0.5$  (after list of German affixes)

English  $0.61 \approx 0.6$  (after list of English affixes)

Icelandic 0.63 (after list of Icelandic affixes)

**3.2.2. Slavonic group**

Czech 0.52 (after list of Czech affixes)

Polish 0.57 (after list of Polish affixes)

**3.2.3. Celtic group**

Irish 0.67 (after list of Irish affixes)

Welsh 0.35 (after list of Welsh affixes)

**3.2.4. Romance group**

Latin 0.26 (after list of Latin affixes)

Spanish 0.34 (after list of Spanish affixes)

**3.3. Afroasiatic stock****3.3.1. Semitic group****3.3.1.1. Central Semitic subgroup**

Arabic (Classical) 0.26 (The PAI has been calculated after Yushmanov 2008)

Phoenician 0.26 (The PAI of Phoenician has been calculated after Shiftman 2010)

### 3.3.1.2. Eastern Semitic subgroup

Akkadian (Old Babylonian dialect) 0.2 (The PAI of Akkadian has been calculated after Kaplan 2006)

### 3.3.2. Egypt group

Coptic (Sahidic dialect) 0.87 (The PAI of Coptic has been calculated after Elanskaya 2010)

One could probably say that Coptic has broken our hypothesis, but actually PAI just has shown us that group to which Coptic language belongs and the Semitic group diverged very long ago, probably by the Neolithic.

The values of PAI correlate well with the degree of language relation (Figure 1): when we see very close values it means the corresponding languages may be very closely related and then structural and material evidence should be shown alongside the PAI correlation.

If the PAI of certain languages differ seriously (as in the case of the Coptic and Semitic groups) then, if those languages are supposed to be related, very firm evidence to prove that fact should be shown; in the case of the Egyptian group and Semitic stock such structural evidence have been demonstrated.

Hence, in summary, the following can be stated: the PAI is something like a safety valve for comparative linguistics, i.e.: if its values do not differ more than fourfold then there are actually no obstacles for further research on genetic relationships; if the values differ fourfold or more then very serious evidence of a genetic relationship should be shown, otherwise there is a serious reason to think that the languages under consideration are not related genetically.

### 3.3.3. PAI of a group/stock

The PAI of a linguistic group or a stock can be calculated as an arithmetic mean and it is precise enough for rough estimates. Really this procedure is slightly more complicated. First, a reconstruction of the language that was ancestor for that group/stock then should be estimated. Then, the degree of correlation of each language of group to the ancestor and then the PAI of each language of group should be taken into the arithmetical mean with a coefficient of correlation.

However, since this is very complicated and obviously would take much more time than just a simple estimation of the PAI it is possible to use just the arithmetic mean. Also if there is a lack of material, then even a comparison of the PAI of particular languages is not completely useless.

## 4. Applying PAI to some unsettled hypotheses

Since the PAI has been successfully tested and can be considered as a relevant method, it can be applied to some of the ideas mentioned in the introduction.

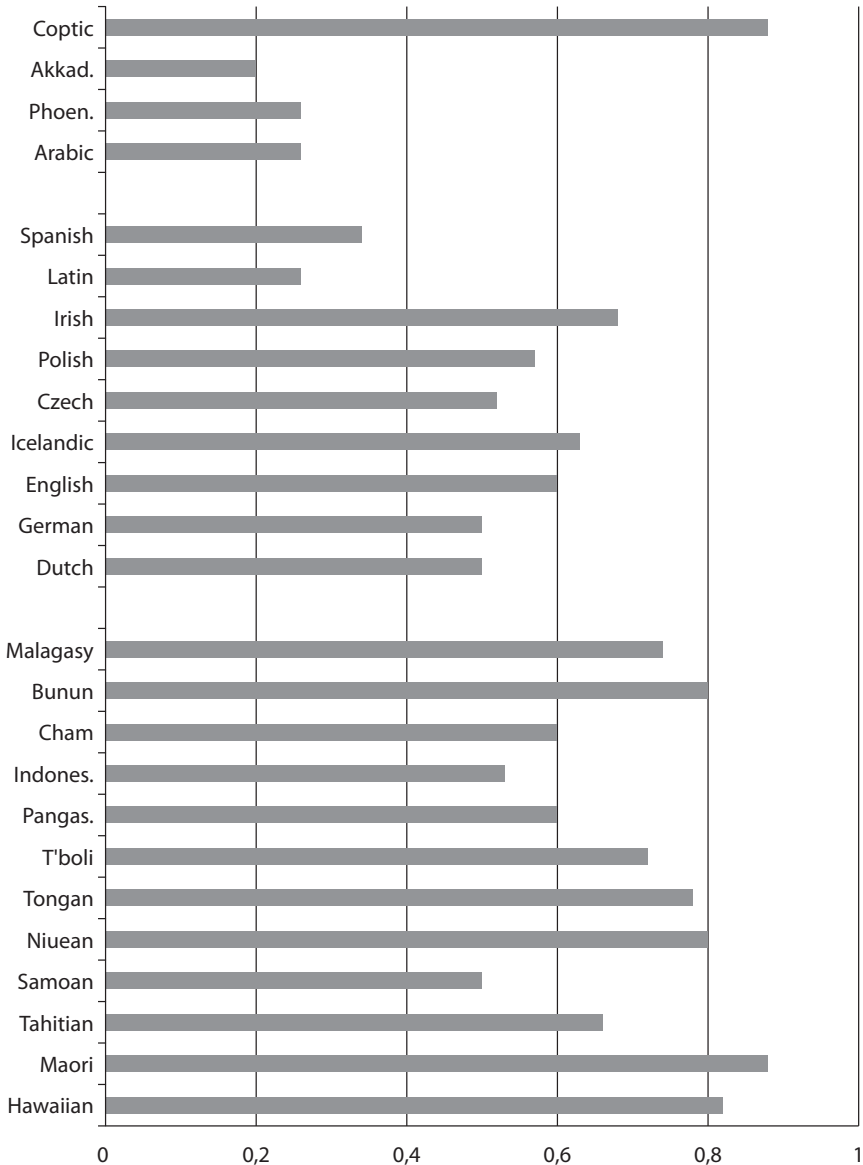


Figure 1: Diagram representing the PAI values of some established stocks

## 4.1. Ainu and Altaic

### 4.1.1. Does Ainu belong to the Altaic stock?

Having compared lexes Patrie states that Ainu is related to Japanese and Korean and is a language of the Altaic stock (Patrie 1982). The Altaic stock itself and whether Japanese and Korean are part of it is still not settled and even the relationship of Japanese and Korean

is still actually questioned. However, leaving those considerations aside, a look at the PAI of these languages reveals the following:

The PAI of Ainu is 0.75 (after Tamura 2000).

The PAI of Japanese is 0.13 (after Lavrent'yev 2002).

The PAI of Korean is 0.13 (after Mazur 2004).

The values of the PAI differ sixfold. Can languages having such seriously divergent PAI values be relatives?

In the case of Coptic and the Semitic group, the values of PAI differ fourfold and if there were no other firm structural evidence of the relationship of Coptic and the Semitic stock would be very problematic.

In the case of Ainu and Altaic, serious differences in PAI values is obviously proof of no relation. Moreover, it should be kept in mind that Japanese and Korean are actually marginal to Altaic and they have, probably, the highest PAI values of Altaic so if Ainu is compared with some “true” Altaic languages the difference is much more striking.

And also the fact there is almost no structural correlation between Ainu and Japanese and between Ainu and Korean corroborates the conclusion made with the use of the PAI. All attempts to establish a hypothetical Ainu – Altaic stock are nonsense and display ignorance of the fundamentals of linguistics.

#### 4.1.2. Japanese and Korean

Meanwhile Japanese and Korean might be more closely related than was thought, since their PAI values completely coincide. And this fact well correlates with their structural correlations.

#### 4.2. Ainu and hypothesis of Mudrak

Mudrak believes that such languages as Ainu, Nivkh, Chukchi-Koryak, Itelmen and Aleut are genetically related (Mudrak 2013). In order to better understand the structure of this hypothetical stock I show a scheme (Figure 2) taken from Mudrak’s report.

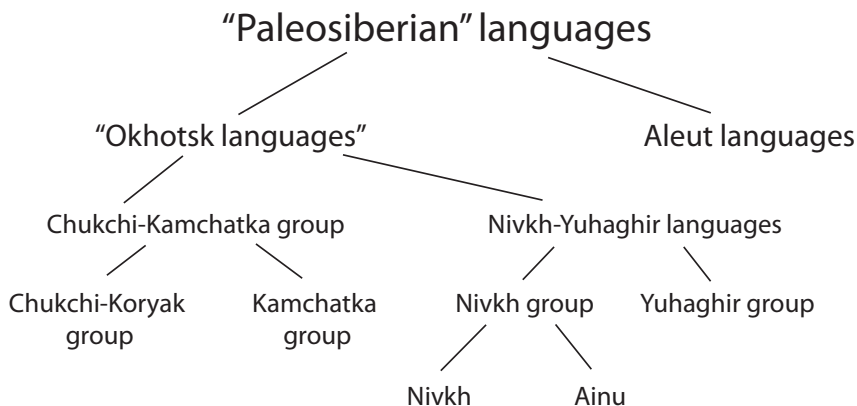


Figure 2: General scheme representing the hypothetical stock created by Mudrak

#### 4.2.1. Ainu and Nivkh

According to Mudrak Ainu and Nivkh not only belong to this hypothetical stock but belong to same group inside the stock.

The PAI of Nivkh is 0.07 (after Gruzdeva 1997).

The PAI of Ainu is 0.75.

Values differ more than tenfold. There are many differences in grammar between Ainu and Nivkh.

The hypothesis regarding a relationship between Nivkh and Ainu is similar to hypothetical hypothesis of a common ancestor of Estonian and Latvian; such a hypothesis could be made by Nivkh or Ainu scholars (if Nivkh and/or Ainu had scholars and European languages were “exotic minority languages”). It should be noted that such hypotheses are completely naïve since they are based only on very perfunctory first sight impression of cultural similarities of Sakhalin Nivkh and Sakhalin Ainu.

#### 4.2.2. Aleut and Ainu

The PAI of Aleut is zero (Golovko 1997: 115) and the same value of PAI is shown by any true relatives of Aleut, for instance the Siberian Eskimo language (Menovschikov 1997: 77); while the PAI of Ainu is 0.75. In (3) We have seen some well assembled groups and stocks and know how values of PAI should differ if languages really to form a stock. As far as current math, which was used to calculate and estimate the correlation of PAI values, does not allow division by zero so an imaginary value for the PAI of Aleut (for instance, 0.000001) must be ascribed in order to show the absurdity of attempts to join Ainu and Aleut in the same stock.

#### 4.2.3. Chukchi-Koryak stock and Itelmen

Another point of Mudrak’s hypothesis is that Itelmen is a quite close relative of the Chukchi-Koryak stock. Traditionally Itelmen is considered as a language of Chukchi-Koryak stock and Mudrak follows that traditional model.

The PAI of the Chukchi-Koryak stock average is about 0.32 (after Volodin Kerekskii yazyk; Volodin & Skorik 1997).

The PAI of Itelmen is 0.12 (after Volodin 1976).

It seems there are no serious objections against this stock. However, it is the point when PAI does not help since its values do not differ much.

I. S. Vdovin, D. S. Worth and A. S. Asinovski pointed to facts against the hypothesis that Itelmen is related to the Chukchi-Koryak stock and have proved that Itelmen is a completely separate stock (Volodin 1997a: 60).

#### 4.3. Ainu and Austronesian

Murayama believed Ainu might be a distant relative of Austronesian. Despite the very naïve lexicostatistic approach the idea seems to be rather more realistic since the PAI of Ainu is 0.75 and Austronesian stock average PAI is about 0.7.

### 4.4. Ainu and Austroasiatic

Vovin tried to show that Ainu is a distant relative of Mon-Khmer. As in the case of Murayama this idea is not completely off base since the PAI of Khmer is 0.66 (after Haiman 2011) which correlates well with that of Ainu.

Scheme below (Figure 3) represents how seriously dissimilar PAI values of languages are forced to be parts of same stock; can be useful and interesting to compare this diagram with Figure 1.

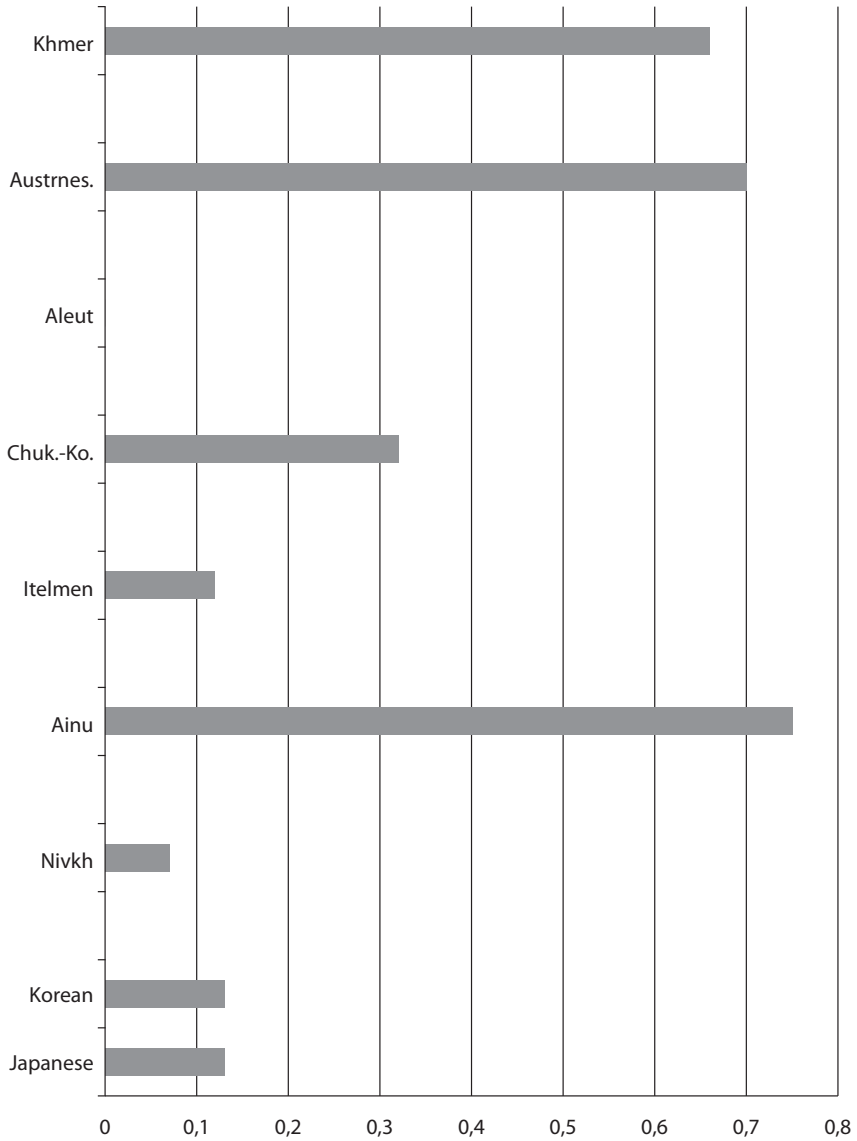


Figure 3: Diagram representing the PAI values of languages that do not belong to established stocks

## 5. Conclusion

The PAI is something like a safety valve for comparative linguistics: if its values do not differ more than fourfold then there is absolutely no obstacles for further research about genetic relationship; if the values differ fourfold or more then absolutely separate, concrete proof of genetic relationships is need. If the values differ sevenfold – tenfold or even more then those languages most probably belong to completely different stocks.

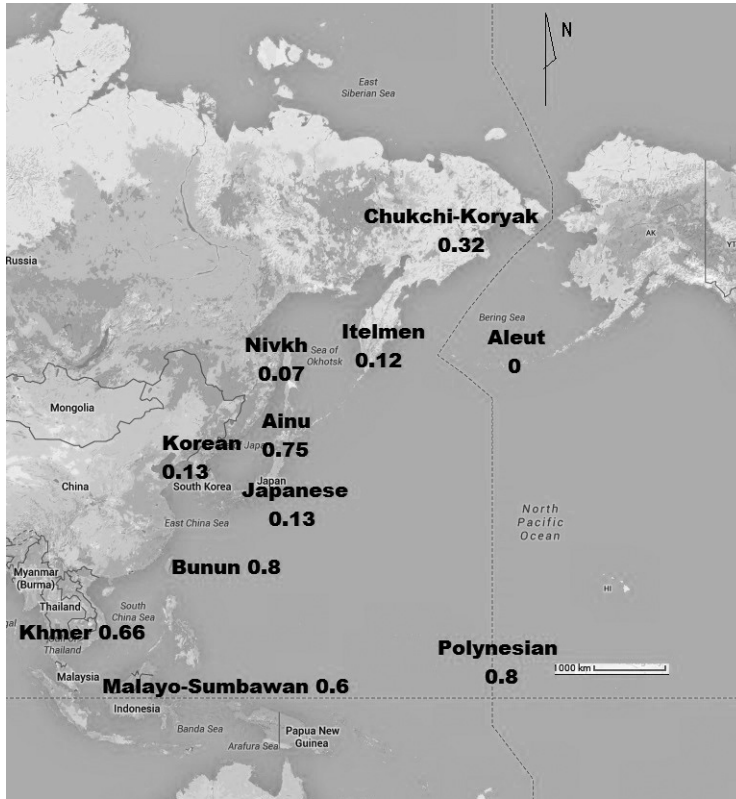


Figure 4: Map representing the geographical distribution of PAI values

The Ainu language has definitely no relation to the Altaic stock or to the Nivkh and Eskimo-Aleut or Itelmen stocks as well; Korean and Japanese are much more closely related than previously supposed; Nivkh could be a potential relative of Itelmen.

Also it should be noted that it would be better to give up the use of the term “Paleosiberian” since it does not aid in comprehension to and discovery but merely inspires megalocomparativists such as Mudrak to invent completely fantastic stocks with no connection with reality. It would be easier to use term “isolated languages and stocks of Siberia and East Asia” than to explain the true meaning of the term “Paleosiberian” each and every time. The term “Paleosiberian” looks much like the name of a stock, it looks too intriguing but is too meaningless for non-specialists to properly understand.

Looking for Ainu relatives in the Southern direction could be a promising field of research (Figure 4).

However, I specially note that such research in the field of linguistics should be correlated with data of other related sciences. “Any hypothesis about the relationship of certain languages should be correlated with corresponding contexts and with data of other related sciences: physical anthropology, population genetics, cultural anthropology and archaeology: if a certain date has been set as an approximate time of existence for Proto-Ainu then how can words of contemporary Ainu be found in preceding epochs? Also if a certain ethnic group is thought to have influenced the Ainu language then this group could hardly influence the terminology of rice cultivation” (Nonno 2015: 44).

I also suppose that the PAI can be a helpful method in those areas where there are many isolated languages/stocks, for instance: in North America, in New Guinea, in Southeast Asia and in Africa.

## References

- Arakin, Vladimir D. (Аракин, Владимир Д.). 1963. *Mal'gashskii yazyk* (Мальгашский язык) [Malagasy language]. Moskva: Nauka.
- Arakin, Vladimir D. (Аракин, Владимир Д.). 1973. *Samoanskii yazyk* (Самоанский язык) [Samoan language]. Moskva: Nauka.
- Aymonier, Etienne. 1889. *Grammaire de la langue cham*. [Grammar of Cham language]. Saigon: Imprimerie colonial.
- De Busser, Rik L. J. 2009. *Towards a grammar of Takivatan Bunun*. La Trobe University, Bundoora, Victoria, Australia. (Doctoral dissertation.)
- Elanskaya, Alla I. (Еланская, Алла И.). 2010. *Grammatika koptskogo yazyka. Saidskii dialect* (Грамматика коптского языка. Саидский диалект) [A Coptic grammar. Sahidic dialect]. Sankt-Peterburg: Nestor-Istoriya.
- Fell J. R. 1918. *A Tonga grammar*. London: Society for promoting Christian knowledge.
- Golovko, Yevgenii V. 1997. Aleutskii yazyk (Алеутский язык) [Aleut language]. In Volodin, A. P. & Vakhtin, N. B. & Kibrik, A. A. (eds.), *Yazyki mira. Paleoaziatskie yazyki* (Языки мира. Палеоазиатские языки) [Languages of the world. Paleosiberian languages], 101-116. Moskva: Indrik.
- Greenberg, Joseph H. 1960. A Quantitative Approach to the Morphological Typology of Language. *International Journal of American Linguistics* 26. 178-194.
- Gruzdeva, Yekaterina Yu. (Груздева, Екатерина Ю.) 1997. Nivhskii yazyk (Нивхский язык) [Nivkh language]. In: Volodin, A. P., Vakhtin, N. B., Kibrik, A. A. (eds.), *Yazyki mira. Paleoaziatskie yazyki* (Языки мира. Палеоазиатские языки) [Languages of the world. Paleosiberian languages], 139-154. Moskva: Indrik.
- Haiman, John. 2011. *Cambodian Khmer*. Amsterdam – Philadelphia: John Benjamins Publishing Company.
- Hoiyer, Harry. 1956. Lexicostatistics: A Critique. *Language* 32(1). 49-60.
- Kaplan, Golda H. (Каплан, Голда Х.). 2006. *Ocherk grammatiki akkadskogo yazyka* (Очерк грамматики аккадского языка) [A sketch of Akkadian grammar]. Saint-Petersburg: Peterburgskoe Vostokovedenie.
- Krupa, Viktor (Крупа Виктор). 1979. *Gavaiskii yazyk* (Гавайский язык) [Hawaiian language]. Moskva: Nauka.
- Krupa, Viktor (Крупа, Виктор). 1967. *Yazyk maori* (Язык маори) [Maori language]. Moskva: Nauka.
- Lavrent'ev, Boris P. (Лаврентьев, Борис П.). 2002. *Prakticheskaya grammatika yaponskogo yazyka* (Практическая грамматика японского языка) [Practical grammar of Japanese]. Moskva: Zhivoi Yazyk.
- List of Czech morphemes represented in Wikipedia. [http://en.wiktionary.org/wiki/Category:Czech\\_morphemes](http://en.wiktionary.org/wiki/Category:Czech_morphemes).
- List of Dutch morphemes represented in Wikipedia. [http://en.wiktionary.org/wiki/Category:Dutch\\_morphemes](http://en.wiktionary.org/wiki/Category:Dutch_morphemes). (Accessed 2015-03-10.)
- List of English morphemes represented in Wikipedia. [http://en.wiktionary.org/wiki/Category:English\\_morphemes](http://en.wiktionary.org/wiki/Category:English_morphemes). (Accessed 2015-03-10.)
- List of German morphemes represented in Wikipedia. [http://en.wiktionary.org/wiki/Category:German\\_morphemes](http://en.wiktionary.org/wiki/Category:German_morphemes). (Accessed 2015-03-10.)



- List of Icelandic morphemes represented in Wikipedia. [http://en.wiktionary.org/wiki/Category:Icelandic\\_morphemes](http://en.wiktionary.org/wiki/Category:Icelandic_morphemes). (Accessed 2015-03-10.)
- List of Irish morphemes represented in Wikipedia. ([http://en.wiktionary.org/wiki/Category:Irish\\_morphemes](http://en.wiktionary.org/wiki/Category:Irish_morphemes)) (Accessed 2015-03-10.)
- List of Latin morphemes represented in Wikipedia [http://en.wiktionary.org/wiki/Category:Latin\\_morphemes](http://en.wiktionary.org/wiki/Category:Latin_morphemes). (Accessed 2015-03-10.)
- List of Polish morphemes represented in Wikipedia. [http://en.wiktionary.org/wiki/Category:Polish\\_morphemes](http://en.wiktionary.org/wiki/Category:Polish_morphemes). (Accessed 2015-03-10.)
- list of Spanish morphemes represented in Wikipedia. [http://en.wiktionary.org/wiki/Category:Spanish\\_morphemes](http://en.wiktionary.org/wiki/Category:Spanish_morphemes). (Accessed 2015-03-10.)
- List of Welsh morphemes represented in Wikipeda. [http://en.wiktionary.org/wiki/Category:Welsh\\_morphemes](http://en.wiktionary.org/wiki/Category:Welsh_morphemes). (Accessed 2015-03-10.)
- Majewicz, Alfred F. 1989. *Ajnu – lud, jego język i tradycja ustna* [Ainu – ethnicity, its language and oral tradition]. Poznań: Wydawnictwo Naukowe Uniwersytetu Adama Mickiewicza.
- Mazur, Yuri N. (Мазур, Юрий Н.). 2004. *Grammatika korejskiego języka (Morfologiya. Slovoobrazovanie)* (Грамматика корейского языка (Морфология. Словообразование)) [Korean grammar (Morphology. Derivation)]. Moscow: Muravey.
- Matisoff, James A. 1990. On Megalocomparison. *Language* 66(1). 106-120.
- Menovschikov, Georgii A. (Меновщиков, Георгий А.) 1997. Aziatskih eskimosov yazyk (Азиатских эскимосов язык) [Asian Eskimo language]. In Volodin, A. P. & Vakhtin, N. B. & Kibrik, A. A. (eds.), *Yazyki mira. Paleoaziatskie yazyki* (Языки мира. Палеоазиатские языки) [Languages of the world. Paleosiberian languages]. 75-81. Moskva: Indrik.
- Mudrak, Oleg A. (Мудрак, Олег А.). 2013. О языках северовосточной Азии (О языках северовосточной Азии) [About languages of North-East Asia]. [http://polit.ru/article/2013/04/22/ps\\_mudrak\\_linguistics](http://polit.ru/article/2013/04/22/ps_mudrak_linguistics). (Accessed 2015-03-10.)
- Murayama, Shichirō (村山七郎). 1993. Ainu go to ōsutoronejia go (アイヌ語とオーストロネシア語) [Ainu language and Austronesian languages]. In: Hanihara, Kazurō (埴原和郎) (ed.) *Nihonjin to Nihon bunka no keisei* (日本人と日本文化の形成) [Forming of Japanese ethnicity and Japanese culture]. 46-72. Tokyo: Asakura shoten.
- Nonno, Tresi. 2015. Critical notes on “A Reconstruction of Proto-Ainu.” *Cultural Anthropology and Ethnosemiotics* 1(1). 36-45. <https://culturalanthropologyandethnosemiotics.files.wordpress.com/2015/03/critical-notes-on-a-reconstruction-of-proto-ainu7.pdf>. (Accessed 2015-03-03.)
- Ogloblin, Alexander K. (Оглоблин, Александр К.). 2008. *Grammatika inodonezijskogo literaturnogo yazyka* (Грамматика индонезийского литературного языка) [Grammar of Indonesian literary language]. Izdatel'stvo Sankt-Peterburgskogo Universiteta.
- Patrie, James. 1982. The Genetic Relationship of the Ainu language. *Oceanic Linguistic Special Publication* 17. 1-174.
- Polinskaya, Mariya S. (Полинская, Мария С.). 1995. *Yazyk niue* (Язык ниуэ) [Niue language]. Moskva: Vostochnaya literartura RAN.
- Porter, Doris. 1977. *A T'boli Grammar*. Manila: Linguistic Society of the Philippines.
- Rayner, Ernest Adolphus. 1923. *Gramatica tan diccionario na salitay Pangasinan* [Grammar and dictionary of the Pangasinan language]. Manila: Methodist Publishing House.
- Shiftman, Ilya Sh. (Шифтман, Илья Ш.). 2010. *Finikiiski yazyk* (Финикийский язык) [Phoenician language]. Moskva: URSS.
- Tamura, Suzuko. 2000. *The Ainu Language*. Tokyo: Sanseidō.
- Volodin, Alexander P. (Володин, Александр П.). 1976. *Itel'menskii yazyk* (Ительменский язык) [The Itelmen language]. Leningrad: Nauka.
- Volodin, Alexander P. (Володин, Александр П.) 1997a. *Itel'menskii yazyk* (Ительменский язык) [The Itelmen language]. In Volodin, A. P. & Vakhtin, N. B. & Kibrik, A. A. (eds.), *Yazyki mira. Paleoaziatskie yazyki* (Языки мира. Палеоазиатские языки) [Languages of the world. Paleosiberian languages], 60-71. Moskva: Indrik.
- Volodin, Alexander P. (Володин, Александр П.) 1997b. *Kerekskii yazyk* (Керекский язык) [Kerek language]. In Volodin, A. P. & Vakhtin, N. B. & Kibrik, A. A. (eds.), *Yazyki mira. Paleoaziatskie yazyki* (Языки мира. Палеоазиатские языки) [Languages of the world. Paleosiberian languages], 53-60. Moskva: Indrik.
- Volodin, Alexander P. (Володин, Александр П.) 1997c. *Paleoaziatskie yazyki* (Палеоазиатские языки) [Paleosiberian languages]. In Volodin, A. P. & Vakhtin, N. B. & Kibrik, A. A. (eds.), *Yazyki mira. Paleoaziatskie*

- yazyki* (Языки мира. Палеоазиатские языки) [Languages of the world. Paleosiberian languages], 8-11. Moskva: Indrik.
- Volodin, Alexander P. (Володин, Александр П.) & Skorik, Pyotr Ya. (Скорик, Петр Я.). 1997. Chukotskii yazyk (Чукотский язык) [Chukchi language]. In: Volodin, A. P. & Vakhtin, N. B. & Kibrik, A. A. (eds.), *Yazyki mira. Paleoaziatskie yazyki* (Языки мира. Палеоазиатские языки) [Languages of the world. Paleosiberian languages], 23-39. Moskva: Indrik.
- Vovin, Alexander. 1993. *A Reconstruction of Proto-Ainu*. Leiden: Brill.
- Yushmanov, Nikolai V. (Юшманов, Николай В.). 2008. *Grammatika literaturnogo arabskogo yazyka*. (Грамматика литературного арабского языка) [Grammar of literary Arabic language]. Moskva: URSS.