

## Some thoughts on the problem of the Austro-Asiatic homeland

In a recent publication (“The Austroasiatic central riverine hypothesis”, JLR 4), Paul Sidwell came up with a new conception of what he considers to be the optimal localization of the original homeland of the Proto-Austro-Asiatic language: the Mekong valley in and around Indochina. The current paper is written in response to Sidwell’s conclusions. These are, first of all, evaluated from a strictly linguistic (lexicostatistical) point of view, and then compared with some of the available evidence on the most likely geographical distribution, cultural and ecological status, and ethnic surroundings of speakers of Proto-Austro-Asiatic. It is concluded that the overall complex of available data much more often contradicts Sidwell’s theory than confirms it, and that, consequently, there is no need to abandon the older hypothesis of a mid-Yangtze valley homeland for Austro-Asiatic (Peiros & Shnirelman), nor is there sufficient evidence to rearrange previously offered models of the internal classification of Austro-Asiatic languages (Diffloth; Peiros) in order to better accommodate the «central riverine» speculation.

*Keywords:* Austroasiatic, comparative method, lexicostatistics, linguistic homeland.

In issue 4 of *Journal of Language Relationship*, Paul Sidwell (further — PS) has published a paper under the title “The Austroasiatic central riverine hypothesis” (pp. 117–134). In the paper the author proposes a new localization for the Austro-Asiatic homeland, challenging some of the previously stated hypotheses on that subject. In this submission, I would like to evaluate PS’s paper along with the currently available data on the topic.

To begin with, there are several important requirements for scientific writing, which, as I am afraid, PS has not been able to satisfy in his paper. Unfortunately, in several respects it is more reminiscent of a commercial prospect or pamphlet than a true work of science. To illustrate my point, I will comment upon some of them.

Each scientific publication must treat its readers with respect, even if they may not have the same experience in the field as the author. This is why it is the author’s responsibility to provide all the necessary information that is needed for the reader to follow and understand the discussion. If a paper contains a certain definitive statement, that statement should be justified by either a reference or an explanation. For instance, it is not enough to say that “... the diversity of Vietnamese dialects in the south is low, but markedly higher in the north, especially, for example, around Vinh” (p. 119). No reference is given here, meaning that the readers are supposed to either trust PS on his word, or to conduct their own investigation.

A suspicious reader who would want to check the reliability of PS’s statement, probably will at least look up Wikipedia, where he/she will find the following information: “There are various mutually intelligible regional varieties, the main four being”: see the table at the next page.

Wikipedia’s sources on Vietnamese variation include: [Alves (forthcoming)], [Alves & Nguyen 2007], [Emeneau 1947], [Hoàng 1989], [Honda 2006], [Nguyễn, Đ.-H. 1995], [Pham 2005], [Thompson 1991(1965)], [Vũ 1982], [Vuong 1981]. Apparently, none of these publications support PS’s view; so what is the source of PS’s statement?

Dialect region	Localities	Names under French colonization
Northern Vietnamese	Hanoi, Haiphong and various provincial forms	Tonkinese
North-central (or Area IV) Vietnamese	Nghệ An (Vinh, Thanh Chương), Thanh Hoá, Quảng Bình, Hà Tĩnh	High Annamese
Central Vietnamese	Hué, Quảng Nam	Low Annamese
Southern Vietnamese	Saigon, Mekong (Far West)	Cochinchinese

Another example of condescendence to not only his readers but to his colleagues as well, is PS's statement that "Peiros applies Starostin's glottochronological method which is controversial" without any justification of this view (p. 129). Naturally, there is a number of publications critical of lexicostatistics and glottochronology in general, but I am not aware of any published criticism of Starostin's method. If PS has any personal constructive objections against it, they should be presented, and it would be very instructive for us to study them.

Some of the footnotes are quite strange. For example, note 7, accompanying the above-mentioned remark about Starostin states that «The software, and Peiros' data and reconstructions are accessible via: <http://starling.rinet.ru>» (p. 129), which is misleading: the statement reads as if only lexicostatistics were presented there. However, the same site also hosts my publicly available Austro-Asiatic etymological dictionary, with about 2,500 etymologies. Is this simply a sign of negligence?

Normally, scientific papers are expected to avoid misrepresenting opposing views and justify the reasons for which the author does not accept them. Unfortunately, PS does not follow this tradition. For example, in discussing the mid-Yangtze location of the AA homeland proposed by Shnirelman and myself [1998], PS says that we use "supposed lexical isoglosses with Hmong-Mien languages", which is absolutely wrong: no such nonsense can be found in any of the mentioned publications or in any of our other publications on this topic. In [Peiros 1998] I claim that Miao-Yao (another name for Hmong-Mien) are related to AA. In [Peiros 1989] I do not discuss homelands at all, and I would be quite surprised to learn that PS has actually seen that particular Russian paper, which had been circulated in no more than a hundred copies.

However, the main problem of the paper is neither its style nor its occasionally misleading notes; rather, it is the general logic of the author. PS's conclusions are: "Three independent lines of inquiry — morphological, phonological, and lexical — have failed to provide decisive indications of nested sub-groupings among AA branches, while the lexical data strongly suggests that there is a contact area centered on Katuic and Bahnaric. Until other indications are forthcoming, the most reasonable hypothesis is a simple radiation out of the Mekong valley" (p. 132).

The major implication of this statement is quite straightforward: all the classifications of the AA family, including those of Pinnow and Diffloth (mentioned on pp. 120–121), are wrong, although PS does not really take the trouble to explain why it is so.

Concerning morphology, it is a well known fact that, as of now, there is no morphological or morphosyntactic reconstruction of either Proto-AA or its descendant, Proto-Mon-Khmer. Therefore, it is premature to talk about clear morphological evidence for AA classification: how could one use non-existing evidence?

Reflexes of several Proto-AA consonants in the daughter-languages<sup>1</sup>, given by PS in Table 1 (p. 128), are also useless for classification, not to mention that the table does not present the

<sup>1</sup> It is unclear why the table is called "correspondences for initial / pre-vocalic oral stops", since it also discusses the reflexes of affricates and \*s-.

whole set of phonemes, lacking some consonants and vowels. Once again I must reiterate that, without a detailed reconstruction of Proto-AA, it is useless to talk about the lack of evidence for subgrouping. Existing etymological collections mainly represent the better known AA languages; therefore, it is also unsurprising that many lexical isoglosses which could be used for classification have not yet been found.

Having dismissed traditional methods of classification, but without explaining, for example, why the AA tree proposed by Diffloth is not acceptable, PS turns his attention to lexicostatistics and my AA classification, mentioning that it is, in some aspects, similar to Pinnow [1960] and earlier version of Diffloth [1974]. To improve upon these results, he decided to check my etymological identifications and also apply the GlotPc.exe package (developed by Jacque Guy) for lexicostatistics<sup>2</sup>. GlotPc is not well known, and, upon conducting an Internet search, I have not been able to find any examples of its applications apart from PS's work on the AA languages; it would have been prudent on the part of the author to explain the basic principles of this practically unknown program which, on top of everything, failed to yield a reasonable classification for the AA family. No explanations have been presented by PS.

Upon first glance, PS's logic seems to be reasonable: he "decided to replicate the study, but with a smaller, more manageable set of data, in this case 36 languages. A major concern being to use only languages where both the sound correspondences and the major contact languages are reasonably well understood / documented" (p. 129). Consequently, PS uses about 30% of languages from my study, ignoring not only recently discovered ones, but also some well known languages, including those that have been studied by PS himself.

Lexicostatistics is not a very sophisticated procedure; still, it is a statistical analysis whose results depend significantly on the amount of data. It would, therefore, not be surprising if the results of studies based on 30 and a hundred languages were somewhat different. However, as we will see later, this is not the case.

PS's classification is generated based on the lexicostatistical matrix (figure 7, p. 131) in which some languages reveal closer links with Katuic / Bahnaric than other languages from the same branch; thus, for example, Semai and Semelai, but not Jahai.

The Aslian part of Figure 7 forms the following matrix:

	Semelai	Jahai
Semai	40	30
Semelai	×	24

Its interpretation suggests the following tree:

1. Semai, Semelai
2. Jahai

PS writes: "My data set, with cognate assignments and commentary is publicly available via my website..." (p. 129). Five Aslian languages are represented in this database, with most of the forms and etymologies identical to [Peiros 2004]<sup>3</sup>.

StarLing analysis of PS's Aslian data from his website (no changes have been made, even in the cases of my original typos) generates the following matrix:

<sup>2</sup> In my own work, I rely on Sergei Starostin's StarLing with its more than 20 years of success.

<sup>3</sup> I have never heard about etymologies changing their authorship upon the actual forms (including some typos) being copied from one source into another.

	Kensiw	Jahai	Ja-Hut	Semai	Temiar	Semelai
Kensiw	×	63	48	37	38	33
Jahai	63	×	53	43	44	30
Ja-Hut	48	53	×	56	50	46
Semai	37	43	56	×	51	46
Temiar	38	44	50	51	×	34
Semelai	33	30	46	46	34	×

The tree constructed based on this matrix is:

1. Southern: Semelai
2. (i) Northern: Kensiw, Jahai;
- (ii) Senoic: a. Ja-Hut, Semai,  
b. Temiar.

This classification of Aslian languages is identical to the one in [Peiros 2004] and very close to [Diffloth 2005]. Predictably, StarLing generates a better result than GlotPc.

The Bahnaric group is obviously the largest in AA, and is formed by at least five branches [Diffloth 2005, Peiros 2004]. In his paper PS uses only three languages: Bahnar, Jru, and Stieng (they belong to three different branches), and gets the following matrix:

	Jru	Stieng
Bahnar	50	47
Jru	×	41

There are nine Bahnaric languages in PS's database. About 15% of the etymologies are different from [Peiros 2004], mainly because PS identifies many forms as borrowings. I do not agree with most of these identifications, but even if we accept them, StarLing generates the same tree as in [Peiros 2004]. Could this be just a coincidence?

	Bahnar	Tampuan	Jeh	Sedang	Jru	Nhaheun	Sre	Stieng
Bahnar	×	67	58	59	54	51	45	54
Tampuan	67	×	54	53	52	49	49	51
Jeh	58	54	×	60	59	58	44	49
Sedang	59	53	60	×	52	50	37	40
Jru	54	52	59	52	×	88	41	46
Nhaheun	51	49	58	50	88	×	40	43
Sre	45	49	44	37	41	40	×	67
Stieng	54	51	49	40	46	43	67	×

## Bahnaric:

- A. Southern: Sre, Stieng.
- B. i. Western: Jru, Nhaheun;
- ii. a. Bahnar, Temiar;
- b. Northern: Jeh, Sedang.

Let us now combine the matrices for the Aslian and the Bahnaric languages:

	Ken	Jah	Ja-H	Semai	Tem	Semel	Bah	Tam	Jeh	Sed	Jru	Nha	Sre	Sti
Kensiw	×	63	48	37	38	33	20	16	19	18	17	15	18	17
Jahai	63	×	53	43	44	30	20	18	19	18	19	15	21	19
Ja-Hut	48	53	×	56	50	46	29	27	33	30	33	29	29	30
Semai	37	43	56	×	51	46	29	23	27	26	25	23	26	29
Temiar	38	44	50	51	×	34	23	23	22	27	22	20	22	23
Semelai	33	30	46	46	34	×	33	33	32	31	32	31	28	29
Bahnar	20	20	29	29	23	33	×	67	58	59	54	51	45	54
Tampuan	16	18	27	23	23	33	67	×	54	53	52	49	49	51
Jeh	19	19	33	27	22	32	58	54	×	60	59	58	44	49
Sedang	18	18	30	26	27	31	59	53	60	×	52	50	37	40
Jru	17	19	33	25	22	32	54	52	59	52	×	88	41	46
Nhaheun	15	15	29	23	20	31	51	49	58	50	88	×	40	43
Sre	18	21	29	26	22	28	45	49	44	37	41	40	×	67
Stieng	17	19	30	29	23	29	54	51	49	40	46	43	67	×

Here we see that Jahai (and Kensiw) have lower percentages with Bahnaric languages than with those of other Asian languages. PS noticed the same pattern in other groups as well and suggested the following explanation: “It is possible that to some extent the data of individual languages includes some cases of more rapid lexical replacement that are complicating the picture. This said, we still have a stark geographic correlation: the closer any language or branch is to Katuic or Bahnaric, the higher percentage of cognates will be found with Katuic and/or Bahnaric” (p. 130). Unfortunately, these two sentences contain at least two misleading statements.

1. It is not clear how the geographical proximity is estimated, but the distance between Jahai and Katuic/Bahnaric is shorter, since Semai and Semelai are actually spoken further to the south from Jahai.
2. PS talks about “rapid or faster rate of change” on two occasions. In note 5 he wrote: “Peiros mentions the possibility of faster rate of change in Nicobarese, but did not take it into account” (p. 129). My Russian text, however, says: “It is not impossible that the suggested 1,500 years gap [between Nicobarese and other AA languages] will be reduced in the process of further etymological studies” [Peiros 2004: 25].

Let us discuss this second occasion. What does “rapid rate change” actually mean? No reliable examples of such phenomena are truly known to linguists<sup>4</sup>, making this expression un-

<sup>4</sup> See the discussion of Blust’s [2000] proposal in [Peiros 2000].

substantiated. Yet, at the same time, in many lexicostatistical matrices one can indeed see cases when one language's percentage of words shared with other related languages seems to be much lower than expected.

Modern lexicostatistics is not only about counting percentages; it also includes a study of relevant etymologies, and there should be no disagreements between these two tasks. Such an approach was introduced by Sergei Yakhontov in his public talks in the 1970s. One of Yakhontov's examples was as follows: "Suppose that we have three languages (A, B, and C) which share 15% similarities, however, etymologically identical words in A and B are not the same as etymologically identical words in A and C. In such cases we cannot claim that all of these languages form the same family; related languages always share words of common origin"<sup>5</sup>.

A (diagnostic) lexicostatistical list of any language consists of several types of words. For example, in English there are:

1. words of the same origin as in other Germanic languages included in the LS study ('all', 'ashes', 'bite', etc.);
2. words with Germanic etymologies whose reflexes in other languages are not represented in their LS lists ('kill', 'meat', 'smoke', etc.);
3. borrowings from Old Norse ('bark', 'skin', etc.) and Old French ('mountain', 'round');
4. words of unknown origin ('big', 'bird', 'dog', etc.).

In the case of English, with its long tradition of etymological studies, most of the words on the list belong to groups 1–3, and only a few have an unknown origin. In lesser known languages one would expect a much higher percentage of words from group 4. Some words could also be missing altogether (if they are not represented in available sources, or if the corresponding notion is not known to the speakers<sup>6</sup>).

Starostin (2000) suggests a uniform treatment for borrowings and missing forms that interprets them as lack of information for LS-calculations: situations in which we are not dealing with uninterrupted development from the previous historical stage. If so, the position of a language in a LS-classification is determined by the ratio of words from group 1 to the combined amount of words in groups 2 and 4 (but not 3).

A word may be called a "borrowing" only if its origin is known, i. e. linguists have identified a word in another language which had served as the source of borrowing. If such a source is not identified, we cannot truly call it a "borrowing" and must place it in group 4 rather than in group 3. In doing this, we increase the amount of unique words, and thus increase the distance of the respective language from other related languages.

Jahai reveals exactly this type of situation, since many of its words are not found in other Aslian or AA languages (*kawaw* 'bird', *ʔam* 'breast', etc.). Therefore, it is possible that at least some of these words were borrowed from unidentified sources, and thus, must be attributed to group 3 rather than group 4. It is currently impossible to identify these words, since their traces are seen only in the LS matrix. No traces of such potential "borrowings" are detected in other Aslian languages which, probably, did not experience similar influence from unknown donor(s). The same explanation can be offered for other similar cases mentioned by PS (Mal and Khasi Wa).

---

<sup>5</sup> This is according to my recollection of Yakhontov's position.

<sup>6</sup> A good example is the notion *horn*, unknown in the languages of Australia and Oceania.

Summarizing PS's lexicostatistical investigation, we can conclude:

1. StarLing provides much better results than GlotPc.
2. When applied to a larger set of languages from PS's electronic database, StarLing generates results similar to those given in [Peiros 2004], even if one accepts PS's own etymological assignments (although I do not agree with most of them).
3. Both matrices (generated by StarLing and GlotPc) reveal no indisputable traces of influence from Katuic / Bahnaric on other AA languages.

The main goal of PS's paper is to localize the AA homeland. In addition to "lexical" arguments discussed above, he also relies on geographical distribution of modern languages to conclude: "... since the greatest number of AA branches is spoken along an axis that runs roughly southeast to northwest along the middle course of the Mekong river, it is reasonable to suggest that the AA language dispersed along and out from this axis. This theory is provisionally called the *Austroasiatic Central Riverine Hypothesis*" (p. 118).

Looking at the flat map of the region like the one given on p. 118 of PS's paper, it may be easy to imagine the Munda migrating to India directly from Laos or Vietnam, forgetting that even today such a journey is not possible due to the rough terrain in between.

Localization of a homeland requires more than simply pin-pointing the place of the highest linguistic diversity. A multifocal approach had been used by Diffloth [2005], but, obviously, was not appreciated by PS. Diffloth mentions three linguistic issues<sup>7</sup> related to localization:

1. "A rich lexicon for rice terminology is reconstructible to PAA, making it evident that the people who spoke this language were thoroughly familiar with rice agriculture" [Diffloth 2005: 77].
2. "In the reconstructed PAA lexicon there are names for animal species which are restricted to humid tropics. The floral vocabulary would probably lead to the same conclusion, but it is more difficult to reconstruct at present" (ibid.).
3. "The geographic distribution of the thirteen branches of AA (Munda plus MK) would imply a center of greatest historical diversity in the region which encompasses the fertile flood plains of the Irrawaddy in Burma and the lower Bramaputra in Assam and Bangladesh" (p. 78).

Remarkably, Diffloth was not able to reconcile his three sources of information, so the question was intentionally left unanswered.

After mentioning Diffloth's hypothesis, PS switches to the localization proposed by Peiros and Shnirelman [1998], saying "[they] have asserted that the AA lexicon indicates a non-tropical, non-coastal location. Combining this idea with data about the mid-Holocene climate, and supposed lexical isoglosses with Hmong-Mien languages (some two dozen or more supposed resemblances), they proposed an origin on the mid-Yangtse river" (p. 119).

However, Shnirelman's and my own view as it was given in the paper is quite different. Having studied the linguistic data, we came to the conclusion that, since (i) the reconstructed AA lexicon does not contain any words associated with the sea coast, and (ii) there are no clear indications of any tropical environment, "...the Austro-Asiatic homeland was located in mainland eastern Eurasia in an area with a non-tropical climate: the sub-tropics or the mountainous region" and, further: "... was located somewhere near the Middle Yangtze River Val-

---

<sup>7</sup> The paper also discusses the somatic diversity of AA people.

ley” [Peiros, Shnirelman 1998: 380]<sup>8</sup>. It is not clear why PS has omitted the suggested possibility of mountainous regions near the Yangtze, the importance of which will be discussed later.

Let us, however, see what else PS has to say on the subject: “This view dovetails with philologists’ suggestions of various AA etymologies for words of uncertain origins in classical languages of China...” (p. 119)<sup>9</sup>. Here PS mentions Schuessler’s [2007] discovery of hundreds of AA loans in Chinese. Instead of offering specific explanations for these loans, PS prefers to brush them off: “Generally these philological arguments have the following character: there are a handful of close phonetic and semantic matches which are intriguing, plus there are scores or more of vague resemblances which cannot be organized into any system of regular correspondences” (*ibid.*). Thus, a handful of good matches is equaled to a total lack of information, simply because “unfortunately this is what one expects to find when comparing unrelated but typologically similar languages” (*ibid.*).

Personally, I am not aware of any such meaningless collections of forms and have reasons to doubt that such a situation is possible. My own experience suggests the opposite situation: it is *very* difficult to find a significant number of good matches between any two languages (comparable in quantity and quality to Schuessler’s collection of Chinese-AA parallels) and show that they have been caused only by chance.

At the end of his discussion PS writes: “After all, the various philological arguments suppose localizing proto-AA in places where little or none of the diversity of the AA phylum is presently found. There are no AA speaking communities along the middle Yangtze or in China’s southeastern provinces” (p. 119).

This is another misleading statement. It is true that there are no *modern* AA speakers in the region, since today, the majority of the population speaks Chinese, and not much is known about the languages spoken there about 7,000 years ago. But from various Chinese sources we know that as late as two thousand years ago several states, such as Yuè, Chǔ, or Bā, were populated by non-Chinese peoples (although not all of them were necessarily speakers of AA). Intense processes of Chinese migration and cultural assimilation that started around that time, eventually wiped out the original ethnic diversity of that region. Thus, PS’s argument about the modern situation is hardly of any value.

According to glottochronology, the disintegration of Proto-AA began about 8,000 years ago or somewhat later, depending on the treatment of Nicobarese material [Peiros 2004: 34]. Diffloth estimates that the disintegration took place about a thousand years later (5,000 BCE), but provides no reasons for this suggestion [Diffloth 2005: 79]. It should be noted that glottochronology provides information on the time range during which the protolanguage has already begun to disintegrate, which often coincides with the beginning of migrations out of the original homeland; a more or less homogeneous protolanguage may be significantly older.

Employing solely linguistic methods, homelands may be localized in three different spaces — geographical, ecological, and cultural — which do not necessarily coincide with each other [Peiros 1997]. An additional, «ethnic», space will be introduced later.

**Geographic space:** to localize a certain proto-language’s homeland in this space we can use the method of «geographic pin-pointing», based on tracing possible migrations of speakers according to their current location on geographic maps. A region in which genetically diverse

---

<sup>8</sup> In his presentation of this suggestion, PS, for some reasons, has omitted the mountainous regions, but added a misleading connection with Miao-Yao.

<sup>9</sup> It is hard to understand what is PS’s exact meaning when he talks about “classical languages of China”; only Classical Chinese is known in the area, although from much later periods.

languages are represented is more likely to have been the original homeland than a linguistically homogeneous region. This method depends on a previously established genetic classification of languages and provides information on the relative distances that separate them.

However, «geographic pin-pointing» also has some serious theoretical restrictions, since the observed diversity does not always represent the original situation:

1. Languages from different branches could have been brought to the same area over the course of a single or several independent migrations.
2. Linguistic diversity may be caused by non-genetic reasons. E. g., Chinese dialects are more diverse in the south of the country, although their common ancestor was spoken in the north.
3. «Refuge zones», where various languages are cramped in a small area, may also be a cause for erroneous interpretation.

Modern distribution of AA languages suggests that the ancient speakers predominantly moved along river valleys and not across mountains: they are found on the Brahmaputra, Irrawaddy, Salween, Mekong, and their tributaries. Only Aslian and Nicobarese languages do not follow this pattern. Later movements, perhaps, followed other routes, as is shown by the spread of Vietnamese or Monic.

Based on general considerations, one can suggest that the starting point of the original AA migrations was the area of modern Southern Sichuan, or somewhat to the west from there, not far from the Yangtze valley. From there, ancient AA speakers were presumably able to reach the valleys of the major rivers of Southeast Asia<sup>10</sup>. It also possible that they moved down the Yangtze, although this area is now occupied by Chinese speakers who migrated there relatively late.

Since many modern AA languages are spoken in the mountains, it is possible that this reflects the original situation: the homeland was located not at the bottom of a tropical valley, but on much higher grounds.

**Ecological space:** localization in a certain ecological space is based on the results of lexical reconstruction, as well as the assumption that, if a certain word is used in a language, its speakers are actually familiar with the exact object/notion that it represents<sup>11</sup>. Analysis of such words can provide us with some information about the natural world which surrounded its speakers, e. g.:

- a list of wild plants typical of a particular zone, like ‘Siberian cedar’, ‘mangrove’, or ‘mulga trees’;
- a list of wild animals, such as ‘crocodile’, ‘polar bear’, ‘koala’, etc.;
- sometimes, names of specific natural phenomena, such as ‘monsoons’, ‘Northern lights’, ‘earthquakes’, etc.

A justified list of such words can, in theory, describe the ecological zone (an intersection of climatic, floral and faunal zones) in which the protolanguage was spoken. It is clear, however, that one cannot expect to find this zone on a modern map; only data from paleoclimatology, paleobotany, and paleozoology would make it possible.

<sup>10</sup> It is still nor clear how they have managed to cross the ridges between the valleys.

<sup>11</sup> This does not imply, however, that the absence of a certain word necessarily marks the absence of the corresponding idea.

«Pin-pointing» an ecological zone for AA requires a thoroughly representative etymological dictionary, which is yet to be completed. As it has been already mentioned above, Diffloth [2005: 77] presents a list of “... names for animal species which are restricted to humid tropics”<sup>12</sup>:

- #*tərkuał* Land, or tree monitor *Varanus bengalensis*, or *V. nebulosus*
- #*(bən-)jo:l*, #*j(ərm)o:l* Ant-eater *Manis javanica*
- #*tənriak* Buffalo *Bubalus bubalis*
- #*tənyu:?* Bear-cat *Arctitis binturong*
- #*kiač* Mountain goat *Capricornis sumatrensis*
- #*kacian* Asian elephant *Elephas maximus*
- #*mra:k* Peacock *Pavo muticus*
- #*rəma:s* Rhinoceros *Dicerorhinus sumatrensis*
- #*dəkan* Bamboo-rat *Rhizomys sumatrensis*

Upon first glance, the list should support a tropical localization, but a more detailed interpretation shows that the consequences are not quite as straightforward. Two problems, in particular, are of major concern:

(a) How precisely can one reconstruct the zoological identifications of these faunal terms?

As a first example, let us take *Varanus*<sup>13</sup>. *Varanus bengalensis* or *Varanus nebulosus* is found in a zone that stretches from Iran to Indonesia, including South China. Is it, however, possible to demonstrate that the semantic reconstruction should be specifically *Varanus bengalensis*? As far as I can tell from known linguistic data<sup>14</sup>, the actual spectrum of the meanings for this item can be ‘any big lizard (including monitors)’, which makes the distribution much wider.

Next, *Manis javonica* is quite similar to *M. pentadactylus* (Chinese pangolin), which is found from East Nepal to Central and South China, including Taiwan. If we reconstruct the meaning as simply ‘ant-eater’, there will no longer be any strong link with the tropics.

Finally, elephants are still found in Yunnan, and pink-tusked elephants (*Elephas maximus rubridens*) were hunted in Central China as late as the 14th century BCE.

(b) Geographical distribution of animal species 7–8 thousand years ago must have been different from what we observe today. It is quite probable that animals listed by Diffloth, or, perhaps, different, but very similar species, lived not only in India and Burma, but also in Sichuan and Yunnan, as well as in Southern China and Vietnam.

**Cultural space:** any speech community is characterized by a certain technological repertoire and a material cultural complex that is based on it. Under this distinction, the fact that a certain community breeds sheep belongs to its technological repertoire, but the exact manner in which this is done is a characteristic of the corresponding material culture. Technological repertoire is not connected with a specific community — rather, many communities share it. The material culture of a community depends on its technological repertoire, but its elements are often community-specific.

---

<sup>12</sup> The forms marked with # are not fully reconstructed, but represent reasonable approximations.

<sup>13</sup> Since I am not a zoologist, this information had to be extracted from Wikipedia.

<sup>14</sup> Diffloth does not list the reflexes of his reconstructions in daughter languages.

It seems that using linguistic methods, one can identify only the technological repertoire of a community, but not its specific material culture. Therefore, if there are several similar archaeological cultures in the same region, it is almost impossible to connect the speech community with a specific one of them, excluding the others.

Any food-producing technological repertoire can be characterized by its achievements in agriculture and stock-raising. Information on agriculture includes:

- a list of cultivated plants and their relative value (for instance, rice is a common crop, millet is not);
- a description of agricultural techniques, such as slash-and-burn practice, etc.

Similarly, information on stock-raising consists of:

- a list of domesticated animals and their relative roles in the community;
- a description of animal usage.

If one can find additional information, such as knowledge of metals, data on hunting, fishing, seafaring, etc., the description of the technological repertoire could be quite detailed.

The AA cultural lexicon has two particularly prominent features:

- (a) there are no names for any metals;
- (b) there is a set of words closely associated with rice agriculture:

Rice plant  $^{\#}(kə)ba:?$   
 Rice grain  $^{\#}rəŋko:?$   
 Rice outer husk  $^{\#}cəŋka:m$   
 Rice inner husk  $^{\#}kəndək$   
 Rice bran  $^{\#}phe:?$   
 Rice-compliment  $^{\#}kəntu:?$   
 (cooked food other than rice)

As of now, we have no indications that proto-AA speakers might have borrowed their agricultural knowledge from another language, or have inherited it from their ancestral group<sup>15</sup>.

It is important to mention that the set indicates a well-developed knowledge of rice agriculture rather than merely an incipient stage. Some scholars believe that rice was domesticated in two different regions: South Asia and the middle Yangtze, from where it was brought to lower Yangtze around 6,000 BCE [Fuller et al. 2008]. The suggested middle Yangtze localization of the AA homeland does not contradict this belief, but possibly indicates a long-term existence of AA speakers not far from the original area of rice domestication.

In my AA database there is also at least one word for a root vegetable ( $^{\#}təro \sim ^{\#}təraw$ ) whose reflexes in daughter languages display meanings such as 'taro' or 'arum'. No words for other crops, such as 'millet', are found.

**Ethnic Space:** localization of a certain protolanguage X in an ethnic space surmises identifying its linguistic neighbors. This can be done by means of analysing borrowings from and into X.

It is possible that in the future we will be able to detect other areas of contacts, but at the current stage of our knowledge only two directions of research are relevant: potential contacts with Proto-Austronesian and Proto-Sino-Tibetan.

The problem of Austronesian contacts cannot be solved before we have worked out the details of the internal classification of Austric.

<sup>15</sup> Terms for rice could not be reconstructed in the etymological list for Proto-Austric that was compiled by Sergei Starostin and myself about ten years ago.

It has been suggested ([Peiros 1998] and elsewhere) that the Austric superfamily is formed by two branches: Austronesian + Tai-Kadai (Austro-Tai) and Austro-Asiatic + Miao-Yao. Proto-Miao-Yao and Proto-Tai-Kadai are much younger than Proto-AA, with only Proto-Austronesian being roughly of the same age. S. Starostin and I have proposed a set of phonological correspondences between Proto-Austro-Asiatic and Proto-Austronesian, based on a few hundred discovered lexical comparisons<sup>16</sup>. Some of these comparisons may, in fact, represent borrowings, but the current level of Austric comparative linguistics does not permit their identification.

The Austric theory, even at its current stage of development, suggests that the AA homeland must be located much further east than India and Burma, where other Austric languages are spoken. Otherwise, we would need to accept migrations of other Austric groups to the east.

There is a significant number of words that are similar between Proto-AA and Proto-Sino-Tibetan, which I plan to discuss in a separate publication. Some of them can also be found in Schuessler's etymological dictionary of Chinese [2007] and in [Peiros 1998: 226–27]. Their preliminary analysis suggests that the protolanguages were in contact, with borrowings travelling in both directions. Important cultural terms, such as the words for rice, were, however, excluded from this process. If my suggestion that the Sino-Tibetan homeland was located somewhere in the Sub-Himalayas [Peiros 1998] is correct, the AA homeland must have been nearby.

As may be seen, the discussed evidence altogether supports the proposal that the AA homeland was located somewhere not far from the mid-Yangtze valley, probably in the nearby mountains in modern Sichuan, as has been suggested by Peiros and Shnirelman in 1998.

## References

ALVES, Mark J. (forthcoming). A look at North-Central Vietnamese. In *Papers from the Thirteenth Annual Meeting of the Southeast Asian Linguistics Society*. Arizona State University Press. Pre-publication electronic version: [http://www.geocities.com/malves98/Alves\\_Vietnamese\\_Northcentral.pdf](http://www.geocities.com/malves98/Alves_Vietnamese_Northcentral.pdf).

ALVES, Mark J.; & Nguyễn, Duy Huong. (2007). Notes on Thanh-Chưong Vietnamese in Nghệ-An province. In M. ALVES, M. SIDWELL, & D. GIL (Eds.), *SEALS VIII: Papers from the 8<sup>th</sup> annual meeting of the Southeast Asian Linguistics Society 1998* (pp. 1–9). Canberra: Pacific Linguistics, The Australian National University, Research School of Pacific and Asian Studies. Electronic version: [http://paclng.anu.edu.au/catalogue/SEALSVIII\\_final.pdf](http://paclng.anu.edu.au/catalogue/SEALSVIII_final.pdf).

BLUST, R. 2000. Why lexicostatistics doesn't work: the "universal constant" hypothesis and the Austronesian languages. In C. RENFREW, A. MCMAHON and L. TRASK (eds.) *Time Depth in Historical Linguistics*, Cambridge: The McDonald Institute for Archaeological Research, vol. 2: 311–332.

EMENEAU, M. B. 1947. Homonyms and puns in Annamese. *Language*, 23 (3), 239–244.

DIFFLOTH, G. 1974. Austro-Asiatic Languages. *Encyclopedia Britannica*.

DIFFLOTH, G. 2005. The contribution of linguistic palaeontology to the homeland of Austro-Asiatic. In: L. SAGART, et al. (eds.) *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. New York: Routledge. Pp 79–82.

FULLER, D., Ling QIN, and E. HARVEY. 2008. Evidence for a late onset of agriculture in the Lower Yangtze region and challenges for an archaeobotany of rice. In: A. SANCHEZ-MAZAS et al. (eds.) *Past Human Migrations in East Asia*. New York: Routledge, pp. 40 – 83.

HOÀNG, Thị Châu. 1989. *Tiếng Việt trên các miền đất nước: Phương ngữ học* [Vietnamese in different areas of the country: Dialectology]. Hà Nội: Khoa học xã hội.

HONDA, Koichi. 2006. F0 and phonation types in Nghe Tinh Vietnamese tones. In P. WARREN & C. I. WATSON (eds.), *Proceedings of the 11<sup>th</sup> Australasian International Conference on Speech Science and Technology* (pp. 454–459). Auckland, New Zealand: University of Auckland. Electronic version: <http://www.assta.org/sst/2006/sst2006-119>.

<sup>16</sup> The comparisons exist in the form of computer databases and have been used several times in presentations given in Moscow and Santa Fe in 2003 – 2004, but, for the most part, remain unpublished.

NGUYỄN, Đinh-Hoà. 1995. *NTC's Vietnamese-English dictionary* (updated ed.). NTC language dictionaries. Lincolnwood, Illinois: NTC Pub. Press.

PEIROS, I. 1989. Linguistic history of the Asian East (in Russian). In I. PEIROS (ed.), *Lingvisticheskaja rekonstruktsija i drevnejshaja istorija Vostoka* (Linguistic Reconstruction and the Ancient History of the Orient), 4 vols. Moscow: Nauka, vol. 1: 167 – 178.

PEIROS, I. 1997. Is a Correlation between Archaeological and Linguistic Reconstructions Possible? In R. BLENCH and M. SPRIGGS (eds.). *Archaeology and Language*, vol. 1. London and New York: Routledge, pp. 149 – 157.

PEIROS, I. 1998. *Comparative Linguistics in Southeast Asia*. Pacific Linguistics, Series C-142.

PEIROS, I. 2000. Family diversity and time depth. In C. RENFREW, A. McMAHON and L. TRASK (eds.). *Time Depth in Historical Linguistics*, vol. 1: 75–108. The McDonald Institute for Archaeological Research. Cambridge.

PEIROS, I. 2004. *A Genetic classification of Austroasiatic languages* (in Russian). Moscow: RGGU.

PEIROS, I and V. SHNIRELMAN. 1997. Rice in South-East Asia: regional interdisciplinary approach. In R. BLENCH and M. SPRIGGS (eds.), *Archaeology and Language* 2. London and New York: Routledge, pp. 379–389.

PHAM, Andrea Hoa. 2005. Vietnamese tonal system in Nghi Loc: A preliminary report. In C. FRIGENI, M. HIRAYAMA, & S. MACKENZIE (eds.), *Toronto working papers in linguistics: Special issue on similarity in phonology* (Vol. 24, pp. 183–459). Auckland, New Zealand: University of Auckland. Electronic version: [http://r1.chass.utoronto.ca/twpl/pdfs/twpl24/Pham\\_TWPL24](http://r1.chass.utoronto.ca/twpl/pdfs/twpl24/Pham_TWPL24)

PINNOW, H.-J. 1963. The Position of the Munda languages within the Austroasiatic language family. In: H. L. SHORTO (ed.) *Linguistic Comparison in Southeast Asia and the Pacific*. London: SOAS, pp. 140–152.

SCHUESSLER, A. 2007. *ABC Etymological Dictionary of Old Chinese*. Honolulu: University of Hawaii Press.

SIDWELL, P. 2010. The Austroasiatic central riverine hypothesis. *Journal of Language Relationship*, 4: 117–134.

STAROSTIN, S. 2000. Comparative-historical linguistics and lexicostatistics. In C. RENFREW, A. McMAHON and L. TRASK (eds.). *Time Depth in Historical Linguistics*, Cambridge: The McDonald Institute for Archaeological Research, v2, pp. 223–259.

THOMPSON, Laurence C. 1991. *A Vietnamese reference grammar*. Seattle: University of Washington Press. Honolulu: University of Hawaii Press. (Original work published 1965).

VŨ, Thang Phuong. 1982. Phonetic properties of Vietnamese tones across dialects. In D. BRADLEY (ed.), *Papers in Southeast Asian linguistics: Tonation* (Vol. 8, pp. 55–75). Sydney: Pacific Linguistics, The Australian National University.

VƯƠNG, Hữu Lẽ. (1981). Vài nhận xét về đặc điểm của vần trong thô âm Quảng Nam ở Hội An [Some notes on special qualities of the rhyme in local Quang Nam speech in Hoi An]. In *Một Số Vấn Đề Ngôn Ngữ Hoc Việt Nam* [Some linguistics issues in Vietnam] (pp. 311–320). Hà Nội: Nhà Xuất Bản Đại Hoc và Trung Hoc Chuyên Nghiệp.

В одной из своих недавних публикаций («Гипотеза “центрально-речной” прародины австроазиатской семьи», ВЯР 4), П. Сидвелл выдвинул новую теорию, согласно которой исходную прародину праавстроазиатского языка уместнее всего локализовать в долине р. Меконг (Индокитай и прилегающие территории). Основной целью настоящей статьи является анализ выводов, полученных Сидвеллом, как чисто лингвистический (с опорой на лексикостатистику), так и в сравнении с данными о географической дистрибуции, культурной и экологической специфике, а также этническом окружении праавстроазиатского языка и его носителей. Автор приходит к выводу, что общий комплекс данных чаще противоречит теории Сидвелла, чем подтверждает ее, и что, таким образом, нет никакой необходимости ни отвергать старую гипотезу о возможной прародине австроазиатской семьи в районе среднего течения р. Янцзы (И. Пейрос и В. Шнирельман), ни пересматривать существующие модели внутренней классификации языков австроазиатской семьи (Г. Диффлот, И. Пейрос) для того, чтобы лучше согласовать их с «центрально-речной» гипотезой, во многом скорее умозрительной, чем подтверждаемой конкретными фактами.

**Ключевые слова:** австроазиатские языки, сравнительно-исторический метод, лексикостатистика, лингвистическая прародина.

