

Российский государственный гуманитарный университет
Russian State University for the Humanities



RSUH/RGGU BULLETIN

№ 5(127)

Academic Journal

Series:

Philology. Linguistic Studies.
Journal of Language Relationship
Issue 11 (2014)

Moscow 2014

ВЕСТНИК РГГУ

№ 5(127)

Научный журнал

Серия «Филологические науки. Языкознание» /
«Вопросы языкового родства»
Выпуск 11 (2014)

Москва 2014

УДК 81(05)
ББК 81я5

Главный редактор
Е.И. Пивовар

Ответственный секретарь
Б.Г. Власов

Редакционный совет серии
«Филологические науки. Языкознание» / «Вопросы языкового родства»

Председатель Вяч. Вс. Иванов (Москва – Лос-Анджелес)
Х. Айхнер (Вена)
М. Е. Алексеев (Москва)
В. Блажек (Брно)
У. Бэкстер (Анн Арбор)
В. Ф. Выдрин (Санкт-Петербург)
М. Гелл-Манн (Санта Фе)
Ф. Кортландт (Лейден)
А. Лубоцкий (Лейден)
Дж. П. Мэллори (Белфаст)
Л. Хайман (Лос-Анджелес)

Редакционная коллегия серии:

В. А. Дыбо (главный редактор)
Г. С. Старостин (заместитель главного редактора)
Т. А. Михайлова (ответственный секретарь)
К. В. Бабаев
А. В. Дыбо
А. С. Касьян
С. В. Кулланда
И. С. Якубович

ISSN 1998-6769

© Российский государственный
гуманитарный университет, 2014

УДК 81(05)
ББК 81я5

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. В. А. Дыбо. — М., 2014. — № 11. — xii + 225 с. — (Вестник РГГУ: Научный журнал; Серия «Филологические науки. Языкознание»; № 5(127)).

Journal of Language Relationship: International Scientific Periodical / Russian State University for the Humanities; Russian Academy of Sciences. Institute of Linguistics; Ed. by V. A. Dybo. — Moscow, 2014. — No. 11. — xii + 225 p. — (RSUH Bulletin: Scientific Periodical; Linguistics Series; No. 5(127)).

ISSN 1998-6769

<http://www.jolr.ru/>
journal@jolr.ru

Дополнительные знаки: С. Г. Болотов
Add-on symbols by S. G. Bolotov

Подписано в печать 14.01.2014. Формат 60×90/8.
Бум. офсетная.
Печать офсетная. Тираж 1050 экз.
Заказ №

Отпечатано в полном соответствии с качеством
предоставленного оригинал-макета
в «Наша Полиграфия», г. Калуга, ул. Грабцевское шоссе, 126
Лиц. ПЛД № 42-29 от 23.12.99

Table of Contents / Содержание

Table of Contents / Содержание	vii
Contributors / Сведения об авторах	ix
Note for Contributors / Будущим авторам	x
Note from the Editors / От редакции	xi

Articles / Статьи

<i>George Starostin</i> . Macro-comparative linguistics in the 21st century: state of the art and perspectives	1
[Г. С. Старостин. Макрокомпаративистика в XXI веке: текущее состояние и перспективы развития]	
<i>Kirill Babaev</i> . Verbal morphosyntax in Southwest Mande: elements of reconstruction	13
[К. В. Бабаев. К реконструкции глагольного морфосинтаксиса в праюгозападном манде]	
<i>Václav Blažek</i> . Indo-European nominal inflection in Nostratic perspective	19
[В. Блажек. Индоевропейское именное склонение в ностратической перспективе]	
<i>С. Ю. Бородай, И. С. Якубович</i> . Корпусные методы дешифровки анатолийских иероглифов	39
[<i>Sergey Boroday, Ilya Yakubovich</i> . Corpus-based methods for deciphering Anatolian hieroglyphics]	
<i>А. С. Касьян</i> . К формальной генеалогической классификации лезгинских языков (Северный Кавказ)	63
[<i>Alexei Kassian</i> . Towards a formal genealogical classification of Lezgian languages of the North Caucasus]	
<i>Sergei Kullanda</i> . External relations of Scythian	81
[С. В. Куланда. Внешние связи скифского языка]	
<i>Johann-Mattis List</i> . Investigating the impact of sample size on cognate detection	91
[<i>J.-M. List</i> . К вопросу о влиянии размера лексической выборки на обнаружение этимологических когнатов]	
<i>Sergei Nikolaev</i> . Toward the reconstruction of Proto-Na-Dene	103
[С. Л. Николаев. Материалы к реконструкции пра-на-дене]	
<i>Luobbal Sámmol Sámmol Ánte (Ante Aikio)</i> . On the reconstruction of Proto-Mari vocalism	125
[А. Айкио. К реконструкции прамарийского вокализма]	
<i>Alexander Militarev</i> . A complete etymology-based hundred wordlist of Semitic updated: Items 75–100	159
[А. Ю. Милитарев. Новый вариант подробной этимологизации стословного списка семитских языков: элементы 75–100]	

Discussion Articles / Дискуссионные статьи

<i>К. И. Поздняков.</i> О пороге родства и индексе стабильности в базисной лексике при массовом сравнении: атлантические языки	187
[<i>Konstantin Pozdniakov.</i> On the threshold of relationship and the “stability index” of basic lexicon in mass comparison: Atlantic languages]	
<i>Д. Р. Лещинер.</i> Заметки к статье К. И. Позднякова «О пороге родства...»	210
[<i>Dmitry Leshchiner.</i> Notes on Pozdniakov’ paper “On the threshold of relationship...”]	
<i>Г. С. Старостин.</i> Ответ на статью К. И. Позднякова «О пороге родства...»	215
[<i>George Starostin.</i> Reply to Pozdniakov’ paper “On the threshold of relationship...”]	
<i>К. И. Поздняков.</i> От дальнего родства позиций — к ближнему (ответ на отзывы Г. С. Старостина и Д. Р. Лещинера)	220
[<i>Konstantin Pozdniakov.</i> Reply to Starostin and Leshchiner]	

Contributors

Kirill Babaev — candidate of sciences (Philology), lead researcher, Institute of Oriental Studies, Russian Academy of Sciences (Moscow), babaev@yandex.ru

Václav Blažek — professor, Masaryk University, Brno, blazek@phil.muni.cz

Sergei Boroday — post-graduate student at the Institute of Oriental Studies of the Russian Academy of Sciences (Moscow), sergey_voody@inbox.ru

Alexei Kassian — candidate of sciences (Philology), researcher, Center for Comparative Linguistics, Russian State University for the Humanities / researcher, Institute of Linguistics, Russian Academy of Sciences (Moscow), a.kassian@gmail.com

Sergey Kullanda — candidate of sciences (History), senior researcher, Institute of Oriental Studies of the Russian Academy of Sciences (Moscow), kullanda@java.msk.ru

Johann-Mattis List — post-doctoral researcher, Forschungszentrum Deutscher Sprachatlas, Philipps-University (Marburg), mattis.list@uni-marburg.de

Dmitry Leshchiner — candidate of sciences (Computer Science), software engineer, Google Inc. (Mountain View, CA, USA), dmitry_1111@inbox.ru

Luobbal Sámmol Sámmol Ánte (Ante Aikio) — researcher, Giellagas Institute for Saami Studies, University of Oulu (Finland), ante.aikio@oulu.fi

Alexander Militarev — doctor of sciences (Philology), professor, Center of Ancient Eastern research, Institute of Eastern Cultures and Antiquity, RSUH (Moscow), amilitarev@gmail.com

Sergei Nikolaev — doctor of sciences (Philology), lead researcher, Department of Slavic linguistics, Institute of Slavic studies, Russian Academy of Sciences (Moscow/Novosibirsk), sergenicko@mail.ru

Konstantin Pozdnyakov — doctor of sciences (Philology), professor, Institut national des langues et civilisations orientales (Paris), pozdnyakov@free.fr

George Starostin — candidate of sciences (Philology), head of Department of the history and philology of the Far East, Institute of Eastern Cultures and Antiquity, RSUH (Moscow), head of Department of Oriental studies and comparative-historical linguistics, School for Advanced Studies in the Humanities, RANÉPA (Moscow), gstarst@rinet.ru sds

Ilya Yakubovich — candidate of sciences (Philology), research associate, Institute of World Cultures, Moscow State University, sogdiana783@gmail.com

Сведения об авторах

Айкио, Анте — научный сотрудник института саамских исследований Гиеллагас, Университет Оулу (Финляндия), ante.aikio@oulu.fi

Бабаев, Кирилл Владимирович — канд. филол. наук, ст. науч. сотрудник Института востоковедения РАН (Москва), babaev@yandex.ru

Блажек, Вацлав — проф. университета им. Масарика (Брно), blazek@phil.muni.cz

Бородай, Сергей Юрьевич — аспирант Института востоковедения РАН (Москва), sergey_voody@inbox.ru

Касьян, Алексей Сергеевич — канд. филол. наук, науч. сотрудник сектора анатолийских и кельтских языков Института языкознания РАН, преп. Центра компаративистики ИВКА РГГУ (Москва), a.kassian@gmail.com

Куланда, Сергей Всеволодович — канд. ист. наук, ст. науч. сотрудник Института востоковедения РАН (Москва), kullanda@java.msk.ru

Лещинер, Дмитрий Роальдович — канд. техн. наук, инженер-программист, Google Inc. (Маунтин-Вью, Калиф., США), dmitry_1111@inbox.ru

Лист, Йоханн-Маттис — постдокторант, Марбургский университет имени Филиппа, mattis.list@uni-marburg.de

Милитарев, Александр Юрьевич — доктор филол. наук, проф. Центра древневосточных исследований ИВКА (Москва), amilitarev@gmail.com

Николаев, Сергей Львович — доктор филол. наук, вед. науч. сотрудник Отдела славянского языкознания Института славяноведения РАН (Москва), sergenicko@mail.ru

Поздняков, Константин Игоревич — доктор филол. наук, профессор Национального института восточных языков и цивилизаций (Париж), pozdnyakov@free.fr

Старостин, Георгий Сергеевич — канд. филол. наук, зав. кафедрой истории и филологии Дальнего Востока ИВКА РГГУ (Москва), зав. Центром востоковедения и сравнительно-исторического языкознания ШГИ РАНХиГС (Москва), gstarst@rinet.ru

Якубович, Илья Сергеевич — канд. филол. наук, научный сотрудник Института мировой культуры МГУ (Москва), sogdiana783@gmail.com

Note for Contributors

Journal of Language Relationship welcomes submissions from everyone specializing in comparative-historical linguistics and related disciplines, in the form of original articles as well as reviews of recent publications. All such submissions should be sent to the managing editor:

G. Starostin
Institute of Oriental Cultures and Antiquity
Russian State University for the Humanities
125267 Moscow, Russia
Miuskaya Square, 6
E-mail: journal@jolr.ru

Articles are published preferably in English or Russian, although publication of texts in other major European languages (French, German, etc.) is possible. Each article should be accompanied with an abstract (not exceeding 300 words) and keywords.

For more detailed guidelines on article submission and editorial policies, please see our Website at: <http://www.jolr.ru> or address the editorial staff directly at journal@nostratic.ru.

Будущим авторам

Журнал *Вопросы языкового родства* принимает заявки на публикацию оригинальных научных статей, а также рецензий, от всех, кто специализируется в области сравнительно-исторического языкознания и смежных дисциплин. Рукописи можно высылать непосредственно заместителю главного редактора по адресу:

125267 Москва
Миусская площадь, д. 6
Российский государственный гуманитарный университет
Институт восточных культур и античности
Г. Старостину
E-mail: journal@jolr.ru

Предпочтительные языки публикации — английский или русский, хотя возможна также публикация статей на других европейских языках (французский, немецкий и т. п.). К каждой статье обязательно прикладывается резюме (не более 300 слов) и список ключевых слов.

Подробнее о требованиях к оформлению рукописи, редакционной политике журнала и т. п. Вы можете узнать на нашем сайте по адресу: <http://www.jolr.ru>, или же непосредственно обратившись к редакции по электронной почте (journal@nostratic.ru).

Note from the Editors

Although the papers presented to our readers' attention in the 11th issue of JLR cover an expectedly wide range of thematic subjects and subdisciplines of historical linguistics, the main bulk of the articles has evolved out of a series of talks, presented by participants of the international conference "Comparative-Historical Linguistics of the 21st Century: Issues and Perspectives", in memory of Sergei A. Starostin (1953–2005), held in Moscow by the Russian State University for the Humanities on March 20–22, 2013. Responding to the editors' proposal, multiple participants were kind enough to agree on reworking or expanding their presentations into full-blown scientific papers: this includes papers by G. Starostin, K. Babaev, V. Blažek, A. Kassian, S. Kullanda, J.-M. List, S. Nikolaev, as well as the joint paper by S. Boroday and I. Yakubovich. Several other contributions have been postponed for technical reasons and will probably be published in subsequent issues of the Journal.

In addition to papers submitted by participants of the Conference, the 11th issue of JLR also includes two other publications: an innovative treatment of Proto-Mari vocalism by A. Aikio and the last part of the etymological analysis of the Swadesh wordlist for Semitic languages by A. Militarev (the previous three parts have all been published in JLR as well, and a final fifth part, containing comprehensive lexicostatistical analysis of the data, is forthcoming). The «Discussion» section, represented by K. Pozdniakov's paper on the degree of importance of automated lexical comparison for establishing long-distance genetic relationship and several answers to his ideas, was also not initiated at the Conference, but is thematically very close to some of the main issues that were discussed in several talks given by participants (additionally, K. Pozdniakov himself, as well as A. Militarev, also gave their own talks at the Conference, though different from their contributions for this issue).

От редакции

Статьи, предлагаемые вниманию наших читателей в 11-м выпуске ВЯР, как обычно, охватывают широкий набор областей исторического языкознания, но при этом данный выпуск можно считать формально тематическим, т. к. большая часть публикуемых работ основана на докладах, представленных авторами — участниками международной конференции «Сравнительно-историческое языкознание в XXI веке: проблемы и перспективы» (20–22 марта 2013 г., Москва, РГГУ), посвященной памяти С. А. Старостина (1953–2005). На предложение редколлегии ВЯР о публикации результатов конференции откликнулось многие докладчики — тексты своих выступлений переработали в научные статьи Г. С. Старостин, К. В. Бабаев, В. Блажек, А. С. Касьян, С. В. Куланда, Й.-М. Лист, С. Л. Николаев, а также, в рамках совместной статьи, С. Ю. Бородай и И. С. Якубович. По техническим причинам публикацию еще нескольких докладов пришлось отложить; с большой вероятностью их удастся выпустить в свет в следующем номере нашего журнала.

Помимо статей, поданных в печать участниками конференции, 11-й номер ВЯР включает еще две работы — опыт новой реконструкции прамарийского вокализма за авторством А. Айкио и последнюю часть этимологического анализа списка Сводеша для семитских языков, проведенного А. Ю. Милитаревым (предыдущие три части были также опубликованы в ВЯР; заключительная часть, посвященная итоговому лексикостатистическому анализу данных, будет опубликована позже). Раздел «Дискуссии» в 11-м номере представлен статьей К. И. Позднякова о степени важности автоматического сравнения лексики для определения дальнего языкового родства и комментариями к ней ряда специалистов; несмотря на то, что формально эта статья не имеет отношения к конференции 2013 г., в тематическом плане поднимаемые в ней вопросы чрезвычайно близки некоторым из озвученных докладов (а сам автор, равно как и А. Ю. Милитарев, также принимал участие в конференции в качестве докладчиков).

Macro-comparative linguistics in the 21st century: state of the art and perspectives

The paper represents an attempt to explicitly summarize most of the major theoretical and methodological problems that, as of today, hinder significant progress in the field of macro-comparative linguistics (research on distant relationships between the various language families of the world). Among these problems are such issues as: the amount, quality, and nature of linguistic data that is necessary to establish long-distance relationship; methodological priorities of the etymologization process; and the complex interdependencies of “objective” (automated) and “subjective” (manual) data comparison. Partial solutions and/or recommendations of a general character are offered for each of the specified issues.

Keywords: Comparative linguistics, long-range comparison, language macrofamilies, computational methods in linguistics.

Any attempt to exhaustively summarize the state of the art of modern macro-comparative linguistic studies within the limits of a short paper would result in inevitable failure. Not even the most informed linguist these days could lay claim to being fully aware of everything that is going on in this field. As a matter of fact, no responsible linguist would probably *want* to waste precious time on such completism, since, in many respects, the field in question is actually a “minefield”, in that the reasonably accurate long-ranger, attempting to navigate it with proper scientific diligence and caution, does not always understand where the principal opposition is coming from. Should the long-ranger, first and foremost, engage in theoretical and methodological arguments with “traditionally-minded” historical linguists? With philologists? With computational scientists? Or should he/she be primarily concerned about rectifying the methods and hypotheses of his/her other fellow long-rangers, whose research has occasionally made macro-comparison so easily vulnerable to criticism from all the abovementioned groups of people?

With this question in mind, I would like to use this short paper as a pretext to briefly discuss where we stand today in the long-range comparison department from a *theoretical* point of view, rather than (as the title could suggest) simply list a bunch of macrofamily hypotheses and rank them in terms of how much credit they have (if any at all) among mainstream specialists¹. (We will refrain from the delicate discussion of what is actually supposed to constitute the “mainstream” paradigm in today’s historical linguistics — for the purposes of the current paper, it will suffice to equate “mainstream specialists” with the average anonymous peer reviewer who is automatically tempted to give a negative assessment to any paper that dares to mention J. Greenberg or V. M. Illich-Svitych in a positive light).

It is more or less obvious that there is little sense to talk about how “reliable” or “convincing” a particularly bold hypothesis of genetic relationship, such as Nostratic or Amerind,

¹ For a summary of these hypotheses from the point of view of the “Moscow school of comparative linguistics” and the Evolution of Human Languages project (<http://ehl.santafe.edu>), see the detailed, if already slightly obsolete, report, published several years ago as Gell-Mann, Peiros, Starostin 2009.

is for a particular researcher, unless *all* researchers, or, at least, a sizeable and representative group of researchers, has reached a detailed agreement on the *criteria* according to which the hypothesis should be judged. Actually, it is not even the word “criteria”, but rather the word “*detailed*” that should be stressed — time has shown, over and over again, that several people can easily agree on a general principle (such as “regular phonetic correspondences”, or “semantic proximity”, or even “lexicostatistical analysis”), but when it comes to practical research, it often turns out that everyone has one’s own understanding of what these words are supposed to mean in particular situations.

The result is a huge conflict of interest even between the long-rangers themselves. Whenever a macro-hypothesis is being worked on by several people that do not form a coherent team, what one gets is a sub-set of highly dissimilar and often mutually incompatible sub-hypotheses: not “Nostratic” as such, but “Illich-Svitych’s Nostratic”, “Dolgopolsky’s Nostratic”, “Bomhard’s Nostratic”²; not “Nilo-Saharan”, but “Greenberg’s Nilo-Saharan”, “Bender’s Nilo-Saharan”, “Ehret’s Nilo-Saharan”³, etc. This confusion turns both the hypotheses and their authors into easy prey for critics of long-range comparison, who are under no obligation to conduct a proper comparison of all the competing variants — even if they may sometimes be mildly sympathetic to the more accurately stated ones, they still run the risk of being discarded along with the poor ones (for extra security).

But there is an additional problem here, the gravity of which, I believe, is not always fully understood either by the “lumpers” or the “splitters”. Today, it is no longer possible for mainstream linguistics to simply *ignore* the macro-comparative department or to *dismiss* it right out of hand, either explicitly or implicitly⁴. The main reason for this is increasing pressure from the adjoining disciplines — above all, archaeology and human genetics.

Modern genetics, in particular, since it studies human prehistory from a natural science angle, is employing sets of powerful computational tools to construct credible genetic trees of humanity and trace its major migration routes — while it isn’t quite there yet, it is highly likely that only a matter of years, perhaps a few decades, separates us from having at our disposal a mathematically approved “optimal” scenario of the chronological dispersal of *Homo sapiens*. And, naturally, geneticists are all too keen to compare the results that they are getting — for periods that go back dozens of thousands of years — with whatever the linguists have to offer; and, on an informal basis at least, they are frequently quite surprised to discover that the linguists do not have to offer *that* much, even though they have been going at it for a far longer period of time.

In all seriousness, it no longer becomes permissible for the comparative linguist to remain content with hiding behind an agnostic formula, such as “the comparative method cannot go further back than six...” (eight, ten, the precise figure does not matter) “...thousand years”⁵. It is becoming a matter of “do or die”: the comparative method *has* to be able to go further than

² The three competing variants of the Nostratic theory, supported with intersecting, but still widely varying evidence, are best illustrated in their authors’ etymological dictionaries: Illich-Svitych 1971–1984, Dolgopolsky 2008, Bomhard 2003.

³ See Greenberg 1966, Bender 1997, Ehret 2001 for three entirely different methodologies and vastly differing corpora of evidence to back up the hypothesis.

⁴ “Implicit” rejection may, for instance, be found in a recent monograph on language classification (Campbell & Poser 2008), where the authors express mild optimism about the future of demonstrating “remote linguistic relationships”, but it soon turns out that their ideal of a remote linguistic relationship is represented by such families as Uralic or Uto-Aztecan, which no authentic long-ranger could ever define as “remote”.

⁵ Cf.: “the diagnostic kinds of accident... dissipate entirely after about 8000 years (distinctly less in languages of certain non-Indo-European structural types)” (Nichols 1992: 313).

that, and even if someone manages to prove beyond reasonable doubt that it cannot (which, in my opinion, is not very likely), in reality this will only mean that the *current version* of the comparative method is not robust enough to go there — so it is up to us to increase its robustness and succeed where previous efforts have failed.

Giving up on the comparative method when “larger-than-*x*” time periods are concerned means only one thing: something else will inevitably emerge to take its place in linguistic palaeontology, and that “something else” will not necessarily be an improvement. In fact, something else has *already* emerged: the application of computational phylogenetic models, borrowed from biology and other branches of natural science, to linguistic data — a development that has enjoyed ever increasing popularity throughout the last decade ever since Russell Gray’s and Quentin Atkinson’s (2003) landmark publication of their model of Indo-European classification in “Nature”. The so-called “Gray Lab” in Oakland, New Zealand still remains arguably the largest single supplier of such studies, but multiple other teams and individuals have also joined the fray, and while, on the whole, these developments are quite exciting, one generally troubling thing about them is that most of the involved scientists are anything *but* professionally trained historical linguists. Paradoxical as it may sound, the most widely discussed publications in historical linguistics from the last decade have not been published by actual historical linguists! And this fact alone should be enough to make most historical linguists — including “long-rangers” / “lumpers” and “short-rangers” / “splitters” alike — sit up and take notice; especially because, in this particular case, “the most widely discussed” may not necessarily mean “the most useful in advancing real science”.

There is, however, one aspect in all of these recent computational studies that “traditional” historical linguists should probably pay close attention to — all of them approach linguistic reconstruction and classification as *stochastic* in nature, usually generating results by exploiting various models that are based on the principles of maximum likelihood. And while some might object that the stochastic principle in comparative linguistics may serve as a clever trick to make the work formally invulnerable to criticism (“why does this model contradict established historical facts?” — “well, the model does not state that the results are *right*, it just states that this is the most probable outcome given the algorithm and the data fed into the algorithm”), probabilistic reasoning is actually a deeply right methodological foundation that is all too often overlooked in traditional comparative linguistics — namely, that the process of linguistic reconstruction is almost always a process of *choice* of an optimal solution given the circumstances, rather than a process of strict mathematical or logical *proving* of any particular statement. Despite the frequent use of the very loosely understood term “proof” in comparative linguistics, one does not usually resort to rigorously proving anything in this discipline; instead, one considers alternatives, for each of which evidence is gathered, weighed, and the weightier bunch is given precedence. It is perfectly possible that tomorrow new evidence will appear, forcing us to switch from solution A to solution B, making the latter “weightier” than it was yesterday — but this, of course, is just how science is actually expected to work.

So where does today’s macro-comparative linguistics stay on this issue? Let us roughly assume the existence of two principal schools of long-range comparison. One of them insists on the helplessness of the comparative method when it comes to sorting out superfamilies — this is the “mass comparison” approach of Joseph Greenberg and his followers. The other one is the tradition initiated by Russian linguists Vladislav Illich-Svitych and Aharon Dolgopolsky with their “Nostratic” hypothesis and subsequently inherited by the so-called “Moscow school” of comparative linguistics: its main idea is that there are no fundamental differences between long-range comparison and regular short-range comparison of the Neogrammarian type. One look at the voluminous Nostratic dictionary by A. Dolgopolsky (2008) clearly suf-

fices to reveal the stylistic legacy of J. Pokorny's Indo-European dictionary (somewhat obscured as it is by a swarm of additional transcriptional symbols and complex abbreviations). In fact, both schools agree in their assumption that the same methodology, be it "mass comparison" or the classic comparative method, may and should be applied to language families of just about any time depth.

The relations of both approaches to the issue of "probability" and "optimality" is complex. Greenberg, harassed by demands of "rigorous proof" by traditionally-oriented historical linguists throughout his life, recognized and advocated the importance of probabilistic thinking in almost every theoretical publication where he defended mass comparison⁶. Unfortunately, it might be argued that he hardly ever practiced what he preached: in his huge corpora of data, Greenberg rarely discusses alternatives, never really drawing the reader's attention to what it is exactly that makes *his* hypotheses better than anybody else's, not to mention the "null hypothesis" of non-relationship. Hence the superficially convincing "discreditation" of the Amerind hypothesis by Lyle Campbell (1988), who successfully "demonstrated" that Finnish was Amerind, too; all the more easily done since there was no strict procedure to quantify the evidence in Greenberg's paradigm — no way to state explicitly how the evidence for "Finnish as Amerind" relates in quantitative or probabilistic terms to the evidence for "Amerind as Amerind".

On the other hand, the tradition of regarding macro-comparative studies as simply an extension of regular comparative studies, upheld most rigorously in the Moscow school, still shares the same methodological problem as its opponents: some scholars still seem too beset with the idea of "categoricity", the issue of "proof" of relationship understood in strictly binary terms (yes or no) with a dogmatic flair. Pro- and opponents of long-range comparison alike demand a threshold for the evidence, one which they rarely, if ever, try to establish formally and, therefore, one which they all understand differently. For instance, the demands that Lyle Campbell establishes for "proving" a long-range comparison are so strict that, most likely, not a single genuine long-range hypothesis will ever be proven to his satisfaction (Starostin 2009). At the same time, one could also list certain long-rangers, even those who formally accept the importance of the comparative method and regular phonetic correspondences, whose own threshold is intuitively motivated by a small handful of intuitively stunning resemblances between families, around which they build up pseudo-theories that are completely devoid of historical realism.

All of my own experience of working on long-range comparison, mostly in the area of large Eurasian stocks ("Nostratic", "Sino-Caucasian") and African classification, seems to indicate the necessity of admitting it — there *is*, or at least, there *should* be a certain methodological difference between the usual ways in which comparative linguistics has up to now worked with "short-range" families, and the ways in which we should be conducting further research on the macro-comparative level. Furthermore, I believe that the fact that macro-comparative linguistics has somewhat stalled these days, leading to a temporary crisis of such studies, has much to do with our reluctance to accept this.

Hence, if we want real progress in the development of macro-comparative studies, it makes sense to suggest that a certain compromise is necessary between all these positions, one that would take the most promising aspects of "macro-comparative Neogrammarianism" *à la* Illich-Svitych / Dolgopolsky, of Greenberg's "mass comparison", of modern computational

⁶ E. g.: "...in all empirical sciences... all that we *can* get are results so close to certainty that for all practical purposes we can consider them true, that is, a hypothesis which is overwhelmingly better than any other in accounting for the facts" (Greenberg 1995: 207).

approaches, assess them in the light of reasonable, unbiased “traditionalist” criticism, and synthesize a relatively complex, but “accessible” and efficient set of methodological rules. Such a thing is probably easier said than done, as it clearly involves a lot of heavy, often tedious teamwork, where specialists working on different language families should regularly share their experience and help refine each other’s work ideology; but the more tedious that teamwork is, the less space might be left for creative bickering between people who should be advancing science rather than wasting time on highly unproductive debate (as is, unfortunately, quite often the case in historical linguistics).

In an attempt to help facilitate such a compromise, I will try to concisely outline four of the most common points of debate that are regularly observed in works on long-range comparison (actually, all four are relevant for historical linguistics as a whole, but the relevance naturally increases as we proceed further on to deeper time levels).

Problem 1: Quantity or quality?

This dilemma, as a rule, tends to be resolved in extremities. Active and productive proponents of macro-comparative studies tend to stress the importance of compiling huge etymological dictionaries (cf. approximately 2800 etymologies in the comparative Altaic dictionary of Starostin, Dybo, and Mudrak (2003); amazingly, almost exactly the same number — 2800! — in Dolgopolsky’s Nostratic dictionary (2008); more than 3000 etymologies in the Afro-Asiatic database by A. Militarev and O. Stolbova (available online at <http://starling.rinet.ru>); 1606 etymologies in Ch. Ehret’s Nilo-Saharan dictionary (2001), etc.). A common criticism of such “etymological mastodonts” is that these huge numbers are only possible since there are so many languages to choose from — implying that most, if not all, of the individual etymologies simply reflect chance resemblances that accumulate in daughter languages as time passes.

The opposite approach is illustrated by such relatively recent attempts at macro-comparison as Laurent Sagart’s “Sino-Austronesian” (2005) or Edward Vajda’s “Dene-Yeniseian” (2011). These works tend to focus on a relatively small number of examples (around 100 etymologies); to introduce a system of phonetic correspondences that is claimed to be fully regular (something that Eric Hamp (2011) calls “total accountability”); and to exclude any semantic comparisons that could seem improbable or dubious. Such comparisons have a lot of potential charm — above all else, they are particularly seductive for the eye of the outside evaluator, if only because it is a much easier task to browse through 100 reconstructions than 2800, and also because of their superficial “high accuracy” as compared to the superficial “sloppiness” of the large macro-etymological dictionaries.

There are, however, two serious problems with this approach. First, claims of “perfect” regularity of correspondences for such a small corpus are usually exaggerated. The very fact that there is only a small handful of data means that most of the correspondences recur only in a few examples, and there is no easy way to statistically verify the significance of these recurrences (as an example, see Starostin 2012 for an analysis of such a situation in the case of Vajda’s “Dene-Yeniseian”). Second, if we are speaking of macrofamilies (such as Austronesian with its thousand languages, or Sino-Tibetan with its twenty or more nearly equidistant branches), our main problem is the same as in the case of large etymological dictionaries: a near-limitless choice of comparanda, out of which it is not all that difficult to fashion a “mini-package” of seemingly corresponding forms.

Proposed solution: *Size does not matter.* A long-range genetic relationship hypothesis may operate with as many comparisons as it needs, no more and no less — but only provided there

is some sort of objective methodology that helps *range* these comparisons and measure their degree of correspondence to historical and typological expectations. “Hugeness” or “compactness” of the etymological corpus should not be regarded as self-sufficient values — hence, one should refrain from being driven by the desire to either “create as many etymologies as possible” or “purge as many dubious etymologies as possible”.

What really *does* matter is our ability to arrange the evidence in a sort of pyramid, where the strongest comparisons (phonetically, semantically, distributionally) should be clearly positioned at the top, and then propped up by as much supportive evidence as is necessary at the bottom. Such an arrangement, among other things, would facilitate external evaluation of hypotheses — one starts by scrutinizing the top part, then selectively examines the rest of the evidence whenever one feels that the “top” evidence is in need of additional supportive data. This is a principle that, surprisingly enough, has not been so far implemented in any of the large etymological corpora that I am familiar with; the closest analogy would probably be M. L. Bender’s treatment of comparative Nilo-Saharan evidence (e. g. his system of “excellent”, “good”, and “fair” isoglosses introduced in Bender 1997, unfortunately, on rather flimsy and poorly stated criteria), but this is more an analogy of style than of substance.

Problem 2: Grammar or lexicon?

Having ruled out extra-linguistic and typological data as evidence for genetic relationship, J. Greenberg (1966) correctly emphasized the importance of grammatical and lexical isomorphisms between compared languages as the main (in fact, only) valid evidence in this particular sphere. On relatively shallow levels of comparison, as is well known to all historical linguists, grammar and lexicon usually go hand in hand inasmuch as the required isomorphisms are concerned. On deep chronological levels, however, this link is just as frequently broken, giving rise to the problem of what is *more* important for demonstrating genetic relationship — grammar (more precisely, paradigmatic morphology) or lexicon (more precisely, *basic* lexicon in a Swadesh-like understanding)? From both a theoretical and a pragmatic standpoint, the seductive power of morphology often wins out, because:

(a) it is assumed, based on selective evidence — never really proven — that morphology is much more resistant to borrowing than basic lexicon, and therefore constitutes a more reliable source for comparison at large time depths (see, e. g., Vovin 2002);

(b) it is relatively easy to prepare and present comparative morphological evidence, just because the amount of such evidence is, by the very definition of morphology, smaller and easier to assimilate than the corresponding amount of lexical evidence;

(c) an additional important point is that, as a rule, comparative morphological evidence is of much greater interest to specialists in typology and various synchronic subdisciplines of linguistics than comparative lexicology — this is yet another stimulus to produce works on comparative morphology, because they do not seem quite as “dull” for the general audience.

At the same time, however:

— argument (a), even if it *is* true, disregards the fact that, while morphology may be more resistant to externally driven change than basic lexicon, it is almost definitely less resistant to internally driven change, where, in addition to general processes of morpheme replacement, we also have to remember the factor of “phonetic erosion” (loss of morphological elements on word borders due to reduction, sandhi, etc.), far more detrimental towards morphology than towards lexicon. For example, as a rule, modern Indo-European languages from different branches tend to preserve far more common morphemes in their basic lexicon than in their

morphological systems (no less than 25–30 items on the Swadesh list are still the same, whereas finding 25–30 common Proto-Indo-European grammatical morphemes between, say, a present day Germanic and a present day Iranian language would be quite a challenge);

– argument (b), if taken at face value, may easily lead the researcher into the temptation of thinking along the lines of “I really believe in this genetic relationship, but it is also necessary to convince the public at large, so the best thing to do is to produce some impressively looking morphological evidence, whatever the cost”. Analysis of both Laurent Sagart’s “carefully assembled” morphological evidence for Sino-Austronesian and Edward Vajda’s morphological evidence for Dene-Yeniseian⁷ uncovers quite serious subjective distortions of the semantic and distributional properties of the analyzed morphemes, which could have easily been avoided if the researchers in question did not strive so hard to recover the morphological systems (rather than isolated morphemes) of the hypothetical protolanguages.

Proposed solution: *Morphology need not matter*. (Note that this is quite different from “*morphology does not matter*”, which would be completely wrong). The presence of strong, cohesive isoglosses that suggest a common paradigmatic morphology for the ancestor of the compared languages is a very strong argument for relationship (on “macro-comparative” levels, chunks of paradigmaticity may perhaps be glimpsed in Niger-Congo and Afro-Asiatic languages, but even in those macro-families the importance of morphology tends to be exaggerated, mainly through “linguistic hearsay”); however, the *lack* of such evidence essentially means nothing (cf. the case of Russian and English, where, other than a tiny bunch of archaic participial suffixes, transparent morphological isoglosses no longer exist).

The reverse, however, is not true: *basic lexicon always matters* — I am willing to risk stating that there is not even a *single* case of commonly accepted genetic relationship, no matter how close or distant, where it would not be possible to demonstrate it on the basis of basic lexicon alone; and in quite a few cases (e. g. the isolating languages of Southeast Asia), it is *only* possible to demonstrate it on the basis of basic lexicon. Consequently, in generating and testing hypotheses of long-range relationship, it is recommendable to always begin with the basic lexicon, which is precisely the ideology behind “The Global Lexicostatistical Database”, the Moscow School’s latest collective project on language classification. Any long-range comparison that emphasizes morphological reconstruction while downplaying basic lexicon will always look suspicious, since any such scenario would simply violate observed diachronic typology.

Problem 3: Internal solutions or external evidence?

This issue is a rich source for endless debates that usually take place between “narrow” specialists in particular language groups or families and “broader” specialists in macro-comparative studies. The model of the debate is always the same: a Vasconist / Japanologist / Sinologist / Indo-Europeanist / etc. criticizes select long-range etymologies (“Dene-Caucasian”, “Altaic”, “Nostratic”, etc.), concluding that the Basque / Japanese / Chinese / Indo-European / etc. part of the etymology is explainable as secondary on internal grounds, implicitly assuming or explicitly stating that internal etymologization should always take precedence over far-flung attempts at external explanation. Examples are too numerous to list, but see Dybo & Starostin 2007 for a detailed treatment of several such disputes on Altaic etymologies.

⁷ For a serious critical analysis of Sagart’s “Sino-Austronesian” morphology, see Blust 1995; my own critical remarks on Vajda’s “Dene-Yeniseian” morphology have been published as Starostin 2012.

Of all the listed problems, this one is probably the hardest to deal with, because in each individual case there is usually some risk of taking it to almost personal levels — from the point of view of the macro-comparative linguist, the narrow specialist is seen as a retrograde conservative, defending his/her subjective views as a dogma simply because they are supported by “expert authority”; from the point of view of the narrow specialist, it is the macro-comparativist that comes across as an ignorant amateur striving for sensationalism, not interested in real hard work or scientific honesty. (See Manaster-Ramer 1993 for some penetrating insights into the nature of such debates between Nostraticists and their critics).

Proposed solution: This “conflict of interests” may be overcome only when the issue is approached from a thoroughly unbiased position. In particular, flashy, easily memorizable, but overstated phrases like “internal etymologization should always take precedence over external etymologization” should probably be considered just as harmful to doing good science as their opposites.

What is *really* necessary in such cases is an elaborate standard against which the alternatives could be weighed objectively — for instance, a general database of the various types of language change, against the data of which (including statistical data) it would be possible to test the conflicting solutions. One of the most important sections in such a database should be dedicated to semantic change, of which we still know relatively little (e. g. not nearly enough to answer the question, “how likely is it that the noun ‘stone’ would be derived from the verb ‘to roll’, anywhere in the world or in one particular linguogeographical area?”). A database of phonetic change might also be important, but only if it is sufficiently detailed — it is very easy to make an inventory for trivial and widespread types of phonetic change, but the toughest problems of etymologization generally tend to concern non-trivial paths of change for typologically rare segmental and suprasegmental units.

In other words, it is necessary, first and foremost, to recognize that most of the debates over internal vs. external etymologization do not so much reveal the personal flaws and biases of the participants (although these things happen, too) as they highlight the weak spots of general comparative methodology; inasmuch as they stimulate us to think of the possible ways to improve the method, they are quite useful, but the important thing is to not let oneself get carried away by ideological, let alone personal, motives.

Problem 4: Formal objectivity or subjective judgement?

This old problem of etymology (the value and importance of subjective judgement and intuition in linguistic reconstruction) has had a major revival in recent times; today, it does not so much pit “lumpers” against “splitters” as it tends to oppose “traditional comparative linguists” and a new generation of scientists — many of them with backgrounds in “hard science” (biology, physics, general computational studies, theory of information etc.) or anthropology and sociology rather than linguistics. What puts them all together is the employment of formal probabilistic methods, usually based on Bayesian principles, to hunt for automatically generated optimal scenarios of language classification (Gray & Atkinson 2003), models of language evolution (Pagel et al. 2007), and, most recently, even protolanguage reconstruction (Bouchard-Côté et al. 2013). Although, in a way, these studies may be said to continue the computational tradition that had already been introduced to historical linguistics by glottochronologists (Swadesh, Lees, etc.) in the 1950s, this is now done on a much more complex scale, with the added potential of high performance computers assisting researchers in selecting the most parsimonious historical scenario.

Traditionally oriented historical linguists seem rather slow to embrace these new methods — not only because many of them have trouble assimilating the mathematical apparatus, but also because even some of those who do not have any such trouble still feel skeptical about whether machine-based methods are powerful enough to achieve results which the human mind cannot achieve on its own (after all, linguistic classification and reconstruction could hardly be reduced to a small set of computational issues). The immediate advantages of formal mathematical methods may seem obvious to the naked eye, but in reality these advantages turn out to be overrated, for the following reasons:

(a) formal methods may give the impression of filtering out the subjective factor — one assumes that they generally produce a mix of correct and erroneous results, and accordingly tries to precisely define the margins of error, something that is rarely, if ever, done in manual etymological research. However, even in the strictest procedures of this kind subjectivity is never really ruled out completely. For instance, any automatic analysis of lists of words/morphemes automatically depends on how accurately these lists have been collected and compiled. Any automatic analysis that tries to build a genealogical tree or a network based on pre-established etymological cognations depends on the degree of subjectivity already present in these arguments. And, finally, any automatic analysis that tackles the data head on, intentionally ignoring all the work that had previously been carried out in the “dark ages” of pre-computational linguistic science, is useless if it does not take into consideration the phonetic, morphological, and lexical specifics of the analyzed languages — something that is more easily said than done;

(b) despite the steady flow of works describing automated procedures, most of them cover relatively “safe” territory — as a rule, the data come from such families as Indo-European or Austronesian, more rarely, Bantu, Semitic, or Turkic. What all or most of these taxa share in common is the following: (a) they have already been well studied by comparative linguists over a research period of 100–200 years; (b) most of them, with the notable exclusion of many modern Indo-European languages, have a relatively simple story of phonetic change, making them ideal “polygons” for testing out simplistic algorithms.

Consequently, we have yet to see actual situations where formalized automatic methods help achieve a real breakthrough in some issue that has proven too tough for “ordinary” comparative linguistics — including, of course, macro-comparative hypotheses as well. Not only that, but even in the respective areas of the listed families application of computational methods has so far been unable to successfully resolve any of the remaining complicated issues: for instance, the most recent publication on the homeland of Proto-Indo-European (Bouckaert et al. 2012) has largely failed to convince Indo-Europeanists of the adequacy of its methodology or the correctness of its results, and the recent publication on the automatic reconstruction of Proto-Austronesian (Bouchard-Côté et al. 2013), has not served to elucidate any of the remaining ambiguities about the phonological system of the protolanguage. Keeping in mind the rather humble efficiency of all these efforts (many of which, I might add, have received the kind of publicity that is quite disproportionate relative to the achieved results), one would, indeed, be seriously tempted to doubt that such methods could be successfully applied to far more complex problems, such as the demonstration and clarification of genetic links between, e. g., North Caucasian languages, let alone Nostratic, Sino-Caucasian, or Amerind.

Proposed solution: Just like in every other situation described above, a compromise is necessary here. Without denying the usefulness and added potential of computational algorithms borrowed from other branches of natural and social sciences and adapted to the needs of historical linguistics, I would suggest that a truly reasonable approach towards any hy-

pothesis of linguistic “macro-relationship” should always strive to combine automated procedures with manually performed research of the traditional type.

Thus, for instance, the ideology that is currently employed in the “Global Lexicostatistical Database” project, which aims at an improved classification of the world’s languages on different chronological levels based on their basic lexicon, recommends taking the following steps:

(a) accurate manual compilation of 100-wordlists for each language group, guided by a formal standard defined in Starostin 2010 and Kassian et al. 2010; (b) generation of two types of cognation indexes — manual (based on phonetic correspondences where they are available and on phonetic similarity where they are not) and automatic, based on the method of consonant class comparison, with different lexicostatistical trees built for both; (c) comparison of results, with well-argued conclusions on which classification should be regarded as being closer to the truth, and why; (d) reconstruction of the proto-wordlist for the language group; (e) repetition of procedures (a)-(d) for other language groups of the region; (f) comparison of the reconstructed proto-wordlists, and so on (the whole procedure is explained in much detail in Starostin 2013).

The most important role of automatic analysis in this procedure is that in those cases where one ends up with multiple discrepancies between the automatically generated and the manually generated trees, such discrepancies serve as a strong stimulus to look at the data with increased accuracy, so as to understand their nature. In particular, manual analysis may help correct the errors that come from applying the “universal” algorithm of comparison to typologically peculiar situations that were not taken into account at the stage of parameter selection and adjustment. At the same time, automatic analysis may point out some of the subjective weaknesses of “human” etymologization, e. g. reveal cases of fake phonetic irregularity, unjustly postulated by the researcher on the basis of one or two tempting similarities. That way, both approaches complement each other, and regularly switching back and forth between them will unquestionably yield more insights than relying one hundred percent upon only one.

In conclusion, here is a brief list of tasks that I would define as “primary” for the researcher who truly wants to achieve notable progress in the field of linguistic “macro-comparison”:

(1) *Work first and foremost with evidence that may be quantified and statistically assessed*: above all, this means Swadesh-type lexicostatistical lists, but helpful statistical evaluation may also be performed on etymological databases, provided they are complete and organized accurately enough to be easily subjected to various types of automatic analysis.

(2) *Build up reference corpora of typological evidence*: as we move back into the past and multiply our choice options, an important compensatory mechanism for narrowing them back down is correlation with historical-typological databases of the various types of phonetic and semantic change, as well as typological databases that collect and systematize various types of linguistic interaction, primarily loanwords. The construction of a single, unified database of this sort is a major task to be achieved in the future, but even small, “local” databases run by particular teams or individuals may be of use.

(3) *Try to develop and apply universal standards and reference frames*, be it for lexicostatistics (with rigidly defined standard wordlists) or etymology (unified criteria for postulating cognacy). This principle need not be understood simplistically (e. g. as “what is good for Indo-European is also good for Nostratic”), since similar standards should only be applied to similar type objects (Nostratic is, at the very least, much chronologically deeper than Indo-European, so it might require a slightly different approach). But, on the other hand, if every hypothesis of macro-relationship is judged according to its own standard, this renders useless

the best criterion for falsifying such hypotheses — testing alternatives according to the same set of rules. As far as my opinion is concerned, no better universal standard than the lexicostatistical method has been offered so far; however, there is no reason why the lexicostatistical test, relying on a small subset of data, could not be accurately improved in the future by gradually expanding that data.

It goes without saying that not all of the listed recommendations have been properly implemented even for “short-range” families, including such well-studied ones as Indo-European. However, the need to implement them for “long-range” and overall problematic families is much higher — as systems of regular phonetic correspondences between hypothetically related (proto-)languages become harder to establish due to decrease of data and increase of the chronological gaps between different reconstructed states, one must somehow learn to compensate for this added trouble.

At the present time, not a single macro-comparative hypothesis that I know of fully satisfies all of the listed requirements, but some are definitely in better shape than others: e. g. Nostratic fares better in terms of statistical tests and general “standardization” than Austric, while Austric, in its turn, is better than Nilo-Saharan or Amerind, etc. etc. One thing can be said for sure: regardless of whether macro-comparative studies are ever capable of becoming the dominant paradigm in historical linguistics, their possibilities are far from being exhausted, and as long as there are still scholars around who are willing to engage in this exciting field and strive towards “raising the bar”, there is really no telling what unpredictable surprises the future may bring.

Literature

- Bender 1997 — Lionel M. BENDER. *The Nilo-Saharan Languages: A Comparative Essay*. München — Newcastle: LINCOM Europa.
- Blust 1995 — Robert BLUST. An austronesianist looks at Sino-Austronesian // *The Ancestry of the Chinese Language*. Ed. by William S-Y. Wang. Journal of Chinese Linguistics: Monograph Series 8, pp. 283–298.
- Bomhard 2003 — Allan R. BOMHARD. *Reconstructing Proto-Nostratic: Comparative Phonology, Morphology, and Vocabulary*. Charleston: Signum Desktop Publishing.
- Bouchard-Côté et al. 2013 — Alexandre BOUCHARD-CÔTÉ, David HALL, Thomas L. GRIFFITHS, Dan KLEIN. Automated reconstruction of ancient languages using probabilistic models of sound change // *Proceedings of the National Academy of Sciences*, 10.1073.
- Bouckaert et al. 2012 — R. BOUCKAERT, P. LEMEY, M. DUNN, S. J. GREENHILL, A. V. ALEKSEYENKO, A. J. DRUMMOND, R. D. GRAY, M. A. SUCHARD, Q. D. ATKINSON. Mapping the Origins and Expansion of the Indo-European Language Family // *Science*, 337 (6097), pp. 957–960.
- Campbell 1988 — Lyle CAMPBELL. Review article on *Language in the Americas* by Joseph H. Greenberg // *Language*, v. 64, pp. 591–615.
- Campbell & Poser 2008 — Lyle CAMPBELL, William J. POSER. *Language Classification: History and Method*. Cambridge University Press.
- Dolgopolsky 2008 — Aharon DOLGOPOLSKY. *Nostratic Dictionary*. McDonald Institute for Archaeological Research, Oxford Publishing Press.
- Dybo & Starostin 2007 — A. V. DYBO, G. S. STAROSTIN. In Defense of the Comparative Method, or the End of the Vovin Controversy // *Аспекты компаративистики III*. Moscow: RSUH Publishers, pp. 119–258.
- Ehret 2001 — Christopher EHRET. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln: Rüdiger Köppe Verlag.
- Gell-Mann, Peiros, Starostin 2009 — M. GELL-MANN, I. PEIROS, G. STAROSTIN. Distant Language Relationship: The Current Perspective // *Journal of Language Relationship*, 1, pp. 13–30.
- Gray & Atkinson 2003 — R. D. GRAY, Q. D. ATKINSON. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426, pp. 435–439.
- Greenberg 1966 — Joseph H. GREENBERG. *The Languages of Africa*. Bloomington, Indiana University; Mouton & Co., The Hague.

- Greenberg 1995 — Joseph H. GREENBERG. The Concept of Proof in Genetic Linguistics // *Mother Tongue I*, pp. 207–216.
- Hamp 2011 — Eric HAMP. On The First Substantial Trans-Bering Language Comparison // *The Dene-Yeniseian connection*. Ed. by Jim Kari & Ben Potter, Fairbanks, ANLC, pp. 285–298.
- Haspelmath & Tadmor 2009 — Martin HASPELMATH, Uri TADMOR. The Loanword Typology Project and the World Loanword Database // *Loanwords in the World's Languages: A Comparative Handbook*. Ed. by M. Haspelmath & U. Tadmor. The Hague: Mouton de Gruyter, pp. 1–34.
- Illich-Svitych 1971–1984 — V. M. ILLICH-SVITYCH. Opyt sravnenija nostraticheskix jazykov. Sravnitel'nyj slovar'. [An attempt at a Comparative Dictionary of the Nostratic languages]. Moscow: Nauka. Vol. 1: 1971; vol. 2: 1976; vol. 3: 1984.
- Kassian et al. 2010 — A. KASSIAN, G. STAROSTIN, A. DYBO, V. CHERNOV. The Swadesh wordlist: an attempt at semantic specification // *Journal of Language Relationship*, 4, pp. 46–89.
- Manaster-Ramer 1993 — Alexis MANASTER-RAMER. On Illich-Svitych's Nostratic theory // *Studies in Language*, 17, pp. 205–249.
- Nichols 1992 — Johanna NICHOLS. Linguistic Diversity in Space and Time. University of Chicago Press.
- Pagel et al. 2007 — M. PAGEL, Q. D. ATKINSON, A. MEADE. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449 (7163), pp. 717–720.
- Sagart 2005 — Laurent SAGART. Sino-Tibetan–Austronesian: an updated and improved argument // *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. Ed. by Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas. London: Routledge Curzon, pp. 161–176.
- Starostin 2009 — George STAROSTIN. Review of Lyle Campbell & William J. Poser. Language Classification: History and Method, 2008 // *Journal of Language Relationship*, № 2, 2009, pp. 158–174.
- Starostin 2010 — George STAROSTIN. Preliminary lexicostatistics as a basis for language classification: a new approach // *Journal of Language Relationship*, № 3, pp. 79–117.
- Starostin 2012 — George STAROSTIN. Dene-Yeniseian: A Critical Assessment // *Journal of Language Relationship*, № 8, pp. 117–152.
- Starostin 2013 — G. S. STAROSTIN. *Jazyki Afriki: opyt postrojenija leksikostatisticheskoy klassifikacii. Tom I: Metodologija. Kojsanskije jazyki* [Languages of Africa: an attempt at a lexicostatistical classification. Vol. I: Methodology. Khoisan languages]. Moscow: Languages of Slavic Culture, 2013, 510 pp.
- Starostin, Dybo, Mudrak 2003 — Sergei STAROSTIN, Anna DYBO, & Oleg MUDRAK. *An Etymological Dictionary of Altaic Languages*. Leiden: Brill.
- Vajda 2011 — Edward VAJDA. A Siberian link with Na-Dene languages // *The Dene-Yeniseian connection*. Ed. by Jim Kari & Ben Potter, Fairbanks, ANLC, pp. 33–99.
- Vovin 2002 — Alexander VOVIN. Building a 'bum-pa for Sino-Caucasian: a reply to Sergei Starostin's reply // *Journal of Chinese Linguistics*, v. 30, N 1, pp. 154–171.

Г. С. СТАРОСТИН. Макрокомпаративистика в XXI веке: текущее состояние и перспективы развития.

Статья представляет собой попытку эксплицитно суммировать бóльшую часть теоретических и методологических проблем, на сегодняшний день препятствующих существенному прогрессу в области макрокомпаративистики, изучающей вопросы глубокого родства языковых семей мира. К числу таких проблем относятся, в частности: (а) вопрос о количестве, качестве и природе языковых данных, требуемых для установления глубокого родства; (б) вопрос о приоритетах этимологизации данных при разработке макрокомпаративистских гипотез; (в) вопрос о сложных взаимозависимостях между «объективным» (автоматическим) и «субъективным» (ручным) сравнением. Для каждой из перечисленных проблем в общем виде предлагаются частичные решения или методологические рекомендации.

Ключевые слова: Сравнительное языкознание, дальнее родство языков, языковые макро-семьи, вычислительные методы в языкознании.

Verbal morphosyntax in Southwest Mande: elements of reconstruction

The article is written after the recent conference report on some of the preliminary results of the collective research project “Towards Proto-Niger-Congo: Comparison and Reconstruction”. It presents some comparative data on Southwest Mande verb morphosyntax and thoughts on the possibility of reconstructing the system of verbal markers in the proto-language.

Keywords: Mande languages, reconstruction, verbal system, morphology, syntax, African languages.

In September 2012, the author of this paper was glad to take part in The International Congress “Towards Proto-Niger-Congo: Comparison and Reconstruction”, organized in Paris by a group of scholars who are still keen on reconstructing distant proto-languages despite the fact that the world of mainstream linguistics seems to be gradually losing interest in language kinship and shifting its focus to other sub-disciplines. The participants presented some of their recent achievements in the comparative analysis of language groups and families belonging to the largest genetic grouping ever proposed: the Niger-Congo macrofamily, consisting of some 1 500 tongues spread all over Sub-Saharan Africa, from the Senegal to the Orange River. They also agreed to prolong their cooperation so as to undertake a thorough and contemporary reconstruction of what could be reasonably defined as Proto-Niger-Congo, a language that was proposed by Joseph Greenberg some fifty years ago, yet has never been substantially approved through strict comparative analysis.

One of the teams that took shape during the Congress became specifically targeted on the reconstruction of Proto-Mande, the ancestral language of a relatively large West African family whose members have been relatively well studied. The team, led by the prominent Mandeist Prof. Valentin Vydrin, are to present a valid outlook of the family protolanguage, based on the analysis of phonology, morphosyntax and lexicon of the modern Mande languages. The current paper should be seen as a small step towards achieving this goal.

The Southwestern group of Mande languages (SW Mande from here on), consisting of six units, is spread across three republics of West Africa. Kpelle and Looma are both spoken in Liberia and Guinea, Bandi is found only in Liberia, and Zialo is across the border in Guinea, while Mende and Loko are closely related languages of Sierra Leone. Descriptions of the majority of these languages have existed already since the early 20th century, and their internal classification, according to which Kpelle is considered to have been the first offshoot from the proto-language, is, to the best of my knowledge, accepted by all the relevant scholars. Within Mande, the closest relatives for SW Mande languages are Vai-Kono, Soso-Jalonke and other languages that belong to the Western branch of the family.

Comparative research on SW Mande languages is represented by just a few papers, which include Dwyer 1973 with a complex tonology analysis, Выдрин 2006 that deals mostly with

Table 1.

Series	1sg.	2sg.	3sg.	1pl.	2pl.	3pl.
Basic	*ɲé	*í / *é	*ì / *è	*mù	*wó	*tì
Imperfective	*ɲáá	*yáá	*ɲàà	*màà	*wáá	*tàà
Focalised	*ɲá	*íyá / *bíyá	*à	*mùyà	*wóyá	*tiyà
Comitative / Instr.	*ɲ-yé	*íyé	*lá	*mùyè	*wóyè	*tiyè
Object	*ɲ-	*í-	*ɲ-	*mù-	*wó-	*tì-
Possessive	*ɲí- / *ɲé-	*í-	*ɲì-	*mù-	*wó-	*tì-
Imperative	—	*ø	—	*mù (du.), *kàmù (pl.)	*kà	—
Negative	*ɲéè	*yéè	*ɲèè	*mùèè	*wóèè	*tìèè

Table 2.

Original semantics	Meanings	Examples
Motion	prospective, future, inchoative, conditional	Loko <i>ba</i> ‘come’, Vai <i>na</i> ‘come’, Gban <i>nù</i> ‘come’, Kagoro <i>na</i> ‘come’, Mano <i>ló</i> ‘go’, Wan <i>zò</i> ‘come’
Phase	perfective, resultative	Jalonke <i>banta</i> ‘finish’, Kagoro <i>máɲ</i> ‘close’
Existence	durative, habitual	Seenku <i>sín</i> ‘be’, Dyula <i>bé</i> ‘be’, Busa <i>gɔ</i> ‘stay’
Perception	prospective	Looma <i>ká</i> ‘see’
Volition	prospective, future	Zialo <i>pɔ</i> ‘want’

phonology and nominal morphosyntax, and, most recently, papers by K. Babaev (2010, 2011) that compare person marking systems in SW Mande.

The SW Mande tongues are known for their mostly isolating grammar structure and quite scant morphology. Naturally, these features are combined with super-short morphs (mostly of the CV or CVN types) and a vast and complicated tonology. Most languages of the group share a system of two level underlying tones, but they tend to mutate and interchange according to a variety of complicated rules within a syntactic group, so the surface tone picture may almost never reflect the original lexical tones.

The morphosyntax of SW Mande is therefore hard to analyse even on the synchronic level, but diachronic analysis reveals even more serious obstacles. Since morphemes are short, verbal grammatical markers and auxiliaries often become impossible to compare. Some of the verbal markers and clitics have a V / N structure which cannot be found among other parts of speech; this makes it impossible to compare diachronic changes in grammatical morphemes with the ones that take place in verbal or nominal roots. Moreover, it seems that the markers are quite unstable in the particular languages and tend to be replaced by newer morphemes very quickly. At least some of them in certain languages of the group look brand new, and their genesis is easy to trace. But the older markers that they replace can sometimes be found only in the status of “rudiments”. There are cases of fusion of several (up to three) verbal markers into one, as well as alternating usage of several (older and newer) markers in the same meaning.

SW Mande languages are characterised by a number of morph boundary sandhi processes that look similar to the initial consonant mutations in Celtic languages of Europe. These happen to be caused by initial or final *N, the homorganic nasal which was dropped almost everywhere (except for Kpelle, where it was preserved in word-final position). The drop of *N

has resulted in a number of consonant alternations not only word-initially, but also on the boundaries of morphs within a word.

For these and other reasons, it is sometimes advisable to reconstruct proto-language constructions rather than morphs. Such constructions consist of the verb root combined with grammatical markers. The latter will include both verbal meanings (aspect / tense, modality, polarity) as well as the meanings referring to the subject of the clause, namely, the meanings of person and number, expressed by the so called predicative person markers (PPM), introduced in detail in the works by Vydrin [Выдрин 2010] and Zheltov [Желтов 2012].

In such verbal constructions, the structure is usually fixed and stable. There are three possible positions for verbal markers: initial (placed at the very beginning of the clause), post-subjective (placed between the predicative subject marker and the verb), and post-predicative (following the verb). For details of their syntax, one of the author's previous papers on the topic (Babaev 2011) may be consulted.

The post-subjective markers, locked in their position between the PPM and the verb, as a rule, become gradually fused with the former, generating a variety of pronominal sets or paradigms which in SW Mande are the only grammatical markers of both the subject and the predicate of the clause. Thus, the Zialo sentence *gèy va* 'I will not come' incorporates both the person and the number of the subject *and* the TAMP meanings of the verb within the pronominal marker.

Therefore, the most obvious thing to reconstruct in the verbal sub-system of the morphosyntax would be a system of pronominal sets; the listing below generally follows Babaev 2010, with but a few slight modifications: Table 1.

Analysis of these fused forms can give only a few ideas about how their elements, previously syntactically independent from each other, may have looked like. The following predicative markers can be identified:

- *à (imperfective), originally possibly *Nà
- *yá (focus / topic)
- *yé (comitative / instrumental)
- *éè (negative), originally possibly *Néè

Since we may be relatively sure that the origins of most of these morphemes lie within the verbal sphere, it would be interesting to suggest some verbal roots which could have been the source of the markers on this list.

Typologically, five semantic groups of verbs are well known to serve as the source of grammatical auxiliaries, and these can be traced across the Mande languages as well as other language families of the world. The following chart shows some examples of those roots, which, in some Mande idioms, can today be employed both as predicates and auxiliaries within the same clause: Table 2.

Comparative data allow to conclude that the imperfective morpheme *Nà can be traced back to *na 'to come', reconstructed for Proto-Mande. However, the ultra-short structure of the majority of the markers leaves little space for reliable hypotheses.

The following chart demonstrates some of the post-subjective predicative markers in SW Mande which may be projected onto the protolanguage level. The constructions are marked with the abbreviation of the pronominal set (BAS, FOC, IMP). The "Representation" column lists those SW Mande languages where the construction is actually witnessed. The external correspondences are taken from across the Mande language family (see the list of abbreviations at the end of the paper): Table 3.

For the moment, this seems like an exhaustive list of verbal markers which can be reconstructed for Proto-SW Mande. The absolute majority of them should be traced back to an even

Table 3.

Meaning	Construction	Representation	Correspondences
Resultative / perfective	BAS V *-á / *Ná	M Lk Z L K	South Mande *-la, Beng -nã, Vai -a / -da, Manden -ta / -la / -ra, Lele -ta / -la / -ra
Preterite	BAS V *-i / *Ni	M Lk B Z L	Soso -xi
Aorist	BAS V *-ɔ / *Nɔ	Lk B Z	
Imperfective	S *a / *ɣa	M Lk B Z L K	Manden *na, Vai na, Bisa ná, Bobo nà
	S V -ma	M Lk B Z	Mande *ma 'on the surface', Soso-Jalonke *-ma, Guro -mã, Yaure -ma / -a
Progressive	FOC *tɔ V (*ma)	M Lk B Z	*tɔ 'stand'
	*na(N) S V	B Z L	*na 'time, when'
Imperative	IMP V	M Lk B Z L K	= Mande
Conditional	S *ɔ V	Lk B L	
Negative	S *ta V	Z L K	Mande *ta 'go, leave'
	S *ε(ε) < lε ?	M B Z L	Vai -le, Manden *tε, Lele té, Jeri tε, Bozo te, Guro tē

higher chronological node, some to Proto-Mande or even further (cf. Proto-Niger-Congo *tε / *té 'not be' as the negation marker).

The following conclusions may be offered:

- several morphs can be reconstructed for TA and negation marking in Proto-SW Mande;
- the origins of these morphs may in most cases be found in either the noun or the verb system;
- the basic opposition PFV / IPFV in the verb system is duly confirmed, the perfective cluster being marked by the basic (unmarked) pronominal set;
- modality is represented mostly by the imperative and conditional moods (probably ir-realis);
- regeneration of the verbal systems and rapid 'cycles of grammaticalisation' make it a tricky task to go deeper than 1500–2000 years in reconstructing the Proto-Mande verb system.

Literature

- Babaev 2010 — БАБАЕВ, К. Person marking in South-West Mande languages: a tentative reconstruction. *Mandenkan*, 46, 2010, pp. 3–48.
- Babaev 2011 — БАБАЕВ, К. On the reconstruction of some tense/aspect markers in Proto-Mande. *Journal of Language Relationship*, 6, 2011. Pp. 1–23.
- Dwyer 1973 — DWYER, David J. *The comparative tonology of Southwestern Mande nominals*. Michigan State University, Ph.D.
- Бабаев 2011 — БАБАЕВ К. В. Личное маркирование в юго-западных языках манде: опыт сравнения и реконструкции. *Труды Института лингвистических исследований*. Том 7, ч. 2. СПб.: Наука, 2011. Стр. 393–449. [БАБАЕВ К. В. Lichnoe markirovanie v yugo-zapadnykh yazykakh mande: opyt sravneniya i rekonstruksii. *Trudy Instituta lingvosticheskikh issledovaniy*. Tom 7, ch. 2. SPb.: Nauka, 2011. Str. 393–449.]
- Выдрин 2006 — ВЫДРИН В.Ф. К реконструкции фонологического типа и именной морфологии пра-манде // *Труды Института лингвистических исследований*. Т. II, ч. 2. СПб., 2006 [VYDRIN, Valentin. On the Recon-

struction of the Phonological Type and Nominal Morphology of Proto-Mande. *Works of the Institute of Linguistic Research*, vol. 2, no. 2. Pp. 9–252. St. Petersburg: Nauka,], pp. 9–252.

- Выдрин 2010 — ВЫДРИН В. Ф. Ещё раз о “субъектных местоимениях” в южных языках манде: местоимения или предикативные показатели? // ВИНОГРАДОВ В. А. (ред.). *Основы африканского языкознания: синтаксис именных и глагольных групп*. М.: Академия. Стр. 385–400. [VYDRIN V. F. Esche raz o “sub’ektnykh mestoimeniyakh” v yuzhnykh yazykakh mande: mestoimeniya ili predikativnye pokazateli? // VINOGRADOV V. A. (red.). *Osnovy afrikanskogo yazykoznaniiya: sintaksis imennykh i glagol’nykh grupp*. М.: Akademiya. Str. 385–400.]
- Желтов 2012 — ЖЕЛТОВ А. Ю. Еще раз о местоименно-предикативных показателях в языках манде: к типологии маркеров персональности // *Африканский сборник 2011*. СПб.: МАЭ РАН, 2012, с. 197–203. [Zhel'tov A. Yu. Esche raz o mestoimenno-predikativnykh pokazatelyakh v yazykakh mande: k tipologii markerov personal'nosti // *Afrikanskij sbornik 2011*. SPb.: MAE RAN, 2012, s. 197–203.]

К. В. БАБАЕВ. К реконструкции глагольного морфосинтаксиса в праюгозападном манде.

Статья написана по материалам доклада о предварительных результатах коллективного исследовательского проекта «Нигеро-конголезский праязык: сравнение и реконструкция». В ней проанализированы сравнительные данные глагольного морфосинтаксиса юго-западных языков манде и определены возможности реконструкции праязыковых глагольных показателей.

Ключевые слова: языки манде, реконструкция, глагол, морфология, синтаксис, языки Африки.

Indo-European nominal inflection in Nostratic perspective*

The paper summarizes some of the current views on the history and origins of Indo-European nominal declension, including a brief comparison of several hypotheses on the mechanism of arisal of the “thematic” type of declension (*o*-stems). The reconstructed paradigmatic system is subsequently compared with the respective systems for other language families that form part of the hypothetical Nostratic macrofamily: Semitic (and Afro-Asiatic in general), Kartvelian, Uralic. It is concluded that, since most of the case endings of Proto-Indo-European are explainable either through internal derivation within Indo-European itself or through comparison with other Nostratic languages, the situation points strongly towards an analytic nature of Proto-Nostratic.

Keywords: Historical linguistics, Indo-European languages, comparative morphology, Nostratic hypothesis.

0. One of the parameters employed in the so-called typological classification of languages is whether the languages in question use or do not use affixation, i. e. whether they express such nominal grammatical categories as case, number, gender, grade, etc., or such verbal categories as person, number, time, aspect, mood etc., by means of suffixes, prefixes, or infixes. Such languages are called **synthetic**. Languages without affixation, called **analytic**, use other grammatical tools to express these categories (if they are differentiated in the first place), such as (frequently) various auxiliaries and particles, as well as more or less firmly fixed word order. Synthetic languages are further divided into **agglutinative** and **flective**. A basic feature of the agglutinative languages is that each of their affixes bears only one function (cf. e.g. Čermák 1997: 178–80). This can be illustrated on the example of three specimen agglutinative languages, Turkish, Finnish, and Georgian, where it is easy to identify the case endings that are universal for both singular and plural numbers: Table 1. In addition, Turkish and Georgian use only one plural suffix. In Finnish, two plural markers are differentiated: *-t* in the nom.-acc. and *-i/-j-* in oblique cases.

0.1. From the “European” point of view, languages that classify nouns into specific semantic fields via so-called **class affixes** may seem exotic. Class affixes are congruentially introduced in adjectives, pronouns, numerals and verbs (sometimes they are lacking in nouns as such, with the exception of kinship terms in Nakh languages or kinship and anatomical terms in Burushaski from Hindukush, where they are productive). Vladimír Skalička (1945[2004]: 284–315) demonstrated that Swahili and other Bantu languages with class systems are flective inasmuch as several functions (e.g., class and number) may be cumulatively expressed with just one prefix. The same may be said about Bats, Chechen, and other languages with class prefixes spoken in the Northeastern Caucasus. On the other hand, the system of class prefixes in Aka-Bea on the South Andaman Island in the Indian Ocean bears apparent features of the agglutinative pattern: all plural prefixes are formed from the singular prefix by means of the universal plural morpheme *-at-*. It should be mentioned that Bats, Chechen, and Aka-Bea have their own case systems of agglutinative structure: Table 2.

* The present study was prepared thanks to the grant of the The Czech Science Foundation, P406/12/0655.

Table 1.

Turkish “house”			Finnish “house”			Georgian “son”		
case	sg.	pl.	case	sg.	pl.	case	sg.	pl.
nom.	<i>ev</i>	<i>ev-ler</i>	nom.	<i>talo</i>	<i>talo-t</i>	nom.	<i>švil-i</i>	<i>švil-eb-i</i>
acc.	<i>ev-i</i>	<i>ev-ler-i</i>	acc.	<i>talo</i>	<i>talo-t</i>	erg.	<i>švil-ma</i>	<i>švil-eb-ma</i>
gen.	<i>ev-in</i>	<i>ev-ler-in</i>	gen.	<i>talo-n</i>	<i>talo-j-en</i>	gen.	<i>švil-is</i>	<i>švil-eb-is</i>
dat.	<i>ev-e</i>	<i>ev-ler-e</i>	allat.	<i>talo-lle</i>	<i>talo-i-lle</i>	dat.	<i>švil-s(a)</i>	<i>švil-eb-s(a)</i>
loc.	<i>ev-de</i>	<i>ev-ler-de</i>	ines.	<i>talo-ssa</i>	<i>talo-i-ssa</i>	dir.	<i>švil-ad(a)</i>	<i>švil-eb-ad(a)</i>
abl.	<i>ev-den</i>	<i>ev-ler-den</i>	abl.	<i>talo-lta</i>	<i>talo-i-lta</i>	voc.	<i>švil-o</i>	<i>švil-eb-o</i>
			instr.	<i>talo-n</i>	<i>talo-i-n</i>	instr.	<i>švil-it(a)</i>	<i>švil-eb-it(a)</i>
			part.	<i>talo-a</i>	<i>talo-j-a</i>			
			es.	<i>talo-na</i>	<i>talo-i-na</i>			
			trans.	<i>talo-ksi</i>	<i>talo-i-ksi</i>			
			elat.	<i>talo-sta</i>	<i>talo-i-sta</i>			
			ill.	<i>talo-on</i>	<i>talo-i-hin</i>			
			ades.	<i>talo-lla</i>	<i>talo-i-lla</i>			
			abes.	<i>talo-tta</i>	<i>talo-i-tta</i>			
			comit.	—	<i>talo-i-neen</i>			

Abbreviations: abes. — abessive, abl. — ablative, adess. — adessive, acc. — accusative, all. — allative, comit. — comitative, dat. — dative, elat. — elative, erg. — ergative, es. — essive, gen. — genitive, ill. — illative, instr. — instrumental, loc. — locative, nom. — nominative, part. — partitive, pl. — plural, sg. — singular, trans. — translative.

Table 2.

Swahili (East Africa)			Bats / Chechen (Northeast Caucasus)			Aka-Bea (Andaman Islands)		
class	sg.	pl.	class	sg.	pl.	class	sg.	pl.
I.	<i>mu-/mw-</i>	<i>wa-</i>	I.	<i>v- / vu-</i>	<i>b- / du-</i>	I.	<i>ab-</i>	<i>at-</i>
II.	<i>mu-</i>	<i>mí-</i>	II.	<i>j- / ju-</i>	<i>d- / du-</i>	II.	<i>ar-</i>	<i>arat-</i>
III.	<i>li- ~ ji-</i>	<i>ma-</i>	III.	<i>j- / ju-</i>	<i>j- / ju-</i>	III.	<i>aka-</i>	<i>akat-</i>
IV.	<i>ki-/ch-</i>	<i>vi-</i>	IV.	<i>b- / bu-</i>	<i>b- / bu-</i>	IV.	<i>ig-/i-</i>	<i>i-/iti-</i>
V.	<i>n-/ny-/∅-</i>	<i>n-/ny-/∅-</i>	V.	<i>d- / du-</i>	<i>d- / du-</i>	V.	<i>ong-</i>	<i>oiot-</i>
VI.	<i>u-/w-</i>	<i>n-/ma-/∅-</i>	VI.	<i>b- / bu-</i>	<i>d- / du-</i>	VI.	<i>ot-</i>	<i>otot-</i>
VII.	<i>ku-</i>		VII.	<i>b- / -</i>	<i>j- / -</i>			
VIII.	<i>pa-</i>		VIII.	<i>d- / -</i>	<i>j- / -</i>			
IX.	<i>ku-</i>							
X.	<i>mu-</i>							

Functions of the class prefixes:

Swahili: I. human beings; II. trees & plants, tools & products, some body parts, abstracts; III. paired body organs, fruits, liquids, big things, foreign words; IV. things; V. animals; VI. abstracts & extracts, names of countries; VII. verbal infinitives; VIII. locatives; IX. locatives; X. locatives.

Bats: I. male beings; II. female beings; III.-VIII. animals, inanimate nature, things, abstracts.

Aka Bea: I. body, back, thighs, lap, shins, calves, groin, elbows, arms, knees, ribs, navel, belly etc.; II. legs, hip, kidney, anus, urine, urinary bladder, peritoneum, gut, scrotum, buttocks; III. mouth, palate, chin, lips, tongue, throat,

pharynx, trachea, jaw, vertebra, saliva, breathe etc.; IV. eyes, eyebrow, eyelids, gum, face, forehead, ears, nose, cheeks, temples, teeth, shoulders, forearm, biceps, breasts, nipples, tears; V. hands, fingers & toes, fingernails, sole of the foot, heels, joints, spleen, small intestine; VI. heart, head, neck, chest, brain, nape, lungs, phlegm.

0.2. Indo-European languages have been cited as a typical example of fleective languages. This supposes the presence of more or less stable word roots which are extended by derivational affixes and grammatical morphemes. These may bear several functions, e.g. Latin *servī*, formed from the nom. sg. *servus* “servant, slave”, expresses both the gen. sg. m. and nom.-voc. pl. m., i.e. a single ending accumulates the grammatical meanings of case, number, and gender. As a sub-variant of the fleective type, the **introfleective** subtype has also been defined. This term refers to the so-called ‘internal inflection’, specified both on the affixes and the root of the word; as a rule, it is represented by vocalic alternations. As an illustrative example, we may cite the so-called ‘broken’ plurals of Semitic nouns, e.g. Classical Arabic *raḡul^{un}* “man”, pl. *riḡāl^{un}*, *ʕayn^{un}* “eye”, pl. *ʕuyūn^{un}*, *qaws^{un}* “bow”, pl. *qusāy^{un}*, *tāḡir^{un}* “merchant”, pl. *tiḡār^{un}*, *rākib^{un}* “rider”, pl. *rakb^{un}* ~ Hebrew *rōkēb*, pl. *rekeb*, *ḥimār^{un}* “ass”, pl. *ḥamīr^{un}* (final *-un* in Arabic represents the so-called *nunation* that performs the function of a determinant and is neither a part of the root nor of the stem). In Indo-European nominal paradigms such vocalic alternations, which can be classified as internal inflection, appear too, and are usually specified as ‘ablaut’ or ‘apophony’. Thus, in the paradigm of the word “water” different variants of the root are present in the nom. sg. **uód-r̥*, loc. sg. **ud-én*, nom.-acc. col. **uéd-ōr* (NIL: 706–07); the basic paradigm of the word “winter” is also reconstructed at least in three apophonic variants: nom. **ḡ^héiōm*, acc. **ḡ^hém-m̥*, gen. **ḡ^him-ós* (Beekes 1995: 178). Even more important is the role of the ablaut in the Indo-European verb.

Naturally, it is very rare for any language to be in complete agreement with the criteria that define concrete language types in the typological classification of Skalička. A more adequate statement is that in any one language one type is usually dominant, but other types may also appear, although they may be rudimentary or, vice versa, *in statu nascendi*.

1. Proto-Indo-European is generally seen as having had the same nominal word classes that are present in historically attested Indo-European languages, i.e. nouns, adjectives, pronouns, numerals. Numerals were declined the same way as nouns and adjectives, although this rule originally covered only numerals from 1 to 4; higher order numerals were indeclinable, just like adverbs, conjunctions, and particles. Pronouns had their own inflection system, frequently different from nominal inflection; both types of inflection could also influence one another. Nouns and adjectives (the latter serving as congruent attributes to governing nouns) differentiated three basic grammatical categories: number, case and gender. There were three numbers – singular, plural and dual, although the dual has not been preserved in all the branches. Traditionally, a system of 8 cases has been reconstructed, based on the situation in Vedic and Avestan. However, more modest case inventories in Greek, Celtic, Germanic, and especially in Anatolian imply the question whether a paradigm with fewer cases could reflect the original situation. The same issue is seen with grammatical gender: Brugmannian reconstruction operated with three genders (masculine, feminine and neuter), but the Anatolian languages differentiated only two genders. This discrepancy may be solved by the assumption of two chronological phases. In the early phase, only two genders would be differentiated, expressing the opposition “active : inactive”, whereas in the later phase, after the separation of the Anatolian branch, the traditional tripartite gender system would be introduced (Vavroušek 2008: 59). An alternative model assumes an original feminine gender and its elimination in the Anatolian branch (cf. Beekes 1995: 174).

1.1. From the point of view of word formation, an Indo-European noun consists of the root (sometimes extended with a derivational suffix) + stem-forming suffix + case ending. A good example may be the word **H₂erH₃-tr-o-m* “plough”, reconstructed on the basis of Arm. *arawr*, Gr. ἄροτρον, Lat. *aratrum*, Mlr. *arathar*, Welsh *aradr*, ON. *arðr*. It consists of the verbal root **H₂erH₃-* “to plough”, the derivational suffix **-tr-*, the stem-forming marker *-o-* classifying the word as a so-called *o*-stem, and the ending *-m* of the nom.-acc. sg. of the neuter gender. Out of all these segments, it is only the stem-forming suffix that determines the inflectional class to which the noun in question belongs. A basic division has been sought between so-called vocalic and consonantal stems. The *o*-stems and *ā*-stems are traditionally understood as vocalic, whereas the so-called *i*-stems and *u*-stems belong to the consonantal type, being resonants. If the *-ā* of the *ā*-stems is reinterpreted in agreement with the laryngeal theory as **-eH₂*, the result will also be a consonantal coda; likewise, in the case of *ī*- and *ū*-stems, which, in their laryngealistic reinterpretation, are really *iH*- and *uH*-stems. In other words, the vocalic stems will be represented only by the *o*-stems, while all the other ones will (at least historically) be consonantal stems. In nominal paradigms with ablaut the cases are called strong if the stem has the full ablaut grade, and weak if the stem has the zero ablaut grade. The full ablaut grade generally appears in nom., acc., voc. sg. and pl. and usually in the loc. sg. (Szemerényi 1996: 161).

1.2. When reconstructing Indo-European nominal inflection, syllabic accent should also be taken in account. This phenomenon, neglected by Neogrammarians, was given a new interpretation by Holger Pedersen (1926: 24), who himself came from the Neogrammarian milieu. For the consonantal stems he defined two types of inflection according to differences of accentuation: *hysterodynamic* with accent (Gr. δύναμις) always on the last syllable, and *proterodynamic* with accent moving between the first syllable and the stem-forming suffix.

In the 1970s his ideas were further developed by Karl Hoffmann, a representative of the so-called Erlangen School. Hoffmann defined 5 types of inflection depending on the position of accent (ἄκρος “at the furthest point”, πρότερος “in front, forward”, ἀμφί “on both sides of”, ὕστερος “latter, behind”, μέσος “in the middle”) and its mobility (*kinetic*) or immobility (*static*): (1) *akrostatic* — accent always remains on the root; (2) *proterokinetic* — accent stays on the root in strong cases, shifts to the stem-forming suffix in weak cases; (3) *hystero-kinetic* — accent always remains on the final syllable, which alternately coincides with the stem-forming suffix in strong cases and the case ending in weak cases; (4) *amphikinetic* — accent moves from the root in strong cases to the case ending in weak cases; (5) *mezostatic* — accent remains on the stem-forming suffix; if this suffix is preceded by another derivational suffix, accent shifts position between them (Szemerényi 1996: 161–62).

2. The so-called **thematic inflection** is limited only to nouns extended by the vowel *-o-* (which shifts to *-e-* in the voc. sg.), the so-called ***o*-stems**. Unless indicated otherwise, in table 3–4 the reflexes of the IE words **u₁lk^{uo}-* “wolf” and **iugó-* “yoke” are cited (see Beekes 1995: 190–92; GHL: 79–83; Fortson 2004: 113–16; Szemerényi 1996: 182–88).

Apart from the singular (Table 3) and plural (Table 4), Proto-Indo-European also had a separate dual category (Tables 5–7), although it was not preserved in all daughter branches. Thus, in old literary Germanic languages the dual disappeared from the nominal morphology but remained in the pronominal and verbal systems. In Latin traces of the dual are preserved only in the numerals *duo* (*duō* by Plautus), *duae*, *duo* “two”, and *ambō*, *-ae,-ō* “both”. In the Anatolian branch the traces of dual have been sought in designations of paired body parts, e.g. Hitt. *sakuwa* “eyes”, Luw. *aruta* “wings”, *issara* “hands”, *pata* “feet” (Krasuxin 2004: 133; Szemerényi 1996: 161). Because of homonymy of endings, the dual paradigm is less rich in comparison with plural even in those languages where the dual number was preserved.

Table 3.

sing.	Vedic	Hittite	Greek	Latin	Old Irish	Gothic	Lith.	OCS.	IE
nom.	<i>vṛkas</i>	<i>ishās</i> “lord”	λύκος ⁶	<i>lupus</i> ¹⁴	<i>fer</i> ²⁴	<i>wulfs</i> ³⁴	<i>vil̃kas</i>	<i>vl̃vk̃v</i>	*-os
vok.	<i>vṛka</i> ¹	<i>ishā</i>	λύκε	<i>lupe</i> ¹⁵	<i>fir</i> ²⁵	<i>wulf</i>	<i>vil̃ke</i>	<i>vl̃ṽče</i>	*-e
acc.	<i>vṛkam</i>	<i>ishān</i>	λύκον ⁷	<i>lupum</i> ¹⁶	<i>fer</i> ²⁶	<i>wulf</i> ³⁵	<i>vil̃ka</i> ⁴⁰	<i>vl̃vk̃v</i>	*-om
nom.- acc. n.	<i>yugám</i>	<i>yugan</i>	ζυγόν ⁸	<i>iugum</i> ¹⁷	<i>scél</i> ²⁷	<i>juk</i> ; ORun. <i>horn</i> “horn”	Pr. <i>lunkan</i> ⁴¹	<i>igo</i>	*-om
gen.	<i>vṛkasya</i>	<i>ishās</i> ⁴	λύκοιο ⁹	-osio ¹⁸ <i>lupī</i> ¹⁹	?-oiso ²⁸ <i>fir</i> ²⁹	ORun. -as ³⁶ <i>wulfis</i> ³⁷	Pr. -as ⁴² = abl.	= abl.	*-os *-os- <i>io</i> *-ī ⁴⁶
abl.	<i>vṛkād</i>	<i>antuhset</i> ⁵	<i>e-ko-me-</i> <i>no</i> ¹⁰	<i>lupō</i> ²⁰	Cib. -u ³⁰		<i>vil̃ko</i> ⁴³	<i>vl̃vka</i> ⁴²	*-ōd < *-o-ed ⁴⁷
dat.	<i>vṛkāya</i> ²	<i>ishī</i>	λύκωι ¹¹	<i>lupō</i> ²¹	<i>fiur</i> ³¹	ORun. -ai ³⁸	<i>vil̃kui</i>	<i>vl̃vku</i>	*-ōī < *-o-eī
loc.	<i>vṛke</i>	<i>ishī</i>	οἴκοι ¹²	<i>domī</i> ²²	Cib. -ei ³²	dat. <i>wulfa</i>	<i>vil̃kè</i> ⁴⁴	<i>vl̃včė</i>	*-oī < *-o-i
inst.	<i>vṛkeṇa</i> ³		<i>a-to-ro-</i> <i>qo</i> ¹³	<i>lupō</i> ²³	Cib. -u ³³	dat. <i>wulfa</i> ³⁹	<i>vil̃kū</i>	<i>vl̃vkomv</i> ⁴⁵	*-oH ₁ < *-o-H ₁

Notes: see Beekes 1995: 190–92; GHJ 79–83; Szemerényi 1996: 182–88; 1) Absence of palatalization of the velar is explainable via paradigmatic levelling. 2) Final *-a* is a particle (Beekes 1995: 192). 3) Adapted from the pronominal inflection, cf. the instr. sg. of demonstratives *ena*, *téna*, *eténa*, interrogative *kéna*, relative *yéna*, reflexive *svéna* (Macdonell 1916[2000]: 77, 106–12). 4) Corresponds to the gen. ending of other inflectional types. In the Luwian sub-branch the genitive is replaced by so-called possessive adjectives, formed by the suffix *-assi-* in Cuneiform Luwian and *-asi-* in Hieroglyphic Luwian. Szemerényi (1996: 184, 187; following Mittelberger) derived them from the *o*-stem gen. in *-os-*io*. 5) Instr. sg. from *antuhšas* “man”. Abl. sg. has the ending *-az*. 6) Cf. Myc. *wo-no* /*woinos*/ “vine”. 7) Cf. Myc. *to-ro-^{qo}* /*trok^won*/ “strap” or *wo-i-ko-de*, *wo-ko-de* /*woikon-de*/ “in the house”. 8) Cf. Myc. *e-ra-3-wo* /*elaitwon*/ “olive oil”. 9) Cf. Myc. *do-e-ro-jo* /*do(h)elojjo*/ “servant”. 10) Myc. toponym, which may preserve the abl. sg. in *-ō*. 11) Cf. Myc. *a-to-ro-^{qo}* /*ant^hrōk^wōi*/; cf. Gr. ἄνθρωπος “man”. 12) Att. “at home”; cf. the Myc. toponym *pu-ro* /*pulōi*/ “in Pylos”. 13) Myc. /*ant^hrōk^wō*/ (notes 6–13: see Bartoněk 2003: 188–210 & Hajnal 1995: 23–24). 14) Cf. OLat. *equos* (Plautus), *duenos* (CIL I, 4), further Fal. *Kaios*, Ven. *Voltiimnos*, Osc. *húrz* “grove”, SPic. *meitims* “gift?”. 15) Cf. Fal. *Uoltene*, Umb. *Šerfe*, Osc. *ἄτοφε*. 16) Cf. OLat. *manom* “hand”, Fal. *vinom* “vine”, Ven. *ekvon* “horse”, Osc. *húrtúm*, *dolom* “intent”, SPic. *meitimúm*. 17) Cf. OLat. *donom* “gift”, Fal. *duenom* “good”, Ven. *donom* & *donon* “gift”, osk. *sakaraklúm*, vol. *pihom* “of pious”. 18) Cf. OLat. *Popliosio Valesiosio* = class. Lat. ‘Publii Valerii’ (Lapis Satricanus, 500 BC), Fal. *Kaisiosio*. 19) Cf. OLat. *Aisclapi* (CIL I, 440), *argentī* (Plautus) “of silver” (adj.), Fal. *Marci*, Ven. *louki* “of sacred grove”. 20) Cf. OLat. *fileod* “to son”, Osc. *sakaraklúd* “to sanctuary”. 21) Cf. OLat. *Numasioi duenoi*, later *Aiscolapio*, Fal. *Kaisioi*, Ven. *murtuvoi* “to dead”, Osc. *húrtúi*, SPic. *brimeqlúí*. 22) Lat. “in homeland”, lit. “at home”; cf. Osc. *húrtéi*. 23) In a form such as *lupō* there was a merger of three cases — dative, ablative & instrumental sg.; cf. also Ven. instr. *Voltiio* (notes 14–23: see Meiser 1998: 134–35; UB: 228–30). 24) Goid. **uiros* “man”; cf. Cib. *ueidos* “witness”, Gl. *tarvos* “bull”, Lep. *Alkouinos*, *Ulkos*. 25) Goid. **uire*, cf. Gl. (Lezoux) *nate* “oh, son!” (LG: 143). 26) Goid. **uirom*; cf. Cib. *elađunom*, Gl. (Alise) *celicnon* “hall”, but (Larzac) *brictom* “magic”. 27) Olir. *scél* “message” < **sketlom* (LEIA: S 39–40); cf. Cib. *Belikiom*, Lep. *uinom* “vine”, Gl. (Chamalières) *ollon* “all”, *meion* “small”. 28) Lep. gen. sg. *Xosioiso*. Metathesis? 29) Goid. **uirī*, Ogam *maqqi* “of son”; cf. Gl. (Couchey) *Segomari*, (Todi) *Trutikni*, Lep. *Aškoneti*, but Cib. *-o* in *karuo* etc. 30) Cib. *Usamuđ*. 31) Goid. **uirū*; cf. Cib. *ueiđui*, Lep. *Pelkui*, Gl. (Cavillon) *βαλαυδοῦ* & (Couchey) *Alisanu*. 32) Cib. *kortonei*, Gl. *-e: uo duno derce* “under the barrow”, in *Alixie* “in Alesia”. 33) Cib. *auku*, Gl. *sunartiu* “by good strength” (notes 24–33: see Lejeune 1971: 467; LG: 51–56; MLH: 396–400). 34) Cf. ORun. *laukaz* “garlic”. 35) Cf. ORun. *staina* “stone”. 36) ORun. *Godagas*. 37) Gmc. **-eso*, adapted from the

pronominal inflection, cf. Goth. *hvis* “what”, *þis* “that”, OCS. *česo* “what”. 38) Cf. ORun. *hahai*. 39) Cf. the instr. sg. in OSas. *dagu*, OHG. *tagu* “by day”. 40) Cf. Prus. *deiwan* “god”. 41) Prus. “bast”. 42) Prus. *deiwas*. 43) **-ād* as in the *ā*-stems (cf. Szemerényi 1996: 183). 44) The diphthong remains fossilized in such adverbs as Lith. *namiē* “at home”, Prus. *bitai* “in the evening”. 45) Levelled after other inflectional patterns (Ehart 1982: 120). The original ending, which should be **-a*, may have been fossilized in the interrogative & relative OCS. adv. *kogda*, *kvogda* “when”, if it is a compound of the pronominal root **ko-*, **kv-* and the word *godv* “time” (Wiedmann apud Brugmann 1911: 189). 46) Cf. also Alb. *et* “of father” < **attī* : nom. *atē*; Toch. A *Mahiśvari* “Mahiśvari’s”, maybe also the genitive of kinship terms such as A *pācri*, B *pātri* “of father” (Klingenschmitt 1992: 98–104). 47) Only the abl. sg. of the *o*-stems differs from the gen. sg., in other inflectional classes gen. & abl. sg. merge (with the exception of Italic, Celtiberian and Anatolian languages, where the dental ablative was extended in other inflectional classes). Its origin has been sought in the pronominal ablative of the type of Ved. *mad*, OLat. (also acc.) *med* (Praeneste), *mēd* (Plautus), Fal. *med*, *met*, similarly OLat. *ted* & *sed*, and perhaps Gl. **med* & **sed* in the syntagms *to-med-ec-lai* (Voltino, North Italy), *met-ingi-set-ingi* “between me and between her” (Châteaubleau; see Lambert 2001: 112). The ending is identified with the particle **ad* < **H₂ed* (Dunkel, Sihler etc.) or **eti* ~ **oti* < **H₁eti* ~ **H₁oti* (Neu, Tichy etc.), see Szemerényi 1996: 187).

Table 4.

plur.	Vedic	Hittite	Greek	Latin	OIrish	Gothic	Lith.	OCS.	IE
nom.	<i>vṛkās</i>	<i>ishēs</i>	λύκοι ³	O.-U. <i>-us</i> ⁹ <i>lupī</i> ¹⁰	<i>fir</i> ¹⁶	<i>wulfos</i> ²⁴	<i>vilkaĩ</i> ²⁹	<i>vlvci</i>	<i>*-ōs</i> < <i>*-o-es</i> <i>*-o₁</i>
acc.	<i>vṛkān</i> ¹	EN.MEŠ- <i>us</i>	λύκουσ ⁴	<i>lupōs</i> ¹¹	<i>firu</i> ¹⁷	<i>wulfans</i>	<i>vilkaĩs</i> ³⁰	<i>vlvky</i> ³⁵	<i>*-oms</i>
nom.- acc. n.	<i>yugā</i>		ζυγά ⁵	<i>iuga</i> ¹²	<i>scél(a)</i> ¹⁸ Cib. <i>-a</i> ¹⁹	<i>juka</i> ²⁵	Prus. <i>warto</i> ³¹	<i>iga</i>	<i>*-(e)H₂</i>
gen.	<i>vṛkānām</i> ²	<i>siunan</i> “god”	λύκων ⁶	<i>lupō-</i> <i>rum</i> ¹³	<i>fer</i> ²⁰	<i>-o/-a</i> ²⁶ <i>wulfē</i> ²⁷	<i>vilkaĩ</i> ³²	<i>vlvkv</i>	<i>*-ōm</i> < <i>*-o-om</i>
dat.- abl.	<i>vṛkebhyaś</i>	<i>ishas</i>		<i>-oibos</i> ¹⁴	<i>fer(a)ib</i> ²¹	<i>wulfam</i> ²⁸	<i>vilkaĩms</i> ³³	<i>vlvkomv</i>	<i>*-o₁-b^h/m-</i> <i>*-o₁-os?</i>
loc.	<i>vṛkeṣu</i>	<i>ishas</i>	λύκοισι ⁷	<i>lupīs</i> ¹⁵	<i>-ei</i> ²²		<i>vilkuosē</i> ³⁴	<i>vlvcčxv</i>	<i>*-o₁-su</i>
inst.	<i>vṛkaīs</i>		λύκοισι ⁸	<i>lupīs</i> ¹⁵	<i>-uis/-us</i> ²³		<i>vilkaĩs</i>	<i>vlvky</i> ³⁶	<i>*-ō₁s</i> ³⁷

Notes: see Beekes 1995: 192; Fortson 2004, 113–16; Szemerényi 1996: 183–88; 1) Long *-ā-* levelled after the nom. pl. *-ās*. Better is preserved the acc. pl. in OAv. *mašiiṇg*, *mašiiqs-cā* “man” (**“mortal”*) < **martianh*, *ans-ca*; cf. Ved. *mārtiyan*, *mārtiyanīs-ca* (Hoffmann & Forssman 1996, 120). 2) Through influence of the *n*-stems. The primary gen. pl. is preserved e.g. in the syntagm *dévāñ janma* “race of gods” (Szemerényi 1996: 185). 3) Cf. Myc. *o-no /onoī/* “donkeys”. 4) The ending of the acc. pl. is preserved in Arg. *υίονς* “sons”; Myc. *si-a₂-ro /sihalons/*; Gr. nom. sg. *σίαλος* “fat hog”. 5) Cf. Myc. *do-ra /dōra/*; Gr. *δῶρον* “gift”. 6) Cf. Myc. *e-ra-po /elap^hōn/*; Gr. *ἔλαφος* “deer”. 7) Cf. Myc. *te-o-i /t^he(h)oi(h)i/*; *θεός* “god”. 8) Cf. Myc. *de-so-mo /desmois/*; Gr. *δεσμός* “strap” (notes 3–8: see Bartoněk 2003: 188–210; Hajnal 1995: 23–24). 9) Osc. *Núvlanús*, SPic. *Safinús*, Umb. *Ikuvinu*. 10) Cf. OLat. *peploe*, besides epigraphic *Virei*. 11) Cf. OLat. *deivos*, Ven. *deivos*, Osc. *feihúss*, Umb. *viltuf*. 12) Cf. Umb. *iuku* & *iuka*. 13) Adapted from the pronominal inflection: *ōrum* < **-ūsom* < **-o₁-s-ōm*, cf. the Vedic demonstrative *tésām*, OCS. *těxv*. The nominal gen. pl. is preserved in OLat. *deom*, Ven. *Oterginon*, Pael. *Cerfum*, CPic. *Safinúm*, Osc. *Núvlanúm*. 14) Ven. *ekvoibos*, besides *louderobos*. 15) In Lat. the dat.-abl. pl. of *o*-stems in *-īs* merged with two other cases, the loc. pl. in **-o₁-su*, and instr. pl. in **-o₁s*. The diphthong is also preserved in OLat. *quouis*, *poplois*, Pael. *puclouis*, Osc. *feihúis*, *zicolouis* (notes 9–15: see UB: 229–30). 16) Goid. **uirī*; cf. Gl. (Brioni) *Tanotalíknoi*, (Mailly-le-Camp) *ταουτανοι* & (Saint-Germain) *Aresequani*, Lep. *Kasiloī*, Cib. *Alaboī*. The original nom. pl. in **-ōs* was preserved in OIrish voc. pl. *firu* < Goid. **uirūs*. 17) Goid. **uirūs*; cf. Gl. (Graufesenque) *tuθθus*, Cib. *?matus*. 18) Goid. **sk^eetla*. 19) Cib. *korta*. 20) Goid. **uirom*; cf. Gl. (Chama-lière) *diiiuion* & *ande-dion*, Cib. *Titum*. 21) Goid. **uiro[i]bi(s)* (Thurneysen 1946: 182) with the ending **-bi(s)*, corresponding to the Gl. instr. pl. *gobedbi*, cf. dat. pl. in Gl. *Rudiobo*, Lep. *Uvltiauobos*, Cib. *Uetikubos*, similarly abl. pl. *Nouantubos*. 22) Cib. *Lutiakei* (Cólera 2005: 124). 23) Gl. (Vaison) *τοουτιους* / (Graufesenque) *Vindulus*. 24) The pronominal nom. pl. in *-ai* appears in the Gothic strong adjective nom. pl. m. *blindai* “blind”. Further cf. ORun.

arbijarjostez “most legitimate-to-inherit”?, besides later *stAinAz* “stones”, ON. *ulfar* “wolves”. 25) Cf. ORun. *hag^alu* “hail”: nom. sg. n. *hag^ala*. 26) ORun. *Wiwio*, later *flAinA*, ON. *daga*, OEng. *dæga*, OSax. *dago* “days” < **-ōⁿ*. 27) Gothic gen. pl. m./n. in *-e* instead of expected **-o* had perhaps to eliminate homonymy with the gen. pl. f. (Brugmann 1911: 238–39; Ringe 2006: 282). Kortlandt apud Beekes (1985: 142) explains *-e* from Gmc. **-ē* < **-e_i-om*, thus from the gen. pl. *i*-stems, whence the ending had to spread. 28) Cf. ORun. *borumz* “to sons”. Kortlandt (apud Beekes 1985: 144) explains the vowel *u* in the OHG. dat. pl. *tagum* via *u*-umlaut caused by vocalization of the ending **-mus* (notes 23–27: see Antonsen 1975: 18–19). 29) Cf. Prus. *wijrai* “men”. 30) Lith. *-us* < **-uos* < **-ōns* (Otrębski III: 16). Cf. Prus. *deiwan* “gods”. 31) Prus. “doors”. 32) Cf. Prus. *grikan* : nom. sg. *grikas* “sin”. 33) Cf. OLith. [Daukša’s Postilla] *waykâmus* = Lith. *vaikâms* “to children”. 34) The Lith. ending of loc. pl. could be the result of contamination with acc. pl. **-uons* and postposition **-en*, cf. Žemaitic loc. pl. in *-unse* (Otrębski III: 16). 35) Sl. acc. pl. *-y* < **-ōns*. 36) Sl. instr. pl. *-y* < **-ōis* (Erhart 1982: 121). 37) Instr. pl. in **-o_is* (Gr., It., Balt.) or **-ō_is* (IL., Celt., Sl.) remains obscure. Maybe it is the instr. sg. in **-oH₁* extended with non-singular **-o_i-* and pluralizing *-s* (otherwise Erhart 1982: 98; Szemerényi 1985: 519–20).

Table 5.

dual.	Vedic	Avestan	Greek	Old Irish	Lith.	OCS.	IE
nom.-acc. m.f.	<i>vṛkā, vṛkau</i>	<i>spādā³</i>	<i>λύκω⁷</i>	<i>fer</i> < * <i>uirā</i>	<i>vilkū</i>	<i>vlvka</i>	* <i>-o-H₁</i>
nom.-acc. n.	<i>yugé</i>	<i>šiiāoθnōi⁴</i>	<i>ζυγώ⁸</i>	<i>scél¹¹</i>		<i>i(d)zě</i>	* <i>-o-iH₁</i>
gen.	= loc.	<i>qsaiiā⁵</i>	= loc.	<i>fer</i> < * <i>uiro_u</i>	<i>dviejau¹²</i>	= loc.	* <i>-o_i-H₁ōs</i>
loc.	<i>vṛkayos¹</i>	<i>zastaiiō⁶</i>	<i>λύκου⁹</i>		<i>dviejau¹²</i>	<i>vlvku</i>	* <i>-o_i-H₁o_u</i>
dat.-abl.-inst.	<i>vṛkābhyām</i> <i>aśvebhyām²</i>	<i>zastōibiā⁶</i> <i>aspaēbiā²</i>		<i>fer(a)ib</i> < * <i>uirobim</i>	<i>vilkām</i>	<i>vlvkoma</i>	* <i>-o_i-b^hiōH₁</i> * <i>-o-moH₁</i>
instr.			<i>o-mo-pi¹⁰</i>		<i>vilkamī</i>		* <i>-o-b^h/miH₁</i>

Notes: see Beekes 1995: 194–95; Brugmann 1911: 282–83; Hoffmann & Forssman 1996; Szemerényi 1996: 183–85. 1) **-o_i-H₁o_u* + *-s*, perhaps levelled after the gen. 2) “horse”. 3) OAv. *spāda-* “army”. 4) OAv. *šiiāoθna-* “action”, cf. YAv. *saite* “200”, *duiie*, *duuaē-ca* “2”. 5) OAv. *qsa-* “share”. 6) OAv. *zasta-* “hand”. 7) Cf. Myc. *po-ro* /*pōlō*/ “two foals”. 8) Ending of the m.-f. du. 9) Rix (1976: 141): **-o_isin* after the dat.-loc. pl. **-o_isi* with final nasal after the instr. du. **-o_i-b^him*; otherwise Beekes (1995: 195): **-o_i-Hu-m?* 10) Myc. /*oimoph^hi*/ : Gr. οἶμος “bundle” (Hajnal 1995: 23). 11) Causes nasalization of the following initial: maybe due to contamination with the ending of nom. sg. n. **-om*. More original is perhaps Gl. *uercobreto* (Thurneysen 1946: 182). 12) Lith. *dvejī* “double”.

Table 6.

<i>eH₂-stems</i>	Vedic	Avestan	Greek	OIrish	Lithuanian	OCS.	IE
nom.-acc.	<i>séne</i> “army”	<i>uruuairē¹</i>	<i>θεᾶ³</i>	<i>túait⁵</i>	<i>ranki</i> “hands”	<i>rqcě</i> “hands”	* <i>-eH₂-iH₁</i>
gen.-loc.	<i>sénayos</i>	<i>uruuaraiiā²</i>	<i>θεαῖν⁴</i>	<i>túath⁶</i>		<i>rqku</i>	* <i>-eH₂-iH₁-H₁o_u</i>
dat.-abl.-ins.	<i>sénābhyām</i>	<i>vqθbābiā²</i>		<i>túath(a)ib⁷</i>	<i>rañkom</i>	<i>rqkama</i>	* <i>-eH₂-b^hiōH₁</i> * <i>-eH₂-moH₁</i>

Notes: see Beekes 1995: 194–95; Brugmann 1911: 284–85; Hoffmann & Forssman 1996: 122; Krasuxin 2004: 140; Szemerényi 1996: 188–90; 1) YAv. *uruuarā-* “plant”; cf. OAv. *ubē*, YAv. *uiie* “both”. 2) YAv. *vqθbā-* “herd”. 3) Gr. “goddess”; cf. Myc. *pte-no* /*pternō*/ : Gr. πτέρνη “heel”, besides *wo-ra-e* /*wōrā^h(h)e*/ : Gr. ὠρη “care” (Bartoněk 2003: 618; Hajnal 1995: 16). Final *-a-e* may reflect **-ā(i)e* < **-eH₂-iH₁* (Rix 1976: 135). 4) Cf. Myc. *wa-na-so-i* /*wanasso(ji)(n)*/ : Gr. (ῥ)άνασσα “lady”, besides Arc. gen.-dat. κρᾶναυον : Att. κρήνη “spring, source” (Bartoněk 2003: 166, 612; Rix 1976: 135). 5) Goid. **toūt-ī* “races”; cf. OIr. *mnaí* “two women” < **g^{un}-eH₂-iH₁*. 6) Goid. **toūtō_u* levelled after the *o*-stems. 7) Goid. **toūtābim* (Thurneysen 1946: 189–90).

Table 7.

cons. stems	Vedic	YAvestan	Greek	Old Irish	Lith.	OCS.	IE
nom.-acc. m.f.	<i>pitárā,-au</i> ¹	<i>nara</i> ³	<i>πατέρες</i> ⁴	<i>athir</i> ⁷	<i>žmune</i> ¹¹	<i>kamēni</i> ¹⁴	*-H ₁ e
nom.-acc. n.	<i>nāmanī</i> ²		<i>ὄσσε</i> ⁵	<i>ainm</i> ⁸	<i>aki</i> ¹²	<i>imeni, -ě, oči</i>	*-iH ₁
gen.-loc.	<i>pitrós</i>	<i>narā</i>	<i>πατέροις</i> ⁶	<i>athar</i> ⁹		<i>kamenu</i>	*-H ₁ ou(s)
dat.-abl.-ins.	<i>pitṛbhyām</i>	<i>narabiia</i>		<i>athr(a)ib</i> ¹⁰	<i>akmeniñ</i> ¹³	<i>каменѣма</i>	*-b ₁ ioH ₁ *-moH ₁

Notes: see Beekes 1995: 194–95; Brugmann 1911: 294–97; 1) “two fathers”; 2) “two names”; 3) “two men”; 4) “two fathers”, cf. Myc. *pi-ri-je-te-re* /*pr̥iēntēre*/ “two sawers” (Bartoněk 2003, 255); 5) “eyes”. 6) Ending of *o*-stems (Rix 1976: 160). 7) Goid. **atere* “two fathers”. 8) Goid. **anmenī* “two names”. 9) Goid. **atrouis*? 10) Goid. **atribim*. 11) OLith. *žmune* “two men” (Fortson 2004: 105). 12) “eyes”. 13) “to stones”. 14) “two stones”.

3. The issue of the origin of *o*-stems has been raised many times. Two hypotheses seem to rest on the best argumentative evidence:

3.1. Pedersen (1907: 152) mentioned that the subject of the transitive verb looked as if it had the form of the genitive (sigmatic case) if it was active, and as if it had the form of the instrumental case, if it was inactive. On the other hand, the subject and object of intransitive verbs seemed to have the form of the absolutive (i.e. asigmatic) case. This asymmetry between the valencies of transitive and intransitive verbs is summarized in Table 8 (see Beekes 1995: 193):

Table 8.

Verb	Role	Nominative system	Ergative system
transitive verb	subject	nominative	ergative
	object	accusative	absolutive
intransitive verb	subject	nominative	absolutive

Beekes (1985: 191–95; 1995: 193) and Kortlandt (2002: 217) have further developed Pedersen’s idea, assuming that the nominative syntax of old Indo-European languages was formed later and that the case system of the Proto-Indo-European language was primarily based on the ergative syntax. The same ending for the nom. and acc. neuter, originally designating inactive nouns, originated from the primary absolutive case, while the ergative was correlated with the active subject. According to Beekes, the sigmatic genitive-ablative developed from the ergative. During the transformation of the ergative system into the nominative one, the form reconstructed as *CC-R-ós* became the nominative, a new case of subject; later, the vowel *-o-* spread to other cases as well. Schmalstieg (1997: 401–07) and Gamkrelidze & Ivanov (1984: 267–91, active typology) formulated their own theories assuming the ergative past of the IE syntax. For the implications of the ergative interpretation see Table 9.

3.2. Jean Haudry (1982: 36–38) formulated the idea that the *o*-stems originated from pronouns with a determining function that were added to a nominal base, thus playing the role of a postpositional article. For this solution there are typological parallels, e.g. in Balto-Slavic languages, where adjectives used in the attributive role are extended with the IE relative stem **i_o-/*i_ā-*; postpositional determination with the use of demonstratives is also active in modern

Table 9.

case	form (*)	comments from the perspective of ergative interpretation
sg. nom.	-∅	in HD inflection and PD neuters (B 1985: 172)
	-s act.	identical with the sigmatic gen.-abl. sg., originally ergative (B 1985: 172–95)
	-om inact.	agens of the transitive verb (Pedersen 1907: 152; Schmalstieg 1997: 405–06: instr. of <i>o</i> -stems)
voc.	-∅	see B 1985: 99–108
acc.	-m	originally directive-terminative (B 1985: 198)
gen.	-(o)s	originally ergative (B 1985: 172–95); in Hitt. used for both gen. sg. & pl.
-"- o-st.	-os _i o	< gen. *-os & relative * _i o (B 1985: 185; Nikolaev 2000)
	-ī	cf. the suffix of affiliation *- _i io- (*-iHo-?); see K 1994: 98–104; Meiser 1998: 135
abl.	-os	identical with the gen. sg. in *(o)s, originally the ergative (B 1985: 172–95)
-"- o-st.	-(o/e)t/d(i)	cf. also Hitt. -az
dat.	-(e)i	on persons; dat. & loc. were originally one and the same case (B 1995: 173)
loc.	-i	on places & inactive nouns; originally identical with the dative (B 1995: 173); cf. Arm. <i>herow</i> , Gr. πέρουσι(v), ON. <i>í fforð</i> “last year”, OIr. <i>ón n-urid</i> “from last year” < * <i>per-uti</i> “year ago”
instr.	-H ₁	cf. instr. pl. of the <i>o</i> -stems *-ō _i s < instr. sg. *oH ₁ + non-singular *-o _i - + pluralizing -s
pl. nom.	-es	
-"- o-st.	-o _i	plural of pronominal origin
	-o _i -	non-singular affix of the <i>o</i> -stems, preceding the case ending
	-H ₂ inact.	collective
voc.	= nom.	
acc.	-ms	= acc. sg. -m + pluralizing -s in congruence to the nom. pl. in -es
gen.	-om	see Kortlandt 1978; in Hitt. used in both gen. sg. & pl.; Schmalstieg 1997: 405–06: ergative
abl.	- _i os	> Arm. - <i>ǰ</i> , II. *-b ^h (i)- _i os
dat.	-mus	> BSl. *-mus; Gmc. *-muz
loc.	-su ~ -si	cf. Lat. <i>mox</i> , MWelsh <i>moch</i> “soon” < * <i>mok-su</i> ; Alb. abl. pl. - <i>sh</i> : <i>malesh</i> “in mountains”; <i>për-posh</i> “down”, <i>posh-të</i> “below” : * <i>posh</i> ^o < * <i>pēd-si</i> ; Phryg. <i>τευτωσι</i> “in villages” (K 1994: 313; H 2003: 129)
instr.	-b ^h i	originally probably without number distinction, cf. Gr. ἰφί “strongly”, Myc. <i>wi-pi-no-o</i> / <i>Wip^hi-no(h)os</i> /, Hom. Ἰλιόφιν ... τείχεα “walls of Ilium” (Rix 1976: 158–59; Ba 2003: 274); II. dat. of personal pronouns: OAv. <i>maibiā</i> , <i>taibiā</i> , <i>ahmaibiā</i> , <i>yūšmaibiā</i> vs. Ved. <i>túbhya(m)</i> , <i>asmábhyam</i> , <i>yusmábhyam</i> ; OCS. dat.-loc. <i>tebě</i> , <i>sebě</i> , instr. <i>tobojo</i> , <i>sobojo</i> (see Br 1911: 187)

Notes: B = Beekes; Ba = Bartoněk; Br = Brugmann; H = Hajnal; K = Klingenschmitt. Identification of non-singular morpheme *-o_i- and pluralizer -s in some cases imply an agglutinative structure.

Balkanian or Scandinavian languages. So-called *mimation* & *nunation* in Semitic languages also generally represent postpositional determination. In Indo-European, a good ancestral candidate could be identified in the Anatolian pronoun -a- of the 3rd person, attested only in post-position (HEG 1–2, 6–7): Table 10.

Table 10.

case	Hittite	Palaic	C. Luwian	Hier. Luwian	Lydian	*
nom. sg. c.	<i>-as</i>	<i>-as</i>	<i>-as</i>	<i>-(a)s</i>	<i>-aś</i>	<i>*-os</i>
nom.-acc. sg. n.	<i>-at</i>	<i>-at</i>	<i>-ata</i>	<i>-(a)ta</i>	<i>-ad, -at</i>	<i>*-ot</i>
acc. sg. c.	<i>-an /-un</i>	<i>-an</i>	<i>-an</i>	<i>-an</i>	<i>-av</i>	<i>*-om</i>
nom. pl. c.	<i>-e</i>	<i>-as</i>				<i>*-o_i</i>
nom.-acc. pl. n.	<i>-e</i>	<i>-e</i>				<i>*-o_i</i>
acc. pl. c.	<i>-us</i>					<i>*-oms</i>
acc. pl. n.	<i>-at</i>		<i>-ata</i>	<i>-(a)ta</i>		<i>*-ot</i>

4. While summarizing the preceding partial reconstructions, it becomes apparent that the most difficult task is to establish the original protoforms of the ablative, dative and instrumental plural. It is natural to suppose some mutual levelling, merging, and contamination. These processes are summarized in Table 11 (see Brugmann 1911: 120; Beekes 1985: 144–46), whence the following case protosystem may be postulated (Table 12) to explain the partial case systems in daughter branches.

5. Internal reconstruction should be followed with external comparison. Interpretations are possible on both typological and genetic levels, provided that the Proto-Indo-European language was not completely isolated.

5.1. Semitic noun inflection uses forms both with and without determination, the latter expressed with the postpositional article in the form **-m* ('mimation') or **-n* ('nunation'). See Table 13 (Dolgopolsky 1991; Lipiński 1997; Blažek 2006).

5.1.1. Semitic languages form one branch of the Afroasiatic macrofamily. Table 14 summarizes the correspondences in case endings, prepositions & postpositions between the (sub)branches of AA.

5.2. Another language family in the neighborhood of Indo-European is Kartvelian. Its nominal declension is a transparent case of the agglutinative model, cf. the paradigm of **kac₁*-“man”: Table 15.

5.3. Another important language family in the neighborhood of Indo-European is Uralic. The agglutinative structure is also typical for Uralic languages.

5.3.1. The nominative has been reconstructed as unmarked in the Uralic protolanguage. However, it is remarkable that in several Fenno-Ugric languages a special nominative form was generated out of demonstratives or demonstrative suffixes of the 3rd person according to the scenario described by Haudry in connection with the *o*-stem IE nominative in **-s* (§3.2): Table 16.

5.3.2. Other cases are represented in the Uralic languages as follows: Table 17.

6. Finally, from the point of view of the Nostratic theory of V. M. Illich-Svitych, A. Dolgopolsky and other scholars, it is legitimate to compare the Proto-Indo-European case system with reconstructed case systems of neighboring language families so as to establish the number of common inherited morphemes: Table 18.

Table 11.

case	IE	II	Arm.	Gr.	Messap.	Italic	Celtic	Gmc.	BSl.
dat. pl.	*-mus								*-mus
>								*-mos	
abl. pl.	*-ios		-j ²						
>		*-b ^h ios			-bis ³	*-b ^h os	*-b ^h os		
ins. pl.	*-b ^h i	*-b ^h is	*-b ^h i(s)	*-b ^h i			*-b ^h i(s)		
>					-bas ⁴		*-mi ⁵	*-mis ⁶	*-miHs
dat. pl.	*-mus								

Notes: 1) The same suffix probably appears in Hitt. *kuwapi* “where, when” (HEG 4: 229–32). 2) Arm. -j in abl. pl. of personal pronouns *mēnj*, *jēnj* is derivable from *-ios (Kortlandt 1984: 103–04 = 2003, 50; Beekes 1985: 144). 3) Messap. abl. pl. *ogrebis*, *tatθebis*, *valebis* (see MLM II). 4) Messap. dat. pl. *Laidehiabas*, *Logetibas*. 5) Cf. the OIr. dat. pl. *n*-stems of the type *anm(a)imm* from *ainm* “name”; already Brugmann (1911: 188) thought about this solution and Hamp (1996) later returned to it. 6) See WGmc. dat. pl. *Vatvims* & *Aflims*, corresponding to Lat. *Vatviabus* & *Afliabus*, and OEng. dat. pl. *ḏæm* from the demonstrative *sē* “that” with *i*-umlaut (Brugmann 1911: 262, 264).

Table 12.

number	singular		plural		dual	
	athematic	thematic	athematic	thematic	athematic	thematic
nom. anim.	*-s	*-os	*-es	*-ōs < *-o-es	*-H ₁	*-o-H ₁
voc.	*-∅	*-e	*-es	*-ōs < *-o-es	*-H ₁	*-o-H ₁
acc. anim.	*-m	*-om	*-ms	*-oms	*-H ₁	*-o-H ₁
nom.-acc. n.	*-∅	*-om	*-H ₂	*-eH ₂	*-iH ₁	*-o-iH ₁
gen.	*-s	*-os(-i∅)	*-om	*-ōm	*-H ₁ ōs	*-o _i -H ₁ ōs
abl.	*-s	*-ōd < *-o-ed	*-ios	*-o _i -os	= dat.	= dat.
dat.	*-e _i	*-ō _i < *-o-e _i	*-mus	*-o _i -mus	*-mōH ₁	*-o _i -mōH ₁
loc.	*-i	*-o _i	*-su	*-o _i -su	*-H ₁ ou	*-o _i -H ₁ ou
instr.	*-H ₁	*-oH ₁	*-b ^h i(s)	*-o _i -b ^h i(s)	*-b ^h iH ₁	*-o _i -b ^h iH ₁

Notes: see Beekes 1995: 173; Fortson 2004: 113.

Table 13.

case	* Sem.	Akk.	* Ebl.	Amarna	Ugar.	Hebr.	Aram.	CIArab.	EpNAr.	EpSAr.	Geez
sg. indet.											
nom.	-u	-∅			-u(m?)	-∅	-∅	-u		-V	
acc.	-a	-∅			-a(m?)	-∅	-∅	-a		-V	-a
gen.	-i	-∅			-i(m?)	-∅	-∅	-i		-V	
loc.	-uma	-um	-ūm	-ama	-VmV	-om	-ū	-u			-u
dir.- dat.	-a/iš	-iš	-iš		-h	-ā(h)					
pred.	-a	-∅	-a			-∅	-∅	-a			-a
sg. det.											
nom.	-u-m	-um	-um		-u(m)	-∅	-∅	-un		-Vm	-∅
acc.	-a-m	-am	-a(m)		-a(m)	-∅; adv. -ām	-∅; adv. -ā	-an		-Vm	-a
gen.	-i-m	-im	-im		-i(m)	-∅	-∅	-in		-Vm	-∅
du. indet.											
nom.	-ā	-ā	-ā		/-ā/			-ā			
acc.- gen.	-ay	-ī			/-ē/	-ē	-ē	-ay	-ay	-y	
pred.	-ā	-ā						-ā			
du. det.											
nom.	-ā-ni	-ān	-ān		-m /-āmi/	-m		-āni	-n	-n(y) Q. -my	
acc.- gen.	-ay-ni	-īn	-ayn	-ēmali	-m /-ēmi/	-áyim		-ayni		-yn	
pl. A indet.											
nom.	-ū	-ū	-ū		-u /-ū/		Yu. -w				
acc.- gen.	-ī	-ī	-ī		-i /-ī/		Yu. -y	-ī			
pred.	-ū	-ū				-ū		-ū		-w	-ū
pl. A det.											
nom.	-ū-ma				-um(a) /-ūma/			-ūna			
acc.- gen.	-ī-ma				-im(a) /-īma/	-īm	-īn	-īna		-VnV	
pl. B indet.											
nom.	-āt-u	-āt			-t	-ōt	-āt	-ātu			
acc.- gen.	-āt-i	-āt			-t	-ōt	-āt	-āti			
pred.	-ā	-ā					-ā	-ā			-ā
pl. B det.											
nom.	-āt-u- m	-ātum	-ātum		-t	-ōt		-ātun			-āt
acc.- gen.	-āt-i-m	-ātim	-ātim		-t	-ōt		-ātin			-āt

Table 14.

Semitic	Egyptian	Berber	Cushitic:	Beja	Agaw	ECush.	Dah.	SCush.	Omotic
*-u ¹	*-u ¹		*-u ²⁴	*-u ²⁴ ?		*-u ²⁴			*-u ¹
			*-i ²⁴	or *-i ²⁴ ?	*-i ²⁴	*-i ²⁴			*-i/-y ¹
*-i ²	*-i ²	*-i ⁵	*-i ²	-i ²	*-i ²				*-i ²
*-a ³	*-a ³	*-a ^{3a}	*-a ²⁵	*-a ²⁵ ?	*-a ²⁵	*-a ²⁵		-a ⁴³ = 3a?	*-a ²⁵
		*i/əy ¹⁹	*-ɾ ²⁶	-i(i) ²⁼³⁰ ?		*-ɾ ²⁶	-ɾi ³¹	*i ⁴¹	
*-a/iš ⁴	js ¹⁴	*əS ²⁰	*-sa/i ²⁷	-S ³⁵	*-S ²⁶	*-sa/i ²⁷		*sa ⁴⁴	*-S ²⁶
*-uma ⁵	m ¹⁵		*-ma ⁴²			*-ma ⁴²		?*maa ⁴⁸	?A. *-m ⁵¹
*ka ⁶			*-kV ²⁸	-ka ³⁶		*-kV ²⁸		*koo ⁴⁴	*-kV ⁴⁵
*la/i ⁷	nj ¹⁶		*-la/i ²⁹		*-li ²⁹	*-la/i ²⁹		*-ɾilay ⁴⁰	
*ɿad ⁱ⁸		*dəy ²¹	*-di ²⁹	-d/-t ³⁷ ?	*-di ⁴⁰	*-d(d)i ²⁹	-da ²⁹	*-r ⁴⁹	
*ɿittul/i ^{a9}			*-tti ³⁰	-d/-t ³⁷ ?		*-tti ³⁰			Ye. -tā ⁵
*ɿina ¹⁰	jn ¹⁷	*Hīn ²²	*-na/i ³¹	-na ³⁸ , -nee ³⁹		*-ni ³¹			*-nV ⁴⁶
*ɿar ⁻¹¹	jr ¹⁸	*Hār ²³	*ɿar ⁻³²			*ɿar ⁻³²		*hari ⁴⁵	*-rV ⁴⁷
*wa ¹²			*-wa ³³		*-wa ⁴¹	*-wa ³²		*wa ³⁹	*-u/-w ²⁶
*ba/i ¹³			*-b ³⁴	-b ³⁴		h. -beyyo ⁵⁰			
*-ay ⁵²⁺⁵³	-j ⁵³				*-j ⁵⁴				

Notes: see Blažek 2006; 1 nominative, 2 genitive, 3 accusative, 3a adverbial accusative, 4 dative-terminative, 5 locative, 6 simulative (“as, like”)-comitative (“with”), 7 “to”, 8 “up to”, 9 comitative (“with”), 10 locative-inessive (“in, on”), 11 Jibbali ɿer “towards”, 12 “and”, cf. Arabic *wa-llāhi* “by God” (Sasse 2003: 139), 13 locative (“in, at, by”), 14 “as”, 15 “in, with, from; as”, 16 “belonging to”, cf. Coptic *la* “possessing”, 18 “by”, cf. Coptic (S) *ara-* “to, at; for”, 19 dative (“to” = “à”), 20 “to, at, by”, 21 “in (the middle), inside”, 22 “in, at, to”, 23 “to, till, as far as”, 24 subject, 25 absolutive = object, 26 dative, 27 dative-benefactive, 28 dative-simulative, 29 locative, 30 locative-ablative-instrumental, 31 instrumental, 32 allative-directive (“towards”), 33 adessive-directive, 34 adessive-ablative, 35 locative-genitive, 36 comparative, 37 simulative (“as”), 38 instrumental, 39 ablative, 40 comitative, 41 directive-allative, 42 Dullay “to”, 43 in adverbs of the type of Proto-Iraqw **gawáa* “on” vs. West Rift **gaba* “summit, top of the mountain”, Proto-Iraqw **afáa* “by, beside” vs. West Rift **afa* “mouth”, 44 causal-benefactive, 45 instrumental-directive, 46 instrumental-directive-locative, 47 comitative-dative, 48 Alagwa *maa* “then, afterwards” (Kiessling 2002: 377), 49 -r in **dír* “by, on”, derived from **dii* “place”, or ɿamór “by, on”, from **amoo* “way” (Kiessling 2002: 424–25), 50 directive, 51 direct object, 52 Hebr. -ē pl. in status constructus m., Aram. -e pl. in status determinus, 53 dual, 54 e.g. Kemant *azi* pl. from *aza* “fish”; cf. Chadic *-ay pl. (Illič-Svityč 1971: 285).

Table 15.

case / number	Georgian	Laz	Mingrelian	Svan	Kartvelian	
nom.	sg.	<i>ḡac-i</i>	<i>ḡoč-i</i>	<i>ḡoč-i</i>	<i>čäš</i>	nom. *-i (F 211)
	pl.	<i>ḡac-eb-i / ḡac-n-i</i>	<i>ḡoč-ep-e</i>	<i>ḡoč-ep-i</i>	<i>čäš-är</i>	pl. *-eb (F 145) pl. *-n (F 311) pl. *-ar (F 38)
erg.	sg.	<i>ḡac-ma</i>	<i>ḡoč-ik</i>	<i>ḡoč-k</i>	<i>čäš-d</i>	
	pl.	<i>ḡac-eb-ma / ḡac-ta</i>	<i>ḡoč-epe-k</i>	<i>ḡoč-e(n)-k</i>	<i>čäš-är-d</i>	
dat.	sg.	<i>ḡac-s(a)</i>	<i>ḡoč-is</i>	<i>ḡos / ḡoc</i>	<i>čäš-s</i>	dat. *-s (F 358)
	pl.	<i>ḡac-eb-s(a) / ḡac-ta</i>	<i>ḡoč-epe-s</i>	<i>ḡoč-e(n)-s</i>	<i>čäš-är-s</i>	
gen.	sg.	<i>ḡac-is(a)</i>	<i>ḡoč-iš(i)</i>	<i>ḡoč-iš(i)</i>	<i>čäš-i(š)</i>	gen. *-is ₁ (F 215)
	pl.	<i>ḡac-eb-is(a) / ḡac-ta</i>	<i>ḡoč-iš(i)</i>	<i>ḡoč-ep-iš(i)</i>	<i>čäš-are(š)</i>	
dir.	sg.	<i>ḡac-ad(a)</i>	<i>ḡoč-iša</i>	tran. <i>ḡoč-o</i> <i>ḡoč-iša</i>	<i>čäš-d</i>	advr. *-ad (F 32) dir. *-is _{1-a} (F 215)
	pl.	<i>ḡac-eb-ad(a)</i>	<i>ḡoč-epe-ša</i>	<i>ḡoč-ep-iša</i>	<i>čäš-är-d</i>	
abl.	sg.	= dir.	<i>ḡoč-iše(n)</i>	<i>ḡoč-iše</i>	—	
	pl.	= dir.	<i>ḡoč-epe-še(n)</i>	<i>ḡoč-ep-iše</i>	—	
instr.	sg.	<i>ḡac-it(a)</i>	<i>ḡoč-ite(n)</i>	<i>ḡoč-it(i)</i>	<i>čäš-šw</i>	instr. *-it (F 213)
	pl.	<i>ḡac-eb-it(a)</i>	<i>ḡoč-epe-te(n)</i>	<i>ḡoč-ep-it(i)</i>	<i>čäš-är-šw</i>	
fin.	sg.	<i>zvel-isad</i>	—	<i>ḡoč-išo(t)</i>	<i>žwintel-išd</i>	dir. *-is _{1-d} (F 215)
	pl.	<i>zvel-eb-isad</i>	—	<i>ḡoč-ep-išo(t)</i>	<i>žwintel-är-išd</i>	
tran.	sg.	= dir.	—	<i>ḡoč-o</i>	= dir.	
	pl.	= dir.	—	<i>ḡoč-ep-o</i>	= dir.	
voc.	sg.	<i>ḡac-o</i>	—	—	—	
	pl.	<i>ḡac-eb-o / -n-o</i>	—	—	—	

Notes: See Jost Gippert <<http://titus.uni-frankfurt.de>>; on reconstructions, see Fähnrich 2007 = F.

Table 16.

language	indet. sg.	det. sg.	indet. pl.	det. pl.	origin of suffix
Mordvin	<i>tolga</i> “feather”	<i>tolgas</i> “that f.”	<i>tolgat</i> “feathers”	<i>tolgat’nä</i> “those f.”	<i>šä, še</i> “that” : <i>nie</i> “these”
Udmurt	<i>iz</i> “stone”	<i>izez</i> “that s.”	<i>izjos</i> “stones”	<i>izjosjz</i> “those s.”	-z suffix of 3rd person sg.
Komi	<i>vok</i> “brother”	<i>vokjs</i> “that b.”	<i>vokjas</i> “brothers”	<i>vokjasjz</i> “those b.”	-s suffix of 3rd person sg.

Notes: See Szinyei 1910: 62.

Table 17.

case	Ural./FU	BFin.	Saami	Mordvin	Mari	Permic	Hung.	Ob-Ugr.	Samoy.
gen.	*-n	*-n	*-n	-ń	-n				*-n
acc.	*-m	*-n	*-m	-ń	-m		Ugr. *-mp ⁴ ; Ma. *-mV		*-m
loc.	*-na/*-nä	*-na/*-nä	*-nē	-ńa/-ńe	-nə	-yn	-n	*-nV	*-nV
loc.	*-tti	*-ttik, *-ten		-t'	-t	tran. -ti	-tt	-t	
abl.	*-ḍa/*-ḍä	*-ḍa/*-ḍä ¹	*-ḍē	-do	-ć				*-ḍ
dat.- lative	*-ńi	Fi. -n ² Est. -ni ³			-n	*-ń	-nyi	*-nə	*-ni
lative	*-ka/*-kä	*-k	*-kē	-va	-kə	*-ǵ	-é, -a	*-γ	
lative	*-j	Fi. -i Est. -ja				*-i	-é	Kh. -j, -i	
lative	*-s ⁵	*-s	*-s	-s	-š				
pl.	*-t	*-t	*-t	-t	-t			*-t	*-t
pl.	*-j ⁶	*-i-	*-i-				-i ⁷		*-i
pl.	*-n ⁸	*-n-							*-n
du.	*-ka/*-kä		*-kâ				kettő ⁹ “2”	*-γ-(ən)	*-kV

Notes: see Collinder 1960: 282–303; Hajdú 1985: 292–311; Hoffirková & Blažek 2012; Szinnyei 1910: 63–84; 1) Partitive. 2) Archaic dative-genitive: *Jumala-n kiitos* “thank god”. 3) Terminative. 4) Ugr. *-mp < acc. *-m + pronoun of the 3rd person sg. *sa/*sä. 5) Szinnyei 1910: 71–72. 6) Plural of oblique cases. 7) Cf. Hung. *kez-e* “his hand” : *kez-e-i* “his hands”. 8) Pronominal. 9) Attributive *ket* : absolutive *kettő* ~ Mansi *kitiy*, E.Khanty *kätkən* “2” with the dual suffix.

Table 18.

case	IE (*)	Afroasiatic	Kartvelian	Uralic	Altaic	Dravidian
	-∅			∅	∅	∅
sg. nom.-voc.	<i>i-</i> & <i>u-</i> stems	* <i>i</i> & * <i>u</i>	* <i>i</i>			* <i>i</i>
	- <i>s</i> act.					
	- <i>om</i> inact.					* <i>m</i> / <i>n</i> inact.
acc. < term.	- <i>m</i>	?loc.-dir. *-(<i>u</i>) <i>ma</i>		acc. * <i>-m</i>	*(-) <i>be</i>	* <i>am</i> / <i>-an</i>
gen.-abl.	-(<i>o</i>) <i>s</i>	?dat.-term.- dir. * <i>-is(a)</i>	gen. * <i>-is₁</i>			
gen. (<i>o</i> -stems)	- <i>i</i>	gen. * <i>-i</i>				gen. * <i>-i</i> (Gond.)
abl. (<i>o</i> -stems)	-(<i>o</i>) <i>e</i> / <i>t</i> / <i>d(i)</i>	loc.-abl.-instr.- com. * <i>-ittV</i>	instr. * <i>-it</i>	loc. * <i>-tti</i>		gen. obl. * <i>-(t)tu</i>
dat.-loc.	-(<i>e</i>) <i>i</i>	dat.-loc.-abl. * <i>-iy</i>		lative * <i>-j</i>		?acc. * <i>-ay</i>
loc. heter.	- <i>en</i> ¹	instr.-dir.-loc.- abl. * <i>-na/i</i>		loc. * <i>-na</i> / <i>*-nä</i>	dat.-instr. * <i>-nV</i>	loc. * <i>-in</i> / * <i>-il</i> instr. * <i>-ān</i>
advr.	- <i>d^hi</i> , - <i>d^he²</i>	loc.-com. * <i>-di</i>	advr. * <i>-ad</i>	abl. * <i>-ḍa</i> / <i>*-ḍä</i>	dat.-loc. * <i>-da</i> / * <i>-du</i>	soc. * <i>-oṭu</i>
instr.	- <i>H₁</i>	dat.-instr. * <i>-?i</i>				instr. * <i>-ē</i> (Gond.)
pl. nom.	- <i>es</i>				* <i>-s-</i>	
- " - <i>o</i> -stems	- <i>oĵ</i>	du. * <i>-ay-</i>				
	- <i>oĵ-</i> obl.	* <i>-ay⁴</i>		pl. obl. * <i>-j⁴</i>		
pl. nom.-acc.	- <i>H₂</i> inact.	pl.-coll. * <i>-āh³</i>				pl. ntr. * <i>-a</i>
acc.	- <i>ms</i> < * <i>-s-m</i> ?					
gen.	- <i>om</i>		?Ge. erg. - <i>ma</i>			
dat.	- <i>mus</i>	?loc.-dir. *-(<i>u</i>) <i>ma</i>				
loc.	- <i>su</i> ~ - <i>si</i>	?dat.-term.- dir. * <i>-is(a)</i>	dat. * <i>-s</i> or Svan instr. -šw < * <i>-s₁wV</i> ?	lative * <i>-s</i>		
instr.	- <i>b^hi</i>	loc.-abl.-dir. *(-) <i>ba</i> / *(-) <i>bi</i>				

Notes: 1) Base of the oblique cases of heteroclitics, originally probably locatives of the type of Vedic *udān* “in water”. 2) Greek οἶκοι “at home”, οὐρανόθι “in heavens”, πόθι “where?”, ἐνθεν “from here”, further Vedic *kúha* “where?”, Old Avestan *kudā*, Old Church Slavonic *куде* id., *суде* “here” etc. (Brugmann 1911: 167, 728). 3) Semitic plural of biradical substantives: Hebrew *šēmāhōt* “she-servants” = Syrian *ʔamhātā*; Syrian *šēmāhē* “names”, *ʔabāhē* & *ʔabāhātā* “fathers” = Arabic *ʔabahāt*; Arabic *sitāh* “backs”, *šifāh* “lips”, *šiyāh* “sheep” (pl.), *miyāh* “waters”; further *ʔilāh* “god” (pl.) = Hebrew *ʔelōhīm*, Syriac *ʔallāhā* (Brockelmann 1908: 455). Cushitic plural in **-a* is probably of the same origin. 4) Illič-Svityč 1971: 285–86: IE+AA+Uralic; Kortlandt 2002: 217: IE. **-oĵ-*+ Uralic **-j-*.

Conclusion. Aharon Dolgopolsky (2005: 35) used to wonder if the original grammatical structure of Nostratic was synthetic or analytic. The present analysis of the Indo-European nominal inflection in Nostratic context confirms his preference of the analytic structure, with regard to the fact that most of the Indo-European case endings are derivable from various deictic or adverbial particles, some on the Indo-European level (usually with Nostratic roots), e.g. loc. sg. in *-en-* (Skt. *udán*) vs. **H₁en-* “in”, others on the Nostratic level at least, e.g. loc. pl. in **-su* vs. Kartvelian **šurwa-* “in the middle” or Central Cushitic **šaw-* “heart” (Dolgopolsky 2005: 17–19).

Abbreviations

A. — Aroid, ab(l). — ablative, abs. — absolute, adj. — adjective, adv. — adverb, advr. — adverbialis, Aeol. — Aeolic, acc. — accusative, Akk. — Akkadian, Alb. — Albanian, act. — active, Ar. — Arabic, Aram. — Aramaic, Arg. — Argive, arch. — archaic, Arc. — Arcadian, Arm. — Armenian, Att. — Attic, Av. — Avestan, Balt. — Baltic, BFin. — Balto-Finnic, Bret. — Breton, BSl. — Balto-Slavic, c. — genus communis, Celt. — Celtic, Clb. — Celtiberian, Cl. — Classic, col. — collective, com. — comitative, cons. — consonantic, Copt. — Coptic, Corn. — Cornish, Cret. — Cretan, Cush. — Cushitic, Cz. — Czech, Dah. — Dahalo, dat. — dative, det. — determined, dir. — directive, Dor. — Doric, du. — dual, E — East, Ebl. — Eblaite, Eng. — English, ep. — epigraphic, erg. — ergative, Est. — Estonian, f. — feminine, Fal. — Faliscan, Fi. — Finnish, fin. — finalis, FU. — Fenno-Ugric, Ge. — Georgian, gen. — genitive, Gl. — Gaulish, Gmc. — Germanic, Goid. — Goidelic, Gond. — Gondwan group, Goth. — Gothic, Gr. — Greek, H. — Hadiyya, HD — hysterodynamic, Hebr. — Hebrew, heter. — heteroclitic, Hitt. — Hittite, Hom. — Homeric, Hung. — Hungarian, id. — idem, II. — Indo-Iranian, inact. — inactive, indet. — indetermined, ins(tr). — instrumental, Ir. — Irish, It. — Italic, Kh. — Khanty, Lat. — Latin, Latv. — Latvian, Laz. — Lazish, Lep. — Leponitic, Lesb. — Lesbian, Lith. — Lithuanian, loc. — locative, Luw. — Luwian, Lyd. — Lydian, Lyc. — Lycian, m. — masculine, M — Middle, Ma. — Mansi, Marr. — Marrucine, Mars. — Marsian, Megr. — Megrelian, Messap. — Messapic, Mord. — Mordvin, Myc. — Mycenaean, n. — neuter, N — North, nom. — nominative, O — Old, obl. — oblique, OCS. — Old Church Slavonic, OHG. — Old High German, ON. — Old Nordic (incl. Old Icelandic), ORun. — Old Runic, Osc. — Oscan, Pael. — Paelignian, PD — proterodynamic, Phryg. — Phrygian, Pic. — Picenian, p- — proto-, pl. — plural, Pr(us). — Prussian, pred. — predicative, Q. — Qatabanian, S — South, Samoy. — Samoyedic, sg. — singular, Skt. — Sanskrit, Sl. — Slavic, soc. — sociative, suf. — suffix, Svan. — svanský, Sx. — Saxon, Syr. — Syriac, term. — terminative, Toch. — Tocharian, tran. — transformative, Ugar. — Ugaritic, Ugr. — Ugric, Umb. — Umbrian, Ved. — Vedic, Ven. — Venetic, voc. — vocative, Vol. — Volscan, W — West, Y — Young, Ye. — Yemsa, Yu. — Yudeo-.

Literature

- ADAMS, Douglas Q. 1999. *A Dictionary of Tocharian B*. Amsterdam-Atlanta: Rodopi.
- ANDRONOV, Mixail S. 1978. *Sravnitel'naja grammatika dravidijskix jazykov*. Moskva: Nauka.
- ANTONSEN, Elmer. 1975. *A Concise Grammar of the Older Runic Inscriptions*. Tübingen: Niemeyer.
- BEEKES, Robert S.P. 1985. *The Origins of the Indo-European Nominal Inflection*. Innsbruck: IBS 46.
- BEEKES, Robert S.P. 1995. *Comparative Indo-European Linguistics. An Introduction*. Amsterdam — Philadelphia: Benjamins.
- BLAŽEK, Václav. 1999. *Numerals. Comparative-etymological analyses and their implications*. Brno: Masarykova univerzita.
- BLAŽEK, Václav. 2006. Traces of a common case system in Afroasiatic. In: *Loquentes linguis. Linguistic and Oriental Studies in Honour of Fabrizio A. Pennacchietti*, ed. by P.G. Borbone, A. Mengozzi, M. Tosco. Wiesbaden: Harrassowitz, 91–101.
- BLAŽEK, Václav. 2010. Indoevropská substantivní deklinace [Indo-European declension of substantives]. *Linguistica Brunensia* 58, 2010/1–2, 51–91.
- BLOCH, Jules. 1949. Préfixes et suffixes en Andaman. *Bulletin de la Société de Linguistique de Paris* 45, 1–46.
- BROCKELMANN, Carl. 1908. *Grundriss der vergleichenden Grammatik der semitischen Sprachen*, Bd. I: *Laut- und Formenlehre*. Berlin: Reuther & Reichard.

- BRUGMANN, Karl. 1906/1911. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*, II.1/2. Strassburg: Trübner.
- CÓLERA, Carlos Jordán. 2005. *Celtibérico*. Zaragoza: Gorfisa.
- COLLINDER, Björn. 1960. *Comparative Grammar of the Uralic Languages*. Stockholm: Almqvist & Wiksell.
- ČERMÁK, František. 1997. *Jazyk a jazykověda*. Praha: Pražská imaginace.
- ČERNÝ, Václav A. 1975. *Základy gruzínštiny*. Praha: Academia.
- DEŠERIEV, Ju. D. 1967. Čečenský jazyk. In: *Jazyki narodov SSSR: Iberijsko-kavkazskie jazyki*. Moskva: Nauka, 190–209.
- DEŠERIEV, Ju. D. 1967. Bachijský jazyk. In: *Jazyki narodov SSSR: Iberijsko-kavkazskie jazyki*. Moskva: Nauka, 228–245.
- DOLGOPOLSKY, Aron. 1991. Two problems of Semitic historical linguistics: I. Mimation and nunation; II. Akkadian reflexes of *ʕ and *ħ. In: *Semitic Studies in honor of Wolf Leslau on the occasion of his eighty-fifth birthday*, Vol. I, ed. by Alan S. Kaye. Wiesbaden: Harrassowitz, 328–339.
- DOLGOPOLSKY, Aron. 2005. Nostratic Grammar: Synthetic or Analytic? *Orientalia et Classica. Trudy Instituta Vostočnyx kul'tur i antičnosti*, vypusk VI: *Aspekty komparativistiki 1*. Moskva: Rossijskij Gosudarstvennyj Gumanitarnyj Universitet, 13–38.
- EDAL = STAROSTIN, Sergei, Anna DYBO, Oleg MUDRAK. 2003. *Etymological Dictionary of the Altaic Languages*. Leiden-Boston: Brill.
- ELIZARENKOVA, Taťjana Ja. 1987. *Vedijskij jazyk*. Moskva: Nauka.
- ERHART, Adolf. 1982. *Indoevropské jazyky. Srovnávací fonologie a morfologie*. Praha: Academia.
- ERHART, Adolf. 1993. *Die indogermanische Nominalflexion und ihre Genese*. Innsbruck: IBS 73.
- EWAI = MAYRHOFER, Manfred. 1986. *Etymologisches Wörterbuch des Altindoarischen*, I-III. Heidelberg: Winter.
- FÄHNRIK, Heinz. 2007. *Kartvelisches Etymologisches Wörterbuch*. Leiden-Boston: Brill.
- FISCHER, Wolfdietrich. 1972. *Grammatik des klassischen Arabisch*. Wiesbaden: Harrassowitz.
- GAMKRELIDZE, Tamas V. & IVANOV, Vjačeslav V. (1984). *Indoeuropejskij jazyk i indoeuropejcy*. Tbilisi: Izdatelstvo Tbilisskogo universiteta. In English: *Indo-European and the Indo-Europeans: a reconstruction and historical analysis*. Berlin: Mouton de Gruyter 1995.
- GHL = *A Grammar of Hittite Language*, Part 1: *Reference Grammar*, by Harry A. HOFFNER, jr. & H. Craig MELCHERT. Winona Lake: Eisenbrauns 2008.
- HAJDÚ, Péter. 1985. *Ural'skie jazyki i narody*, přeložil E. Xelimskij. Moskva: Progress.
- HAJNAL, Ivo. 1995. *Studien zum mykenischen Kasussystem*. Berlin — New York: Walter de Gruyter.
- HAJNAL, Ivo. 2003. methodische Vorbemerkungen zu einer Palaeolinguistik des Balkanraums. In: *Languages in Pre-historic Europe*, ed. by Alfred Bammesberger & Theo Vennemann. Heidelberg: Winter, 117–145.
- HAMP, Eric P.. 1979. Indo-European *g^wen-H_a. *Zeitschrift für vergleichende Sprachforschung* 93, 1–7.
- HAMP, Eric P. 1981. Indo-European *(H.)op-. *Münchener Studien zur Sprachwissenschaft* 40, 39–60.
- HAMP, Eric P. 1987. Varia. *Études Celtiques* 24, 185–189.
- HAMP, Eric P. 1994. The Laryngeal Heteroclitics. In: *Indogermanica et Caucasia. Festschrift für K.H. Schmidt zum 65. Geburtstag*, ed. Roland Bielmeier & Reinhard Stempel. Berlin — New York: Walter de Gruyter, 35–40.
- HAMP, Eric P. 1996. Varia II: 1. On the Old Irish dat. sg. in *mi. *Ériu* 47, 209.
- HAUDRY, Jean. 1982. *Prehistoire de la flexion nominale indo-européenne*. Lyon: Institut d'études indo-européennes de l'Université Jean Moulin (Lyon III).
- HEG = *Hittite Etymological Dictionary*, by Jaan PUHVEL. Berlin — New York: Mouton de Gruyter 1984n.
- HOFFMANN, Karl & FORSSMAN, Bernhard. 1996. *Avestische Laut- und Flexionslehre*. Innsbruck: IBS 84.
- HOFÍRKOVÁ, Lucie & BLAŽEK, Václav. 2012. Jmenná deklinace fino-sámských jazyků v ugrofinské perspektivě [Nominal Declension of the Fenno-Saamic Languages in the Fenno-Ugric Perspective]. *Linguistica Brunensia* 60, 127–146.
- HŘEBÍČEK, Luděk. 1969. *Turečtina*. Praha: Academia.
- ILLIČ-SVITYČ, Vladislav M. 1971–76–84. *Opyt sravnenija nostratičeskix jazykov*, I-III. Moskva: Nauka.
- KLIMOV, Georgij A. 1962. *Sklonenie v kartvel'skix jazykax v sravnitel'no-istoričeskom aspekte*. Moskva: AN SSSR.
- KLINGENSCHMITT, Gerd. 1992. Die lateinische Nominalflexion. In: *Latein und Indogermanisch. Akten des Kolloquiums der Indogermanischen Gesellschaft* (Salzburg, Sept. 1986), ed. Oswald Panagl & Thomas Krisch. Innsbruck: IBS 64, 89–135.
- KLINGENSCHMITT, Gerd. 1994. Das Tocharische in indogermanischer Sicht. In: *Tocharisch (Akten des Fachtagung der Indogermanischen Gesellschaft, berlin, sept 1990)*, ed. B. Schlerath. Reykjavík: TIES Suppl. Series, Vol. 4), 310–411.

- KORTLANDT, Frederik. 1978. On the history of the genitive plural in Slavic, Baltic, Germanic, and Indo-European. *Lingua* 45, 281–300.
- KORTLANDT, Frederik. 1984/2003. Proto-Armenian case endings. In: *International symposium on Armenian linguistics*. Erevan: Akademija nauk Armjanskoj SSR, 97–106 / *Armeniaca*. Ann Arbor: Caravan Books, 45–56.
- KORTLANDT, Frederik. 2002. The Indo-Uralic Verb. In: *Finno-Ugrians and Indo-Europeans: Linguistic and Literary Contacts. Proceedings of the Symposium at the University of Groningen* (Nov. 2001). Maastricht: Shaker (Studia Fenno-Ugrica Groningana 2), 217–227.
- KRAHE, Hans. 1948. *Historische Laut- und Formenlehre*. Heidelberg: Winter.
- KRASUXIN, Konstantin G. 2004. *Vvedenie v indoevropejskoe jazykoznanie*. Moskva: Akademija.
- KUČEROVÁ, Hana. 1962. *Česko-finské rozhovory*. Praha: SPN.
- KULIKOV, Leonid. 2009. Evolution of Case Systems. In: *Oxford Handbook of Case*, ed. by Andrej Malchukov & Andrew Spencer, 439–457.
- LAMBERT, Pierre-Yves. 2001[1998–2000]. La tuile de Châteaubateau (Seine-et-Marne). *Études celtiques* 34, 57–115.
- LEIA = *Lexique étymologique de l'irlandais ancien*, par Joseph VENDRYES & Pierre-Yves LAMBERT. Dublin: Institute for Advanced Studies — Paris: CNRS Éditions 1959–96.
- LEJEUNE, Michel. 1971. Lepontica. *Études celtiques* 12, 356–500.
- LG = *La langue gauloise₂*, par Pierre-Yves LAMBERT. 2003. Paris: Errance.
- LIPÍŃSKI, Edward. 1997. *Semitic Languages. Outline of a comparative Grammar*. Leuven: Peeters (Orientalia Lovanensia Analecta 80).
- MACDONELL, Arthur A. 1916[2000]. *A Vedic Grammar for Students*. Delhi: Motilal Banarsidass.
- MACHEK, Václav. 1968. *Etymologický slovník jazyka českého*. Praha: Academia.
- MAYRHOFER, Manfred. 1986. *Indogermanische Grammatik I.2.: Lautlehre [Segmentale Phonologie des Indogermanischen]*. Heidelberg: Winter.
- MAŽIULIS, Vytautas. 2004. *Prūsų kalbos istorinė gramatika*. Vilnius: Vilniaus universiteto leidykla.
- MERIGGI, Piero. 1980. *Schizzo gramaticale dell'Anatolico*. Roma: *Atti della Accademia nazionale dei Lincei, Memorie: Classe di Scienze morali, storiche e filologiche, Serie VIII, Volume XXIV, Fasc. 3*.
- MLH = *Monumenta Linguarum Hispanicarum IV: Die tartessischen, keltiberischen und lusitanischen Inschriften*, von Dagmar WODTKO & Jürgen UNTERMANN.
- MLM = *Monumenta Linguae Messapicae, I–II*, Carlo de SIMONE & Simona MARCHESINI. Wiesbaden: Reichert, 2002.
- MOSCATI, Sabatino et al. 1964. *An Introduction to the Comparative Grammar of Semitic Languages*. Wiesbaden: Harrassowitz.
- NIKOLAEV, Alexander. 2000. PIE ergativity and the Genitive in *osyo. In: *Proceedings of the Eleventh Annual UCLA Indo-European Conference*, ed. K. Jones-Bley, M.E. Huld & A. Della Volpe. Washington D.C.: Institute for the Study of Man (*Journal of Indo-European Studies Monograph Series No. 35*), 293–309.
- NIL = *Nomina im indogermanischen Lexikon*, von Dagmar S. WODTKO, Britta IRLINGER & Carolin SCHNEIDER. Heidelberg: Winter 2008.
- OLSEN, Birgit A. 1999. *The Noun in Biblical Armenian. Origin and Word-Formation*. Berlin-New York: Mouton de Gruyter.
- OTRĘBSKI, Jan. 1956. *Gramatyka języka litewskiego, Tom III: Nauka o formach*. Warszawa: Państwowe wydawnictwo naukowe.
- PEDERSEN, Holger. 1907. Neues und nachträgliches. *KZ* 40, 129–217.
- PEDERSEN, Holger. 1926. *La cinquième déclinaison latine*. København: Høst & Søn.
- PINAULT, Georges-Jean. 1989. *Introduction au tokharien*. Paris: LALIES (Actes des sessions de linguistique et de littérature 7).
- POKORNY, Julius. 1925. *Altirische Grammatik*. Berlin — Leipzig: Walter de Gruyter.
- PORTMAN, M.V. 1887. *A Manual of the Andamanese Languages*. London: Allen.
- RASMUSSEN, Jens E. 1987. Aspirated tenues and Indo-European phonation series. In: *Acta Linguistica Hafniensia* 20, 81–109.
- RASMUSSEN, Jens E. 1989. *Studien zur Morphophonemik der indogermanischen Grundsprache*. Innsbruck: IBS 55.
- RASMUSSEN, Jens E. 1999. *Selected Papers on Indo-European Linguistics, I–II*. Copenhagen: Museum Tusulanum — University of Copenhagen.
- RINGE, Don. 2006. *From Proto-Indo-European to Proto-Germanic*. Oxford: University Press.

- RIX, Helmut. 1976. *Historische Grammatik des Griechischen. Laut- und Formenlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- RŮŽIČKA, Karel F. 1970. *Úvod do swahilštiny*. Praha: Academia.
- SAVČENKO, Aleksej N. 1974. *Sravnitel'naja grammatika indoevropskix jazykov*. Moskva: Vysšaja škola.
- SCHINDLER, Jochem. 1975. L'apophonie des thèmes indo-européens en *-r/n*. *Bulletin de la Société de Linguistique de Paris* 70, 1–10.
- SCHMALSTIEG, William R. 1997. The Origin of the Neuter Nominative-Accusative Singular In **-OM*. *Journal of Indo-European Studies* 25, 401–407.
- SCHMITT, Rüdiger. 1981. *Grammatik des Klassisch-Armenischen mit Sprachvergleichenden Erläuterungen*. Innsbruck: IBS 32.
- SIHLER, Andrew L. 1995. *New Comparative Grammar of Greek and Latin*. New York — Oxford: Oxford University Press.
- SKALIČKA, Vladimír. 1945[2004] Über die Typologie der Bantusprachen. Citováno podle překladu Víta Dovalila: O typologii bantuských jazyků. In: Vladimír Skalička — *Souborné dílo I*, ed. František Čermák, Jan Čermák, Petr Čermák, Claudio Poeta. Praha: Karolinum, 284–315.
- STREITBERG, Wilhelm. 1920. *Gotisches Elementarbuch*. Heidelberg: Winter.
- SZEMERÉNYI, Oswald. 1985. Strukturelle Probleme der indogermanischen Flexion. Prinzipien und Modellfälle. In: *Grammatische Kategorien. Funktion und Geschichte. Akten der VII. Fachtagung der Indogermanischen Gesellschaft* (Berlin, Februar 1983), hrsg. Bernfried Schlerath & V. Rittner. Wiesbaden: Reichert, 515–533.
- SZEMERÉNYI, Oswald J.L. 1996. *Introduction to Indo-European Linguistics*. Oxford: Clarendon Press.
- SZINNYEI, Josef. 1910. *Finnisch-ugrische Sprachwissenschaft*. Leipzig: Göschen.
- THURNEISEN, Rudolf. 1946. *A Grammar of Old Irish*. Dublin: Institute for Advanced Studies.
- TICHY, Eva. 2004. *Indogermanistisches Grundwissen für Studierende sprachwissenschaftlicher Disziplinen₂*. Bremen: Hempen (1st edition 1999).
- TREMBLAY, Xavier. 2004. Die Ablautstufe des Lokativs der akrostatischen Nomina. In: *Per Aspera ad Asteriscos. Studia Indogermanica in honorem Jens E. Rasmussen*, ed. by Adam Hyllested et al. Innsbruck: IBS 112, 573–590.
- UB = URBANOVÁ, Daniela & BLAŽEK, Václav. 2008. *Národy starověké Itálie, jejich jazyky a písmo*. Brno: Host.
- VAVROUŠEK, Petr. 2008. *Rekonstrukce? Rekonstrukce!* Praha: Ústav srovnávací jazykovědy Filozofické fakulty Univerzity Karlovy.
- de VAAN, Michiel. 2008. *Etymological Dictionary of Latin and the other Italic Languages*. Leiden-Boston: Brill.
- WATKINS, Calvert. 1997. Luvo-Hittite *:lapan(a)-*. In: *Studies in Honor of Jaan Puhvel*, Part I, ed. D. Disterheft et al. Washington D.C., 29–35.
- ZVELEBIL, Kamil V. 1990. *Dravidian Linguistics. An Introduction*. Pondicherry: Institute of Linguistics and Culture.

В. БЛАЖЕК. Индоевропейское именное склонение в ностратической перспективе.

В статье обобщены текущие представления об истории и происхождении системы именного склонения в индоевропейских языках, а также вкратце сравниваются несколько гипотез о причинах возникновения т. н. «тематического» склонения. Реконструированная парадигматическая система далее сравнивается с соответствующими системами, восстановленными для других ветвей гипотетической ностратической макросемьи — семитской (и, шире, афроазиатской), картвельской, уральской. Автор заключает, что, поскольку большинство падежных окончаний праиндоевропейского языка этимологизируются либо на основе самостоятельных индоевропейских лексем, либо на внешнем уровне (при сравнении с другими ностратическими языками), это можно расценивать как серьезный аргумент в пользу аналитического характера праностратического языка.

Ключевые слова: Историческое языкознание, индоевропейские языки, сравнительная морфология, ностратическая гипотеза.

Корпусные методы дешифровки анатолийских иероглифов*

Анатолийские иероглифы использовались в Малой Азии и Сирии для передачи преимущественно лувийских текстов в XIV—VIII вв. до н. э. Относительно скромный корпус иероглифических надписей из Анатолии в сочетании с малочисленностью билингв является причиной незавершенности дешифровки данной письменности. Тем не менее, значительный прогресс в дешифровке анатолийских иероглифов был достигнут после публикации первых томов Корпуса лувийских иероглифических надписей в самом конце прошлого тысячелетия. Многие недавние открытия в данной области основаны на применении комбинаторных методов и никак не зависят от публикаций новых текстов. В данном обзоре мы подвергаем подробному анализу три группы открытий: во-первых, разграничение между идеограммами, фонетическими комплементами и фонетическими индикаторами, во-вторых, фонетическая интерпретация идеограмм и, в-третьих, уточнение фонетических значений силлабограмм. Авторы постарались показать, что продолжающаяся работа по интерпретации анатолийских иероглифов также имеет определенное значение для индоевропеистики в целом.

Ключевые слова: анатолийские иероглифы, лувийский язык, дешифровка, корпусная лингвистика.

1. Методы дешифровки древних письменностей

Криптография, или теория вторичного языкового кодирования, является широко распространенным направлением прикладной математики, традиционно востребованным разведчиками и дипломатами, а в последнее время активно задействованным в целях обеспечения безопасности компьютерных систем передачи информации. Формальные алгоритмы дешифровки непрерывно усложнялись по мере того, как совершенствовались криптографические алгоритмы. На фоне этого развития следует констатировать ограниченное количество приемов, имевших решающее значение при дешифровке первичных систем кодировки (т. е. древних письменностей). По существу, эти приемы можно разделить на три категории: использование билингв и квазибилингв, угадывание лексем или формул, постулируемых для определенных культур, и идентификация генетической принадлежности языка с дальнейшим применением сравнительного метода.

Наиболее известным примером дешифровки с опорой на двуязычный текст является, несомненно, использование Розеттского камня для определения структуры египетской иероглифики [Daniels 1995: 84]. Не менее важна роль дешифрованной ранее древнеперсидской версии Бехистунской надписи для определения содержания вавилонской и эламской версий соответствующего текста, которая, в свою очередь, привела к определению структуры различных вариантов месопотамской клинописи [Daniels 1995: 85—87].

* Настоящее исследование выполнено при поддержке гранта Президиума РАН «Полный аннотированный корпус лувийских текстов» в рамках программы фундаментальных исследований Президиума РАН «Корпусная лингвистика».

Последним успешным примером дешифровки с опорой на квазибилингвы является фонетическая интерпретация карийского алфавита с использованием древнеегипетских и карийских надписей, найденных на одних и тех же объектах. Интересно, что окончательное подтверждение данной дешифровки оказалось возможным в результате обнаружения двуязычной карийской и греческой проксении в девяностых годах прошлого века [Adiego 2007: 166—204].

Классическим примером использования языковых формул для целей дешифровки является идентификация фрагментов надписей, содержащих стереотипную ахеменидскую титулатуру «X, царь великий, царь царей, сын (царя) Y». Шаблон данной формулы был уже известен из дешифрованных ранее сасанидских надписей на среднеперсидском языке. Однако Георгу Фридриху Гротефенду удалось в самом начале девятнадцатого столетия обнаружить ее эквивалент в древнеперсидских клинописных текстах. Решение последующей комбинаторной задачи привело к идентификации цепочек, соответствующих именам отдельных ахеменидских царей, известных из древнегреческих источников. В свою очередь, сравнение царских имен друг с другом позволило Гротефенду предположить фонетические значения для ряда символов древнеперсидской клинописи [Kent 1953: 10]. Впоследствии эти чтения были уточнены в результате идентификации более пространной формулы, а именно списка древнеперсидских сатрапий, также в основном известных по трудам Геродота и других древнегреческих авторов [Daniels 1995: 83b].

Вместе с тем анализ Гротефенда является далеко не единственным примером дешифровки посредством угадывания культурного контекста. Хотя микенское «линейное письмо Б» часто рассматривается как классический пример первичного языкового кода, дешифрованного исключительно формальными методами, в действительности дело обстоит сложнее. Первичные наблюдения Майкла Вентриса и его предшественников действительно носили формальный характер и привели к классификации микенских силлабограмм в форме таблиц («решеток») с фонетически идентичными инициалами и финалями. Однако интерпретации «решеток» с самого начала способствовало генетически обусловленное сходство ряда знаков линейного письма Б и уже дешифрованного кипрского слогового письма. Решающий же прорыв в определении фонетических значений силлабограмм последовал за гипотетической идентификацией критских топонимов Кносс и Амнис. Согласно первоначальной догадке Вентриса, эти топонимы просто не могли не быть упомянуты в критских хозяйственных документах. Фактом же, подтвердившим дешифровку Вентриса в глазах классических филологов всего мира, стало обнаружение новых табличек из Пилоса, в которых греческие композиты, указывающие на сосуды с различным количеством ножек, сопровождалась детерминативами с изображениями соответствующих сосудов [Chadwick 1967].

Применение сравнительного метода редко играло решающую роль на начальных этапах дешифровки, но часто приводило к ее успешному завершению и способствовало интерпретации соответствующих языковых систем. В частности, идентификация языка линейного письма Б как греческого диалекта не только дала ключ к пониманию большинства текстов, но и позволила предложить более точные значения для многих силлабограмм, например для сочетаний с лабиовелярными [Chadwick 1967]. Лабиовелярные согласные отсутствуют в других греческих диалектах, но были реконструированы для общегреческого еще до дешифровки линейного письма Б. Прогресс в чтении древнеперсидских текстов за пределами стереотипных формул был достигнут в результате систематического сравнения древнеперсидского языка с авестийским и санскритом [Daniels 1995: 83b]. Сравнение между аккадским и другими семитскими языками, в первую оче-

редь с древнееврейским, помогло ученым в сборе и понимании парадигм аккадского глагола, содержащих внутреннюю флексию [Daniels 1995: 86a].

Роль языкового родства для понимания первичных систем кодировки можно проиллюстрировать следующим негативным соображением: пока не известно ни одного примера успешной дешифровки, который привел бы к идентификации и описанию нового языка-изолята. Разумеется, в случае шумерского, эламского или хаттского языков не приходится говорить о дешифровке как таковой, поскольку для их фиксации использовались варианты месопотамской клинописи, а описание их грамматических структур было в основном произведено с опорой на двуязычные источники. Подобные билингвы отсутствуют, например, в случае линейного письма А, что делает пока невозможным понимание соответствующих текстов, несмотря на то что структура линейного письма А в целом экстраполируется из линейного письма Б. Рискнем, однако, предположить, что если бы линейное письмо А использовалось для фиксации индоевропейского или семитского языка, его грамматика уже была бы частично описана, несмотря на отсутствие билингв.

Вероятно, именно тот факт, что все древние системы письма с обширными корпусами текстов случайным образом оказались связанными с большими языковыми семьями низкого уровня (семитской, индоевропейской, сино-тибетской, майя) или, по крайней мере, с языками, чьи потомки засвидетельствованы в иной традиции (древнеегипетский > коптский), сыграл свою роль в истории дешифровки первичных систем кодировки. Многие любители искренне верят, что дешифровать, например, письмо Фестского диска или этеокипрских документов означает идентифицировать его с одним из описанных на сегодняшний день языков (древнегреческим, ликийским, хурритским и т. д.) или, по крайней мере, доказать его принадлежность к одной из широко известных языковых семей (как правило, индоевропейской или семитской).

Как известно, многочисленные попытки дешифровать Фестский диск сравнительным методом потерпели фиаско. С другой стороны, применение формальных алгоритмов дешифровки к интерпретации сверхмалых письменных корпусов, подобных корпусу Фестского диска, также оказывается крайне затруднительным. К сожалению, археология пока не предоставила нам достаточно материала для решения идеальной задачи по контекстно-независимой дешифровке неизвестного языка-изолята. Существуют, однако, промежуточные случаи, когда дешифровка более обширных корпусов с уже установленной языковой принадлежностью может быть уточнена комбинаторными методами.

Темой настоящей статьи является применение методов корпусной лингвистики в целях уточнения чтений лувийских иероглифических текстов. Мы посвящаем ее памяти Сергея Анатольевича Старостина, чье многогранное научное наследие включает в себя фундаментальное исследование по фонетической интерпретации древнекитайских иероглифов [Старостин 1989].

2. История дешифровки анатолийских иероглифов

Анатолийские иероглифы представляют собой оригинальную слоговую систему письма, употреблявшуюся на территории Малой Азии и северной Сирии в период приблизительно с 1500 по 700 г. до н. э. для записи, как правило, лувийских текстов. Суммарный инвентарь анатолийских иероглифов, известный на сегодняшний день, насчитывает более 500 знаков. Наиболее частотны слоговые знаки вида CV, знаки CVCV использовались

реже, также употреблялись идеограммы и детерминативы. Таким образом, мы имеем дело с иерархической системой письма, типологически сходной в данном отношении с египетскими иероглифами или месопотамской клинописью. Важное отличие анатолийских иероглифов от двух вышеупомянутых систем — это существенно меньший размер корпуса (около 20 тыс. слов).

Разумеется, подобный корпус не может отражать реального уровня письменной активности в хеттском царстве и постхеттских княжествах. Скорее всего, его скромные размеры вызваны утратой вероятных основных носителей анатолийских иероглифов: навощенных деревянных дощечек. Но каковы бы ни были причины данной ситуации, она объективно приводит к тому, что компетенция современных исследователей данной системы письма не является адекватной компетенции лувийских писцов. Поэтому дешифровка анатолийских иероглифов исторически носила постепенный характер и не является вполне завершённой даже спустя 150 лет после первых шагов в данном направлении.

Первые известия об анатолийском иероглифическом письме были получены европейцами после экспедиции 1861 г. французского археолога Жоржа Перро в Малую Азию. Перро сфотографировал наскальную надпись, обнаруженную им недалеко от Богазкёя, и опубликовал ее в работе 1862 г. [Perrot 1862]. Десятью годами позже Ричардом Бертоном и Тирвитом Дрейком были опубликованы надписи на камнях из Хамы [Burton, Drake 1872]. Анализ этих надписей был проделан Гайдом Кларком, который, используя чисто статистический метод, продемонстрировал, что мы имеем дело с письмом, а не с орнаментом. Кларк выделил 59 знаков и показал их распределение по тексту: самый частотный знак встречался 27 раз, далее знаки встречались 26, 24, 21, 15, 11, 9 раз и т. д., при этом 17 знаков встретились в тексте лишь однажды. Кларк полагал, что письмо было алфавитным, однако он допускал, что некоторые знаки являются детерминативами или служат для целей пунктуации.

В 1876 году оксфордский исследователь преподобный Арчибальд Сейс подробно рассмотрел все надписи из Хамы и заключил, что письмо было слоговым, но с идеографическими элементами; его вывод базировался на длине вычлененных им отдельных слов, а также на сравнении с недавно расшифрованным кипрским слоговым письмом [Sayce 1876]. По мнению Сейса, вначале рассматриваемое письмо было идеографическим, однако со временем оно эволюционировало в слоговое. В докладе 1880 г., прочитанном в «Обществе библейской археологии», Сейс предположил, что обнаруженное письмо принадлежало «великой хеттской расе», известной из греческих и библейских источников [Sayce 1880a]. С тех пор обозначение «хеттское иероглифическое письмо» стало общепринятым среди анатолистов, и оно просуществовало в этом статусе вплоть до выхода работы Хокинса, Морпурго-Дэвис и Нойманна в 1974 г. [Hawkins et al. 1974].

В докладе 1880 г. Сейс выразил сожаление по поводу отсутствия билингв, столь необходимых для дешифровки неизвестного письма. Однако уже в том же году в одном из немецких журналов была опубликована серебряная печать, купленная на рынке Смирны. Надпись на печати была выполнена с помощью клинописи и иероглифики. Печать привлекла внимание Сейса, который дал ей название «Печать Таркондемоса» и окрестил ее «Розеттским камнем хеттской дешифровки» [Sayce 1880b]. В клинописной части Сейс прочитал имя и титул *Tar-rik-tim-me šar mat Er-me-e* «Таркондемос, царь страны Эрмы» (имя «Таркондемос» было известно из греческих источников). В иероглифической части Сейсу удалось правильно прочитать логограммы для «царя» (*17 = REX) и «страны» (*228 = REGIO). Хотя роль данной билингвы для дешифровки анатолийских иерог-

лифов не сопоставима с ролью Розеттского камня для египетской дешифровки, все же с ее помощью удалось сделать первый прорыв в данной области. Стоит отметить, что личное имя «Таркондемос» было правильно прочитано как Таргаснава лишь в самом конце XX в. [Hawkins, Morpurgo-Davies 1998].

Изучение анатолийских иероглифов в начале XX века характеризовалось быстрым накоплением материала. В период с 1900 по 1906 гг. выходит первое собрание иероглифических надписей (около 40 больших текстов) [Messerschmidt 1900—1906]. Большое число новых текстов было обнаружено при раскопках Каркемиша, проводившихся Британским музеем в 1911—1914 гг. Открытие богазкейского клинописного архива экспедицией под руководством Гуго Винклера (1906) и установление индоевропейского характера хеттского языка Бедржихом Грозным (1915) также способствовали увеличению потенциального материала для сопоставительного анализа. При этом следует отметить, что клинописные таблички из Богазкёя также содержали значительное количество лувийских заклинаний, однако особая близость языка анатолийских иероглифов к языку клинописных лувийских текстов была осознана лишь к 30-м гг. XX века.

В 30-е гг. XX века были также установлены значения многих фонетических знаков анатолийского иероглифического письма. Как это произошло впоследствии с линейным письмом Б, решающую роль здесь сыграло угадывание культурного контекста, т. е. идентификация топонимов и личных имен, которые, как предполагалось, должны встречаться в иероглифических надписях из определенных областей. Вот только некоторые примеры соответствий между именами собственными в современной иероглифической и клинописной передаче: лув. <tu-wa/i-na-wa/i-ni-sa(URBS)> «относящийся к Тиане» vs. хетт. *Tiwanuwa* «Тиана»; лув. <ku+ra/i-ku-ma-wa/i-ni-i-sa(URBS)> «относящийся к Гургуму» vs. ассир. *Gurgume*; лув. <i-ma-tu-wa/i-ni(URBS)> «относящийся к Хаме» vs. ассир. *Hammatti*; лув. <mu-wa/i-ta-li-si-sa> «принадлежащий Муваталли» vs. ассир. *Muttallu*; <tu+ra/i-hi-li-na> «Урахилина» vs. ассир. *Irhuleni*; <wa/i+ra/i-pa-la-wa/i-sa> «Варпалава» vs. ассир. *Urballa*. Результаты данного этапа дешифровки нашли выражение в работах Мериджи [Meriggi 1934], Гельба [Gelb 1931—1942], Боссерта [Bossert 1932] и Форрера [Forrer 1932]. К началу Второй мировой войны совместными усилиями этих исследователей удалось правильно определить фонетические значения 55 знаков, а также прочесть многие логограммы и детерминативы.

О результатах дешифровки в данный период можно составить представление на основе Таблицы 1. Очевидно, использование комбинаторного подхода с опорой на ограниченное количество личных имен дало неоднозначные результаты. Несмотря на достигнутый прогресс, многие важные чтения по-прежнему оставались неверными, и в некоторых случаях по принципиальным вопросам отсутствовал консенсус.

Дальнейшее продвижение в дешифровке анатолийских иероглифов стало возможным благодаря открытию в Восточной Киликии в 1946 г. билингвы KARATEPE. Этот памятник содержал три версии текста на финикийском языке и две версии, записанные анатолийскими иероглифами. Первооткрывателем памятника Т. Боссертом он былнесен к концу VIII века до н. э. Надписи были составлены от имени местного князя Ацативады, и в них сообщалось о том, что Ацативада является основателем крепости и творцом многочисленных благ для ее жителей. Как было показано в исследованиях Барнетта [Barnett 1953], Фридриха [Friedrich 1953], Гельба [Gelb 1950], Альпа [Alp 1950], Гютербока [Güterbock 1949] и Мериджи [Meriggi 1951], билингва KARATEPE дала подтверждение многим фонетическим и семантическим интерпретациям, считавшимся до этого гипотетическими. С помощью этой билингвы удалось установить чтения нескольких знаков, а также значения нескольких лексем и логограмм.

Таблица 1. Эволюция чтений анатолийских иероглифов (с сокр. из [Pore 1999: 140]).

№ знака	До 1931 года	Форрер 1931/1932	Гельб 1931	Гельб 1935	Гельб 1942	Ларош 1960	Hawkins 2000
*209	a? (Коули, 1917) a (Мериджи, 1929)	a	a			a/e	i
*210			ā	ā			ia
*450			e				a
*376	i (Пайзер, 1892)	i	wa	i		i	zi
*215		ha		ha		ha/he	ha
*413	hi (Томсон, 1913)		hi			hi	hi
*307				hu		hu	hu
*434	ka (Йенсен, 1894; Сейс, 1903)	ka	ka			ka	ka
*446				r	ri	ki	ki
*423	ku (Томсон, 1912)	ku	ku			ku	ku
*175				la		la	la
*278	li (Томсон, 1912)	li	li			li	li
*110	ma (Томсон, 1912)	ma	ma			ma	ma
*107	mu (Коули, 1917)	mu		mu		mu	mu
*35	n (Сейс, 1893)	na		na		na	na
*334	p (Коули, 1917)		pi	pa		pa	pa
*66			ne	pi		pi	pi
*328					pu	pu	pu
*415	es (Йенсен, 1894)		si			sa/s	sa
*104	s (Томсон, 1912)		sa		se	sà	sà
*370			mi	lu	su	su	su
*41					za	tà	tà
*29			ta		ti	tá	tá
*100		ta	ta		te	ta ₄	ta
*319			tu	ki	zi	te/ti	ta ₄
*439	u/w (Коули, 1920)	wa	i	wa		wa/wi	wa/i

Следующий этап в изучении лувийского языка характеризуется последовательным применением сравнительного метода. Здесь нужно, прежде всего, отметить работы французского исследователя Эммануэля Лароша [Laroche 1958, 1960a, 1967]. Признавая значения хеттского языка для дешифровки «хеттских иероглифов», Ларош вместе с тем подчеркивал, что ближайшими родственниками последнего являются язык клинописных лувийских заклинаний, а также ликийский язык, сохранившийся в алфавитной передаче. Таким образом, Ларош имплицитно исходил из существования лувической языковой подгруппы, хотя сам он никогда не употреблял данного термина. Систематически сравнивая отдельные подсистемы «клинописного лувийского», «иероглифического лувийского» и ликийского языков, такие как ваккернагелевские клитики, пространственные наречия и т. д., Ларош предоставил еще одно независимое подтверждение предшест-

вующих результатов дешифровки анатолийских иероглифов, а также предложил ряд новых интерпретаций.

В начале 1960-х гг. вышли два фундаментальных компендиума, посвященных анатолийским иероглифам, в которых суммировались результаты исследований на тот период. Работа Эммануэля Лароша [Laroche 1960b] включала перечень из 497 знаков, для которых (где это возможно) были определены фонетические чтения, а также представлена краткая информация об их открытии. В книгу Лароша были также включены результаты его собственных исследований. Вышедшая в 1962 г. работа Мериджи [Meriggi 1962] представляла собой усовершенствованную версию его глоссария 1934 г. Для каждой лексемы глоссария Мериджи указал фонетическое чтение, а также произвел подробный грамматический анализ. Поскольку чтения, представленные в работах Лароша и Мериджи в целом совпадали, работа по дешифровке казалась завершённой.

Дальнейшие исследования, однако, показали, что представление о языке, стоявшем за анатолийскими иероглифами, было еще довольно смутным. Проблемы, которые стояли перед исследователями в 1960-е — нач. 1970-х гг., хорошо отметили Хокинс и его коллеги: «Мы все еще не знаем слоговые или логографические значения редких знаков, которые часто могут являться индивидуальными или местными вариантами. В целом в рамках системы письма отсутствует консенсус по поводу вокалических значений консонантных знаков; можно ли приписать некоторым знакам двойную или даже тройную вокализацию (*Ca/i* или *Ca/i/u*) или нам следует считать, что обычно каждый знак характеризуется только одной вокализацией (т.е. либо *Ca*, либо *Si*, либо *Cu*)? В данный момент для иероглифического письма мы не можем установить силлабическую решетку, сформированную в соответствии со строгой логикой, которая была бы сопоставима с логикой линейного письма Б или кипрского письма. Многие консонантные серии, такие как *ya* (*yí*), *yi*, *za* и *zi*, по-видимому, полностью отсутствуют. В случае с этими и другими отдельными проблемами слишком часто остается неясным, связана ли неоднозначность с отсутствием систематической организации в самом письме или с неадекватностью современной дешифровки» [Hawkins et al. 1974: 8].

Начиная с 1970-х гг., центр работы по дешифровке анатолийских иероглифов перемещается в Великобританию. Организатором рабочей группы по изучению анатолийских иероглифических надписей стал оксфордский филолог-классицист Леонард Палмер, полагавший, что лувийский язык даст ключ к пониманию критского линейного письма А. Несмотря на эту, по-видимому, ошибочную, посылку, Палмеру удалось сплотить группу молодых исследователей, среди которых выделялись ассириолог Дэвид Хокинс и классицист Анна Морпурго-Дэвис. Опираясь на исследования своих предшественников, оксфордские филологи приступили к созданию корпуса лувийских иероглифических текстов и публикации филологических изданий отдельных памятников. Однако прежде чем методы корпусного исследования принесли свои результаты, случайная находка эквивалентных клинописных и иероглифических граффити на пифосах при раскопках урартской крепости Алтын-Тепе позволила совершить новый рывок в чтении анатолийских иероглифов. Знак, читавшийся до этого как $\langle \bar{i} \rangle$, оказался соответствующим силлабограмме с дентальной аффрикатой в клинописной передаче. Это потребовало новых чтений $\langle \bar{i} \rangle \rightarrow \langle za \rangle$ и $\langle i \rangle \rightarrow \langle zi \rangle$, а поскольку место знаков серии $\langle i \rangle$ оказалось таким образом вакантным, возникла необходимость в реинтерпретации $\langle a \rangle \rightarrow \langle i \rangle$ и $\langle \bar{a} \rangle \rightarrow \langle ia \rangle$.

«Новые чтения» анатолийских иероглифов были независимо предложены немецким филологом Гюнтером Нойманном и английским тандемом Хокинс — Морпурго-Дэвис. Трое ученых решили объединить усилия для подготовки совместной публикации с обоснованием своего открытия [Hawkins et al. 1974]. Однако поскольку анатолийские

иероглифы уже считались к тому времени дешифрованными, доказательство «новых чтений» не могло ограничиться демонстрацией их преимущества на отдельных примерах, а потребовало фронтальной росписи всего иероглифического корпуса. При этом в качестве основного критерия оценки надежности интерпретации выступал сравнительный метод. Выяснилось, что последовательное применение «новых чтений» позволяет устранить большинство формальных расхождений между лувийским языком клинописных заклинаний и языком хеттских иероглифов. Например, лув. <za-> (иер.) «этот», заменившее прежнее чтение $**\langle\bar{i}-\rangle$, оказалось в точности соответствующим лув. <za-> (клин.) «этот». Окончание именительного падежа множественного числа общего рода <-i-zi> (иер.), предложенное вместо прежнего $**\langle-a-i\rangle$, оказалось формально совместимым с синонимичным <-in-zi> (клин.). Были, разумеется, и такие случаи, когда «новые чтения» создавали новые проблемы, например глагол <i-zi-> (иер.) «делать» оказался отделенным от <a-ya-> (клин.) «делать», с которым его сближало старое чтение $**\langle a-i\rangle$. Тем не менее, кумулятивные результаты работы Хокинса и его соавторов хорошо сочетались с прямым свидетельством граффити из Алтын-Тепе, и поэтому «новые чтения» получили быстрое признание в мировом научном сообществе.

В последующий период Гюнтер Нойманн сосредоточился на изучении анатолийских языков в алфавитной передаче, а Хокинс и Морпурго-Дэвис продолжили работу по дешифровке анатолийских иероглифов. Все новые открытия проверялись английскими учеными на полном корпусе иероглифических текстов. Прогресс в уточнении значений иероглифических логограмм достигался как при помощи сравнения с клинописными лувийскими и алфавитными ликийскими текстами, так и комбинаторным методом. Успешным примером последнего стала, в частности, идентификация лувийских отрицаний [Hawkins 1975]. К этому же периоду относятся первые серьезные попытки использования лувийских текстов в иероглифической передаче для уточнения анатолийской реконструкции. Так, иероглифические тексты были привлечены наравне с клинописью и алфавитными ликийскими текстами для реконструкции общелувической лениции [Morpurgo-Davies 1982/1983]. Подобная практика свидетельствовала о зрелости дешифровки и способствовала повышению интереса к анатолийским иероглифам среди широкого круга индоевропейцев.

Вместе с тем изучение лувийских иероглифических текстов в конце двадцатого столетия серьезно тормозилось отсутствием современной справочной литературы. Открытия семидесятых годов сделали устаревшими более ранние вторичные источники, а корпус иероглифических текстов в современной транслитерации долгое время оставался достоянием узкого круга исследователей, причастных к «новым чтениям». Это делало затруднительным участие в дешифровке молодых филологов, поскольку их гипотезы, основанные на частичном материале, могли быть опровергнуты с привлечением малодоступных текстов. В течение двадцати пяти лет научная общественность с нетерпением ждала публикации собрания лувийских иероглифических текстов, основанного на «новых чтениях». Эта лагуна стала ощущаться особенно остро после публикации корпуса лувийских клинописных текстов [Starke 1985] и глоссария лувийских форм в клинописной передаче [Melchert 1993].

Переход к современному этапу в исследовании анатолийских иероглифов ознаменовался выходом монументального «Корпуса иероглифических лувийских надписей» [Hawkins 2000]. Издание Хокинса включает в себя все известные к 2000 г. иероглифические тексты, относящиеся к постхеттскому периоду. В первых двух томах тексты даны в транслитерации. Каждый текст снабжен историческим предисловием, переводом и подробным лексико-грамматическим анализом. В третьем томе тексты даны в оригина-

ле, при этом имеются как фотографии, так и прорисовки. Издание Хокинса содержит довольно подробное историческое предисловие, обширный список литературы, перечень иероглифов, логограмм и небольшой конкорданс. С выходом изданий Хокинса и словаря Мельчерта стал возможен полноценный количественный анализ лувийской иероглифики.

Корпус Хокинса не претендует на абсолютную полноту. С одной стороны, из него эксплицитно исключены иероглифические тексты хеттского периода (до конца XIII в. до н. э.). Это, впрочем, создает дополнительные удобства для статистического анализа, поскольку орфография древнейших текстов еще не была вполне стабилизирована. С другой стороны, в корпус, разумеется, не могли быть включены надписи, обнаруженные в XXI веке. Поскольку публикация новых иероглифических находок из Анатолии осуществляется в целом достаточно оперативно, можно констатировать, что объем недавно опубликованных надписей достигает 10–15% корпуса Хокинса. Дополнительная обновляемая библиография по иероглифическим лувийским текстам железного века доступна по адресу: <http://agyagrap.blogspot.de/2010/10/list-of-iron-age-luwian-inscriptions.html>.

Вместе с тем можно констатировать, что обнаружение новых памятников пока не привело к принципиальному изменению существующей дешифровки. Последние шаги в интерпретации анатолийских иероглифов, реферируемые в настоящей статье, скорее являются результатом применения комбинаторных методов анализа к репрезентативному корпусу лувийских текстов. Ниже мы рассмотрим некоторые наиболее показательные работы, посвященные (1) различению идеограмм, фонетических компонентов и индикаторов, (2) уточнению чтений и интерпретаций идеограмм, (3) уточнению значений слоговых знаков. Из обзора должно быть ясно, что изучение анатолийской иероглифической письменности вышло на новый уровень, для которого характерна растущая независимость прогресса в данной области от внешних стимулов.

3. Различение идеограмм, фонетических компонентов и индикаторов

Анатолийские иероглифы, как и другие иерархические системы письма, часто используют омографичные написания для идеограмм (логограмм) и силлабограмм. Так, например, идеограмма ⟨PES⟩, использовавшаяся для обозначения ноги и глаголов движения, имеет фонетическое чтение ⟨ti⟩, выделившееся из хеттского глагола *tiya-* «ступать», а идеограмма ⟨SIGILLUM⟩ «печать» имеет фонетическое чтение ⟨sa₅⟩, полученное при посредстве лувийского существительного *sanza* «печать». Особенность анатолийских иероглифов — это употребление фонетических индикаторов, т. е. специальных силлабограмм, которые используются для уточнения чтения идеограмм, не передавая при этом точный фонемный состав или положение соответствующего слога. Для их обозначения используется заглавный курсив, например ⟨PRAE-*ia* AUDIRE+MI-ma-ti-mi-i-sa⟩ = *parriyatummantimmis* «прославленный», ⟨INFANS.NI-sa⟩ = *nimuwizzas* «сын». Вопрос о том, является ли тот или иной знак силлабограммой, фонетическим компонентом или фонетическим индикатором, не всегда имеет очевидный ответ.

Недавним примером применения количественного подхода при различении идеограмм и фонетических написаний является разбор, проведенный Якубовичем в рамках работы по происхождению анатолийских иероглифов [Yakubovich 2008]. Знак *41 (здесь и далее нумерация следует каталогу [Laroche 1960b]) имеет слоговое значение ⟨tā⟩, и в статье Рикен [Rieken 2008] убедительно показывается, что он используется для записи ленированного /da/ в интервокальной позиции (обзор этой работы см. ниже). Однако

имеется одно исключение — лувийский глагол со значением «братъ». При записи этого глагола знак *41 стоит в начальной позиции, т. е. в позиции нейтрализации фонетического противопоставления между напряженными и ненапряженными согласными. В этой позиции мы не можем ожидать /da-/, но, скорее, должны ожидать /tta-/ (в конвенциональной записи /ta-/). Тем не менее стремление связать лувийский глагол с хеттским *dā-* «братъ», привело анатолистов к введению мнимой лексемы лув. **/ta-/ = ⟨tā-⟩ «братъ».

Между тем в клинописных текстах имеется глагол лув. *la-* «братъ», а также его редуцированная форма *lala-* «братъ» [Melchert 1993: 120—121]. В качестве лувийского вкрапления в хеттский текст засвидетельствована производная основа *lalam(i)-* «получение, квитанция» [Melchert 1993: 122]. В иероглифических текстах также имеется глагол лув. *la-* «братъ». При записи знака *41 с фонетическим компонентом данный компонент иногда содержит *-la(-la-)* и никогда не содержит ***-ta-*. Особенно показательным является чередование в одном и том же тексте между фонетической записью ⟨la-i⟩ (KÖRKÜN §11) и логографическим написанием ⟨CAPERE-i⟩ (KÖRKÜN §8). Все это позволяет свести формы с начальным *41 (= ⟨CAPERE⟩) и формы глагола *la-* в единую парадигму (табл. 2):

Таблица 2. Парадигма глагола *la(la)-* «братъ».

Pres.	2sg.	⟨la-si⟩	ISKENDERUN §6
	3sg.	⟨la-i⟩	KÖRKÜN §11 (+ARHA)
		⟨CAPERE-i⟩	KARKAMIŠ A11a §22, A2+3 §20 (+ARHA), A6 §26, A15b §12, KÖRKÜN §8 (+ARHA), BOROWSKI 3 §9 (+ARHA), ALEPPO 2 §13, 18 (+ARHA), KÖTÜKALE §5 (+ARHA), BOYBEYPINARI 2 §19 (+CUM-ni ARHA), ANCOZ 7 §4 (+ARHA), 9 (+PRAE)
		⟨CAPERE-i?⟩	KELEKLİ §2
		⟨CAPERE-ia⟩	KARKAMIŠ A6 §27, 28, 30 (+CUM-ni ARHA), A31+ §12, HAMA 4 §8, 5 §1 (+ARHA)
		⟨CAPERE⟩	KARKAMIŠ A11b+c §22
		⟨CAPERE?-i⟩	ASMACIK 1.1—2
	3pl.	⟨CAPERE-ti-i⟩	KARKAMIŠ A11a §27, A11b+c §28, 29
Pret.	1sg.	⟨la-ha⟩	MARAŞ 13, l. 2—3 (+ARHA), BOHÇA §13
		⟨CAPERE(-)la-ha⟩	MARAŞ 4 §4,12 (+INFRA-ta)
		⟨CAPERE-ha⟩	KARKAMIŠ A11b+c §30 (+ARHA), A12 §5, A7 §3, KARAHÖYÜK §19, TELL TAYINAT 2 l. 2 frag 2b-a (+ARHA)
	3sg.	⟨CAPERE-ta⟩	TELL AHMAR 1 §12 (+SUPER+ra/i-a), TELL AHMAR fr. 1, TULLEIL 1 l.2 (or 3 pl?)
		⟨CAPERE-tá⟩	KARAHÖYÜK §7
		⟨CAPERE?-ta⟩	TELL TAYINAT 1, frag 3—5, l.5
		⟨CAPERE-ta _x ⟩	TOPADA §28
	3pl.	⟨(“CAPERE”)la-la-ta⟩	MARAŞ 1 §9 (+PRAE-i)
Imv.	3sg.	⟨CAPERE-tu⟩	EĞRIKÖY §1
	3pl.	⟨CAPERE-tú⟩	KARKAMIŠ A25a §3
Inf.		⟨CAPERE(-)la-l na⟩	MARAŞ 3 §8 (+ARHA)
		⟨CAPERE(-)la/i/u-na-´⟩	BOHÇA §3, 9
Part.	acc.sg.n.	⟨CAPERE-ma-za-´⟩	KULULU 2 §2
		⟨CAPERE-ma-z[a]⟩	BULGARMADEN §13

Предположение о чтении знака *41 в начальной позиции как *CAPERRE* и о связи с ним корня *la-* «братъ» высказывалось еще Морпурго-Дэвис [Morpurgo-Davies 1987: 211–212]. Однако проблематичность подобной интерпретации виделась в том, что с ее помощью сложно объяснить значение ⟨tā⟩ для знака *41. Якубович дает следующее объяснение: знак *41 изначально использовался как логограмма ⟨*CAPERRE*⟩ для глагола со значением «братъ»; затем на основе хетт. *dā-* «братъ» произошло акрофоническое развитие, в результате которого знак *41 получил значение ⟨tā⟩ = /da/; после установления конвенции, касающейся чтения лувийских иероглифов, знак *41 был вторично ассоциирован с лув. *la(la)-* «братъ», но его фонетическое значение сохранилось. Такая эволюция знака *41, по мнению Якубовича, является дополнительным аргументом в пользу идеи о том, что фонетический компонент анатолийского иероглифического письма изначально развивался в смешанном хетто-лувийском окружении. Таким образом, формальный анализ распределения анатолийских иероглифов может способствовать содержательным выводам социолингвистического характера.

Другим примером использования квантитативного подхода при различении идеограмм, фонетических комплементов и фонетических индикаторов является анализ лувийской лексемы со значением «брат». Предложенное решение было кратко упомянуто в работе [Rieken, Yakubovich 2010: 211], но впервые детально аргументируется в настоящей публикации. Знак *45 используется при записи лувийских лексем со значением «брат» и «сын, ребенок». В первом случае он условно транслитерируется как ⟨*FRATER*⟩, а во втором — как ⟨*INFANS*⟩. Лувийская лексема *nimuwizza-* «сын» отличается от лексемы со значением «брат» с помощью фонетического индикатора ⟨*NI*⟩, кодирующего начальный слог. Традиционно считалось, что слово для «брата» сопровождается фонетическим комплементом ⟨-*la-*⟩, однако Рикен и Якубович показали, что в данном случае мы имеем дело не с комплементом, а с индикатором.

В пользу записи ⟨*FRATER.LA*⟩, вместо традиционного ⟨*FRATER-la-*⟩, говорят несколько соображений. Во-первых, знак ⟨*FRATER*⟩ при маркировании соответствующей лексемы никогда не встречается без ⟨*LA*⟩. Во-вторых, использование фонетического индикатора со знаком *45 выглядит более последовательным с учетом аналогичного употребления ⟨*INFANS.NI*⟩. В-третьих, запись ⟨*FRATER-la-*⟩ предполагает существование нерегулярной парадигмы, тогда как запись ⟨*FRATER.LA*⟩ позволяет реконструировать обычное склонение с *i*-мутацией (см. табл. 3).

Таблица 3. Парадигма существительного ⟨*FRATER.LA*⟩ «брат».

nom.sg.	⟨ <i>FRATER.LA-sa</i> ⟩	TELL AHMAR 2 §18
	⟨ <i>FRATER.LA-i-sa</i> ⟩	ALEPPO 2 §3
	⟨[-F]RATER.LA-i-sa⟩	EĞRİKÖY §1
acc.sg.	⟨ <i>FRATER.LA-na</i> ⟩	ALEPPO 2 §9
gen.sg.	⟨ <i>FRATER.LA-sa</i> ⟩	TELL AHMAR 2 §18
	⟨ <i>FRATER.LA-sa_s</i> ⟩	KULULU 4 §15
dat.sg.	⟨ <i>FRATER.LA-ni</i> ⟩	KARKAMIŞ A5a §3
	⟨ <i>FRATER.LA-i</i> ⟩	KULULU lead strip 1 l. 38, 44
nom.pl.	⟨ <i>FRATER.LA-zi-i</i> ⟩	KARKAMIŞ A7 §2
	⟨ <i>FRATER.LA-i-zi</i> ⟩	JISR EL HADID l. 2
acc.pl.	⟨ <i>FRATER.LA-zi-i</i> ⟩	KARKAMIŞ A6 §13, KARKAMIŞ A15b §15
dat.pl.	⟨ <i>FRATER.LA-za</i> ⟩	TELL AHMAR 1 §16, MARAŞ 6 l. 1, KIRÇOĞLU §4
poss.adj.dat.sg.	⟨ <i>FRATER.LA-sa-na</i> ⟩	KARKAMIŞ A4a §2

Кроме того, данная основа также встречается в качестве компонента следующих личных имен: ⟨^lá-là/i-|FRATER.LA-⟩ (KARKAMIŠ A18j), ⟨ara/i-FRATER.LA-⟩ (TÜNP 1 §1), ⟨^lmu-a-|FRATER.LA-⟩ (KARATEPE 3 §1), ⟨^lmu-sa-FRATER.LA-⟩ (KULULU 1/s 1 1.13), ⟨^lsà-tá-|FRATER.LA-⟩ (CEKKE §6c), ⟨FRATER.LA-wa/i+ra/i-⟩ (CEKKE §6d), ⟨OMNIS-mi-|FRATER.LA-⟩ (CEKKE §9m), ⟨^lREL-za-FRATER.LA-⟩ (KULULU 1/s 1 1.17). По-видимому, функциональная причина введения фонетического индикатора заключалась в необходимости разрешения неоднозначности между морфемами для «брата» ⟨FRATER.LA⟩ и для «сына, ребенка» ⟨INFANS.NI⟩, которые записывались с помощью одного иероглифического знака в железном веке. Хотя фонетическая форма, стоящая за FRATER.LA, остается неизвестной, в данном случае сочетание комбинаторных и функциональных аргументов представляется достаточным для доказательства.

Употребление анатолийских иероглифов в качестве фонетических индикаторов фиксируется даже в тех случаях, когда подобная стратегия кажется функционально избыточной. В качестве примера можно рассмотреть лексему для «войска, армии»: acc.sg. ⟨EXERCITUS.LA//U-za⟩ = /kwalanza/ (KARATEPE 1 (Hu.) §9), dat.sg. ⟨EXERCITUS.LA//U-ní⟩ = /kwalani/ (KARATEPE 1 (Hu.) §9 и т. д.), instr.sg. ⟨EXERCITUS.LA//U-ti⟩ = /kwalanadi/ (KÖRKÜN: Rev. §9 и т. д.). В том случае, если бы полифоничный знак ⟨la/i/u⟩ являлся не фонетическим индикатором, а фонетическим компонентом, пришлось бы допустить существование нерегулярной формы instr.sg. ****/kwaladi/**. Между тем существование лувийской основы *kwalan-* «войско, армия» подтверждается клинописными данными [Melchert 1993: 119]. Поэтому в данном случае самодостаточный комбинаторный анализ подтверждается этимологическими соображениями, несмотря на отсутствие функциональных коррелятов.

4. Уточнение интерпретаций идеограмм

Полноценный количественный подход, ставший возможным благодаря корпусу Хокинса, применяется и при уточнении фонетических чтений идеограмм. Примером сочетания количественного и этимологического подхода является статья Петры Худехебурэ [Goedegebuure 2010], посвященная пространственным наречиям *zanta* и **anni*. Для наших целей интересен в первую очередь вклад Худехебурэ в интерпретацию лувийской иероглифической формы ⟨INFRA-ta⟩. Раньше считалось (ср. [Plöchl 2003]), что лувийское иероглифическое ⟨INFRA-ta/tá⟩ соответствует хеттскому *katta* «вниз» и потому должно читаться как **katta*, хотя прямых доказательств в пользу такого чтения не существовало. Худехебурэ приводит многочисленные аргументы, показывающие, что ⟨INFRA-ta/tá⟩ не может читаться как ***katta*, и доказывает, что их следует читать как *zanta*.

Традиционное предположение о чтении ⟨INFRA-ta/tá⟩ как **katta* всецело базируется на экстраполяции хеттской формы *katta*. В самих иероглифических текстах ⟨INFRA-ta/tá⟩ никогда не дается в фонетической записи, которая могла бы свидетельствовать о начальном /k/. В клинописных лувийских текстах мы также не находим ничего похожего на **katta*. Между тем имеется клинописная форма *zanta*, которая традиционно рассматривалась как лексикализованный инструменталис от *za-* «этот» [Carruba 1982: 44; Starke 1990: 42; Melchert 1993: 278]. Вместо этого Худехебурэ предлагает рассматривать *zanta* как пространственное наречие, являющееся синтаксическим и семантическим аналогом хеттского *katta* «вниз». Это предположение выглядит привлекательнее, чем старая гипотеза, хотя бы потому, что *zanta* стабильно занимает синтаксическую позицию проверба по всему корпусу лувийских клинописных текстов. Такая же прилагательная позиция,

разумеется, характерна как для хеттского преверба *katta*, так и для иероглифического <INFRA-ta/tá>.

Однако Худехебурэ идет дальше, последовательно подтверждая комбинаторные соображения этимологическими аргументами. Она анализирует многочисленные лувийские и хеттские контексты, позволяющие установить полный параллелизм между лув. *zanta* и хетт. *katta*. Ее анализ может быть резюмирован в следующей Таблице 4 (в английских переводах можно явственнее передать роль наречий *zanta / katta* «down»).

Таблица 4. Соответствия между лув. (клин.) *zanta* и хетт. *katta*.

Лувийский	Хеттский	Перевод
-tta zanta ār-	-kan katta ar-	«спускаться с» / «come down from»
-tar zanta kiš(ai)-	-kan katta kiš(ai)-	«счесывать» / «comb down»
[-tta] zanta mana	-ašta/-kan katta auš-	«смотреть вниз с» / «look down from»
zanda pašš-	katta pašš-	«проглатывать» / «swallow down»
zanta ša-	katta šai-/šiya-	«отправить вниз» / «send down»
-tar zanta tarmai-	-šan/-ašta katta tarmai-	«прибивать на» / «nail down on»
zanta dupai-	(-kan) katta walh-t	«поражать» / «strike down, afflict»
zanta(l)lanu-	tepnu-	«унижать» / «belittle, diminish»

Установление соответствия между клинописным *zanta* и хетт. *katta*, в свою очередь, позволяет составить более точное представление об индоевропейской этимологии этих форм. Хеттское *katta* может восходить как к и.-е. **kmt-o*, так и к и.-е. **k'mt-o*, поскольку в хеттском языке совпали палатальные и простые велярные [Kloekhorst 2008: 463—464]. Однако в лувийском языке было сохранено различие между двумя сериями велярных: и.-е. **k* > лув. *k*, и.-е. **k'* > лув. *z* [Melchert 1987]. Поскольку мы видим в лув. *zanta* начальное /z/, то анатолийские формы должны восходить к и.-е. **k'mt-o*.

Другим примером применения корпусного анализа к системе превербов является статья Якубовича [Yakubovich 2012], посвященная лувийскому пространственному наречию со значением «прочь». Это наречие обычно записывается с помощью знака *216, который принято транслитерировать как <ARHA> и интерпретировать как ***arha*, что связано с существованием аналогичного хеттского наречия *arha* «прочь». Между хеттским и лувийским наречием отмечен семантический и синтаксический параллелизм: ср. хетт. *arha da-* «забирать» vs. лув. *ARHA la-* «забирать»; хетт. *arha pai-* «уходить» vs. лув. *ARHA i-* «уходить»; хетт. *arha ed-* «съесть» vs. *ARHA ad-* «съесть». Другим традиционным аргументом в пользу фонетического чтения лувийского наречия как **/arha/* является то, что рассматриваемое наречие записывалось в имперский период с помощью знака <FINES>, который в иероглифических надписях железного века маркирует существительное *irh(i)-* «граница». При этом логограмма <ARHA>, используемая для соответствующего пространственного наречия в текстах железного века, состоит из <FINES+ha>. Отсюда делался вывод, что лувийское наречие родственно существительному *irhali-*, и оно, скорее всего, должно читаться как ***/arha/* = *arha*. Якубович полагает, что традиционные аргументы в пользу такого чтения являются слабыми, и стремится показать, что рассматриваемое лувийское наречие должно читаться как */axha/* = *ahha*.

Прежде всего, Якубович обратил внимание на отсутствие формы ***/arha/* в лувийских клинописных текстах, где все известные пространственные наречия записываются фонетически. Ранее это объяснялось тем, что соответствующая лексема отсутствует в

«клинописном лувийском», но сейчас такая интерпретация уже не представляется правдоподобной, поскольку показано, что клинописные и иероглифические тексты — это тексты на одном языке, записанные разными системами письма; диалектное членение проходит не между корпусами, а внутри клинописного корпуса (иероглифический корпус более однороден). К этому можно добавить, что все пространственные наречия, зафиксированные в иероглифических текстах, либо находят прямые соответствия, либо имеют прозрачные дериваты в клинописи. Было бы странно предполагать, что пространственное наречие *ARHA*, засвидетельствованное более 200 раз в иероглифике, по воле случая не сохранилось в клинописном корпусе.

В действительности удается обнаружить формы соответствующего пространственного наречия в клинописных текстах, которое выглядит там как *ahha*. Сегмент ⟨(a)ah-ha⟩ зафиксирован в клинописном корпусе 8 раз [Melchert 1993: 4]. Согласно традиционной интерпретации во всех случаях *ahha* является союзом со значением «когда, как» (в иероглифическом корпусе такой союз не засвидетельствован). Проанализировав все контексты, Якубович заключает, что только в одном случае *ahha* может интерпретироваться в качестве союза, при этом в трех случаях данная лексема выполняет функцию пространственного наречия, а в других четырех случаях ее функция неясна, что связано с фрагментарностью текста. Синтаксические критерии играют ключевую роль при разграничении вероятных функций *ahha*. Например, фрагментарная клауза *ahha=da tuwatta hatta[z-... «он убрал их прочь...»* (КВо 29.6 obv. 13) содержит преверб *ahha* непосредственно перед глагольной формой *tuwatta* «положил, убрал», от которой он отделяется лишь ваккернагелевской клитикой *=(a)da* «их».

Однако, учитывая малочисленность клинописных примеров, предложенное чтение не было бы вполне убедительным, если бы не удалось обнаружить свидетельства в пользу чтения *ahha* в иероглифическом корпусе. К счастью, когнат хеттского *arha* не записывался как ⟨*ARHA*⟩ в тех случаях, когда он образовывал одну морфему с предшествующим превербом. В частности, сегмент ⟨*wa/i-na-ha*⟩ (ISKENDERUN §6, HAMA 4 §8) должен интерпретироваться как составной преверб *wanahha* (< */awan/ + */axxa/), что является соответствием хетт. *awan arha*. Якубович также показывает, что преверб ⟨*SUPER+ra/i-ha*⟩ (SULTANHAN §19—20, ASSUR e §25) должен читаться как *sarihha* (< */sarri/ + /ahha/) или *sarahha* (< */sarra/ + /ahha/), а его хеттским эквивалентом является *ser arha*. Наконец, сегмент ⟨*PRAE-i-ha*⟩ (ANCOZ 7 §9, BULGARMADEN §6) может быть интерпретирован как составной послелог *parihha* < */pari/ + /axxa/, которому формально соответствует хеттское пространственное наречие *para arha*. Гипотеза Мельчерта [Melchert 2013], согласно которой ассимиляция */arxa/ > /axxa/* ограничена хеттскими заимствованиями в имперский лувийский, теоретически могла бы объяснить случай *wanahha*, но никак не *parihha*, где мы явно имеем дело с унаследованным когнатом хетт. *para arha*, а не с заимствованием.

Таким образом, Якубович подтверждает фонетическую реальность лув. *ahha* независимыми примерами из иероглифических и клинописных текстов. Хотелось бы, однако, подчеркнуть, что вопрос о пересмотре фонетической интерпретации ⟨*ARHA*⟩, вероятно, не возник бы, если бы не выявленное отсутствие ***arha* в лувийских клинописных текстах. Не случайно, что корпусные методы фонетической интерпретации с опорой на статистические аргументы оказались особенно продуктивны для лувийских превербов, образующих закрытую систему из частотных элементов.

Вместе с тем уточнение чтений идеограмм на основе корпусного анализа коснулось и других частей речи. В качестве примера можно привести статью Тео ван ден Хаута [van den Hout 2002], посвященную лувийским обозначениям для «личности» и «души», часто скрывающимся за идеограммой ⟨*COR*⟩. Лувийская лексема *atr(i)- / atl(i)-* соответствует ликийскому *atra-/atla-* «личность, сам(ость)» и карийскому *ot2r* «сам». Добавляя к этому

мнимую лувийскую *n*-основу *at(a)ni-*, некоторые исследователи (Дэвид Хокинс, Иво Хайнал) высказывали точку зрения о том, что в данном случае следует реконструировать древнюю гетероклитическую *r/n*-основу среднего рода. Ван ден Хаут привел убедительные аргументы против этой позиции. Ниже приводится роспись обеих форм по корпусу [Hawkins 2000], из которой очевидно, что две лексемы не находятся в дополнительном распределении по падежным формам, и поэтому о синхронной гетероклизе не может быть и речи. Кроме того, для первой лексемы засвидетельствован графический вариант с выписанной начальной гласной, а во втором случае он отсутствует, поэтому форма ****COR-tani** может читаться с детерминативом как (COR)*tan(i)-*.

Таблица 5. Лувийские квазисинонимы *atr(i)-* и *tan(i)-*.

<i>atr(i)-</i> «личность, образ»	acc.sg.	⟨á-tara/i-i-na⟩	KULULU 1 §12
		⟨“COR”-tara/i-(na)⟩	KARKAMIŠ A14a §6, KARKAMIŠ A2 §§2, 3, ALEPPO 2 §10
		⟨COR-tara/i-na⟩	MARAŞ 2 §2, KULULU 4 §4, TELL AHMAR 5 §15, TOPADA §37
		⟨COR-tara/i-i-na⟩	TELL AHMAR 5 §11, IZGIN 2 §8, TOPADA §35
		⟨(“COR”)a-tara/i-i-na⟩	KARKAMIŠ A15b §10
		⟨(“COR”)á-tara/i-i-na⟩	MARAŞ 4 §14
		acc.pl.	⟨COR-tara/i-zi⟩
<i>tan(i)-</i> «душа»	instr.	⟨COR-la-ti-i-’⟩	KULULU 4 §5
		acc.sg.	⟨COR-ni-na⟩
<i>tan(i)-</i> «душа»	dat.sg.	⟨(“COR”)ta-ni⟩	ASSUR f+g §29
		⟨COR-ni-i⟩	TELL AHMAR 2 §13
	instr.	⟨“COR”-na-ti⟩	KÖRKÜN §3

Формальный анализ ван ден Хаута подкрепляется этимологическими аргументами. Лув. *tan(i)-* является соответствием хетт. *istan-za(n)-* «душа» и может реконструироваться на индо-хеттском уровне как **psten-* «грудь». Поэтому в данном случае для реконструкции начальной гласной не находится не только синхронных, но и этимологических оснований. Вместе с тем можно предполагать, что *tan(i)-* и *atr(i)-* / *atl(i)-* могли контаминировать в позднелувийском языке. Это следует из последовательного употребления обоих существительных по отношению к одному и тому же референту в надписи VIII в. до н. э. KULULU 4. За предложением ⟨wa/i-mu-ta (LITUUS)á-za-mi-na COR-tara/i-na a-ta tu-tá⟩ «Они (боги) вложили в меня милую душу» (KULULU 4 §4) следует с некоторым интервалом развитие сюжета ⟨wa/i-mu-ta DEUS-ni-zi-i (LITUUS)á-za-mi-na || COR-ni-na a-ta tu-wa/i-mi-na-’ la-ta | wa/i-li?-ia-wa/i-ti-na⟩ «Боги взяли у меня вложенную милую душу, полную благорасположения» (KULULU 4 §9). Подобная контаминация неудивительна, учитывая, что для позднейших лувийских памятников характерна как афреза начального **a-*, так и спорадический переход **-n- > -r-* в интервокальной позиции. Поэтому не исключено, что подчеркнутые формы произносились соответственно как **atrin > t(a)rin* и **tanin > tarin* в VIII в. до н. э., что и способствовало их смешению.

Таким образом, упор на отдельные контексты действительно затрудняет различение существительных *tan(i)-* и *atr(i)-* / *atl(i)-*. Для того чтобы добиться правильного лексикографического описания и фонетической интерпретации обеих лексем, в данном случае, как и во многих других случаях, необходим репрезентативный корпусный анализ.

5. Уточнение значений слоговых знаков

В предыдущих двух секциях обсуждались достижения корпусного метода, имевшие важное значение для дальнейшей дешифровки анатолийских иероглифов, но едва ли способные оказать влияние на развитие анатолистики и индоевропеистики в целом. Ниже речь пойдет о реинтерпретации силлабограмм, которая изменяет наши представления о фонетической системе лувийского языка, и тем самым способствует его исторической реконструкции. Ведущая роль в уточнении лувийской фонетики методами корпусного анализа принадлежала в последние годы Элизабет Рикен.

Статья [Rieken 2008] посвящена знакам *100, *29 и *41, которые транслитерируются как ⟨ta⟩, ⟨tá⟩ и ⟨tà⟩ соответственно. До ее выхода считалось, что употребление этих знаков ничем не мотивировано, и все они могут указывать как на фонетическое /tta/, так и на фонетическое /da/. В результате подробного количественного анализа Рикен удалось обнаружить, что в иероглифических текстах первого тысячелетия до н. э. ⟨ta⟩ и ⟨tá⟩ соответствуют /(t)ta/, а ⟨tà⟩ — /da/. Рикен обратила внимание на то, что знак ⟨tà⟩, в противоположность знакам ⟨ta⟩ и ⟨tá⟩, никогда не встречается в начале слова, что связано с нейтрализацией фонетического противопоставления в данной позиции. Мнимым исключением является глагол *SAPERE* «брать», о котором уже шла речь выше при обзоре статьи [Yakubovich 2008]. Заметим, что данное распределение является чисто синхронным и не опирается на сравнительно-исторический анализ. Вместе с тем *a posteriori* можно отметить, что предполагаемая нейтрализация унаследованной оппозиции между звонкими и глухими (или напряженными и ненапряженными) согласными в начальной позиции в слове с высокой вероятностью относится к контактно обусловленным явлениям лувийского языка. Данное распределение, в целом характерное для языков анатолийской группы, находит близкую параллель в хурритском языке, но не в близкородственном ему урартском. С типологической точки зрения оно является достаточно необычным, и поэтому его можно объяснять влиянием общего субстрата.

Кроме того, Рикен проанализировала написания пространственного наречия *anta* «в, вовнутрь»: в 38 случаях оно записывается как ⟨a-ta⟩, в 32 случаях как ⟨a-tá⟩ и в одном случае как ⟨á-tá⟩; иначе говоря, оно никогда не записывается с ⟨tà⟩, что обусловлено нейтрализацией оппозиции в консонантном кластере и фонетической реальностью /ta/ в /anta/. Напротив, многочисленные иероглифические формы, записываемые с ⟨tà⟩, имеют надежные параллели в клинописных текстах, где они даются в «ленированной» записи с ⟨(C)a-ta⟩ или ⟨(C)a-da⟩, в противоположность ⟨(C)a-at-ta⟩ или ⟨(C)a-ad-da⟩, в соответствии с законом Стертеванта. Некоторые примеры приведены в Табл. 6.

Наконец, Рикен заметила, что некоторые лексемы, подвергшиеся «ротацизму», т. е. позднелувийскому переходу /d/ > /r/ (см. [Morpurgo-Davies 1982/1983]), регулярно записываются с ⟨tà⟩, например, ⟨(LITUUS)á-za-ti-wa/i-tà-sá⟩ vs. ⟨(LITUUS)á-za-ti-wa/i-ra/i-sá⟩ «Ацативада»; ⟨(FLUMEN.REGIO)ha-pa-tà-si⟩ vs. ⟨(“FLUMEN”)há-pa+ra/i-sá⟩ «речная долина; область» и др. Особое значение имеют формы глагольного окончания третьего лица единственного числа прошедшего времени, иллюстрируемые в табл. 7. В этой связи существенно отметить, что соответствующие окончания третьего лица множественного числа прошедшего времени никогда не записываются с помощью ⟨tà⟩ и не подвергаются ротацизму. Подобное распределение вполне естественно, учитывая, что соответствующие окончания реконструируются как *-to и *-nto на индо-хеттском уровне [Yoshida 1993/1994].

Последняя группа примеров иллюстрирует также важную роль открытия Рикен для реконструкции анатолийской глагольной морфологии. Как показала Анна Морпурго-

Таблица 6. Формы с <tà) и связанные с ними ленированные формы из клинописи.

Лексема	Иероглифическая форма	Клинописная форма
ad- «есть»	<(á-tà-tu-u) SULTANHAH §33a и др.	<a-da-an-du) KBo 13.260 III 10'
huhadall(i)- «дедовский»	<(AVUS-ha-tà-li-sa) KARKAMIŠ A11 b+c §2	<(hu-ha-da-al-l[a]) KUB 33.106+ II 61
pad(i)- «нога»	<(“PES”)pa-tà-za) KARATEPE §XXII	<(pa-a-ta-an-za) KUB 9.6+ III 29”
tadall(i)- «отцовский»	<(tá-tà-li-sa) KARKAMIŠ A11 b+c §2	<(da-da-al-la) KUB 33.106+ II 61
tiwad- «Тивад (бог Солнца)»	<(DEUS)SOL-tà-ti-i) KARKAMIŠ A 15b §1	<(ti-wa-da-ni-in-ti) KBo 12.89 III 9

Таблица 7. Глагольные формы, записанные с <tà) и с ротацизмом.

Лексема	Нормальная форма	Форма с ротацизмом
i- «идти»	<(“PES ₂ ”)i-tà) KAYSERİ §19	<(PES ₂)i+ra/i) CEKKE §20
awi- «приходить»	<(“PES”)á-wa/i-tà) İSKENDERUN §2 и др.	<(á-wa/i+ra/i) PALANGA §11
izziya- «делать»	<(i-zi-i-tà) BABYLON 2 §1 и др.	<(i-zi-i-ri+i) KULULU 5 §4
mana- «видеть»	<(LITUUS+na-tà) KARKAMIŠ A2 §6 и др.	<(LITUUS+na-ri+i) ASSUR letter f+g §17

Дэвис, далеко не все лувийские глагольные формы 3sg. подвергаются лениции. Распределение между «сильными» и «слабыми» (ленированными) окончаниями зависит от типа глагольной основы и задается достаточно сложными правилами [Morpurgo-Davies 1982/83]. Однако поскольку ротацизм «слабых» окончаний лишь спорадически отражается в иероглифической графике, зачастую остается невозможным определить по данному критерию, индуцирует ли леницию та или иная глагольная основа. Напротив, распределение между претеритными окончаниями 3sg. <ta)/<tá) и <tà) имеет категорический характер, и поэтому характер глагольной основы, как правило, можно установить по одной-единственной форме 3sg.pret. Например, «слабая» основа <(i-zi-(i)-) ‘делать’ стабильно противопоставляется «сильной» основе <(á-ri-(i)-) ‘поднимать’. Последовательное различие между «сильными» и «слабыми» основами в позднелувийском важно для внешнего сравнения, поскольку правила лениции достаточно схожи в близкородственном ликийском языке, а ликийские «сильные» и «слабые» формы всегда различаются в алфавитной передаче. Результат сравнения между лувийским и ликийским языками, в свою очередь, помогает реконструировать модели глагольной деривации, характерные для пралувиического и праанатолийского языков.

Другим примером сочетания количественного и этимологического подходов для уточнения значений слоговых знаков является недавняя статья Элизабет Рикен и Ильи Якубовича [Rieken, Yakubovich 2010], посвященная знакам *319 и *172. Со времен Лароша считалось, что эти знаки обозначают открытый слог, начинающийся с переднеязычного согласного. Ларош [Laroche 1960] предложил для них транслитерацию **<té/tí) и **<tì) соответственно; сходной точки зрения придерживался и Мериджи [Meriggi 1962], предлагавший записывать эти знаки как **<tí) и **<tì). В публикации 1975 г. Хокинс и Морпурго-Дэвис [Hawkins, Morpurgo-Davies 1975: 130–132] привели многочисленные свидетельства в пользу -a- вокализма для обоих знаков, при этом, не отвергая полностью -i- вокализм, они отмечали его скудную представленность. В той же статье отмечалась возмож-

ность латеральных инициалей для соответствующих силлабограмм, но аргументы в пользу данной интерпретации были признаны недостаточными. В итоге авторы предложили для знаков *319 и *172 транслитерацию **⟨ta₄⟩ и **⟨ta₅⟩. В ходе дальнейшего развития дешифровки было сделано предположение, что данные знаки отражают слоги с ленированным дентальным смычным [Melchert 1994: 38]. Рикен и Якубович пересматривают традиционную точку зрения и приводят многочисленные материалы, показывающие, что знаки *319 и *172 должны транслитерироваться как ⟨la/i⟩ и ⟨lá/i⟩ для лувийских надписей железного века.

Прежде всего, Рикен и Якубович обращают внимание на то, что гипотеза, согласно которой знаки *319 и *172 отличаются от знаков ⟨ta⟩, ⟨tá⟩ и ⟨tà⟩ лишь признаком звонкости или слабости, уже не актуальна. Как уже упоминалось, Рикен показала, что в иероглифических надписях в случае лениции используется знак ⟨tà⟩, а в случае напряженного /tta/ записываются ⟨ta⟩/⟨tá⟩ [Rieken 2008], поэтому контраст между «сильными» и «слабыми» дентальными смычными обеспечивается иными средствами. С другой стороны, еще в 1995 году Хокинс [Hawkins 1995: 114–115] привел аргументы в пользу чтения знаков *416 (предшественник *319) и *172 как ⟨ali⟩ и ⟨ala⟩ в лувийских надписях бронзового века. Поэтому гипотеза о латеральных инициалах силлабограмм *319 и *172 нуждается в дополнительной проверке. Результаты анализа Рикен и Якубовича, с учетом последующих исследований, можно резюмировать в Таблице 8.

Пять из шести частотных позднелувийских лексем, содержащих знаки *319 и *172, имеют внешние параллели или варианты написания, указывающие на их произношение с латеральным сонантом. Например, ⟨á-172/319-ma-⟩ «имя» имеет надежные анатолийские когнаты: хетт. *laman-*, лик. *alāma-* «имя». Эти формы однозначно говорят в пользу того, что лувийская форма отражает основу /alaman-/. Исходя из подобных аргументов, Рикен и Якубович предлагают новую транслитерацию знаков *172 и *319 как соответственно ⟨la/i⟩ или ⟨lá/i⟩, что хорошо согласуется с транслитерацией их предшественников, как ⟨ali⟩ и ⟨ala⟩. Единственным системным исключением, пока так и не получившим удовлетворительного объяснения, является лувийская лексема ⟨LOCUS—319/172-t°⟩, соответствующая лик. *pddānt-* «место». В данном случае остается предполагать употребление логограммы с фонетическими индикаторами, напр. LOCUS.LA/I, которая была перенесена с какой-то другой синонимичной лексемы, содержащей латеральный элемент. Несмотря на нерешенность данной проблемы, кумулятивные данные, приведенные выше, говорят в пользу новой интерпретации.

Существует, однако, другая группа аргументов, которые традиционно приводились в пользу чтения знаков *319 и *172 как **⟨ta₄⟩ и **⟨ta₅⟩ [ср. Hawkins, Morpurgo-Davies 1975: 130–131]. Речь идет о единичных вариантных формах, отраженных в Таблице 9. Предложенные приблизительные чтения в нормализованной записи отражают наши представления о том, как соответствующие лексемы могли бы произноситься в «стандартном» лувийском языке.

Согласно Рикен и Якубовичу, исследователи, транслитерировавшие во всех представленных случаях знак *172 как **⟨ta₅⟩, упустили из виду то, что здесь мы имеем дело с примерами из поздних текстов (IX — VIII вв. до н. э.). Кроме того, ни одна из представленных форм не записывается с **⟨ta₅⟩ чаще, чем один раз. Рикен и Якубович объясняют такую запись фонетического /da/ графическим смешением и случаями гиперкоррекции, которые обусловлены позднелувийским фонетическим переходом /d/, /l/, и /r/ во флэп /r/. Об этом переходе имеется довольно много свидетельств. С одной стороны, хорошо известен позднелувийский феномен «ротацизма», описанный Морпурго-Дэвис (см. выше). С другой стороны, отмечены многочисленные случаи перехода /l/ > /r/: ср. ⟨wa-la-⟩

Таблица 8. Фонетические значения *319 и *172 в частотных лексемах.

Закономерные соответствия	Исключения
MALUS-319/172/la- ~ (клин.) <i>adduwal(i)</i> - «злой»	
á-172/319/la-ma- ~ лик. <i>alāma</i> - «имя»	
(FEMINA) á-172- ~ Hitt. ^(d) <i>Ala</i> (теоним)	
*a-319-ni- = *a-319-lu/a/i-ni- «враг»	
SERVUS-319/172- ~ CLuw. <i>hutarla/i</i> - «слуга»	SERVUS = ⟨mí⟩ (1×)
	LOCUS-319/172-t° ~ Lyc. <i>pddānt</i> - «место»

Таблица 9. Свидетельства в пользу традиционного чтения знаков *319 и *172.

	Источник	Транслитерация и чтение	Перевод
A	KARKAMIŠ A6 §4	⟨wa/i-ma-172⟩ = <i>wa=m(u)=ada</i>	(клитич. группа)
B	KARKAMIŠ A6 §11	⟨⟨“PES ₂ +PES”⟩⟩tara/i-pa-172⟩ = <i>tarpada</i>	«топал»
C	KARKAMIŠ A 15b §12	⟨⟨“PES”⟩⟩pa-172- ‘) = <i>padi</i>	«у (ее) ноги»
D	KARKAMIŠ A 15b §21	⟨⟨“VIA”⟩⟩ha+ra/i-wa/i-ta-hi-172-ti-i) = <i>harwattahidadi</i>	«путешествием»
E	KARATEPE (Hu.) §49	⟨sa-pi-sá-172-ri+i) = <i>saPisadadi</i>	«со здравием»
F	KARATEPE (Hu.) §67	⟨i-zi-172) = <i>izida</i>	«сделал»
G	MARAŞ 3 §4	⟨EDERE?-172-ti-wa/i-li-na) = <i>adantiwalin</i>	«жирный (acc.sg.)»

(KARKAMIŠ A23 §9) vs. ⟨wa/i+ra/i-⟩ (KULULU 2 §3) «умирать», ⟨pa-la-sa-⟩ (KARKAMIŠ A2+3 §22) vs. ⟨pa+ra/i-si) (KARKAMIŠ A6 §19) «путь», ⟨ka-la/i/u-na-⟩ (MARAŞ 8 §7) vs. ⟨ka-ru-na-⟩ (KARATEPE §7) «житница», ⟨MALLEUS-la/i/u-⟩ (KARKAMIŠ A14b §3) vs. ⟨MALLEUS-x+ra/i-⟩ (MARAŞ 8 §12) «разбивать», ⟨tu-ni-ka-la-⟩ (KARKAMIŠ A2+3 §22) vs. ⟨tu-ni-ka-ra+a-⟩ (ASSUR f+g §45) «пекарь».

Если предположить, что нейтрализация /d/, /l/ и /r/ имела регулярный характер в определенных фонетических позициях в позднейший период, то иероглифические писцы должны были испытывать определенные трудности при выборе исторических написаний. Это могло приводить не только к отражению «ротацизма» на письме, но и к отдельным случаям гиперкоррекции, когда писцы неправильно реконструировали слоги, синхронно начинавшиеся с одноударного /r/. Случаи, собранные в Таблице 9, вероятно, относятся к последней категории и поэтому не могут быть использованы для определения исходных фонетических значений иероглифа *172. Однако для того, чтобы прийти к такому выводу, и вообще для изоляции гиперкорректных написаний, необходим был квантитативный анализ с учетом датировки текстов. Разумеется, ни то ни другое не было возможно до публикации корпуса [Hawkins 2000], что и объясняет задержку с обоснованием правильных чтений ⟨la/i) и ⟨lá/i).

Наконец, большое значение для уточнения чтений слоговых знаков и для индоевропейской реконструкции имеет недавняя статья Элизабет Рикен [Rieken 2010], посвященная знаку *104. Этот знак традиционно читается как ⟨sà), причем принято думать, что его фонетическая реализация ничем не отличается от фонетической реализации других знаков, маркирующих /sa/ (*415 ⟨sa), *433 ⟨sá), *402 ⟨sa₄, *327 ⟨sa₅, *223 ⟨sa₆, *316 ⟨sa₇, *380 ⟨sa₈). Рикен выписала все формы из текстов начала I тысячелетия до н. э. (приблизительно 1000—850 гг. до н. э.) и показала, что в текстах этого периода знак ⟨sà) употребля-

ется для фонетического /ʃa/ или /ʃ/; при этом звук /ʃ/ является фонетическим развитием */s/ в позиции после /i/ и /u/ в лексемах с этимологическим сочетанием */swō/ > лув. /ʃa/ и в ряде других позиционно или исторически обусловленных контекстов. В позиции перед /i/ и /u/ знак ⟨sà⟩ зафиксирован, например, в следующих формах:

Таблица 10. Некоторые формы со знаком ⟨sà⟩ в позиции перед /i/ и /u/.

Лексема	Засвидетельствованная форма	Источник
hwisar- «животное»	⟨(BESTIA)hwa/i-sà+ra/i-sa⟩	ALEPPO 2 §5
haristana/i- «балкон»	⟨(DOMUS)ha+ra/i-sà-tá-ni-zi⟩	KARKAMIŠ A11a §19
isnu(wa)- «устанавливать»	⟨(SOLIUM)i-sà-nu-wa/i-ha⟩	KARKAMIŠ A11a §20
asharimmis- «жертва»	⟨(*350)á-sa-ha+ra/i-mi-sà⟩	KARKAMIŠ A18b §18a
harnissa- «крепость»	⟨CASTRUM-sà⟩	KARKAMIŠ A12 §7
kistama/i- «восток»	⟨(DEUS.ORIENS)-ki-sà-ta-ma-si-zi⟩	TELL AHMAR 6 §5
karkamisa- «Каркемиш»	⟨ka+ra/i-ka-mi-sà(REGIO)⟩	KARKAMIŠ A4b §1
kinussa- «сжигать»	⟨DELERE-nu-sà-ha⟩	TELL AHMAR 6 §6
kusa- «нападение?»	⟨ku-sà-ti⟩	KARKAMIŠ A12 §7

Подобное использование ⟨sà⟩ именно в текстах 1000—850 гг. до н. э. должно, вероятно, объясняться тем, что в данную эпоху иероглифическая система письма достигла наиболее сложного и дифференцированного вида. В период после 850 г. до н. э. происходит упрощение, и знак ⟨sà⟩ начинает использоваться для /sa/, при этом другие знаки, исконно маркировавшие /sa/, начинают появляться в позициях, где мы бы ожидали ⟨sà⟩. В качестве примера смешения можно указать билингву KARATEPE §20, где одна и та же лексема /ussall(i)-/ «вор» в nom.pl.c. записывается как ⟨(“*271?”)u-sa-li₄-[zi]⟩ (Hu.) и как ⟨(*217)u-sà-lì-zi⟩ (Ho.). Разумеется, отбор репрезентативных памятников для анализа был бы также невозможен без опоры на корпус иероглифических текстов. Вместе с тем многие тексты, не попавшие в выборку Рикен по формальным параметрам, также могут показывать четкое распределение между ⟨sà⟩ и другими знаками серии ⟨sa⟩.

Фонемный статус /ʃ/ в лувийском языке предполагается на основании случаев с этимологическим сочетанием */sw/, например ⟨(LIGNUM)sà-la-ha-za⟩ (TELL AHMAR 1 §12 и др.), восходящее к и.-е. корню *swelH- «набухать» [LIV 2001: 698]. С точки зрения индоевропейской реконструкции интереснее, однако, случаи, собранные в таблице 10, которые являются частичной типологической параллелью для «правила RUKI», известного из индоиранских и балто-славянских языков. Если выйти за пределы обозначенного Рикен корпуса, можно сравнить императив ⟨sà-ti-nu-i⟩ «устрани, утоли» (ASSUR f §§4,8) и хеттскую основу *kistanu-* со сходным значением. Данную этимологию можно интерпретировать как аргумент в пользу действия лувийского варианта «правила RUKI» также и в позиции после этимологического *k.

Открытие Рикен позволяет задаться вопросом о том, не является ли «правило RUKI» общеиндоанатолийским архаизмом. Поскольку переход *s > š в позиции после r, u, k и i характерен для меньшинства индоевропейских групп, он традиционно рассматривается как балто-славянская и индоиранская инновация. Высказывалась также точка зрения, что данное нетривиальное развитие можно рассматривать как изоглоссу, объединяющую балто-славянские и индоиранские языки в одну подгруппу внутри позднеиндоевропейского [Ringe et al. 2002: 104]. Более того, можно утверждать, что «правило RUKI»

является единственной внешне убедительной общей инновацией балто-славянской и индоиранской групп, постулируемой в указанной работе («сатемное» развитие веллярных является тривиальным и могло происходить независимо в обеих группах, тогда как балто-слав. **wišo-* и ин.-ир. **wiśwo-* ‘весь’ не являются фонетически сопоставимыми). Однако если принять гипотезу о существовании «правила RUKI» в праиндоанатолийском языке как фонетического распределения между аллофонами [s] и [ʃ] внутри одной фонемы, то его выравнивание вполне могло происходить независимо в различных группах и даже языках. Исключением стали лишь те языки, где по различным причинам произошла фонемизация /ʃ/, т. е. языки индоиранской и балтославянской групп, а также лувийский. Данный сценарий является релевантным для проблемы индоевропейской филиации, поскольку устраняет необходимость постулировать балто-славо-арийскую языковую общность.

Таким образом, исследования по интерпретации анатолийских иероглифов могут иногда иметь неожиданные импликации для индоевропеистики в целом. Однако даже не беря в расчет эти результаты, можно констатировать, что хотя дешифровка иероглифических надписей из Малой Азии еще далека от завершения, она вступила в новую фазу, характеризующуюся последовательным применением корпусных методов. В качестве практического инструмента, облегчающего работу с анатолийскими иероглифическими текстами, хотелось бы обратить внимание на электронную версию корпуса [Hawkins 2000], доступную для автоматического поиска по нескольким параметрам (словоформы, лексемы, лексические и грамматические значения). Ресурс подготовлен в 2012—2013 гг. авторами настоящей статьи в сотрудничестве с А. С. Касьяном и Т. А. Архангельским и доступен по адресу: http://web-corpora.net/LuwianCorpus/search/?interface_language=ru.

Литература

- СТАРОСТИН, С. А., 1989, *Реконструкция древнекитайской фонологической системы*. Москва: Наука. [Starostin, S. A., 1989, *Rekonstruktsiya drevnekitajskoj fonologicheskoy sistemy*. Moskva: Nauka.]
- ADIEGO LAJARA, Ignacio-Javier, 2007, *The Carian Language*, HdO 1/86, Leiden.
- ALP, Sedat, 1950, *Zur Lesung von manchen Personennamen auf den hieroglyphenhethitischen Siegeln und Denkmälern*, Ankara.
- BARNETT, R. D., 1953, Karatepe, the key to the Hittite Hieroglyphs. *An. St.* 3: 53—95.
- BOSSERT, Helmuth Th., 1932, *Šantaš and Kupapa*, Leipzig.
- BURTON, R., DRAKE, T., 1872, *Unexplored Syria*, London.
- CARRUBA, Onofrio, 1982, Beiträge zum Luwischen, *Serta Indogermanica. Festschrift für Günter Neumann zum 60. Geburtstag*, ed. Tischler J., pp. 35—51, Innsbruck.
- CHADWICK, John, 1967, *The Decipherment of Linear B*, Cambridge.
- DANIELS, Peter, 1995, The Decipherment of Ancient Near Eastern Scripts, *Civilizations of the Ancient Near East*, ed. Sasson J. et al., pp. 81—93, New York.
- FORRER, Emil, 1932, *Die hethitische Bilderschrift*. Chicago.
- FRIEDRICH, Johannes, 1953, Zur Lesung der hethitischen Bilderschrift, *Ar. Or.* 21: 114—139.
- GELB, Ignace, 1931—1942, *Hittite Hieroglyphs*. Volumes: I, II and III, Chicago.
- GELB, Ignace, 1950, The Contribution of the new Cilician bilinguals to the decipherment of Hieroglyphic Hittite, *Bi. Or.* 7: 129—141.
- GOEDEGEBUURE, Petra, 2010, The Luwian Adverbs *zanta* “down” and **ānni* “with, for, against”, *Acts of the VIIth International Conference of Hittitology*, ed. Süel A., pp. 299—318, Ankara.
- GÜTERBOCK, H.G., 1949, Die Bedeutung der Bilinguen vom Karatepe für die Entzifferung der hethitischen Hieroglyphen, *Eranos* 47: 93—115.
- HAWKINS, John David, 1975, The Negatives in Hieroglyphic Luwian, *Anatolian Studies* 25: 119—156.

- HAWKINS, John David, 1995, *The Hieroglyphic Inscription of the Sacred Pool Complex at Hattusa (Südburg)*. (Studien zu den Boğazköy-Texten Beiheft 3), Wiesbaden.
- HAWKINS, John David, 2000, *Corpus of Hieroglyphic Luwian Inscriptions*. Volume I. Part I, II: Texts; Part III: Plates. Berlin — New York.
- HAWKINS, John David, Anna MORPURGO-DAVIES, Günter NEUMANN, 1974, Hittite Hieroglyphs and Luwian: New evidence for the connection, *Nachrichten der Akademie der Wissenschaften in Göttingen (Philologisch-historische Klasse)* 6: 145—197.
- HAWKINS, John David, Anna MORPURGO-DAVIES, 1975, Hieroglyphic Hittite: Some new readings and their consequences, *Journal of the Royal Asiatic Society* 2: 121—147.
- HAWKINS, John David, Anna MORPURGO-DAVIES, 1998, Tarkasnawa King of Mira, “Tarkondemos”, Boğazköy sealings and Karabel, *Anatolian Studies* 48: 1—31.
- KENT, Roland J., 1953, *Old Persian: Grammar, Texts, Lexicon*, New Haven.
- KLOEKHORST, Alwin, 2008, *Etymological Dictionary of the Hittite Inherited Lexicon*, Leiden.
- LAROCHE, Emmanuel, 1958, Comparaison du louvite et du lycien, *Bulletin de la Société Linguistique* 53: 159—97.
- LAROCHE, Emmanuel, 1960a, Comparaison du louvite et du lycien, *Bulletin de la Société Linguistique* 55: 155—85.
- LAROCHE, Emmanuel, 1960b, *Les hiéroglyphes hittites*, Paris.
- LAROCHE, Emmanuel, 1967, Comparaison du louvite et du lycien. *Bulletin de la Société Linguistique* 55: 155—85; 62: 46—66.
- RIX, Helmut, et al. (eds.), 2001, *Lexicon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen*, Wiesbaden.
- MELCHERT, H. Craig, 1993, *Cuneiform Luwian Lexicon*. Chapel Hill.
- MELCHERT, H. Craig, 1994, *Anatolian Historical Phonology*. Amsterdam-Atlanta.
- MELCHERT, H. Craig, 2013, Hittite and Hieroglyphic Luwian *arha* ‘away’: Common Inheritance or Borrowing, *Journal of Language Contact* 6: 300—312.
- MERIGGI, Piero, 1951, La bilingue di Karatepe in Cananeo e geroglifici etei. *Athenaeum* 29:25—99.
- MERIGGI, Piero, 1962, *Hieroglyphisch-hethitisches Glossar*. Second, fully revised edition. Wiesbaden.
- MESSERSCHMIDT, Leopold, 1900—1906, *Corpus Inscriptionum Hettitarum*, Berlin.
- MORPURGO-DAVIES, Anna, 1987, “To put” and “to stand” in the Luwian Languages, *Studies in Memory of Warren Cowgill*, ed. Watkins C., pp. 205—228, Berlin — New York.
- MORPURGO-DAVIES, Anna, 1982/1983, Dentals, rhotacism and verbal endings in the Luwian languages. *Zeitschrift für vergleichende Sprachforschung* 96: 245—270.
- PERROT, Georges, 1862, *Exploration archéologique de la Galatie et de la Bithynie*. Paris.
- PLÖCHL, Reinhold, 2003, *Einführung ins Hieroglyphen-Luwische (DBH 8)*. Dresden.
- POPE, Maurice, 1999, *The Story of Decipherment: From Egyptian Hieroglyphs to Maya Script*, London.
- RIEKEN, Elisabeth, 2008, Die Zeichen ⟨ta⟩, ⟨tá⟩ und ⟨tà⟩ in den hieroglyphen-luwischen Inschriften der Nachgrossreichzeit, *VI Congresso Internazionale di Ittologia, Roma, 5—9 settembre 2005, Parte II (Studi Micenei ed Egeo-Anatolici 50)*, eds. Archi A., Francia R., pp. 637—648. Roma.
- RIEKEN, Elisabeth, 2010, Das Zeichen ⟨sà⟩ im Hieroglyphen-luwischen, *Acts of the VIIth International Congress of Hittology, Çorum, August 25—21, 2008. Volume II*, pp. 651—660, Ankara.
- RIEKEN, Elisabeth, Ilya YAKUBOVICH, 2010, The New Values of Luwian Signs L 319 and L 172, *Luwian and Hittite Studies Presented to J. David Hawkins on the Occasion of his 70th Birthday*, ed. Singer I., pp. 199—219, Tel Aviv.
- RINGE, Don, Tandy WARNOW, and Ann TAYLOR, 2002, Indo-European and Computational Cladistics, *Transactions of the Philological Society* 100/1: 59—129.
- SAYCE, Archibald Henry, 1876, The Hammathite Inscriptions, *TSBA* 5: 22—32.
- SAYCE, Archibald Henry, 1880a, The Monuments of the Hittites, *TSBA* 7: 248—293.
- SAYCE, Archibald Henry, 1880b, The Bilingual Hittite and Cuneiform Inscription of Tarkondimos, *TSBA* 7: 294—308.
- STARKE, Frank, 1985, *Die keilschrift-luwischen Texte in Umschrift (StBoT 30)*, Wiesbaden.
- STARKE, Frank, 1990, *Untersuchung zur Stammbildung des keilschrift-luwischen Nomens*, Wiesbaden.
- VAN DEN HOUT, Theo, 2002, Self, Soul and Portrait in Hieroglyphic Luwian, *Silva Anatolica. Anatolian Studies Presented to Maciej Popko on the Occasion of his 65th Birthday*, Ed. Taracha P., pp. 171—186, Warsaw.
- YAKUBOVICH, Ilya, Hittite-Luwian Bilingualism and the Origin of Anatolian Hieroglyphs, *Acta Linguistica Petropolitana* 4: 9—36.

- YAKUBOVICH, Ilya, 2012, The Reading of Luwian *ARHA* and Related Problems, *Altorientalische Forschungen* 39: 321–339.
- YOSHIDA, Kazuhiko, 1993/1994, Notes on the Prehistory of Preterite Verbal Endings in Anatolian, *Historische Sprachforschung* 106: 26–35.

Sergey BORODAY, Ilya YAKUBOVICH. Corpus-based methods for deciphering Anatolian hieroglyphics.

The Anatolian hieroglyphs were used in Asia Minor and Syria for writing mostly Luwian texts between 14–8th centuries BC. Because of a relatively small corpus of the Anatolian hieroglyphic inscriptions and the limited number of bilinguals, their decipherment has not been fully completed up to now. Nonetheless, significant progress has been made in this area in the recent years following the publication of the first volumes the Corpus of Hieroglyphic Luwian Inscriptions at the very end of the last millennium. Many recent studies made in in this area are based on the application of formal methods and do not rely on new textual data. In this survey we are going to dwell on three groups of discoveries made in the last dozen or so years: (1) discrimination between logograms, phonetic complements, and phonetic indicators, (2) phonetic interpretation of logograms and (3) refining phonetic values of syllabograms. The authors have endeavoured to demonstrate that the ongoing work on interpreting the Anatolian hieroglyphs has broader repercussions for Indo-European comparative linguistics.

Keywords: Anatolian hieroglyphs, Luwian language, decipherment, corpus linguistics.

К формальной генеалогической классификации лезгинских языков (Северный Кавказ)*

В статье предлагается лексикостатистическая классификация 20 языков и диалектов лезгинской группы северокавказской семьи, выполненная на основе высококачественных 110-словных списков проекта «Глобальная лексикостатистическая база данных / The Global Lexicostatistical Database». К лексическому материалу последовательно применяются основные филогенетические методы, как дистантные, так и дискретные: метод ближайших соседей, реализованный в программе Starling (Starling neighbor joining), стандартный метод ближайших соседей (Neighbor joining), метод попарного внутригруппового невзвешенного среднего (Unweighted pair group method with arithmetic mean), метод Монте-Карло с цепями Маркова (Markov chain Monte Carlo), метод невзвешенной максимальной бережливости (Unweighted maximum parsimony). Все методы, кроме последнего, породили деревья, достаточно совместимые между собой, чтобы на их основе можно было составить сводное филогенетическое дерево лезгинских языков. Полученное сводное дерево согласуется с традиционной и некоторыми предшествующими формальными классификациями этой языковой группы. Вопреки теоретическим ожиданиям, метод максимальной бережливости предложил наименее правдоподобное дерево из всех.

Ключевые слова: генеалогическая классификация языков, лексикостатистика, филогения, лезгинские языки.

Данные

В рамках международного научного проекта «Глобальная лексикостатистическая база данных / The Global Lexicostatistical Database»¹ были составлены 110-словные списки базисной лексики для 20 языков и диалектов лезгинской группы (северокавказская языковая семья): удинский (2 диалекта), арчинский, крызский (2 диалекта), будухский, цахурский (3 диалекта), рутульский (3 диалекта), агульский (5 диалектов), табасаранский (2 диалекта), лезгинский, см. [Kassian 2011—2012]. Это максимальное количество лезгинских идиомов, для которых можно составить сводешевский список без полевой работы. Двадцать рассматриваемых списков полностью соответствуют лексикографическим стандартам проекта «Глобальная лексикостатистическая база данных».

* Я выражаю искреннюю благодарность Валерию Запорожченко (Москва) и Йоханну-Маттису Листу (Johann-Mattis List; Марбург) за консультации по компьютерным программам филогенетического анализа, а также Дмитрию Лещинеру (Москва) за консультации по ряду математических вопросов. Кавказоведческие аспекты статьи обсуждались с Тимуром Майсаком (Москва). Кроме того, данная работа вряд ли была бы возможна без бесед на смежные или более широкие темы с Георгием Старостиным, Михаилом Живловым, Анной Дыбо, Филиппом Минлосом и другими моими коллегами по московской школе компаративистики.

¹ <http://starling.rinet.ru/new100/main.htm>

1) Отбор слов производился в соответствии с семантическими спецификациями сводешевского списка, предложенными в [Kassian et al. 2010].

2) Были использованы фактически все релевантные источники по соответствующим языкам — словари, грамматики, собрания текстов, — причем не только современные публикации, но и материалы П. К. Услара, А. Дирра, А. Старчевского и других кавказоведов кон. XIX — нач. XX в.

3) Все языковые формы были единообразно транскрибированы фонетическим алфавитом, базирующимся на системе IPA; формы в традиционных кириллических орфографиях приводятся в скобках.

4) Лексические списки аннотированы. При языковых формах дается ссылка на источник, а в примечаниях эксплицитно обсуждаются существенные фонетические, морфологические и семантические особенности используемых форм и их синонимов. Также в примечаниях цитируются лексические данные из языков и диалектов, опубликованные материалы по которым недостаточны для составления полноценных 110-словных сводешевских списков (например, удинский список сопровождается формами из кавказско-албанских палимпсестов).

В полученных таким образом синхронных списках когнации размечались по этимологическому принципу. Я основывался на пралезгинской реконструкции С. А. Старостина [Starostin & Nikolayev 1994: 122 ff.; S. Starostin 1994; С. Старостин б. д.], по необходимости внося в нее определенные уточнения и изменения, см. [Kassian 2011—2012]. Публикации С. А. Старостина — это единственная на сегодняшний день обнародованная полноценная фонетическая и лексическая реконструкция лезгинского праязыка. Не так давно немецкий кавказовед В. Шульце [Schulze 1988; 2001; Gippert et al. 2008] объявил о разработке своей версии пралезгинской реконструкции. Совокупность лексических этимологий, уже опубликованных в работах В. Шульце, пока недостаточна для окончательных выводов, но я вынужден отметить, что многие диахронические идеи и решения В. Шульце не кажутся мне удачными или приемлемыми.

Для укоренения деревьев в сравнение был введен 110-словный список чеченского литературного языка [G. Starostin 2011]. Чеченский язык был выбран, с одной стороны, как родственник лезгинским в рамках северокавказской семьи, а с другой — как заведомо не входящий в лезгинскую группу. Этимологическое сопоставление чеченского списка с лезгинскими проводилось по [Starostin & Nikolayev 1994] с уточнениями из [G. Starostin 2011].

Метод

Генеалогические деревья строились несколькими методами.

1. Модифицированный метод ближайших соседей, разработанный С. А. Старостиным для лексикостатистического анализа и реализованный в программе Starling (метод Starling neighbor joining, далее — StarlingNJ). См. описание StarlingNJ в [Бурлак & Старостин 2005: 163 сл.] (в настоящей статье используется порог, после которого начинается усреднение значений, не в 70%, а в 75%, что является значением по умолчанию в последних версиях Starling). Дерево строилось в программе Starling (v. 2.5.3; см. [С. Старостин 1993/2007; Бурлак & Старостин 2005: 270 ff.]) из лексикостатистической базы данных, представляющей собой многозначную матрицу с возможностью синонимии (узлы датировались так называемым «экспериментальный методом», при котором сводешевским словам присваиваются индивидуальные индексы стабильности, [С. Старостин 2007а;

G. Starostin 2010)). Дерево укоренялось методом иерархической кластеризации, что стандартно для алгоритма StarlingNJ. Для данных, обработанных в программе Starling, приводятся деревья двух типов: дерево, допускающее только бифуркацию (как того требует метод ближайших соседей), и это же дерево, но в котором соседние узлы сведены в один, если временной промежуток между ними составляет 300 или менее лет² (300 лет соответствуют замене приблизительно в полтора слова в каждом из двух идиомов). Датировки узлов определялись по принципу строгих молекулярных часов, подробнее об этом и о калибровке шкалы см. [С. Старостин 1989/2007; S. Starostin 1999/2000]. Деревья визуализировались в программе Starling.

2. Стандартный метод ближайших соседей (Neighbor joining, далее NJ), см. [Saitou & Nei 1987; Makarenkov et al. 2006: 65–66]. Дерево строилось в программе SplitsTree4 (v. 4.13.1, см. [Huson & Bryant 2006]) из лексикостатистической бинарной матрицы формата NEXUS, которая была получена из многозначной путем кодирования каждого задействованного в списке пракорня (всего 481 пракорень) как присутствующего («1») или отсутствующего («0») в данном сводешевском значении в соответствующем языке; «?» значит, что в данном языке данное сводешевское значение выражается через иноязычное заимствование или же выражение для него не известно (не найдено в источниках или отсутствует в языке). Для оценки устойчивости топологии дерева использовался непараметрический бутстреп-тест: 10 000 реплик. Дерево укоренялось через внешнюю группу, т. е. через чеченский список. Дерево не датированное. Дерево визуализировалось в программе FigTree (v. 1.4.0). Также в SplitsTree было построено дополнительное дерево методом BioNJ [Gascuel 1997], которое оказалось идентичным дереву NJ.

3. Метод попарного внутригруппового невзвешенного среднего (Unweighted pair group method with arithmetic mean, далее — UPGMA), см. [Sneath & Sokal 1973: 230–234; Makarenkov et al. 2006: 65–66]. Дерево строилось в программе SplitsTree4 (v. 4.13.1) из бинарной матрицы, описанной выше. Для оценки устойчивости топологии дерева использовался непараметрический бутстреп-тест: 10 000 реплик. Дерево укоренялось через внешнюю группу, т. е. через чеченский список. Дерево не датированное. Дерево визуализировалось в программе FigTree (v. 1.4.0).

4. Метод Монте-Карло с цепями Маркова (Markov chain Monte Carlo, далее — MCMC; см. [Makarenkov et al. 2006: 68–69]), впервые примененный в рамках байесовского подхода к лингвистическим данным в [Gray & Atkinson 2003]. Дерево строилось в программе MrBayes (v. 3.2.1, см. [Huelsenbeck & Ronquist 2001]) из бинарной матрицы, описанной выше. Программа запускалась 4 раза, каждый раз с 4 цепочками; чеченский список был указан как внешняя группа. При каждом запуске порождались 5 000 000 случайных конфигураций дерева (поколений), из них каждое 500-е дерево сэмплировалось; при каждом запуске первые 25% деревьев использовались для отжига и исключались из дальнейшего анализа. Дерево укоренялось через внешнюю группу, т. е. через чеченский список. Дерево не датированное. Дерево визуализировалось в программе FigTree (v. 1.4.0).

5. Метод невзвешенной максимальной бережливости (Unweighted maximum parsimony, далее — UMP), см. [Makarenkov et al. 2006: 66–67]. Деревья строились в программе TNT (Willi Hennig Society edition of TNT, v.1.1, 08 May 2013, см. [Goloboff et al. 2008]) из бинарной матрицы, описанной выше, по строгому принципу ветвей и границ (Branch &

² На нынешнем этапе разработки метода затруднительно строго определить погрешность датирования, но представляется, что интервал погрешности вряд ли составляет менее чем несколько сотен лет. Также отмечу, что дистантные методы типа StarlingNJ, NJ, UPGMA в принципе способны порождать только бинарные деревья, и это само по себе предполагает, что близкие узлы логично объединять в общий узел.

bound / Implicit enumeration). Принудительная бинаризация узлов была запрещена (Collapse trees after the search); чеченский список был указан как внешняя группа. Было получено 4 оптимальных дерева одинаковой стоимости, на их основе было построено строгое консенсусное дерево. Для оценки устойчивости топологии консенсусного дерева использовался непараметрический бутстреп-тест: 1000 реплик. Деревья укоренялись через внешнюю группу, т. е. через чеченский список. Деревья не датированные. Деревья визуализировались в программе FigTree (v. 1.4.0).

Результаты

Были получены следующие деревья:

- рис. 1a, метод StarlingNJ без объединения узлов;
- рис. 1b, метод StarlingNJ с объединением близких узлов;
- рис. 2, метод NJ;
- рис. 3, метод UPGMA;
- рис. 4, метод МСМС;
- рис. 5, метод UMP;
- рис. 6, сводное дерево, составленное вручную.

Если исключить дерево UMP (о чем см. ниже), разница между деревьями по большей части не представляется принципиальной. Прокомментируем расхождения.

1) Все дистантные методы, т. е. StarlingNJ, NJ, UPGMA (рис. 1a, 2, 3), предполагают последовательные бифуркации с отделением сначала удинской ветви, затем арчинской и соответствующим обособлением узколезгинского (самурского) праязыка. Дистанция между двумя узлами (отделение удинского и отделение арчинского), однако, минимальна на всех деревьях, что хорошо видно из графического представления деревьев и вероятностной оценки узлов, и при введении временной погрешности в 300 лет в методе StarlingNJ (рис. 1b) первичное разделение пралезгинского языка оказывается тернарным: удинский, арчинский и узколезгинский. Напротив, используемый дискретный метод (МСМС, рис. 4) сразу предлагает тернарное разделение на удинский, арчинский и узколезгинский. Следует помнить, что дистантные методы StarlingNJ, NJ, UPGMA в принципе способны породить только бинарные деревья. Дерево UMP здесь существенно расходится с остальными деревьями, см. ниже.

2) Все методы дают принципиальное членение узколезгинской подгруппы на три ветви: (1) западнолезгинскую (цахурский и рутульский языки); (2) южнолезгинскую (крызский и будухский языки); (3) восточнолезгинскую (агульский, табасаранский и лезгинский языки). Разница обнаруживается в иерархии членения. Методы StarlingNJ и NJ (рис. 1a, 2), а также UMP (рис. 5) указывают на первое отделение западнолезгинской ветви и последующую бифуркацию на южнолезгинскую и восточнолезгинскую ветви. Метод UPGMA (рис. 3) указывает на первое отделение южнолезгинской ветви. Наконец, метод МСМС (рис. 4) указывает на первое отделение восточнолезгинской ветви. Дистанция между двумя узлами (последовательные бифуркации между западнолезгинским, южнолезгинским и восточнолезгинским праязыками), однако, минимальна на всех деревьях, что хорошо видно из графического представления деревьев и вероятностной оценки узлов, и при введении временной погрешности в 300 лет в методе StarlingNJ (рис. 1b) разделение узколезгинского праязыка оказывается тернарным: западнолезгинский, южнолезгинский и восточнолезгинский.

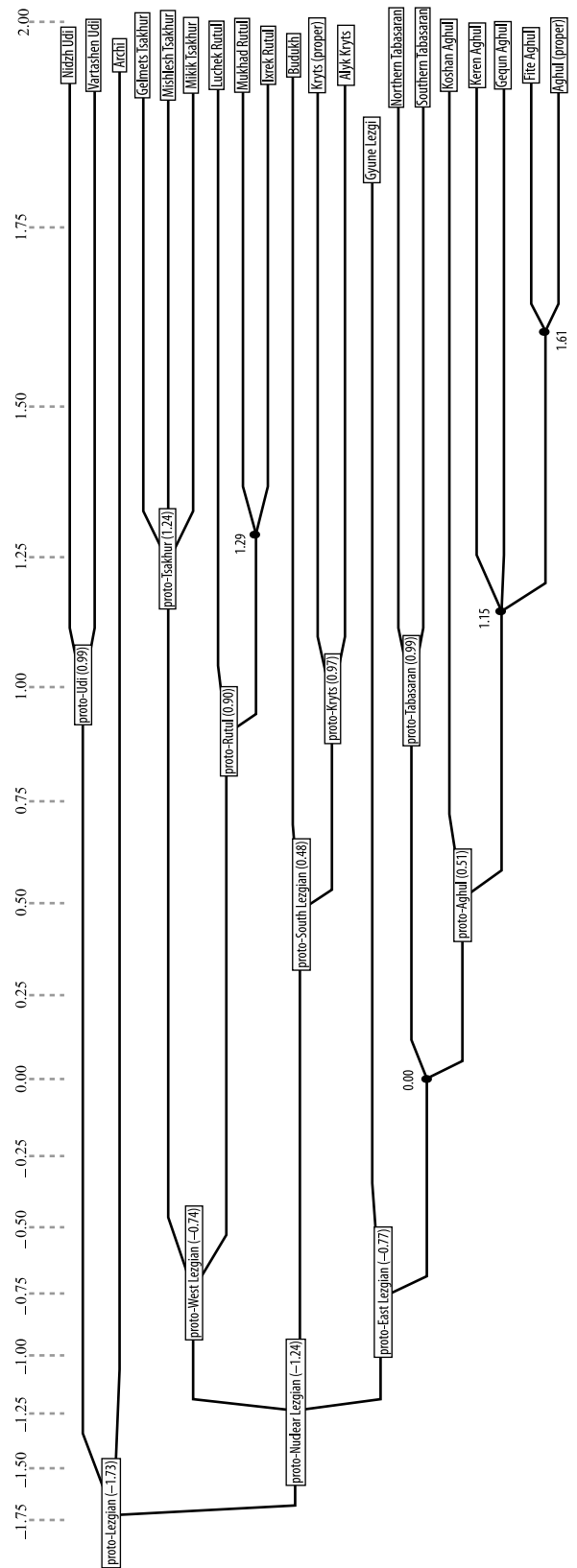
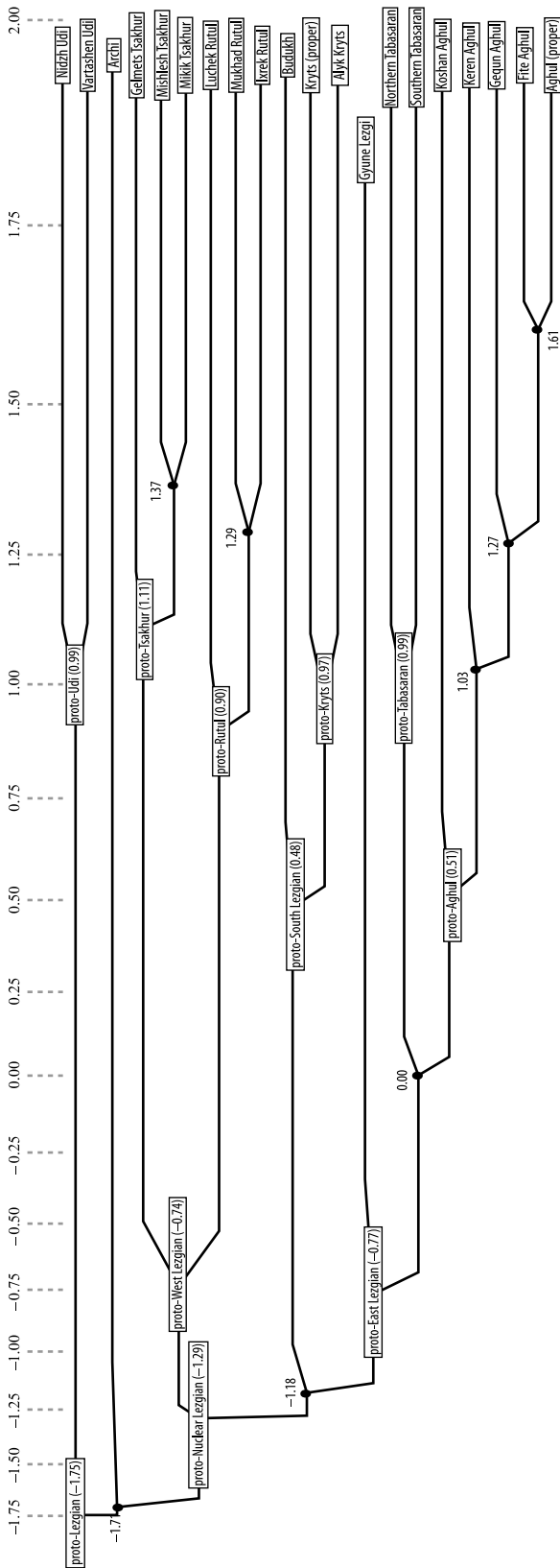


Рис. 1а (слева). Филогенетическое дерево лезгинских идиомов, полученное методом StarlingNJ в Starling без объединения близких узлов из многозначной матрицы. Даты даны в тысячах лет (например, -1.72 означает 1720 г. до н. э.).

Рис. 1б (справа). Филогенетическое дерево лезгинских идиомов, полученное методом StarlingNJ в Starling с объединением близких узлов (дистанция 300 или менее лет) из многозначной матрицы. Даты даны в тысячах лет (например, -1.72 означает 1720 г. до н. э.).

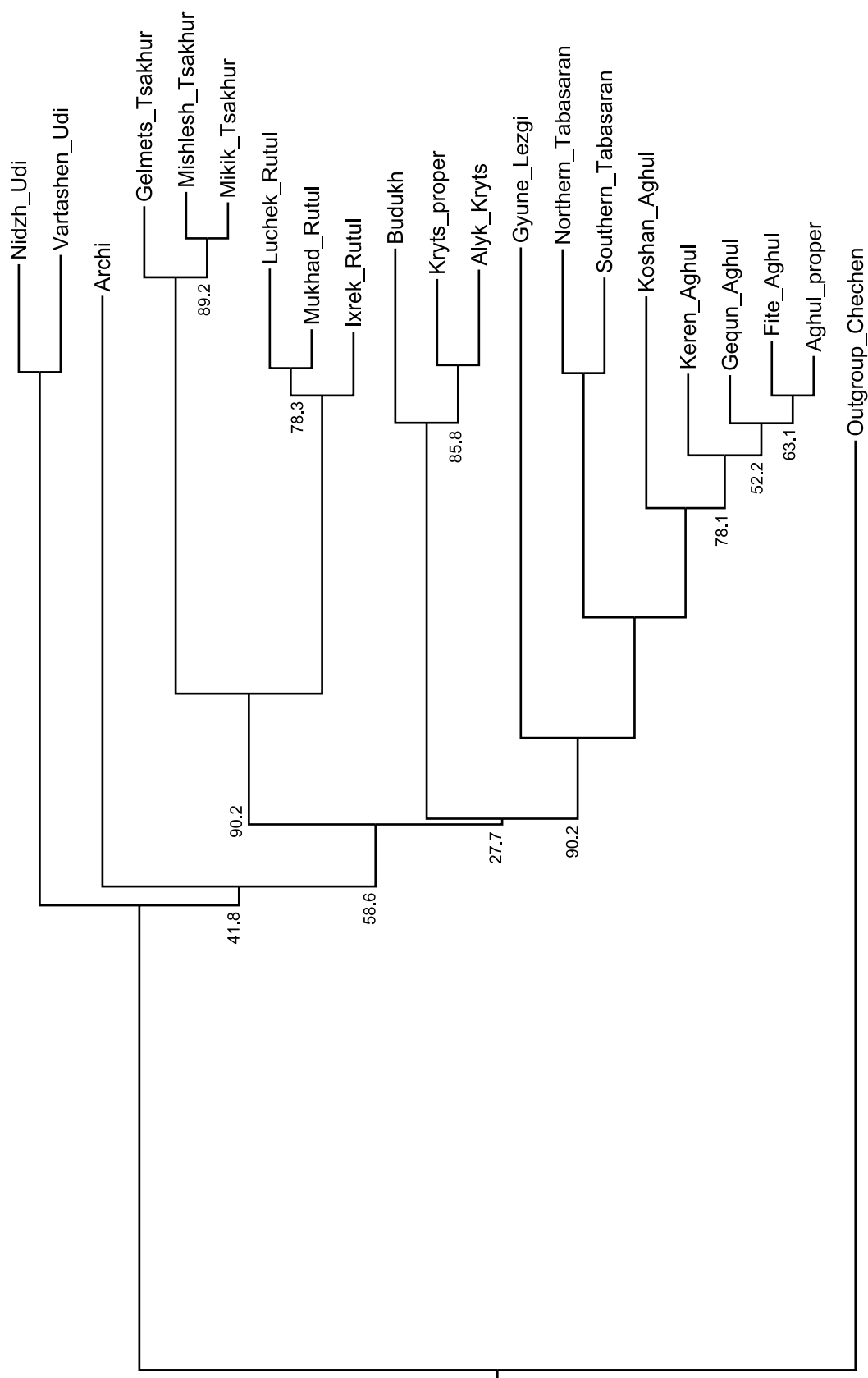


Рис. 2. Филогенетическое дерево лезгинских идиомов, полученное методом NJ в Splits-Tree4 из бинарной матрицы. Статистическая поддержка той или иной ветви в процентах реплик бутстрепа обозначена числом при соответствующей ветви (для сверхустойчивых ветвей с поддержкой $\geq 95\%$ этот параметр не указывается). Длина ветвей выражает относительное количество предполагаемых машиной лексических замен. Метод BioNJ дает идентичную топологию.

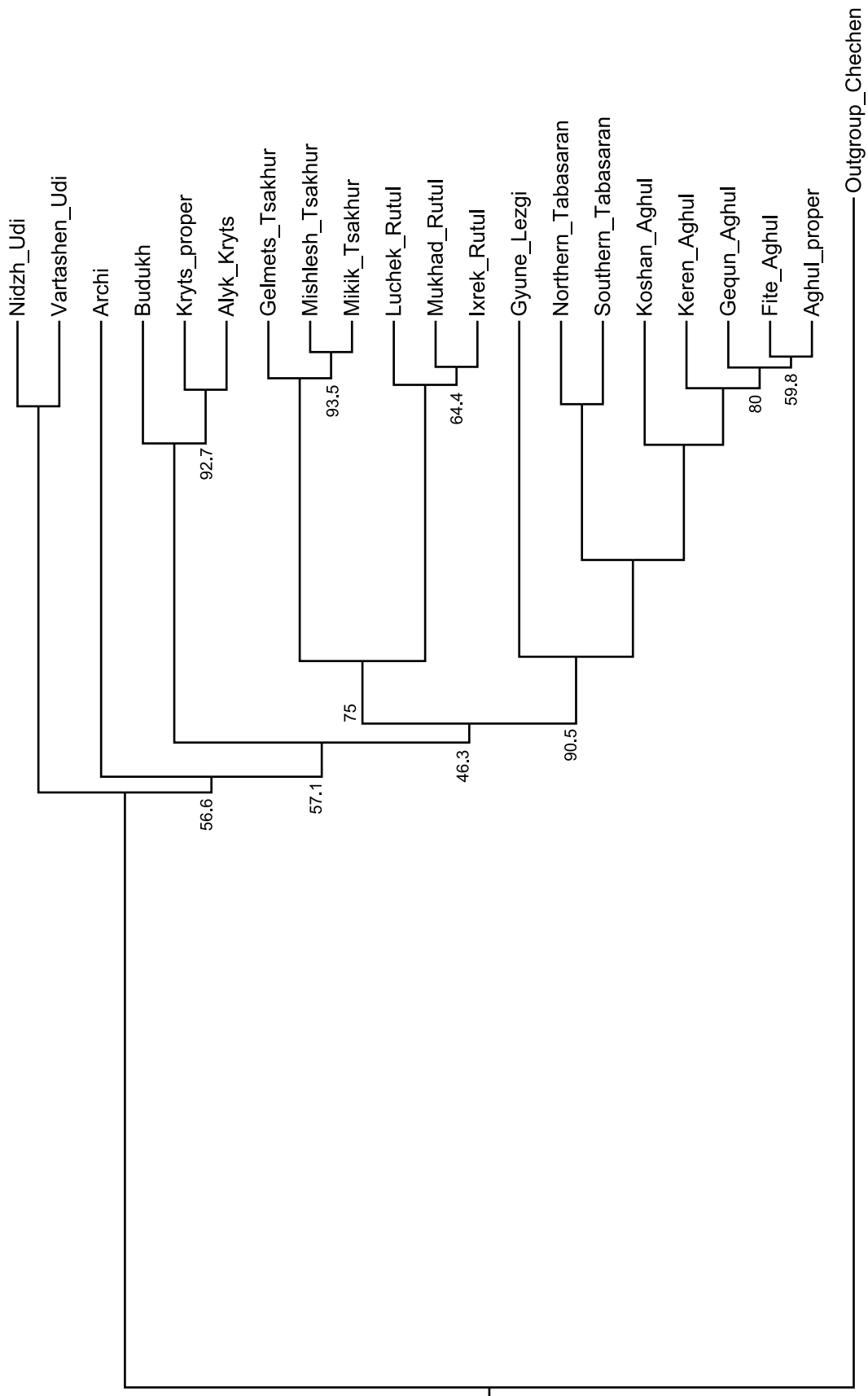


Рис. 3. Филогенетическое дерево лезгинских идиомов, полученное методом UPGMA в SplitsTree4 из бинарной матрицы. Статистическая поддержка той или иной ветви в процентах реплик бутстрапа обозначена числом при соответствующей ветви (для сверхустойчивых ветвей с поддержкой $\geq 95\%$ этот параметр не указывается). Длина ветвей выражает относительное количество предпологаемых машиной лексических замен.

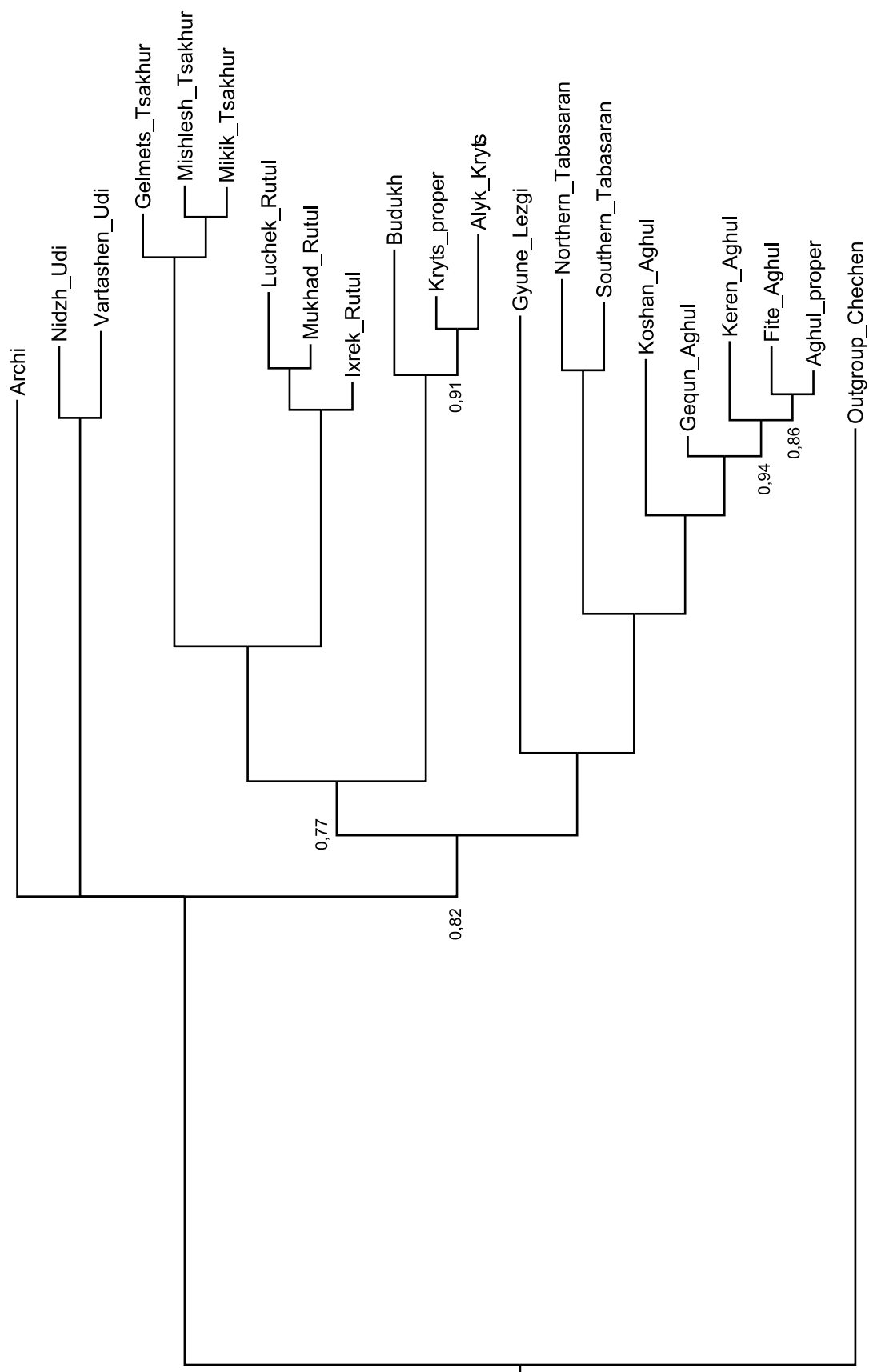


Рис. 4. Консенсусное филогенетическое дерево лезгинских идиомов, полученное методом MCMC в MrBayes из бинарной матрицы. Апостериорная вероятность той или иной ветви обозначена числом над соответствующей ветвью (вероятность для сверхустойчивых ветвей с $P \geq 0,95$ не указывается). Длина ветвей выражает относительное количество предполагаемых машиной лексических замен.

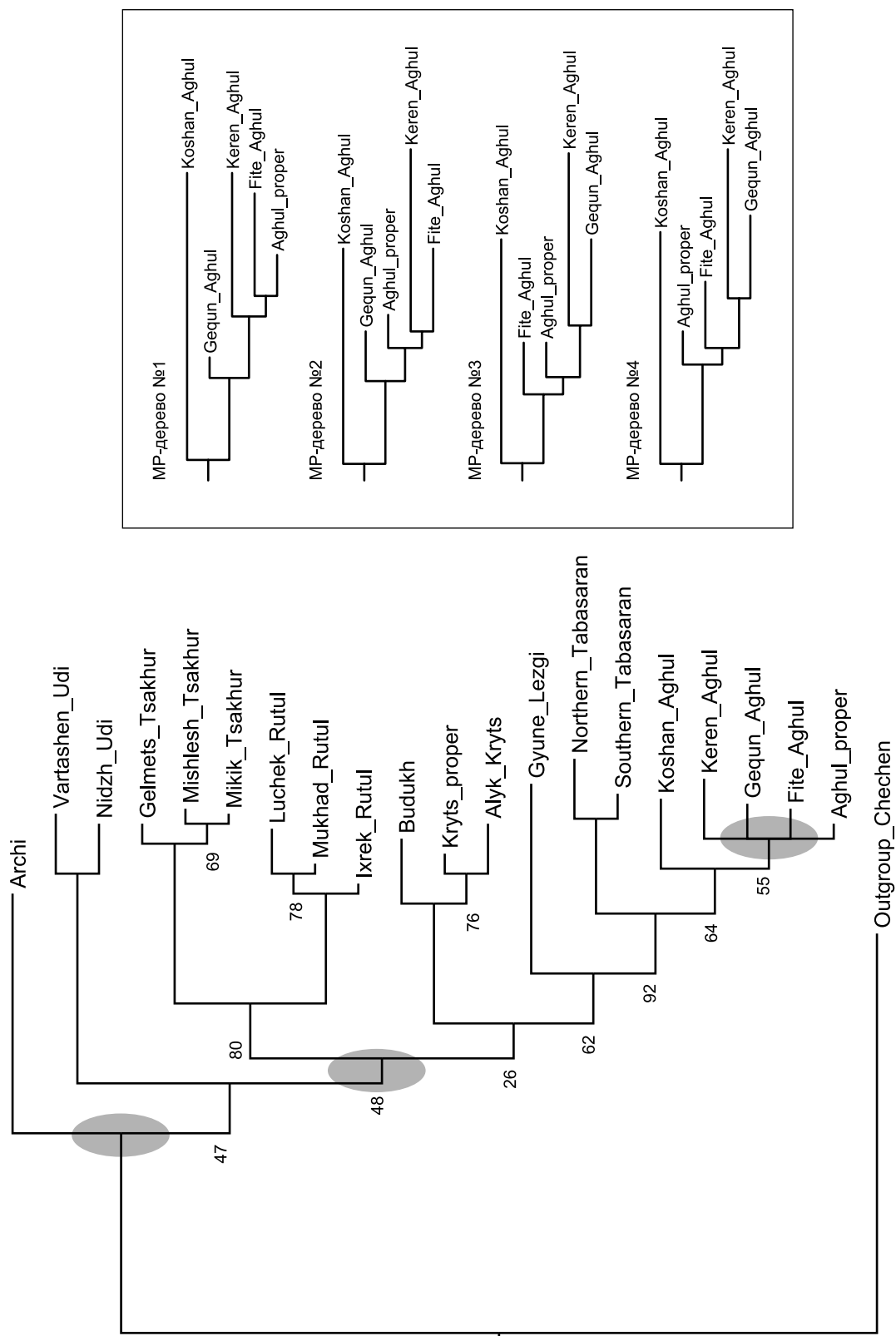


Рис. 5. Строгое консенсусное филогенетическое дерево лезгинских идиомов, полученное методом UMP в TNT из бинарной матрицы. Статистическая поддержка той или иной ветви в процентах реплик бутстрапа обозначена числом при соответствующей ветви (для сверхустойчивых ветвей с поддержкой $\geq 95\%$ этот параметр не указывается). Длина ветвей выражает относительное количество предполагаемых машиной лексических замен. Четыре оптимальных дерева имеют различия только в агульском узле, что продемонстрировано на врезке. Серым выделены фрагменты дерева, проблемные при сопоставлении с другими методами.

3) Агульские диалекты. Все методы реконструируют первичное отделение кошанского диалекта (что соответствует интуитивным ожиданиям), но далее начинают противоречить друг другу. Дистантные методы, т. е. StarlingNJ, NJ, UPGMA (рис. 1а, 2, 3), предполагают последующее отделение керенского диалекта и далее гехюнского диалекта, а используемый дискретный метод (МСМС, рис. 4) наоборот — сначала отделение гехюнского диалекта и затем керенского. Дистанция между двумя узлами (последовательные бифуркации между керенским, гехюнским и собственно агульским/фитинским), однако, минимальна на всех деревьях, что хорошо видно из графического представления деревьев, и при введении временной погрешности в 300 лет в методе StarlingNJ (рис. 1б) разделение праагульского после отделения кошанского диалекта оказывается тернарным: керенский, гехюнский и собственно агульский/фитинский. Дерево UMP здесь существенно расходится с остальными деревьями, см. ниже.

4) Пожалуй, наиболее серьезное расхождение между имеющимися деревьями (исключая дерево UMP) касается членения трех рутульских диалектов. Методы StarlingNJ и UPGMA (рис. 1а, 3) предполагают, что первым отделился лучекский диалект. Напротив, NJ и МСМС (рис. 2, 4) предполагают, что первым отделился ихрекский диалект. При этом на рис. 1а (StarlingNJ) узлы находятся хронологически достаточно далеко друг от друга и не объединяются в один при введении временной погрешности в 300 лет (рис. 1б). Как видно из таблиц дистанций, и для многозначной, и для бинарной матрицы в рутульской части дерева получаемые лексикостатистические дистанции не удовлетворяют условию аддитивности: табл. 1, 2. При постулате постоянной скорости изменения сводешевского списка рутульская ситуация является аномальной, и разные методы классификации предлагают в таком случае разные решения. Лингвистически рутульские данные могут объясняться двумя возмущающими факторами: (1) междиалектными заимствованиями и контактно обусловленной гомоплазией (выявить такие случаи пока не представляется возможным); (2) несовершенством имеющихся лексикографических описаний, не позволяющих составить сводешевские списки более аккуратно. Сложно сказать, какая из двух топологий рутульских диалектов адекватнее соответствует исторической реальности (ср. [Ибрагимов 1978: 15]), однако в любом случае математические методы классификации вряд ли в полной мере применимы к ситуации взаимопонятных контактирующих идиомов (диалектного континуума), как то наблюдается на рутульской территории.

Табл. 1. Таблица обратных дистанций для трех рутульских диалектов при многозначной матрице.

	Ихрекский	Лучекский
Мухадский	0,96	0,94
Ихрекский	—	0,91

Табл. 2. Таблица обратных дистанций для трех рутульских диалектов при бинарной матрице.

	Ихрекский	Лучекский
Мухадский	0,92	0,91
Ихрекский	—	0,87

5) Наконец, в изолированном положении оказывается метод UMP, который порождает дерево, слабо совместимое как с деревьями, полученными остальными методами, так и с нашими неформальными интуитивными представлениями о членении лезгинской

языковой группы (см. рис. 5, где проблемные узлы выделены серым). Во-первых, арчинский язык оказывается первым отделившимся кластером, а удинский, напротив, хоть формально и отделяется вторым, стремится объединиться со следующим узколезгинским (самурским) узлом. Во-вторых, не удается выстроить осмысленную консенсусную топологию агульских диалектов. В-третьих, бросается в глаза неустойчивость консенсусного дерева в некоторых принципиальных узлах (вроде отделения арчинского и удинского), где бутстреп-тест демонстрирует поддержку $< 50\%$. По этим причинам я был вынужден исключить дерево UMP из рассмотрения при составлении сводного лезгинского дерева (рис. 6).

С учетом вышеперечисленных расхождений, отбросив результаты анализа методом UPM, можно предложить сводное филогенетическое дерево лезгинских идиомов: рис. 6 (составлено вручную). На этом дереве объединены соседние узлы, (1) хронологическое расстояние между которыми ≤ 300 лет по подсчетам методом StarlingNJ (см. рис. 1a, 1b) или (2) топология которых зависит от используемого метода классификации. Серым цветом отмечены 4 объединенных тернарных узла, за которыми скрываются бинарные ветвления, различающиеся в зависимости от метода: три из этих узлов автоматически получаются при введении указанной временной погрешности, а четвертый узел — это рутульские диалекты, обсуждаемые выше. Как можно видеть, сводное дерево (рис. 6) идентично дереву StarlingNJ (см. рис. 1b), за исключением дополнительного объединения в тернарный узел трех рутульских диалектов.

Преыдушие классификации

Полученное сводное дерево лезгинских языков и диалектов (рис. 6) с двумя аутлайерами (удинский и арчинский) и многочисленной узколезгинской или самурской подгруппой, делящейся на три кластера (западный, южный, восточный), согласуется с такими предлагаемыми ранее филогенетическими реконструкциями лезгинской языковой группы:

1) Традиционная неформальная классификация, см., напр., [Талибов 1980: 11—16] с дальнейшей литературой.

2) Предшествующие более грубые лексикостатистические подсчеты, упомянутые в [Алексеев 1984: 91 сл.], на основе 100-словных списков, проэтимологизированных и обработанных методом типа UPGMA; списки не удовлетворяют современным критериям проекта «Глобальная лексикостатистическая база данных / The Global Lexicostatistical Database».

3) Формальная классификация в проекте «The Automated Similarity Judgment Program», см. [Müller et al. 2010], где используются непроэтимологизированные 40-словные списки, суммарное измерение расстояний Левенштейна между которыми дает матрицу дистанций между языками, из которой строится дерево методом NJ в программе MEGA 4³.

Напротив, предшествующие лексикостатистические классификации, согласно которым арчинский оказывается четвертым кластером внутри узколезгинской подгруппы, не подтверждаются и, видимо, должны быть отвергнуты. Речь идет о [Алексеев 1985: 17—23] (100-словные списки, проэтимологизированные и обработанные методом типа UPGMA) и [Коряков 2006: 21] (100 или 110-словные списки, проэтимологизированные и обработанные методом StarlingNJ в программе Starling); в обеих публикациях списки не удовлетворяют современным критериям проекта «Глобальная лексикостатистическая база данных / The Global Lexicostatistical Database».

³ Отметим, что в [Müller et al. 2010] в лезгинскую группу попал и хиналутский язык, что, видимо, неверно.

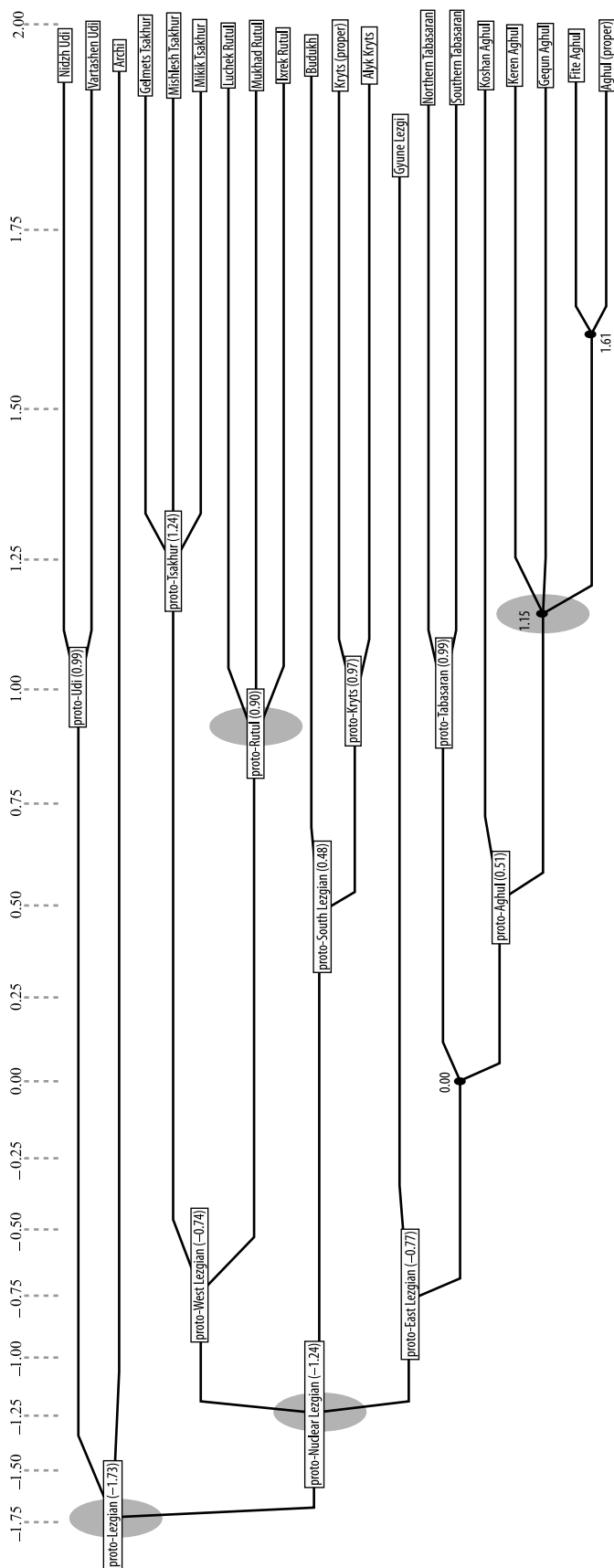


Рис. 6. Сводное дерево лезгинских идиомов, отражающее классификацию методами StarlingNJ, NJ, BioNJ, UPGMA, MCMC (но не UMP). Серым отмечены 4 объединенных тернарных узла, за которыми скрываются бинарные ветвления, различающиеся в зависимости от метода. Датировки даются по методу StarlingNJ в тысячах лет (например, -1.72 означает 1720 г. до н. э.).

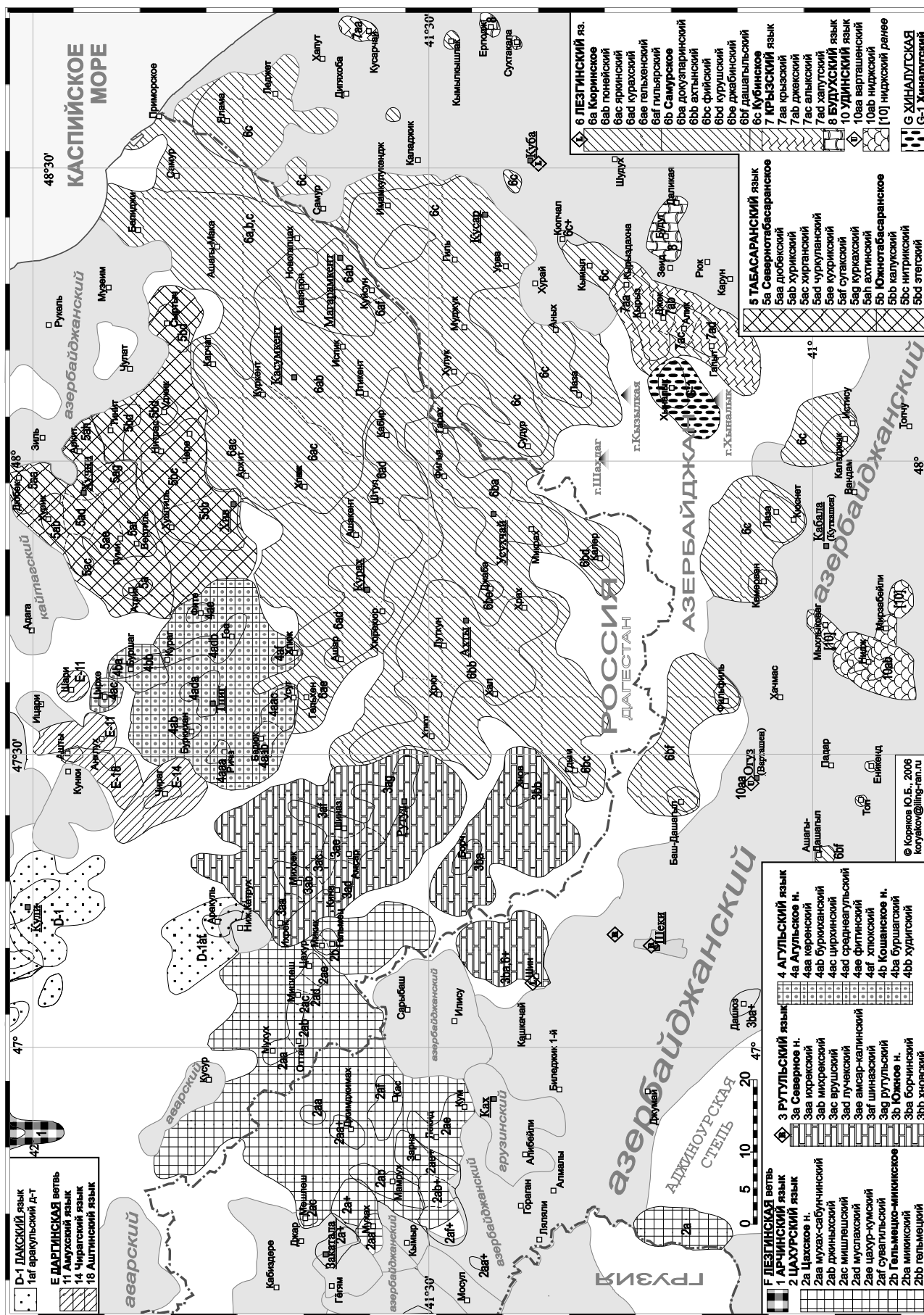


Рис. 7. Карта современного распределения лезгинских языков (по кн.: [Коряков 2006: карта № 13]).

Заведомо не находит поддержки идея В. Шульце [Schulze 2005; Gippert et al. 2008: II-65—75; против Schulze-Fürhoff 1994: 450] о том, что удинско-кавказско-албанская ветвь входит в восточнолезгинский кластер вместе с агульским, табасаранским и собственно лезгинским. В. Шульце [Gippert et al. 2008: II-65—75] опубликовал составленные им сводешевские списки для удинского и кавказско-албанского в сравнении с данными других лезгинских языков. К сожалению, В. Шульце не приводит никаких объяснений своей особой версии лексикостатистической процедуры, а в то же время лексикографическое качество его сводешевских списков весьма низко. Всё это позволяет заключить, что В. Шульце не смог представить какие-либо формальные аргументы в пользу своей филогенетической классификации. С неформальной, т. е. интуитивной точки зрения идея В. Шульце о месте удинско-кавказско-албанского внутри лезгинской группы также не представляется приемлемой.

Выводы

Как было частично указано выше, база данных с 110-словными лексическими списками лезгинских языков и диалектов [Kassian 2011—2012] обладает рядом важных свойств:

1) база включает в себя достаточно большое число идиомов: 20 единиц, причем среди них есть как языки, длительное время развивающиеся в изоляции, например арчинский, так и языки, активно контактирующие с другими языками данной группы (что потенциально дает подскок лексикостатистических совпадений благодаря контактно обусловленной гомоплазии), например, агульский;

2) не будет большим преувеличением сказать, что среди кавказоведов имеется консенсус относительно членения лезгинской группы (аутлайеры удинский и арчинский, отделившиеся первыми, плюс узколезгинская, или самурская, подгруппа, состоящая из трех кластеров: западного, южного и восточного);

3) лезгинскую группу, наверное, можно охарактеризовать как среднюю или чуть выше среднего по параметру надежности и подробности имеющихся лексикографических и грамматических описаний среди языковых групп мира;

4) общее качество, так сказать, степень «очистки» лезгинских лексикостатистических списков (равно как и списков других языков в проекте «Глобальная лексикостатистическая база данных / The Global Lexicostatistical Database») беспрецедентно высока для мировой лингвистики.

Всё это делает лезгинскую базу данных прекрасным полигоном для лингвистической апробации различных филогенетических методов.

В теоретической статье [Barbañon et al. 2013] симулированием различных лингвистических ситуаций сравнивается адекватность нескольких филогенетических методов. Авторы приходят к выводу, что по всем параметрам наиболее надежным является метод максимальной бережливости (MP), далее следует метод Монте-Карло с цепями Маркова (MCMC), затем метод ближайших соседей (NJ) и существенно менее точным методом оказывается метод попарного внутригруппового невзвешенного среднего (UPGMA). Оставляя в стороне некоторые спорные аспекты статьи⁴, можно видеть, что авторами явным

⁴ Одной из причин трудностей в применении метода MP с бинарными признаками к лингвистическим данным может быть неадекватность соответствующей модели представлениям об эволюции естественных языков. Метод MP сильнее других зависит от гомоплазии (т. е. от обратного или же параллельного развития), соответственно, для минимализации эффекта гомопластических возмущений [Barbañon et al.

образом отдают предпочтение дискретным методам (MP, MCMC) перед дистантными (NJ, UPGMA), и это, в принципе, является главным выводом публикации [Barbaçon et al. 2013]⁵. В качестве количественной оценки в [Barbaçon et al. 2013: 166] предлагается считать, что все протестированные методы, кроме UPGMA, реконструируют порядка 90% ребер истинного с исторической точки зрения дерева.

Эксперименты с лезгинской лексикостатистической базой данных показывают, однако, еще более отрадную картину, если учесть, что каждое из ребер «истинного» дерева реконструировано хотя бы одним из методов, исключая UMP (т. е. отражено хотя бы на одном из деревьев на рис. 1–4). При введении небольшого доверительного интервала (и объединении соседних узлов, попавших в него, см. рис. 6 и комментарии к нему) видно, что протестированные методы StarlingNJ, NJ, UPGMA, MCMC противоречат друг другу только в иерархии трех рутульских диалектов. На сводном дереве на рис. 6 всего 33 ребра плюс, если мы не будем объединять рутульские диалекты в тернарный узел, дополнительное 1 ребро, таким образом 34 ребра. Расхождение между методами в членении рутульских диалектов дает ошибку в 1 ребре из 34, и это предполагает, что все методы (за исключением UMP) правильно реконструировали от 97% до 100% ребер. Неожиданным результатом лезгинского теста оказалась невысокая правдоподобность дерева, полученного методом максимальной бережливости (UMP, рис. 5), что прямо противоречит выкладкам [Barbaçon et al. 2013].

Рассмотренные лезгинские данные подтверждают некоторые положения, составляющие идеологическую основу проекта «Глобальная лексикостатистическая база данных / The Global Lexicostatistical Database»:

1) При классификации языков лучше воздерживаться от использования грамматических (фонетических, морфологических, синтаксических) признаков⁶, т. к., во-первых, эти признаки не универсальны; во-вторых, они легко могут образовывать вторичные ареальные изоглоссы (особенно если речь идет о языках, чье родство еще ощущается носителями), причем выявить источник инновации часто оказывается затруднительно; в-третьих, грамматические признаки образуют систему, т. е. изменение одного признака

2013] предлагают вводить индивидуальные веса признаков (Weighted maximum parsimony), присваивая больший внутренний вес признакам, не демонстрирующим гомоплазию на данном лингвистическом материале. На бинарных данных с одинаковым весом перехода между состояниями наличие гомоплазии (параллельное или же обратное развитие какого-либо признака) математически эквивалентно присутствию в исходной матрице двух несовместимых (incompatible) признаков, т. е. принимающих все четыре возможные пары состояний «00», «01», «10» и «11» (см., напр., [Semple & Steel 2003: 69 сл.]). И использованный в [Barbaçon et al. 2013] метод MP трактует переходы между состояниями «0» и «1» как равнозначные. Но если мы работаем, как в данной статье, с бинарной матрицей, где «1» значит маркированное значение признака, а «0» — немаркированное (например, «1» = наличие, а «0» = отсутствие данного пракорня в данном сводешевском значении в данном языке), то переход $1 > 0$ (утрата корня) не является значимым событием, он может происходить параллельно и независимо в разных языках, и такую параллельную утрату лингвист вряд ли должен считать гомопластичной. Таким образом, чтобы обнаружить факт лингвистической гомоплазии, надо произвести реконструкцию значений используемых признаков для праязыка исследуемой языковой группы, что совсем не тривиальная теоретическая и практическая задача, в частности реконструкция невозможна без знания филогенетического дерева [Kassian 2013], — мы попадаем в порочный круг.

⁵ Схожий взгляд на иерархию точности классификационных методов постепенно возобладает и в молекулярной биологии.

⁶ Ср., например, [Nakhleh et al. 2005], где различные методы филогенетического анализа тестируются на материале индоевропейской семьи, при этом на вход подаются как лексические, так и грамматические признаки. Аналогично смешанный набор лексических и грамматических признаков используется в моделировании [Barbaçon et al. 2013].

с высокой вероятностью влечет за собой изменение других признаков. Для лексических же признаков эти недостатки характерны в значительно меньшей степени.

2) Точность филогенетического дерева зависит в первую очередь не от математического метода, а от степени очистки входных данных, иными словами, не от труда компьютера, а от труда лингвиста, кропотливо анкетизирующего индивидуальные диалекты по принятому списку признаков (хотя отдельные филогенетические методы, вроде максимальной бережливости / *maximum parsimony*, вызывают сомнения).

Дополнительные материалы по исследованию доступны по адресам:

- <http://jolr.ru/article.php?id=133>
- http://starling.rinet.ru/~kass/Lezgian_classification_RUS_2013.zip

Архив включает в себя:

- lez.xls, многозначная матрица в формате MS Excel;
- lez.nex, бинарная матрица в формате NEXUS;
- lez.tnt, бинарная матрица в формате NEXUS для программы TNT;
- lez-reverse-distances-multistate.xls, таблица обратных расстояний, полученная из многозначной матрицы в программе Starling;
- lez-distances-binary.txt, таблица расстояний, полученная из бинарной матрицы в программе Splits-Tree4;
- *.tre, некоторые обсуждаемые в статье деревья в формате NEWICK.

Литература

- Алексеев 1984 — М. Е. АЛЕКСЕЕВ. К вопросу о классификации лезгинских языков // *Вопросы языкознания*, 1984, № 5. С. 88—94. [М. Е. ALEKSEEV. K voprosu o klassifikatsii lezginskikh yazykov // *Voprosy yazykoznaniiya*, 1984, № 5. S. 88—94.]
- Алексеев 1985 — М. Е. АЛЕКСЕЕВ. *Вопросы сравнительно-исторической грамматики лезгинских языков. Морфология, Синтаксис*. Москва, 1985. [М. Е. ALEKSEEV. *Voprosy sravnitel'no-istoricheskoy grammatiki lezginskikh yazykov. Morfologiya, Sintaksis*. Moskva, 1985.]
- Бурлак & Старостин 2005 — С. А. БУРЛАК, С. А. СТАРОСТИН. *Сравнительно-историческое языкознание*. 2-е изд. Москва, 2005. [S. A. BURLAK, S. A. STAROSTIN. *Sravnitel'no-istoricheskoe yazykoznanie*. 2-e izd. Moskva, 2005.]
- Ибрагимов 1978 — Г. Х. ИБРАГИМОВ. *Рутульский язык*. М., 1978. [G. Kh. IBRAGIMOV. *Rutul'skij yazyk*. M., 1978.]
- Коряков 2006 — Ю. Б. КОРЯКОВ. *Атлас кавказских языков. С приложением полного реестра языков*. Москва, 2006. [Yu. B. KORYAKOV. *Atlas kavkazskikh yazykov. S prilozheniem polnogo reestra yazykov*. Moskva, 2006.]
- С. Старостин 1989/2007 — С. А. СТАРОСТИН. Сравнительно-историческое языкознание и лексикостатистика // Старостин 2007. С. 407—447. [S. A. STAROSTIN. *Sravnitel'no-istoricheskoe yazykoznanie i leksikostatistika* // Starostin 2007. S. 407—447.] [Впервые опубли. в: *Лингвистическая реконструкция и древнейшая история Востока*. М., Наука, 1989: 3—39. Статья опубликована также по-английски, см. S. Starostin 1999/2000.]
- С. Старостин 1993/2007 — С. А. СТАРОСТИН. Рабочая среда для лингвиста // Старостин 2007. С. 481—496. [S. A. STAROSTIN. *Rabochaya sreda dlya lingvista* // Starostin 2007. S. 481—496.] [Впервые опубли. в: *Базы данных по истории Евразии в средние века*, вып. 2. М., Институт востоковедения РАН, 1993: 50—64. Перепечатано в: *Гуманитарные науки и новые информационные технологии*. М., РГГУ, 1994: 7—23.]
- С. Старостин 2007 — С. А. СТАРОСТИН. *Труды по языкознанию*. Москва, 2007. [S. A. STAROSTIN. *Trudy po yazykoznaniiyu*. Moskva, 2007.]
- С. Старостин 2007a — С. А. СТАРОСТИН. Определение устойчивости базисной лексики // Старостин 2007. С. 827—839. [S. A. STAROSTIN. *Opredelenie ustojchivosti bazisnoj leksiki* // Starostin 2007. S. 827—839.]
- С. Старостин б. д. — С. А. СТАРОСТИН. *Историческая фонетика лезгинских языков*. Машинопись, 1980-е гг. [S. A. STAROSTIN. *Istoricheskaya fonetika lezginskikh yazykov*. Mashinopis', 1980-e gg.]
- Талибов 1980 — Б. Б. ТАЛИБОВ. *Сравнительная фонетика лезгинских языков*. Москва, 1980. [B. B. TALIBOV. *Sravnitel'naya fonetika lezginskikh yazykov*. Moskva, 1980.]

- Barbançon et al. 2013 — François BARBANÇON, Steven N. EVANS, Luay NAKHLEH, Don RINGE, Tandy WARNOW. An experimental study comparing linguistic phylogenetic reconstruction methods // *Diachronica* 30/2 (2013). P. 143—170.
- Gascuel 1997 — O. GASCUEL. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data // *Molecular Biology and Evolution* 14 (1997). P. 685—695.
- Gippert et al. 2008 — J. GIPPERT, W. SCHULZE, Z. ALEKSIDZE, J.-P. MAHÉ. *The Caucasian Albanian Palimpsests of Mt. Sinai*. 2 vols. Brepols, 2008.
- Goloboff et al. 2008 — P. A. GOLOBOFF, J. S. FARRIS, K. C. NIXON. TNT, a free program for phylogenetic analysis // *Cladistics* 24/5 (2008). P. 774—786.
- Gray & Atkinson 2003 — Russell D. GRAY & Quentin D. ATKINSON. Language-tree divergence times support the Anatolian theory of Indo-European origin // *Nature* 426 (27 November 2003). P. 435—439.
- Huelsenbeck & Ronquist 2001 — J. P. HUELSENBECK, F. RONQUIST. MrBayes: Bayesian inference of phylogenetic trees // *Bioinformatics* 17/8 (2001). P. 754—755.
- Huson & Bryant 2006 — D. H. HUSON & D. BRYANT. Application of phylogenetic networks in evolutionary studies // *Molecular Biology and Evolution* 23/2 (2006). P. 254—267.
- Kassian 2011—2012 — *Annotated Swadesh wordlists for the Lezgian group (North Caucasian family)*. Database compiled and annotated by A. KASSIAN (November 2011 — October 2012). The Global Lexicostatistical Database project: <http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\ncc\lez&limit=1>
- Kassian 2013 — A. KASSIAN. The Lezgian linguistic group within the framework of the Global Lexicostatistical Database. Talk at the conference *Comparative-Historical Linguistics of the 21st Century: Issues and Perspectives*, Moscow, March 20—22, 2013: <http://www.academia.edu/3040336/>
- Kassian et al. 2010 — A. KASSIAN, G. STAROSTIN, A. DYBO, V. CHERNOV. The Swadesh wordlist. An attempt at semantic specification // *Journal of Language Relationship*, No. 4 (2010). P. 46—89.
- Makarenkov et al. 2006 — Vladimir MAKARENKOV, Dmytro KEVORKOV, Pierre LEGENDRE. Phylogenetic Network Construction Approaches // Dilip K. ARORA, Randy M. BERKA, Gautam B. SINGH (eds.). *Applied Mycology and Biotechnology*. Vol. 6: *Bioinformatics*. Elsevier, 2006. P. 61—98.
- Müller et al. 2010 — André MÜLLER, Søren WICHMANN, Viveka VELUPILLAI, Cecil H. BROWN, Pamela BROWN, Sebastian SAUPPE, Eric W. HOLMAN, Dik BAKKER, Johann-Mattis LIST, Dmitri EGOROV, Oleg BELYAEV, Robert MAILHAMMER, Matthias URBAN, Helen GEYER, Anthony GRANT. *ASJP World Language Tree of Lexical Similarity*. Version 3 (July 2010). http://email.eva.mpg.de/~wichmann/language_tree.htm
- Nakhleh et al. 2005 — Luay NAKHLEH, Tandy WARNOW, Donald RINGE & Steven N. EVANS. A Comparison of Phylogenetic Reconstruction Methods on an IE Dataset // *The Transactions of the Philological Society* 103 (2005). P. 171—192.
- Saitou & Nei 1987 — N. SAITOU & M. NEI. The neighbor-joining method: A new method for reconstructing phylogenetic trees // *Molecular Biology and Evolution* 4 (1987). P. 406—425.
- Schulze 1988 — W. SCHULZE. *Studien zur Rekonstruktion des Lautstandes der südostkaukasischen (lezgischen) Grundsprache*. Habilitationsschrift Universität Bonn. Unpubl. ms.
- Schulze 2001 — W. SCHULZE. *The Udi Gospels. Annotated text, etymological index, lemmatized concordance*. München/Newcastle: Lincom, 2001.
- Schulze 2005 — W. SCHULZE. *A Functional Grammar of Udi*. Unpubl. ms., 2005. Available on demand at author's site: <http://www.lrz.de/~wschulze/FGU.htm>
- Schulze-Fürhoff 1994 — W. SCHULZE-FÜRHOFF. Udi // R. SMEETS (ed.). *The Indigenous Languages of the Caucasus*, vol. 4. Caravan Books, 1994. P. 447—514.
- Semple & Steel 2003 — Ch. SEMPLE, M. STEEL. *Phylogenetics*. Oxford University Press, 2003.
- Sneath & Sokal 1973 — P. H. A. SNEATH & R. R. SOKAL. *Numerical Taxonomy*. San Francisco: W.H. Freeman and Company, 1973.
- G. Starostin 2010 — G. S. STAROSTIN. Preliminary lexicostatistics as a basis for language classification: A new approach // *Journal of Language Relationship*, No. 3 (2010). P. 79—116.
- G. Starostin 2011 — *Annotated Swadesh wordlists for the Nakh group (North Caucasian family)*. Database compiled and annotated by G. STAROSTIN (last revision: October 2011). The Global Lexicostatistical Database project: <http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\ncc\nah&limit=1>
- S. Starostin 1994 — S. A. STAROSTIN. *Lezgian Etymological Database*. Published in [Starostin & Nikolayev 1994]; available at: <http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl>

- S. Starostin 1999/2000 — S. A. STAROSTIN. Comparative-historical linguistics and lexicostatistics // *Historical Linguistics and Lexicostatistics*. Melbourne, 1999. P. 3–50 [republ. in: *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford, 2000, p. 223–259.]
- Starostin & Nikolayev 1994 — S. A. STAROSTIN, S. L. NIKOLAYEV. *A North Caucasian Etymological Dictionary*. Moscow, 1994 [reprinted: 3 vols. Ann Arbor: Caravan Books, 2007]. Available online at the Tower of Babel project as Cautet.dbf: <http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl>

Alexei KASSIAN. Towards a formal genealogical classification of Lezgian languages of the North Caucasus.

A lexicostatistical classification is proposed for 20 languages and dialects of the Lezgian group of the North Caucasian family, based on meticulously compiled 110-item wordlists, published as part of the *Global Lexicostatistical Database* project. The lexical data have been subsequently analyzed with the aid of the principal phylogenetic methods, both distance-based and character-based: Starling neighbor joining (StarlingNJ), Neighbor joining (NJ), Unweighted pair group method with arithmetic mean (UPGMA), Markov chain Monte Carlo (MCMC), Unweighted maximum parsimony (UMP). All these methods, with the exception of UMP, have yielded trees that are sufficiently compatible with each other to generate a summary phylogenetic tree of the Lezgian lects. The obtained summary tree agrees with the traditional classification as well as some of the previously proposed formal classifications of this linguistic group. Contrary to theoretical expectations, the UMP method has suggested the least plausible tree of all.

Keywords: language classification, lexicostatistics, phylogeny, Lezgian languages.

External relations of Scythian

The paper is a case study of an ancient language known only from foreign traditions. It is argued that a fairly adequate description of its phonetics and contacts with other languages can still be achieved. The Scythian word stock shows certain exceptions to the regular phonetic correspondences that are probably due to borrowing. It seems likely that in Ciscaucasia and the North Pontic area the Scythians and their ancestors encountered bearers of West Iranian or other Aryan, North Caucasian, and unidentified Indo-European languages and partially embraced their onomasticon, theonyms, and names of some *realia*. On the other hand, Scythian loanwords can be detected in East Caucasian, Median, Avestan and Thracian.

Keywords: Scythian, Thracian, Avestan, North Caucasian, language contacts.

My paper is a case study of an ancient language known only from foreign traditions. I believe that a fairly adequate description of its phonetics and contacts with other languages can still be achieved. The most distinctive trait of Scythian phonetics is the change $d > \delta > l$, cf. Scythian Παράλαται, the name of the Scythian royal caste, and Avestan *paradāta*, the mythical dynasty of kings. There is also the change of the $*xš-$ cluster into a sibilant in the initial position. Thus, the famous Olbian decree in honour of a certain Protogenus dealing with the events of the late 3rd century BC mentions the tribe of Σάιοι (in the plural genitive case Σαίων, from Iranian $*xšaya-$ ‘king’) and its king Saitapharnes (Σαιταφάρνης, from Iranian $*Xšaitafarna-$, ‘[Possessing of] a bright/regal *farnah*,’ the text recorded only the genitive case Σαιταφάρνου) (IOSPE I²: 32, A, lines 10, 34, 83). The Iranian origin of both names is unmistakable, and so is the Scythian change $*xš > s$ in the initial position. Therefore, both the ethnonym and the anthroponym of the Olbian decree should be regarded as Scythian and not Sarmatian, as it has been occasionally believed. Moreover, as Tomaschek (1888: 721) rightly pointed out, the ethnic name can be identified with that of the ‘regal Scythians’ (οἱ βασιληῖοι Σκύθαι, the text recording only the plural genitive case, τῶν βασιληίων Σκυθέων) mentioned by Herodotus (IV, 20). The same change can be seen in the name of Σατράκης, the king of Central Asian Scythians, a contemporary of Alexander the Great mentioned by Arrian (Anab. Alex. IV, 4, 8). These examples can neither chronologically nor/or geographically pertain to the Sarmatians. Moreover, the initial $Xš-$ of Sarmatian proper names is invariably rendered as Ξ: cf., for instance, Ξάρθανος (Tanais, CIRB 1287, AD 244), Ξαιορσάκης (Pantikapaion, CIRB 689, 2nd century AD) as well as the Ossetic reflex $xs-$. Therefore it seems likely that the change $*xš- > s-$ did take place in Scythian.

The second part of the names of Scythian kings recorded by Herodotus, Ἀριαπίθης (Herod. IV 76, 78) and Σπαργαπίθης, being clearly the reflex of Common Iranian $*paiša-$ ‘outlook’, shows the change $*ś > \theta$ instead of the standard change $*ś > s$ reflected, for instance, in the cognate name of the prince of the Massagetae — Σπαργαπίσης (Herod. I 211, 213). True, it is believed that $-\theta-$ emerged in these names under the influence of Greek names with the element $-\piείθης$ (Διοπέιθης, etc.) (Vasmer 1923: 12; Schwyzler 1968: 206 and Schmitt 2002: 149; 2003: 5). Yet some words showing the change $*ś > *θ$ (and finally t) were borrowed by Os-

setic (for instance, *færæt* ‘axe’ from **paraθu-*, going back to common Iranian **paraśu-*, etc.) where they should be treated as Scythian loanwords: the regular Ossetic reflex of **ś* is *s*. Therefore I think that *-θ-* in the names of Ariapeithes and Spargapeithes was the regular reflex of Common Iranian **ś*.

Given the above-discussed change $d > \delta > l$, the emergence of intervocalic *-d-* in Scythian calls for an explanation. Judging from the rendering of the name of the Scythians both in Greek and Akkadian, by the time of their acquaintance Common Iranian **d* had already changed to Scythian δ . However, Herodotus (I 103) mentions Μαδύης, the Scythian king of the 7th century BC. It means that at the same time Scythian had a plosive *d* not going back to Common Iranian **d*. This fact can be accounted for by the change **-nt-* > *-d-*, typical of Southeast Iranian languages, cf. Yidga *lad* ‘tooth’ < **danta*, etc. (Эдельман 1986: 163). In this case the name of Μαδύης should be explained as the reflex of Indo-Iranian **mantu-* (cf. Avestan *mantu-*, Old Indian *mántu-* ‘counsellor; ruler’), not of Iranian **madu-* ‘honey; mead’. Scythian forms showing, according to generally accepted etymologies, *-d-* < **-d-* and *-nt-* < **-nt-* can be explained differently. Thus, the name of the Scythian king Idanthyrso is recorded as Ἰνδᾶ(ν)θυρσοσ in some of Herodotus’ manuscripts (Common Iranian **d* survived in Scythian in the **-nd-* cluster). Moreover, intervocalic *d* of the forms Ἰδάνθυρσοσ and Ἰδανθούρας can go back to **-nd-* (in this case the first part of the name can be traced to **vinda-* ‘acquiring’). Multiple forms of this name (cf., for instance, Грантовский 1998: 251; Иванчик 2009: 69–71) imply that the phoneme/consonant cluster rendered alternately as δ or $\nu\delta$ in Greek could be pronounced in Scythian in a specific way, hindering its unambiguous perception by foreign speakers. It could be the cluster *-nd-*; cf. its reflexes in modern East Iranian languages showing both *-nd-* and *-d-*: Munji *vond-*, Upper Munji *vod-*, Yidga *vad-* < **band-* ‘to tie’ (Эдельман 1986: 169), Pashto *āyund-* < **āgunda-* ‘to dress’, but Pashto Afridi *adūm* < **handāma* ‘thigh’ (Skjærvø 1989: 403). Given that Iranian *-nd-* could be rendered by Greek *-ντ-* as well (cf. the rendering of Old Persian *Vīdāfarnah-* by Greek Ἰνταφέρνης / Ἰνταφρένης), it can be surmised that the element *-αντ-* in such names as, for instance, Ἀριάντας (Herod. IV 81) renders Scythian **-anda-* (as in the name of a certain Κολανδάκης from the Olbian graffito on the wall of a black-figure kylix dated to the first half of the 4th century BC — Толстой 1953: 11, № 7) (cf. the name of the Persian satrap Ἀρσάνδας mentioned by Polyaeus [VII 11, 7; VII 34; VIII 47]). It seems all the more plausible since the widely accepted explanation of such names as Iranian adjectives or participles ending in *-ant-* does not take into account that the final *t* of the latter was dropped in the nominative and vocative cases, and Greek could hardly have borrowed the oblique stem. Yet another example of the alleged survival of Common Iranian **-nt-* in Scythian, the hydronym Παντικάπησ, also does not necessarily contradict our thesis of the change **-nt-* > *-d-* in Scythian, since the names of big rivers are prone to be substrate.

The Scythian wordstock shows, however, certain deviations from regular phonetic correspondences that are probably due to borrowing, i.e. West Iranian or Aryan *κτ* (*kt/xt*)- instead of East Iranian *-γd-*, *s* instead of *θ* < Common Iranian **ś*, etc. These linguistic phenomena and the ensuing historical and ethnocultural conclusions will be dealt with below. Let us first turn to the relevant names and words recorded by Herodotus.

Ὀκταμασάδης is the name of the king of Scythia mentioned by Herodotus (IV 80). Its first part (Ὀκτα-) seemingly renders either Aryan **ukta* or West Iranian **uxta-* ‘said, proclaimed’ (past participle from *vak-*) (Müllenhoff 1866: 567 = Müllenhoff 1892: 116; Vasmer 1923: 15). Humbach and Faiss’s etymology (Humbach and Faiss 2012: 7): from Iranian **hu-kr̥ta* ‘well-built; well-shaped’ implies an otherwise unattested disappearance of the vocalic *r* and can therefore hardly be accepted. O. Trubachev pointed out that *-kt-* became *-xt-* in Iranian and held that the name of Octamasades (where *-κτ-*, according to him, rendered *-kt-*) was Indo-

Aryan (Трубачев 1999: 199–201 = Трубачев 1985: 140–143). Actually, it is not necessarily so. Iranian *-xt-* was invariably rendered by Greek *-κτ-*, e.g. Βάκτρα from *Bāxtri-* or Ἀρταῦκτης from **Arta-uxta-*, and never by *-χτ/θ-*. It is true, however, that Greek *-κτ-* never renders the East Iranian reflex of Aryan **-kt-*, i.e. *-γδ-*, always transcribed as *-γδ-* (Ἰγδαμπαῖς, Σόγδιος, etc.). S. Tokhtas'ev suggested in this connection that in some East Iranian languages, including Scythian, Indo-Iranian *-kt-* could become *-xt-* without a subsequent transition to *-γδ-* (Тохтас'ев 2005: 98–99, n. 227). It seems, however, more likely that Scythian kings could bear non-Scythian names. There is also some circumstantial evidence in favour of a non-Scythian origin of Octamasades' name, i.e. the fact that it was borne by the son of the king of the Sindoi mentioned in an inscription from the Semibratnee site on the south bank of the Kuban river on the Taman peninsula, dated palaeographically to the 4th century BC (Блаватская 1993; Тохтас'ев 1998; Яйленко 2008).

Herodotus (IV 53) narrates that ἀντακάϊος is a great spineless fish from the mouth of the Borusthenes river, used for salting. Trubachev (1999: 64–65, first published in Трубачев 1977) compared it to Old Indian *ánta-* 'end; limit' (present in Iranian as well — ЭСИЯ I: 173–175) and *kāya-* 'body', meaning something like 'sharp-bodied.' If Greek *-ντ-* stood for *-nt-* (and not *-nd-*, which is also possible, cf. above) of the source language, the word was probably non-Scythian — as stated above, Common Iranian **-nt-* apparently became *-d-* in Scythian. As Trubachev already noted, ἀντακάϊος was reflected in West Caucasian and Kartvelian languages of the Black Sea Caucasian coast. Thus, Giorgio Interiano, an Italian traveller of the late 15th — early 16th century in his description of the Adygs wrote that there was in their land a kind of sturgeon called *anticei* just as it was called by Strabo (Кавказ 2010: 29). Interiano's compatriot Archangelo Lamberti, who was a missionary in Megrelia from 1635 to 1653, noted that a kind of local sturgeon was called *Angiachia* (Ламберти 1913: 207). The latter name is also recorded in Kipshidze's Megrelian dictionary (Кипшидзе 1914: 195) (I am indebted to Yakov Testelet's for this reference). The word has neither a West Caucasian nor a Kartvelian etymology and it seems likely that the Adygs and Megrelians borrowed the name for sturgeon from the early inhabitants of the east coast of the Black Sea, judging from the area, either Sindiian or Maeotian. Given that the name of the Maeotians, i.e. Μαῖται¹ / Μαῖῶται² / Μαῖῆται,³ is reminiscent of the name of the Near Eastern kingdom of Mitanni where the ruling élite was Aryan, it can be surmised that Maeotian was an Aryan language.

Ταβιτί, according to Herodotus (IV 59), is the Scythian name of Hestia, the goddess of the hearth. It is traditionally traced to the Iranian stem **tap-* 'to burn; to heat' (Zeuß 1837: 286 und Anm. *; Müllenhoff 1866: 558, Anm. 1 = 1892: 108, Anm. ***; Mayrhofer 2006: 17; Humbach and Faiss 2012: 7) though Vasmer (Vasmer 1923: 16–17) rightly doubted that *-β-* could represent Iranian *-p-*⁴ and suggested sagaciously that the initial voiceless stop of the name of Tabiti corresponded to the Iranian voiced consonant. However, his comparison of the Scythian goddess's name with the causative participle **dāvañantī-* 'burning; heating' from the Old Indian stem *dū-* 'to burn' does not explain the intervocalic beta in Greek rendering. In this connection

¹ This form is recorded in epigraphy, cf. CIRB 10, 25, 40, 180, 971, 972, 1015, 1039. This list does not include nos. 8, 9, 11, 39, 1040 where the ethnonym can be safely reconstructed in accordance with the standard formula (except for no. 9 showing a somewhat unusual formula).

² The most widespread form of Greek literature.

³ The Ionic form used by Herodotus (IV 123).

⁴ Indeed there are no examples of voicing of intervocalic *-p-* in Scythian, cf. the names of Api, Lipoxais, Papiaios, Spargapeithes, etc. Tokhtas'ev (2005: 78–79, n. 109) suggested a dissimilation of voiceless stops, but there are no parallels for such a dissimilation in Scythian.

Holzer's hypothesis of the existence of a certain Indo-European language reflecting Indo-European (including Indo-Iranian) unvoiced consonants as voiced and vice versa⁵ in early Eastern Europe (Holzer 1989: 13 et passim) is worth noting. It seems likely that the name of Tabiti was a substrate borrowing cognate to the Indo-European stem **dh₂p-* 'sacrificial meal' (cf. Latin *daps* 'sacrificial meal, feast', Armenian *tawn* 'feast,' Old Icelandic *tafn* 'sacrificial animal, feast,' etc.),⁶ which does not contradict the image of the hearth goddess.

Hesychius recorded the word **σανάπη* (*σάναπτιν* in the manuscript, *σανάπην* is Scali-ger's conjecture) as Scythian for 'drunken woman'. Its first part *σανά-*, meaning 'wine' (cf. Ossetic *sæn/sænæ* 'id.'), can be traced to Proto-North Caucasian **swīnē* 'barberry; currants' (Avar *saní* 'barberry', Lak *sunū* 'pomegranate', Adyghe *sāna*, Kabardian *sāna* 'grapes; wine' — NCED: 971) and, judging from its phonetic form and semantics, is a West Caucasian borrowing.⁷ The second part seems to be a reflex of the Indo-European stem **peh₃ / ph₃* 'to drink' (Vasmer 1923: 50.) The same word, according to Hecataios, existed in Thracian (*ἐπεὶ δὲ οἱ μέθυσοι σανάπαι λέγονται παρὰ Θραξίν...* "Since drunkards are called *σανάπαι* among the Thracians..." — FGH I F 34), and was probably a Scythian loanword. It is all the more plausible since the Thracians had their own word for wine: *ζίλαι* (Hes. *ζίλαι ὁ οἶνος παρὰ Θραξί*).

The word *ἀκινάκης* used by Herodotus to call Persian (VII 54) and Scythian (IV 62; 70) swords (cf. also Sogdian *kyn'k* 'id.') cannot belong to the inherited Indo-Iranian wordstock where **ki* would have become *ci*. In Greek it is regarded as a loanword, either Iranian (Frisk 1960–73, I: 53) or 'Pre-Greek rather than Iranian' (Beekes 2009, I: 51), with no suggested etymology. The word can be traced back to Proto-Nakh **hāhķīnekV* 'iron knife': **hāhķi* 'iron,' cf. Chechen *ēčig*, Ingush *äšk*⁸, Batsbi *šajhķi* 'id.' + suffix forming adjectives *-in* + *nekV* 'knife,' cf. Batsbi *nek* 'id.' — for North Caucasian etymology see NCED: 851–852⁹. The combination **hāhķi-in* accounts for the long iota in the Greek loanword *acīnacēs* implied by Latin prosody: it is the last word in the first line of Alcaic hendecasyllabic verse where the antepenultimate syllable can only be long (*Hor. Od. I, 27, 5*) (cf. Chantraine 1999: 47; Beekes 2009, I: 51).

The Scythians proper also contacted with the speakers of East Caucasian languages as evidenced by the phonetics of an Iranian loanword in East Caucasian, reconstructed as **μῆλι* 'felt; felt cloak' (Archi *warti* 'felt cloak,' Tabasaran *verč* 'felt,' Lezghi *lit* 'felt,' Dargwa Akushi

⁵ Holzer traces voiceless consonants of the postulated substratum only to Indo-European voiced aspirates and not to the voiced unaspirates of the traditional reconstruction, resp. glottalized consonants of Gamkrelidze — Ivanov's and tense consonants of Shevoroshkin's and Starostin's reconstructions. However, in at least one of his examples (see below, s.v. *Τάναϊς*), the substratum *t-* is clearly the reflex of the Indo-European unaspirated / glottalized / tense stop. As to *Τάναϊς* and *Ταβιτί*, it can be accounted for either by the fact that the reflexes of unaspirated / glottalized / tense stops were unvoiced as well — as far as I can judge, there is no contradicting evidence — or by the presence of a laryngeal as it was the case with the formation of Indo-Iranian aspirates — cf., for instance, **ratha-* 'chariot' from **rót-h₂-o-* — EWAia II: 429–430.)

⁶ On this stem and its reflexes see Toporov 2012: 167–169.

⁷ Contrary to a widely accepted hypothesis (ИЭСОРЯ I: 512–513; III: 66–68, with literature; EWAia II: 605; Тохташев 2005: 102 and n. 251), the word can hardly be cognate with Old Indian *śanā*, Middle and New Persian *šan* 'cannabis' going back to the etymon with a sibilant/stop alternation in the word-initial position whose reflexes are German *Hanf*, Ossetic *gæn/gænæ*, Russian *конопля* 'id.' etc.

⁸ Chechen and Ingush forms are the results of the addition of the diminutive suffix **-k* and a regular palatalization (NCED: 851)

⁹ Cf. NCED: 851–852. Borrowing of Proto-Nakh *hāhķin-ig* 'made of iron' or **=iāχ-in* ("=" denotes a class marker, cf. Batsbi *d-aχē*, Chechen *d-eχa*, Ingush *d-šāχa* 'long') 'long' + *nekV* 'long knife' seems less plausible phonetically, though by no means impossible.

warhi, Dargwa Chirag *warse* ‘felt cloak,’ Lak *warsi* ‘felt cloak,’ Avar *burtina* ‘felt cloak; felt,’ Chechen *werta* ‘felt cloak,’ etc.) (Климов 1972: 54; Старостин 1988: 113 = Старостин 2007: 314) — cf. Avestan *varāsa-* ‘hair,’ Old Indian *vālśa-* ‘shoot; branch; twig.’ As indicated by S. Starostin, the phoneme * θ can be reconstructed for a small number of East Caucasian stems, probably borrowings (Старостин 1988: 154, n. 3 = Старостин 2007: 314, n. 2). In the source-language etymon in question it clearly goes back to * \acute{s} < Indo-European * k . Among the recorded Ancient Iranian languages the change * \acute{s} > θ is characteristic only of Old Persian and Scythian. But the language of the ancestors of Persians still had * \acute{s} when they had come to the Near East; otherwise they would not adopt the name of Assyria, i.e. *Aššur* as **Ašura*, hence Old Persian *Aθura*, given the presence in their language of *s* and *š* sibilants far more resembling the Akkadian phoneme.¹⁰ It appears therefore that the East Caucasian form could not be borrowed from Proto-Old Persian and Scythian remains the only likely donor language.¹¹

The ethnic name *Skudra* known from Old Persian inscriptions is usually identified with the Thracians since they are named after the “Overseas Scythians” (*Sakā tayaiy para draya*) and before the “Greeks bearing a shield[-like hat]”¹² (*yaunā takabarā*) on the list of nations/provinces of the Achaemenid Empire. Such an ethnic attribution seems to be on the whole well-grounded (see below), yet it is substantiated only by geographic arguments. Both *Skudra*’s name and their clothes as depicted by Achaemenid reliefs¹³ are reminiscent of the Iranians.¹⁴ True, the word *Skudra* has been compared to such toponyms as *Skodra* in Illyria, *Skydra* in Macedonia, etc., yet these etymological exercises are no more reliable than the attempts at linking together the Scandinavian ethnonym *Rus* and the Slav hydronym *Ros*’. Now, the etymology proposed by Szemerényi (1980) is far more convincing. Basing himself on Akkadian (*ašguzāia*, *asguzāia*, *iškuzāia*¹⁵) and Greek (Σκούθαι) renderings of the Scythian proper name for themselves he has shown that both Greek θ and Akkadian *z* reflect the interdental voiced fricative δ (δ in the conventional Iranological transcription), the characteristic Eastern Iranian reflex of the Common Iranian * d , and has reconstructed the Early Scythian ethnic name as **Skudā*.¹⁶ Then he confronted this proper name with that of the *Skudra* and surmised that the latter was a Median derivative of the verbal root **skud-*, formed with the aid of the suffix *-ra* and therefore none

¹⁰ True, it can be surmised that the Persians adopted the name of Assyria from Median in the regular Median form **Asura* and, knowing that Median *s* often corresponds to Persian θ , changed *s* to θ . However, in a number of positions Median *s* corresponds to Persian *s*, so such a shift does not seem plausible.

¹¹ Indeed, the word could originate from an extinct Iranian language showing, like Old Persian and Scythian, the change * \acute{s} > θ , but I prefer to deal with established facts.

¹² That is, the wide-brimmed πέτασος hat. The Akkadian versions of Achaemenid inscriptions make it clear by pointing out that they have shields “on their heads” (*ina qaqqadišunu*).

¹³ On the clothes of personages either labelled as *Skudra* (*iyam Skudra*, “this is a *Skudra*”) or identified as such see Walser 1966: 54–55, 60–61, 65–66, 85, 95–97.

¹⁴ According to Diakonoff (1981: 140, note 65), V.A. Livshits “toyed with the idea that the land *Skudra*- of the Bisutun inscription may stand for European Scythia, and not for Thrace as is usually assumed (or for both...)”.

¹⁵ In ancient Semitic languages no syllable could begin with two consonants, hence a prothetic vowel in the Akkadian word (see, e.g., Diakonoff 1991: 80; Коган 2009: 35), *a-* in Assyrian and *-i-* in Babylonian dialect (Дьяконов 2008: 255). On Akkadian rendering of Iranian words with initial consonant clusters see Грантовский 1970: 73 = Грантовский 2007: 84–85.

¹⁶ Later the spirant δ/δ became *l* in Scythian as well as in South-Eastern Iranian languages, namely Bactrian, Pashto, Munji and Yidga. This fact accounts for the name Σκόλοτοι recorded by Herodotus, i.e. **skula-ta*, where *-ta* is a plural marker, and is one of the isoglosses linking Scythian together with South-Eastern Iranian idioms and not with Sarmato-Alano-Osetic ones as it has been believed (see Szemerényi 1980; Витчак 1992; Кулланда 2011).

other than the Median rendering of the Scythians' proper name akin to the Akkadian *ašguzāia*, *asguzāia*, *iškuzāia* and the Greek Σκύθαι. The only contestable point of this etymology is, in my opinion, the assumption that the word *Skudra* is not a mere phonetic rendering of a foreign ethnic name but its reinterpretation by means of singling out the root and providing it with a new suffix. True, Iranian-speaking Medians or Persians would probably have been able, in contradistinction to the Assyrians and Greeks, to define the root of a Scythian word. However, in order to do that they should have had a reflex of the hypothetical root **skud-* in their own tongue, and there are no such reflexes in Western Iranian languages. All the proposed etymologies of the word **Skudā* are either not substantiated at all by the Iranian data¹⁷ or based solely on Eastern Iranian materials.¹⁸ It seems more likely that the Medians would have tried to render the sound of a foreign phoneme with the aid of their proper phonetic means, all the more so since there is yet another example of Eastern Iranian *δ* (*ḍ*) rendered by the “dental + *r*” cluster in a Western Iranian language, namely the ‘Median’ *Bāxtrī-*¹⁹ corresponding to the Avestan *Bāxδī-*. The irregular correspondence *-xtr-* — *-xδ-* has not been explained convincingly. It has been believed that the original Eastern Iranian form had been **Bāxδrī-*, and later *-r-* was dropped (Humbach 1966: 52). However, there are no parallels for such a development in Eastern Iranian languages. Given the two above-mentioned Median renderings of Eastern Iranian proper names, would it not be more natural to suppose that the Medians, like the Greeks and the Assyrians, merely tried to imitate the sound of the foreign phoneme *ḍ* rendering it in an intervocalic position as a cluster of a voiced dental stop and a resonant (*-dr-*), and in a position after a voiceless consonant *x* as a cluster of a voiceless dental stop and a resonant (*-tr-*)?

How was the Scythian ethnic name transferred to the Thracians? The point is that, according to classical sources, there existed close links between the Scythians and the Thracians. For instance, among Thracian tribes one encounters Iranian ethnic names showing a specific Scythian phonetic development. Thus, Hecataios (St. Byz., fr. 128 = FGH I F 157) and Herodotus (VII, 110–113) mention the Thracian tribe Σάτραι, whose name is clearly a reflex of Iranian *xšaθra* with a Scythian *xš-* > *s-* change in the initial position (cf. Tomaschek 1893: 68).²⁰ A poem by Archilochus written in the second half of the 7th century BC mentions the ethnic name Σάϊοι (in the plural genitive case Σαῖων).²¹ The island of Thasos, where Archilochus lived, was situated near the Thracian shores and its inhabitants constantly warred with the Thracians. Moreover, the *scholia* to this fragment identify the Σάϊοι as a Thracian tribe (see Bonnard’s commentary in Archiloque 1958: 5–6). However, one simply cannot help remembering the already cited Olbian decree in honour of Protogenus mentioning a homonymous tribe. As To-

¹⁷ As Leibniz’s hypothesis, supported by Justi and Szemerényi (Justi 1896–1904: 441; Szemerényi 1980: 21 and n. 44), stating that this word goes back to the Indo-European root meaning ‘to shoot.’

¹⁸ As Cornillot’s hypothesis (Cornillot 1981) linking the name **Skudā* to Wakhi *skid* ‘skull-cap,’ formerly ‘pointed cap,’ and that of Trubachev (1999: 137, first published in 1980) comparing it to the Ossetic participle *sk’yd* / *sk’ud* ‘torn; split.’

¹⁹ A proper Old Persian form where *tr* > *š* has survived in the Elamite rendering *ba-ak-ši-iš*, cf., for instance, the Elamite version of the Bisutun inscription (DB I, 13; III, 80; 85) and the so-called Persepolis fortification tablets (PF 1555, 8–9 — Hallock 1969: 434).

²⁰ This phonetic trait of some North Pontic Iranian idioms has been pointed out by Harmatta (1951: 308–309), although he did not consider it as pertaining to Scythian.

²¹ This passage (Archilochi frgm. 13 in Archiloque 1958) has been repeatedly cited in connection with North Pontic ethnonymy and anthroponymy — cf. Tomaschek 1893: 43–44; CIRB: 95. Unfortunately, the context of Archilochus’ quatrain does not enable one to define the ethnicity of the Σάϊοι: it is merely stated that one of them (Σαῖων τις) is boasting of Archilochus’ shield lost in battle.

maschek (1888: 721) rightly pointed out, the ethnic name can be identified with that of the ‘regal Scythians’ (οἱ βασιλήϊοι Σκύθαι, the text recording only the plural genitive case, τῶν βασιληίων Σκυθέων) mentioned by Herodotus (IV, 20),²² just as the Σαυδάραται (< Iranian **saw-dar-a-ta*, ‘Wearing black’), mentioned in the same inscription, were identified with Herodotus’ Μελάγχλαινοι, ‘[Wearing] black cloaks.’²³ All the similarities of Scythian and Thracian ethnic names can hardly be mere coincidences.

We do not know under what circumstances the Thracians embraced Scythian ethnic names, yet it can be surmised that the process was somewhat similar to that of the Eastern Slavs acquiring the Scandinavian name *Rus*, the Slavic Bulgarians acquiring their Turkic name or the Romanized population of the would-be France acquiring their Germanic name. Probably a small Scythian group formed the élite of Thracian society, labelled the local population with their name(s) and dissolved among them like Scandinavians and Turcs among the Slavs or Franks among the Romanian-speaking French.

As stated above, both linguistic and extra-linguistic considerations seem to imply that the name *Skudra* used by the Medians and the Persians as the designation of the Thracians can be traced to the North Pontic Scythian own name for themselves, **Skuḍa*. Moreover, the presumed Median rendering of a specific Eastern Iranian phoneme *ḍ/δ* through a dental + *r* cluster in the word *Skudra* is similar to that in the word *Bāxtrī-*, corresponding to the Eastern Iranian toponym *Bāxδī-*. The only difference is that in the latter case the dental became voiceless, being assimilated to a preceding voiceless fricative. Thus, it seems likely that *ḍ/δ* in Eastern Iranian loanwords in Median was regularly — although, of course, two examples is a far cry from regularity — rendered by a dental + *r* cluster.

Words of Scythian origin can also be traced in Avestan. They show phonetic evolution differing from the Avestan proper. Thus, Kersaspa is called both *gaēsu* ‘curly’ (Y. 9, 10) and *gaēθu* ‘id.’ (Yt. 13, 61; 136), besides *sūra-* ‘strong, powerful’ (cf. Old Indian *śūra-* ‘id.’) one encounters *aīβi-θūra-* ‘mighty’ (Yt. 10, 6; 78; 13, 40; 75; Y 1, 18; 4, 6; 65, 12), etc. The above-mentioned reflexes go back to Indo-Iranian **ś* (< Indo-European **k*). As stated above, among the recorded Ancient Iranian languages the change **ś* > *θ* is characteristic of Old Persian and Scythian. It can be shown, however, that these words are not Old Persian. The initial cluster *θr-* of the relevant Avestan lexical layer, going back to Indo-Iranian **sr-*, is reflected differently in Old Persian where it shows the change **sr-* > **hr-* > *r-*. Suffice it to compare Avestan *θraotah-* ‘stream,’²⁴ Old Indian *srótas* ‘id.’ and Old Persian *rautah* ‘id.’.

Thus, it appears that in Ciscaucasia and the North Pontic area the Scythians and their ancestors encountered the speakers of yet unidentified Indo-European and non-Indo-European languages and partially embraced their onomasticon, theonyms, and names of some *realia*. On the other hand, Scythian loanwords can be tentatively detected in East Caucasian, Avestan, Median and Thracian.

²² Harmatta, who also confronted these ethnic names (op. cit.: 309), arbitrarily identified the Σάϊοι with the Sarmatians.

²³ It is worth noting that such an etymology of the ethnic name Μελάγχλαινοι is in perfect accord with Herodotus’ claim (IV, 20) that the Μελάγχλαινοι are not Scythians: ...Μελάγχλαινοι, ἄλλο ἔθνος καὶ οὐ Σκυθικόν (“The Μελάγχλαινοι ... are another tribe, not Scythian”), since Common Iranian **d* would have become *l* in Scythian.

²⁴ The Avestan word survived in such compounds as *θraotō.stak-* and *θraotō.stāt-* ‘flowing; running’, usually in the feminine gender and in the plural, since it is applied to waters, *āp-*: *yejñhā paiti θraotō.stācō āpō tacinti nāuuaiiā* (“over which [earth] flowing navigable waters run” — Yt 13, 10); [*yazamaide...*] *vīspā āpō xā paiti θraotō.stātasca* (“[we sacrifice] to all the waters in sources and streams” — Y. 71, 9.).

Abbreviations

- CIRB — *Corpus Inscriptionum Regni Bosporani*. Moscow — Leningrad: Nauka 1965.
- CLI — *Compendium Linguarum Iranicarum*. Hrsg. von Rüdiger Schmitt. Wiesbaden, 1989: Dr. Ludwig Reichert Verlag.
- EWAia — Manfred MAYRHOFER. *Etymologisches Wörterbuch des Altindoirischen*. Bd. I–III, Heidelberg: Carl Winter Universitätsverlag, 1992–1996–2001.
- FGH — F. JACOBY. *Die Fragmente der Griechischen Historiker*. Leiden — New York — Köln: E.J. Brill 1995– (Photomechanischer Nachdruck des vermehrten Neudruckes von 1957).
- GiPh — *Grundriß der iranischen Philologie*. Hrsg. von W. Geiger und E. Kuhn. Bd. I, Abschnitt 1, 1895–1901, Abschnitt 2, 1898–1901. Bd. II, 1896–1904. Strassburg: Karl J. Trübner.
- IOSPE I² — *Inscriptiones Antiquae Orae Septentrionalis Ponti Euxini Graecae et Latinae iussu et impensis Societatis Archaeologicae Imperii Russici*. Iterum edidit Basilius Latyshev. Volumen primum. Inscriptiones Tyrae, Olbiae, Chersonesi Tauricae, aliorum locorum a Danubio usque ad regnum Bosporanum continens. Ed. alt., Petropoli MCMXVI.
- NCED — S. L. NIKOLAYEV, S. A. STAROSTIN. *A North Caucasian Etymological Dictionary*. Ed. by S.A. Starostin. Moscow 1994: Asterisk Publishers.
- SPAW — *Sitzungsberichte der philosophisch-historischen Klasse der Kaiserlichen Akademie der Wissenschaften*. Wien.
- ЭСИЯ — В.С. Расторгуева, Д.И. Эдельман. *Этимологический словарь иранских языков*. Москва: Восточная литература. Т.1 — 2000, т. 2 — 2003, т. 3 — 2007, т. 4 — 2011 (to be continued).

Literature

- Archiloque, 1958 — *Archiloque. Fragments*. Texte établi par François Lasserre, Privat-Doctent à l'Université de Genève, traduit et commenté par André Bonnard, Professeur à l'Université de Lausanne. Paris.
- Beekes 2010 — Robert BEEKES with the assistance of Lucien VAN BEEK. *Etymological Dictionary of Greek*. Leiden–Boston: Brill (Leiden Indo-European Etymological Dictionary Series. Ed. by Alexander Lubotsky. Vol. 10/1–2)
- Chantraine 1999 — Pierre CHANTRAINE. *Dictionnaire étymologique de la langue grecque*. Histoire des mots. Nouvelle édition mise à jour. Avec un Supplément sous la direction de : Alain Blanc, Charles de Lamberterie, Jean-Louis Perpillou. Paris: Klincksieck.
- Cornillot 1981 — François CORNILLOT. L'origine du nom des Scythes. — *Indo-Iranian Journal*, vol. 2/3. Dordrecht/Boston, 7–52.
- Diakonoff 1981 — I. M. DIAKONOFF. The Cimmerians. — *Monumentum Georg Morgenstierne I* (Acta Iranica 21). Leiden: E.J. Brill, 103–140.
- Frisk 1960–73 — Hjalmar FRISK. *Griechisches etymologisches Wörterbuch*. Bd. I–III. Heidelberg: Carl Winter Universitätsverlag.
- Haas 1966 — O. Haas. *Die phrygischen Sprachdenkmäler*. Sofia: Académie bulgare des sciences.
- Hallock 1969 — Richard T. HALLOCK. *Persepolis Fortification Tablets*. Chicago Illinois. The University of Chicago Oriental Institute Publications. Volume XCII.
- Harmatta 1951 — Harmatta JANOS. Studies in the Language of the Iranian Tribes in South Russia. — *Acta Orientalia Academiae Scientiarum Hungaricae* I, fasc. 2–3. Budapest: Akadémiai Kiadó, 261–314.
- Humbach 1966 — Helmut HUMBACH. *Baktrische Sprachdenkmäler I*. Mit Beiträgen von Adolf Grohmann. Wiesbaden.
- Humbach & Faiss 2012 — Helmut HUMBACH and Klaus FAISS. *Herodotus's Scythians and Ptolemy's Central Asia. Semasiological and Onomasiological Studies*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Justi 1896–1904 — Ferdinand JUSTI. *Geschichte Irans von dem ältesten Zeiten bis zum Ausgang der Sāsāniden*. — GiPh, Bd. 2.
- Mayrhofer 2006 — Manfred MAYRHOFER. *Einiges zu den Skythen, ihrer Sprache, ihrem Nachleben*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Müllenhoff 1866 — Karl MÜLLENHOFF. *Über die herkunft und sprache der pontischen Scythen und Sarmaten*. Auszug aus dem Monatsbericht der Königl. Akademie der Wissenschaften zu Berlin, VIII: 549–576.

- Müllenhoff 1892 — Karl MÜLLENHOFF. *Deutsche Altertumskunde III*. Berlin: Weidmannsche Buchhandlung.
- Schmitt 2002 — Rüdiger SCHMITT. *Die iranischen und Iranier-Namen in den Schriften Xenophons*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Schmitt 2003 — Rüdiger SCHMITT. Die skytischen Personennamen bei Herodot. — *Annali dell'Istituto Universitario di Napoli* [AION], 63/1–4, 1–31.
- Schwyzler 1968 — Eduard SCHWYZER. *Griechische Grammatik*. I–IV, München: C.H. Beck'sche Verlagsbuchhandlung.
- Skjærvø 1989 — Prods O. SKJÆRVØ. Pashto. — *CLI*, 384–410.
- Szemerényi 1980 — Oswald SZEMERÉNYI. *Four Old Iranian Ethnic Names: Scythian — Skudra — Sogdian — Saka*. — Sitzungsberichte der Österreichische Akademie der Wissenschaften. Philosophisch-historische Klasse. 371. Band. Veröffentlichungen der Iranischen Kommission. Band 9. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Tomaschek 1888 — W. TOMASCHEK. *Kritik der ältesten Nachrichten über den Skytischen Norden*. I: Ueber das Arimasische Gedicht des Aristaeas. — *SPAW*, Bd. 116.
- Tomaschek 1893 — W. TOMASCHEK. *Die alten Thraker. Eine ethnologische Untersuchung*. I: Uebersicht der Stämme. — *SPAW*, Bd. 128.
- Vasmer 1923 — M. VASMER. *Die Iranier in Südrußland*. Leipzig: In Kommission bei Markert & Petters.
- Walser 1966 — Georg WALSER. *Die Völkerschaften auf den Reliefs von Persepolis. Historische Studien über den sogenannten Tributzug an der Apadanatreppe*. Berlin: Verlag Gebr. Mann.
- Zeuß 1837 — Kaspar ZEÜß. *Die Deutschen und die Nachbarstämme*. München: Bei Ignaz Joseph Lentner.
- Блаватская 1993 — Т. В. БЛАВАТСКАЯ. Посвящение Левкона I. — *Российская археология* 2, 34–48. [Т. В. BLAVATSKAYA. Posvyaschenie Levkona I. — *Rossijskaya arkheologiya* 2, 34–48.]
- Витчак 1992 — К. Т. ВИТЧАК. Скифский язык: опыт описания. — *ВЯ*, № 5, 50–59. [К. Т. VITCHAK. Skifskij yazyk: opyt opisaniya. — *VYa*, № 5, 50–59.]
- Грантовский 1970 — Э. А. ГРАНТОВСКИЙ. *Ранняя история иранских племен Передней Азии*. Москва: Наука, Главная редакция восточной литературы. [Е. А. GRANTOVSKIJ. *Rannnyaya istoriya iranskikh plemen Perednej Azii*. Moskva: Nauka, Glavnaya redaktsiya vostochnoj literatury.]
- Грантовский 1998 — Э. А. ГРАНТОВСКИЙ. *Иран и иранцы до Ахеменидов*. М: Восточная литература. [Е. А. GRANTOVSKIJ. *Iran i irantsy do Akhemenidov*. M: Vostochnaya literatura.]
- Грантовский 2007 — Э. А. ГРАНТОВСКИЙ. *Ранняя история иранских племен Передней Азии*. Издание второе, исправленное и дополненное. М.: Восточная литература. [Е. А. GRANTOVSKIJ. *Rannnyaya istoriya iranskikh plemen Perednej Azii*. Izdanie второе, ispravlennoe i dopolnennoe. M.: Vostochnaya literatura.]
- Дьяконов 1991 — И. М. ДЬЯКОНОВ. Аккадский язык. — *Языки Азии и Африки*. Т. IV, кн. 1: *Афразийские языки*. М: Наука, Главная редакция восточной литературы, 70–109. [И. М. D'YAKONOV. Akkadskij yazyk. — *Yazyki Azii i Afriki*. T. IV, kn. 1: *Afrazijskie yazyki*. M: Nauka, Glavnaya redaktsiya vostochnoj literatury, 70–109.]
- Иванчик 2009 — А. И. ИВАНЧИК. К вопросу о скифском языке. — *ВДИ*, № 2, 62 — 88. [А. И. IVANCHIK. K voprosu o skifskom yazyke. — *VDI*, № 2, 62 — 88.]
- Кавказ 2010 — *Кавказ: европейские дневники XIII–XVIII веков*. Сост. В. Аталиков. Нальчик: Издательство М. и В. Котляровых. [*Kavkaz: evropejskie dnevniki XIII–XVIII vekov*. Sost. V. Atalikov. Nal'chik: Izdatel'stvo M. i V. Kotlyarovykh.]
- Кипшидзе 1914 — И. КИПШИДЗЕ. *Грамматика мингрельского (иверского) языка с хрестоматиею и словарем*. Санкт-Петербург. [И. KIPSHIDZE. *Grammatika mingrel'skogo (iverskogo) yazyka s khrestomatieyu i slovarem*. Sankt-Peterburg.]
- Климов 1972 — Г. А. КЛИМОВ. О некоторых словарных общностях картвельских и нахско-дагестанских языков. — *Этимология 1970*. Москва: Наука, 349 — 355. [Г. А. KLIMOV. O nekotorykh slovarnykh obschnostyakh kartvel'skikh i nakhsko-dagestanskikh yazykov. — *Etimologiya 1970*. Moskva: Nauka, 349 — 355.]
- Коган 2009 — Л. Е. КОГАН. Семитские языки. — *Языки мира: Семитские языки. Аккадский язык. Северозападносемитские языки*. Москва: Academia, 15–112. [Л. Е. KOGAN. Semitskie yazyki. — *Yazyki mira: Semitskie yazyki. Akkadskij yazyk. Severozapadnosemitskie yazyki*. Moskva: Academia, 15–112.]
- Кулланда 2011 — С. В. КУЛЛАНДА. Уроки скифского. — *Вопросы языкового родства*. № 5: 48–68. [S. V. KULLANDA. Uroki skifskogo. — *Voprosy yazykovogo rodstva*. № 5: 48–68.]

- Ламберти 1913 — Арканджело ЛАМБЕРТИ. *Описание Колхиды, называемой теперь Мингрелией*. Перевел с ита-
лианского К. Ф. Ган. Тифлис. [Arkandzhelo LAMBERTI. *Opisanie Kolkhidy, nazываемой teper' Mingreliej*. Pere-
vel s italianskogo K. F. Gan. Tiflis.]
- Старостин 1988 — С. А. СТАРОСТИН. Индоевропейско-севернокавказские изоглоссы. — *Древний Восток: эт-
нокультурные связи*. Москва: Наука, Главная редакция восточной литературы, 112–163. [S. A. STAROSTIN.
Indoeuropejsko-severnokavkazskie izoglossy. — *Drevnij Vostok: etnokul'turnye soyazi*. Moskva: Nauka, Glav-
naya redaktsiya vostochnoj literatury, 112–163.]
- Старостин 2007 — С. А. Старостин. *Труды по языкознанию*. Москва: Языки славянских культур. [S. A. STAROS-
TIN. *Trudy po yazykoznaniiyu*. Moskva: Yazyki slavyanskikh kul'tur.]
- Толстой 1953 — И. И. ТОЛСТОЙ. *Греческие граффити древних городов Северного Причерноморья*. Москва–Ленин-
град: Издательство Академии Наук СССР. [I. I. TOLSTOJ. *Grecheskie graffiti drevnikh gorodov Severnogo*
Prichernomor'ya. Moskva–Leningrad: Izdatel'stvo Akademii Nauk SSSR.]
- Топоров 2012 — В. Н. ТОПОРОВ. *Пиндар и Ригведа. Гимны Пиндара и ведийские гимны как основа реконструкции*
индоевропейской гимновой традиции. *Orientalia et Classica*. Труды Института восточных культур и антич-
ности. Выпуск XLIV. М. [V. N. TOPOROV. *Pindar i Rigveda. Gimny Pindara i vedijskie gimny kak osnova rekon-*
struktsii indoeuropejskoj gimnovoј traditsii. *Orientalia et Classica*. Trudy Instituta vostochnykh kul'tur i antich-
nosti. Vypusk XLIV. M.]
- Тохтасьев 1998 — С. Р. ТОХТАСЬЕВ. К чтению и интерпретации посвячительной надписи Левкона I с Семи-
братнего городища. — *Hyperboreus* 4, fasc. 2, 286–301. [S. R. TOKHTAS'EV. K chteniyu i interpretatsii pos-
vyatitel'noj nadpisi Levkona I s Semibratnego gorodischa. — *Hyperboreus* 4, fasc. 2, 286–301.]
- Тохтасьев 2005 — С. Р. Тохтасьев. Проблема скифского языка в современной науке. — *Ethnic Contacts and Cul-
tural Exchanges North and West of the Black Sea from the Greek Colonization to the Ottoman Conquest*. Edited by
Victor Cojocaru. Iași: Trinitas, 59–108. [S. R. TOKHTAS'EV. Problema skifskogo yazyka v sovremennoj nauke.]
- Трубачев 1977 — О. Н. ТРУБАЧЕВ. Лингвистическая периферия древнейшего славянства. Индоарийцы в Се-
верном Причерноморье. — *ВЯ*, № 6, 13–29. [O. N. TRUBACHEV. Lingvisticheskaya periferiya drevneyshego
slavyanstva. Indoarijtsyv Severnom Prichernomor'e. — *VYa*, № 6, 13–29.]
- Трубачев 1985 — О. Н. ТРУБАЧЕВ. INDOARICA в Северном Причерноморье. Этимологии. — *Этимология*
1982. М., 140–148. [O. N. TRUBACHEV. INDOARICA v Severnom Prichernomor'e. Etimologii. — *Etimologiya*
1982. M., 140–148.]
- Трубачев 1999 — О. Н. ТРУБАЧЕВ. *INDOARICA в Северном Причерноморье*. Москва: Наука. [O. N. TRUBACHEV.
INDOARICA v Severnom Prichernomor'e. Moskva: Nauka.]
- Эдельман 1986 — Д. И. ЭДЕЛЬМАН. *Сравнительная грамматика восточноиранских языков*. Фонология. Москва:
Наука. [D. I. EDEL'MAN. *Sravnitel'naya grammatika vostochnoiranskikh yazykov*. Fonologiya. Moskva: Nauka.]
- Яйленко 2004 — В. П. ЯЙЛЕНКО. Вотив Левкона I из Лабриса. — *Древности Боспора* 7. Москва: Институт ар-
хеологии РАН. [V. P. YAJLENKO. Votiv Levkona I iz Labrisa. — *Drevnosti Bospora* 7. Moskva: Institut ark-
heologii RAN.]

С. В. КУЛАНДА. Внешние связи скифского языка.

В статье предпринято исследование скифского — древнего бесписьменного языка, из-
вестного только по упоминаниям его слов в иноязычных текстах. Представляется, тем
не менее, что основные фонетические черты, характеризующие скифский и отличаю-
щие его от прочих иранских, могут быть установлены достаточно надежно. Отклоне-
ния от регулярных фонетических соответствий наблюдаются в заимствованиях, судя по
которым скифы и их предки контактировали в Предкавказье и Северном Причерно-
морье с носителями западноиранских или иных арийских, северокавказских и неустан-
новленных индоевропейских языков и заимствовали ряд теонимов, личных имен и на-
звания некоторых реалий. С другой стороны, скифские заимствования можно обнару-
жить в восточнокавказских языках, мидийском, авестийском и фракийском.

Ключевые слова: скифский, фракийский, авестийский, северокавказские языки, языковые
контакты.

Investigating the impact of sample size on cognate detection

The paper deals with the question of how many words are needed to successfully apply different methods for cognate detection. In order to investigate this question, a large gold standard consisting of 550 concepts translated into 4 languages (English, German, Dutch, and French) was compiled and divided into subsets of increasing sample size. Applying automatic methods for cognate detection on this gold standard shows that the accuracy of language-specific cognate detection methods clearly depends on the sample size. However, given that sample size depends on various different factors such as the genetic closeness of the languages or the degree of contact between the languages under investigation, no general lower or upper bound can be determined from the analysis.

Keywords: Comparative method, lexicostatistics, etymology, computational linguistics.

1. Cognate Detection in Historical Linguistics

In historical linguistics, the problem of cognate detection is traditionally approached within the framework of the *comparative method* (Trask 2000: 64–67, Fox 1995). The most important aspects of this traditional method for cognate detection are a language-specific notion of word similarity, which is derived from previously identified regular sound correspondences, and the iterative character of the method, by which proposed lists of cognates and sound correspondences are constantly refined and updated (Durie 1996: 6 f.). Being a non-automatic method which was never really laid out in a strict algorithmic way, there are many parameters which were never specified in the methodological literature. It is left open

- (a) how many languages should be compared,
- (b) whether or not the genetic relatedness between these languages should have been already proven,
- (c) whether or not the cognate sets to be identified should be restricted to semantically similar words, and
- (d) how many word pairs of all languages should be included in the survey (henceforth referred to as *sample size*).

For the successful application of the method it is irrelevant whether the first three parameters (a, b, and c) are specified or not. The method is indifferent regarding the number of languages being compared, it has its own procedure to determine genetic relatedness between languages, and semantically different but formally similar words have seldom posed a problem for historical linguists. The last parameter (d), the size of the word lists, however, is of crucial importance for the method, although nobody has so far been able to determine how many items a word list should at least contain in order to be applicable.

That the popular Swadesh-200 word lists (Swadesh 1952) are surely not enough when questions of remote relationship have to be solved can be easily demonstrated when consid-

ering the amount of cognate words in these word lists for some genetically related languages such as Armenian, English, French, and German (see Table 1): Given that there are maximally 20 cognates between Armenian and the other three languages, it is hardly possible that these cognates are enough to set up a satisfying set of sound correspondences between these languages. In order to prove the genetic relationship of Armenian with English, French, and German, it is unavoidable to expand the sample.¹ It might also be questioned whether the number of cognates attested between French and the Germanic languages is enough for a rigorous application of the comparative method.

Table 1: Number (lower triangle) and proportion (upper triangle) of cognates within Swadesh wordlists of 200 items for four Indo-European languages. Cognate counts are based on the data given in Kessler (2001).

	Armenian	English	French	German
Armenian	200 \ 1.0	0.07	0.10	0.10
English	14	200 \ 1.0	0.23	0.56
French	20	46	200 \ 1.0	0.23
German	20	111	46	200 \ 1.0

However, as can also be seen from the examples of shared cognate percentages in Table 1, the question of how many items constitute the “ideal sample size” for the successful application of the comparative method also depends on the genetic closeness of the languages under investigation: While Swadesh lists may not be enough to prove genetic relationship between Armenian and, say, German, they surely provide enough evidence to prove the genetic relationship between German and English. In cases of remote relationship, however, it becomes increasingly difficult to find enough initial pairs of cognate words to establish sound correspondences rigorously (Starostin 2013: 57–65).

But this is again only part of the whole problem, since cognate words are not the only *historical similarities* that can be detected when applying the comparative method. That similarities arising from language contact can seriously influence the results of the comparative method has been noticed by linguists for a long time. Increasing the sample size also increases the chance of finding contact-induced similarities. In cases of heavy contact, only the rigorous stratification of cognate candidates and proposed sound correspondences can help to disentangle borrowed from inherited traits. As a result, the sample size does not only have a theoretical minimum, but also a theoretical maximum. If the sample is too small, one may fail to detect the relevant similarities between genetically related languages. If the sample is too large, one may detect similarities which are not the result of genetic inheritance.

2. Sample Size and Cognate Detection

Given that sample size is crucial for the success of the comparative method, it would be desirable to have at least a rough estimate regarding the lower bound of how many words are needed for the task of cognate detection. Stating that a word list of 200 items is not enough for

¹ Note that “sample size” here means all words-comparisons that are needed to establish an initial set of sound correspondences between two languages. Thus, what I to address here is the question of the size of the “additional material” which the Moscow school of historical-comparative linguistics requires as a backbone to establish genetic relationship with help of Swadesh’s (1955) list of 100 items (Dybo and Starostin 2008, Starostin 2013: 30–44).

the comparative method to successfully prove the genetic relationship between Armenian and English does not really solve the question. We still don't know how many words are needed for a successful application of the comparative method, neither in general, nor in this specific case. Such an estimate would, of course, depend on the genetic closeness of the languages being compared, and it would surely vary accordingly. Nevertheless, it would be helpful to know how many items one needs *at least* in order to successfully compare languages as divergent as, say, German and French.

Given the “manual” character of the comparative method, it is not easy to investigate the problem by simply applying the method to randomly varying sizes of a given word list. Not only would it be too time-consuming to conduct all the analyses, it would also be difficult to maintain objectivity when having the same sample of languages being investigated again and again by the same scholar. Fortunately, there are alternative ways to investigate the impact of sample size on cognate detection. We can, for example, use methods which do not rely on a manual application of the comparative method. Since the reason, why the comparative method relies so heavily on sample size is its language-specific similarity notion, it is enough to employ an automatic method for cognate detection that closely mimics the comparative method regarding the underlying notion of word similarity, and apply it to varying samples of a large gold standard containing cognate judgments taken from the literature.

2.1. Language-Specific and Language-Independent Similarities.

It is useful to make a distinction between language-specific and language-independent notions of word similarity. Language-specific similarity is hereby understood as similarity between words which is reflected in regular sound correspondences. Lass (1997: 130) calls this kind of similarity *genotypic* as opposed to *phenotypic similarity*, which is based on surface resemblances of phonetic segments. However, the most crucial aspect of this kind of similarity is that it is *language-specific*. It is never defined in general terms but always with respect to the language systems which are being compared. Correspondence relations can therefore only be established for individual languages, they can never be taken as general statements.

As an example, consider the two words English *mouth* [mauð] and German *Mund* [mont] “mouth”. From a language-specific perspective, these two words are maximally similar, since all correspondences, which are reflected in the alignment of the words, occur regularly, even the null-correspondence German [n] ≈ English [-] (Starostin 2010: 95). From a *language-independent perspective*, however, there are phonetically much more similar candidates to compare in both languages, such as, e.g., English *mount* [maunt], or German *Maus* [maus] “mouse”. In contrast to language-independent phenotypic similarities, language-specific similarities can never be proposed by relying on one word pair alone. This is the reason why the comparative method so heavily relies on the sample size: The smaller a sample is, the greater the possibility that it does not contain enough cognate words that make it possible to detect these specific similarities.

2.2. Language-Independent Approaches to Cognate Detection.

Most of the current automatic approaches to cognate detection employ a language-independent notion of similarity. The method by Turchin et al. (2010), for example, builds on Dolgopolsky's (1964) idea of *sound classes*. All words passed to the method are first converted to their respective Dolgopolsky sound classes and all words whose first two consonant classes match are assumed to be cognate. As an example, consider the two words English *mouth* [mauð] and German *Mund* [mont] “mouth”. Converting the words into their Dolgopolsky sound classes (vowels

being ignored), this yields the two strings “MT” and “MNT”. Since the first two consonant classes do *not* match, the method by Turchin et al. (2010) assumes that the words are not cognate.

As an alternative, alignment algorithms can be used to calculate the *edit distance* between two words. The edit distance between two words is defined as the smallest number of *edit operations* (*deletion, insertion, substitution*) needed to transform one word into the other (Levenstein 1965). This is equivalent to the Hamming distance of the alignment of two words (Hamming 1950). It can further be normalized by dividing it by the length of the longer word. Once the pairwise normalized edit distance (NED) is computed for a given pair of words, one can define a specific threshold below which the words are judged to be cognate. As an example, consider again the two words English *mouth* [mauð] and German *Mund* [mont] “mouth”. The optimal alignment of both words is:

$$\begin{array}{cccc} m & a & u & - & \theta \\ m & \upsilon & n & & t \end{array}$$

and the edit distance between both words is thus 3 (since they differ in three positions in the alignment). The normalized edit distance (NED) is $3 / 4 = 0.75$. Assuming a threshold of 0.6, the NED approach will also assume that both words are not cognate.

2.3. Language-Specific Approaches to Automatic Cognate Detection.

LexStat (List 2012a) is a new method for automatic cognate detection based on language-specific similarities. The method is implemented as part of a larger Python library for quantitative tasks in historical linguistics (List and Moran 2013) and can be downloaded from <http://www.lingpy.org>. LexStat takes multilingual (usually semantically aligned) word lists in IPA transcription as input and returns the same list with additional cognate judgments as output. The basic working procedure of the method consists of five stages:

- (1) sequence conversion,
- (2) preprocessing,
- (3) scoring-scheme creation,
- (4) distance calculation,
- (5) sequence clustering.

In stage (1), the input words are converted into tuples consisting of *sound classes* and *prosodic strings* (cf. List 2012b regarding the idea behind sound classes and prosodic strings). In stage (2), a simple language-independent method is used to derive preliminary cognate sets. In stage (3), a Monte-Carlo permutation test is used to create language-specific log-odds scoring schemes for all language pairs. In stage (4) the pairwise distances between all word pairs, based on the language-specific scoring schemes, are computed. In stage (5), the sequences are clustered into cognate sets whose average distance is beyond a certain threshold.

In addition to these five stages, all cognate sets detected by the method are aligned, using the SCA method for multiple phonetic alignment (List 2012b). As was shown in List (2012a), LexStat largely outperforms alternative methods that rely on language-independent similarities, such as the above-mentioned sound-class-based method proposed by Turchin et al. (2010), or alignment-based methods, such as normalized edit distance (NED). Given that LexStat closely mimics the comparative method regarding the underlying notion of word similarity, it seems to be a good candidate to test the impact of sample size on cognate detection.

3 Testing the Impact of Sample Size

3.1 Gold Standard.

In order to test to which degree language-specific methods for cognate detection depend on the samples size, an analysis of different, randomly created partitions taken from a newly compiled large gold standard was carried out. The gold standard consists of 550 items translated into four languages (German, English, Dutch, and French) which were taken from the Intercontinental Dictionary Series (Key & Comrie 2007). The orthographic entries in the original were converted into IPA transcriptions by the author, relying on one dictionary source for each language in order to maintain consistency. Cognate judgments were applied manually by consulting the respective literature (Kluge and Seebold 2002, Meyer-Luebke 1911, Pfeifer 1993, Vaan 2008, Wodtke 2008). Borrowings were coded in two ways: In the first coding, borrowed words were assigned to separate cognate sets. In the second coding, borrowings were assigned to the cognate sets to which they would belong if they were *not* borrowed. The second coding procedure is common in evolutionary biology where the term *homology* is used to indicate that two genes share a common history without specifying whether this common history is due to vertical inheritance or lateral transfer (Fitch 2000). For our experiment, it may be interesting to code borrowed words as cognates, since it may give us some hints whether and to which degree borrowing influences the results of language-specific cognate detection algorithms. For the downloadable gold standard, see Supplementary materials.

3.2 Test Samples.

With its 550 glosses translated into four languages, this gold standard is much larger than other publicly available datasets with respect to sample size. The data for the test was created as follows: Starting from the basic gold standard containing all 550 items, 550 new subsets of the data were created by randomly deleting 5, 10, 15, etc. items from the original dataset and taking 5 different samples for each distinct number of deletions. This process yielded 550 datasets, covering the whole range of possible sample sizes between 5 and 550 in steps of 5. These datasets were then analyzed, using the LexStat method, the method by Turchin et al. (2010), and the NED method (see List 2012a for details). In contrast to the NED and the LexStat method, the method by Turchin et al. (2010) does not need to be passed a specific threshold for cognate detection, since the threshold (two matching consonant classes) is inherent in the method itself. Choosing optimal thresholds for automatic cognate detection methods is not trivial, and no methods to automatically infer optimal thresholds are available. In order to apply a consistent criterion for threshold selection, the thresholds for NED and LexStat were calibrated on the results of the Turchin method. This was done by applying LexStat and NED to the largest sample, using several varying thresholds. Of all results, those thresholds were picked in which the number of false positives proposed by LexStat and NED came closest to the results of the Turchin method. This calibration procedure yielded an “optimal” threshold of 0.65 for NED, and a threshold of 0.625 for LexStat.

3.3 Evaluation Measures.

In applications of information retrieval it is common to evaluate algorithms by calculating their precision and recall. Precision refers to the proportion of items in the test set that also occur in the reference set. Recall refers to the proportion of items in the reference set that also occur in the test set (Witten and Frank 2005: 171). In the context of automatic cognate detection, a

high precision is equivalent to a low proportion of false positives, and a high recall is equivalent to a high proportion of correctly identified cognates. Precision and recall can be summarized by calculating their harmonic mean, the so-called *F-scores*, using the formula $2 \times (P \times R) / (P + R)$, where *P* is the precision and *R* is the recall.

Among different evaluation measures which have been proposed to estimate the accuracy of automatically induced cognate judgments (see Bergsma & Kondrak 2007), *B-Cubed* scores were chosen. *B-Cubed* scores were originally introduced as part of an algorithm by Bagga and Baldwin (1998), but Amigó et al. (2009) could show that they are especially apt as a clustering evaluation measure, and Bergsma & Kondrak (2007) showed that they are very useful to estimate the performance of cognate detection algorithms.

4. Results

4.1 Sample Size and General Accuracy of Automatic Cognate Detection.

The results of the general analysis (precision, recall, and F-scores for the cognate detection task) are plotted in Figure 1. As can be seen from the figure, the results of the two language-independent methods are quite similar regarding their tendency. After an initial phase of scattered results in those tests where the sample size is low, they stabilize and remain constant regardless of the sample size. The results for the language-specific LexStat analysis, on the other hand, clearly depend on the sample size. Both recall and F-Scores grow logistically and converge around a sample size of 200 items and 300 items, respectively. This nicely reflects the language-specific character of the LexStat method: If the word lists fed to the algorithm are too small, no language-specific similarities can be inferred, and no cognates can be detected, as reflected by the low recall and F-scores for small word lists. This changes dramatically once the sample size increases. Comparing the scores for a sample size of 50 items (F-score ca. 0.90) with those of 100 items (F-Score ca. 0.915), an increase of about 0.015 points can be attested, and between 100 and 300 items (F-Score cs. 0.93), there is still an increase of more than 0.02 points.

The scores for precision seem also to show a logistical growth, although it is not possible to determine a definite point of convergence for the given range of sample sizes. The drastic initial decrease of precision is a relic of the *B-Cubes* measure: If no cluster decision is being made, i.e. if all words are assigned to different cognate sets, the *B-Cubed* precision is 1, since no erroneous cluster decisions have been made. Since LexStat tends to leave most of the words unclassified if not enough evidence can be found to assign them to the same cluster, it automatically commits only a few erroneous decisions when dealing with small samples.

4.2 Optimal Sample Size and Genetic Closeness.

Figure 2 shows the results of the analyses for the Germanic languages in the sample. Basically, the results show a similar tendency as was observed for the analysis of all four languages. However, the increase in accuracy for the LexStat method is accelerated, and the convergence of the F-scores is reached at about 250 items (in contrast to 300 items in the full analysis). On the one hand, this illustrates the trivial fact that sample size directly depends on genetic closeness. On the other hand, it may seem surprising that the difference between the Germanic and the full sample is rather small (250 vs. 300 items in the F-Scores). One might argue that this is due to the fact that the Germanic languages also constitute the majority of the full sample. However, even when comparing further subsets like, for example, Dutch and Ger-

Figure 1: Comparing the performance of the methods for the cognate detection task. Y-axis shows the scores of the analyse, X-axis shows the size of the sample (number of basic vocabulary items).

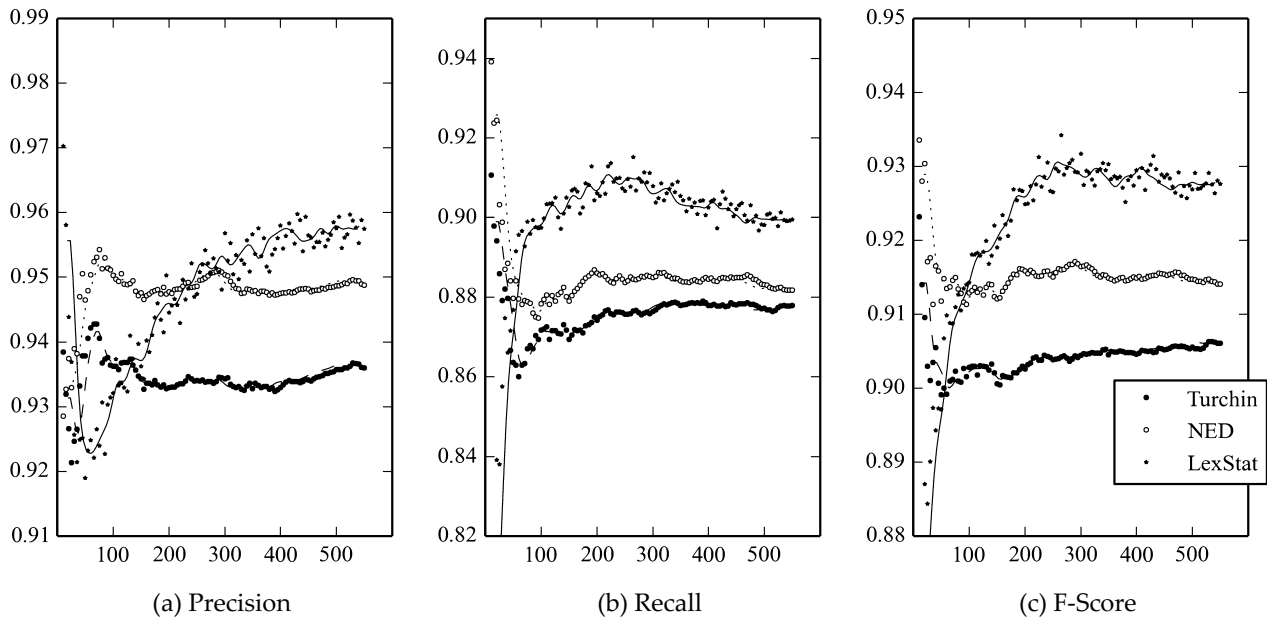
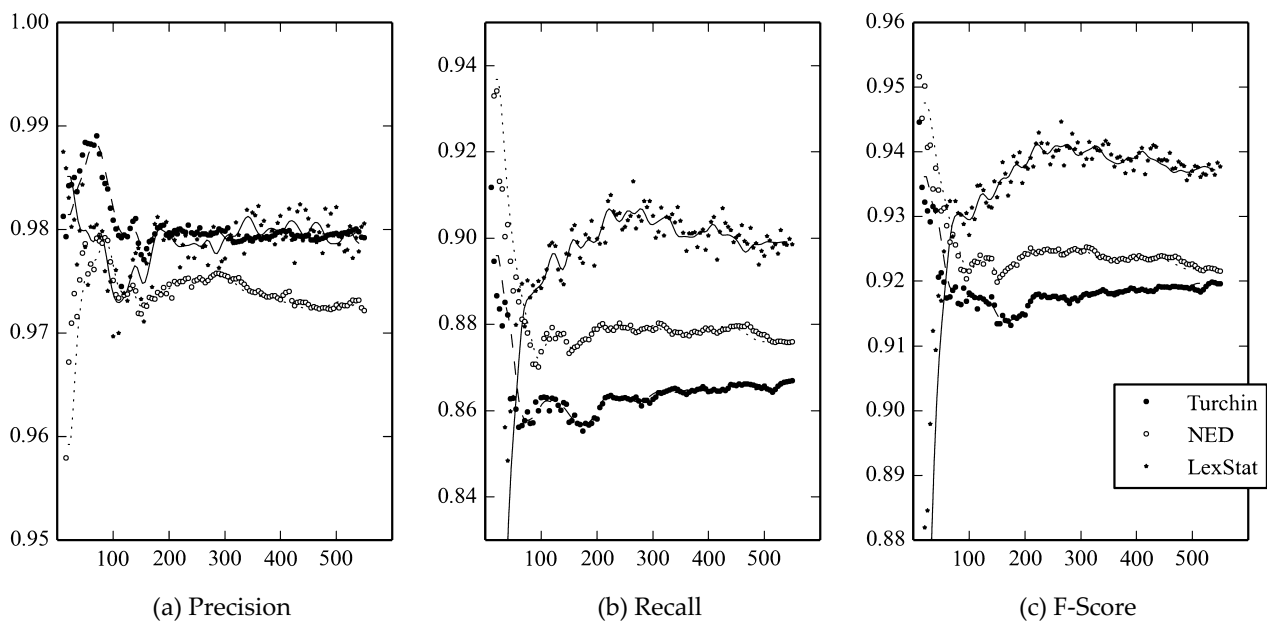


Figure 2: Comparing the performance of the methods on the subset of Germanic languages (English, German, Dutch).



man with English and French, the number of items when LexStat reaches convergence does not differ too much (250 vs. 300). It is possible that the number of 550 test items is still too small to conduct realistic tests on the impact of sample size on cognate detection. The current analyses might have missed interesting results which only show up when further increasing the sample size. Nevertheless, a striking difference between the full analysis and the Germanic subset is the increase in precision: While the precision of LexStat steadily increases in the full analysis along with the increase in sample size, it does not show this tendency in the Germanic subset where it quickly (at around 150–200 items) reaches a rather steady state. Since an increase in precision points to a decrease in false positives, this shows that for genetically close languages a much smaller sample suffices to achieve stable results.

Figure 3: Comparing the performance of the methods on the subset of English and French.

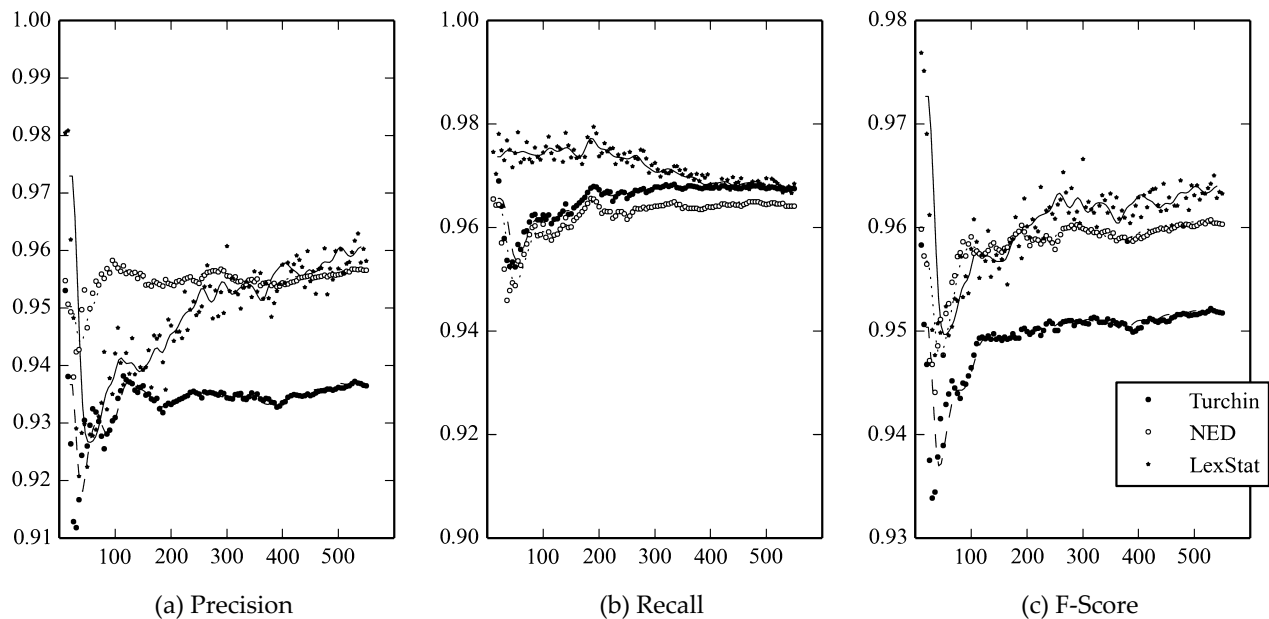


Figure 3 shows the specific results for the subset of English and French. These results are rather surprising: While precision steadily increases, recall shows a specific downtrend which starts at around 250 items. A decrease in recall corresponds to an increase in false negatives. With an increasing sample size, LexStat gets better in avoiding false positives, but at the same time gets worse in finding true cognates. One might think that this trend is somehow related to the increase of noise introduced by the large amount of borrowings from French into English. However, the same downtrend can also be observed when comparing the results for German and French, where the number of borrowings is much lower. Unfortunately, no further explanation for the results can be given at the moment.

4.3 Undetected Borrowings.

Given that undetected borrowings can yield a set of “wrong correspondences”, they need to be identified and filtered out when applying the comparative method. That borrowings have a definite influence on the results of automatic cognate detection analyses is illustrated in Table 2. Here, the evaluation scores of the application of the three methods to the largest sample (550 items) of the gold standard are given in two “flavors”. The first flavor is the performance on the traditional cognate detection task. The second flavor is the performance on the *homolog detection task*. In contrast to pure cognate detection, borrowings are explicitly included in this task, and the failure of a method to correctly identify borrowed words as homologs is penalized. As can be seen from the table, all three methods achieve a higher precision for the homolog detection task and a higher recall for the cognate detection task. This shows that both methods yield less false positives but more false negatives when no difference between borrowings and cognates is being made. The difference in precision, however, is much smaller for the LexStat method, and the F-Scores are higher for the cognate than for the homolog detection task, while they are identical in case of the NED and the Turchin method. This shows that the LexStat method handles noise arising from language contact much better than the NED method. Nevertheless, it also shows that LexStat can definitely be betrayed by large amounts of borrowings in the data. In terms of concrete numbers, of 176 borrowings with a direct donor in one of the languages, LexStat wrongly identifies 104 as cognates, NED 117, and the Turchin

Table 2: Comparing the performance of NED, Turchin, and LexStat in the cognate and the homolog detection task.

Method	Task	Precision	Recall	F-Score
NED	Cognates	0.95	0.88	0.91
	Homologs	0.98	0.86	0.91
Turchin	Cognates	0.94	0.88	0.91
	Homologs	0.97	0.86	0.91
LexStat	Cognates	0.97	0.90	0.93
	Homologs	0.98	0.86	0.92

method 140. This shows that borrowing definitely constitutes a problem for automatic cognate detection analyses.

An interesting question is whether stratification can make a difference in automatic cognate detection. In order to test this, a further test with the LexStat method was carried out. While the original LexStat method draws the attested distribution of possible sound correspondences from the *whole sample* it is given, the initial sample for the attested distribution was now restricted to basic vocabulary items drawn from the 100 and the 200 concept list proposed by Swadesh (1955 and 1952). The results for these analyses were compared to random trials. In these trials, the sample size was also restricted to 100 and 200 concepts, but the selection of concepts was carried out at random. The trials were repeated 50 times each, and the average of the results were compared with those obtained for the analyses based on sound correspondences derived from a stratification of the data.

Table 3: Comparing the impact of stratification on erroneous classification of English borrowings.

Items	Stratification	F-Score	Erroneously classified borrowings
100	random	0.85	0.28
	basic	0.86	0.17
200	random	0.88	0.35
	basic	0.87	0.19

Table 3 gives the proportion of misclassified French borrowings in English in the two analyses. As can be seen clearly, the number of erroneously classified borrowings is much lower in the analyses in which the initial sample was based on proper “basic vocabulary” than for randomly selected words pairs. This seems to indicate that stratification can indeed make a difference, also in automatic cognate detection. However, comparing the low F-Scores with those obtained for analyses in which the full sample was used also shows that a lot of interesting signal is lost. Further research is needed to find the right balance between signal loss resulting from stratification and unwanted noise resulting from large samples.

5. Discussion

The results reported in this study may be a bit disappointing, since it is not clear what they actually tell us. We still don’t know the lower bound of words needed for a successful applica-

tion of the comparative method. We also don't find direct evidence for an upper bound, not to speak of the specific results of the analyses, which are generally difficult to explain. However, what the results definitely show is that word list size definitely *has* an impact on the results and that stratification *cannot be ignored*. More research with larger samples (both regarding the number of languages and the number of test items) is needed to shed light on the problems that were discussed in this study.

Supplementary materials are available from:

- <http://johr.ru/article.php?id=134>
- <https://gist.github.com/LinguList/8235795>

The zip-archive includes:

- `readme.md`, a short description of the data-format;
- `ids.qlc`, the gold standard in QLC-format.

Literature

- AMIGÓ, E., J. GONZALO, J. ARTILES, and F. VERDEJO (2009). “A comparison of extrinsic clustering evaluation metrics based on formal constraints”. In: *Information Retrieval* 12.4, 461–486.
- BAGGA, A. and B. BALDWIN (1998). “Entity-based cross-document coreferencing using the vector space model”. In: *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. “COLING-ACL ’98” (Montréal, Quebec, Canada, Aug. 10–14, 1998)*. Association of Computational Linguistics, 79–85.
- BERGSMA, S. and G. KONDRAK (2007). “Multilingual cognate identification using integer linear programming”. In: *RANLP Workshop on Acquisition and Management of Multilingual Lexicons*. Ed. by The International Conference on Recent Advances in Natural Language Processing. Borovets, Bulgaria.
- DURIE, M., ed. (1996). *The comparative method reviewed. Regularity and irregularity in language change*. With an intro. by M. D. Ross and M. Durie. New York: Oxford University Press.
- DYBO, A. and G. STAROSTIN (2008). “In defense of the comparative method, or the end of the Vovin controversy”. In: *Aspekty komparativistiki*. Vol. 3. Ed. by I. S. Smirnov. Orientalia et Classica XI. Moscow: RGGU, 119–258. <http://starling.rinet.ru/Texts/compmeth.pdf>.
- FITCH, W. M. (2000). “Homology. A personal view on some of the problems”. In: *Trends in Genetics* 16.5, 227–231.
- FOX, A. (1995). *Linguistic reconstruction. An introduction to theory and method*. Oxford: Oxford University Press.
- HAMMING, R. W. (1950). “Error detection and error detection codes”. In: *Bell System Technical Journal* 29.2, 147–160.
- KESSLER, B. (2001). *The significance of word lists. Statistical tests for investigating historical connections between languages*. Stanford: CSLI Publications.
- KEY, M. R. and B. COMRIE, eds. (2007). *IDS — The Intercontinental Dictionary Series*: <http://lingweb.eva.mpg.de/ids/>.
- KLUGE, F., found. (2002). *Etymologisches Wörterbuch der deutschen Sprache*. Cont. by E. Seebold. 24th ed. Berlin: de Gruyter.
- LASS, R. (1997). *Historical linguistics and language change*. Cambridge: Cambridge University Press.
- LIST, J.-M. (2012a). “LexStat. Automatic Detection of Cognates in Multilingual Wordlists”. In: *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH (Avignon, France, Apr. 23–24, 2012)*. Association for Computational Linguistics, 117–125.
- LIST, J.-M. (2012b). “SCA. Phonetic alignment based on sound classes”. In: *New directions in logic, language, and computation*. Ed. by M. Slavkovik and D. Lassiter. LNCS 7415. Berlin and Heidelberg: Springer, 32–51.
- LIST, J.-M. and S. MORAN (2013). “An open source toolkit for quantitative historical linguistics”. In: *Proceedings of the ACL 2013 System Demonstrations. (Sofia, Bulgaria, Aug. 4–9, 2013)*. Association for Computational Linguistics, 13–18. <http://aclweb.org/anthology/P/P13/P13-4003.pdf>.
- MEYER-LÜBKE, W., comp. (1911). *Romanisches etymologisches Wörterbuch*. Sammlung romanischer Elementar- und Handbücher 3.3. Heidelberg: Winter.
- PFEIFER, W., ed. (1993). *Etymologisches Wörterbuch des Deutschen*. 2nd ed. 2 vols. Berlin: Akademie. <http://www.dwds.de/>

- STAROSTIN, G. (2010). "Preliminary lexicostatistics as a basis for language classification: A new approach". In: *Journal of Language Relationship* 3, 79–116.
- STAROSTIN, G. S. (2013). *Jazyki Afriki. Opyt postroenija leksistatističeskoj klassifikacii* [The languages of Africa. Experience in establishing a lexicostatistical classification]. Vol. 1: Metodologija. Kojanskie jazyki [Methodology. Khoisan languages]. Moscow: Jazyki Russkoj Kul'tury
- SWADESH, M. (1952). "Lexico-statistic dating of prehistoric ethnic contacts. With special reference to North American Indians and Eskimos". In: *Proceedings of the American Philosophical Society* 96.4, 452–463.
- SWADESH, M. (1955). "Towards greater accuracy in lexicostatistic dating". In: *International Journal of American Linguistics* 21.2, 121–137.
- TRASK, R. L., comp. (2000). *The dictionary of historical and comparative linguistics*. Edinburgh: Edinburgh University Press.
- TURCHIN, P., I. PEIROS, and M. GELL-MANN (2010). "Analyzing genetic connections between languages by matching consonant classes". In: *Journal of Language Relationship* 3, 117–126.
- VAAN, M. (2008). *Etymological dictionary of Latin and the other Italic languages*. Leiden Indo-European Etymological Dictionary Series 7. Leiden and Boston: Brill.
- WITTEN, I. H. and E. FRANK (2005). *Data mining. Practical machine learning tools and techniques*. 2nd ed. Amsterdam et al.: Elsevier.
- WODTKO, D., B. IRSLINGER, and C. SCHNEIDER, eds. (2008). *Nomina im Indogermanischen Lexikon*. Heidelberg: Winter.

Й.-М. ЛИСТ. К вопросу о влиянии размера лексической выборки на обнаружение этимологических когнатов.

В статье исследуется вопрос об оптимальном размере словарного списка, на котором можно было бы апробировать различные методы детекции этимологических когнатов. Чтобы получить ответ на этот вопрос, был разработан «золотой стандарт» из 550 концептов, переведенных на 4 языка (английский, немецкий, голландский, французский); внутри этого списка было выделено несколько последовательно увеличиваемых подмножеств. Применение автоматических методов детекции когнатов к этому стандарту показывает, что степень точности методов, разработанных для конкретных языковых типов, явно зависит от размера списка. Учитывая, однако, что оптимальный размер зависит от столь различных факторов, как степень генетической близости языков и масштаб ареальных контактов между сравниваемыми языками, нельзя сказать, что анализ позволяет определить универсальную верхнюю или нижнюю границу списка.

Ключевые слова: сравнительный метод, лексикостатистика, этимология, компьютерная лингвистика.

Toward the reconstruction of Proto-Na-Dene

The paper presents the author's current version of the reconstruction of the phonological system of Proto-Na-Dene (PND = Proto-Athabaskan-Eyak-Tlingit in J. Leer's terms), based on comparison of three groups of languages: 1) Tlingit dialects, 2) Eyak and 3) Athabaskan languages (Proto-Athabaskan). Eyak and the Athabaskan languages are quite close to each other and are traced back to an intermediate Proto-Eyak-Athabaskan language. Regular phonetic correspondences between Eyak and PA have received an original interpretation by Michael E. Krauss and Jeffrey Leer, including very complicated correspondences of sonorants. In his works, J. Leer proposed a PND reconstruction that explained most of the regular sound correspondences between the Na-Dene languages. Although Leer's reconstruction is quite seductive with its apparent simplicity, in some aspects this simplification is unwarranted, as the real situation turns out to be a lot more complicated. This is possibly a consequence of the number of the roots involved: Leer's reconstruction is based on a relatively short list of cognate sets (ca. 300), whereas the author of the current paper has tried to take into account the entire comparative corpus (ca. 800 sets). Due to volume restrictions, the paper consists of only a brief summary of the reconstruction and an illustrative subset of the comparative material, dealing with certain complicated sound correspondences between front and lateral affricates/fricatives, previously analysed in a different light by J. Leer.

Keywords: Na-Dene languages, Athabaskan languages, historical phonology.

Reconstruction of Proto-Na-Dene (= Proto-Athabaskan-Eyak-Tlingit, PAET in Jeff Leer's terms) is based on comparison of three groups of languages: 1) Tlingit dialects (TI), 2) Eyak (E) and 3) Athabaskan languages (PA = Proto-Athabaskan)¹. Eyak and the Athabaskan languages are close to each other and are traced back to an intermediate Proto-Eyak-Athabaskan language (PEA = PAE of Jeff Leer). The regular phonetic correspondences between Eyak and PA were interpreted by Michael E. Krauss and Jeffrey Leer, including very complicated correspondences of sonorants (Krauss & Leer 1981). Leer (1992; 2008; 2008a) has proposed a PND reconstruction, explaining most of the regular sound correspondences between the Na-Dene languages. Leer's reconstruction is quite seductive with its apparent simplicity, but in some aspects this simplification is unwarranted, as the real situation turns out to be a lot more complicated.

Volume restrictions permit me little more than a brief digest of the Proto-Na-Dene reconstruction; the material listed below deals only with certain complicated sound correspondences of front and lateral affricates/fricatives that have been, in general, previously analysed by Leer 2008; 2008a.²

¹ A close genetic relationship between Tlingit, Eyak and Athabaskan has been proven to general satisfaction (Krauss 1964, 1965, 1968, 1969; Krauss & Leer 1981; Leer 1979, 2008, 2008a). The affiliation of Haida with the Na-Dene family remains rather speculative. John Enrico has presented an annotated list of presumable cognates (Enrico 2004), for the most part consisting of early Athabaskan (or even Eyak-Athabaskan) borrowings. This means that Haida data may at least be of additional help with the Eyak-Athabaskan reconstruction.

² My Na-Dene data base, available online at <http://starling.rinet.ru>, contains a collection of relevant data (PND and PA reconstructions; external comparison; detailed comments and references). The main sources of my

The Proto-Athabaskan forms in this paper are generally established on the basis of the “regional” protoforms in Leer 1996. Positional alternations of voiced and voiceless fricatives have been unified: *S = *S/*Z. The secondary clusters of the *ʔS type are presented as *C. *N = *n ~ *ń ~ *m. Krauss and Leer’s PA final glottalized sonorants (R’) are presented as Rʔ. PA orin nasal and glottalized (“constricted”) long vowels are transliterated as a vowel plus N/? : *V:N instead of *V̥:; *V:ʔ, V:Nʔ instead of *V̥:; V̥: . PA *ñ/ỹ/ŋ₁ and *ŋ^w/w̃/ŋ₂ (in various works by Krauss and Leer) are replaced with *ń, *m respectively.

The sign “~” in PND and PA reconstructions denotes alternative rather than alternating variants. Etymological variants are divided with a comma.

In the Proto-Na-Dene forms: root-medial E = *i ~ *ä; root-final E = *i ~ *ə; root-medial A = *a ~ *ə.

Tlingit forms in Leer 1975 contain information on Pre-Tlingit prosody: three phonations on long vowels. In my notation: *h* — fading phonation; ? — clipped phonation; “:” — sustained phonation³; “;” — Pre-Tlingit phonation may have been sustained or glottalized; [?] — Pre-Tlingit phonation is not determined.

Vowel correspondences

Straightforward comparison of Tlingit vowels with Eyak and Athabaskan generates an enormous number of non-interpretable series. Comparison of the PEA reconstruction with Tlingit is not more informative: any PEA vowel can correspond to any vowel in Tlingit, because any Eyak vowel can correspond to any Athabaskan vowel (except for the CV-type roots, where correspondences are trivial). Only comparison of Tlingit and Eyak gives a reasonable system. Provisionally, for PND roots I reconstruct final vowels that are almost completely lost in descendant languages, but may have left traces in the coloring of PA vowels.

I assume that PND roots had a basic shape of CV or (CV)CV(C)CV (there was possibly a CVC structure, but there is no way to distinguish it from CVCV). In most contemporary Na-Dene languages stress is placed on the first root vowel; unstressed vowels are reduced up to complete deletion. Secondary voiced fricatives and glottalized consonants are reconstructed in PA not only between vowels, but also in root-final position; this supposes an original intervocalic position. Many PA noun roots end in *-ə which probably reflects PND root-final vowels, e.g. *č^wəʔə ‘underbrush; willow’, *čəwʒə ‘blowfly maggot’, *gi:ʔgə ‘berry, berries’, *gəš^wə ‘fork’, *tu:qə ‘fish, salmon’, *še:nə ‘summer’, *t’a:qə ‘three’, *təńə ‘path, trail’, *təʒə ‘night’, *t’e:də ‘teenage girl’, *χayə ‘winter’ (Krauss 1979; Krauss & Leer 1981). Root-final vowels are admittedly preserved before suffixes in Tlingit, Eyak and PA; because of this, many suffixes have a secondary shape of VC. However, it is hard to directly reconstruct the quality of root-final vowels, since they are usually represented by a neutral sound. In PA suffixes with the structure VS a unified *ə is reconstructed. But in some Athabaskan languages reflexes of the “subordinate ə” differ from reflexes of the “root ə”. This may be a generalizing of different root-final vowels in the

reconstruction are: **Tlingit**: default source is Leer 1975 (a morphophonological representation of Tlingit wordforms with a reconstruction of phonations). Abbreviations: ITI = Inner Tlingit; NTL = Northern Tlingit; Yak = Yakutat; Tg = Tongass. Additional sources on NTL: Gillian L. Story, Constance M. Naish. Tlingit verb dictionary. Fairbanks, 1973; Constance Naish, Gillian Story. English-Tlingit dictionary. Nouns. Fairbanks, 1963; Keri Edwards. Dictionary of Tlingit. Juneau, 2009. **Eyak**: Krauss 1970. **Proto-Athabaskan**: default source is Leer 1996. My additional reconstructions are based on the data of Athabaskan dictionaries.

³ The same notation of these phonations is used for Eyak.

history of separate languages. Eyak possibly preserves ancient root-final vowels in semantically “empty” nominal suffixes *-eh*, *-uh*, *-ah*. The Tlingit and Eyak roots (except for CV) consistently end in consonants; therefore, I rely on Proto-Athabaskan, where preserved final vowels could cause Umlaut on the root-medial vowel.

Three PND root-final vowels are reconstructed, assuming assimilation of the PA root-medial vowels to three root-final vowels (**i*, **a*, **ə*). The PND medial vowels are freely combined with any of the final ones.

My reconstruction of vowels is shown in Table 1.

Table 1. Vowel correspondences and PND reconstruction.

PND	Tlingit	Eyak	Proto-Athabaskan			
			in mono-syllables	before * <i>-i</i>	before * <i>-ə</i>	before * <i>-a</i>
* <i>i</i>	<i>i, e (u)</i> ⁴	<i>i, e (u) / ə (u)</i>	<i>e: (u): / ə (u)</i>	<i>i: (u): / ə (u)</i>	<i>e: / ə (u)</i>	<i>a: / a</i> ⁵
* <i>ä</i>	<i>i, e (u)</i>	<i>a / ə</i>	<i>a: / a</i>	<i>i: (u): / ə (u)</i>	<i>e: / ə (u)</i>	<i>a: / a</i>
* <i>ə</i>	<i>a</i>	<i>i, e (u) / ə (u)</i>	<i>e: (u): / ə (u)</i>	<i>i: (u): / ə (u)</i>	<i>e: / ə (u)</i>	<i>a: / a</i>
* <i>a</i>	<i>a</i>	<i>a / ə</i>	<i>a: / a</i>	<i>i: (u): / ə (u)</i>	<i>e: / ə (u)</i>	<i>a: / a</i>

Tlingit *i/e* may reflect two different “Pre-Tlingit” phonemes, but the distribution of these vowels in dialects seems to be chaotic. A dialectal (?) development *a(:) > e(:)* in some roots remains unexplored. Historical distribution of *i/e* in Eyak is also unclear, contrary to Leer’s opinion.

The correspondence Tl *i/e* — E *a* is considered by Leer as a manifestation of qualitative ablaut. But there are almost no traces of such an ablaut in separate groups, in contrast to the quantitative one (Krauss & Leer 1981: 78–93), and there is no reason to explain a trivial vowel correspondence with its help.

There are no grounds for the reconstruction of PND rounded vowels. The vowel *u* does not form an independent series of correspondences. Leer’s series **o* (Tl *u* — E *a*) is a particular case of development of PND **ä* (Tl *i/e* — E *a*). The correspondence “Tl, E, PA *u*” (Leer’s PND **u*) is too scarce: usually *u* is observed in one or two languages, when the other(s) has (have) an unrounded vowel. Thus, the *u(:)* is not original in any of the groups, being a result of assimilation to adjacent labialized and labial consonants, including a hypothetical sequence *CV_wCV. Vowels turn into an *u(:)* independently in all three groups, but the details of this process remain unstudied. Labialization of obstruents often disappears before [e(:)] in Tlingit and is preserved more often before [a(:)]. Variants such as *iK^w/[uK^w]*, *K^wi/[K^wu]*⁶ fluctuate from one Tlingit dialect to another. Tlingit also has a regular reflex **wi > wu*⁷. A labialization process in history of Eyak and Athabaskan was apparently similar to the one in Tlingit. It is important to note that there is no vowel labialization in the vicinity of PND sibilants and lateral affricates / fricatives either in Tlingit or in Eyak-Athabaskan, unless they develop into labialized palatal/velar consonants — this fact may determine relative chronological frames of vowel labialization.

Cf. Leer’s reconstruction in Table 2 (according to Leer 2008a: 173 in my transliteration).

⁴ The slash mark separates reduced variants from full ones; labialized variants of vowels are given in brackets.

⁵ Reflexes of PA **a* are unstable in many Athabaskan languages and are often replaced with reflexes of **a*. Therefore, a PA reconstruction of **ə* is often encountered in the place of **a* because of insufficient data.

⁶ Velar and uvular consonants in Tlingit automatically receive labialization before and after the *u*.

⁷ The shift *aK^w > uK^w*, *wa > wu* is strictly dialectal.

Table 2. Vowel correspondences and PAET reconstruction by Leer.

PAET	Tlingit	PAE	Eyak	PA
*i(:)	i	*i(:)	i(:), ə	*i(:), *ə
*i(:) next to rounded consonant	u	*e(:)	e(:), ə	*e:, *ə
*i(:) next to uvular	i~e	*e(:)	e(:), ə	*e:, *ə
*e(:)	a	*e(:)	e(:), ə	*e:, *ə
*a(:)	a	*a(:)	a(:), ə	*a:, *a
*o(:)	u	*a(:)	a(:), ə	*a:, *a
*u(:)	u	*u(:)	u(:), ə	*u:, *u

I omit here any reconstruction of PND prosody — vowel quantity, tones and/or phonations. Tlingit and Eyak are very similar in respect to this: short/long (or full/reduced) vowels and three phonations (clipped, fading and sustained) in long vowels are symmetric with three laryngeal features of occlusives: clipped≈glottalized, fading≈aspirated and sustained≈neutral. This system is preserved in Eyak and in some Tlingit dialects. These phonations turn into tones in Tlingit dialects: clipped phonation turns into rising tone; fading phonation into falling tone; sustained phonation turns into one of the tones depending on dialect. Proto-Athabaskan had short/long vowels and clipped/neutral phonations in vowels of either quantity. In Athabaskan languages clipped phonation turns into rising tone and sustained phonation into falling tone (in “high-marked” languages), or vice versa (in “low-marked” languages). There is no evidence of any interaction between phonations and laryngeal features of obstruents in Tlingit and Eyak. In Athabaskan *-V(:)C' > *-V(:)?S before consonants and in word-final position; a glottal stop precedes glottalized obstruents after short vowels and arises before the so called glottalized sonorants (clusters *R?* in my notation⁸) and more or less functions as a segmental phoneme. In other words, the Eyak-Athabaskan clipped phonation was no more than an ordinary word-medial / final *ʔ after vowels and sonorants. But for the time being, a straightforward comparison of the Tlingit and Eyak-Athabaskan phonations gives no basis for a reliable PND reconstruction.

Sonorant correspondences

A velar sonorant /Y/ has presumably recently gone extinct in Tlingit; this was a consonant distinct from /y/ at least in root-initial position. Nowadays they differ only on the morpho-phonological level: *Y > {Y} (represented as positional alternation between *y* and *w*), *y- > {y-} (non-alternating *y-*). There are examples of secondary development *y- > {Y-} in several Tlingit dialects.

My reconstruction of sonorants is shown in Table 3.

The clusters *nC, *mC keep their nasalization in Eyak and/or Athabaskan, but *NC > C in Tlingit (except for NC < *NVC). Rules of preservation/loss of nasals in NC-clusters both on the Eyak and the Athabaskan side remain to be ascertained. Since PEA nasal sonorants go back not only to PND *n, *m, but also to PND *l, *ŋ, and *ŋ^w, reconstruction of PND *nC is theoretically equal to *lC ~ *nC ~ *ŋC and reconstruction of *mC is equal to *mC ~ *ŋ^wC. PND *mC is re-

⁸ These PA units could be reconstructed as PND glottalized sonorants if they had consistent parallels in Tlingit. In this case Eyak initial clusters ?l-, ?n, ?y-, ?w- would be worth interpreting as reflexes of PND *l', *n' etc.

Table 3. Sonorant correspondences and PND reconstruction.

PND	Tlingit	Eyak ⁹	PA
*l	Y-, Y/0 ¹⁰	l, -V̥	*n-, n(0)
*m	n	w-, 0	*m
*n	n	n/l, -V̥	*n
*ń	?-, n	?-, -V̥	*ń
*ŋ	Y-, Y/0	?-, 0	*ń
*ŋ ^w	Y-, Y/0	w-, 0	*m
*w	w-, 0/w	w-, 0/w	*w-, w/0
*y	y-/Y-, 0/Y	y-, 0	*y-, y/0

Table 4. Sonorant correspondences and PAET reconstruction by Leer.

PAET	Tlingit	PAE	Eyak	PA
*ń	Y	*ń	y	*ń
*n	n	*n	l~n	*n
*ŋ	Y	*n	l~n	*n
*ŋ ^w	w	*n	l	*n
*y	Y	*y	y	*y (~*ń)
*w	w, Y	*w	w	*w

constructed whenever 1) **m* labializes the preceding vowel; 2) there is a development “early PA” **Ci(:)mC* > PA **Cu:nc* (= **Cu:C* in Leer’s notation). Elsewhere I reconstruct PND root-medial clusters like *wN*, *yN* in a few roots where non-trivial sound correspondences force me to do so. Despite the fact that the sound correspondences are generally established, defining the quality of medial-root sonorants is often difficult, unless they have evident reflexes (such as *-n-* in all the subgroups). Many PND roots reconstructed as **CV* may have really had a more complicated structure, i.e. **CVyV*, **CVwV*, **CV?(^w)V*, **CVh(^w)V*. Truly monosyllabic nominal roots often show metathesis in Tlingit: *?uhχ* (cf. PA **χu*: ‘tooth’); *?ihχ* ‘oil’, cf. PA **χe*: ‘grease, oil, tallow’, etc. Sometimes a former bisyllabic structure may be guessed on the basis of irregular vowel correspondences. But reconstruction of all remaining “monosyllabic” roots as **CVHV* [*H* = *y*, *w*, *h(^w)*, *?(^w)*] seems superfluous. Cf. Leer’s sonorant reconstruction in Table 4 (according to Leer 2008a: 172 in my transliteration).

Obstruent correspondences

My interpretation of sound correspondences is presented in Table 5. A transcription is accepted for the Na-Dene languages, in which “voiced” letters denote non-aspirated voiceless or

⁹ Eyak has complicated reflexes of sonorants in intervocalic position. In particular, hiatus (when a sonorant is deleted) is eliminated by means of *-y/-w/-/?-* depending on preceding vowels; *-n-* > *-m-* after certain labializing segments. Note also that PND **wVN(C)* > E *mV(C)* > *mV(c)* (Krauss & Leer 1981).

¹⁰ Here and further in the tables initial, medial and final reflexes are shown for roots rather than wordforms (where root-initial reflexes can be found in medial position).

Table 5. Obstruent correspondences and PND reconstruction.

PND	Tlingit	Eyak	PA	Statistics (initial/ medial)
*d	t-/d-, d	d	*d	52 ¹¹ /30
*t-	t-/d-	t-	*t-	41
*t'	t', t'/d	t'	*t'	20/30
*ʒ	š-, ʒ	ʒ	*ʒ	3/11
*c-	š-	c-	*c-	12
*c'	c', c'/ʒ	c'	*c'	6/4
*s	s-	s	*s	16/5
*s'	s'	c'	*c'-, c'/s	16/21
*ʒ ^w	g ^(w)	ʒ	*ʒ	6/6
*c ^w	k ^(w) -, k ^(w) /g ^(w)	c'	*c'	3/3
*c ^w -	k ^(w) -/g ^(w) -	c-	*c-	11
*s ^w	x ^(w)	s	*s	14/3
*s ^w	x' ^(w)	c'	*c'-, c'/s	2/10
*ʒ	c-/ʒ-, ʒ	ʒ	*ʒ	8/4
*č-	c-/ʒ-	č-	*č-	7
*č'	c'	č'	*č'	2/7
*ś	s	š	*ś ^w	6/7
*ś' see *ś ^v 12				
*ś ^w (?)	s-/š-	x-	*x'-	3/—
*ʒ	č-/ʒ-, ʒ	ʒ(/ʒ) ¹³	*ʒ	11/12
*č-	č-	č-	*č-	19
*č'	č'	č'(/c')	*č'	15/1
*š	š	š(/s)	*š ^w	19/19
*ś' (and *ś ^v ?)	s'	č'(/c')	*č'-, č'/ś ^w	15/19
*ʒ ^w	k ^(w) -/g ^(w) -, g ^(w)	g ^(w)	*ʒ ^w	12/4
*č ^w -	k ^(w) -	k ^(w) -	*č ^w -	15
*č ^w	k ^(w) -, k ^(w) /k ^(w)	k ^(w) -	*č ^w -	8/—
*ś ^w	x ^(w) -	x ^(w) -	*ś ^w -	4/—
*ś ^w	x' ^(w)	x' ^(w)	*č ^w -, č ^w /ś ^w	8/4
*L	l-/λ-, l/ł	l/ł	*L/ł	4/8
*λ-	λ-/ł-	λ-	*λ-/ł-	15

¹¹ An enormous abundance of initial *d- and *t- can be explained by assuming the presence of the fossilized classifier d- in verbal roots.

¹² Reflexes of presumed *ś' and *ś^v do not differ from each other. In PEA, they had to merge in *č'; Tlingit predictably has no /ś'/, since it must have merged with /s'/ at some point in Tlingit history. Thus, some parts of roots with PND *ś' may indeed have had PND *ś^v.

¹³ Eyak sibilants correspond to Tlingit and Athabaskan hissing sounds in a few roots and are unlikely to form an additional series.

PND	Tlingit	Eyak	PA	Statistics (initial/ medial)
*λ'	λ', λ'/L	*λ'	*λ'	9/10
*ɬ	ɬ	ɬ	*ɬ	20 ¹⁴ /34
*ɮ	ɮ	λ'	*λ'-, λ'/ɬ	13/25
*L ^w	L-, l(/ɬ)	c-, ʒ/s	*c-, ʒ/s	2/4
*λ ^w -	λ-	c-	*c-	8
*λ ^w	λ'	c'	*c'	5/4
*ɮ ^w	ɬ	s	*s	8/5
*ɮ ^w	ɮ	c'	*c'-, s'/s	5/5
*g	k-/g-, g	g	*g'	14/10
*k-	k-/g-	k-	*k-	14
*k'	k'-, k'/k	k'	*k'	4/6
*x	x	x	*x'	9/3
*x'	x'	k'	*k'-, k'/x	4/7
*g ^w	k ^(w) -/g ^(w) -, g ^(w)	g ^(w)	*g' ¹⁵	18/18
*k ^w -	k ^(w) -/g ^(w) -	k ^(w) -	*k-	18
*k ^w	k ^(w) '-, k ^(w) '/k ^(w)	k ^(w) '	*k'-, k'/x	5/12
*x ^w	x ^(w)	x ^(w)	*x'	17/6
*x ^w	x ^(w) '	k ^(w) '	*k'	7/12
*G	G-/q-, G	G	*G	32/19
*q-	q-	q-	*q-	22
*q'	q'-, q'/q	q'	*q'	6/5
*χ ¹⁶	χ	χ	*χ	15/13
*χ'	χ'	q'	*q'-, q'/χ	15/13
*G ^w	G ^(w)	G	*G ^(w)	7/8
*q ^w -	q ^(w) -	q-	*q ^(w) -	11
*q ^w ?	?	q'	*q ^(w) '	(PEA 1/1)
*χ ^w	χ ^(w) '	q'	*q ^(w) '-, q ^(w) '/χ ^(w)	5/14
*χ ^w	χ ^(w)	χ	*χ ^(w)	15/8
*ʔ ¹⁷	ʔ-	ʔ-	*ʔ-	41/—
*ʔ ^w - ¹⁸	ʔ-	w-	*w-	8/—
*h-	h-	ʔ-	*h-	15/—
*h ^w - ¹⁹	h-	w-	*h-	11/—

¹⁴ Some verb roots with initial *ɬ- may contain the fossilized classifier ɬ-.

¹⁵ Original labialization is reflected as a change *i:, *o > PA *u(:).

¹⁶ PND *χ, *χ^w > Tl h/0 before a consonant in clusters.

¹⁷ Reflexes of root-medial *ʔ, *ʔ^w, *h, *h^w do not allow any reliable reconstruction, as they have been obscured by clipped and fading phonations.

¹⁸ Since there is no separate PND *q^w- with uvular reflexes, this correspondence may also be interpreted as reflecting PND *q^w-.

¹⁹ This correspondence can also be interpreted as PND *f-.

Table 6. Obstruent correspondences and PAET reconstruction by Leer.

PAET	Tlingit	PAE	Eyak	PA
*d	d	*d	d	*d
*t	t	*t	t	*t
*t'	t'	*t'	t'	*t'
*t̥	t̥	*t̥	t̥	*t̥-λ
*λ'	ʔ, λ'	*λ'	λ'	*λ'
*s	s	*s	s	*s *z
*c	c	*c	c	*c
*c'	s', c'	*c'	c'	*c'
*š	š (s)	*š	š (s)	*š ^w ~ *ž ^w
*ž	ž	*ž	ž	*ž
*č	č (c)	*č	č	*č
*č'	s', č' (c')	*č'	č'	*č'
*x	x	*s	s; š	s ~ z
*g	g	*ž	ž	*ž, *s ~ *z
*k	k; š	*c	c	*c
*k'	k'	*c'	c'	*c'
*x ^w	x ^(w)	*x	x	*x ~ *y
*g	g ^(w)	*g	g	*g
*k	k ^(w)	*k	k	*k
*k'	x' ^(w) , k' ^(w)	*k'	k'	*k'
*x ^w	x ^(w)	*x ^w	/x ^w / > x	*š ^w ~ *ž ^w
*g ^w	g ^(w)	*g ^w	/g ^w / > g	*ž ^w
*k ^w	k ^(w)	*k ^w	/k ^w / > k	*č ^w
*k' ^w	x' ^(w) , k' ^(w)	*k' ^w	/k' ^w / > k'	*č ^w
*χ	χ ^(w)	*χ	χ	*χ ~ *ʔ
*G	G ^(w)	*G	G	*G
*q	q ^(w)	*q	q	*q
*q'	χ' ^(w) , q' ^(w)	*q'	q'	*q'
*χ ^w	χ ^(w)	*χ ^w	χ	*/χ ^w / > *χ ~ *ʔ
*G ^w	G ^(w)	*G ^w	G	*/G ^w / > *G
*q ^w	q ^(w)	*q ^w	q	*/q ^w / > *q
*q' ^w	χ' ^(w) , q' ^(w)	*q' ^w	q'	*/q' ^w / > *q'
*ʔ	ʔ	*ʔ	ʔ	*ʔ
*h	h	*(h)	0	*(h)
*ʔ ^w	ʔ	*ʔ ^w	w	ʔ
*hw	h	*w	w	*w
*ʔ ^y	ʔ	ʔ	ʔ	*y

voiced occlusives²⁰ (D = [T] or [D]), whereas “voiceless” letters denote voiceless aspirated occlusives (T = [T^h]). The same orthographical principle is used in reconstructions, especially because we do not know whether the non-aspirated occlusives were voiced or voiceless. PND aspirated occlusives were present only in root-initial position along with non-aspirated and glottalized occlusives; in root-medial position only non-aspirated and glottalized occlusives were opposed²¹.

The Tlingit root-initial non-aspirated (D) or aspirated (T) reflexes of occlusives and affricates alternate even within the same dialect (e.g. NtI *túqʔ* ~ *dúqʔ* ‘young spruce or hemlock’). This dialectal fluctuation is accounted for by postulating a positional distribution of allophones [T] in word-initial position, [D] in intervocalic position and, subsequently, their random phonological redistribution. As a result, Tlingit almost completely lost the historical distinction between aspirated and non-aspirated occlusives. Fortunately, the distinction between PND root-initial D- and T-series remains stable in Eyak-Athabaskan.

Glottalization of PND root-medial (root-final in the contemporary languages) occlusives is often lost in Tlingit and occasionally in Eyak. PA is stable in regard to occlusives, but shows numerous examples of deglottalization of root-final fricatives. Final deglottalization, although its origins remain unknown²², will not be considered below as an irregularity. In general, it is difficult to ascertain a PND distribution of laryngeal features in roots containing two obstruents, both or even one of which is glottalized in descendant languages.

A not particularly numerous row, not included in Table 5, consists of Tl labialized velars that correspond to PEA uvulars²³. I consider Eyak affricates (as well as any affricate/fricative fluctuation) in place of Tlingit and PA fricatives as irregularities. Multiple reflexes of $*_L(w)$, $*_λ(w)$ in all the subgroups demand further research. Cf. Leer’s reconstruction in Table 6 (according to Leer 2008a: 170 in my transliteration).

My interpretation of the numerous series of front and lateral affricates/fricatives seriously differs from Leer’s, except for the evident series $*_č-$ and $*_λ-$. The difference between our reconstructions is due primarily to Leer’s neglect of some series, which he combined, considering the multiple Tlingit reflexes to be a result of internal splitting. Maybe it is a consequence of the number of the roots involved: Leer’s reconstruction is based on a relatively short list of cognate sets (ca. 300), whereas I have tried to take into account the entire comparative corpus (ca. 800).

Leer does not distinguish PND glottalized affricates from fricatives, although they are strongly opposed in most Tlingit dialects, necessarily reflecting a “Pre-Tlingit” situation, since glottalized affricates and fricatives have different reflexes if they lose their glottalization in Athabaskan.

The series $*_č^w-$ can hardly be interpreted as $*_K^w-$, contra Leer’s opinion²⁴. The series $*_K^w-$ is reconstructed by me as an independent one; it yields labialized velar reflexes both in Tlingit and Eyak-Athabaskan. The series $*_C^w-$ and $*_č^w-$ are reconstructed by me on the grounds that Tlingit here has labiovelar reflexes (K^w)²⁵, which can be explained as a result of a backward

²⁰ Here and further below the term “occlusives” always denotes “occlusives and affricates”, and the term “obstruents” embraces occlusives, affricates and fricatives.

²¹ At the very least, modern languages show no traces of PND word-medial aspirates.

²² A fortuitous choice of one of two alternate allomorphs may have been the cause.

²³ E.g., Tl *guʔs* ‘cloudy sky’ vs. E *q’ahs* ~ *q^wahs*, PA **q^was* ‘cloud’. This root also shows metathesis of glottalization.

²⁴ Leer’s interpretation goes back to Krauss’s reconstruction of PEA $*_č^w$ -series as $*_K^w$ (“rounded palatals”). Krauss consistently distinguishes palatals and velars, regardless of whether they are rounded or unrounded. Leer takes a further step, transforming the PEA palatal series into a velar one in PND, but disregarding the evident rounded velars with their stable back reflexes, including Athabaskan.

²⁵ $K^w > K/K^w$, with fluctuation of labialization, also observed in velar and uvular labialized series.

shift of labialized front affricates/fricatives. Leer assumes a forward shift of labialized velars in Athabaskan ($*K^w > \check{C}^w/\check{C}$), whereas I suggest a backward shift of labialized front affricates / fricatives in Eyak and Tlingit: $*C^w$ and $*\check{C}^w > *C^w > K^w$ in Tlingit; $*c^w$ lost its labialization in PEA, but $*\check{C}^w > *[\check{C}^w] > \check{C}^w/\check{C}$ in Athabaskan; $*\check{C}^w > *C^w$ and subsequently $> K^w$ in Eyak. Labialized velar reflexes of $*C^w$ in Tlingit are as frequent as the same reflexes of $*\check{C}^w$ ($*K^w$ in Leer's terms). There are several examples of PND $*\check{C}^w/\check{C}^w$ duplicates²⁶. Development of the PA series $*\check{K}$ - into Athabaskan front affricates / fricatives is an entirely different process, since it deals with non-labialized palatal phonemes.

My presumably intermediate (“sibilant-hissing”) series $*\check{C}$ - is represented with scant examples and has no evident labialized counterpart — except for a quite hypothetical $*\check{s}^w$, i.e. Leer's “extrasystematic fricative correspondences”, which he interprets as $*sx$ and $*\check{s}x$ (Leer 2008a: 170)²⁷.

My series $*\lambda^w$, in which PEA sibilants correspond to Tlingit laterals, is interpreted as “labialized laterals” based on the principle of “what else?”, especially since the non-labialized series $*\lambda$ otherwise remains without its labialized counterpart. Leer regards Tlingit laterals, corresponding to PEA sibilants, as a result of “lateralization” of the latter due to fusion with l -prefixes, but this does not explain the presence of the series $*\lambda^w$ in root-medial position. Eyak-Athabaskan reflexes of the series $*\lambda^w$ show that it has effectively merged with $*C^w$ -, including subsequent delabialization.

As is evident from Leer's solutions, he avoids reconstructing any phonemes that would exceed the typological limits of the well-known Na-Dene systems. I, however, hold the opinion that one of the main difficulties in reconstruction may be overcome only if we refute the widespread conception that the ancestral language *must* look like its descendants on the phonemic level, even though their systems may actually form a typological unity with those attested in genetically unrelated, but contiguous languages. The complexity of sound correspondences between Tlingit and Proto-Eyak-Athabaskan aroused the suspicion that Tlingit was “a portmanteau language family” (Leer 1990), but this hypothesis seems untenable.

A characteristic feature of the Na-Dene family is absence of velar obstruents, and only external comparison with other members of the Dene-Caucasian macrofamily may elucidate their reflexes in PND. Some hope remains that external comparison will also help us solve the problem of interpreting Na-Dene correspondences between front affricates/fricatives.

Some series of front and lateral affricates and fricatives

As an example of my reconstruction, I have chosen three disputable series, namely, the series $*C^w$, $*\check{C}$ and $*\lambda^w$. Restrictions on volume have forced me to minimize adduced data, comments

²⁶ For example, $*\check{z}i\check{c}i / *z^wi\check{c}^wi$ ‘punch, knock’: PA $*\check{z}\check{a}\check{c}$ ‘hit with fist, punch’ and perhaps TL $-\check{z}eh\check{c}$ - ‘move? O’, $Y-\check{z}eh\check{c}$ - ‘become surprised’, NTL $ya-li-\check{z}e:\check{c}$ (tr.) ‘surprise, astonish, amaze’ vs. E $O-gu\check{z}k$ ‘punch, knock, strike with closed fist’; $*\check{c}a\check{s}V / *c^wa\check{s}^wV$ ‘thick’: E $-\check{c}ah\check{s}$ [adj.] ‘thick (in least dimension), coarse’ [with a secondary final fricative] vs. TL $s-kahk$ -, $-ka:k$ ‘be thick’, etc. Such controversial cases as TL $gu-k$ vs. PEA $*\check{z}e:-\check{\chi}$, $*\check{z}e:j$ - ‘ear’; TL $k'u\check{\chi}$ vs. PA $*\check{z}e:q$ ‘pitch’ and TL $ga:n$ ‘menstrual discharge’, E $g\check{a}m\check{a}-G$ ‘(soft wet) mud’ vs. PA $*\check{z}a:n$ ‘muddy water, soft mud’ continue this row of irregularities, centered around the fluctuation between $*\check{C}^w$ and $*\check{C}$.

²⁷ Some remnants of the old series $*\check{C}^w$ are possibly reflected in PA alternations $*K^w/*\check{C}^w$, corresponding to TL E $K^{(w)}$, formally reconstructed as $*K^{(w)}$ with irregular PA variants. For example: TL $x'u\check{h}k$ - ‘become steamed, steamy’; $x'uk-\check{z}a$ ‘steam’ ▪ E $x^wa:g \sim xa:g$ ‘steam, steaminess’ ▪ PA $*\acute{x}i:\check{c}^w / *x'i:k$ ‘breath, breathe; spirit, life; shadow, ghost’ < PND $*x^w\acute{a}\check{c}^wi$ (?) ‘steam, breath’; TL $=Ya-\check{z}a:k$ ‘place for, room for, time for’ ▪ E $Op-\check{z}a:g$ [postposition] ‘middle, center of O; Op's waist’ ▪ PA $*ni:\check{z}^w/*ni:\acute{g}$ ‘middle; waist; seed, pit’ < PND $*\acute{n}\acute{a}\check{z}^wi$ (?) ‘middle’.

and references; a detailed account of my reconstruction, including the entire data corpus, thorough analysis of PND sound correspondences and an accurate review of previous versions of PND reconstruction will be published elsewhere.

Semantic reconstruction of PND roots, given below, is rather approximate. A question mark after a protoform does not denote my doubt as to the very fact of comparison, but only points to some irregularity of sound correspondences (as a rule, concerning the distribution of glottalized obstruents, affricate/fricative fluctuation and reconstruction of root-medial sonorants).

Syntactical abbreviations: S = subject, O = object.

*C^w-series

PND *ʒ^w > Tl. g^(w), E ʒ, PA *ʒ

Root-initial: 1. *ʒ^wans^wV ‘plead, pray’ (#1107²⁸): Tl *χ'e-d-gaʔx'~* ‘pray’; *χ'a-ga;x'* ‘prayer’; *d-gaʔx'~* ‘become annoyed by noise’ ▪ E *d-də-ʒa:c'* ‘S pleads, implores, begs, prays’; (O?)*d-t-ʒa:c'* ‘S begs (O?)’ ▪ Leer 2008: 11; 2008a #14²⁹.

2. *ʒ^wAt'E ‘shake, move upside down’ (#1134) ▪ Tl *k-g^waʔt'~* ‘(container) moves upside down’ ▪ PA *ʒəʔ ‘shake, rattle’.

3. *ʒ^wäng^wV ‘act with end of stick-like instrument’ (#1108): Tl *-guhG~* ‘push’; *tu-guhG~* pl. ‘run’; *k-ʒ-guhG~* ‘(hair, plants) push up (out of scalp, ground)’ ▪ E *d-ʒahG ~ -ʒahG-t* ‘cane, staff, walking-stick, prop, wand’, etc.³⁰ ▪ Leer 2008: 11; 2008a #16.

4. *ʒ^wELV ‘hump, hill’ (#1110): Tl *guhL* ‘bump, hump’ ▪ PA *ʒəʔ ‘mountain (above timberline)’ ▪ Leer 2008: 12.

5. *ʒ^wiwV ‘good, pleasant’ (#319): Tl *s-gu:* ‘be joyful, fun’; *sa-gu* ‘joy, fun’ ▪ E *-ʒu:* ‘good, nice, pretty, etc.’; *tə-ʒuʔ* ‘S improves (his lot) somewhat’ ▪ Leer 2008: 11; 2008a #15. Cf. Leer 1992 #100.

6. *ʒ^wis^wV ‘cup, receptacle, calyx’ (#1113): Tl *gux'-a* ‘cup’; *-guʔx'~* ‘dip (water) out’ ▪ E *-d-t-ʒic'-t* ‘receptacle and/or calyx, sepals, “stem” (of berry)’; *O-lχ-t-ʒic'* ‘S removes calyx (etc.) from O (berry)’.

Root-medial: 1. *ʔəʒ^wə ‘move the foot, step’ (#32): Tl *Y-ʒ-ʔaʔg* ‘stagger (of wounded animal or person)’ [with *ŋə-(c-)də- ‘mis-’] ▪ E *O-ʔeʔʒ* ‘S touches, affects, acts upon O with foot’; *O-t-ʔeʔʒ* ‘S moves, fixes position, substantially affects O with foot’ ▪ PA *ʔe:ʔʒ ‘move the foot, step’, *ŋə-(c-)d-ʔe:ʔʒ ‘misstep’ ▪ Leer 2008: 18–19; 2008a #46a, b.

2. *dEmʒ^wi ~ -wʒ^w- ‘skin, fur’ (#209): Tl *=duhg* ‘skin’ ▪ PA *du:(ʔ)ʒ ‘fur’.

3. *ʔäʒ^wV ‘hard, stiff’ (#458): Tl *ka-t'u[?]g~* (ITl) ‘(snow) has thin hard crust’ ▪ E *-ʔ'ahʒ ~ -ʔ'aʔʒ* ‘S is tightly packed, firm, rigid, stiff, hard, intractable, immobile’.

4. *t'äʒ^wV ‘stiff, numb’ (#824): Tl *-t'iʔg~* ‘become stiff’ ▪ E *-t'aʔʒ* ‘S (place) becomes impassable’; *də-t'aʔʒ* ‘S (place) becomes impassable’.

5. *χaʒ^wV ‘finger-, toenails, claws’³¹ (#979): Tl *=χahg^w* ‘nails’; *=χahg-u* ‘claws’; *L-χaʔg^w~* ‘hang on with nails’ ▪ E *-y-t-χahʒ-t* ‘nails, claws’; *-qi:y-t-χahʒ-t* ‘toenails, claws’ ▪ Leer 2008: 18; 2008a #45.

²⁸ The number refers to the record number of the etymology in the database.

²⁹ References are given without any mark if Leer’s and my own comparison of Tlingit with Eyak and/or Athabaskan data coincides; they are preceded by “Cf.” in case of difference.

³⁰ Cf. also *ʔad-tə-ʒihG* ‘S moves self along in boat by poling’ with a different vowel.

6. *χ^wiʒ^wa ‘shoulder, upper arm’ (#1002): TI =χehg, =χihg ‘upper arm’ ▪ E O-*t-χeʒ* ‘S shoulders O, S picks O up onto shoulders, S carries O on shoulders, on head’ ▪ PA *χ^waʒ, *χ^wəʒ ‘shoulder; carry on shoulders’ ▪ Leer 2008: 18; 2008a #44.

PND *c^w- > TI k^(w)-/g^(w)-, E c-, PA *c-

Only root-initial: 1. *c^wa ‘change, distribute’ (#76): TI *k-t-ga*~ ‘distribute, hand out, pass out (esp. definite amount given out at party)’ ▪ E O-*t-ca*? ‘S gives O (possession of deceased) away at potlatch’; O-*ʔ-ca* ‘S buys O (with cash, not on credit)’.

2. *c^wäna ‘see, look’ (#77): TI *s-kuh*~ ‘come to know’ ▪ E O-*ʔ-l-ta-ca* ‘S stares (penetratingly) at O, looks hard at O’; O-*t-ca* ‘O is, becomes (clearly) visible, O shows’, etc. ▪ PA *ca:n ‘see, look’ ▪ Leer 2008: 12; 2008a #21.

3. *c^wEgE ‘paired with, whole’ (#79): TI =*kig*, (ITL) *kih*g ‘paired with’; =*kig-i* ‘one of a pair, half (of symmetrical obj.)’ ▪ PA *cəg ‘whole, entire, all’.

4. *c^wELE ‘hole’ (#80): TI =*kuht* ‘navel’, (NTL) *hà:t kù:t* ‘whirlpool’ ▪ PA *cəL(ə) ‘small opening, gap; anus, rectum’; *ta:-cəLə ‘open hole in ice’.

5. *c^wEśV ‘extinguish’ (#84): TI *k-kis*~ ‘become extinguished’ ▪ PA *cəs (perf. *cas) ‘extinguish (intr.)’ ▪ Leer 2008a #25.

6. *c^wEx’a ‘dry’ (#819): TI *k-t-kuhx*~ ‘dry up’; *ša-kuhx*~ ‘become thirsty’; *ša-kuhx* ‘thirst’; *k-s-kuhx*~ ‘bail’³² ▪ PA *ca:k’ [and *ʒa:k’ < *d-ca:k’] ‘dry (by evaporation), become dessicated’ ▪ Leer 2008: 8; 12–13; 2008a #3b.

7. *c^wina ‘breachclout’ (#89): TI =*ku:n* ‘hem’ ▪ E *d-ciʔ-g* ‘S is naked’; O-*t-ciʔ-g* ‘S undresses O, strips O naked’ [-g ‘negative/privative’] ▪ PA *ca:n ‘breachclout’ ▪ Leer 2008: 12; 2008a #20.

8. *c^winʔə ‘ashes’ (#90): TI *keʔ-t* ‘cedar bark ashes’; *keʔ* ‘dandruff’; *ši-keʔ*[ʔ]ʔ (tr.) ‘burn to ashes, make into ashes (for snuff)’ ▪ E *ciʔλ-g* ‘ashes; soot’; *tə-ciʔλ-g* ‘S makes ashes’ ▪ PA *ce:λ ‘ashes, embers’ ▪ Leer 1992 #20; 2008: 12; 2008a #19.

9. *c^winE ‘flee, run to safety’ (#91): TI *-keʔ-t*~ pl. ‘run away’; caus. O-*t-keʔ*~ ‘chase pl. O away’ ▪ E *ciʔ(-l)* ‘disorder, confusion’ ▪ PA *cəń- (pref.) ‘flee, run to safety, for one’s life, in fear/panic’.

10. *c^wimdV ~ -wd- ‘snore, sleep’ (#88): TI *ʔa-kihd*~ ‘snore’ ▪ E *cuʔd* [n.] ‘sleep’; *cuʔd-t* [n.] ‘sleep’; *-cuʔd* ~ *-cuhd* ‘S sleeps, goes to sleep’ ▪ Leer 1992 #174; 2008: 13; 2008a #23.

11. *c^wiʒV ‘thin, loose (snow)’ (#87): TI *k-keʔ*[ʔ]ʒ~ ‘(dry snow) becomes loosely piled up’; =*ka-keʒ-g^w* ‘light dry snow’³³ ▪ E *-ciʒ-g* [adj.] ‘narrow, thin’, *-ciʒ-g* ‘S is thin’, etc. ▪ Leer 2008: 13; 2008a #24.

PND *c^w > TI k^(w)-, k^(w)/g^(w)-, E c’, PA *c’

Root-initial: 1. *c^wAwa ‘crosspiece, crossbar’ (#69): TI =*Yaχa-k’a;wu* ‘crosspieces of canoe’ ▪ PA *c’a:ʔ ‘crosspiece, crossbar’ ▪ Leer 2008: 16; 2008a #36.

2. *c^wEčE (?) ‘nape, occiput’ (#71): TI =*ta-k-iʔč* ‘back of head’ ▪ PA *c’əc ‘nape, occiput’ [with c’ instead of *č by assimilation] ▪ Cf. Leer 2008: 17; 2008a #42.

³¹ Not to be confused with similar roots: PA *χu:(ʔ)k’ ‘fingers, toes; soft bones’; *ǵu:nʒ ‘finger’; *k^we:c’ ‘claw, finger-, toenail, thumb’; E *-y-ku:c*’ ‘thumb’; PEA *ʒ^waga ‘thumb, big toe; claw, finger-, toenail’; PND *λ^winGE ‘finger’.

³² And *-xuhk* ‘to dry’; =*xuhk* ‘dry’ with a metathesis, perhaps under the influence of *qu-k-xa*[ʔ]k ‘weather becomes dry and crisp’, *quh-xahk* ‘dry weather’, cognate with PA *xa:k’ ‘hard and dry’ < PND *xak’a ‘dry’ (#919).

³³ Cf. also TI *kex^w-k*, (ITL, EN) *-kex^w-t* ‘be light, insubstantial’; =*ka-kex^w-k* ‘light dry (snow)’; *k-ʒ-kex^w-k* ‘be porous’; *kex^w-k*, *gex^w-k* ‘pumice’, formally deduced from PND *c^wis^wV or *kix^wV.

3. *c^wəxki, c^wəxk'i (?) 'body, inside of skin' (#72): Tl =kahg (ITl) 'flesh of inside of skin' ▪ E -c'ehχk' [a misrecord of -xk'?] 'inside of pelt' ▪ PA *ci:[xX] 'body, outer skin' [with a cluster impeding precise reconstruction] ▪ Cf. Leer 1992 #173.

Root-medial: 1. *sEc^wE 'belt' (#882): Tl (NTl) *sí:k* 'belt'; Y-si?g~ 'be kept over, detained' ▪ PA *səc' 'belt'³⁴.

2. *s^wəc^wV (?) 'squeeze, mash (berries)' (#479): Tl *k-t-xahg^w~* 'grind, whip up (soapberries)'; *ka-xag^w-a* 'pestle'; *xag^w-t'i* 'soapberries' ▪ E *O-t-c'e?c'* 'S smashes, mashes, squeezes, expresses O; S makes O (berrymash)'; *c'e?c'* 'berrymash' ▪ Leer 1992 #181.

3. *tEc^wə 'thorny/stinging plant' (#825): Tl *t'u:k* 'nettles' ▪ PA *t'e:c' 'conifer'.

PND *s^w > Tl x(^w), E s, PA *s

Root-initial: 1. *s^wak^wa (?) 'shell' (#672): Tl =ša-ka-xahg-u 'skull'; =xahg-i, *xa:k* 'skeleton'; *xa:g* 'empty shell', etc. ▪ E *sahx^w ~ sahx* 'cockle, pecten' [with irregular fricative] ▪ PA *-sa:ʔk' 'shell, pot' ▪ Tl (Yak) *sa:x^w* 'cockles' is an Eyak loan ▪ Leer 2008: 8; 2008a #2.

2. *s^wAg^wa 'sand, gravel' (#918): Tl *xag^w* 'sandbar' ▪ PA *sa:x (< *sa:g) 'sand, gravel' ▪ Leer 2008: 8; Leer 2008a #3.

3. *s^wEdE 'fall (animate subj.)' (#1109): Tl *ʒ-gihd~* 'anim. moves rapidly, falls'; Adv=ʒ-gihd~ 'act thus'; *k-u-ʒ-gid-ag^w*, *k-u-s-gid-ag^w* 'show off' [< *ʒ^wEdE < *d(ə)-s^wEdE] ▪ PA *səd 'fall; spend period of time; undergo an experience' ▪ Cf. Leer 2008: 10; Leer 2008a #13a, b.

4. *s^wəc^wV 'squeeze, mash (berries)' (#479) — see under the root-medial *c^w.

5. *s^wədE 'extend, pull' (#686): Tl -*xaʔt~* 'pull, tighten'; sg. 'become fastened, extend, hand'; *k-xaʔt~* sg. 'resemble'; *xahd-a* 'veil' ▪ E -*sid* pl. 'S (of definite length) extend'; *χd-sid* pl. 'S (of definite length) extend'; *t-sid* pl. 'S extend comparatively', etc. ▪ PA *səd 'be far, distant, deep, long in duration' ▪ Leer 2008: 9. Cf. Leer 1992 #70; 2008a #7.

6. *s^wEna ~ -n-, -m- 'frown (with displeasure), punish' (#684): Tl *?a-s-xi?n~* 'frown (with bad-temper, displeasure)' ▪ PA *sa:N? 'beat one's wife'.

7. *s^wEnE ~ -n- 'mark, recognize' (#678): Tl (Y-) *xuhm~* pl. 'move faces'; (NTl) *ya-di-xu:n* 'show faces (as entering in ceremonial dance); peer, peep' ▪ PA *sən? ~ -n' 'mark, recognize'.

8. *s^wəs'E 'skin' (#687): Tl =*xaʔs'(-i)* 'fish skin'; -*xaʔs'~* 'scrape, slime (fish)' ▪ E -*sic* [ti:l-class] 'skin (of fish)' ▪ PA *səc' 'skin, hide, pelt; skin bag' ▪ Leer 2008: 8.

9. *s^wEt'i (?) 'sweep, scrape, rake' (#682): Tl -*xi:t~* 'sweep'; *t-xi:t~* 'scrape cambium'; *xit'-a* 'broom, bark sharper' ▪ E *O-še:t* 'S scrapes O (inner side of bark) for cambium' [with irregular š- perhaps under the influence of *O-t-šu:t* 'S pries O loose, moves O in arc with pole'] ▪ PA *si:ʔt' 'rake sand, dirt; pour grain-like objects' ▪ Leer 1992 #123; 2008: 9; 2008a #6, 11.

10. *s^wEwfi 'drool, snort' (#683): Tl *χ'a-xeʔ(-g)* 'drool' ▪ PA *su:(ʔ)ʔ 'snort, blow nose'.

11. *s^wEx'E 'fall, move rapidly' (#685): Tl *s-xihx~* 'compact obj. moves through space, moves rapidly, falls'; Y-s-xihx~ 'occur'; *šu-s-xihx~* 'become depleted'; *ʒ-s-xihx~* 'run' ▪ PA *sək' 'fall, move through space (sg.)' ▪ Leer 2008: 8; 2008a #4.

12. *s^wi 'dusk, evening' (#688): Tl *xeh=*, *xih=* 'dusk'; =*xehY-i* '(mountain) shadow'; cf. *xáʔ-nah* 'evening' ▪ E *se:-t* 'evening, twilight, dusk', *t-seʔ-t* 'S (impersonal) becomes evening, twilight falls' ▪ Leer 2008: 9–10; 2008a #9a, b.

13. *s^widE 'draw, make furrow' (#689): Tl -*xihd~* 'make furrow, plow'; *xi:d* 'furrow'; *k-xihd~* 'push, poke with stick'; *ʒ-xihd~* 'paint design, write'; *kù-xid-a*, -*xiʔd-a* 'brush, pen', etc. ▪ PA *səd 'dig, draw, make furrow'³⁵.

³⁴ This root is homonymous with PA *səc' 'skin, hide, pelt; skin bag', for notes on it see below under *s^w.

³⁵ It is unclear whether E *O-sehd* 'S trips, stumbles over O' belongs here.

14. *s^widV ‘ribs together with spine, comb’ (#690): Tl *xe:d-u* ‘comb’ ▪ E *-dl-sj:d ~ -sj:t* ‘ribs, ribs together with spine’

Root-medial: 1. *č^wEs^wa ‘edible root’ (#155): Tl *ku:x* ‘Kamchatka lily root [Fritillaria camtschatcensis]’ ▪ PA *č^wa:s ‘Hedysarum alpinum, Indian potato’³⁶ ▪ Leer 2008a #76.

2. *f^wEs^wE ‘stone’ (#476): Tl *ʔehx* (ITI) ‘“aluminium”, a gray rock’ ▪ PA *c^ʷəs ‘rock, stone’.

3. *näs^wE, *nä(n)ʒ^wE (?) ‘downriver, out to sea’ (#579): Tl *ʔiʔx- (ʔix-)* [dir.] ‘downriver, south’ ▪ PA *nəsə ‘ahead, forward, out on open water, over fire’ ▪ Irregular reflex in Eyak: *lahʒ* [preverb] ‘forward; forward out to sea, out at sea, south’; *Op-lahʒ ~ -lahʒ ~ -nahs-* [postposition] ‘forward, in front of, on one side of, out to sea from, south of’; *χə-lahs-d* ‘in front, out at sea, south, “outside” (over the Gulf of Alaska) < *näs^wE/ *näʒ^wE (?) and *ʔi:ʒ-iʔ-* [postpositional phrase] ‘in bow of boat’, *ʔi:ʒ-iʔ-χ* ‘forward, in front, out to sea, overseas, south, to Seattle’ < *nänʒ^wE (?)³⁷ ▪ Leer 2008:17; 2008a #43a.

PND *s^w > Tl x^(ʷ), E c^ʷ, PA *c^ʷ-, c^ʷ/s

Root-initial: 1. *s^wE[j]i (?) ‘straight, true’ (#667): Tl *x'e:-ga* ‘true’ ▪ PA *ci:= ‘straight, directly’ ▪ Leer 2008: 16; 2008a #39.

2. *s^windE ‘fall, sink, etc.’ (#898): Tl *ša-s-x'ihd~* ‘become uprooted’; *x'i:dadi, x'i:tadi* ‘uprooted stump’ ▪ PA *c^ʷəd ‘fall, sink; event occurs, time passes’ ▪ Leer 2008: 16, 37–38; Leer 1992 #79.

Root-medial: 1. *ʔis^wV ~ n- (?) ‘call, summon’ (#38): Tl *-ʔiʔx~* ‘call; invite’; *qu-ʔi:x* ‘party’ ▪ E *O-ʔ-ʔehʒ* ‘S calls, summons, invites (for any reason, or to potlatch, party)’ [with irregular loss of glottalization].

2. *dims^wE ‘cover (mouth, face, eyes), plug’ (#224): Tl *-diʔx~* ‘plug’, (NTI) = *χ'a-dí:x-i* n. ‘cork, plug’; *O-χ'e-S-0-dí:x~* ‘for S to cork up O (bottle); for S to shut, cover mouth of O’ ▪ E *-l-du:c* ‘skin of seal’s face from nose to above eyes’; *-lχ-du:c* ‘inside corner (?) of eye’ ▪ PA *du:nʔc^ʷ, *di:nʔc^ʷ ‘eyelid, eyelash, eyebrow’.

3. *λ^wis^wa ‘trash, dirt’ (#514): Tl *-λ'ihx~* ‘become dirty (bestrewn with trash)’; *λ'ihx* ‘trash’; *t-λ'ihx-i* ‘be trashy’ ▪ E *λ'ic(-g) ~ λ'ac(-g)* ‘dirt, dirtiness, dust’; *tə-λ'ic* ‘S is dirty’; *lχ-(t)λ'ic* (~ *-λ'ac*) ‘slush, wet snow’ ▪ PA *λ'a:(ʔ)c^ʷ ~ -ʒ³⁸ ▪ Leer 2008: 21; 2008a #53a.

4. *ʔis^wi ‘soak, rain, drink up’ (#470): Tl *k-t-ʔuʔx~* ‘drink up’ ▪ E *O-λ'i:c* ‘S soaks O’; *λ'i:c* ‘sth. is thoroughly soaked’ ▪ PA *λ'i:(ʔ)c^ʷ ‘rain clouds, shower moves’.

5. *q^wEs^wə ‘enclosed/sticklike object moves’ (#643): Tl *s-quʔx~* ‘sticklike obj. falls over’; *qu:x* ‘fort’; (MS) *O-quʔx~* ‘(anim.) rolls (on the ground)’ ▪ PA *q^we:[ʔ]c^ʷ, *quc^ʷ ‘enclosed object moves independently’ ▪ Leer 2008:4, 20; Leer 2008a #50.

6. *tAs^wa ‘cut’ (#817): Tl *-t'aʔx~* ‘cut in small pieces; flick with fingers or thumb’; *kù-t'a;x'-a, -t'ax'-a* ‘marble; inch’ ▪ PA *t'a:c^ʷ ‘cut (out, into shape)’.

7. *tis^wV ‘ice; freeze’ (#837): Tl *l-t'i:x~* ‘become frozen’, (NTI) *t'i:x* ‘ice’ ▪ E *t'ic* [d-class] ‘ice’; *lχ-t'ic* ‘hail’; *g_L-t'ic* ‘icicle’; *də-t'ic* ‘S freezes’, etc. ▪ Leer 2008: 19–20; 2008a #49a, b, c.

8. *χ^wEs^wi ‘round, round object’ (#1030): Tl *χ'ix* ‘eel eggs’ ▪ PA *χi:c^ʷ, *χe:s, *χəs ‘cylindrical; square; turn, roll’.

³⁶ E *ča:s* ‘Hedysarum alpinum plant; edible Hedysarum root, Indian potatoes’; Tl *ca:c* ‘plant with edible root (Hedysarum?)’ are borrowed from Athabaskan (Leer 2008a #76).

³⁷ Cf. the similar appearance of an affricate in Eyak in PND *ŋiwsⁱ, *ŋiwʒi ‘long fiber, meat slab’ > Tl *Yihs* ‘horse clam [Tresus capax]’ and PA *nú:(ʔ)s ‘long fiber of meat, fish gut with stringy end, long tail’, as opposed to E *də-ʔe:ʒ-g* ‘dried slices or slabs of fish-meat’. See also *ʒəʒV, *ʒəʒV ‘skinny, exhausted’ under *ʒ- and *χ'imʒE, *χ'imʒE ‘sour, stinking’ under *-ʒ-.

³⁸ Koyukon *q'ix=λ'ol-əʔ* ‘decayed birch wood (soft and crumbling)’; *də-lə-λ'ol* ‘(rotten birch wood) crumbles’.

9. *ʒ^wans^wV ‘plead, pray’ — see above (under *ʒ^w).
 10. *ʒ^wis^wV ‘cup, receptacle, calyx’ — see above (under *ʒ^w).

*Ć-series

PND *ǰ > Tl ʒ-/c-, ʒ, E ǰ, PA *ǰ

Root-initial: 1. *ǰAmAʔE ‘wart’ (#1097): Tl *Ya-ʒaʔ* (Yak) ‘lumpy face’ ▪ PA *ǰəməʔ ‘wart’.

2. *ǰEnE ‘muskrat’ [Ondatra zibethicus] (#1098): Tl *cin* ‘muskrat’ ▪ PA *ǰən ‘muskrat’.

3. *ǰEǰk^wi (?) ‘owl’ (#1099): Tl *ʒisk^w, cisk^w* ‘owl’ ▪ PA *ǰi:[xX] ‘owl’ [with a cluster that impedes precise reconstruction].

4. *ǰEχE ‘oddly, erroneously’ (#1100): Tl (NTl) *ya-li-ʒé:χ-a*: ‘miss the target (when shooting in basketball)’ ▪ PA *ǰ^(w)əχ ‘oddly, strangely, badly’.

5. *ǰəśV ‘thong, lace’ (#1102): Tl *k-ʒa[?]s~* ‘lace’; *ʒàs, ʒa;s* ‘thonging, lacing’ ▪ E *O-ǰiš* ‘S makes O (platform, or rack for drying fish?’).

6. *ǰəśV, *ǰəǰV (?) ‘skinny, exhausted’ (#1101): Tl *k-ʒa[?]s~* ‘be exhausted, short of breath, tired out’; *ka-li-ʒa:s* (tr.) ‘exhaust, tire out’ ▪ E *-ǰiǰ-g* ‘S is very narrow, thin, skinny’ [irregular affricate, see footnote to *č^wEs^wa above].

7. *ǰik^wV ‘crosspiece, thwart; back’³⁹ (#1103): Tl =*ʒug* ‘back (of person)’ ▪ E *ǰuʔk^l-t* ‘crosspiece (of canoe), thwart’; *O-t-ǰuʔk^l* ‘S makes or installs O (thwart, crosspiece)’.

8. *ǰindE ‘round, oval, round object’ (#1104): Tl *ceʔd* [also *ceʔt*], (NTl) *cé:t* ‘plant with edible root’ ▪ PA *ǰi:nd ‘spherical, round, oval; ball-like gut, pyloric caecum’

Root-medial: 1. *ćiǰi ‘waterfowl, sandpiper’ (#285): Tl *hi-χu-ka-ʒi;ʒi* ‘sandpiper’ ▪ PA *ći:ǰə, *ǰi:ǰə ‘waterfowl’⁴⁰.

2. *hEǰa ‘singe’ (#337): Tl *-hiǰ~* ‘singe fur’ ▪ PA *ha:ǰ ‘singe’.

3. *wAǰEnxE ‘elk, caribou’ (#840): Tl *wǰix, wacix* ‘caribou’⁴¹ ▪ PA *wəǰənʔx ‘caribou, elk, deer’⁴².

4. *yaǰa, *yaśa (?) ‘small’ (#1079): Tl =*Yaǰ-k^u*, pl. *-xⁱ* ‘small’ ▪ E *yahš* ‘doll’; *-yahš* ‘(woman’s) child, (woman’s) sister’s child’; *də-yahš* ‘S gives birth’ ▪ PA *ya:š^w(ə) ‘young, small, little; (woman’s) child’ ▪ The Tl affricate contrary to the PEA spirant is irregular.

PND *ć- > Tl c-/ʒ-, E č-, PA *č-

Only root-initial: 1. *ćA ‘beaver, hair seal’ (#61): Tl *cah* ‘hair seal’ ▪ PA *ča:? ‘beaver, seal’.

2. *ćEkE ‘stick, pole’ (#62): Tl *-cihk~* ‘roast on stick’; *ci:k* ‘roasting stick’ ▪ PA *čǰk^l ~ -yʔ ‘mainbeam, rail (of sled)’.

3. *ćEmE ‘strong, hard’ (#63): Tl *t-cihn~* ‘make strong’; *t-cihn* (neg. *-čihn*) ‘be strong’, etc. ▪ PA *čəmʔ ‘hard, packed’.

4. *ćEχ^wE ‘stake, dam up’ (#64): Tl *-cuʔχ~, -ciʔχ^w~* ‘dam up’ ▪ PA *čəq ‘stake O; put/hang O on pole (/line)’.

³⁹ Cf. Ukrainian *nonépék* ‘waist’, dial. *nonépék* ‘lower part of back, back’ as a development of the earlier meaning that still remains in *nonépéuka* ‘crosspiece, thwart’, Russian *nonépék* ‘across, crosswise’, etc.

⁴⁰ E *ʒi:ʒi*: ‘sandpiper’ is borrowed from Tlingit.

⁴¹ Here *ǰ- is treated as a root-initial consonant.

⁴² Tsimshian (only Gitksan?) *wiǰix* ‘caribou’, North Wakashan Heiltsuk (Klemtu) *w’əǰik*, Haisla *w’əǰik* ‘moose’ (without any South Wakashan cognates) are borrowings from Tlingit.

5. *ćacV ‘push (with) sticklike object end forward’ (#65): Tl *cahc~* ‘push (with) sticklike obj. end forward’; *cač-a* ‘pushing instrument: pole, rod, antlers’, etc. ▪ E -čəc ‘S uses a fork’; *O-t-čac* ‘S lifts, handles O with a fork’; *čacł* ‘fork, forked stick, table-fork’.

6. *ćinjE (?) ‘bad, evil, difficult’ (#1105): Tl *-zi*: ‘be difficult’ ▪ Eyak: *-šiy-ah* [adj.] ‘bad, evil, anger, offense, ugly, mean, unclean, spoiled, naughty, nasty, queer, sickly, disgusting’ [with an irregular fricative š- instead of č-] ▪ PA *čəñ? ‘bad; stingy’.

7. *ćíži ‘waterfowl, sandpiper’ — see under root-medial *-ž-.

PND *ć > Tl c', E č', PA *č

Root-initial: 1. *ćíc'χ^wE (?) ‘snail’ (#66): Tl *c'esχ^w* ‘beach snail’ ▪ E *č'u:č'* ‘snail, conch’ ▪ PA *čə[χX] ‘snail’ [with a cluster that impedes precise reconstruction].

2. *ćíc'g^wE (?) ‘small songbird (generic)’ (#67): Tl *c'izg^w* ‘songbird (generic)’ ▪ E *ca:di-čič'g, ca:di-čizg* ‘(any) small non-aquatic songbird’ ▪ PA *čəč'g ‘small songbird (“American sparrows”)

Root-medial: 1. *žic'ig^wa, -x^w- (?) ‘sneeze’ (#37): Tl *ž-a-l-c'ix-à, -c'isx^w-a* ‘sneeze’ [the latter form with reduplication] ▪ E *ž-žəšg* ‘S sneezes’; *žəšg* ‘sneeze’ ▪ PA *žac'g ‘sneeze’.

2. *ćíc'g^wE (?) ‘small songbird (generic)’ -- see above (under *ć-).

3. *ćíc'χ^wE (?) ‘snail’ -- see above (under *ć-).

4. *g^wimVćV, *k^w- (?) ‘a plant with edible roots’ (#314): Tl *k'unc* ‘swamp plant (death camas?), potato’ ▪ E *guč'* ‘Indian potato’ [Apios americana?]⁴³.

5. *q^wäxćE ‘tickle’ (#646): Tl *k-t-qełc~* ‘tickle’; *ka-qełc* ‘ticklishness’ ▪ E *O-t-χałχć-χ-g* ‘S tickles O’; *ž-χałχć-χ ~ -g* ‘S is ticklish’ ▪ PA *χ^wəč ‘be ticklish; tickle (tr.)’.

6. *x^wic'V ‘charred, rough wood’ (#946): Tl *y-t-xułž~* ‘char’; *xuhzi* ‘charred wood, charcoal’ ▪ E *xułč* ‘rough wood’; *k'u-xułč* ‘smth. rough’ or ‘smth. is rough’; *də-xułč* ‘S (especially wood) is, becomes rough from abrasion’ ▪ Cf. Leer 1992 #41.

7. *χEmVćə (?) ‘egg, testicle’ (#1058): Tl (NTl) *-k'unc-i* ‘testicles (of moose, caribou)’ [with irregular *k^w-* instead of *χ^w- due to contamination with *k'unc* ‘swamp plant’, see above *g^wimVćV, *k^w-] ▪ PA *χe:n?č ‘egg, testicle(s)’.

PND *ś > Tl s, E š, PA *š^w

Root-initial: 1. *śAwni ‘good, normal’ (#658): Tl *sa:n-i* ‘medium amount’ ▪ PA *š^wu:n ‘good, well, nice’.

2. *śEnχə ‘wet snow, rain coming down in long streaks’ (#661): Tl *ka-se;χ-žā* ‘rain coming down in long streaks’ ▪ PA *š^we:χ ‘wet, slushy snow’.

3. *śə[j]i, *wəśə[j]i (?) ‘name’ (#662): Tl *-sa:~* ‘to name’; *sà* ‘name’; *-sałY~* ‘name off’ ▪ E *wəšeh, -wəšeh* ‘name’ ▪ PA *žu:-š^wi: ‘name; call/mention by name’ ▪ Leer 1992 #126; 2008: 5.

4. *śEji ‘voice’ (#659): Tl *=se* ‘voice’ ▪ PA *š^wi: ‘voice’ ▪ Cf. Leer 1992 #127.

5. *śi[lw]V (?) ‘flowing water’ (#664): Tl *si:w* ‘rain’ ▪ E *ši: ~ šj:* ‘creek, stream’.

6. *śiwV ‘sleep’ (#663): Tl *t-sižs, -sužs* ‘lull to sleep’ [with reduplication] ▪ E *žə-šu:-kih* ‘asleep’ (of baby) [-kih diminutive suffix] ▪ Leer 1992 #166.

⁴³ This PND root as well as the separate PA root *Gu:n?c ‘wild celery (Heracleum lanatum), wild rhubarb (Polygonum alaskanum)’ are ancient, but probably independent borrowings from the Penutian languages, cf. Tsimshian, Gitksan *s-gusi:t* ‘potato(es)’ and Nez Percé *qém'es* (borrowed as Engl. *camas*), Sahaptin *χma(:)š*, Yakima *χmá:š* (borrowed as Engl. *quamash*) ‘camas (Camassia quamash)’.

- Root-medial:** 1. *däsV ‘foam, (foamy) juice’ (#183): Tl *de;s-a* ‘clam juice used for seasoning’ ▪ E *da:š* ‘foam, froth, white foam as found on beach, on beer’; *də-da:š* ‘S foams’.
2. *g^wiśi ‘vagina’ (#317): Tl =*gu;s* ‘vagina’ ▪ PA **gu:š* ‘vagina’.
3. *šEmsi ~ -ws- ‘beak, crooked stick’ (#702): Tl *ši[?]s* (Swanton) ‘stick game’ [also *ši[?]šd* (Swanton) ‘lucky gambling stick’ with irregular -š-] ▪ PA **šu:(?)š* ‘beak; snow playing-stick’.
4. *wəsV ‘dry, roast’ (#862): Tl *t-wahs~* ‘become scorched, roasted by fire’ ▪ E *we:š(-g)* ‘rack of sticks for drying fish (outdoors)’.
5. *χ^wimśE, *χ^wimžE (?) ‘sour, stinking’ (#1036): Tl *k-t-χ^wis~* ‘smell like pee’ ▪ E *də-q^wihž* ‘S is, becomes rancid, bitter, sour, spoiled’⁴⁴ ▪ PA **q^wu:nš ~ -ž-* ‘sour, rancid’.
6. *žəsV ‘thong, lace, tie’ — see under root-initial *ž.
7. *žəsV, *žəžV (?) ‘thin, skinny, exhausted’ — see under root-initial *ž.

PND *ś^w (?) > Tl s-/š-, E x-, PA *x-

- Only root-initial:** 1. *ś^wa ‘south, summer’ (#179): Tl *sah-nYah* ‘Southern (Craig-Klawock) Tlingits’; *sa:-naχ* ‘south wind’ ▪ Eyak *t-xa?* ‘it becomes summer’ ▪ Leer 1992 #165.
2. *ś^w[a]na ‘old (person, animal)’ (#920): Tl *d-šahn~* ‘become old, gray’; *šahn* ‘old age, old person’ ▪ PA **xa:n* ‘old (person, animal); old age’⁴⁵ ▪ Cf. Leer 1992 #128; 2008: 5.
3. *ś^wənE ‘cure shamanistically, medicine-man’ (#883): Tl *-sažn~* ‘cure shamanistically’ ▪ E *xi:l* ‘shaman, medicine-man, indian-doctor’ ▪ PA **xən* ‘spiritual power or medicine, medicine song’ ▪ Leer 1992 #168; 2008: 4.

*λ^w-series

PND *L^w > Tl L-, L(/ł), E c-, ž/s, PA *c-, ž/s

- Root-initial:** 1. *L^wEL^wə ‘hang’ (#521): Tl *-le?L~* ‘hang slackly’ ▪ PA **cəs* ‘hang (on string, rope)’.
2. *L^wEx^wE ‘peel’ (#533): Tl *-Lihx^w~* ‘skin peels off’; *t-Lehx^w~* ‘skin peels off; eat by removing from skin or shell with thumb’ ▪ PA **cəx-* ‘peeling bark, skin’.
- Root-medial:** 1. *g^wiL^wi ‘frail, small’ (#312): Tl *k-d-gel* ‘be frail’ ▪ E *-gehž ~ -gihž(-g) ~ -gehs* ‘S is very poor, miserable, wretched, pitiable’; *d-tə-gihž-g* ‘S (board) is useless’; *lχ-t-guhž* ‘S (berries) are poor and small’, etc. ▪ ? PA **gi:(?)ž, *g^wəž* ‘small’ [irregular *ž instead of *ž].
2. *L^wEL^wə ‘hang’ — see under root-initial *L^w.
3. *ńaL^wE ‘old (man, animal)’ (#576): Tl =*?a:ti* ‘grandparent’ ▪ PA **ńəž* ‘old (man, animal)’.
4. *x^wäl^wa ‘fear, danger; monster’ (#928): Tl (NTl) *xè:L* ‘thunderbird’; *xè:L tá:x^wa:yi* ‘horsefly’ ▪ E *x^wa:s ~ xa:s* ‘lucky, ominous, tabu, dangerous, strange’; *O-ž-t-xa:s* ‘S abstains from, avoids tabu O’; *ləχisx^wa:s* [gerundive] ‘fear, danger’ ▪ PA **xa:ž ~ y-* ‘fear, startle’.

PND *λ^w- > Tl λ-, E c-, PA *c-

- Only root-initial:** 1. *λ^wa ‘stout, thick’ (#494): Tl *-lah~* ‘be stout’ ▪ PA **ca:ž-s* ‘stout, thick in diameter’ ▪ Leer 2008: 29–30.

⁴⁴ See footnote to *ńäs^wE.

⁴⁵ E *xa:nih* ‘very old salmon, ready to die, red and turning yellow in spots’ is borrowed from Tl *xehn* ‘spawned-out salmon with white scabs, ready to die’, a derivative of *-xežn* ‘scabby; scratch oneself (esp. of dog)’. At the same time, E *-a:-* instead of *-e:-* points to an extinct root **xa:n* ‘old’, contaminated with the phonetically similar Tlingit word.

2. * λ^w ag^wa ‘old, long ago’ (#497): Tl λag^w ‘always (in past)’; $\lambda ag-u=$ ‘old’ ▪ E *cahg(-t)* ‘legend, myth’; *O-cahg* ‘S tells legend about O’ ▪ PA *ca:ǵ⁴⁶ ‘old, of old, long ago’ ▪ Leer 2008: 35–37.
3. * λ^w Aq^a ‘ferrule’ (#520): Tl $\lambda a:g$ ‘arrowhead, spear point’ ▪ PA *caq ~ -k̄ ‘(finger) ring’.
4. * λ^w E ‘run away, to ruin’ (#522): Tl *ka-t- λe* ~ ‘run away (avoiding responsibility)’ ▪ PA *ce:- ‘away, to ruin, forebodingly’.
5. * λ^w əli ‘head’ (#523): Tl = $\lambda a-$ ‘head’ in = $\lambda a-gehYi$ ‘brain’ [=*gehYi* ‘inside (of a shell)’] ▪ E -*ci?* ‘neck, nape’; *ci?*- ~ *ci?l-* ~ *ci?l-* ‘head, neck, nape’ ▪ PA *ci:N? ‘head’ ▪ Cf. Leer 2008: 13.
6. * λ^w iV ‘pull along, slide’ (#525): Tl *k-t-ti?* ‘pull along length of sticklike object’ ▪ E -*cəλ* ~ -*ciλ* ‘S slides, slips’ ▪ Leer 2008: 40.
7. * λ^w imG^wE ‘grab, carry by hand’ (#526): Tl (*ša-*) $\lambda e:g^w$ ~, - λehG^w ~ ‘grab one at a time’ ▪ E *O-cihG* ‘S handles, grabs, carries (quantity of) O by hand, in handful’; *O-t-cihG* (?) ‘S carries (quantity of) O in hands a long distance’ ▪ ? PA *cu:ʔG, *cug ‘stuff inside; crush’.
8. * λ^w ix^wə, * λ^w ix^wə (?) ‘shove, pull’ (#527): Tl - $\lambda u?$ x~ ‘move by pulling self with arms or rolling’; *Ya- $\lambda u?$ x*~ (ITl) ‘carry obj. clutched to self’ ▪ E *O-cu:x* ‘S thrusts O, shoves O (into sth.)’; *O-d-cu:x* ‘S trusts O, shoves O (into sth.)’; *O-dl-cu:x* ‘S bastes (in sewing)’; *O-t-cu:x* ‘S moves O by thrusting with e. g. stick’ [with irregular -x instead of -k] ▪ PA *ce:ǵ, *cəǵ ‘stick, poke, shove, thrust, point sticklike or elongated object, body part’.

PND * λ^w > Tl λ , E c’, PA *c’

- Root-initial:** 1. * λ^w ak’a ‘wet’ (#528): Tl *d- $\lambda a?$ k*~ ‘become wet’⁴⁷ ▪ E *t-c’ak* ‘(liquid) drips, leaks (by drops)’; *?u:nahd qe?**tca?**k* ‘April’ (“when water drips down from the trees”) ▪ PA *c’ak ‘leak, be soaking wet’⁴⁸ ▪ Leer 1992 #64; 2008: 38.
2. * λ^w ämG^E ‘rot, disintegrate’ (#516): Tl *d- $\lambda u?$ g*~ ‘become rotten and smelly, develop sore’; *λu hG* ‘sore’ ▪ E *c’qhg* ‘be extremely soft and flimsy, is disintegrating’, etc. ▪ PA *c’əG ~ - χ ‘rot, disintegrate’ ▪ Leer 1992 #61.
3. * λ^w EĪE, * λ^w EĪE (?) ‘milt; fat around large intestine’ (#517): Tl = $\lambda e?$ t-i ‘milt; fish guts’ ▪ PA *cəλ ‘fat around large intestine of moose’.
4. * λ^w im λ^w E ‘suck’⁴⁹ (#518): Tl *λe λ*, *λe [?]λ* ‘suckerfish [Catostomus catostomus]’ ▪ E *O-c’u?**c’-g* ‘S sucks O (draws O into mouth by suction)’, etc. ▪ PA *c’u:nc’ ‘suck, kiss’.
5. * λ^w ingE ‘finger’ (#519): Tl *λihG* ‘finger’; *t- λihG* ~ ‘carry with pincers’; *$\lambda ig-a$* ‘pincers’ ▪ E -*y(-t)-c’ihG(-t)* ‘fifth (or fourth?) finger’; *O-t-c’ihG* ‘S marks O’ ▪ PA *c’əG ‘finger, toe’ ▪ Leer 2008: 36; cf. Leer 1992 #80.

- Root-medial:** 1. * λ^w ə λ^w V ‘soak, be wet’ (#46): Tl -*na?* λ ‘become affected by being in water (wrinkle, swell)’ ▪ E -*?li?**c*, *da-?li?**c* ‘S is, becomes wet’; *li?**c* ‘dampness’.
2. * λ^w E λ^w G^E (?) ‘soft’ (#436): Tl =*te* λ ’k ‘soft (solid)’ ▪ PA * λ əčǵ ‘soft’ [with irregular č’ instead of *c’, possibly under the influence of * λ^w :(?)č^w ‘soft’].
3. * λ^w im λ^w E ‘suck’ — see under root-initial * λ^w .
4. * χ E λ^w E, * χ EĪ^wE (?) ‘be afraid, startle’ (#992): Tl *?a-k-u-l- χih λ* ‘be afraid’; *?a-k-uh-t- χi :λ* ‘fear’, etc. ▪ PA * χe :s, * χ əs ‘start (when startled), be afraid’ ▪ Cf. Leer 2008: 19.

⁴⁶ Also *sa:ǵ under the influence of PA *sa:- ‘far, long, distant, ancient’, see below on PND * λ^w ə ‘far’.

⁴⁷ Cf. also *k-d- λu hG*~ ‘drip drop by drop’; *ka- λu G- λa* ‘slow drip’; *k-tu;χ*.

⁴⁸ Cf. also some similar, but quite irregular E *t-ca?* λ ‘(liquid) drips, leaks (by drops)’ and PA *šək ‘drip’.

⁴⁹ Perhaps an expressive reduplicated form of * λ^w əwE ‘breast, nipple, suck’, see below.

PND *ɬ^w > Tl ɬ, E s, PA *s

Root-initial: 1. *ɬ^wa ‘mouth (interior)’ (#485): Tl =*lah-ka*, =*la-* ‘inside of mouth’; =*χ’as’-lah-Yih* ‘underneath chin’; Tl =*ša-la-χ* ‘skull’ ▪ E *saʔ* [preverb] ‘at the mouth’; *-saʔ* ‘mouth’; *O-saʔ* [post-position] ‘into, to O’s mouth’; *-saʔ-d* ‘mouth’; *-sa-q’-d* ‘palate, velum, roof of mouth’ ▪ PA **sa:ʔ(-d)* ‘mouth (interior); tongue; language’; **se:q* ‘throat, gullet’ [< **sa:ʔ(-yəq)*] ▪ Leer 2008: 34.

2. *ɬ^waq’asgV ‘a kind of edible seaweed, dulse’ [*Palmaria palmata*] (#486): Tl *lahq’asg* ‘black seaweed, dulse’ ▪ E *sa:q’sg* ‘dulse’⁵⁰.

3. *ɬ^wAña ‘tan(ned) skin’ (#483): Tl *gaχ=tahn* (ITl) ‘rabbit blanket’ ▪ PA **sa:n* ‘tan (skin), to soft (wet hide)’.

4. *ɬ^wAnx^{wi} ‘numb, nauseated, hungry’ (#482): Tl (NTl) *lâ:x^w* ‘famine’ ▪ PA **si:nx* ‘numb, tingle, nauseated’ ▪ Leer 2008: 9; cf. Leer 2008a #8.

5. *ɬ^wä ‘far’: Tl *-leh~* ‘be far’; *na-ti-ye/ -Ye* ‘distant place’ ▪ E *sah-d-χ* [adverb] ‘for a long time’ ▪ PA **sa:-* ‘far, long, distant, ancient’.

6. *ɬ^wEdE ‘glide, slide’ (#487): Tl *lih-d~* ‘throw pl. objects so as to scatter’; *-lih-d~* ‘slide’ ▪ PA **səd* ‘steer (a boat), soar’ ▪ Leer 2008: 35, 39.

7. *ɬ^wEmLE ‘hollow, hollow object, (plant with) hollow stem’ (#488): Tl *tu:t* ‘fireweed’ ▪ PA **su:ʔnL* ‘hollow, hollow object’ ▪ Leer 1996.

8. *ɬ^wənχV, *ɬ^wənχV (?) ‘sapwood, pitch’ (#491): Tl *laχ* ‘sapwood; sappy inner bark (of a tree)’ ▪ E *sihχ* ‘resin (hardened sap inside grain of wood of any tree)’ ▪ Loss of final glottalization in Eyak is irregular.

Root-medial: 1. **daʔ*^wa ‘heavy’ (#182): Tl *-daht~*, *-daʔ* ‘be heavy’⁵¹ ▪ E *t-da:s* ‘S is heavy’; *t-da:s-g* [-g ‘negative’] ‘S is light’ ▪ PA **da:ʔs/ *da:s* ‘heavy’.

2. **giʔ*^wV, **giʔ*^wV (?) ‘roast on stick’ (#265): Tl (EN) *-giht* ‘roast skewered on stick’; *L-giht* ‘become shriveled from roasting’ ▪ E *O-gis* ‘S roasts, warms, burns O on stick by fire’ ▪ The Eyak word-initial velar, disagreeing with the Tlingit uvular, is an inexplicable irregularity ▪ Cf. Leer 1992 #152.

3. **k^wiʔ*^{wi} ‘fabric-like object (blanket, apron, breechclout, sack)’ (#419): Tl *g^weʔt* ‘sack’ ▪ E *kuhs-t* ‘breechclout, loincloth; apron (worn at home or e. g. at work in cannery)’; [χd-class] ‘fringes’; *-ku:l-kuhs-t* ‘brisket’ (and perhaps ‘plate, flank’) ▪ PA **ku:ʔs* ‘fabric-like, flat, flexible obj.’ ▪ Leer 1992 #46.

4. **χiʔ*^wE ‘pus, semen, viscous matter, foam’ (#1011): Tl *š-χiht~* ‘to foam’; *χiht* ‘foam’ [also irregular *t-χihš~* ‘to foam’] ▪ E *χəs* [gl-cl.] ‘pus’; *də-χeʔs* ‘S is, becomes infected, pussy’ ▪ PA **χəs* ‘pus, semen, viscous matter’.

5. **yiʔ*^wV ‘a large bird’ (#1094): Tl *ye:t* ‘raven’; *šah-ya:t* ‘hawk’ ▪ E *yehs* ‘loon (*Gavia* sp.)’.

PND *ɬ^w > Tl ɬ, E c, PA *c-, s/s

Root-initial: 1. *ɬ^wEs^wE ‘stone’ — see above under root-medial *-s^w-.

2. *ɬ^wənE ‘bone’ (#475): Tl =*k’iq-l’an*, =*ciq-l’an*, =*k’iχ-l’an* ‘palate’; *waq-l’an-dah* ‘edge (?) of eye’ ▪ E *c’al*, archaic *c’al-ih* ‘bone’ (anatomical, not as part of food); *-c’al-ih* ‘bone’ (anatomical); *-t-c’al-ih* (~ *-c’al*) ‘bone (anatomical), shell, pit, structural hard part of animal or fruit’ ▪ PA **c’an* ‘bone; leg’ ▪ Leer 1992 #34.

⁵⁰ Tsimshian *laʔask* ~ *laʔask* ~ *laʔqask* ‘seaweed’; Giktsan *laq’asx^w* ~ *laʔasx^w* ‘edible seaweed, sea lettuce’ are Tlingit loans. The Giktsan form may have kept the final labialization of Old Tlingit, in this case PND *ɬ^waq’asg^wV.

⁵¹ Cf. also the antonym *k-daʔs-k* ‘be too light’; see Leer 2008: 30–34 on the alternation *t* ~ *s* in Tlingit word derivation.

3. *ɬ^wəwE ‘breast, nipple, suck’ (#477): Tl ɬah ‘breast’; -ɬah~ ‘suck’ ▪ E c’u: [l-class] ‘breast, teat, nipple’; [gl-class] ‘milk’; -c’u: (?) ‘breast, teat’; O-c’uh ‘S (especially infant) sucks O’ ▪ PA *c’u: ‘(woman’s) breast; milk’.

4. *ɬ^wiχV ‘strike, pound, stump’ (#472): Tl ʔa-ɬehχ~ ‘to dance’; ɬehχ ‘dance’ ▪ E O-c’əχ ‘S throws, hurls (usually sg.) O’; O-t-c’əχ, O-ʔ-(l)-t-c’əχ ‘pound’; O-d-c’əχ ‘S throws O’; O-t-c’əχ ‘S strikes, hits, pounds, breaks O (with held or thrown object)’ ▪ Leer 1992 #196.

5. *ɬ^wiχ^wE ‘move swiftly, spurt’ (#480): Tl ʔa-d-ɬu?χ’~ ‘spray milt, (herring) spawn’ ▪ E -c’əq’ ‘S moves very swiftly through air’ ▪ PA *c’əq’ ‘smoothe, plane, comb O’ ▪ Leer 1992 #179.

Root-medial: 1. *cEɬ^wə ‘dried (skin, fish)’ (#264): Tl š-gil’ ‘become dry and rigid [hide]’ ▪ E gac’ ‘dry-salmon; some kind of smoked fish; salmon dried, split up the back’; S-t-gac’ ‘S makes O (dry-salmon), S cuts O (salmon) dry-salmon style’ ▪ PA *ge:(ʔ)s, *gəs ‘king salmon [On-corhynchus tschawytscha]’ ▪ Leer 1992 #65.

2. *k^wEɬ^wE ‘take apart, shed feathers’ (#418): Tl -keɬ’~ ‘take apart; untie’ ▪ PA *kuc’ ‘molt, shed feathers’ ▪ Leer 1992 #29.

3. *wäɬ^wV ‘peel’ (#850): Tl k-wuɬ’ ‘peel’ ▪ E O-waɬc’(-g) ‘S whips O’; waɬc’-t, waɬc’-g-t ‘whip’.

4. *χ’iɬ^wV ‘slippery, boggy’ (#1035): Tl k-χ’iɬ’~, -χ’iɬ’~ ‘become slippery’; kuh-χ’iɬ’a ‘iron’; kuh-χ’iɬ’a=Yeht ‘sled’; š-χ’iɬ’~ ‘slip’; k-š-χ’iɬ’-k ‘be slippery’ ▪ E q’ec’ ‘slough’.

5. *ʒ^wAɬ^wa ‘tangled (rope, hair)’ (#1133): Tl k-g^waɬ’ ‘curl’ ▪ PA *ʒ^wa:(ʔ)c’ ‘tangled (rope, hair)’.

Literature

- Enrico 2004 — John ENRICO. Toward Proto-Na-Dene. *Anthropological Linguistics* 46, No. 3. P. 229–302.
- Krauss 1964 — Michael E. KRAUSS. Proto-Athapaskan-Eyak and the problem of Na-Dene: The phonology. *IJAL* 30. P. 118–136.
- Krauss 1965 — Michael E. KRAUSS. Proto-Athapaskan-Eyak and the problem of Na-Dene II: Morphology. *IJAL* 31. P. 18–29.
- Krauss 1968 — Michael E. KRAUSS. Noun-classification systems in Athapaskan, Eyak, Tlingit and Haida verbs. *IJAL* 34. P. 194–203.
- Krauss 1969 — Michael E. KRAUSS. On the [Classifiers] in the Athapaskan, Eyak, and Tlingit Verb. *Indiana University Publications in Anthropology and Linguistics*, Memoir 24 of *IJAL*, Supplement to 35 (4). P. 53–83.
- Krauss 1970 — Michael E. KRAUSS. *Eyak dictionary & Indexes*. Item EY961K1970b. Ms., Alaska Native Language Archive. Online: <http://www.uaf.edu/anla/collections/search/resultDetail.xml?id=EY961K1970b>
- Krauss 1979 — Michael E. KRAUSS. Athabaskan tone. Item CA961K1978. Ms., Alaska Native Language Archive. Online: <http://www.uaf.edu/anla/collections/search/resultDetail.xml?id=CA961K1978>
- Krauss & Leer 1981 — Michael E. KRAUSS, Jeff LEER. *Athabaskan, Eyak, and Tlingit sonorants*. Alaska Native Language Center Research Papers. Number 5. Fairbanks.
- Leer 1975 — Jeff LEER. Tlingit stemlist, typed with numerous annotations by the author, including comparanda in Athabaskan and Eyak, making this document valuable for Comparative Athabaskan-Eyak-Tlingit as well. Item TL962L1975m. Ms., Alaska Native Language Archive. Online: <http://www.uaf.edu/anla/collections/search/resultDetail.xml?id=TL962L1975m>
- Leer 1979 — Jeff LEER. *Proto-Athabaskan Verb Stem Variation*. Part One: *Phonology*. Fairbanks: Alaska Native Language Center.
- Leer 1990 — Jeff LEER. Tlingit: A portmanteau language family? In: Philip BALDI (ed.), *Linguistic change and reconstruction methodology*. P. 73–98. Berlin
- Leer 1992 — Jeff LEER. March 17, 1992. *Na-La-Dene cognate sets*. Item CA965L1992b. Ms., Alaska Native Language Archive. Online: <http://www.uaf.edu/anla/collections/search/resultDetail.xml?id=CA965L1992b>
- Leer 1996 — Jeffrey LEER. *Comparative Athabaskan Lexicon*. Item CA965L1996. Ms., Alaska Native Language Archive. Online: <http://www.uaf.edu/anla/collections/ca/cal/>
- Leer 2008 — Jeffrey LEER. Recent advances in AET comparison. Paper prepared for the Dene-Yeniseian Symposium. Fairbanks, Feb. 26, 2008. Item CA965L2008b. Ms., Alaska Native Language Archive. Online: <http://www.uaf.edu/anla/item.xml?id=CA965L2008b>

Leer 2008a — Jeff LEER. 2008. The palatal series in Athabascan-Eyak-Tlingit, with overview of the basic sound correspondences. In: Kari, James; Potter, Ben. *The Dene-Yeniseian Connection*. Anthropological Papers of the University of Alaska 5 (new series). P. 168–193.

С. Л. НИКОЛАЕВ. Материалы к реконструкции пра-на-дене.

Предлагаемая реконструкция пра-на-дене (ПНД) основана на сравнении трех групп языков: 1) диалекты языка тлингит, 2) язык эяк, 3) атабаскские языки (праатабаскский = ПА). Эяк и атабаскские языки близки между собою и восходят к промежуточному праэяк-атабаскскому языку. Регулярные соответствия между эяк и ПА были интерпретированы Майклом Крауссом и Джефффри Леером, включая очень сложные соответствия сонантов. Дж. Леер в своих работах предложил ПНД реконструкцию, привлекательную своей простотой, однако в некоторых отношениях заметно упрощенную. Реальная ситуация представляется нам более сложной. Возможно, это следствие того, что реконструкция Леера основана на приблизительно 300 корнях, тогда как автор использует полный сравнительный корпус, включающий ок. 800 корней. Из соображений объема автору приходится ограничиться лишь кратким обзором своей версии ПНД реконструкции. Сравнительный материал, приводимый в статье, иллюстрирует только ряд сложных звуковых соответствий между передними аффрикатами и спирантами, интерпретация которых отличается от интерпретации Дж. Лира.

Ключевые слова: языки на-дене, атабаскские языки, историческая фонетика.

On the reconstruction of Proto-Mari vocalism*

Two different theories regarding the Proto-Mari vowel system have been put forward by Erkki Itkonen and Gábor Berczki. This paper critically evaluates these theories and aims to establish a solidly argued reconstruction of Proto-Mari initial-syllable vocalism. It is argued that 11 distinct vowel phonemes must be reconstructed for Proto-Mari, as opposed to 13 reconstructed by Itkonen and 7 reconstructed by Berczki.

Keywords: Uralic languages, Mari language, language reconstruction, historical phonetics.

1. Introduction

Two quite different theories regarding the Proto-Mari vowel system and the development of first-syllable vocalism in Mari varieties have been put forward by Itkonen (1954) and Berczki (1994). The purpose of this paper is to critically evaluate these theories and to establish a solidly argued reconstruction of Proto-Mari vocalism. The older research history of Mari historical vocalism is reviewed by Itkonen (1954) and will not be discussed here.

Itkonen (1954) postulates thirteen vowels for Proto-Mari, whereas Berczki (1994) assumes only seven (Table 1).

Table 1. The Proto-Mari vowel system according to Itkonen (1954) and Berczki (1994).

	Itkonen 1954			Berczki 1994		
full:	i	ü	u	i	ü	u
	e	ö	o	e	ö	o
	ε		ǎ			a
	ä		a			
reduced:	ĩ (~ə)	ǔ	ũ			

It is interesting that two scholars have reached dramatically different conclusions regarding the Proto-Mari vowel system, as the Mari varieties are genetically very closely related to each other and show rather straightforward vowel correspondences. There are two key points of disagreement: the series of close reduced vowels postulated by Itkonen but rejected by Berczki, and the number of open or open-mid vowels (Itkonen's four vs. Berczki's one). Both questions will be reviewed below. In addition, the evidence for the reconstruction of the phoneme *ö will be examined, as in all Mari varieties /ö/ shows distributional peculiarities that suggest its secondary origin.

All Mari lexical material cited in this study derives from Moisio & Saarinen's *Tscheremisches Wörterbuch* (2008; henceforth TschWb). Forms are rendered in a broad near-phonologi-

* I am obliged to Juho Pystynen and an anonymous referee for remarks that have helped to improve this paper.

cal transcription, as a thorough analysis of Mari dialect phonology cannot be attempted in this paper. For the purpose of phonological reconstruction this seems justified, even though the possibility of errors in the simplification of the phonetic transcriptions in TschWb is a drawback. For ease of reading, reduced vowels other than schwa (*ə*, *â*) are indicated with a breve (*ĭ*, *ĭ̃*, *ũ*, *ö*) instead of sideways and rotated characters traditionally used in the Uralic Phonetic Alphabet. Mari can hardly be regarded a single language, but because the purpose of this study is not to address questions of dialect and language boundaries, the term ‘variety’ is used to refer to the traditional Mari regional lects.

2. The close reduced vowels

All the Mari varieties documented in TschWb, except for Bolshoj Kil’mez (Ki), possess reduced vowels in initial syllables that contrast phonologically with full vowels. The number of contrasting reduced vowels varies between one and three, as seen in Table 2.

Table 2. The reduced vowel phonemes in Mari varieties.

variety	reduced vowels
E (B Ka Kr M MU S)	/ə/
U	/ə ǔ ǔ̃/
V	/ǔ̃ ǔ̃ ǔ̃̃/
Nw	/ə ǔ̃ ö/
W	/ə â/

According to Itkonen (1954) the contrast between full and reduced vowels was already present in Proto-Mari. The most common reflexes of Itkonen’s Proto-Mari reduced and close full vowels in the attested varieties are shown in Table 3. As the table shows, in Ki all reduced vowels merged with full vowels, in E labial reduced vowels merged with full vowels, whereas U, V, Nw and W kept the reflexes of full and reduced vowels consistently apart (in W *ĭ and *ǔ merged, however).

Table 3. The most common reflexes of Proto-Mari close full and reduced vowels according to Itkonen’s (1954) reconstruction.

Proto-Mari	E	Ki	U	V	Nw	W
*ĭ (~ə)	ə	i	ə	ǔ̃	ə	ə
*ǔ̃	ũ	ũ	ǔ̃	ǔ̃	ǔ̃	ə
*ũ̃	u	u	ũ̃	ũ̃	ö	â
*i	i	i	i	i	i	i
*ü	ü	ü	ü	ü	ü	ü
*u	u	u	u	u	u	u

Bereczki (1994: 65–84) rejects Itkonen’s series of Proto-Mari reduced vowels, and maintains that attested reduced vowels represent secondary developments of original full vowels. His arguments for this view consist of two main points. First, according to Bereczki reduced vowels show irregular correspondences between Mari varieties, which would support their

secondary origin. Second, Berezcki claims that reduced vowels originated in loanwords from Turkic languages and then spread to native vocabulary; neighboring Chuvash and Tatar, which have heavily influenced Mari both structurally and lexically, possess a very similar contrast between full and reduced vowels. He summarizes his arguments for the secondary origin of reduced vowels as follows:

“In den benachbarten türkischen Sprachen, im Tschuwassischen und Tatarischen entwickelten sich aus den ursprünglichen Vokalen *i, *u, *ü die Vokale ə, ǔ, ǖ, hier aber im gesamten Sprachgebiet und im jedem Wort. Im Tscheremissischen erscheinen diese Vokale in erster Linie in den tschuwassischen und tatarischen Lehnwörtern, in den ursprünglichen, eigenen Wörtern kommt sie seltener vor und oft variieren dialektal.” (Berezcki 1988: 337)

Berezcki’s claim that reduced vowels are rare in inherited Mari vocabulary is false, however. Lists of Uralic etymologies for Mari words with reduced vowels are given in Appendix B, and they include a total of 85 examples of inherited words with reduced vowels. In fact, reduced vowels are more common than close full vowels (*i, *ü, *u) in inherited vocabulary. We shall return to the Uralic background of reduced and close full vowels below.

Berezcki’s account of the history of reduced vowels in Mari presupposes that they are a product of irregular change, as his postulated changes *i > *ə, *ü > *ǖ and *u > *ǖ cannot be attributed to any conditioning factors. This is evident from the occurrence of words which in Itkonen’s reconstruction form minimal and semiminimal pairs between Proto-Mari close full and reduced vowels:

***u vs. *ǖ**

- E Ki *luḁo*, Nw W *luḁə* ‘gray’ (< *luḁə) vs. E Ki V (!) *luḁo*, U *lüḁö*, Nw *löḁö*, W *läḁə* ‘duck’ (< *lüḁə)
- E V U Nw W *šur* ‘horn’ (< *šur) vs. E Ki *šur*, V U *šür*, Nw *šör*, W *šär* ‘shit’ (< *šür)
- W *tul* ‘storm’ (< *tul) vs. E Ki *tul*, V U *tül*, Nw *töl*, W *täl* ‘fire’ (< *tül)
- E Ki U *šulḁo*, Nw W *šulḁə* ‘cheap’ (< *šulḁə) vs. E *šulḁor*, Ki *šulḁur*, V *šülḁür*, U *šulḁər* (!), Nw *šölḁör*, W *šälḁär* ‘wing’ (< *šülḁär)
- E Ki V U Nw W *kuškam* ‘I grow up’ (< *kuškam) vs. E Ki *kuškeḁam*, V U *küşkeḁam*, Nw *köškeḁäm*, W *käškeḁäm* ‘I tear apart’ (< *küşkeḁam)

***ü vs. *ǖ**

- E Ki V U *šüḁö*, Nw W *šüḁə* ‘hundred’ (< *šüḁə) vs. E *šüḁər*, Ki *šüḁür*, V U Nw *šǖḁür*, W *šəḁər* ‘spindle’ (< *šǖḁər)

***i vs. *ǖ**

- E Ki V U Nw W *kiš* ‘resin’ (< *kiš) vs. E U Nw W *kəškem*, V *kǖškem*, Ki *kiškem* ‘I throw; I pour’ (< *kǖškem)

Let us first consider Berezcki’s argument that many words with first-syllable reduced vowels show irregular vowel correspondences between Mari varieties, and that the irregularities would support the idea that reduced vowels emerged through irregular sound change. In the examples above, for instance, there are two occurrences of irregular full vowels in cases where a reduced vowel is expected: V *luḁo* ‘duck’ and U *šulḁər* ‘wing’.

Berezcki (1994: 72) states that approximately 30 % of the words with reduced vowels show an irregular vowel reflex in at least one variety, which in his view supports the idea that vowel

reduction is a secondary phenomenon. This argument is fallacious, however: even if 30 % of the words show an irregular vowel in some variety, it does not follow that every reduced vowel in every variety is a product of irregular change. Berczki (1994: 69) also refers to varieties where there is synchronic variation between reduced and full vowels, and provides examples of his own fieldwork data from the Arbor variety, where variant forms such as *pūra* ~ *pura* ‘(s)he comes in’ and *īlāš* INF ~ *ilen* GER ‘live’ are attested even within a single idiolect. It is, however, difficult to see why such variation would support his conclusion that the distinction between full and reduced vowels arose through irregular sound change. It is a basic sociolinguistic fact that variation can result from many factors, including ongoing sound change and dialect coalescence. These weaknesses in Berczki’s argumentation already cast serious doubts on the validity of his account of the history of Mari reduced vowels.

It must be noted that if one intends to make statistical claims about regularity or irregularity, it is entirely uninformative to calculate the percentage of word-roots that show an irregular form in any arbitrary variety, because such a figure reveals nothing of the actual frequency of regular and irregular forms. Instead, the reflexes of assumed Proto-Mari reduced vowels must be separately studied in each Mari variety. In order to perform such a study, the word-roots for which a Proto-Mari reduced vowel (*ī, *ū, *ü) could be postulated in Itkonen’s system of reconstruction were compiled from TschWb, and the vowels attested for each root in each variety were put in tables. Recent borrowings (from Chuvash, Tatar, Permic and Russian) were excluded, as they include words that have spread between already differentiated Mari varieties and may thus exhibit irregularities. The data is given in Appendix A.

Let us first consider reduced rounded vowels. As only V, U, Nw and W contrast the reflexes of reduced rounded vowels with full ones, data from other varieties is irrelevant here. Figures indicating the absolute number and relative frequency of irregular reflexes of PMari *ū and *ü in the four varieties are given in Tables 4 and 5. As we are primarily interested in the phenomenon of vowel reduction here, numbers and percentages of cases of failed vowel reduction (i.e., a full vowel occurring where a reduced one is expected) have also been separately calculated. A part of the irregularities in the material are not directly related to vowel reduction: there are also instances of unexpected vowel quality (e.g., W *kāpš* ‘nap, fuzz’ instead of expected **kāpš* < PMari **kūpš*, cf. E *kūpš*, V Nw *kūpš*) and complete vowel loss (e.g., the reflexive pronoun *ške* ~ *aške*, Ki *iške* < **iške*). The rare instances of vowel loss have been counted as occurrences of reduced vowels, as it appears evident that they have involved a reduced vowel as an intermediate stage.¹ Instances where both a regular and an irregular form is attested (e.g., Nw *kūžū* ~ *küzü* ‘knife’ < PMari **kūčə*, cf. E *küzö*, W *kəzə*) have been counted as half an occurrence of both.

As Tables 4 and 5 show, the presence of a reduced vowel as the reflex of Itkonen’s PMari *ū and *ü is in most cases quite consistent, with reduction regularity rates in the range of 87–95%. A curious exception is formed by the reflexes of *ü in the Upsha variety, where there is remarkable inconsistency between *ü* (26,5 occurrences) and *ū* (45 occurrences). This variation

¹ As pointed out by the anonymous referee, some words that would superficially seem to exhibit vowel loss rather seem to involve loanwords in which a prothetic vowel was added for phonotactic reasons in some varieties of Mari: cf. E *ūžara*, V *ūžara*, U *žara*, Nw *žärä*, W *žerä* ‘red sky, dawn, dusk’ (< Russian *заря* ‘dawn, dusk’), E *žawa*, *užawa*, V *žawa*, U *ūžawa*, Nw *žawa*, *āžawa*, W *žawa* ‘toad’ (< Russian *жаба* ‘toad’). To this group one can also count E *uškal*, V *uškal*, U *skal*, *uškal*, Nw W *āškal*, *skal* ‘cow’. This word has been considered cognate with Mordvin *skal* and Udmurt *iskal*, *āskal*, *skal*, *sikal* ‘cow’, supposedly reflecting a proto-form **uskal*V (UEW: 805; HPUL: 552). In light of the irregular sound correspondences this is not justified; the forms rather look like recent nativizations of a non-Uralic word of the shape **skal*, but the source of the word remains unknown.

Table 4. The regularity of reflexes of PMari *ũ in V, U, Nw and W Mari.

Proto-Mari *ũ	U	V	Nw	W	all varieties
total reflexes	135	98	126	129	488
irregular reflexes	16	14,5	21,5	13	65
irregular full vowels	12	9,5	6	6	33,5
regularity rate	88%	85%	83%	90%	87%
reduction regularity rate	91%	90%	95%	95%	93%

Table 5. The regularity of reflexes of PMari *ű in V, U, Nw and W Mari.

Proto-Mari *ű	U	V	Nw	W	all varieties
total reflexes	72	63	67	71	273
irregular reflexes	27	11	13,5	7	58,5
irregular full vowels	26,5	8	4,5	4	43
regularity rate	63%	83%	80%	92%	79%
reduction regularity rate	63%	87%	93%	94%	84%

Table 6. The regularity of reflexes of PMari *ĩ in Mari varieties.

Proto-Mari *ĩ	B	Kr	Ka	S	M	MU	U	V	Nw	W	all
total reflexes	110	75	68	109	105	82	113	89	105	108	964
irregular reflexes	19	16,5	16	20,5	12,5	9	17,5	9,5	16	7	143,5
irregular full	13	13,5	14	19,5	11,5	6	5	2	6	5	95,5
irregular reduced	4,5	0,5	1	—	—	2	—	—	—	—	8
regularity rate	83%	78%	76%	81%	88%	89%	85%	89%	85%	94%	85%
reduction regularity	84%	81%	78%	82%	89%	90%	96%	98%	94%	95%	89%

obviously has some special explanation, but as the Upsha material in TschWb was originally recorded by Yrjö Wichmann from a single informant, it is difficult to say anything definite about the issue without additional sources of data.

Next, let us consider the case of Itkonen's PMari *ĩ, which can be reflected as a reduced vowel (*ɘ*; in V *ĭ*) in all varieties except for Ki which lacks reduced vowels altogether. However, as discussed by Itkonen (1954: 225–238), there is also a notable number of cases which show *i* as the reflex of *ĩ in several East Mari varieties. Judging from the material in Appendix A, the regular reflex of Itkonen's *ĩ would seem to be B Kr Ka *i* when followed by PMari *ć and B Kr Ka M MU S *i* when followed by the cluster *ńć. There are also instances of E *i* in other environments: in initial position, after *j-, before palatalized consonants and adjacent to *š, for example. These do not seem to be explicable as results of regular development, even though there obviously are tendencies based on the consonant environment. Taking *ć and *ńć as regular conditioning factors into account, the regularity rates of the reflexes of *ĩ in Mari varieties turn out to be as shown in Table 6.

As the figures in Tables 4–6 are put together, we have 1725 attested reflexes of Itkonen's PMari reduced vowels *ĩ, *ű and *ũ in Mari varieties that contrast these reflexes with full vowels. Of these reflexes 1545 (90%) are regular in terms of vowel reduction, i.e. a reduced vowel

occurs where expected, and in East Mari a full vowel as the reflex of *ĩ occurs in those environments where expected.

These figures reveal the inadequacy of Berezcki's account of the history of Mari reduced vowels: in his model all occurrences of reduced vowels have to be interpreted as having undergone an irregular change, whereas only the exceptional occurrences of full vowels would represent regular development. In Berezcki's theory, of course, also those word-roots that have close full vowels according to Itkonen's reconstruction could be added to the instances of regular development, but because such roots are fewer in number than roots displaying reduced vowels, the presence of vowel reduction would still remain more common than its absence. While Itkonen's reconstruction successfully explains 90% of the material, the explanatory power of Berezcki's model is virtually nil: by postulating rampant irregular sound change, it provides no actual explanation to either the presence or the absence of a reduced vowel in any single word-root.

It bears mentioning that there is yet another fatal problem in Berezcki's argument: as already shown by Itkonen (1954), Mari reduced and close full vowels have different Uralic origins. This becomes evident by examining the etymologies for Mari words which have the PMari vowel *i, *ĩ, *ü, *ũ, *u or *ũ according to Itkonen's reconstruction. In this connection we can ignore Mari monosyllabic vocalic roots of Uralic origin, as due to a phonotactic restriction reduced vowels do not occur in Mari stems of the type *CV-. The remainder of the etymological material (see Appendix B) allows the following two generalizations to be made.

- 1) PMari *ĩ and *ũ are reflexes of PU close and mid front vowels (*i, *ü and *e), whereas PMari *i and *ü are reflexes of PU *ä and *j. There appears to be one minor conditioned exception: PU *jä- and *nä- can be reflected as PMari *jĩ-.
- 2) PMari *ũ is the reflex of PU *u, whereas PMari *u occurs as the reflex of PU *o. There is a conditioned exception: PMari *ũ can also reflect PU *o when adjacent to a labial consonant (*p, *m or *w).

Most of the regular correspondences between Mari reduced vowels and their Uralic predecessors that underlie these generalizations had already been discovered by Itkonen (1954), but they are completely ignored by Berezcki (1988; 1994). Obviously, no such regularities could occur if reduced vowels had developed from full vowels through irregular sound change.

Berezcki's account of the origin of Mari reduced vowels is flawed, as it explains neither the regular vowel correspondences between Mari varieties nor the regular vowel correspondences between Mari and other Uralic languages. Thus, both internal and external comparison clearly indicates that Mari initial-syllable reduced vowels go back to Proto-Mari vowel phonemes that were distinct from the sources of close full vowels *i*, *ü* and *u* in the same varieties, as maintained by Itkonen (1954). Thus, two series of close vowels must be reconstructed for Proto-Mari: *i, *ü, *u vs. *ĩ, *ũ, *ũ.

Even though Berezcki's claim that Mari reduced vowels originated through irregular sound change under Turkic influence is untenable, it cannot be denied that the reduced vowels in V, U, Nw and W are strikingly analogous to those in the neighboring Turkic languages. Thus, it appears highly likely that language contact has nevertheless played a role in the development of Mari reduced vowels. The original distinctive feature between Proto-Mari *i, *ü, *u and *ĩ, *ũ, *ũ was not necessarily full vs. reduced articulation (e.g, IPA [u] vs. [ũ]). More probably, the opposition has become phonetically restructured as such under the influence of Turkic languages with reduced vowels. It could also have originally been a distinction of length (IPA [u:] : [u]), height (IPA [u] : [ũ]), or a combination of length and height (IPA [u:] : [ũ]).

It is worth noting that there are many examples of similar restructurings of vowel systems due to language contact. For example, Salminen describes how the Forest Nenets vowel system has become restructured under East Khanty influence so that the resulting system is in essential respects identical to the model in the contact language:

“...at least three of the main characteristics of the Forest Nenets vowel system, namely the nature of the quantity opposition, the smaller number of contrasts in non-initial or unstressed syllables, as well as the presence of metaphony leading to morphological vowel alternations, bear striking similarities to what is found in Eastern Khanty, which is also known to have provided Forest Nenets with the bulk of its loanwords. A look at Honti’s (1984: 20) description of the vowel system in the Tromagan dialect of Eastern Khanty shows that when the three peripheral vowels, which could not possibly be reflected in Forest Nenets where the frontness of the vowel depends on the palatality of the preceding consonant and the syllable as a whole, are eliminated, the remaining system of cardinal vowels consists of exactly six long vowels and four short vowels, and even the phonetic match is very close, especially if recent sound changes in Eastern Khanty dialects are taken into account. A claim can therefore be put forward that the Forest Nenets vowel system was reorganized following a specifically Eastern Khanty model.” (Salminen 1997: 368)

Thus, Berezki (1994) is still probably right in attributing the **nature** of the opposition of Mari full and reduced vowels to influence of Turkic languages. Contrary to his claim, however, this influence cannot account for the **origin** of this opposition, which must go back to Proto-Mari as maintained by Itkonen (1954). The original phonetic nature of the opposition in Proto-Mari remains unclear, and the present way of marking the Proto-Mari predecessors of Mari reduced vowels with a breve (**ĩ*, **ũ*, **ü*) is merely a practical solution for indicating the opposition in reconstructed forms.

3. Proto-Mari **ö*

Both Itkonen (1954) and Berezki (1994) postulate a vowel phoneme **ö* for Proto-Mari. Nevertheless, this vowel shows distributional peculiarities which require its status in Proto-Mari to be reassessed. Disregarding Tatar and Chuvash loanwords which may have spread between already distinct Mari varieties, TschWb contains 22 roots which uniformly have the front mid rounded vowel *ö* in all Mari varieties; for the sake of brevity, only E and W forms are cited here:

- E *öram*, W *öräm* ‘I am amazed’
- E, W *ördem* ‘I grow fat’
- E, W *örδəž* ‘side (body part)’
- E *öraš*, *örš*, W *öraš* ‘bullfinch’
- E *öraš*, *örš*, W *öraš* ‘mustache’
- E, W *jörem* ‘I put out (a fire)’
- E, W *jöraktem* ‘I knock over’
- E *köryö*, W *korya* ‘inside’
- E *löča*, (B Ka) *löča*, W *löčä* ‘it swells (due to moisture)’
- E, W *mör* ‘strawberry’
- E *nörtmö*, *mörtñö*, W *mörtñi* ‘roe’
- E *nöryö*, W *norya* ‘cartilage’
- E *nörö*, W *nöra* ‘pliable’
- E *pöčəž*, (B) *pöčöz*, W *pöča* ‘lingonberry’
- E *pördam*, W *pörtäm* ‘I go round’

- E *pörš*, W *pöršänjšə* ‘rime’
- E *pöraž*, *pörž*, W *pöraž*, *pörža* ‘brother-in-law’
- E, W *šör* ‘side’
- E, W *šörem* ‘I guess, I solve’
- E *šöryaltem*, W *šöryältem* ‘I tie loops’
- E *šörtnö*, W *šörtni* ‘gold’
- E *törya*, W *töryä* ~ *təryä* ‘it gallops’

Both Itkonen (1954: 213–215) and Bereczki (1994: 116–118) acknowledge that the reconstructed phoneme **ö* shows a distributional peculiarity: it is almost completely restricted to the position before **r*. This contextual factor was already noted by Räsänen (1920: 97), who suggested that Proto-Mari had no vowel **ö*, and that modern Mari *ö* is the reflex of Proto-Mari **ü* before **r*. Itkonen, however, supports the reconstruction of a distinct Proto-Mari vowel **ö* by noting that not only Finno-Ugric **ü* but also other Finno-Ugric vowels may be reflected as *ö* in Mari, for example:

- E W *örδəž* ‘side (body part)’ ~ SaaN *erttet*, MdE *ird’es*, Komi *ord-li* ‘rib’ (< **ertä* acc. Itkonen)
- E W *mör* ‘strawberry’ ~ SaaN *muorji*, Fi *marja* ‘berry’ (< **marja* acc. Itkonen)

Itkonen’s argument is entirely unconvincing, however. One must note that also Proto-Mari **ü* — when reflected as *ü* in all varieties — can appear as the reflex of several Uralic vowel phonemes, including for instance **e* and **a* in Itkonen’s reconstruction: cf., e.g., E, W *müks* ‘bee’ (< **mekši* acc. Itkonen), E *šüδö*, W *šüδə* ‘hundred’ (< **śata* acc. Itkonen). Hence, there is no reason why Mari *ö* in words such as *örδəž* ‘side (body part)’ and *mör* ‘strawberry’ could not go back to Proto-Mari **ü* as well, regardless of the actual Uralic source of this sound in each word. Thus, Itkonen’s argument for treating **ö* as an independent Proto-Mari phoneme is based on a confusion between the Proto-Mari and the Uralic levels of reconstruction.

If we restrict our analysis to the actual Mari material, it appears impossible to find clear evidence for reconstructing a phonological opposition **ü* : **ö* into Proto-Mari. Almost all the cases that uniformly show *ö* in all varieties involve the position before *r*, and in the same environment *ü* is not found. While the sequence *ür* does occur in East Mari, in such cases it goes back to PMari **ür* and regularly corresponds to V U Nw *ür* and W *ər*, as in the following examples:

- E *küram*, V U *küräm*, Nw *küräm*, W *kəräm* ‘I tear off’ (< PMari **küräm*)
- E *šüryö*, V U *šüryö*, Nw *šüryü*, W *šəryə* ‘cheek, face’ (< PMari **šürgə*)
- E *türwö*, V U *türwö*, Nw *türwü*, W *tərwə* ‘lip’ (< PMari **türwə*)
- E *wür*, V U Nw *wür*, W *wər* ‘blood’ (< PMari **wür*)

Hence, it appears evident that there has been a regular shift **ü* > *ö* before *r* in all varieties. In East Mari the opposition between [ö] and [ü] became phonologized through the merger of reduced labial vowels with full ones: PMari **ür* > E *ör*, but PMari **ür* > E *ür*.

In addition, *ö* is found in two words before the affricate **č*: E *löčä*, W *löčä* ‘it swells (due to moisture)’ and E *pöčəž*, W *pöčə* ‘lingonberry’. These cases have no straightforward explanation. One cannot assume a regular lowering **ü* > *ö* before **č*, because the sequence **-üč-* is preserved in four cases: E *tüčäm* ~ *čüčäm*, U *tücam*, Nw *tüčäm*, W *čüčäm* ‘I close (a door)’ (< PMari **tüčä-* ~ **čüčä-*), V *čüčem*, W *čüčem* ‘I make a hole’ (< PMari **čüče-*), E *čüčä*, B Kr *čüčä*, Nw *tüčä* ‘it drips’ (< PMari **čüče-*), and E *čüčkem*, B *čüčkem*, U *cuckem* (!), Nw *cüčkem* ‘I shake my body’ (< PMari

*čüčke-). Thus, *löča* and *pöčəž* could theoretically support the reconstruction of an opposition *ü : *ö. It would, however, be an implausible solution to reconstruct a phonological opposition of two vowels that was only realized before the consonant *č, especially as the number of examples supporting the postulation of such an opposition is limited to two word-roots.

It is worth noting that even in loanwords it is difficult to find examples of widespread Mari *ö* in contexts other than before *r*. Rare exceptions are the Tatar loanwords E W *jön* ‘means, way’, E W *kök* ‘gray (horse)’, and E *töwa*, W *töwä* ‘hill’. More commonly, Tatar loanwords with *ö* show irregular vowel correspondences: e.g., B Ka *pölek*, Kr *pölek* ~ *pelek*, S *pelek*, M *pölok*, U *pelak*, Nw *peläk* ‘gift’. All this points to the conclusion that *ö* was established as an independent phoneme only after the breaking up of Proto-Mari and the independent development of Mari varieties. However, [ö] may have existed as an allophone of *ü before *r already in Proto-Mari.

There is also a handful of words in which some varieties show *ö* corresponding to *ü* in other varieties. As already suggested by Itkonen (1954: 222), these correspondences seem to reflect Proto-Mari *ü. Most noteworthy is a group of words which show the correspondence E Ki V U *ö* ~ Nw W *ü*:

- B *nölaš*, Kr V *nölaš*, Ki *nölaš*, M *nölaš* ~ *nelaş* (!), MU *nelaş* (!), V *nölaš*, U *nelaş* ~ *lelaš* (!), Nw *lüläkš*, W *lüläš* ‘spindle ring, wheel’
- E *nölpö* ~ *lölpö*, V *nölpö*, U *lölpö*, Nw *lülpü*, W *lülpa* ‘alder’
- E *nöltem* ~ *löltem*, V U *nöltem*, Nw W *lültem* ‘I raise, lift’
- E *nönčäk*, B *nönčök*, V *nönčük*, U *nöncük*, Nw *nüncük*, W *nünčäk* ‘dough’
- E *nöšmö*, Nw *nüšmü*, W *nüšmü* ~ *nüšmä* ‘seed’
- E *nöštalam*, Nw W *nüštäläm* ‘I knead dough’

In addition, one can also count M *nölä pikš* ‘arrow with a bone head’ (*pikš* ‘arrow’) in this group. The word is not attested in other varieties, but we can assume it goes back to Proto-Mari *nülə. This is suggested both by the apparent regularity of the correspondence on account of the six cases listed above, and by the fact that the form *nülə is the regular reflex of PU *ñjli ‘arrow’; other examples of the development PU *j > PMari *ü can be found in Appendix B.

It is notable that the phonological context is similar in all six words: there is a preceding coronal sonorant (*n* ~ *l*) and a following coronal continuant (*l*, *n* or *š*). Regardless of how exactly we choose to define the context for the shift *ü > E *ö* in these words, the correspondence is evidently regular: there are no counterexamples showing a retention of *ü in a comparable context. It must be noted that the consonant *d (> *δ*) apparently did not count as a coronal continuant at the time of lowering, as no lowering is attested in E *lüdam*, W *lüdäm* ‘I am afraid’, E *šüdam*, W *šüdäm* ‘I clear (forest)’, E, W *šüdem* ‘I order’, and E *šüdü*, W *šüdü* ‘hundred’. Also, the phonetic motive for the lowering *ü > *ö* in the defined context remains obscure, but this does not prevent us from recognizing the regularity of the development.

There are also two other possible cases of the same correspondence in different environments:

- E Ki *möŋgö*, U *müŋgö*, Nw *müŋgü*, W *məŋgə* (!) ‘back (= zurück)’
- B Kr Ki MU S V *kö*, M *ke* (!), U (!) Nw W *kü* ‘who’

The first word seems to have undergone the same East Mari change *ü > *ö, even though here we find *ü* in Upsha, perhaps due to dialect borrowing. However, in W there is a reduced vowel *ə* (< *ü). From an internal point of view it would seem plausible to treat this as a secondary case of irregular vowel reduction, but external data suggest otherwise: the word goes back to PU *müŋä- (UEW: 276–277; HPUL: 546), and W *ə* < *ü is the expected reflex of PU *ü.

The interrogative pronoun ‘who’ is even more obscure. Here a vowel change *ü > ö has taken place in all Eastern varieties except for M and U, but this definitely cannot be regular, as in other CV-stems *ü is preserved throughout Mari:

- B Ka Ki *küj*, Kr S M MU V U Nw W *kü* ‘stone’ (< PMari *kü)
- B Ka Ki *müj*, Kr S M MU V U Nw W *mü* ‘honey’ (< PMari *mü)
- B Ka Ki *püj*, Kr S M MU V U Nw W *pü* ‘tooth’ (< PMari *pü)
- B Ka *šüj*, Ki *šüj*, Kr S M MU U Nw W *šü* ‘pus’ (< PMari *sü)
- B Ki *šüj*, Kr S M MU V U Nw W *šü* ‘charcoal’ (< PMari *šü)
- B Ka Ki *šüj*, Kr S M MU V U Nw W *šü* ‘neck’ (< PMari *šü)

Bereczki (1992) reconstructs these words as glide stems (*küj, *müj, etc.) on the basis of the B Ka and Ki forms. If this was correct, it could account for their difference to *kö* ‘who’. It is highly probable, however, that B Ka Ki *-j* originated as a hiatus-filling sound before suffixes beginning with a vowel, and was then analogically generalized as a part of the stem. The secondary status of *-j* in these varieties is also evident from the fact that it appears in the place of any deleted PU consonant: cf. Ki *küj* ‘stone’ (< PU *kiwi), *püj* ‘tooth’ (< PU *piŋi), *šüj* ‘pus’ (< PU *säji), *šüj* ‘charcoal’ (< PU *šüdi), *šüj* ‘neck’ (< PU *šepä). The word *müj* ‘honey’ (< *mü), in turn, is a loan from Proto-Udmurt *mü ‘honey’ (> Udmurt *mu*, dial. *mü*; cognate with Komi *ma*). It cannot be a direct reflex of PU *meti ‘honey’, because no loss of intervocalic *t has taken place in Mari, as opposed to Permic (cf. Bereczki 1992: 40).² Moreover, it is not even clear that the B Ka Ki glide *j* in these words is a phonological segment; Sebeok and Ingermann (1961: 7) describe the East Mari phoneme /ü/ as having the allophones [ü] ~ [ɥ] ~ [ü:] ~ [ɥ:] word-finally and [ü] ~ [ɥ] before another vowel. The vowel /i/ is described as having the same kind of phonetic glide component in the same environments.

The vowel change *ü > ö in *kö* ‘who’ thus must be irregular, and its reason remains unclear; perhaps it is related to the high frequency of the pronoun. An even more mysterious question is why the pronoun contains a labial vowel in the first place, as it goes back to PU *ke (cf. North Saami *gii*, Finnish *ke-n*, Mordvin *ki*, Komi, Udmurt *ki-n*, Hungarian *ki*, Nganasan *sĭ-lĭ* ‘who’). The M form *ke* is noteworthy in this connection. Despite its limited distribution, it is tempting to consider M *ke* an archaism, as it is exactly the expected reflex of PU *ke ‘who’: cf. M *me* ‘we’ < PU *me (> North Saami *mii*, Finnish *me*, Mordvin *mi-ń*, Komi, Udmurt *mi*, Hungarian *mi*, Nganasan *mĭ-ŋ*) and M *te* ‘you (pl.)’ < PU *te (> North Saami *dii*, Finnish *te*, Mordvin *ti-ń*, Komi, Udmurt *ti*, Hungarian *ti*, Nganasan *tĭ-ŋ*). Thus, the forms *kö* and *kü* in other varieties would seem to be the result of some post-Proto-Mari irregular development.

In addition to the words discussed above, there are three nouns that seem to show a shift *ü > ö in a more restricted set of varieties:

- M V *šön*, MU *šön* ~ *šin* (!), S *šön* ~ *šün*, Ka Kr U Nw W *šün*, Ki *šün*, B *šün* ~ *sün* (!) ‘vein’
- V *šöm*, M *šöm* ~ *šüm*, B Ka Ki Kr S MU *šüm*, U Nw *šüm* (!), W *süm* (!) ‘scale’
- S *löda*, M *löda* ~ *lüda*, V *lödö*, B *rüdü* ~ *lüdü*, Ka *rüdü*, Ki *lüdü*, Kr MU U *lüdü*, NW W *lüda* ‘trap (e.g. for mice)’

² According to an alternative etymology, PMari *mü ‘honey’ is unrelated to Udmurt *mu*, *mü*, and instead cognate with Proto-Khanty *mäy (VVj *mäy*, Irt *mäy*, Ni Kaz *maw* ‘honey’) and Proto-Mansi *mäy (E *möäy*, N *māy* ‘honey’) (? < *mäkV; UEW: 266). This seems unlikely, however, because there is a perfect match between PMari and Proto-Udmurt *mü ‘honey’, and Mari has a large number of Udmurt loanwords in any case. Also, the vowel correspondence Proto-Khanty *ä ~ Proto-Mansi *ä is atypical of inherited Uralic vocabulary.

In these cases one can hardly provide a regular account of the forms with *ö*, as in each case the dialect distribution of these variants is unique. It is also worth noting that each of these words also displays some other irregularities: MU *šin* (irregular *i*) ‘vein’, U Nw *šüm* (irregular *ü*), W *süm* (irregular *s-*) ‘scale’, and B Ka *rüdü* ‘trap’ (irregular *r-*).³ Due to the limited distribution of the forms with *ö* they are no doubt secondary, even though the reasons behind their development remain unclear.

4. Proto-Mari *e and *ε

Itkonen (1954: 207–211) reconstructs a PMari opposition between *ε (> W *ä*, other varieties *e*) and *e (> *e* in all varieties). He points out, however, that the reconstructed phoneme *ε shows a skewed distribution: in nearly all cases it occurs before either *r or a velar (*ŋ or *k). In light of this distribution, Bereczki (1994: 92–95) rejects the opposition *ε : *e, and regards W *ä* as the result of a change *e > *ä* in these particular environments. As the common Mari words in TschWb in which W *ä* corresponds to *e* in other varieties are examined, the correspondence is indeed found to be almost exclusively restricted to three kinds of environments:

1) Before a velar:

- W *ängər* ~ E V U Nw *eŋer* ‘river’
- W *ängər* ~ E V *eŋer*, U Nw *eŋər* ‘fishing hook’
- W *ängəremša* ~ E *eŋeremše*, *eŋəremaš*, V *eyremaš*, U Nw *eŋəremša* ‘spider’
- W *jäh* ‘soul’ ~ E *jeŋ*, *jəŋ*, V *jŋ* (!), U Nw *jeŋ* ‘person’⁴
- W *kängəž* ~ E V *keŋež*, U Nw *keŋəž* ‘summer’
- W *läktäm* ~ E V U *lektam*, Nw *lektäm* ‘I go out’
- W *läŋgəš* ~ E V *leŋež*, U *leŋəž*, Nw *leŋəš* ‘wooden pail’
- W *mäkš* ~ E V U *mekš* ‘rotten spot in a tree’
- W *mähgə* ~ E V *meŋge*, U Nw *meŋgə* ‘post, stake’
- W *pähgeš* ~ E U Nw *peŋeš* ‘it smolders’
- W *šäkš* ~ E V U Nw *šekš*, Ki *šekš* ‘gall’
- W *šähgə* ‘dead branch’ ~ E V *pu-šeŋge*, Ki U Nw *pu-šeŋgə* ‘tree’
- W *šähgäm* ~ E V U *šeŋam*, Ki *šeŋam*, Nw *šeŋäm* ‘I rummage, search through’
- W *täktə* ~ E *tekte*, U *tektə* ‘beehive’

2) Before *r:

- W *ärδə* ~ E V *erde*, Ki U Nw *erδə* ‘thigh’
- W *äryem* ~ E Ki V U Nw *eryem* ‘I coil yarn’
- W *ärtñi* ~ E V *ertñe*, U *ertñä*, Nw *ärtñä* (!) ‘birch-bark container’
- W *jär* ~ E Ki V U Nw *jer* ‘lake’
- W *käryältem* ~ E Ki V U *keryältem*, Nw *keryältem* ‘I roll up’
- W *pärtñä* ~ E *pertñä*, Ki *partñi* (!), U *pertñä* ‘bud’
- W *šär* ~ E U Nw *šer*, Ki *šer* ‘vein’
- W *šäryem* ~ E *šeryem* ‘I open, disperse, scatter’

³ It has been suggested that *rüdü* ~ *lüdü* ~ *lödü* is etymologically related to Skolt Saami *râä'tt*, Finnish *rita* ‘cage trap’ and Komi *ri* ‘well sweep; trigger of a trap’ (UEW: 746; HPUL: 553); if correct, then B Ka *r-* must be an archaism. However, this lexical set is a Wanderwort at best; the vowel correspondences are irregular, and hence the proposed reconstruction *rita cannot be justified.

⁴ Nw *jäh* ‘soul’ appears to have been borrowed from W.

- *W šártńi* ~ *E šertńe, šartńe (!)*, *Ki šartńi (!)*, *V šertńe*, *U šertńa*, *Nw šerńa* ‘willow’
- *W wär* ~ *E Ki V U wer*, *Nw wär* ~ *wer (!)* ‘place’
- *W wäryǵa* ~ *E V werye*, *Ki U Nw weryǵa* ‘kidney’

3) In CV-stems:

- *W mä* ~ *E Ki V U Nw me* ‘we’
- *W tä* ~ *E Ki V U Nw te* ‘you (pl.)’

If we disregard Tatar and Chuvash loanwords, the only case in TschWb falling outside these conditions is *W kädä* ~ *E keǵe*, *U Nw keǵa* ‘dove’. In Turkic loanwords there are more examples of *W ä* in other environments, e.g. *W kǵpšǵl* ~ *E kǵpšǵl* ‘fetter’. Such words, however, have probably been borrowed after the conditioned change **e* > *W ä*. In this case also the irregular correspondence *W l* ~ *E l* points to the word having entered West and East Mari separately.

Next, we need to consider possible counterexamples where PMari **e* would have been retained in *W* in environments where *ä* is expected. In the case of CV-type stems none occur. Before velars *W e* only occurs in loanwords that have obviously diffused between already distinct Mari varieties (e.g., *W jǵǵä* ~ *E jǵǵa* ‘older brother’s wife’ < Tatar). TschWb cites no etymology for *W wek-ät*, *U wek-at* ‘at all (in negated sentences)’, *E wek* ‘constantly; at all (in negated sentences)’, but this must be a loan from Russian *век* ‘century, age, lifetime; for ages, always’ (Komi *vek* ‘always, constantly’ derives from the same source).

However, there are some examples of *W a*, *i*, and *ö* corresponding to *e* in other varieties:

- *W aǵǵǵ* ~ *E Ki Nw eǵǵǵ*, *V eǵǵǵ*, *U eǵǵ* ‘raspberry’
- *W aǵǵeš* ~ *E Ki V U Nw eǵeš* ‘it is scorched’
- *W šaǵǵal* ~ *E V šǵǵel, šǵǵal*, *U šǵǵel* ~ *šǵǵel*, *Ki šǵǵel* ‘back side’
- *W piǵǵada* ~ *E V U peǵǵade*, *Nw peǵǵada* ‘hard, tight, strong’
- *W töǵǵal* ~ *E U Nw teǵǵal*, *Ki teǵǵal* ‘bench’

The last two examples are unique correspondences, and as such clearly irregular. The three cases of *W a* could theoretically result from some kind of regular development, but this seems unlikely, as further cases of *W a* corresponding to *e* in other varieties are not found in other environments. Hence, these three words hardly serve as counterevidence to the assumption of a regular sound change **e* > *W ä* before velars.

The situation is quite different when we take the examples involving *r* into consideration. There are many cases where all varieties, including *W*, show *e* in this environment:

- *E V erye*, *Ki U Nw W erya* ‘son, boy’
- *E Ki V U keram*, *Nw W keräm* ‘I stick in’
- *E V kerye*, *Ki U Nw W kerya* ‘woodpecker’
- *E Ki kertam*, *V ketam (!)*, *U kerǵam*, *Nw W kertäm* ‘I can, am able’
- *E šu-kerte*, *-kerše*, *Ki šu-kerta*, *-kerša*, *U šu-kerǵa*, *Nw W šu-kerǵa*, *-kerša* ‘a long time ago’
- *E V merčem*, *B nerčem (!)*, *U Nw W mercem* ‘I am unwell, sickly’
- *E Ki V U Nw W ner* ‘nose’
- *E nerye*, *U W nerya* ‘badger’
- *E Nw W šer* ‘pearl’
- *E V U šeram*, *Ki šeram*, *Nw W šeräm* ‘I brush’
- *E V šerye*, *Ki U Nw W šerya* ‘expensive’
- *E werč(ǵan)*, *Ki werčǵan*, *V werčǵ*, *U Nw W werc(ǵan)* ‘instead of’

Thus, there are 12 examples of *W -er-* and 11 examples of *W -är-*. The distinction cannot be attributed to any conditioning factor, as revealed by the minimal pair *šer* ‘pearl’ : *šär* ‘vein’ and the semiminimal pairs *keryə* ‘woodpecker’ : *käryältem* ‘I roll up’ and *eryə* ‘son’ : *äryem* ‘I coil yarn’. Thus, in this environment a contrast between Itkonen’s PMari **e* and **ε* really can be established. The reconstruction of a phoneme **ε* is an unsatisfactory interpretation of this contrast, however. From a typological point of view it appears implausible that Proto-Mari possessed a phonological contrast between /*e*/ and /*ε*/, but that this contrast was only realized before the consonant /*r*/ (in addition, there is the unique word *käδə* ‘dove’ which could perhaps be reconstructed as **kεdə*). Such a claim would also make it quite impossible to understand why the specific sound correspondence that **ε* is reconstructed to account for can be explained as a result of conditioned development of **e* in another context, namely before a velar (**ŋ* or **k*).

On the other hand, Bereczki’s (1994: 92–95) solution of reconstructing **e* in place of both Itkonen’s **e* and **ε* is unacceptable, because it leaves it a mystery why **er* is sometimes reflected as *W är* and sometimes as *W er*. Importantly, *W er* and *W är* have different Uralic sources, which shows that the opposition must represent an archaism. *W är* appears in the reflexes of Uralic **ä*-stems and *W er* in the reflexes of **i*-stems:

- *W ärδə* ‘thigh’ < **ärtä* < PU **ertä* (cf. Itkonen 1954: 178; UEW: 625; HPUL: 552)⁵
- *W jär* ‘lake’ < PU **jäwrä* (UEW: 633)
- *W šär* ‘vein’ < PU **särä* (UEW: 437; HPUL: 548)
- *W šärtüi* ‘willow’ < PU **särnä* (UEW: 752)⁶

- *W keräm* ‘I stick in’ < PU **käri-* (Aikio 2002: 18)
- *W keryə* ‘woodpecker’ < PU **kärki* (UEW: 652)
- *W kertäm* ‘I can, am able’ < PU **kärti-* (UEW: 652)
- *W ner* ‘nose’ < PU **näri* (UEW: 303–304; HPUL: 552)

Hence, the opposition between *W er* and *W är* must date back to Proto-Mari, but a distinct vowel phoneme **ε* is unlikely to be the source of this opposition. Instead, we can postulate the hypothesis that *W är* reflects PMari **er*, whereas *W er* reflects PMari **ir*. This implies that there has been a change **i* > *e* before *r* in all varieties of Mari. This would of course remain an ad hoc claim, were it not that two circumstances support this idea. First, there appear to be no common Mari words with the sequence *ir*. Second, the assumed change **i* > *e* / *_r* is completely analogous with the change **ü* > *ö* / *_r*, which in section 3 above was shown to have taken place in all varieties of Mari. Hence, we can assume that Proto-Mari close full front vowels (**i* and **ü*) became lowered to mid vowels (*e* and *ö*) before **r*.

⁵ Traditionally Mari E *W örδəž* ‘side (body part)’ (< PMari **ürdäž*) has been considered the reflex of PU **ertä*, and thus cognate with Inari Saami *ertti* ‘side’, Erzya Mordvin *ird’es* ‘rib’, Udmurt *urdes* ‘side’ and Komi *ord-lj* ‘rib’ (UEW: 625; HPUL: 552). Itkonen (1954: 178) cautiously suggests that also Mari E *erδe*, *W ärδə* ‘thigh’ could belong in this connection. However, there is no regular internal relationship between the Mari words for ‘thigh’ and ‘side’, and PMari **ürdäž* is not a regular reflex of PU **ertä* because of its vowel **ü*. Hence, it seems that Mari E *erδe*, *W ärδə* ‘thigh’ is the true inherited reflex of PU **ertä*, and that **ürdäž* was borrowed from some other branch, probably from Pre-Proto-Permic **ördVs* (> Proto-Permic **ordēs*).

⁶ UEW reconstructs the proto-form as **saríe*, but the front vowel in Mari *W šärtüi*, E V *šertüe*, U *šertüa*, Nw *šerüa* cannot be a reflex of PU **a*. Instead, the form **särnä* can be reconstructed. Finnish *saarni* ‘ash’ shows secondary back vocalism like, e.g., Finnish *sappi* ‘gall’ (< PU **säppä*) and *talvi* ‘winter’ (< PU **tälwä*). The irregular back vocalic forms in some Mari varieties (B Ka *šartüe*, Ki *sartüi* ~ *šärtüi*) also seem to be secondary.

There are, however, two words which seem to serve as counterevidence to the idea of a change *i > e / _r in all varieties. In the following cases the change *i > e has occurred in B Ka Kr Ki S M V, whereas *i* is retained in U Nw W:

- B Kr V *šere*, Ki *šera*, M S *šera*, MU *šire*, U Nw W *šira* ‘unleavened’
- B Ka Ki Kr S M MU *ter*, U Nw W *tir* ‘sled’

These items belong to a larger group of words that appear to have undergone a change *i > e in all eastern varieties except for U (in MU the reflexes are inconsistent, perhaps as a result of dialect borrowing):

- B Ka Kr M MU (!) S V *lewa*, Ki *l'ewa*, U *liwa*, Nw W *livä* ‘it becomes warm’
- B Ka Ki Kr M S V *mež*, U Nw W *miž* ‘wool’
- B Ka Kr *peče*, Ki M S *pečā*, V *peče*, MU *piče* ~ *peče* (!), U Nw *pica*, W *pičā* ‘fence’
- Ki M *pemba*, U *pimba*, W *piŋgā* (!) ‘finch’
- B M S V *šem*, Ki *šim* (!), Ka (!) MU U Nw W *šim* ‘black’
- B Ka Ki Kr M S V *šen*, MU U Nw W *šin* ‘tinder’
- E *ter*, Ki *teř*, U Nw W *tir* ‘sled’
- B Ka Kr M S V *welš*, Ki *wel'eš*, MU U Nw W *wileš* ‘it falls’
- B Ki Kr S M V *wem*, MU *wime*, U *wimā*, Nw *mimā*, W *wim* ‘marrow’
- B Ka V *wenē*, Ki Kr M S *wenā*, MU U Nw *wiŋā*, Nw *wiŋgā* ‘son-in-law’

This group of words has been discussed by Itkonen (1954: 219–221), who presents cogent arguments for interpreting *i* as the original vowel, and speaks of a “tendency” (“Neigung”) of *i to change to *e* before the consonants *l*, *r* and *m* in eastern varieties. It is worth noting that in all cases except for *peče*, *peče* ‘fence’ and *mež* ‘wool’ the following consonant is a sonorant. In contrast, however, there are only three common Mari words that are not known to be recent borrowings and that uniformly show *i* before a sonorant:

- E W U *ime*, Ki Nw *imā*, W *im* ‘needle’
- E *im'ne*, Ki *im'na*, V *im'ni* (!), U Nw W *im'ni* ‘horse’
- E V U *kinde*, Ki Nw W *kindā* ‘bread; grain’

Common Mari *i* occurs, however, in many words before Proto-Mari obstruents:

- E V *iye*, U Nw W *iya* ‘young (of animals)’ (< PMari *igə)
- E Ki V U Nw W *ik* ‘one’ (< PMari *ik)
- E *iksa*, Ki *iks*, V *iksā*, U *ijsa* (!), Nw W *iksā* ‘small bay’ (< PMari *iksā)
- E *ize*, *izi*, V *izi*, U Nw W *izi* ‘small’ (< PMari *icə)
- E V U *iške*, Ki *iskā*, Nw W *iškā* ‘wedge’ (< PMari *iskə)
- E V U Nw W *kiš* ‘resin’ (< PMari *kiš)
- E *kit*, (M MU) *kət* (!), Ki U Nw W *kit*, V *kīt* (!) ‘hand’ (< PMari *kit)
- Ki S M MU U *lišan*, Nw W *lišän* ‘near’, B Ka Kr *lašān* (!) (< PMari *lišan)
- E V U *pidam*, Nw W *pidām* ‘I bind’ (< PMari *pidam)⁷
- E V U Nw W *pikš*, Ki *pikš* ‘arrow’ (< PMari *piks)

⁷ This verb seems to be an Indo-European loan: PMari *pida- < Pre-PMari *pänti- < Indo-European *b^hend^h- ‘bind’. The suggested comparison to Hungarian *fűz* ‘laces, strings, threads’ (UEW: 386) is not phonologically regular.

- E U Nw *piktem*, W *püktem* (secondary *pi- > pü-), V *pĭktem* (!) ‘I strangle’ (< PMari *piktem)
- E *pište*, *piste*, V *piste*, U Nw W *pištə* ‘linden’ (< PMari *pistə)
- E Ki U Nw W *piž*, V *pež* (!) ‘woolen mitten’ (< PMari *piž)
- E V *šište*, Ki U Nw W *šištə* ‘woodpecker’ (< PMari *šištə)
- V U Nw W *widem*, E Ki *wüdem* (secondary *wi- > wü-) ‘I take (somewhere), lead’ (< PMari *widem)
- E U *wiškađe*, Ki *wiškiđe*, Nw W *wiškaðə* ‘fluid, washy’ (< PMari *wiškəðə)

As there are six examples of lowering of *i before sonorants in eastern varieties and only three cases showing *i retained in the same environment, the former treatment appears to be the regular one. The lack of lowering in *ime* ‘needle’ and *imúe* ‘horse’ might perhaps be explained by the fact that the vowel appears in initial position. At least *ime* ‘needle’ is very probably an old word and goes back to PU *äjmä (UEW: 22; HPUL: 536), even though one cannot completely disregard the possibility that it was borrowed from Permic (cf. Komi *jem*, Jazva Komi *i-m* ‘needle’ < Proto-Permic *im < PU *äjmä). Also, why the vowel was lowered in *peče*, *peće* ‘fence’ and *mež* ‘wool’ lacks an explanation, but it is worth noting that *mež* ‘wool’ seems to be a Permic loanword and as such may have spread between Mari varieties: its loan original is Komi *mež* ‘ram’ (< Proto-Permic *miž < Aryan, cf. Sanskrit *meṣá-* ‘sheep, ram’, Avestan *maēša-* ‘sheep’). Berczki (1992: 92) considers the Mari and Komi words to be independent borrowings from Aryan, but this notion lacks supporting evidence. The word *peče*, *peće* ‘fence’ seems to be of Uralic origin (cf. Finnish *piha* ‘yard’, Udmurt *puč*, Komi *poč* ‘stick, bar’), but the vowel correspondences between the cognates are deviant, and the proposed reconstructions *piča (UEW: 729) and *pičča (HPUL: 553) lack justification. Due to the scarcity of examples and the multiple etymological possibilities regarding some of the words, the details of the eastern Mari change *i > e remain rather unclear.

Let us now return to common Mari *-er-*. The interpretation that *-er-* reflects PMari **-ir-* raises a question: why did the lowering *i > e / *_r* take place only in eastern varieties in the words *šere* (~ U Nw W *širə*) ‘unleavened’ and *ter* (~ U Nw W *tir*) ‘sled’, but in all varieties in other words? This question cannot be given a definite answer for the time being, but we must assume that the common Mari vowel lowering *i > e / *_r* is older than the eastern Mari lowering *i > e before sonorants. This implies that any new loanword with the sequence *-ir-* that was introduced after the first lowering would then have undergone the second lowering in eastern varieties, but not in the west. This is, indeed, confirmed by Chuvash loanwords:

- E V *er*, MU *er* ~ *ir*, V Nw W *ir* ‘morning, early’ < *ir < Chuvash *ir* ‘morning, early’
- E MU *erək*, V *erĭk*, U Nw W *irək* ‘freedom’ < *irək < Chuvash *irək* ‘freedom’
- E *ser*, *šer*, Ki *šer*, U Nw W *sir* ‘shore’ < *šir < Chuvash *šĭr* ‘precipice, ravine, slope on a shore’
- E *serem*, *šerem*, Ki *šerem*, MU U Nw W *sirem* ‘I write’ < *šire- < Chuvash *šĭr* ‘write’

Thus, we can speculate that also *šere* ‘unleavened’ and *ter* ‘sled’ could be loanwords from some as yet unidentified source. A Uralic etymology has been proposed for *ter* ‘sled’; UEW (517) regards it cognate with Estonian *tari* ‘wickerwork, basketwork’, Finnish *tarjat* ~ *tärjät* and East Khanty *tārās* ‘lathwork bottom in a type of sled’, but due to irregular vowel correspondences this fails to convince. While the idea of recent loan origin of *šere* ‘unleavened’ and *ter* ‘sled’ remains a conjecture, the hypothesis put forward here explains the twelve attested cases of W *-er-* as regular and accounts for the fact that W *-är-* and *-er-* have distinct Uralic sources: PU **-är-* developed

to PMari *-er- in *ä-stems and to PMari *-ir- in *i-stems. The cost of this explanation is that the phonological development of two words, Nw W *širə* ‘unleavened’ and *tir* ‘sled’, turns out difficult to account for.

5. Proto-Mari *å and *o

Itkonen (1954) reconstructs an opposition between PMari *å (> E Ki V U *o*, Nw W *a*) and *o (> *o* in all varieties). Bereczki (1994: 92–95) does not accept this postulated opposition, and reconstructs PMari *o for both correspondence patterns. The key reason for these different interpretations is connected with Itkonen and Bereczki’s different views of the Uralic sources of the Mari vowels. Itkonen (1954: 191–195) maintains that PMari *o can reflect either PU *a or *o, but PMari *å only occurs as a regular reflex of PU *a. However, his material also contains a number of exceptions to this postulated rule, which he explains away as results of “sporadic” developments. As noted by Bereczki (1994: 93), this position is untenable: because Itkonen’s PMari *å and *o both occur as reflexes of PU *a and *o alike, Itkonen’s account of the development of these PU vowels in Mari is inadequate. Even though Bereczki is right in pointing this out, his reasoning is faulty, too: it does not follow from this that the reconstruction of an opposition between PMari *å and *o is incorrect.

A major problem is that Bereczki (1994) postulates no conditioning factors that would account for the distribution of Nw W *a* and *o* as reflexes of his reconstructed PMari *o. Numerous semiminimal pairs, such as the following, demonstrate that no such conditioning factors can be established:

- E *koðama*, W *kaðama* ‘gudgeon’ (< *kåðama) vs. E Nw W *koðam* ‘I remain’ (< *kodam)
- E *oškəl*, W *aškəl* ‘step’ (< *åškəl) vs. E *oško*, W *oškə* ‘poplar’ (< *oškə)
- E *podəm*, Nw *padəm*, W *padəm* ACC ‘pot’ (< *pådəm) vs. E Nw *podəlam*, W *podəlam* ‘I eat with a spoon’ (< *podəlam)
- E *polðalye*, W *palðalyə* ‘crooked’ (< *påldalgə) vs. E *polðəš*, W *polðəš* ‘button’ (< *poldəš)
- E *šongo*, Nw *šonğa*, W *šonğə* ‘old’ (< *šonğə) vs. E *šonšo*, Nw *šonşə* ‘hedgehog’ (< *šonşə)

Hence, there seems to be no viable alternative to reconstructing an opposition between PMari *å and *o. It ought to be noted, however, that the distribution of PMari *å and *o seems to show clear tendencies based on the consonant environment, even though these tendencies do not amount to absolute rules. Judging from the material in TschWb, PMari *å tends to occur after initial glides (*w, *j) and before the coronal consonants *ć, *č, *t, *d, *r and *n. PMari *o, in turn, is much more common before velars (*k, *ŋ). These distributional tendencies probably reflect the phonological conditions that once caused a single Pre-Proto-Mari vowel to split into PMari *å and *o, but a further examination of the issue is beyond the scope of this paper.

6. Proto-Mari *a and *ä

Itkonen (1954: 185–188, 203–207) assumes that Proto-Mari *a is uniformly preserved as *a* in the Mari varieties, whereas the correspondence E Ki V U *a* ~ Nw W *ä* reflects a distinct Proto-Mari vowel *ä. According to Bereczki (1994: 88–92), however, both correspondences reflect an original *a, and there has been a sound change *a > ä in Nw and W. His argument for this interpretation is that Nw W *ä* can correspond to foreign *a in loanwords: e.g., E *waraš*, V U *warakš*, Nw *wäräkš*, W *wäräš* ‘hawk’ < Proto-Permic *variš (> Komi and Udmurt *variš* ‘hawk’).

The problem with this interpretation is that Bereczki postulates no conditions under which the assumed change PMari *a > Nw W *ä* would have taken place. An examination of the relevant lexical data in TschWb reveals that no conditions can be postulated, as illustrated by the following examples.

PMari *-ašk-:

- E Ki V U Nw W *paškar* ‘small block of wood, plug’ (< PMari *paškar)
- E *raskalta*, *raškalda*, Ki *raskalta*, V U Nw W *raškalta* ‘(lightning) strikes’ (< PMari *raškalta)

PMari *-äšk-:

- E Ki V U *laška*, Nw W *läškä* ‘noodle’ (< PMari *läška)
- E U *šaške*, Nw W *šäškə* ‘otter’ (< PMari *šäškə)

The initial consonants do not serve as conditioning factors, as there are also examples of PMari *pä-, *la- and *ša-:

- E Ki V *parča*, U *parca*, Nw W *pärcä* ‘ear of corn’ (< PMari *pärcä)⁸
- E U *laštartem*, Ki *laštirtem* ‘I crush into pieces’, W *laštârtem* ‘I splinter (wood)’ (< PMari *laštartem)
- E *šaraŋge*, Ki *šaraŋgi*, U *šoraŋge* (!), Nw W *šaraŋgə* ‘willow’ (< PMari *šaraŋgə)

Hence, an application of the comparative method leaves no alternative to reconstructing two Proto-Mari phonemes: *a and *ä. How the vowel substitution patterns observed in loanwords that Bereczki brings up are to be explained is a complex question which cannot be addressed in the present study. It must be noted, however, that it is not justified to mechanically interpret sound correspondences in loanwords so that if the donor form contains an *a, then this sound also must have been adopted as an *a in the recipient language. Sound substitutions are influenced by a multitude of factors, including phonological and allophonic differences between the contacting languages and rival sound substitution strategies based on conflicting phonetic, etymological and systemic motives (Aikio 2007).

7. Conclusion

The main results of this study can be summarized as follows:

- As maintained by Ikonen (1954), the initial-syllable reduced vowels in Mari varieties clearly go back to Proto-Mari vowel phonemes that were distinct from the source of Northwest and West Mari close full vowels *i*, *ü* and *u*. Thus, two series of close vowels must be postulated for Proto-Mari: *i, *ü, *u vs. *ĩ, *ũ, *ũ. The alternative view of Bereczki (1994) that reduced vowels developed from full vowels through irregular sound change motivated by extensive Turkic influence is untenable, because this assumption does not account for the regular vowel correspondences between the Mari varieties, and because reduced and full close vowels can be shown to have distinct Uralic sources.
- The original distinctive factor between Proto-Mari *i, *ü, *u and *ĩ, *ũ, *ũ was not necessarily full vs. reduced articulation. It is likely that the opposition became phonetically

⁸ PMari *pärcä is apparently a loan from some Iranian source, cf. Young Avestan *parša-* ‘ear of corn’, Sanskrit *parśā-* ‘sheaf, bundle’.

- restructured as such under the influence of Turkic languages. The original distinctive factor could also have been length or height, or a combination of these features.
- c) No phoneme **ö* can be reconstructed to Proto-Mari. The vowel **ö* reconstructed by both Itkonen (1954) and Bereczki (1994) can be interpreted as an allophone of Proto-Mari **ü*; there has been a regular change **ü* > *ö* before **r* in all Mari varieties.
 - d) As maintained by Itkonen (1954; *contra* Bereczki 1994), a phonological opposition between PMari **â* (> E *o*, Nw W *a*) and **o* (> E Nw W *o*) must be reconstructed; the occurrence of both Nw W *a* and *o* as correspondents of E *o* cannot be attributed to any conditioning factors. However, contrary to what is claimed by Itkonen (1954), the PMari opposition **â* : **o* does not appear to be a continuation of the Proto-Uralic opposition **a* : **o*.
 - e) As maintained by Bereczki (1994; *contra* Itkonen 1954), no Proto-Mari vowel phoneme **ε* needs to be reconstructed to account for the correspondence E Nw *e* ~ W *ä*. The open vowel in W is a result of sound change **e* > *ä* that took place 1) before velars, 2) before *r*, and 3) in CV-type stems.
 - f) In cases where W *e* occurs before *r*, Proto-Mari **i* must be reconstructed (*contra* Bereczki 1994). There appears to have been a change **i* > *e* before **r* in all varieties of Mari.
 - g) As maintained by Itkonen (1954; *contra* Bereczki 1994), a Proto-Mari vowel **ä* must be reconstructed to account for the correspondence E *a* ~ Nw W *ä*.

On the basis of these results we can postulate the vowel inventory shown in Table 7 for Proto-Mari initial syllables. The regular reflexes of the Proto-Mari initial syllable vowels in the Mari varieties are shown in Table 8.

On a more general level, the results of the present study provide an example of how the development of vowel systems in the Uralic languages can — and should — be explained as an effect of regular sound change. In the field of Uralic comparative linguistics there has been an unfortunate tradition of resorting to ‘sporadic’ changes to explain the data that do not conform to some preconceived view of historical phonology. In the case of Mari this is most conspicuous in the work of Bereczki (1994), who operates with wholesale irregular change that is assumed to have affected individual word-roots on a random basis. However, also Itkonen’s (1954) attempt to account for the Uralic background of Mari vocalism frequently evokes the mysterious force of ‘sporadic’ change to explain away those parts of data that are incompatible with his own reconstruction of Uralic vocalism.

Thus, the notion of ‘sporadic’ sound change is a source of ad hoc hypotheses whose methodological purpose seems to be to avoid addressing flaws in theory: ‘sporadic’ changes are brought to account for those regular sound correspondences that lack an explanation in the framework of Uralic historical phonology endorsed by the author. However, if regular sound correspondences are thought to result from both regular and irregular change, the basis for a consistent application of the comparative method is lost. Hence, the alleged instances of ‘sporadic’ change should rather be viewed as indicators of weak points in theories of Uralic historical phonology. An essential task for future research in comparative Uralic linguistics is to examine whether regular accounts of supposedly ‘sporadic’ phonological developments can be provided, and the explanatory power of theories of Uralic historical phonology increased accordingly.

Table 7. The Proto-Mari initial syllable vowels.

i	ü	u
ĩ	ǔ	ǘ
e		o
ä	a	ã

Table 8. The reflexes of Proto-Mari initial syllable vowels in the Mari varieties.

Proto-Mari	E	Ki	V	U	Nw	W
*ä	a	a	a	a	ä	ä, e ¹
*e	e	e	e	e	e	e, ä ²
*i	i, e ³	i, e ³	i, e ³	i, e ⁴	i, e ⁴	i, e ⁴
*ü	ü, ö ⁵	ü, ö ⁵	ü, ö ⁵	ü, ö ⁵	ü, ö ⁶	ü, ö ⁶
*a	a	a	a	a	a	a
*ã	o	o	o	o	a	a
*o	o, u ⁷	o, u ⁷	o, u ⁷	o, u ⁷	o, u ⁷	o
*u	u	u	u	u	u	u
*ĩ	ə, i ⁸	i	ə	ĩ	ə	ə
*ǔ	ü	ü	ǔ	ǔ, ü ⁹	ǔ	ə
*ǘ	u	u	ǘ	ǘ	ǘ	ê

Notes:

¹ e before *j² ä before a velar or *r and in *CV-type roots³ e before sonorants, except in word-initial position (?)⁴ e before *r⁵ ö before *r and between a coronal sonorant and a coronal continuant (not δ)⁶ ö before *r⁷ u word finally and before a hiatus⁸ i before the cluster *ńć, and in B Ka Kr also before *ć⁹ usually ǔ, but numerous exceptional forms with ü are attested

Abbreviations for Mari varieties

B	Birsk	Ki	Bolshoj Kil'mez	MU	Mari-Ushem	V	Volga
E	East (any of the varieties B Ka Kr M MU S)	Ka	Kaltasy	Nw	Northwest	W	West
		Kr	Krasnoufimsk	S	Sernur		
		M	Morki	U	Upsha		

Literature

- AIKIO, Ante 2002. New and Old Samoyed Etymologies. *Finnisch-Ugrische Forschungen* 57: 9–57.
- AIKIO, Ante 2006. New and Old Samoyed Etymologies (Part 2). *Finnisch-Ugrische Forschungen* 59: 9–34.
- AIKIO, Ante 2007. Etymological Nativization of Loanwords: a Case Study of Saami and Finnish. In: Ida TOIVONEN & Diane NELSON (eds.), *Saami linguistics*. Amsterdam & Philadelphia: John Benjamins. Pp. 17–52.
- AIKIO, Ante 2012. On Finnic long vowels, Samoyed vowel sequences and Proto-Uralic *x. In: Tiina HYYTIÄINEN, Lotta JALAVA, Janne SAARIKIVI & Erika SANDMAN (eds.), *Per Urales ad Orientem: Iter polyphonicum multilingue*.

Festschrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012. Mémoires de la Société Finno-Ougrienne 264: 227–250.

- AIKIO, Ante (forthcoming). Studies in Uralic etymology III: Mari etymologies.
- BERECZKI, Gábor 1988. Geschichte der wolgafinnischen Sprachen. In: Denis SINOR (ed.), *The Uralic languages. Description, history and foreign influences*, pp. 314–350. Leiden & New York & København & Köln: E.J. Brill.
- BERECZKI, Gábor 1992. *Grundzüge der tscheremissischen Sprachgeschichte II.* Studia Uralo-Altaica 34. Szeged.
- BERECZKI, Gábor 1994. *Grundzüge der tscheremissischen Sprachgeschichte I.* Studia Uralo-Altaica 35. Szeged.
- HONTI, László 1984: *Chrestomathia Ostiacica.* Budapest: Tankönyvkiadó.
- HPUL = SAMMALLAHTI, Pekka 1988. Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permian. In: Denis SINOR (ed.), *The Uralic languages. Description, history and foreign influences*, pp. 478–554. Leiden & New York & København & Köln: E.J. Brill.
- ITKONEN, Erkki 1954. Zur Geschichte des Vokalismus der ersten Silbe im Tscheremissischen und in den permischen Sprachen. In: *Finnisch-Ugrische Forschungen* 31: 149–345.
- RÄSÄNEN, Martti 1920. *Die tschuwassischen Lehnwörter im Tscheremissischen.* Mémoires de la Société Finno-ougrienne 48. Helsinki: Société Finno-ougrienne.
- SALMINEN, Tapani 2007. Notes on Forest Nenets phonology. In: Jussi YLIKOSKI & Ante AIKIO (eds.), *Sámit, sánit, sátnehámít.* Riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007. Mémoires de la Société Finno-Ougrienne 253. Pp. 349–372.
- SEBEOK, Thomas A. & INGERMANN, Frances J. 1961. *An Eastern Chereemis Manual: Phonology, Grammar, Texts, Glossary.* Indiana University Publications, Uralic and Altaic Series vol. 5. Bloomington: Indiana University.
- SSA = Erkki ITKONEN & Ulla-Maija KULONEN (eds.) 1992–2000. *Suomen sanojen alkuperä. Etymologinen sanakirja.* Helsinki: Kotimaisten kielten tutkimuskeskus & Suomalaisen kirjallisuuden seura.
- TschWb = MOISIO, Arto & SAARINEN, Sirkka 2008. *Tscheremissisches Wörterbuch.* Aufgezeichnet von Volmari Porkka, Arvid Genetz, Yrjö Wichmann, Martti Räsänen, T. E. Uotila und Erkki Itkonen. Lexica Societatis Fenno-Ugricae XXXII. Helsinki: Suomalais-ugrilainen seura & Kotimaisten kielten tutkimuskeskus.
- UEW = RÉDEI, Károly 1988–1991. *Uralisches Etymologisches Wörterbuch.* Budapest: Akadémiai Kiadó.

Appendix A. Correspondences of reduced vowels between Mari varieties

This Appendix includes lists of those Proto-Mari word-roots for which Proto-Mari *i, *ǔ or *ũ can be reconstructed according to Itkonen's (1954) system of Proto-Mari vocalism. For each word, the reflexes of the first syllable vowel in the Mari varieties documented in TschWb are given. Reflexes judged irregular appear underlined and in bold font weight.

Proto-Mari *i:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*ćícá 'nipple'	i	<u>u</u>	i	i	ə, <u>i</u>	ə, <u>i</u>	ə, <u>i</u>	ə	ǔ	ə	ə
*ćǵalte- 'tickle'	ə	ə	ə	i	ə	ə			ǔ	ə	ə
*ćilä 'all'	ə	ə	ə	i	ə	ə	ə	ə	ǔ	ə	<u>i</u>
*ćirkæne- 'hurry'	ə							ə		ə	ə
*ćiwälte- 'bathe'	ə		ə		ə	ə		ə	ǔ	ə	ə
*ćiwäšte- 'pinch'	ə	ə	ə	i	ə	ə		<u>i</u>		<u>i</u>	ə
*ćiwätan 'pinchers'	ə	ə		i		ə		ə	ǔ	ə	ə
*ćime- 'stretch'	ə	ə	ə	i	ə	ə		<u>ũ</u>	<u>ũ</u>	ə	ə
*íca 'older brother'	i	i	i	i	<u>i</u>	<u>i</u>	<u>i</u>	ə	ǔ	ə	ə
*íle- 'live'	<u>i</u>	<u>i</u>	<u>i</u>	i	<u>i</u>	<u>i</u>	<u>i</u> , ə	ə	ǔ	ə	ə
*índəŋšə 'nine'	<u>i</u>	<u>i</u>	<u>i</u>	i	<u>i</u>	<u>i</u>	<u>i</u>	ə	ǔ	<u>ũ</u>	ə
*írwećá 'young'	ə	<u>ø</u>	ə	<u>u</u>	ə	ə, <u>ø</u>	ə	<u>ø</u>	ǔ	ə	ə

Proto-Mari *i:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*iške (reflexive pronoun)	∅	ə, ∅	∅	i	∅	ə, ∅	∅	∅	∅	∅	∅
*ište- ‘make, do’	ə		ə		ə	ə	ə	ə	ĩ	ə	ə
*ištar ‘footwrap’	ə	ə	ə	i	ə	ə	ə	ə	ĩ	ə	ə
*jīče- ‘hurt’	ə			i	ə	ə		ə	ĩ	ə	ə
*jīčke- ‘pick, pluck’				i					ĩ	ə	ə
*jīčkəšte- ‘get dislocated’	ə			i	ə	ə		ə		ə	ə, ı
*jīdan ‘bowstring’	ə		ə	i		ə		ə		ə	ə
*jīde ‘every’	<u>i</u> , <u>e</u>		ə		<u>e</u>	<u>e</u> , ə	ə	ə	ĩ	ə	ə, ı
*jīdəm ‘threshing floor’	<u>i</u>	<u>i</u>	<u>i</u>	i	<u>i</u>	<u>i</u>	ə	ə	ĩ	ə	ə
*jīgəлта- ‘mock’	<u>i</u>			i	<u>i</u>				ĩ		ə
*jīgəžge- ‘be disgusted’	ə		ə	i	ə	ə	ə	ə	ĩ		
*jīl ‘earthworm’	ə	ə, <u>e</u>		i	ə	ə	ə	ə		ə	ə, ı
*jīle ‘quick, soon’	ə	ə		i				ə		ə	ə
*jīlgəža- ‘shine, glitter’	ə	ə		ı	ə	ə		ə	ĩ	ə	ə, ı
*jīlmə ‘tongue’	ə	ə	ə	<u>ü</u>	ə	ə	ə	ə		ə, ı	ə, ı
*jīlt ‘entirely’	ə				ə	ə		ə	ĩ	ə, <u>ö</u>	ı
*jīme- ‘go numb’	ə	ə	ə	i	ə	ə		ə	ĩ	ə	ə
*jīnəše- ‘whimper’	ə, <u>e</u>			<u>e</u>	ə	ə		ə			ı
*jīngəl/r ‘bell’	ə, <u>o</u>					ə				ə	ə, ı
*jīp ‘spark’					ə		ə	<u>ü</u>		<u>i</u>	<u>i</u>
*jīps ‘spear, sharp stick’	ə	ə		i			ə	ə			ı
*jīr ‘around’	ə	ə	ə	ı	ə	ə	ə	ə	ĩ	ə	ə
*jīwəžge ‘still, quiet’					ə	ə		ə		ə	ə
*jīžəŋ ‘joint’	ə	ə	ə	ı	ə	ə	ə	ə	ĩ	ə	<u>e</u>
*jīžge ‘quiet, slow’	ə				ə	ə		<u>ü</u>		ə	
*kīcət ‘now’	<u>i</u> , <u>a</u>	<u>i</u>		<u>i</u>	ə	ə	ə	ə		ə	ə
*kīčke- ‘harness’	<u>i</u> , <u>a</u>	<u>i</u>	<u>i</u>	<u>i</u>	<u>i</u> , ə	ə	ə	ə	ĩ	ə	ə
*kīčkə ‘seed’	<u>i</u>	<u>i</u>	<u>i</u>	<u>i</u>	<u>i</u>	ə	ə	ə	ĩ	ə	ə
*kīčala- ‘search’	ə	ə	ə	i	ə	ə	ə	ə	ĩ	ə	ə
*kīčək, *kīča ‘fresh, fine snow’	ə	ə		ı	ə	ə	ə	ə	ĩ	ə	ə
*kīdal ‘waist’	ə	ə	ə	i	ə	ə	ə	ə	ĩ	ə	ə
*kīdež ‘room’	ə			i	ə	ə	ə	ə	ĩ	ə	ə
*kīlakš ‘best linen fibers’	<u>ü</u>	<u>ü</u>	<u>ü</u>	<u>ü</u>	<u>ü</u>	ə	ə	ə	ĩ	ə	ə
*kīlde- ‘bind’	ə	ə	ə	i	ə	ə	ə	ə	ĩ	ə	ə
*kīlōmdə ‘navel’	ə	ə	ə	i	ə	ə	ə	ə	ĩ	ə	ə
*kīlmə ‘cold’	ə	ə		i	ə	ə	ə	ə	ĩ	ə	ə
*kīńe ‘hemp’	ə			i	ə	ə	ə	ə	ĩ	ə	ə
*kīńela- ‘stand up, rise, wake up’	ə			i	ə	ə	ə	ə	ĩ	ə	ə
*kīńer ‘ell’	ə			i	ə	ə	ə	ə	ĩ	ə	ə

Proto-Mari *i:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*kīre- 'hit'	ə	ə	ə	i	ə	ə		∅	ǃ, ∅	ə	ə
*kīrəkš 'ruffe'							ə	ə		ə	ə
*kīrme- 'grab'				ǃ	ə	ə		ə	ǃ	ə	ə
*kīša 'track'	ə	<u>i</u>	ə	i	<u>i</u>	ə	ə	ə	ǃ	<u>i</u>	<u>i</u>
*kīsəja 'great tit'	<u>i</u>	<u>i</u>	<u>i</u>	i	<u>i</u>	<u>i</u>		ə	∅	ə	ə
*kīškar 'swift (for yarn)'	ə	ə	ə	i	ə	ə	ə	ə	ǃ	ə	ə
*kīške- 'throw'	ə	ə	ə	i	ə	ə	ə	ə	ǃ	ə	ə
*kīškə 'snake'	<u>i</u>	<u>i</u>	<u>i</u>	i	<u>i</u>	<u>i</u>	ə, <u>i</u>	ə	ǃ	ə	ə
*kīškə-wuj ~ *kīškə-wi 'cowry'	ə			i	ə	ə	ə	ə	ǃ	ə	ə
*līgə 'scale'	ə	<u>i</u>	<u>i</u>	i	ə	ə	ə	ə	ǃ	ə	ə
*līŋ 'much, very'	ə	ə			ə	ə		ə		ə	
*līp/wə 'butterfly'					ə	ə	ə	ə	ǃ	<u>ǃ</u>	ə
*līštaš 'leaf'	ə	ə		i	ə	ə	ə	ə	ǃ	ə	ə
*līšte- 'make, do'		ə		i	ə			ə		ə	
*līwərgə 'flexible'	ə			i	ə	ə	ə	ə, <u>ǃ</u>	ǃ	<u>ǃ</u>	<u>ǃ</u>
*līwəžge- 'wilt'	ə	ə	ə	i	ə	ə	ə	ə	ǃ	ə	ə
*līžga 'mild, calm'				ǃ	ə	ə		ə	ǃ	ə, <u>ǃ</u>	ə
*mīn/jə 'I'	ə, <u>i</u>	<u>i</u> , <u>e</u> , <u>o</u>	<u>i</u>	i	ə	ə	ə	ə	ǃ	ǃ	ə, ǃ
*mīndər 'happy'								ə	ǃ		ə
*mīžar 'coat'	ə	ə	ə	i	ə	ə	ə	ə	ǃ	ə, <u>ǃ</u>	ə
*nīl 'four'	ə	ə	<u>i</u>	i	ə	ə	ə	ə	ǃ, ǃ	ə	ə
*nīžəlgə 'mild, soft'	ə		ə		ə		ə	ə		ə	ə
*nīcəlge- 'get scratched (skin)'	<u>ǃ</u>		<u>ǃ</u>						ǃ		ə
*nīgəlge- 'get scratched (skin)'	ə			i	ə	ə		ə	ǃ		ə
*nīkta- 'skin'	ə		ə	i	ə	ə, <u>e</u>	ə	<u>i</u>	ǃ	<u>i</u>	ə
*nīmərge- 'get squashed'	ə				ə			ə	ǃ		ə
*pīce- 'adhere'	<u>ǃ</u>			i	ə	ə		ə		ə	
*pīcə 'meat'	i				<u>i</u> , ə	ə	ə	ə		ə	
*pīcəre- 'squeeze'	<u>ǃ</u> , i	<u>ǃ</u> , i	i	i	ə	ə	ə	ə	ǃ	ə	ə
*pīclə, *pīcəlmə 'rowan'	<u>ǃ</u> , i	i	i	i	ə, <u>i</u>	<u>i</u>	ə	ə		i	ə
*pīčək 'vagina'	ə				ə	ə	ə	ə		<u>ǃ</u>	ə
*pīdala- 'defend, rescue'				i	<u>i</u> , ə		ə	ə			
*pīl 'cloud'	ə	ə		i	ə	ə	ə	ǃ	<u>ǃ</u>	ə	ə
*pīldəra 'lewd'	ə			ǃ	ə	ə			ǃ		ə
*pīləkš 'ear'	ə	ə	ə	i	ə	ə	ə	ə	ǃ	ə	ə
*pīrdəš/ž 'wall'	ə			<u>ǃ</u>	ə	ə	ə	ə	<u>ǃ</u>		
*pīšte- 'put'	ə	ə	ə	i	ə	ə	ə	ə	ǃ	ə	<u>i</u>
*pīstəl 'feather'	ə	ə	ə	i	ə	ə	ə, <u>i</u>	<u>i</u>	<u>i</u>	<u>i</u>	
*pīzakš 'nest'	<u>u</u> , ə	ə	ə	i	ə	ə	ə, <u>ǃ</u>	<u>ǃ</u>	<u>ǃ</u> , ǃ	<u>ǃ</u>	ə

Proto-Mari *i:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*r̥wəʒ ‘fox’	ə	ə	ə	i	ə	ə	<u>ü</u>	<u>ü̃</u>		ə	ə
*sila- ‘esape, hide’	ə	ə	ə	i	ə	ə	ə	ə	ǐ	ə	ə
*šir ‘nature, character’	ə				ə			ə			ə
*šire- ‘get angry’	ə	ə	ə	i	ə	ə	ə	ǐ	<u>ə</u>		
*šim ~ *šišəm ‘seven’	ə	ə	ə	i	ə	ə	ə	ə	ǐ	ə	ə
*šidə ‘hatred’	<u>i</u> , ə	<u>i</u>	<u>i</u>	i	ə	ə	ə	ə	ǐ	ə	ə
*šigəʎə ‘wart’	ə	ə	ə	i	ə	ə	ə	ə	ǐ	ə	ə
*šiks ‘smoke’	<u>i</u>	<u>i</u>	<u>i</u>	i	<u>i</u>	<u>i</u> , ə	<u>i</u>	<u>i</u>	ǐ	ə	ə
*šil ‘meat’	ə	ə	ə	i	ə	ə	ə	ə	ǐ	ə	ə
*šiləʒ ‘lower back’	ə	ə	ə		ə	ə	ə	ə		ə	ə
*šima ‘dark (of clouds)’	ə			i	ə	ə		ə		ə	ə
*šínca ‘eye’	i	i	i	i	i	i	i, <u>ə</u>	ə	ǐ	ǐ	ǐ, ə
*šínca- ‘sit down’, *šínce- ‘sit, stand’	i	i	i	i	i	i	i, <u>ə</u>	ə	ǐ	ǐ	ǐ, ə
*šínce- ‘know’	i	i	i	i	i	i	i	ə	ǐ	ǐ	ǐ
*šinde- ‘put, set’	ə	ə	ə	i	ə	ə	ə	ə	ǐ	ǐ	ǐ, ə
*šine- ‘check, keep an eye on’	ə				ə						ə
*šire- ‘grimace’	ə			i		ə	ə	ə			
*širkama ‘brooch’	ə	ə		i	ə	ə		<u>ü</u>	ǐ	<u>ü</u>	ə
*širt ‘evil spirit (of illness)’	ə				ə	ə		ə		ə	
*šistə ‘wax’	<u>i</u>	<u>i</u>	<u>i</u>	i	<u>i</u>	<u>i</u> , ə	ə	ə	ǐ	ə	ə
*šizə ‘drizzle’ (+ derivatives)	ə	ə		i	ə	ə	ə	<u>ü</u>	ǐ	ə	ə
*šizə ‘autumn’	ə, <u>i</u>	ə	<u>i</u>	i	<u>i</u>	ə	ə	<u>ü</u>	ǐ	ə, <u>ü</u>	ə
*t̥iləcə ‘moon’	ə	ə	ə	i	ə	ə	ə	ə	ǐ	ə	ə
*t̥in/jə ‘you’	ə	<u>i</u>	<u>i</u>	i	ə	ə	ə	ə	ǐ	ǐ, ə	ǐ, ə
*t̥iŋga ‘gadfly’	ə	ə		i	<u>i</u>	ə		ə	ǐ	<u>e</u>	ə
*t̥irtəʃ ‘round’	ə				ə	ə	ə	ə	ǐ	ə	ə
*wíc ‘five’	i	i	i	i	<u>i</u>	<u>i</u>	<u>i</u> , ə	ə	<u>i</u>	ə	ə
*wíčkəʒ ‘thin’	i	i	i	i	<u>i</u>	ə	ə	ə	ǐ	ə	ə
*wínər ‘linen’	<u>ə</u> , i	i	i	i	ə	ə	ə	ə	ǐ	ə	ə
*wirləŋə ‘wagtail, titmouse’	ə				ə	<u>ü</u>		ə		<u>ü</u>	ə

Proto-Mari *ú:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*čüme- 'tread'	u	u		u	u	u	u	ê		ê	ê
*jügǵáca- 'have hiccups'	ü	ü	ü	ü	ü	ü	u	ǔ		ê	ə
*jüle- 'burn'	ü	ü	ü	ü	ü	ü	u	ǔ	ê	ê	ə
*jümə 'god'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*jüt 'night'	ü	ü	ü	ü	ü	ü	ü	ü	ǔ	ö	ê, ə
*kü- (interrogative pronoun root)	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*küče- 'hold, grab'	u	u	u	u	u	u	u	ǔ	ǔ	ê	ê
*küdala- 'run (animal); ride fast'	u	u	u		u	u	u	ǔ	ǔ	ê	ê
*küdakša- 'take off'	u		u	u	u	u	u	u	ǔ	ê	ê
*küm 'three'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*küm̄da 'broad'	u	u	u	u	u	u		ǔ		ö	ê
*küme- 'close the eyes'	u	u	u	u	u	u	u	ǔ		ö	ê
*kümək 'upside down'	u	u			u	u	u	ǔ	ǔ	ö	ê
*küməž 'birch-bark'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*künəž 'dirt'	u							ǔ			ê
*küpe- 'get moldy'							u	ǔ	ǔ	ö	ê
*küpšələ 'jay'		u		u	u	u	u	ǔ		ö	ê
*küptərge- 'get wrinkled'	u	u		u	u	u		ǔ	ǔ	ö	ê
*kürala- 'plough'	u	u	u	u	u	u	u	ǔ	ǔ	ê	ê
*kürək 'mountain'	u	u	u	u	u	u	u	ǔ		ö	ê
*kürəkš 'bark basket'	u	u				u	u	ǔ	ǔ	ö	ê
*kürgə 'food, fodder'					u			ǔ		ö	
*kürgəža- 'run'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*kürməcak 'woodcock'	u			u		u	u	ǔ	ǔ		ê
*küškeda- 'tear'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*küškəža- 'mount (horse)'	u	u	u	u	u	u	u	ǔ			ê
*küt 'length'	u			u	u	u		ǔ	ǔ	ö	ê
*kütka 'ant'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*küwa 'old woman'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*küwəl 'bubble'				u	u	u		ǔ		ö	ê
*küž 'urine'							u	ǔ	ǔ	ö	ê
*lüda- 'count'	u	u		u	u	u	u	u	ǔ	u	ê
*lüdə 'duck'	u	u	u	u	u	u	u	ǔ	u	ö	ê
*lüge- 'mix'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*luj 'marten'	u		u	u	u	u	u	u	u	u	ê
*lük 'corner, bend'	u	u	u	u	u	u		ǔ	ǔ	ö	ê
*lükta- 'take out'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê
*lüm 'snow'	u	u	u	u	u	u	u	ǔ	ǔ	ö	ê

Proto-Mari *ú:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*lúmej ‘blackfly’	u, <u>ü</u>	u		u		u	u	ǔ		ǒ	ê
*lúpš ‘dew’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*lúpš ‘whip’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*lúške- ‘loosen’	u	u		u	u	u	u	<u>u</u>	ǔ	<u>u</u>	ê
*múnčalte- ~ *púnčalte- ‘slide’	u	u	u	u	u	u	u	ǔ		ǒ	
*múč ‘end’	u	u	u	u	u	u	u	ǔ, <u>u</u>	ǔ	ǒ, ê	ê
*múčə ‘hazel grouse’	u			u	u	ê, <u>i</u>	u	ǔ	<u>i</u>	ǒ	ê
*múčə-wuj ‘tussock’	u	u	u	u	u		u	ǔ, <u>u</u>		ǒ	ê
*müğəłə ‘gnarl’	u	u		u	u	u, ê	u	ǔ	ǔ	ǒ	ê
*müğər ‘bend’						u		ǔ			ê
*mündəra ‘ball (of yarn)’	u	u	u	u	u	u	u	ǔ		ǒ	ê
*münə ‘egg’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*múnəj ‘toad’	u	u	u	u				ǔ			ê
*mürə ‘song’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*múška- ‘wash’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*múškəndə ‘fist’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*múžeda- ‘tell the fortune’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	<u>u</u>
*múžə ‘illness; evil spirit’	u	u		u	u		u				ê
*múšge- ‘chew something soft’						u		ǔ		ê	ê
*núčəl- ‘scratch’	u			u	u	u		ǔ		ê	ê
*nügədə ‘thick (of fluids)’	u		u	u	u	u	u	ǔ	ǔ	ǒ	ê
*núle- ‘lick’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*núlgə ‘silver fir’	u	u		u	u	u		<u>u</u>		ǒ	
*núnə ‘they’	u	u			u	u	u	<u>ü</u>	ǔ	<u>ü</u>	<u>a</u>
*núr ‘field’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*núška-, *núškəšta- ‘crawl’	u	u		u	u	u		ǔ		ǒ	
*púč ‘stalk, tube’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ, ê	ê
*púče- ‘fall (of water level)’	u	u		u	u	u	u	ǔ	ǔ	ǒ	ê
*púčəšte- ‘itch’	u				u	u		ǔ	ǔ	ǒ	ê
*púdešta- ‘burst’	u	u		u	u	u	u	ǔ	ǔ	ǒ	ê
*púdərge- ‘break’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*púl-wuj ‘knee’	u	u	u	u	u	u	u	ǔ	<u>u</u>	<u>u</u>	ê, <u>u</u>
*pün ‘hair’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*púnčala- ‘wring’	u				u	u	u	ǔ	ǔ	ǒ	ê
*pündaš ‘bottom’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*püne- ‘braid’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*púnəlmə ‘bumblebee’							u	ǔ		ǒ	ê
*pünəške- ‘get moldy’	<u>ü</u>		<u>ü</u>		<u>ü</u>	ê		ǔ	ǔ	ǒ	ê
*püra- ‘bite, chew’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*püre- ‘enter’	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê

Proto-Mari *ú:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*pürgeda- 'hoe, dig up, burrow'	u				u	u	u	ǔ	ǔ	ǒ	ê
*pürgəšte- 'snow over'	u			u	u	u	u	ǔ		ǒ	ê, <u>u</u> , <u>ü</u>
*püskəla- 'sting'	<u>ü</u>	<u>ü</u>	<u>ü</u>	<u>ü</u>	<u>ü</u>	u, <u>ü</u>	u	ǔ	ǔ	ǒ	ê
*püş 'boat'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*püşkeda- 'have diarrhea'	u		u	u	u	u	u	ǔ	ǔ	ǒ	ê
*püşkədə 'soft'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*püt, *pütərak 'strong'	u			u	u	u		ǔ	<u>ı</u>	<u>ê</u>	ê
*püžar 'plane'	u	u		u	u	u	u	ǔ	ǔ	ǒ	ê
*rüde- 'unbind'	u	u	u	u	u	u	u	ǔ	<u>u</u>	ǒ	ê
*rümbək 'mud'	u	u		u	u	u	u	ǔ		ǒ	ê
*šüwan 'boil, abscess'	<u>ü</u>	<u>ü</u>	<u>ü</u>	<u>ü</u>		<u>ü</u>		ǔ		<u>ê</u>	<u>e</u>
*šügəñə 'lever'	u		u	u	u	u		ǔ	ǔ	ǒ	ê
*šüldər 'feather'	u	u	u	u	u	u	u	<u>u</u>	ǔ	ǒ	ê
*šüle- 'melt'	u	u	u	u	u	u		ǔ	ǔ	ǒ	ê
*šüləkš 'boot leg'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	
*šüma- 'get tired'	u	u		u	u	u		ǔ		ǒ	ê
*šüme- 'whet'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*šüŋalta- 'fall head-on'	u	u	u	u	u	u		ǔ	ǔ		ê
*šüpša- 'pull, suck'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*šür 'shit'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*šüre- 'pound, crush'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*šüwəkš 'leather sack'	u	u		u	u	u	u	ǔ	ǔ, <u>ı</u>	ǒ	<u>u</u>
*tüdə 'this'	u	u	u	u	u	u	u	<u>ü</u>	ǔ	<u>ü</u>	<u>e</u>
*tügər 'shirt'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*tüjə 'sick, lean'	u				u	u		<u>u</u>	<u>u</u>	ǒ	
*tül 'fire'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*tünəma- 'learn'	u	u	u	u	u	u	<u>ü</u>	<u>ü</u>	<u>ü</u>	ǒ	ê
*tünğər 'thick tree bark'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*tüp 'back'	u	u	u	u	u	u	u	ǔ	ǔ	<u>u</u>	ê
*tüpka 'heckled flax or hemp'						u	u	ǔ	ǔ	ǒ	ê
*türta- 'shrink'	u	u	u	u	u	u	u	<u>u</u>	<u>u</u>	ǒ	ê
*türəža- 'trample'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê
*tüške- 'glue'	u			u	u	u		ǔ	ǔ	ǒ	ê
*tüšte- 'ask a riddle'	u	u	u	u	u	u	u	<u>u</u>	<u>u</u>	ǒ	ê
*tütəš 'often, constantly'	u				u	u					ê
*tűwəle- 'defend, rescue'							u	ǔ			ê
*tűwərgə- 'curdle, turn sour'	u	u	u	u	u	u		ǔ	ǔ	ǒ	ê
*üdəla- 'pray for'	u	u		u	u	u		ǔ			ê
*üdəre- 'rake'	u	u	u	u	u	u	u	ǔ	ǔ	<u>ê</u>	ê
*üla- 'be'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ê

Proto-Mari *ú:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*úlde- 'ask for'						u		ǔ		ǒ	
*ǔlmə 'man'	u	u						<u>ǔ</u>		ǒ	
*ǔmša 'mouth'	u	u	u	u	u	u	u	ǔ		ǒ	<u>ə</u>
*ǔmər 'warm'	u	u		u	u	u		ǔ		ǒ	
*ǔmbal 'distant'	u	u		u		u	u	ǔ		ǒ, <u>ə</u>	ə
*ǔre- 'put in the ground'	u		u	u	u	u	u	ǔ		ǒ	ə
*ǔrə 'two handfuls'	u		u	u				ǔ		ǒ	ə
*ǔrge- 'sew'	u	u	u	u	u	u	u	ǔ		ǒ	ə
*ǔžar 'green'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ə, <u>ə</u>
*ǔžga 'fur-coat'					u	u	u	ǔ	ǔ	ǒ	ə
*wǔcək 'much'	u, <u>ǔ</u>			u	u	u		ǔ			ə
*wǔče- 'wait'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ə
*wǔle- 'get spoiled'	u					u		ǔ			ə
*wǔlnə 'tin'	u	u	u	u	u	u	u	ǔ	<u>ə, ǐ</u>	ǒ	<u>ə</u>
*wǔrde- 'tend'		u						ǔ		ǒ	<u>ə</u>
*wǔrgem 'clothes'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ə
*wǔrgəže- 'be restless'	u			u	u	u	u	ǔ		ǒ	ə
*wǔrt 'heddle'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	<u>ə</u>
*wǔž (onomatopoetic root)					u	u		ǔ		ǒ	ə
*wǔžale- 'buy'	u	u	u	u	u	u	u	ǔ	ǔ	ǒ	ə

Proto-Mari *ú:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*čüčə ‘maternal uncle’	ü		ü	ü	ü	ü	ü	ü	ü	ə	ə
*čüdə ‘lack, need’	ü			ü	ü			ü			ə
*čünge- ‘peck (of birds)’	ü	ü	ü	ü	ü	ü		ü	ɨ	ü	ə
*küč ‘nail’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*küče- ‘beg’	ü	ü	ü	ü	ü	ü	ü	ü		ü	ə
*küćə ‘knife’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü, ü	ə
*küdə ‘black grouse’	ü	ü	ü	ü	ü	ü		ü	ü	ü	ə
*küdəte- ‘thunder’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*künčä- ‘dig’	ü	ü	ü	ü	ü	ü	ü	ü		ü	
*küps ‘nap, fuzz’	ü				ü	ü		ü	ü	ü	ə
*kür ‘bast’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*kūra- ‘tear, rip’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*kürtə ‘iron’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*küśedək ‘lapwing’	ü	ü	ə	∅	ü	ü		ü		ə	ə
*küžgə ‘thick’	ü	ü	ü	ü	ü	ü	ü		ü	ü	ə
*lügəšte- ‘itch’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*lükə ‘boggy area’						ü		ü		ü	
*lüm ‘name’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*lümə ‘scab’	ü			ü	ü	ü	ü	ü	ü	ü	ɨ
*lüŋge- ‘rock’	ü	ü		ü	ü	ü	ü	ü	ü	ü	ə
*lüškə ‘shake, swing’			ü				ü	ü		ü	
*lüšte- ‘milk’	ü	ü	ü	ü	ü	ü	ü		ü		ə
*mündə ‘far’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*müškə ‘belly’	ü	ü	ü	ü	ü	ü	ü	ü		ü, ə	ə
*nüštala- ‘blow one’s nose’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə	ə
*nüškə ‘blunt’	ü	ü		ü	ü	ü	ü	ü	ü	ü	ə
*nüža- ‘scrape’	ü	ü		ü	ü	ü	ü	ü	ü	ü	ə
*püčka- ‘cut off’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*püncə ‘pine’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	
*pürde- ‘cover with a cloth’	ü				ü	ü	ü	ü	ü	ü	ə
*rüće- ‘shake’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*rüdaŋa- ‘rust’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*rüde- ‘pick, pluck’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*rüdə ‘core’	ü	ü		ü	ü	ü	ü	ü	ü	ü	ə
*rüm(b)alge- ‘get dark’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*rüpe- ‘rock’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*rüškə ‘quake, rumble’					ü	ü					ə
*rüškə ‘gnarl’	ü					ü		ü	ü	ü	ə
*südə(n)e- ‘drag’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə
*süke- ‘shove’	ü	ü	ü	ü	ü	ü	ü	ü	ü	ü	ə

Proto-Mari *ů:	B	Kr	Ka	Ki	S	M	MU	U	V	Nw	W
*sŭlə ‘fathom’	ü	ü	ü	ü	ü	ü	ü	<u>ü</u>	<u>ü</u>	<u>ü</u>	ə
*sŭm ‘sense of touch’	ü			ü	ü	ü			ŭ		ə
*sŭre- ‘smear’	ü	ü	ü	ü	ü	ü		ŭ	ŭ	ŭ	ə
*sŭrtŭe- ‘trip, tumble’	ü	ü		ü	ü	ü	ü		ŭ		ə
*sŭwǎce- ‘shell (nuts)’					ü			ŭ	ŭ	ŭ	
*sŭc ‘soot’	ü	ü	ü	ü	ü	ü	ü	<u>ü</u>	ŭ	<u>ə</u>	<u>ə</u>
*sŭdǎkš ‘barrel hoop’	ü	ü	ü	ü	ü	ü	<u>ə</u>	ŭ	<u>ə</u>	<u>ə</u>	
*sŭdər ‘spindle’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ, <u>ə</u>	ə
*sŭgə ‘bark beetle’		ü		ü			ü	ŭ		ŭ	ə
*sŭm ‘heart’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	<u>ü</u>
*sŭm ‘scale’	ü	ü	ü	ü	ü	ü, <u>ö</u>	ü	ŭ	<u>ö</u>	ŭ	<u>ü</u>
*sŭrgə ‘cheeks, face’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*sŭrgə ‘forest’				ü		ü			ŭ		ə
*sŭrtə ‘yarn’	ü	ü	ü	ü	ü	ü	ü	<u>ü</u>	ŭ	ŭ	ə
*sŭška- ‘stuff’	ü	ü	ü	ü	ü	ü		ŭ	ŭ	ŭ	ə
*sŭštə ‘leather’	ü	ü	ü	ü	ü	ü	ü	<u>ü</u>	ŭ	<u>ə</u>	ə
*sŭwala- ‘spit’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*tŭj ‘base’	ü	ü, <u>ə</u>	ü	ü	ü	ü	ü	ŭ, <u>ü</u>	ŭ	ŭ	ə
*tŭr ‘edge; blade’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*tŭreda- ‘harvest’	ü	ü	ü	ü	ü	ü		ŭ, <u>ø</u>	ŭ	<u>ə</u>	ə
*tŭrəs ‘full’	ü	ü		ü	ü	ü	ü	ŭ	ŭ		ə
*tŭrwə ‘lip’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*tŭrwǎŭca- ‘sneeze’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*tŭška ‘group (of people), herd’	ü		ü	ü		ü	ü	<u>ü</u>	ŭ		ə
*tŭwət ‘entirely’		ü		ü	ü	ü		ŭ			ə
*tŭžem ‘thousand’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*ŭdər ‘girl, daughter’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*ŭškərt ‘stubborn’	ü				ü	ü	ü	ŭ, <u>ü</u>		ŭ	
*ŭštə ‘belt’	ü	ü	ü	ü	ü	ü	ü	<u>ü</u>	<u>ü</u>	<u>ü</u>	ə
*ŭžgar ‘thing’	ü	ü			ü	ü, <u>u</u>	ü	ŭ	<u>ə</u>	ŭ, <u>ə</u>	ə
*ŭžəwər ‘common swift’	ü	ü		ü		ü		ŭ			ə
*wŭčə ‘cut, notch’		ü		ü			ü	<u>ü</u>			ə
*wŭl- ‘on, up, over’							ü	ŭ		ŭ, <u>ə</u>	ə
*wŭl/lə ‘mare’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*wŭr ‘blood’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*wŭrgeŭə ‘copper’	ü	ü		ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*wŭt ‘water’	ü	ü	ü	ü	ü	ü	ü	ŭ	ŭ	ŭ	ə
*wŭtelə ‘snipe’	ü				ü	ü	ü	ŭ, <u>ü</u>		ŭ	ə

Appendix B. Uralic etymologies of Mari words with close full and reduced vowels

PMari *i < PU *e

- E *ilem*, Ki *ilem*, V *ilem*, U Nw W *alem* ‘I live’ < PMari *ile- < PU *elä- (UEW: 73; HPUL: 536)
- E *iza*, Ki *iza*, V *iza*, U *za* (!), Nw W *azä* ‘older brother’ < PMari *iça < PU *içä ~ *eçä (UEW: 22; HPUL: 541)
- E Nw W *kaldem*, Ki *kil'dem*, V *kıldem* ‘I bind, lash’ < PMari *kilde- < PU *keltä- (UEW: 657)
- E U *laštaš*, Ki *lištaš*, V *ljštaš*, Nw *laštäš*, W *laštäš* ~ *alaštäš* (!) ‘leaf’ < PMari *lištaš < PU *lešti (UEW: 690)
- E U Nw W *nal*, Ki *nil* ‘four’ < PMari *nıl < PU *neljä (UEW: 315–316; HPUL: 547)
- E *palaš*, Ki *pil'iš*, V *pil'iš*, Nw *pələkš*, W *palaš*, *pələš* ‘ear’ < PMari *pələkš < PU *peljä (UEW: 370; HPUL: 547)
- E *pəžaš*, Ki *pizaš*, V *pžakš*, U *pžžaš* (!), Nw *pžžakš* (!), W *pžžäš* ‘nest, den’ < PMari *pžakš < PU *pesä (UEW: 375; HPUL: 539)
- E Nw *rəwəž*, Ki *riwiž*, U *rüwüž* (!), W *rəwəž* ~ *ərwəž* ‘fox’ < PMari *rəwəž < PU *repäs (UEW: 423–424; HPUL: 553)
- E U Nw *šalaž* ‘small of the back’ < PMari *šiləž < PU *šelkä (UEW: 772)

PMari *i < PU *i

- E *iza*, Ki *iza*, V *iza*, U *za* (!), Nw W *azä* ‘older brother’ < PMari *iça < PU *içä ~ *eçä (UEW: 22; HPUL: 541)
- E *káčək*, Ka *kačək*, U *kacək*, Nw *kacä löm*, W *káčä* ‘newly fallen fine snow’ < PMari *kičək, *kičä < PU *kičV- (SSA s.v. *kide*)
- E *kərem*, *krem*, Ki *kirem*, V *krem*, Nw *kərem* ‘I hit, forge, hammer, thresh’ < PMari *kire- < PU *kirä- (UEW: 666; HPUL: 552)
- E Nw W *kəškem*, Ki *kiškem*, V *kjškem* ‘I throw; I pour’ < PMari *kiške- < PU *kiški- (HPUL: 552; cf. UEW: 667)
- Ki *jičke-*, W *jačke-* ‘pick, pluck, pull’ < PMari *jičke- < PU *ńičkä- (Aikio, forthcoming). — The change *ń- > *ji- has a parallel in PMari *jilmə ‘tongue’ < PU *ńälmä (see below). Note that E Nw W *nal*, Ki *nil* ‘four’ (< PMari *nıl) is not a counterexample to this change, because this numeral originally had *n- instead of *ń- (PU *neljä). The unpalatalized nasal is also preserved in Hungarian *négy* and many Saami languages (e.g., Ume Saami *nelja*, Inari Saami *nelji*). The initial *ń-* of some cognates (e.g., Komi *ńol*, East Khanty *ńělə*, North Mansi *ńila*) seems to have developed due the assimilatory influence of word-internal *j-.
- E Nw *pəzem*, Ki *pižem* ‘I hold on, cling to’ < PMari *piče- < PU *pitä- (Aikio, forthcoming)
- E *məń*, *məj*, *məjə*, Ki *miń*, U *məń*, *məńe*, V *mijń*, Nw *miń*, W *mińə*, *məńə* ‘I’ < PMari *mińə < PU *minä (cf. UEW: 294)
- E *šińčä*, Nw *šincä*, W *šincä*, *səncä* ‘eye’ < PMari *šin-čä < PU *šilmä (UEW: 479; HPUL: 540)
- E *təjə*, *təj*, Ki *tiń*, U *təńə*, *təń*, V *tij*, Nw *təń*, *tij*, W *təńə*, *tij* ‘you (sg.)’ < PMari *tińə < PU *tinä (cf. UEW: 539)

PMari *i < PU *ü

- E *šəže*, *šiže*, Ki *šiža*, U *šüža*, Nw *šəža*, *šüžü*, W *šəža* ‘autumn’ < PMari *šüzə < PU *sükiš (UEW: 443; HPUL: 549)
- E *káčkem*, *kičkem*, Ki *kičkem*, V *kjškem*, U Nw W *káčkem* ‘I harness’ < PMari *kičke- < PU *kütki- (UEW: 903)
- E *kəlme*, Ki *kil'me*, V *kil'me*, Nw W *kəlmə* ‘frozen’ < PMari *kilmə < PU *külmä (UEW: 663; HPUL: 552)
- E Nw W *kəńer*, Ki *kińer*, V *kjńer* ‘ell’ < PMari *kińer < PU *küńärä (UEW: 158–159; HPUL: 544)
- E *ńakta-*, Nw *ńiktä-*, W *ńaktä-* ‘skin’ < PMari *ńihta- < PU *ńühtä- (UEW: 715)
- E *šəyal'e*, Ki *šiyil'ə*, U Nw *šəyal'ə*, W *šəyal'* ‘wart’ < PMari *šigəľə < PU *šüklä (UEW: 36–37; HPUL: 549)

PMari *ji- < PU *jä-, *nä-

- E *jədaŋ*, Ki *jiđaŋ*, U *jaŋaŋ* (!), Nw W *jədaŋ* ‘bowstring’ < PMari *jidaŋ < PU *jäntiŋ (UEW: 92; HPUL: 537)
- E U Nw *jəžəŋ*, Ki *jižin*, W *ježəŋ* (!) ‘joint’ < PMari *jižəŋ < PU *jäsin (UEW: 95; HPUL: 552)
- E *jəma*, Ki *jima* ‘(a body part) grows numb; (an eye) goes blind’, U *jəma*, Nw W *jəmä* ‘hides oneself; (a tooth) aches severely; (an eye) goes blind’ < PMari *jime- ‘go numb’ < PU *jämə- (Aikio, forthcoming)
- E V *jəlme*, Ki *jülmö* (!), Nw *jilma*, W *jilma*, *jəlma* ‘tongue, language’ < PMari *jilmə < PU *ńälmä (UEW: 313; HPUL: 546).

PMari *i < PU *ä

- E *ime*, Nw *imə*, W *im* ‘needle’ < PMari *imə < PU *äjmä (UEW: 22; HPUL: 536)
- W *kiže* ‘feel cold’ < PMari *kiže- < PU *känšV- (Aikio 2002: 21; cf. UEW: 648, HPUL: 552)
- E *kit*, *kət* (!), Nw W *kit* ‘hand’ < PMari *kit < PU *käti (UEW: 140; HPUL: 545)

- E *liš-*, *laš-* (!), Nw W *liš-* ‘near’ < PMari *liš- < PU *läsi- (UEW: 687)
- E *piste*, *pište*, Nw W *pistə* ‘linden’ < PMari *pistə < PU *päksnä (UEW: 726)
- E *šište*, Nw W *šišta* ‘woodpecker’ < PMari *šišta < PU *šäsna (UEW: 772; HPUL: 554)
- E *tič*, U Nw *tic*, W *cic* ‘full’ < PMari *tič < PU *täwdi (cf. UEW: 518; HPUL: 550; Aikio 2002: 31–34)
- E V *lewa*, Ki *lewa*, U *liwa*, Nw W *liwä* ‘it becomes warm’ < PMari *liwe- < PU *lämpi- (UEW: 685; Aikio, forthcoming)
- E V *šen*, MU U Nw W *šin* ‘tinder’ < PMari *šin < PU *šänä (UEW: 494–495; HPUL: 548)
- E *wenə*, Ki *wenə*, MU U Nw *winə*, Nw *winğə* ‘son-in-law’ < PMari *winə < PU *wäniw (UEW: 565–566; HPUL: 541)
- E *wüdem* (!), Nw W *wiðem* ‘I take (somewhere), lead’ < PMari *widem < PU *wätä- (UEW: 569; HPUL: 551).
— Traditionally reconstructed as *wetä-, but both Moksha Mordvin *vätä-* ‘lead, guide’ and Hungarian *vezet* ‘leads, guides’ point to PU *ä instead of *e; the vocalism of Finnish *vetää* ‘pulls’ is irregular.

PMari *ü < PU *ü

- E *künčem*, (Ka Kr) *künčem*, Nw *künčem* ‘I dig’ < PMari *künče- < PU *künčä- (UEW: 663–664)
- E *küč*, (B Ka Kr) *küč*, Nw *küč*, W *káč* ‘nail’ < PMari *küč < PU *künči (UEW: 157; HPUL: 554)
- E *küžyö*, Nw *küžyü*, W *kəžyə* ‘thick, fat, stout’ < PMari *küžgə < PU *küsä (UEW: 161; HPUL: 544)
- E *küðar*, V U Nw *küðür*, W *kəðar* ‘black grouse’ < PMari *küðar < PU *kütVrV (UEW: 231)
- E *šülö*, Nw *šülü* (!), W *šəl* ‘fathom’ < PMari *šülə < PU *süli (UEW: 444; HPUL: 540)
- E *šüškam*, Nw *šüškäm*, W *šəškäm* ‘I cram, pack’ < PMari *šüška- < PU *süskä- (UEW: 768)
- E *šüwəl*, Nw *šüwül*, W *šəwəl*, *šəwəl* ‘spit’ < PMari *šüwəl < PU *šülki (UEW: 479–480; HPUL: 549)
- E *tünj*, Nw *tünj*, W *təj* ‘base (of a tree)’ < PMari *tünj < PU *tünji (UEW: 523–524; HPUL: 550)
- U *wül-*, Nw *wül-*, *wəl-*, W *wəl-* ‘on’ < PMari *wül- < PU *wüli- (UEW: 573–574; cf. HPUL: 536)

PMari *ü < PU *i

- E *lüm*, Nw *lüm*, W *ləm* ‘name’ < PMari *lüm < PU *nimi (UEW: 305; HPUL: 538)

PMari *ü < PU *e

- E *čüčö*, (B Ka) *čüčö*, U *cüci*, Nw *cəcə*, W *čəčə* ‘maternal uncle’ < PMari *čüčə < PU *čečä (UEW: 34–35; HPUL: 536)
- E *kür*, Nw *kür*, W *kər* ‘bast’ < PMari *kür < PU *keri (UEW: 148–149; HPUL: 543)
- E *pünčö*, (B Ka Kr) *pünčö*, Nw *pünči* ‘pine’ < PMari *pünčə < PU *pe(n)čä (UEW: 727; HPUL: 553)
- E *tür*, Nw *tür*, W *tər* ‘blade; edge; shore’ < PMari *tür < PU *terä (cf. UEW: 522, 795)
- E *wür*, Nw *wür*, W *wər* ‘blood’ < PMari *wür < PU *weri (UEW: 576; HPUL: 551)
- E *wüt*, Nw *wüt*, W *wət* ‘water’ < PMari *wüt < PU *weti (UEW: 570–571; HPUL: 541)

PMari *ü < PU *j

- E W *üp* ‘hair on the head’ < PMari *üp < PU *jpti (UEW: 14–15; HPUL: 536)
- E *ülö-*, W *ülä-* ‘under-’ < PMari *ülə- < PU *jla (UEW: 6; HPUL: 536)
- E *lüda-*, Nw W *lüdä-* ‘be afraid’ < PMari *lüda- < PU *ljdi- (Aikio, forthcoming)
- E W *šüm* ‘scale’ < PMari *šüm < PU *šjmi (UEW: 476; 549)
- E *šün*, *šön*, Ki *šün*, W *šün* ‘vein, sinew’ < PMari *šün < PU *šjni (UEW: 441; HPUL: 548)
- E V *šüdam*, Ki *šüdam*, W *šüdam* ‘I clear (e.g., road, field, forest)’ < PMari *šüda- < PU *šjinti- (Aikio, forthcoming)
- E *šüdü*, W *šüdə* ‘hundred’ < PMari *šüdə < PU *šjta (UEW: 467; HPUL: 549)
- M *nölä pikš* ‘arrow with a bone head’ < PMari *nülə < PU *njli ‘arrow’ (UEW: 317; HPUL: 539)
- E W *mör* ‘strawberry’ < PU *mjrja ‘berry’ (UEW: 264–265)
- E *nörö*, W *nöra* ‘flexible, pliable’ < PMari *nürə < PU *njri (Aikio 2012: 234)
- E *nörjö*, W *nörjə* ‘cartilage’ < PMari *nürgə < PU *njrki (UEW: 317; HPUL: 546)

PMari *ü < PU *ä

- E Nw *jükšem*, W *ükšem* ‘I get cold’ < PMari *jükše- < PU *jäksi- (UEW: 90–91)
- E Nw W *pükš* ‘nut’ < PMari *pükš < PU *päski (UEW: 726–727; HPUL: 553)
- E *šükšö*, Nw W *šükšə* ‘rag; worn-out, bad, unsuitable; rotten’ < PU *säksä (Aikio, forthcoming)

PMari *ũ < PU *u

- E *jumo*, Nw *jōmo*, W *jâmâ* ‘god; heaven’ < PMari *jũmæ < PU *juma (UEW: 638)
- E *kum*, V U *kũm*, Nw *kōm*, W *kâm* ‘three’ < PMari *kũm < PU *kolmi ~ *kulmi (UEW: 174; HPUL: 543)
- E *kumæk*, V U *kũmuk*, Nw *kōmøk*, W *kâmæk* ‘upside down’ < PMari *kũmæk < PU *kuma- (UEW: 201–202; HPUL: 537)
- E *kumða*, U *kũmða*, Nw *kōmða*, W *kâmða* ‘broad’ < PMari *kũmða < PU *kumta (UEW: 203–204)
- M *kužam*, V U *kũžam*, Nw *kōžam*, W *kâžam* ‘I urinate’ < PMari *kũžā- < PU *kunsi- (UEW: 210; HPUL: 537)
- E *kuwəl*, U *kũwul*, Nw *kōwol*, W *kâwəl* ‘blister’ < PMari *kũwəl < PU *kupla (UEW: 212–213)
- E *kurmazak*, W *kârmâzak* ‘Eurasian woodcock’ < PMari *kũrmæcak < PU *kurmićca (UEW: 676–677). — This etymology is not accepted in SSA (s.v. *kurppa*), but the correspondence between Fi *kurmitsa* ‘Eurasian woodcock’ and PMari *kũrmæc- is fully regular.
- E *kuræk*, U *kũruk*, Nw *kōrok*, W *kâræk* ‘hill, mountain’ < PMari *kũræk < PU *kuri (UEW: 677)
- E *kutko*, V *kũtko*, Nw *kōtko*, W *kâtkâ* ‘ant’ < PMari *kũtkæ < PU *kutki (UEW: 678; HPUL: 552)
- E *luđam*, V *lũđam*, Nw *luđam* (!), W *lâđam* ‘I count; I read’ < PMari *lũda- < PU *luki-(ta-) (UEW: 253; HPUL: 545)
- E *lum*, V U *lũm*, Nw *lōm*, W *lâm* ‘snow’ < PMari *lũm < PU *lumi (UEW: 253; HPUL: 538)
- E *šínćam lumem*, W *sancām lāmēm* ‘I bewitch, cast a spell on’ (*šínćam*, *sancām* ACC ‘eye’) < PMari *lũme- < PU *lumV- (UEW: 694)
- E V *luđo*, U *lũdo*, Nw *lōđo*, W *lâđâ* ‘duck’ < PMari *lũdæ < PU *lunta (UEW: 254; HPUL: 545)
- E *lupš*, Ki *luφš*, V U *lũpš*, Nw *lōpš*, W *lâpš* ‘dew’ < PMari *lũpš < PU *lupsa (UEW: 261; HPUL: 538)
- E *muč-*, V *mũč-*, *mūs-*, Nw *mōc-*, *mæc-*, W *mâč-* ‘end’ (in compounds) < PMari *mũč- < PU *muča (UEW: 283)
- E *muno*, M *munâ*, V U *mũno*, Nw *mōno*, W *mânâ* ‘egg; testicle’ < PMari *mũnâ < PU *muna (UEW: 285; HPUL: 538)
- E *nužem*, U *nũžem* ‘I scrub (my skin); I rub off, scrape off’ < PMari *nũže- < PU *nusi- (UEW: 309)
- E *puč*, (B Ka Kr) *puč*, V *pũč*, U *pũc*, Nw *pōc*, *pâc*, W *pâč* ‘hollow plant stalk, tube; shepherd’s horn’ < PMari *pũč < PU *pučki (UEW: 397; HPUL: 539)
- E *pudešteš*, V U *pũdešteš*, Nw *pōdešteš*, W *pâdešteš* ‘it bursts’ < PMari *pũdæšte- < PU *puđa- (Aikio 2006: 22–23)
- E *pun*, V U *pũn*, Nw *pōn*, W *pân* ‘body hair, animal hair, down’ < PMari *pũn < PU *puna (UEW: 402; HPUL: 547)
- E *punem*, V U *pũnem*, Nw *pōnem*, W *pânem* ‘I braid, plait, twine, twist’ < PMari *pũne- < PU *puna- (UEW: 402–403; HPUL: 539)
- E *puram*, V U *pũram*, Nw *pōram*, W *pâram* ‘I chew, bite’ < PMari *pũra- < PU *puri- (UEW: 405; HPUL: 539)
- E *purγeđam*, V U *pũrγeđam*, Nw *pōrγeđām*, W *pârγeđām* ‘I hoe up, dig up, burrow’ < PMari *pũrgeđā- < PU *purka- (UEW: 741)
- E *purγašta*, U *pũrγušta* ‘covers in snow’, Nw *pōrγōšta*, W *pũrγâšta*, *pârγâšta* ‘whirls (of snow, dust, etc.)’ < PMari *pũrgešta- (UEW: 406; HPUL: 547)
- E *pũškeš* (!), Ki *pũškeš* (!), Nw *pōškeš*, W *pâškeš* ‘stings (of an insect)’ < PMari *pũska- < PU *puski- (UEW: 408; HPUL: 547)
- E *šula*, V U *šũla*, Nw *šōla*, W *šâla* ‘it melts’ < PMari *šũla- < PU *sula- (UEW: 450–541; HPUL: 548)
- E *šurem*, V U *šũrem*, Nw *šōrem*, W *šârem* ‘I pound, crush, trample; I prick (with a pin), thrust (with a knife)’ < PMari *šũre- < PU *šurwV- (UEW: 491)
- W *tâktâ* ‘boat rib’ < PMari *tũktæ < PU *tuktV (UEW: 534; HPUL: 550)
- E *tul*, V U *tũl*, Nw *tōl*, *tâl*, W *tâl* ‘fire’ < PMari *tũl < PU *tuli (UEW: 535; HPUL: 540)
- E *tunemam*, V *tũnemam*, U *tũnemam* (!), Nw *tōmejām*, W *tâmeiām* ‘I practice, learn’ < PMari *tũnema- < PU *tuni- (UEW: 537; HPUL: 550)

PMari *ũ < PU *i (in disharmonic roots)

- E *kuškeđam*, V U *kũškeđam*, Nw *kōškeđām*, W *kâškeđām* ‘I tear off, tear in two’ < PMari *kũškeda- < PU *kiška- (Aikio, forthcoming; cf. HPUL: 552)
- E *užar*, V U *ũžar*, Nw *ōžar*, W *âžar*, *žar* ‘green’ < PMari *ũžar < PU *wiša(-ra) (UEW: 823; HPUL: 554)

PMari *ũ < PU *o (adjacent to labial consonants)

- E *kum*, V U *kũm*, Nw *kõm*, W *kâm* ‘three’ < PMari *kũm < PU *kolmi ~ *kulmi (UEW: 174; HPUL: 543)
- E *kuməž*, U *kũmuž*, Nw *kõmož*, W *kâməž* ‘birch-bark’ < PMari *kũməž < PU *kolmis (Aikio, forthcoming)
- E *kuwo*, V *kũwo* ‘husk’ < PMari *kũwə < PU *kopa (UEW: 180; HPUL: 537)
- E *kuwəlčo*, U *kũwũlzo* ‘wood grouse hen’ < PMari *kũwəlčo < PU *koppala (UEW: 181)
- E *kupa*, V U *kũpa*, Nw *kõpa*, W *kâpa* ‘gets moldy’ < PMari *kũpe- < PU *koppi- (UEW: 680)
- E *muškam*, V *mũškam*, Nw *mõškam*, W *mâškam* ‘I wash’ < PMari *mũška- < PU *moški- (UEW: 289; HPUL: 538)
- E *šumam*, U *šũmam*, Nw *šõmam*, W *šâmam* ‘I get tired, languish’ < PMari *šũma- < PU *šoma- (Aikio, forthcoming)
- E *ulam*, V U *ũlam*, Nw *õlam*, *alam*, W *âlam* ‘I am’ < PMari *ũla- < PU *woli- (UEW: 580–581; HPUL: 551)
- E *uryem*, U *ũryem*, Nw *õryem*, W *âryem* ‘I sew’ < PMari *ũrge- < PU *worka- (UEW: 584–585; HPUL: 551)
- E *wučem*, (B Ka Kr) *wučem*, V *wũčem*, U *wũcem*, Nw *wõcem*, W *wâcem* ‘I wait’ < PMari *wũče- < PU *woča- (UEW: 334)
- E *užalem*, V U *ũžalem*, Nw *õžalem*, W *wâžalem* ‘I sell’ < PMari *wũžale- < PU *wosa (UEW: 585; HPUL: 551)

PMari *u < PU *o

- E V U Nw W *ukš*, Ki *uks* ‘branch’ < PMari *ukš < PU *oksa (UEW: 716; HPUL: 552)
- E *ukšəńcam*, *ukšĩńcam*, V *ukšjńcam*, U *ukšəncam*, Nw *ukšəncam*, W *ukšəncam*, *ukšənzam* ‘I vomit’ < PMari *ukšəńca- < PU *oksinta- (UEW: 716; HPUL: 552)
- E *ur* (Kr M S *ur* ~ *ər!*), Ki V U Nw W *ur* ‘squirrel’ < PMari *ur < PU *ora (UEW: 343; HPUL: 552)
- E Ki V U Nw W *wuj* ‘head; end; ear (of corn); tree top’ < PMari *wuj < PU *ojwa (UEW: 336–337; HPUL: 536)
- E *kutkaž*, W *kučkəž* ‘eagle’ < PMari *kut/čkəž < PU *kočka (UEW: 668; HPUL: 552)
- E Ki V U *kuđo*, Nw *kuđə*, W *kuđə* ‘Mari summer house’ < PMari *kuđə < PU *kota (UEW: 120; HPUL: 543)
- E *püčö*, (B Ka) *püčö* (!), Ki *pučo*, V *pjče* (!), U *puco*, Nw *puca*, W *pučə* ‘(wild) reindeer’ < PMari *pučə < PU *počaw (UEW: 387–388; HPUL: 553)
- E *pulaš*, V *pũliš*, Nw *pulõkš*, W *pulaš* ‘shoulder’ < PMari *pulaš < PU *pola (UEW: 734; HPUL: 553)
- E V Nw *pul-wuj*, U *pũl-wuji* (!), W *pul-wuj*, *pəl-wuj* (!) ‘knee’ < PMari *pul- < PU *polwi (UEW: 393; HPUL: 539)
- E Ki V U Nw W *šukš* ‘worm’ < PMari *šukš < PU *soksi (UEW: 764)
- E Ki V U Nw W *šudalam* ‘I scold; I curse’ < PMari *šudala- < PU *šođa- (UEW: 777)

Unexpected instances of full vowels

- E Ki V U *pižam*, Nw W *pižäm* ‘I grab; I get stuck’ < PMari *piže- < PU *pisi- (UEW: 732)
- E *wiste*, Nw *wištə* ‘spelt’ < PMari *wistə < PU *wešnə (UEW: 821)
- E Nw W *üps*, Ki *üφs* ‘smell’ < PMari *üps < PU *ipsi (UEW: 83–84; HPUL: 536)
- E Ki V U Nw W *müks* ‘bee’ < PMari *mükš < PU *mekši (UEW: 271; HPUL: 545)
- W *juž-wət* ‘sweat; fluid in a blister’ (< PMari *juž-; *wət* ‘water’) < PU *jiša ‘skin’ (UEW: 636–637; HPUL: 552)

A. Айкио. К реконструкции прамарийского вокализма.

На настоящий момент существуют две теории устройства вокалической системы в прамарийском языке, выдвинутые соответственно Э. Итконеном и Г. Беречки. В статье дается критический анализ обеих теорий, после чего автор предлагает собственную, детально аргументированную реконструкцию прамарийского вокализма первого слога. В частности, обосновывается необходимость реконструкции 11 отдельных фонем, в отличие от 13-фонемной системы Итконена и 7-фонемной системы Беречки.

Ключевые слова: уральские языки, марийский язык, праязыковая реконструкция, историческая фонетика.

A complete etymology-based hundred wordlist of Semitic updated: Items 75–100

The paper represents the fourth part of the author's etymological analysis of the Swadesh wordlist for Semitic languages (the first three parts having already appeared in Vols. 3, 5 and 7 of the same Journal). Twenty six more items are discussed and assigned Proto-Semitic reconstructions, with strong additional emphasis on suggested Afrasian (Afro-Asiatic) cognates.

Keywords: Semitic, Afrasian, etymology, glottochronology, lexicostatistics.

This study, the fourth portion of an etymologized 100-wordlist of Semitic, concludes the author's second attempt to compile a complete one hundred word list ("Swadesh list") for most Semitic languages, one that would thoroughly represent all branches, groups and subgroups of this linguistic family and provide the etymological background wherever possible. It is another step toward figuring out the taxonomy and building a detailed and comprehensive genetic tree of this family and, eventually, of the Afrasian (Afroasiatic) macro-family with all its branches on a lexicostatistical / glottochronological basis.

Several similar attempts, including those by the author (Mil. 2000, Mil. 2004, Mil. 2007, Mil. 2008, Mil. 2010, Mil. 2011, Mil. 2012), have been made since M. Swadesh introduced his method of glottochronology (Swadesh 1952; Swadesh 1955). In this paper, as well as my previous studies in genetic classification, I rely on Sergei Starostin's method of glottochronology and lexicostatistics (see Starostin 2000), which represents a radically improved and further elaborated version of the Swadesh method.

It should be noted that my main interest in working on these four portions has been in providing new and unrevealed or understudied Semitic etymologies and Afrasian parallels to the 100-wordlist items, rather than in adducing well-known etymologies (which are also given in a feasible measure) or in a strict selection of one most appropriate lexeme for each item, reducing synonymy to a minimum. The latter task will be addressed in the fifth portion which, besides addenda and corrigenda, is going to contain a final scoring of cognates to build up an updated and, hopefully, more refined lexicostatistical classification and chronology of the branching of Semitic. As for Afrasian parallels, my aim, again, was not to adduce the most detailed and complete data for all the items, but to demonstrate the Afrasian origin of as many Semitic words as possible, *inter alia* to eliminate the possibility of the latter to have been borrowed from non-Afrasian languages.

This study was carried out within the frame of two projects: "Bounds of the Afrasian linguistic macrofamily: chronology of branching to be checked against historical dates obtained by natural-science methods; genealogical tree; contacts with Non-Afrasian African languages" (project No. 12-06-00214, supported by the Russian Foundation for Sciences) and "Elaboration of a complete system of regular inter-Afrasian consonantal correspondences and a pilot reconstruction on its base of common Afrasian terms featuring the inner world of the Early Neolithic Man" (project No.12-04-00293, supported by the Russian Foundation for the Humani-

ties). Besides these foundations, my gratitude also goes to my colleagues and collaborators in different projects: Prof. O. Stolbova (with whom we work on the Afrasian Database within the “Evolution of Human Languages” project, wherefrom I draw most of the data) and Drs. L. Kogan and G. Starostin for consultations and discussions. I am also indebted to Dr. M. Bulakh for obtaining for me a 100-word list from a Tigre speaker and to L. Kogan for sharing with me the Soqotri list (namely, of the dialect spoken by a Bedouin tribe of Darho in the central part of Socotra) compiled by him during his expedition to Socotra in November 2010, which allowed me to rectify a few items and cancel several synonyms that, for lack of more precise data, I previously had to treat on par with the corresponding main term in my previous list. Additionally, this portion of the 100-wordlist includes terms from one more MSA language — Hōbyot, thanks to Prof. Robert Ratcliffe, who in 2013 edited the last work of late Prof. Aki'o Nakano (see Nak. 2013) and kindly sent it to me; I am planning to adduce items 1–74 of the Hōbyot list in the fifth portion as well as include the Hōbyot data into the final scoring of cognates.

The list below is based on the following main sources (not referred to in the text except for special cases): Akk. — CAD and AHw; Ugr. — DUL; Hbr. and Bib. — HALOT; Pho. — Tomb.; Pal. — Sok. JP; Syr. — Brock. and Sok. Syr.; Mnd. — DM; Urm. — Tser. and Sarg.; Qur. — Pen. and BK; Leb., Mlt. — native speakers, Mec. — Sat.; Sab. — SD and Stein; Gez. — LGz; Tna. — native speakers and Kane T; Tgr. — a native speaker and LH; Amh. — native speakers, Baet. and Kane A; Arg. — LArg; Gaf. — LGaf; Sod. and Cha. — native speakers and LGur; Har. — a native speaker and LHar; Wol. — LGur; Hob. — Nak. 2013; Hrs. — a native speaker and JH; Mhr. — native speakers, JM and Nak.; Jib. — native speakers, JJ and Nak.; Soq. — data collected by L. Kogan and LS.

The Data.

The data consist of items 75–100 of Swadesh’s 100-word list of 28 Semitic languages representing all groups within the family. Each item unites all synonyms differing by their origin, i.e. forming within the same item different entries preceded by an entry number in round brackets. Each entry, in its turn, consists of one or several cognate lexemes divided by a semicolon; the etymological comments including a Sem. protoform follow after a double slash; non-Semitic Afrasian parallels and a suggested Afrasian proto-form follow after the symbol □. Note that for cases when a choice of only one representative lexeme in the same language is random or unbiased, Starostin’s procedure provides for scoring several synonyms in this language; in this case, synonyms from the same language are present in two (and, in a couple of instances, even three) entries. Within each item there may occur two kinds of cases which are not scored at all (i.e. scored *zero*): borrowing and lack of a corresponding term in the available sources. Such cases form a separate section within the item preceded by the symbol ◊; in each item this section is completed with a Semitic proto-form(s) on the deepest level of reconstruction available (Proto-Semitic, Common West and South Semitic, etc.) in accordance with my genealogical classification of Semitic.

The following dates (some of them average or conventional, chosen after much hesitation and discussions with specialists in individual languages, and still liable to changes before a final scoring) have been attributed to the individual languages: Akkadian, 1450 BCE; Ugaritic, 1350 BCE; Hebrew, 650 BCE; Phoenician 850 BCE; Biblical Aramaic, 200 BCE; Palestinian Judaic, 200 CE; Syrian Aramaic, 200 CE; Mandaic, 750 CE; Urmian Aramaic, 1900 CE; Qur’anic Arabic, 600 CE; Lebanese Arabic, 2000 CE; Meccan Arabic, 2000 CE; Maltese Arabic, 2000 CE; Sabaic, 200 BCE; Geʿez, 500 CE; Tigray, 2000 CE; Tigre, 2000 CE; Amharic, 2000 CE; Argobba, 2000 CE; Gafat, 1900 CE; Soddo, 2000 CE; Harari, 2000 CE; Wolane, 2000 CE; Chaha, 2000 CE; Hobyot, 2000 CE; Harsusi, 2000 CE; Mehri, 2000 CE; Jibbali, 2000 CE; Soqotri, 2000 CE.

Abbreviations of languages, language periods and ancient sources.

A — Assyrian; Afras. — Afrasian (Afroasiatic, Semito-Hamitic); Akk. — Akkadian; Amh. — Amharic; Arb. — Arabic; Arg. — Argobba; Arm. — Aramaic; Ass. — Assyrian, B — Babylonian, BD — Book of the Dead; Brb. — Berber; Bib. — Biblical Aramaic; C. — Central; Cha. — Chaha; Chad. — Chadic; Clas. — Classical; CT — Coffin Texts; Cush. — Cushitic; Daṭ — Daṭīna Arabic; Dem. — Demotic; Doḫ. — Doḫar Arabic; Dyn. — Dynasty; E. — East; Egyp. — Egyptian; ESA — Epigraphic South Arabian; Eth. — Ethiopian; Frahang-i-Pahlavik (Aramaic); Gaf. — Gafat; Gez. — Geʿez; Gur. — Gurage; Har. — Harari; Ḥḫr — Ḥḫramawt Arabic; HEC — Highland East Cushitic; Hbr. — Hebrew; Hob. — Höbyot; Hrs. — Harsusi; Jib. — Jibbali; Jud. — Judaic Aramaic; LL = lexical lists; Leb. — Lebanese Arabic; LEC — Lowland East Cushitic; Mlt. — Maltese Arabic; Mec. — Meccan Arabic; Med. — Medical Texts; met. — metathesis; Mhr. — Mehri; MK — Middle Kingdom; Min. — Minean; Mnd. — Mandaic Aramaic; Mod. — Modern; MSA — Modern South Arabian; N — New, N. — North; NK — New Kingdom; O — Old; OK — Old Kingdom; Omot. — Omotic; Pal. — Palestinian Aramaic; pB. — post-Biblical; Pho. — Phoenician; Pyr. — Pyramid Texts; Qat. — Qatabanian; Qur. — Qur’anic Arabic; RT — Royal Tombs texts, S — Standard, S. — South; Sab. — Sabaic; Sem. — Semitic; Sod. — Soddo; Soq. — Soqotri; syn. — synonym; Syr. — Syrian Aramaic; Tna. — Tigriñña (= Tigray); Tgr. — Tigre; Ugr. — Ugaritic; Urm. — Urmian Neo-Aramaic; W. — West.; Wol. — Wolane, Y — Young.

Transcription and transliteration.

\dot{p} — bilabial emphatic voiceless stop; \dot{b} — bilabial emphatic voiced stop; \dot{b} — bilabial voiced fricative; \dot{t} — dental emphatic voiceless stop; \dot{d} — dental emphatic voiced stop; \dot{t} — voiceless interdental fricative (in Egyp., a conventional symbol most likely conveying \check{c}); \dot{d} — voiced interdental fricative (in Egyp., a conventional symbol most likely conveying \check{z}); c — alveolar voiceless affricate [ts]; z — alveolar voiced affricate [dz]; \check{c} — palato-alveolar voiceless affricate [tʃ]; \check{z} — palato-alveolar voiced affricate [dʒ]; \mathfrak{s} — hissing emphatic voiceless fricative; \mathfrak{c} — emphatic voiceless affricate; \mathfrak{z} — conventionally stands for what was likely \dot{d} , emphatic voiced interdental fricative, or \dot{t} , emphatic voiceless interdental fricative; \check{c} — palato-alveolar emphatic affricate; \hat{s} — lateral voiceless fricative (denoted by \hat{s}_x in Sem. reconstructed proto-forms); \hat{c} — lateral voiceless affricate; $\hat{\mathfrak{s}}$ — lateral voiceless emphatic fricative; $\hat{\mathfrak{c}}$ — lateral voiceless emphatic affricate; \hat{z} — lateral voiced emphatic sibilant (or perhaps affricate); \hat{z} — lateral voiced sibilant; \bar{g} — voiced velar fricative (in Brb.), k or q — emphatic velar stop; q — hypothetical velar affricate [kʰ] (only in reconstructed Afrasian proto-forms); γ — uvular voiced fricative (Arabic “ghain”); h — uvular voiceless fricative; \underline{h} — presumably velar voiceless fricative (only in Egyptian); \mathfrak{h} — pharyngeal voiceless fricative; $\mathfrak{ʕ}$ — pharyngeal stop (“ayin”), h — laryngeal voiceless fricative; $ʔ$ — glottal stop (“aleph”, “hamza”), y — palatal resonant; \mathfrak{z} and \mathfrak{i} — conventional transcription symbols accepted in Egyptology.

Conventions for reconstructed protoforms.

- V renders a non-specified vowel, e.g. $*bVr-$ should be read “either $*bar-$, or $*bir-$, or $*bur-$ ”.
- H renders a non-specified laryngeal or pharyngeal; S, Z render a non-specified sibilant.
- $/$ when separates two symbols means “or”, e.g. $*?i/abar-$ should be read “either $*?ibar-$ or $*?abar-$ ”.
- () a symbol in round brackets means “with or without this symbol”, e.g. $*ba(w)r-$ should be read “ $*bawr-$ or $*bar-$ ”.
- \sim means “and” pointing to two or more co-existing proto-forms.

75. SKIN

- (1) Akk. *mašk-*; Pal. *mšk*; Syr. *mešk-*; Mnd. *mišk* // < Sem. **mašk-* ‘skin’: Hbr. *māšāk* ‘leather pouch’, Arb. *mask-* ‘peau ôtée récemment d’un agneau ou d’un chevreau’ (SED 1 #190), Har. *miskāt* ‘buttocks’, Sel., Wol., Zway *maskāt* ‘back of body; clitoris, female genital organs’ (compared *ibid.* with a question mark).
- < Afras. **mask-* (ADB): Egyp. (Pyr.) *mskʔ* ‘skin, hide’ (*ʔ* is not quite clear: probably conveys the *a* vowel by means of graphic metathesis). The comparison can be tentatively extended, on the one hand, to Brb.: Zenaga *a-mašk* (*-k* < **-k*) ‘haut de la cuisse, cul’; C.Chad.: Logone *māska*, Kuseri *msāka* ‘leg’; Cush.¹ E.: Somali *mísig*, pl. *mískó* ‘meat of the hip’ (< **misk-*), S.: Burunge *mičoko* ‘calf of leg’ (if *-č-* here continues Afras. **-s-*); N.Omot.: Gamu, Dorze *miskata* ‘buttocks’² and, on the other, to Brb. (met.) **kasum-* ‘meat’: Ghadames *a-ksəm*, Audjila *ksûm*, Qabyle *a-ksum*, etc. (likely also Canarian: Gran Canaria *ta-quasem* ‘pig’)³. In Proto-Afrasian, the probability of simple homonymy of triconsonantal roots with the same consonantal composition and compatible meanings being extremely low, all the quoted lexemes are likely to go back to a single Afrasian noun meaning something like ‘(meat and hide/skin of) hip, calf of leg (of an animal)’; borrowings within the Eth.-Sem., Cush. and N.Omot. area are, of course, possible.
- (2) Ugr. *yr*; Hbr. *ŷôr*; Pho. *ŷr* // < Sem. **ya/ur-/*yaryar-* ‘(loose) skin, dewlap’ (SED I #105): Arb. *yaryarat-*, *yuryurat-* ‘gésier d’oiseau’, etc.
- < Afras. **yur-* (highly tentative): Egyp. (OK) *hʔ.w* (if — *ʔ* renders **-r*, and *h-* corresponds to Sem. **γ-* before **r*) ‘skin’; Chad. C.**yar-*: Dghwede *hára* ‘to skin’, Gisiga *gar-ak* ‘skin’, E. **gVr-*: Somrai *gàré*, Ndam *gərə* ‘skin’, Mokilko *gúnḡurú* ‘egg-shell, bark’; E. Cush. **ḡVr-/*ḡur-* (hypothetical, since reflexes of **γ* in Cush. are still obscure) ‘skin’: LEC: Elmolo *úrat* (**ḡ-* > Elmolo *θ* is regular), Dullay: Tsamay *ḡúuro* (ADB).
- (3) Syr. *geld-* (syn.), Urm. *gild-*; Qur. *žild-*; Leb. *žild*; Mec. *žild*; Mlt. *žilda* // < Sem. **gilald-* (SED I #78; Hob. *žóód*, pl. *žiiéléé*, Hrs. *gōd*, Mhr. *gēd*, Jib. *gód*, Soq. *gad* are probably Arabisms).
- (4) Gez. *māʔ/ʕs* // < Sem. **maʕaš-* or **maʕas-*: Tna. *maʕsi*, Tgr. *māʕas*, Amh. *mas* ‘skin, leather’, Arb. *mʕs* ‘rub (leather) vigorously’ (LGz 324; E.Cush.: Saho *maʕas* ‘tan hide’, quoted *ibid.* is an Arabism).
- (5) Tna. *ḡʔərbät*; Tgr. *ḡərbät*; Amh. *ḡʔərbät* // < Sem. **ḡʔirb-at-*: Gez. *ḡʔərbābit* ‘leather bag’ (LGz 440: “probably reconstructed from Amh *ḡʔərbabičča*”), Arb. *ḡirbat-* ‘grande outre à lait ou à eau faite d’une seule peau cousue au milieu’ (BK 2 704)
- < Afras. **ḡʔirab-* ‘bark, skin’: Cush. N.: Beja *kurbe* ‘skin’ (or < Eth.?), C.: Qwara, Kemant *kʔərbäy* ‘skin’, Khamta *qärvoì*, Khamtanga *qärbi/qäbri* ‘bark’; N. Omot.: Male *ḡurubi* ‘bark’ (ADB).
- (6) Amh. *ḡodä* (syn.) // Arg. *ḡoda*, *oda*, Gaf. *ḡodä*, Eža, Muher *ḡʔäda* ‘skin of body’, Arb. *ḡadd-* ‘lamb skin’, *ḡidd-* ‘strip of skin; skin of body’.
- < Afras. **ḡʔad-*: W. Chad. **ḡad-*: Tangale *kada*, Pero *kádà*, Kirfi *kàdà* ‘skin’, Jimi *káda* ‘leather loin cloth’ (St. 2011 #315)⁴.
- (7) Mhr. *bāšarēt*; Jib. *bāšarét* // < Sem. **bi/ašar-* ‘flesh, (human) body’ (SED I #41).

¹ N. Cush.: Beja *mesík* ‘skin, hide’ is an Arabism.

² Data quoted after EDE III 599, where a comparison is hypothetically, but still inconclusively, made with Egyp. *msd.t* ‘haunch’ (not with *mskʔ*).

³ Cf. an identical semantic link between ‘skin’ and ‘body’: Sem. **bišr-/*bašar-* ‘meat’ vs. Arb. *bašar-* ‘(human) skin’ (SED I #41).

⁴ An Arabism in W.Chad. is less likely because of the vocalism *-a-* vs. *-i-* in Arb. *ḡidd-* ‘skin of body’, whereas Arb. *ḡadd-* ‘lamb skin’ does not quite fit as a source word for semantic reasons.

- (8) Soq. *šírhi* // < Sem. **šir-* ‘flesh’ (SED I #238).
 ◇ Gaf. *g^wagä*, Sod. *goga*, Har. *gōga*, Wol. *gogä* and Cha. *goga* are most likely < Oromo *googa*. No term registered in Bib. and Sab.
 → Common North and West Semitic **mašk-* (#1).

76. SLEEP

- (1) Akk. *šalālu* // also ‘to fall asleep, be at rest; remain inactive; sleep with a woman’. Most likely < Sem. **šll* ‘to lie or go down, be in a horizontal position’: Gez. *šll* ‘float upon, come to the surface, swim’, Tna. *šälälä* ‘to float’, Amh. *čällälä* ‘spread (intr.) over a surface (butter or oil)’, Gur. *čälläll* ‘float, soar’ (the Akk. and Eth. forms compared in LGz 555), Hbr. *šll* (hapax) ‘to sink’ (HALOT 1027), Syr. *šll* ‘se inclinavit, inclinatus est; flexit, disposuit; insidiatus est’ (Brock. 628)⁵. See also No. 47 ‘lie’ (1) in Mil. 2011.
- (2) Akk. (syn.) *itūl-* (< **nyl* : *nālu*, *niālu*, *i/utūlu* ‘to lie (down)’ // likely met. < Sem. **lyn*: Ugr. *ln* ‘spend the night’, Hbr. *lyn* ‘spend the night, stay overnight’, etc. (HALOT 529), Arb. *laynat-* ‘coussinet’ (BK 2 1051; unless from the homonymous *lyn* ‘ê. tendre, mou’). Same as No. 47 ‘lie’ (2) in Mil. 2011.
- (3) Ugr. *yšn*; Hbr. *yšn* // < Sem. **šin(-at)-* ‘sleep, dream’ (cf. SED I Verb #182).
 □ < Afras. **wisan-* ~ **sawin-* ‘sleep, dream’: Chad. W.: Bole *ʔi-súnná*, Karekare *súunà*, Kirfi *súnnà*, Guruntum *suní* ‘to dream’, Warji *ussanna* ‘to sleep’, Ngizim *sáunú* ‘to dream’, C.: Musgu *wéseŋ*, Kotoko *wisàn* ‘to sleep’, *súwáné* ‘to dream’, E.: Kabalai *surwán*, Kera *sócné*, Jegu *suun-* ‘to dream’, etc.; E.Cush.: Saho, Afar *sonoo* ‘dream’ (n.), Dullay: Dolpena *šann-*, Dihina *šenn-* ‘to lie, repose’ (ADB).
- (4) Pal. *dmk*; Syr. *dmk*; Urm. *dmk* // No visible etymology unless somehow linked to Soq. *deme* < Sem. **dVm-/*dVwVm-* (v. below 13) with an unexplainable *-k*.
- (5) Pal. (syn.) *nwm*; Qur. *nwm*; Leb. *nēm*; Mec. *nām*; Gez. *nwm* // < Sem. **nwm* ‘to sleep, slumber’ (SED I Verb #52).
 □ With parallels in Afras. (ADB): Eyp. (Pyr.) *mm.t* (< *mV-nVm-t*) ‘bed’, (Amarna) *nmḫ* ‘to sleep’ (-ḫ, since it is not clear, is to be treated as a root extension); E.Chad.: Mokilko *mòné* ‘to sleep’ (met.?).
- (6) Syr. *škb* (syn.), Mnd. *škb*; Tgr. *sākāb hälla* // < Sem. **škb* ‘lie (down), sleep’. See No. 47 ‘lie’ (3) in Mil. 2011.
- (7) Urm. (syn.) *ḫlḫ* // Syr. *ḫūlāḫ-* ‘somnus altus’. Cf. Arb. *ḫly* ‘ê. faible ou fatigué’ (BK 2 99).
- (8) Mlt. *raʔat* // < Arb. *rḫd* ‘to sleep’. No visible Sem. parallels.
- (9) Tna. *dākḫäsä* // Gez. *daḫḫasa* ‘be sleepy, fall asleep, slumber’ (LGz 140); acc. to Leslau, related to Arb. *dukās-* ‘sleepiness’ (I prefer to qualify such cases as “variant roots”). No Sem. parallels.
- (10) Amh. *tāññä*; Arg. *teñña*, *eñä*; Sod. *aññäm*; Har. *ñēḫa*; Wol. *aññe*; Cha. *nəyäm* // < Sem.: Eth. **nhy*: Gez. *nəhya* ‘to repose, recover; be quiet’ (LGz 394), Ugr. *nhmmt* (*nhm* + *mt*) ‘drowsiness, fainting fit’ or ‘deep sleep (of death)’ (DUL 626); cf. with metathesis Arb. *hwn* ‘se reposer, se calmer, s’endormir’ (BK 2 1460). Same as No. 47 ‘lie’ (13) in Mil. 2011
 □ < Afras. **nhy* ~ **hwn*: Chad. W.: Sha, Kulere *nya*, Daffo-Butura *nyah* ‘to sleep’, C. **hwVn-* ~ **hany-*: Gabin *ʔyenè*, Kilba *hány*, Margi *ʔanyi*, Gudu *wiyènù* ‘sleep (n.)’, etc., E. **waHVn-*: Kwang *wéŋ* ‘to sleep’, Migama *wáànò* ‘to dream’ (ADB).
- (11) Gaf. *(tä)gäddälä*; Cha. *(tä)gätäräm* // Related to Arb. *ždl* ‘jeter, renverser par terre’ (BK 1 265), same as No. 47 ‘lie’ (14) in Mil. 2011.

⁵ There are two less plausible etymologies for Akk. *šalālu*: < **ḫall-* ‘shadow’ (> ‘to lie in the shadow’), cf. Gez. *šalālot* ‘shadow, shade’, *ʔašlala* ‘to seek shade, lie down in the shade, take a rest, etc.’ (LGz 555); and < **šll*, cf. Sab. *qll* ‘to fall sick (of an epidemic)’ (SD 41).

- (12) Hob. *šu-ukóof*; Hrs. *še-wkōf*; Mhr. *š-ūkūf*; Jib. *šef* // < Sem. **wkʷ* ~ **kʷ* ‘to lie, sit’, see No. 74 ‘sit’ (5) in Mil. 2012.
- (13) Soq. *déme* // < Sem. **d(w)m* ‘to dwell, last, stay a long time in one place’: Ugr. *dm* ‘to remain still’ (DUL 272), Hbr. *dāmā* ‘to be silent, still’ (HALOT 225), Arb. *dwm* ‘durer, continuer d’exister’ (BK 1 752), Mhr. *adyīm*, Hrs. *adīm* ‘stay a long time in one place’ (JM 75). Cf. DRS 236 #1, where several apparently unrelated roots are united under one entry.
- < Afras. **dam-* ‘to dwell, stay, rest’: Eg. *dmʔ* ‘lie’ (if < **dVmʔ-*, with -ʔ metathetically conveying -a-, cf. *mskʔ* in No. 75 ‘skin’ #1); Chad. W. **di/am-* ‘live, sit’: Geji *dām-ti* ‘live, sit’, Boghom *tam* ‘sit’, *tām-ún* ‘live’, Zakshi *dīm*, Guus *dám* ‘sit’, E.: Somrai *dāma* ‘sit’, Sibine *dām* ‘dwell’; S. Cush.: Iraqw *dam-* ‘wait’ (ADB).
- (14) Soq. (syn.) *henod* // < Sem **hnd* ‘to be languid, slack, drowsy’: Mhr. *hōnad* ‘to feel drowsy’ (JM 158), Jib. *ohúnd* id. (JJ 97), Arb. *hnd* II ‘n’avoir pas de force à faire quelque chose, être très lent à faire quelque chose’ (BK 2 1452), Tgr. *hanad gāʔa, tāhannādā* ‘träge, schlaff, betäubt sein’ (LH 19; not quite adequately rendered into English as ‘to be lazy, amazed’, rather to be translated as ‘to be languid, slack; be in stupor, stupefied’; *gāʔa* is a truncated form of *gābʔa* ‘to be, become’ ibid. 585).
- ◇ No terms registered in Pho., Bib. and Sab.
- Common West Semitic **nwm* (#5).

77. SMALL

- (1) Akk. *šeḫru*; Pho. *šr* ; Qur. *ṣayīr-*; Leb. *zḡr* (< **šyr*); Mec. *ṣayīr*; Mlt. *zā(gh)ir*; Sab. *šyr* // < Sem. **ṣayīr-* (Kog. DD): Ugr. *šyr-* ‘young’, Hbr. *šāīr* ‘the smaller one, little’ (HALOT 1041), Syr. *šāīr-* ‘contemptus, turpis’ (Brock. 634).
- (2) Akk. (syn.) *daḫḫ-*; Gez. *daḫiḫ* // < Sem. **da/iḫḫ-* ~ **daḫiḫ-* ‘small, tiny, fine’: Ugr. *dḫ* ‘tiny, fine; weak’ (DUL 276–7), Arb. *diḫḫ-* ‘menu, fin, etc.’ (BK 1 715), Amh. *dāḫḫākā* ‘be fine, minute’ (LGz 140), Soq. *dḫḫ* ‘devenir mince’ (LS 133)⁶.
- Cf. Afras.: W. Chad.: Boghom *kō-dók*, Dwot *nḏak* ‘short’; E.Cush. **d/ṭi/ukḫ-*: LEC: Somali *dīq-* ‘to become faint, tenuous’, Oromo *diqqa* ‘small’, HEC: Hadiya *ṭuḫḫa*, Burji *ḫiḫḫo* ‘narrow’ (ADB).
- (3) Hbr. *qāṭān*; Mnd. *qotān-*; Sab. *ḫṭn* ‘small’ // < Sem. **ḫatan-* (Kog. DD): Akk. *ḫatnu* ‘thin, fine, narrow’ (CAD q 173), Syr. *ḫṭn* ‘grow thin’, Arb. *ḫatīn-* ‘thin iron of a spear’, Gez. *ḫṭn* ‘be thin, fine, lean, etc.’ (LGz 453), Mhr. *ḫáyṭen*, Jib. *ḫéṭen* ‘become thin’, Soq. *ḫáṭḫon* ‘thin’ (JM 245).
- with extension -n (see Mil. RE) < Afras. **ḫ(w)Vt-* ‘small, thin: Chad **k/ḫ(w)Vt/ṭ-*: W.: Ngas *kwiit* ‘small, narrow’, Tangale *kate* ‘diminish, decrease, reduce, abate’, C.: Tera *katà* ‘decrease’, Mafa *kwitee* ‘small’, Zime-Batna *kwétété* ‘étroit’ (St. 2011 #110; compared with Agaw); Cush.C. **ṭaq^wt-* ‘be small, few’: Bilin *ṭaḫ^wt-aw* ‘few’ (in App CDA 66 compared with E.Cush. with metathesis — see #2, but not with Sem. **ḫatan-*), E.: Dullay: Harso *tiḫḫ-assa* (rather metathesis from **ḫit-* than related to E.Cush. **d/ṭi/ukḫ-* in #2, with *t-* hard to explain) ‘small’ (ADB).
- (4) Bib. *zāfēr*; Pal. *zḡwr*, *z^ʿḡ(y)r*; Syr. *zāfōr-*; Urm. *zāfōr-* // < Sem. **zḡr* ‘be small, thin’: Hbr. *zāīr* ‘a little’ (< Aram? HALOT 276), Arb. *zḡr* ‘ê. clairsemé (se dit des plumes, du poil)’ (BK 1 990).

⁶ More than likely related to Sem. **dkḫ* ‘crush, pound’ (LGz. 140) and further to Afras **dVḫ-* ‘to pound, crush’: Brb.: Qabyle *duqq* ‘strike, pound’; Chad. W.: Hausa *dán-dàkā* ‘pound, beat’, Polchi *ḏok*, Ngizim *ḏakú* ‘pound’, E.: Mawa *ḏuku-ḡ* ‘hit (a person), thrash millet’; S. Cush.: Dahalo *ḏuk-ud-* ‘destroy’ (ADB). In this case, the Proto-Afrasian term must have at one point become bisemantic, meaning both ‘to pound, crush’ and ‘small, tiny, fine’, eventually evolving from ‘to pound, crush finely’.

- < Afras.: Chad. *zVr- ‘to be narrow’ (St. 2009 #563; I would add ‘thin, lean’): W.: Hausa *zúirúrii* ‘narrowness’, Duwai *zàràwo* ‘reduce’, C.: Gude *zár* ‘long and thin’, Podoko *záraka* ‘lose weight’, Mafa *n-zaraʔa* ‘thin’, E.: Migama *zíráw* ‘lose flesh’, East Dangla *zèr-zirē* ‘reduce’.
- (5) Mnd. *zuṭ-* // < Aram. *zṭ (DRS 769). No parallels.
- (6) Gez *nəʔus*; Tna *nuʔus*; Tgr. *nəʔuš*; Arg. *mans, mass*; Gaf. *ənsä*; Sod. *maläs*; Wol. (y)*anäsä*; Cha. (syn.) *ərs* // < Sem. *ʔVnVš- ‘be weak, sick, small’: Akk. *enēšu* ‘be(come) weak’ (CDA 73), Soq. *ʔenes* ‘ê. petit’ (DRS 26).
- (7) Amh. *ṭannəš* // Gur. **ṭänäsä* ‘to shrink because of a disease or lack of food, lose strength, be extinguished (glowing wood)’ (LGur 625); not compared with the Amh. term in spite of an obvious phonetic and semantic similarity and commented upon as “represents *ṭällāsā* (with the same meaning) with *l:n*”, compared, in its turn with Amh. *ṭällāsā* ‘die out (fire)’ with the comment “related to Tgr. *ṭārsa* ‘emaciate, become weak’” (rather variant roots with *-n-/l-/r*)⁷.
- (8) Har. *ṭit*; Wol. (syn.) *ṭit* // According to LGur 635, < **ṭəʔit* < **ṭəḳit* < S. Eth.: Amh. *ṭəḳit* ‘few, a little’⁸.
- (9) Cha. *ḳəsi* // Acc. to LGur 504, < S.Eth. *ḳəssən* id., related to Arb. *ḳsm* ‘divide’ (ibid. 505).
- (10) Hob. *ḳanuún*; Hrs. *ḳennōn* // Mhr. *ḳənnēt* ‘child, little, young’ (JM 232), Soq. *ḳéyhen* ‘petit’ (LS 478), Arb. *ḳunḳuḳ* ‘petit’ (BK 2 825; an uncommon form: partial reduplication with the extension -ʔ?).
- < Afras. **ḳayn-*: Chad. W.: Hausa *ḳàḳánè* ‘small, little’, Mburku *kíyìn*, C.: Mwulien *kiéŋ*, Lame *kʷáníyāù*, E.: Kwang *končé*, Jegu *kóyán* ‘small’ (compared with the MSA terms in ND 1088).
- (11) Jib. *nišan* // No etymology that I could find.
- Common North and West Semitic. **ṣayīr-* (#1) and **dal/iḳḳ-* ~ **daḳiḳ-* (#2).

78. SMOKE

- (1) Akk. *ḳutr-*; Ugr. *ḳṭr*; Mnd. *guṭr-* // < Sem. **ḳuṭār-*: Ebl. /*ḳuṭturū*/, Hbr. *ḳəṭōrāt* ‘incense’ (> Eg NK *ḳdr.t*), *ḳṭr* (pi) ‘make a sacrifice, go up in smoke’ (HALOT 1094), Sab. *m-ḳṭr* ‘incense-altar’ (SD 109), Gez. *ḳəttār(e)* ‘incense, fumigation’, *ḳṭr* ‘fumigate, give off an odor’, Tgr. *ḳətare* ‘fragrance, spice’, Amh. *ḳəttārä* ‘burn incense in church’, *ḳəttärä* ‘bathe in steam or in incense smoke’ (LGz 452).
- If < **ḳuṭ-r*, with the root extension *-r*, cf. either 1) Chad. **ḳut-* (*-t* may convey Afras. **-ṭ*): W.: Sura *kwut* ‘wind’, Chip *kuṭ*, Tangale *kudo*, ‘harmattan’, Burma *kuṭ-küt*, Mangas *ku-kut* ‘dust’ (note also Jimi *kwatir* id., with *-r!*), Zul *kwutu-kwütù* ‘fog’, C.: Gabin *kuṭkuṭ* ‘dust’, Zime *küt* ‘fog’, Lame *kwátú*, Masa, Banana *kuṭ-na* ‘harmattan; fog’, Mesme *kuḍ* ‘fog’ (St. 2001 #103) or 2) Chad. **k/ḳut/ṭ-*: W.: Karekare *kuṭò*, Ngamo *kütò* ‘ashes’, E.: Mawa *koto* ‘cold ash’ (from St. 2011 #113a, with a different interpretation). Unrelated if the Chad. forms are akin to E.Cush. **ḳut-* ‘smoke’ (see footnote 14).
- (2) Hbr. *ṣāšān*, Soq. *ṣeto* (with a loss of *-n?*) // < Sem. **ṣVṭan-⁹*: Arb. *ṣaṭān-* ‘fumée’, *ṣuṭān-* ‘fumée; poussière’ BK 2 172.

⁷ Cf. also E.Cush: LEC: Oromo *ṭinnā* ‘small’, however, without any signs of *-š/s*.

⁸ LGur 635 also compares Gez. *ṭəḳ(k)*, translating its main meaning as ‘minuteness’. However, the meanings quoted for *ṭəḳka* in LGz 596 do not confirm this. Better comparable with S.Eth. **ṭəḳ-it* ‘little’ is Gez. *ṭəḳawa* ‘beat, pound’ (LGz 596); for the meaning shift see footnote 6. As a whole, Eth. **ṭḳ(w)* looks like a late variant root of Sem. **dḳḳ* ‘crush, pound’.

⁹ Soq. *ṣeto* can be alternately compared with Arb. *ḡayṭ-* ‘pluie abondante et qui s’étend aux environs’ (BK 2 522), phonetically immaculate, though semantically less so.

- Compared in Tak. **h* 256 with various Afras. forms meaning ‘smell’ and ‘nose’ going back to Afras. **sVn*- ‘smell’ and **sa/in*- ‘nose’.
- (3) Pal. **n**nan*; Syr. *tān*-; Urm. *tin*-; Gez. *tann*; Tgr. *tānan*; Arg. *tan*, Gaf. *tännä*; Sod. *tān*; Har. *tān*; Wol. *tān*; Cha. *tān* // < Sem. (Arm.-Eth.) **tan*¹⁰.
- Cf. isolated W.Chad.: Pero *tuuṅa* ‘smoke’ (ADB).
- (4) Qur. *duḥān*-; Leb. *dhan*; Mec. *duḥḥān*; Mlt. *duḥḥān*; Hob. *ndóoh*; Hrs. *nedēh*; Mhr. *nidiḥ*; Jib. *māndoh* (met. in all MSA) // A controversial case: acc. to HALOT 218, Sem. terms for ‘millet’ (Hbr. *dōḥan*, Syr. *duḥn*-, Arb. *duḥn*-, Akk. *duḥn*- < Sem. **duḥn*-) actually mean ‘smoke-colored’¹¹, in which case color may be a primary meaning, cf. Arb. *duḥnat*- ‘couleur foncée’, *dahnān*- ‘sombre, couvert, nuageux’ and Hrs. *edḥān* ‘couleur foncée, pourpre’ (DRS 250), likely an Arabism. However, in view of the triconsonantal roots obviously related in Arb. and MSA (while metathesis speaks against borrowing into MSA from Arb.), the primary meaning ‘smoke’ seems more likely¹².
- (5) Gez. (syn.) *ṭis*; Amh. *ṣas*; Arg. *ṣis* (syn.) // Only Eth. **ṭis* according to LGz. 601, where Brockelmann’s opinion of a loan from C.Cush. (Aungi *tiša*, Bilin *tida*, Kemant *taḥsā*) is quoted; however, the case is much more intricate¹³. In view of Arb. *ṭassān*- ‘poussière soulevée et qui voltige et tourbillonne dans l’air’ (BK 2 81), likely related with the Eth. forms (for the meaning shift see Chad. above in #1), the Aungi and Kunfal (*ṭiši*) terms are Ethiopisms.
- (6) Tna. *takk-i* // Tigre *tāka* ‘put on fire’. No Semitic parallels¹⁴.
- < Afras. **tVk*- ‘burn, put on fire’: Eyp. (RT) *tkʔ* ‘burn’ (< **tak*-, *ʔ* conveying *-a-* with graphic metathesis? Or < **tVkVr/l-* with a root extension?); Chad. W.: Hausa *tōká* ‘cold ash’, Pero *tókkò* ‘put on fire’, C.: Matakam *tákwá-r* ‘burn’, E.: Dangla *tuki-nà*, Bidiya *tokí-nà* ‘stove’; Cush. C.: Bilin *tāk^wād*- ‘burn’ (< **tak^w-ad*-?)
- ◇ No term registered in Pho., Bib. and Sab.
- Common North and West Semitic **kuṭr*- (#1).
- Common South and West Semitic (Arabic) **duḥān*- ~ **nidVh*- (#4).

79. STAND

- (1) Akk. *u/izuzz-* // Probably cognate with Arb. *wzy/yzy* ‘è. rassemblé, aggloméré sur un seul point’, III ‘è. placé dû côte opposé, vis-à-vis d’un autre’, IV ‘appuyer, p. ex. le dos contre quelque chose’, *mustawz-* ‘dressé, qui est debout (homme, object)’ (BK 2 1532)¹⁵.

¹⁰ In LGz 577, the Eth. root is compared with the Arm. one, though Brockelmann’s suggestion of the Arm. forms coming from **ʕn* (see 7 below) is regarded as an alternative, while in Tak. **h* 256, the latter, much weaker, if not completely wrong, etymology is quoted as the only one, the Eth. forms left without mention.

¹¹ In DRS 250, ‘millet’ and ‘smoke’ are quoted in different entries, though with the comment: “Le nom de la plante s’expliquerait par sa couleur”.

¹² It is somewhat dubious whether ‘smoke’ could have transferred its color to ‘millet’ without leaving any traces of the original meaning (or at least of a similar color) in those languages where **duḥn*- ‘millet’ is attested.

¹³ Bilin *tāda*, Khamir *ṭiyā* < North Agaw **ṭiz*-, Aungi *tiši*, Kunfal *ṭiši* < South Agaw **ṭiš*-. The forms reconstructed for North and South Agaw do not correspond to each other regularly (it is *-s* rather than *-š* that corresponds to NAgaw **z* according to correspondences in App. CDA 13), which would rather point to a borrowing into Agaw. As for Kemant *taḥza* and Falasha *tāksa*, those forms seem to belong to a different root (according to Appleyard, *-h*- remains unexplained if the forms are indeed related).

¹⁴ Coincides with Saho *tika* ‘smoke’ with no direct cognates either, unless the latter is related through metathesis with E.Cush. **kut*- (in which case it should be considered a loan in Tna.). LEC: Mashile *kutayt* (according to Black, also *tutayt*), Dullay: Gawwada, Gollango, Gobeze, Harso *kuute* (ADB).

¹⁵ In Kog. Ug. 2 #17, Akk. *u/izuzz-* is compared (after Poebel and others) with Ugr. *n-dd* ‘to stand’, pB Hbr. *zww* ‘to move, go away, depart’ (acc. to Kog. *ibid.*, “well compatible semantically”, which to me does not look so), im-

- (2) Ugr. *ḵm* ('stand up'); Bib. *ḵwm*; Pal. *ḵwm*; Syr. *ḵwm*; Mnd. *ḵwm*; Qur. *ḵwm*; Sab. *ḵwm*; Gez. *ḵwm*; Tna. *ḵomä* (the other term, *däw* in *däw bälä*, is a loanword from Bilin *däw y-* or Saho *dau* id.); Amh. *ḵomä*; Arg. *ḵoma, oma*; Gaf. *ḵomä*; Sod. *ḵomäm*; Cha. *ḵämäm* // < Sem. **ḵwm* 'stand (up), rise': Hbr. *ḵwm* 'to rise, get up, stand up' (HALOT 1086).
- < Afras. **ḵaw/ym-* 'stand up, rise': Berb. **ḵaym* 'to sit' (with a meaning shift, perhaps primarily 'to sit up, rise to a sitting position'): Ghadames, Siwa, Rif *qim*, Ahaggar, Taneslemt *ḵaym*, Zenaga *imi* (< **iḵmi* < **iḵmi*), etc.; W.Chad: Geji *kàmi*, Buli *kəmu* 'lift' (ADB).
- (3) Hbr. *ʕmd* // < Sem. **ʕmd*: Akk. *emēd-* 'lean against, impose', Arb. *ʕmd* 'to support; intend'.
- Cf. W.Chad. **Hind/d-* (with assimilation of *-n-* to the dental?) 'stand up': Karekare, Ngamo *hind-*, Bele *hindú-kò*, etc. (ADB).
- (4) Urm. *kl?* // No etymology, suspicious as a borrowing.
- (5) Qur. *wḵf* (syn.); Lbn. *wəʔəf*; Mec. *wagaf*; Mlt. *wiʔf* // < Sem.: Sab. *wḵf* 'attach, fix', *mḵf* 'base' (SD 161), Mhr. *awōḵəf* (JM 427), Jib. *ōḵuf* 'to stop doing st.' (JJ 291)¹⁶.
- (6) Tgr. *bātra* // Looks like a singular semantic development from 'to be arrogant, haughty': Amh. *tānbāṭārrārā* 's'avancer avec une démarche orgueilleuse', Arb. *bṭr* 'être arrogant' (DRS 61).
- (7) Har. *ḵānāna*; Wol. *ḵanānā* // Arb. *ḵnn* VIII 'se reposer', X 'faire halte avec ses troupeaux pour en traire les femelle et boire le lait' BK 2 817 (In LGur 488, Arb. *ḵanna, ʔiḵnānna* 'stand up' are quoted).
- (8) Hob. *šoor*; Hrs. *šōr*; Mhr. *šōr*; Jib. *šōr* // < Sem. **šw/yr-*: Sab. *šrr* 'to stay, stand' (SD 147), Arb. *šyr* II 'arrêter (p.ex., l'eau) dans son cours' (BK 2 1390). Perhaps related, with a meaning shift, are Soq. *šwr* 'carry', Gez. *šwr* 'to bear, carry, support (old parents), bear up' (LGz 567), Tgr. *šora* 'carry, support', Tna. *šorä* 'carry', Amh. *ṭorä*, Gur. *ṭorä* 'carry, support'.
- If the primary meaning is 'to stand' rather than 'carry, support', related to Chad. W. **çVr-* 'to stand': Hausa *çáyà*, Pa'a *ciràà* 'to stand', Warji *çər-w-*, Miya, Kariya, Mburku, Cagu *šər-*, Siri *còuru* 'stand, stop, wait', Tangale *seri* 'stand (up), stop', Geruma *sora*, Daffa-Butura *soor* 'stand' (St. 2009 #772).
- (9) Soq. *ḵəšaʕ* // Cf. Arb. *ḵšʕ* II 'ne pas quitter la maison' (BK 2 753).
- ◇ No term registered in Pho.
- Common West Semitic **ḵwm* (#2).

80. STAR

- (1) Akk. *kakkab-*; Ugr. *kbkb*; Hbr. *kōkāb*; Pho. *kkb*; Pal. *kebkāb*; Syr. *kaḵkāb-*; Mnd. *kukb-*; Urm. *kāḵkāb-*; Qur. *kaḵkāb-*; Sab. *kwkb*; Gez. *kokāb*; Tna. *ḵ^wakob*; Tgr. *kokāb*; Amh. *kokāb*; Arg. *kokāb*; Gaf. *kokobä*; Sod. *kokāb*; Cha. *ḵ^wāk^wab*; Hob. *kābkīb*; Hrs. *kebkāb*; Mhr. *kebkib*; Jib. *kābkēb*; Soq. *kibšib* // < Sem. **kabkib-*, perhaps derived with the original meaning 'a round object' < Sem. **kVbb-*, *kVbkVb-* '(to be) round': Hbr. *kirkeb* 'encircle', Arb. *kbb* 'form into a ball', *kubbat-* 'ball, ball of thread, lump', *kābkāb-* 'ball of thread', Gez. *kābb, kābub* 'round', etc. (ADB).
- (2) Qur. (syn.) *nažm-*; Lbn. *nižmi*; Mec. *nažma* // No etymology that I could find.
- (3) Har. *ṭūy* // No etymology.
- ◇ Mlt. *stilla* < Italian; Wol. *kālbāžo* (see LGur for a source of borrowing). No term registered in Bib.
- Common Semitic **kabkib-* (#1).

plying Sem. **(y/w)dd* as an underlying root. At the same time, Ugr. *ndd* 'to go, move; disappear, be extinguish; prepare, hurry, rush, launch oneself' (DUL 620) is by far better compatible both semantically and phonetically with Hbr. *ndd* 'to flee, escape', Jud. Arm. *ndd*, Arb. *ndd* 'to flee', etc. (HALOT 672), going back to Sem. **ndd*.

¹⁶ Modern Arm. *wakḵef* 'se tenir debout' (compared in DRS 609), absent in other Arm., must be a loan from Arb.

81. STONE

- (1) Akk. *abn-*; Ugr. *ʔbn*; Hbr. *ʔäbän*; Pho. *ʔbn*; Bib. *ʔäbän*; Pal. *ʔäbēn*; Sab. *ʔbn*; Gez. *ʔäbän*; Tna. *ʔämn-i*; Tgr. *ʔämən*; Sod. *əmmayyü*; Har. *ün*; Wol. *un*; Cha. *əməṛ*; Soq. *ʔöben* // < Sem. **ʔabn-*.
- < Afras. **(ʔa)bun-* ‘stone, millstone’: Egyp. (ME) *bnw.t* ‘stone, millstone’; Chad. W. **ʔabun-* ‘millstone’: Bolewa, Dera, Ngamo *bùní*, Kariya *vun*, Mburku *avəna*, Zaar *vuun* (*v* points to **ʔab-*), C: Dghwede *vurá*, Mandara *úvra* ‘millstone’, Zime-Batna *və̀nà* ‘stone’, Gidar *búúna nà ara* ‘stone for grinding’ (ADB)¹⁷.
- (2) Syr. *kəp-*; Urm. *kīʔp-* // only Arm. (see HALOT 493).
- (3) Mnd. *gīal-* // Arm. (Off. *gll*, Palm. *gllʔ*) HJ 224 < Sem. **ga/ul-* ~ **gīal-* ‘stone’: Hbr. *gal* ‘heap, heap of stones’ (HALOT 190), Arb. *ʔūl-*, pl. *ʔīlāl-* ‘rocher dans l’eau’ (BK 1 358). Perhaps < Sem. **gll* ‘be round’ (see DRS 125–6). If so,
- < Afras. **g^walVl-* ‘be round’: Brb. **gelell-et* ‘be round’; Chad. W.: Hausa *gùlùlù* ‘ball’, C.: Gulfei *ngolō* ‘round’, Mofu *-ngá-ngál-* ‘make a ball’. Cf. also isolated E. Chad.: Somrai *gálá*, Mubi *zan-gala* ‘stone’ (ADB).
- (4) Qur. *ḥaʒar-*; Lbn. *ḥaʒar*; Mec. *ḥaʒar* // No direct cognates. Cf. Tgr. *ḥəgrat* ‘a sort of red stone or red earth’ (LGz 228), which may be an Arabism, and Gez. *ḥəg^wər* ‘grape seeds’ (ibid.: “probably from Ar. *ḥaʒar* ‘stone’”).
- Probably related (with metathesis) to Brb. **Hirg-* ‘stone’: Ghadames *irəg*, Zenaga *t-iʔrg-ət* ‘stone’.
- (5) Mlt. *ʒeblla* // < Sem. **gVbVl-* ‘mountain; boundary, border’, see No. 55 ‘mountain’ (4) in Mil. 2012.
- (6) Amh. *dəngay*; Arg. *dəngay*, *gənzela* (< **gVndVl-*, met. < **dVngVl-*); Gaf. *dānga* // Tgr., Tna. *dāng^walla* ‘rocher’ (DRS 279). No cognates in Sem.; looks like a loan.
- Hardly a loan from Oromo *ḍagaa* ‘stone’ < E.Cush.: LEC **ḍagaḥ-* id. Obviously connected with C.Cush. **dal^{ing}Vr-* ‘stone’ (Bilin *dāng^wəra*, Khamtanga *dəg^wra*, *dengúr*, Khamir *dīgūrā*, Waag *digura*), but why *-l*?
- (7) Hob. *šoor*; Hrs. *šewwer*; Mhr. *šowwer* // < Sem. **šVwwVr-* ~ **šir(a)r-*: Hbr. *šərōr* ‘stone or pebble’, Jud. Arm. *šərār-* ‘pebble, flint’ (HALOT 1055; not related to *šōr* < Sem. **ʔul/ir-* ‘flint, rock’), perhaps also Syr. *šūr-* ‘pulvis’ (Brock. 536), Arb. *širrat-* ‘poussière’ (BK 1 1326).
- (8) Jib. *fúdún* (also ‘rock’) // Soq. *félídehon* ‘montagne’ (LS 333) < Sem. **pad(d)an-* ‘path in the mountains’, see No. 55 ‘mountain’ (10) in Mil. 2012.
- Common Semitic **ʔabn-* (#1).

82. SUN

- (1) Akk. *šamš-*; Hbr. *šämāš*; Pho. *šmš*; Bib. *šəmaš*; Pal. *šmš*; Syr. *šəmš-*; Mnd. *šam(i)š-*; Urm. *šimš-*; Qur. *šams-*; Lbn. *šaməs*; Mec. *šams*; Mlt. *šemeš*; Sab. *s₂ms₁*; Soq. *šam* // < Sem. **š_xam(š_xam)-*
- < Afras. **šam-* ‘burning heat’: Brb. **ʔi-sam* ‘lightning’: Ndir, Figig *i-ssim*, Ahaggar *e-ssam*, etc.; Egyp. (OK) *šm*, *šmm* ‘be hot, be burning’; (?) W.Chad.: Angas *lem* ‘sun’ (questionable as Angas *l-* can reflect **l-*, **š-*, **č-*); S. Cush: Alagwa *čehemu* (with a secondary extension *-h-*, see Mil. RE) ‘sun, daylight’, Burunge *čema* ‘sun’¹⁸ (perhaps also Asa *demo-k* ‘morning’). ADB.

¹⁷ The status of **ʔa-* is not quite clear. Presumably, it is a prefixal root extension, not preserved in Egyp. On the other hand, cf. Cush. **ʔab-* (suggesting a segmentation **ʔab-n-*?): Cush. N.: Beja *ʔawi* ‘stone, rock’ (< **ʔab-*?), C.: Bilin *ambá*, Khamta *abaa*, Khamir, Qwara *abaa* ‘mountain’, E.: LEC: Baiso *eʔebo* ‘stone’ (ADB).

¹⁸ According to Tak. 2011 131, S.Cush. **č* (= *tl*; rendered ibid. as **č₁* vs. S.Cush. **č₂* continuing Afras. **č*) points to Afras. **č[a]m-*, not **šam-* as suggested here. However, the situation with S. Cush. laterals is not so straightforward and requires further investigation. According to O. Stolbova, “S.Cush. **š* corresponds to both Chadic non-

- (2) Ugr. *špš* // (cf. Ebl. *sí-piš*, cf. also Hbr. pl. *šabūs-ūm* ‘small sun disc used as jewellery or amulet’ HALOT 1392–93; strangely with *-b-*).
- < Afras. **šip-* ‘sunlight, brightness’: Egyp. (Pyr.) *sšp* ‘to be light’ (*sš-* reflects **š-*?), Brb. **a-sif* ‘day’: Ghat *a-sef*, Ahaggar *esef*, etc.; Chad. W.: Hausa *záafí* ‘heat’, *zúffā* ‘hot weather’, Mupun *mì-lēp* ‘shining thing’, Angas *mu-lep* ‘shine; star’, Ankwe *mìye-lap* ‘lightning’, Tangale *ki-lip*, *ŋi-lip* ‘shining, brilliant; lightning’ (the former four words can be < **žyap-* or **-lyap-*), C.: Kapsiki *žàvu* (< **šap-*) ‘white’ (ADB; cf. otherwise St. 2007 #437).
- (3) Gez. *dāḥay*; Tna. *šāḥay*; Tgr. *šāḥay*; Amh. *ṭay*; Arg. *čāhed* // < Sem. **šVḥy-*: Arb. *duḥā*, *dīḥḥ-* ‘sun’.
- < Afras. **čVḥay/w-* ‘burning sun’: C.Chad. **čaw-* ‘red’: Zime *šīēw*, Masa *cáw*, Mesme *šeo*, Banana *šawi* (St. 2007 #304); S.Cush.: Dahalo *šaḥ-* ‘to burn’ (ADB).
- (4) Gaf. *äymärä*, Sod. *yimər*, *imər* // < Eth.: Gogot *imir* ‘sun’, Gez. *ʔamir* ‘sun, day, time’ (according to LGur 51, the Gurage and Gafat terms are probably from the Gez. root *ʔamärä* ‘indicate, show’, which seems far-fetched), Tgr. *ʔammära* ‘be bright’.
- < Afras. **ʔamir-* ‘daylight’: Egyp. (18 Dyn.) *mšw.t* (if *š* < **-r-*) ‘rays, beams, brightness’; Brb.: Iznasen *ta-miri* ‘moonlight’, Ahaggar *é-mmar* ‘chaleur rayonnée’, *a-sa-mmer*, Ayr *i-səssə-mmar* ‘rayon de soleil’, etc.; C.Cush.: Bilin *amār* ‘klar, licht sein’, **ʔämär-* ‘tomorrow’ (in App. CDA 138, called “a borrowing in EthSem, Gz. *ʔamir* ‘day’”), which is as semantically impossible as, vice versa, a borrowing from Gez.: Agaw ‘tomorrow’ is a normal meaning shift from ‘morning’). See also EDE III, 68–9.
- (5) Har. *īr*; Wol. *ayr*, Cha. *eyat* // < Eth.: Gez. (Sawasew) *ʔer*, Selti *ayər*, Zway *arīt*, Muher, Masqan, Gogot *aret* ‘sun’. Unless from E.Cush.¹⁹ < Sem. **ʔVr-* ‘daylight’ 2, ‘sun’ 3: Akk. *urru* ‘daytime’ (CDA 426), Ugr. *a/ir* ‘light’ (DUL 94), Hbr. *ʔōr* ‘daylight’ (HALOT 24).
- < Afras. **ʔay/wr-* ‘sunlight’: Chad. C. **wur-* / **yirVw-*: Gudu *wür* ‘morning’, Gude *wurə* ‘go to do something or go somewhere early in the morning’, Buduma *yirow* ‘day’, E. **ʔiray-*: Dangla *èrìyó* ‘noon’, Bidiya *ʔira* ‘day’; E. Cush.: Saho, Afar *ayro*, LEC: Jiddu *ʔariye?*, Baiso *ʔareti*, HEC: Darasa *ari-ššo*, Sidama *arri-ššo*, Qabenna *arru*, Tembaro *arri-čču* ‘sun’ (ADB).
- (6) Hob. *ḥyúúm*; Hrs. *ḥeyom*; Mhr. *ḥeyum*; Jib. *yum* // < Sem. **yawm-* ‘day’: Akk. *ūmu*, Ugr. *ym*, Hbr. *yōm*, Arb. *yawm-* (HALOT 399).

emphatic laterals” (St. 2007 11), i.e. to **č* (corresponding to Sem. **š* < Afras. **č*) and **š* (corresponding to Sem. **š*_x < Afras. **š*). In two sets of correspondences adduced by Takács, the opposition **š* vs. **č* in Proto-Afras. is based on the same opposition in S.Cush., Chad. and Sem. (following, without mentioning it, the correlation first established by the present author in an unpublished manuscript of 1979 and later described in detail in SED I XCVIII–CV). As for Sem., the difference between the somewhat rare **š*_x (< Afras. **š*) and the much more common **č* (< Afras. **č*), suggested in these studies and apparently accepted by Takács, lies in the opposition of two sets of correspondences: (1) **š*_x > Hbr. *š* ~ **-s*, MSA **š* and (2) Hbr. *š*, MSA *š*, with Arb. *š* in both sets. However, out of all the examples illustrating Afras. **š* and **č* that are adduced in Tak. 2011 128–132, I could not find a single convincing Semitic case supporting the opposition **š* vs. **č* in S.Cush. (I will analyze these examples in my forthcoming review of Tak. 2011).

¹⁹ According to LGur 118, forms with *-r* are either borrowed from Cushitic or represent the Ethiopic root *ʔmr* with weakened *m*, including the possibility of Eth. *ʔmr* “taken over as *ir* into Cushitic and then borrowed under this form by the various Ethiopic languages”. The idea of “weakened” *-m-* does not hold water, if only because *ʔer* is attested in Sawasew; the Eth. forms are hardly the source of borrowing into East Cushitic, where it is a widespread term for ‘sun’. As for Cha. *eyat* (and similar forms in other Gurage), it is “perhaps to be identified with *aret*, by palatalization of *r*”, according to LGur 118; though I am somewhat skeptical about the loss of *-r-* “by palatalization”, I am including the Cha. term into this entry for lack of a better etymology (another possibility is a borrowing from Omotic: cf. Chara *oyá*, Dime (*ʔiyy-u*, etc.; cf. also E.Cush.: LEC: Arbore *ʔawaté*, Elmolo *áóte?* < **ʔawajš-*, all meaning ‘sun’). ADB.

- < Afras.: Egyp. (MK) *imy* ‘sun (as an eye)’; W.Chad.: Hausa *yàmmā* ‘west; afternoon (when the sun is in the west)’ (unless an Arabism). ADB.
- Common Semitic *š_xam(š_xam)- (#1).

83. SWIM

- (1) Hbr. *šhy*; Syr. *shy*; Urm. *shy* // < West Sem. (Hbr.-Arm.) *š_hy or *š_hy: Chr-Pal. *sh?* ‘wash oneself’, Jud. *shy* ‘swim’, Mnd. *saa* ‘wash, perform ablution’ (HALOT 1314).
 - < Afras. *čah/š- ‘bathe, wash oneself, swim’: Chad. *šVw/y- ‘to bathe, wash’: W.: Sayanchi *šaa* ‘to swim’, Tala *šau* ‘wash’, C.: Gidar *šay* ‘to bathe’, Mafa *šah-* ‘wash a calabash’, Musgu *šaf* ‘swim’ (ADB after Lukas, not included into St. 2007 #19), E.: Mokilko *súuyè* ‘to bathe, wash oneself’, Saba *so?e* ‘wash’, etc. (cf. *ibid.*); S.Cush.: Iraqw *šawš-aas* ‘to rinse, clean’.
 - (2) Pal. *šwt* // No etymology I could find.
 - (3) Qur. *sbh*; Lbn. *sabaḥ* // Hob. *soóbəḥ* must be an Arabism. No etymology.
 - (4) Mec. *šām*; Mlt. (gh)ōm // No etymology.
 - (5) Gez. *ḥ/ḥammaša*; Tna. *ḥambäsä*; Tgr. *ḥämmäsa* // Cf. Arb. *ḥāmišat-* ‘petit ruisseau ou cours d’eau’ (BK 1 633).
 - Cf. isolated W.Chad.: Gwandara *amiši* ‘swimming’.
 - (6) Gez. *wānaya*; Amh. *waññe*; Arg. *wañña*; Sod. *waññäm* // Only Eth. (LGz 615) with this meaning.
 - Cf. Egyp. (Gr.) *iwny* ‘waters’; Chad. C.: Gude *yinwa* ‘dip into’, E.: Migama *ʔányáw* ‘to bathe’, Mubi *ʔányý*, Kajakse *ʔanyiinyà* ‘to swim’; N.Omot.: Mocha *wāna* ‘swim’ (looks an Amharism). All < Afras. *wny/*ynw/*nw/y ?
 - (7) Cha. *darag’ä* // According to LGur 219, lit. ‘hit a hitting’ < Common Gurage *dārrägä ‘hit, beat’, perhaps to be identified with Amh. (a)dārrägä ‘make, do, act’. Another possibility is a meaning shift from Sem. *drg ‘to march, travel’, *darag- ‘way, path’: Akk. (YB) *daragg-* ‘way, track’ (CDA 56), Arb. *drž* ‘aller, marcher’, *daraž* ‘chemin, sentier’, etc. (DRS 308).
 - (8) Jib. *reḥ* (< *rbḥ*) // Soq. *rābaḥ* ‘se baigner’ (LS 392). No etymology.
 - (9) Soq. *geb* // Likely < Sem. *gVbb-, *gab?- ‘well, water-basin’: Mhr., Hrs. *gēbyət* ‘pond, round cistern’, Ugr. *gb* ‘well’, Hbr. *gēb* ‘pit, ditch’, *gäbā?* cistern, puddle’, Syr. *gubb-*, *güb-* ‘well’, Arb. *žubb-* ‘well’, *žab?* ‘lake, pool’, Gez., Tgr. *gabb* ‘pit, ditch, hole’ (LGz 176).
 - < Afras.: Egyp. (Pyr.) *ʔgby* ‘flood’; cf. also Chad. C.: Lamang *óghbù* ‘hole’, *ghùvàyà* ‘well (n.)’, Bachama *gòháḃi* ‘well (n.)’, Gidar *ógbàʔá* ‘to pour’, E.: Somrai *gúbu* ‘hole’; N.Cush.: Beja *gabōy* ‘to flow’ (ADB; cf. EDE I, 68).
 - ◇ Har. (tä)wāka and Wol. *wače* are < HECush. (see LGur 650); Hrs. *sebōḥ* and Mhr. *sūbaḥ* must be Arabisms. No term registered in Akk., Ugr., Pho., Bib., Mnd., Sab. and Gaf.
- No common Sem. term.

84. TAIL

- (1) Akk. *zibbat-*; Ugr. *ḏnb*; Hbr. *zānāb*; Syr. *dunb-*; Mnd. *dinabt-*; Mlt. *demp*; Gez. *zānāb*; Tna. *zānāb*; Tgr. *zānnab*; Hob. *ḏenuúb*; Hrs. *ḏenēb*; Mhr. *ḏənūb*; Jib. *ḏúnub*; Soq. *dínob* // < Sem. *ḏalinab-(at-) (SED I #64).
- (2) Pal. *ḫkwš* // Most likely < West Sem. *ḫkš* ‘curl, braid’: Syr. *ḫaḫušḫuš-*, *ḫaḫušḫus-* ‘crispus’ (Brock. 543), Arb. *ḫiḫšat-* ‘tresse, natte de cheveux’ (BK 2 318).
 - Cf. isolated, but immaculately compatible N.Omot.: Bworo *ḫaččea* ‘tail’.
- (3) Urm. *ḫipr-* // Perhaps a meaning shift < Sem. *ḫip(V)r-* ‘claw’: see No. 13 ‘claw’ (1) in Mil. 2010.
- (4) Lbn. *ḏayl*; Mec. *dayl* // with a metaphoric meaning shift < Sem. *ḏyl ‘to be low, humble’: Arb. *ḏyl* ‘ê. peu estimé, bas; traîner par terre (vêtement)’, *ḏayl-* ‘partie inférieure, bas, queue’, Syr. *dāl* ‘ê. humble’ (DRS 331).

- (5) Tna. (syn.) *mälaläs* // No etymology (if < *mä-laslas-, note a distant and curious parallel in Arb. *lt?* ‘introduire la tête dans un vase et y boire en remuant la queue (se dit d’un chien)’ BK 2 965).
- (6) Har. *ḳänāwa* // < Sem.: Gur. **ḳən* (in LGur 516, the notation is *q’an*) ‘buttocks, anus, bottom of a thing’, Tgr. *ḳən* ‘lower/back part; vulva’ (LGur 516), Akk. (NB) *ḳinnatu* ‘buttocks, rump’ (CDA 289), Arb. *ḳaynat-* ‘derrière, le bas du dos’ (BK 2 848)
- < Afras. **ḳayn-* ~ **ḳany-* ‘lower part of back’: Chad. W.: Hausa *ḳāniyā* ‘perineum, female pudenda, anus’, Sura *ḳānòk*, Montol *ḳóḳ* ‘back’, C.: Glavda *kúnya*, Tera *ganà* ‘back’ (*g-* < **k* or *g*); Cush. E.: LEC: Oromo *ḳinti*, *ḳinti* ‘spine (esp. at the waste line)’ (ADB, cf. ND 1078b).
- (7) Sod. *wədinna* // No etymology.
- ◇ Amh., Arg. and Gaf. *čəra* are probably loans from E.Cush. (Oromo *čəṛā*, Sidamo *čira*, etc. LGur 187); Wol. *goññä*, also meaning ‘penis’, is a Cushitism: HEC: Hadiya *gonnāṛä* ‘penis’ (LGur 286); Cha. *žəwä* is, according to LGur 319, from E.Cush.: Oromo *dubo*, etc., with *d* > *ž* and *b* > *ḳ* > *w*. No term registered in Pho., Bib., Qur. and Sab.
- Common Semitic **da/inab-(at-)*- (#1).

85. THAT²⁰

- (1) Akk. *ulliw*; Qur. *dālika* // The elements *-ll-* and *-l-* supposedly denote distal deixis as opposed to the proximal one, marked with *-nn-* in Akk. and what seems to be a zero marker in Qur. (ADB).
- (2) Ugr. *hnk*; Bib. *dēk*; Qur. *dālika*; Lbn. *hadak*; Mec. *dak*, *hādāk*; Mlt. *dak*; Gez. *zəkkū*; Hob. *dēek*; Hrs. *dēk*; Mhr. *ḳik*; Jib. *dōkun* // Related by the element *-k*, apparently denoting distal deixis as opposed to the proximal one (seemingly expressed by *-d* in Ugr., *-n* in Gez. and MSA, \emptyset in Lbn. and Mec., etc.) (ADB).
- (3) Hbr. *hahū(?)*; Pho. *h?*; Pal. *hāhū(?)*; Syr. *haw*, *huw*; Urm. *hō*; Sab. *h?*; Tgr. *loha* // Related by the element *h*, apparently denoting distal deixis as opposed to the proximal one, expressed by the elements *-nd/-dn* in Hbr., Pal., Mnd., etc., by the element *-n* in Syr., by the element *?* in Urm. and Tgr., etc. (ADB).
- (4) Mnd. *hanata*; Tna. *ʔati*, *ʔatuy* // Related by the element *-t-*, apparently denoting distal deixis as opposed to the proximal one, expressed by the element *-z-* (< **d*) both in Mnd. and Tna. (ADB).
- (5) Amh. *zzəya*, *ya*, Gaf. *aññä*; Har. *ya?*; Wol. *annä*; Sod. *za*; Cha. *za* // United by the vocalism *-a-*, apparently denoting distal deixis, as opposed to the proximal one, expressed by the vocalism *-i-* in most S.Eth. (in Amh. also by the element *-h* < **k*). ADB.
- (6) Arg. *hod*, *wod* // Very likely to be qualified with other S.Eth. (#5) by the vocalism *-o-* (< **a*) as opposed to the proximal deixis, expressed by the vocalism *-u-* (ADB).
- (7) Soq. *degen* // With the element *-n*, probably denoting distal deixis as opposed to the proximal one expressed by a zero marker, and the element *-g-* of an obscure origin (ADB).
- Common West and South Semitic **dVḳ-*.

²⁰ In view of the compound character of deictic pronouns in Sem. and after much hesitation and failed efforts to organize the entries in different ways, I have opted for the present method. It relies on the binary opposition between proximal and distal deixis in the same language (e.g. Akk. *anniw* ‘this’ vs. *ulliw* ‘that’) allowing to single out proximal and distal deixis markers (Akk. *-nn-* vs. *-ll-*) and then give identic scores to those of them that coincide phonetically in different languages, disregarding those other pronominal elements that apparently do not convey the notion of proximal or distal deixis. Afrasian parallels to Sem. deictic pronouns are not adduced, since this would require a separate full-scale study.

86. THIS

- (1) Akk. *anniw*; Bib. *dānā*; Syr. *hānā*; Mlt. *dan*; Sab. *dn*; Gez. *zantu*; Hrs. *den* (also *dah*, *dī*); Jib. *denu* // Related by the element *-n*, apparently denoting proximal deixis as opposed to the distal one, expressed by *-ll-* in Akk., *-k* in Bib., Gez., Hrs. Jib., etc. (ADB).
 - (2) Ugr. *hnd* (*d* < **d*); Hbr. *hazzē* (< **hand-*); Pho. *zn* (and *z*); Pal. *dē(y)n*, *hādē(y)n*; Mnd. *hazin*; Hob. *dāanah* // Related by the element *-nd/-dn*, apparently denoting proximal deixis as opposed to the distal one expressed by *-k* in Ugr., *-h-* in Hbr., Pho. Pal., *-t-* in Mnd. (ADB).
 - (3) Urm. *āhā*; Tgr. *ʔalli/a* // United by the element *ʔ-*, apparently denoting proximal deixis as opposed to the distal one, expressed by a zero marker in Urm. and *-h-* in Tgr. (cf. also changes in the vocalic patterns). ADB.
 - (4) Qur. *hādā*; Lbn. *hēdā*; Mec. *hāda* // Featuring the element *h-*, supposedly denoting proximal deixis as opposed to the distal one, probably expressed by the elements *-l-* and *-k* (or one of them). ADB.
 - (5) Tna *ʔazi*, *ʔazuy* // With the element *-z-* (< **-d-*), apparently denoting proximal deixis as opposed to the distal one expressed by *-t-* (ADB).
 - (6) Amh. *zzih* (< **zik*), *yih* (< **yik*); Gaf. *āññə* (*ə* < **i*); Har. *yīʔ*; Wol. *ənnä* (*ə* < **i*); Sod. *zi*; Cha. *zə(k)* (*ə* < **i*) // United by the vocalism *i-* (also with the *-k* element in Amh. and Cha.), apparently denoting proximal deixis as opposed to the distal one expressed by the vocalism *-a-* (ADB).
 - (7) Arg. *hud*, *hudatte* // Distinguished by the vocalism *-u-*, apparently denoting proximal deixis as opposed to the distal one, expressed by the vocalism *-o-* (likely < **a-*) (ADB).
 - (8) Mhr. *dome* // Distinguished by the element *-m-*, supposedly denoting proximal deixis, as opposed to the distal one, probably expressed by the element *-k* (ADB).
 - (9) Soq. *da* // Distinguished by the zero marker, apparently denoting proximal deixis as opposed to the distal one expressed by the elements *-g* and/or *-n* (ADB).
- Common West and South Semitic **dVn-*.

87. THOU (m.)

- (1) Akk. *atta*; Ugr. *ʔat*, *ʔattā*; Hbr. *ʔattā*; Pho. *ʔt*; Bib. *ʔantā*; Pal. *ʔat(t)*; Syr. *ʔa(n)t*; Mnd. *anat*; Urm. *ʔa(n)t*; Qur. *ʔanta*; Lbn. *ʔint(ə)*; Mec. *ʔinta*; Mlt. *int*; Sab. *ʔnt*; Gez. *ʔanta*; Tna. *ʔanta* (only in addressing); Tgr. *ʔanta*; Amh. *antä*; Gaf. *ant*, *antä*; Wol. *atä* // Sem. **ʔanta* (ADB).
 - < Afras. **ʔa(n)tV*: Eyp. (Pyr.) *ntk(y)*; Cush. C. **ʔant-*, E. **ʔati*, S.: Dahalo *ʔāta*; N. Omot.: Dizoid **yetu/a* (ADB).
 - (2) Tna (syn.) *nəssə-ka*; Arg. *ank(u)*; Har. *akā-*; Cha. *akä-* // As a personal pronoun in Mod. Eth. only (Tna. *nəssə* is assimilated < **nəfsə* ‘soul’). Likely of the same origin as the Sem. possessive pronoun **-k* of the 2nd person (ADB).
 - < Afras. **ʔa-(n)kV(k)-* ~ **kVn-* ‘thou’: Eyp. *kw* (m.), *tw* (f.); Brb. **kay* ~ **kunn* ~ **kVkk*; Chad. W.: Hausa *kay/kē*, Bole *kai*, Kiir *gay* (< **kay?*), C.: Mandara (*ba*)-*ka*, Gisiga *ka*, Mofu-Gudur *ka(h)*, Musgu *tuku(nu)*, E.: Migama *kîtà*; *ki/a*, Jegu *kē*, Mokilko *keŋ/koŋ*; Cush. E.: LEC: Dase-nech *kúuni*, Arbore *ké*, etc., S.: Iraqw *kuŋ/kiŋ*, Qwadza *-ku* (ADB).
 - (3) Sod. *dāhā* // No etymology.
 - (4) Hob. *hēt*; Hrs. *hēt*; Mhr. *hit*; Jib. *het*; Soq. *het* // No etymology (Proto-MSA **h-* cannot be explained from **ʔ-* in *ʔanta*).
 - (5) Soq. (syn.) *ʔē/ī* // No reliable etymology proposed.
- Common Semitic **ʔant-* (#1).

88. TONGUE

- (1) Akk. *lišān-*; Ugr. *lšn*; Hbr. *lāšōn*; Bib. *liš(š)ān*; Pal. *lyšn*, *lāššān*; Syr. *leššān-*; Mnd. *lišān-*; Urm. *lišān-*; Qur. *lisān-*; Lbn. *lsen*; Mec. *lisān*; Mlt. *lsin*; Sab. *ls₁n*; Gez. *lassan*; Tgr. *nəssal*

(met.); Hob. *lšáan*; Hrs. *lēšen*; Mhr. *ewšēn*; Jib. *elšén*; Soq. *léšin* // Sem. *liš(š)ān- (SED I #181).

- < Afras. *lis- ‘tongue’: Egyp. (Pyr.) *ns*; Brb. *ilVs- (< *?V-lis); Chad. W. *ha-lis-um-, C. *?V-lyas-, E. *lyas-an- (ADB).
 - (2) Tna. *mālḥās*; Amh. *mālas*; Arg. *mālas*; Gaf. *mālasū* // < Sem. *lḥs ‘to lick’: Akk. *lāš-* ‘to taste, lick’ (otherwise < *lḥš, cf. Arab *lḥs*), Arb. *lḥs*, Gez. *laḥasa*, Mhr. *laḥās* ‘to lick’ (LGz 311, ADB).
 - ◇ Sod. *alāmāt*; Har. *arrāt*; Wol. *arāmāt*; Cha. *anābat* are loans from E.Cush. (see LGur 89). No term registered in Pho.
- Common Semitic *liš(š)ān- (#1).

89. TOOTH

- (1) Akk. *šinn-*; Ugr. *šn*; Hbr. *šēn*; Bib. *šēn*; Pal. *šn*; Syr. *šenn-*; Mnd. *šin-*; Qur. *sinn-*; Lbn. *sən*; Mec. *sinn*; Mlt. *sinna*; Gez. *sənn*; Tna. *sənni*; Arg. *sən*; Gaf. *sənā*; Sod. *sənn*; Har. *sən*; Wol. *əsən*; Cha. *sən*; Jib. *šnin* // Sem. *šinn- (SED I #249).
 - < Afras. *sin- ‘tooth’: Brb. *sīn- ‘tooth’; Chad. W. *sin-, C. *šyan- (with a “secondary lateralization”), E. *syān-; S.Cush. *siḥin- (with the presumptive body part marker *ḥ inserted, see Tak. *ḥ): Iraqw, Gorowa, Alagwa, Burunge *siḥino* (ADB).
 - (2) Urm. *kīk-* // Sem. *kVkk- ‘tooth, thorn’: Syr. *kakk-* ‘dens molaris’ (Brock. 326), Akk. *kakk-* ‘weapon, tool, shaft, barb, thorn’ (CAD *k* 50); Gez. *kokaṭwa* ‘to break or chip the teeth; set the teeth on edge’, Tna (*ʔa*)*kok^wäyā* ‘set the teeth on edge’ (LGz 281) are denominal verbs. < Afras.
 - Cf. Cush.: E.: Saho, Afar *ikoó* ‘tooth’, S.: Maʔa *iʔike* ‘tooth’ (< Afras. *ʔik- ~ *kak(k)-?).
 - (3) Sab. *ṭn*, Hob. *mṭənyóot* (also ‘incisor’; no other term for ‘tooth’ in Nak.Hob.); Soq. (syn.) *mitnióh* (quoted as ‘dent’ in LS 443) // Mhr. *məṭənyēt* (JM 418), Jib. *muṭnút* ‘incisor tooth’ (JJ 285), Arb. *ṭaniyyat-* ‘dent de devant (l’une de quatre, dont deux supérieures et deux inférieures)’ (BK 1 239), Sab. *ṭny* ‘front-teeth’ (SD 151), with a very likely semantic development from Sem. *ṭinay- ‘two’ < Afras. *čīn- ‘two’: see below No. 91 ‘two’ (#1).
 - (4) Tgr. *nibāt*, pl. *ʔänyab* // < Sem. *nīb-, *nāb- ‘(canine) tooth’: Akk. *nayabtu* ‘floating rib or cartilage at the tip of the rib’ (CAD *n1* 151; according to AHw 717, ‘ein innerer Körper. nahe der Galle’), Hbr. pB *nīb šəpātayim* ‘upper lip’ (Ja 902), Syr. *nīb-, nāb-* ‘dens maxilaris’ (Brock. 427), Arb. *nāb-* (pl. *nuyūb-*) ‘dent canine’ (BK 2 1375).
 - (5) Amh. *ṭərs*; Hrs. *mežréḥ*; Mhr. *məžrāḥ* // non-etymological -ḥ may be the body part marker *ḥ in postposition; < Sem. *šīrs- ‘molar tooth’: Eblaitic *za-ra-sa-tum /qaraštum/* ‘scheggiatura dei dente’ (after Conti), Syr. *šaršā*, Arb. *ḥīrs-*, Sab. (pl.) *ʔḥrs₁*, Gez. *ḥərs* (SED I #275).
 - (6) Soq. *šále* (the main term for ‘tooth’, according to Kogan’s list) // In LS 309, quoted as *šélhe* ‘dents’ (“originaiement ‘dents supérieures’”), deriving it from *šélhe* ‘haut’ < Sem. *šly ‘être haut’, which seems the only plausible etymology for lack of better ones.
 - (7) Soq. (syn.) *šásal* // One more term for ‘dent’ adduced in LS 431 with no Sem. etymology, save for a tentative, according to Leslau (and impossible) comparison with Arb. *sinn-*.
 - Cf. Brb. *ta-zal-t (rather *ta-zHal-t) ‘dent canine’: Ahaggar *tā-hal-at*, Ayr *ta-zal-at*, etc. (compared with a question mark in Tak 2001 85 with Egyp. OK *zʔ.t* (also *zʔr.t*) ‘ein Meissel’) which is ideally compatible with the Soq. term both phonetically (< Afras. *čāš(a)l-) and semantically.
 - ◇ No term registered in Pho.
- Common Semitic *šinn- (#1).

90. TREE

- (1) Akk. *iš(š)-*; Ugr. *šs*; Hbr. *šēš*; Gez. *šəḥ*; Tgr. *šəčūt*; Gaf. *ənča*; Wol. *ənče*; Cha. *äčä* // < Sem. *šī/ušš-: Bib. *ʔāšā* ‘wood’ (HALOT 1821), Arb. *šudḥ-* ‘gros chicot d’arbre; gros morceaux de bois’, *ʔašdād-* (pl.) ‘arbres ou arbrisseaux à épines’ (BK 2 277)

- < Afras. *ʕi(n)ĉ-: Egyp. (BD) *ʕd* ‘k. of wood’; Omot. **HinĈ-* (< **ʕinĉ-*) ‘tree’: Gimirra (Bench) *inĉ*, Mao (Hozo) *ʕiinc*, Dizi (Sheko) *iinĉu*, Ongota *hanĉa*, *hanša* (ADB). Cf. also Brb.: Zenaga *te-iḏuḏ* ‘big stick’ and N.Cush.: Beja *ʔaḏa* ‘pole, long stick’ compared with Egyp. and Sem. in EDE I 95.
- (2) Bib. *ʔilān*; Pal. *ʔilān*; Syr. *ʔilān-*; Mnd. *alan-*; Urm. *ʔilān-* // < Sem. **ʔalyān-* ‘(oak)tree’: Akk. *aliānu* ‘oak; acorn’ (CAD *a1* 354), Ugr. *ʔaln* ‘oak grove’ (DUL 58), Hbr. *ʔallōn* ‘any big tree’, pB ‘oak, acorn’ (HALOT 54)²¹.
- < Afras.: (?) Egyp. (OK) *wʕn*, *wʕr* (conveys *wʕl*; I suspect here a change *ʔ > ʕ, perhaps caused by the vicinity of **l*) ‘Art Nadelholz’ (ADB); Chad. **lV(wV)ηV*: W.: Daffa-Butura *lan* ‘Wildnis, Busch’, C.: Mbara *luḡ* ‘tree’, Munjuk *luḡ* ‘bois, arbre’, Musgu *lúuḡ* ‘tree’, E.: W. Dangla *lèlnyó* ‘arbuste’, Bidiya *lèeliḡko* ‘arbuste sp.’ (St. 2005 #303).
- (3) Qur. *šažar-*; Lbn. *šažar*; Mec. *šažara*; Mlt. *sižra* // < **šagar-*. Hob. *šəžrét* ‘tree’ (syn.) and Hrs. *šegerét* ‘plant name’ may be Arabisms, though *š* in both cases as well as -*g-* (instead of the expected *ž*) and a different meaning in Hrs. leave a grain of doubt.
- < Afras. **ĉagVr-*: Chad. W.: Dott *šakar* ‘Deleb-palm’ (irregular -*k-*?), E. **ĉVgVr-* ‘palm Hyphaene thabaica’: Sokoro *sóger*, West Dangla *čárgó*, Bidiya *čargò* (met.). ADB (in view of regular correspondences in the Sem. and E.Chad. triconsonantal roots their similarity is hardly accidental).
- (4) Tna. *ʔom* // No etymology that I could find.
- (5) Hob. *hərmít*; Hrs. *herōm*; Mhr. *hərmáyt*; Jib. *heyrōm* // < Sem. **harūm-* ~ **harm-ay-*: Soq. *hərcm*, Akk. (MB) *urūmu* ‘a tree in mountains’ (CDA 427), Arb *harmī* ‘bois sec’ (BK 2 1415)²².
- (6) Soq. *širohom* // < **širVm-* or **š_xirVm-*. No regular parallels in Sem., but it is possible to suggest two debatable ones: Jib. *šerm* ‘kind of thorn-bush’ JJ 255 (with irregular correspondence of sibilants) or, with metathesis, Sem. **šami/ur-* ‘thorny bush or tree’ (Hbr. *šāmīr*, Arb. *samur* HALOT 1562); the eventual connection of all of them as variant roots cannot be ruled out either.
- Cf. Egyp. (18 Dyn.) *š* ‘tree’ (if < **šr*); C.Chad. **šarVm-*: Mofu *šarəma* ‘Balanites aegyptica’, Daba *šərmə* ‘soap-tree’ (Chad. **š* seems to contradict Soq. *š*, if < **š_x*). ADB.
- ◇ Amh., Arg. and Sod. *zaf* are loans from C.Cush. (LGur 703 quotes Qemant *zaf* and Qwara *žafa*; see App CDA 140 for more Agaw forms); Har. *läfu* without Eth. parallels looks like a loan from deverbal Arb. *laff-* ‘arbre touffu, plante touffu’ (BK 2 1008) from *lff* ‘è. épais, touffu (se dit des herbes)’ (ibid. 1007). No term registered in Pho. and Sab.
- Common North and West Semitic **ʕi/ušš-* (#1).

91. TWO

- (1) Akk. *šinā*; Ugr. *tn*; Hbr. *šənayim*; Pho. *šnm*; Qur. *ʔitnāni*; Lbn. *tneyn*; Mec. *ʔitnayn*; Mlt. *tnēyn*; Sab. *tny* // Sem. **tinay-*

²¹ In view of Chadic parallels, Sem. **ʔalyān-* goes back to **ʔa-lyān-* (supported by Arb. *līnat-* ‘palmier en gen.’ BK 2 1051 and *lūnat-* ‘sorte de palmier’ ibid. 1045, compared by Stolbova) rather than to **ʔaly-ān-* (supported by Ugr. *ʔlt* ‘support, pillar’ or ‘footstool, base’ DUL 66, Hbr. *ʔēl-īm* (pl.) ‘mighty tree’ HALOT 40, *ʔallā* ‘majestic tree’ ibid. 52).

²² While *h* in Hrs., Mhr. and Soq. may continue **š*, seemingly supported by Soq. *širohom* (#6), *h* in Jib. continues only **h*, confirming the protoform reconstruction with **h-* (to which the Arb. and, to a lesser degree, Akk. parallel also contribute), thus proving that the two partially — and misleadingly — similar Soq. forms, *hərcm* and *širohom*, are unrelated. One wonders if the meaning ‘tree’ in MSA could be a shift from Proto-Afras. ‘wood’, which could have given rise to Cush. **hVram-* ‘ashes’: E.: Afar *rama*, LEC: Arbore *rómm*, Elmolo *rôm*, Yaaku *hroon*, pl. *hroómê* (allegedly from Masai).

- < Afras. *čín- ‘two’: Egyp. (Pyr.) *sn*; Brb. **sin*; C. Chad.: Gisiga *mu-čunay* ‘the second’ (ADB).
- (2) Bib. *tārē(y)n*; Pal. *tray(y)n*, *tryy*; Syr. *tārē(y)n-*; Mnd. *trin*; Urm. *tārī*; Hob. *itróh*; Hrs. *terō*; Mhr. *terōh*; Jib. *troh*; Soq. *tiro* // < Sem. **tiray/w-* (ADB).
- (3) Gez. *kālʔe*; Tna. *kālattä, kältä*; Tgr. *kālʔot m., kālʔe f.*; Amh. *hulät*; Arg. *ket, hāʔat*; Gaf. *alattä*; Sod. *kitt*; Har. *koʔot*; Wol. *hoyt*; Cha. *kʷet* // < Sem. **kilʔ-* ‘both’-: Akk. *kilallū*, Ugr. *kla-t* ‘both’, Hbr. *kilʔ-ayim* ‘of two kinds’, Arb. *kilā*, Sab. *klʔy*, Soq. *keʔala* ‘both’ (ADB).
- < Afras. **ki/alVʔ-* ‘the two equal, one and the other’: Egyp. (Pyr.) *ky* ‘other’ (if < **kVl-*); C.Chad.: Mada *kal* ‘to be equal’, Mofu *kal kal* ‘pareil’ (considered by some Chadacists a loan from Fulfulde); S.Cush.: Burunge *kalel-* ‘to be similar’ (ADB).
- Common North and West Semitic **tīnay-* (#1)²³.
- Common West (Arm.) and South Semitic **tīray/w-* (#2)²⁴.

92. WALK

- (1) Akk. *alak-*; Ugr. *hlk*; Hbr. *hlk*; Pho. *hlk*; Bib. *hlk*; Pal. *hlk* // < Sem. **hlc-*: Arb. *hlc* ‘périr’, Tgr. *halkä* ‘s’efforcer de, mourir’, Tna. *halakä* ‘se fatiguer; aller et venir d’un endroit à l’autre sans raison’, etc. (DRS 413)
 - Most likely, with the root extension **h-* (see Mil. RE) < Afras. **lkw ~ *lwk* ‘to walk’: C.Chad. **lakw-*: Buduma *lokúí* ‘Wohnsitz verändern’, Bura *laku*, Kilba *lākù*, Ngwahyi *lakù*, Banana *lakù* ‘road’; N.Omot.: Basketo *lukk-*, Dokka *luk(k)-* ‘to walk’ (ADB)²⁵.
- (2) Pal. *ʔzl* (syn.); Syr. *ʔzl*; Urm. *ʔzl* // < Sem.: Hbr. *ʔzl* ‘go away, disappear’ (HALOT 27), Arb. *zll* ‘glisser sur un sol glissant; passer rapidement, passer, se terminer (vie)’ (BK 1 1002).
 - < Afras. **(ʔa)ʔVl-* (?) Eg. (ME) *zny* ‘come’ (if < **zly*) or > *zʔ* ‘betake oneself to’ (if < **zl*) or both; Brb. **HazVl* ‘run’: Qabyle *azzal-*, Ahaggar *ahal*, Ayr *azal*, etc.; C. Chad. **ʔVl-*: Higi *zulu-tò*, Higi-Nkafa *zulu*, Kapsiki *zul-tè* ‘follow’, Logone *zalan* ‘road’ (ADB; cf. St. 2009 #468).
- (3) Mnd. *sga* // < Sem. **šxgy*: Syr. *ʔasgā* ‘ivit’ (Brock. 458), Hbr. *šgy* ‘to stray, stagger; do wrong’ (HALOT 1413), Pho. *šgy* ‘to roam, stray’ (HJ 1109), Arb. *šgy* ‘s’en aller, s’eloigner de quelqu’un’ (BK 1 1195), Mhr. *hegg* (*h* < **š*) ‘to stray, wander aimlessly’ (JM 96).
 - < Afras. **šVg-* ‘to roam, stray, err’: Chad. W.: Hausa *záagà* ‘go around’, Ngizim *žàg-dú* ‘move off road, go round’, C.: Mofu *žəg* ‘to err, go without aim’, Mafa *žəgám-žəgámmaža* ‘to be lost’, E.: Mawa *sugun* ‘to move (bouger)’ (distributed in different ways among the entries in St. 2007 #481, 486, 488).
- (4) Qur. *mšy*; Lbn. *māša*; Mec. *mišiy*; Mlt. *meša* // No Sem. etymology that I could find.
 - Cf. Egyp. (NK) *mšʕ* ‘to march, journey, travel’, compared in EDE III 619–20 and Tak. 2011 55 with its exact match Arb. *mšʕ* ‘marcher doucement’ (root extension -ʕ both in Egyp. and

²³ This root also occurs with derived meanings in those languages where the numeral “2” is conveyed by the forms continuing **tīray/w-*: Off. Arm. *tnyn* ‘second’ (HJ 1223), Syr. *tenyān-* ‘secundus’ (Brock. 828), MSA **mi-tnay* ‘(incisor) tooth’ (see No. 89 ‘tooth’ #3 above) and Jib. *məṭənyéh* ‘ring twisted like the strands of a rope’ (JJ 285) as well as in Eth. where “2” is conveyed by **kilʔ-* (Gez. *sanuy* ‘the second day of the week or month, two days’, etc. LGz 509).

²⁴ One wonders if **tīray/w-* (reflecting Afras. **čiray-*) could continue Afras. **cir(ay)-* with **c-* > **č* by contamination with **tīnay-*, cf. Chad. **cVr-* ‘two’ (so St. 2009 #632 where some forms seem to point to **čir-*): W.: Paa *čírù*, Jimbin *šír*, Ngizim *šírín*, etc., C.: Podoko *sará*, etc., E.: Somray *sir*, Mokilko *siré*, etc.

²⁵ Plausibly related to Afras. **likw-* ~ **luk-* ‘leg, foot, hip, thigh’: Sem.: Akk. (YB) *ilku* ‘hip area’ (CDA 127); C.Chad.: Muyang *kō-lōkó* ‘hip’ (with a body-part prefix); Cush. C. **likw-* ‘leg, foot’: Bilin, Khamir, Qwara, Dembea *likw*, Khamta *lukw*, Kemant *läkwə*, Aungi *lakw*, E.: Saho, Afar *lak* ‘hip, thigh, leg’, LEC **lukk-* ‘leg, hip’: Somali *lug*, Oromo *luka*, Baiso *lukka*, Arbore *lukk*, Gidole *lukkə-to*, HEC: Sidamo *lekkee*, Burji *lukka*, etc., Dullay **luk-* ‘leg, foot’: Gawwada *luhte*, Dobase *luh-*, etc., S.: Dahalo *luka* ‘thigh’ (ADB).

Arb.?) and W.Chad.: Tangale *mεε* ‘journey, travel, walking’; although *s* in Tangale may continue Chad. *š (< Afras. *ĉ), the term may be an Arabism. Could Afras. *mĉy/ɣ be reconstructed on this basis?

- (5) Sab. *mḏ?* (< *mz?) // < Sem. *mṭ?: Ugr. *mz?* ‘meet, run into’ (DUL 608), Hbr. *mš?* ‘reach; meet accidentally’ (HALOT 619), Mnd. *mṭa* ‘arrive, reach’, Gez. *mašʔa* ‘come, happen to, etc.’, Tgr. *mäsʔa* ‘come’ (LGz 370), Jib. *mīḏī* ‘reach’ (JJ 169), Soq. *mṭy* ‘venir, arriver, atteindre’ (LS 241), etc. — see also No. 16 ‘come’ (#2) in Mil. 2010.
- (6) Gez. *ḥwr*; Gaf. (a)horä; Har. ḥāra; Cha. *wərwər, wärä* // < Sem. *ḥwr: Arb. *ḥwr* ‘revenir, retourner’ (BK 1 509).
- < Afras.: Eyp. (Pyr.) *ḥry* ‘be far from; go far from, move away’, *ḥr.t* ‘road’; E. Chad. *HVr- ‘go’: Nanchere *āre*, Gabri *ōre*, Kabalai *ārrā*, Dormo *erauá*; E.Cush.: HEC: Sidamo *haʔr* ‘go’ (ADB).
- (7) Tna. *kädä, kedä*; Tgr. *kedä*; Amh. *heddä*; Arg. *heda*; Wol. *hedä* // < *kyd ~ *kdkd: Arb. *kdkd* ‘marcher lentement et lourdement’ (BK 2 875), cf. also *kawdaʔa* (a somewhat odd form) ‘se mettre à courir’ (ibid. 941)
- Cf. Afras.: Cush. E.: Saho *ku(u)d-*, Afar *kuud-* ‘run (away)’; N. Omot.: Dizi *kudu* ‘road’ (ADB).
- (8) Sod. *alläfam* // < Sem. *ḥlp ‘to pass’: Arg. *halläfa*, Tgr. *ḥalfa*, Gez. *ḥälafa*, Hbr. *ḥlp* ‘pass’, Arb. *ḥlf* ‘follow, succeed’, etc. (LGz 260).
- (9) Jib. *ayad* // Arb. *wyḏ* ‘marcher d’un pas régulier’ (BK 2 1571). Borrowing into Jib. is also possible.
- (10) Soq. *ʔód* // < Sem. *ʔdw/y: Hbr. *ʔdy* ‘walk along’, Jud. Arm. *ʔdy* (pe.) ‘pass by’, Arb. *ʔdw* ‘walk, pass by’, etc. (HALOT 789).
- < Afras. *ʔad- ‘walk, come and go’: Brb.: Zenaga *edda*, Senhadja *addu*, Semlal *əddu* ‘go’, Izayan *əddu* (*d*) ‘come’, etc.; Chad. W.: (?) Sha *du-* ‘come’, C. *wuHad-: Hwona *wúḏà* ‘go out’, Bachama *wudó*, Bata *wādo* ‘go’, E. *Had-: Dangla *àḏē* ‘follow’, Migama *ʔáḏ-*, Bidiya *ʔàḏ* ‘go’, Mokilko *ʔáḏḏē* ‘(make) go away’, *ùḏó* ‘come’; Cush. E.: Saho *-ad-/ed-*, HEC: Somali *aad-* ‘go’ (questionable since the expected ʔ is missing), S.: Asa *adi* id., Dahalo *ḏaaʔ-* (met.?) ‘leave in a hurry’; Omot. N.: Male *ad-* ‘come’ (Bla. Om. #16.4. after Fleming), S.: Ari *ad-*, Ubamer *aad-* ‘come’ (ADB).
- (11) Soq. *ṭaher* (syn.) // Quoted as ‘to walk’ by native speakers (Kog. MS) alongside *ʔód*, while in LS translated as ‘aller, s’en aller’ and compared with Jib. *ḏhr* ‘sortir’ and Arb. *ḏhr* ‘paraître, sortir contre quelqu’un’; an alternate comparison, also plausible both phonetically and semantically, is with Arb. *ṭrr* ‘faire marcher d’un pas accéléré (les chameaux)’ (BK 2 264).
- ◇ Hob. *syóor*, Hrs. *seyōr*; Mhr. *səyūr* are most likely from Arab. *syr* ‘aller, se rendre vers...’ (BK 1 1390)
- Common North and West Semitic *h_lk (#1).

93. WARM

- (1) Akk. *emm-*; Hbr. *ḥām*; Pal. *ḥmym, ḥmyn*; Syr. *ḥamīm-* // < Sem. *ḥamm- ‘heat’: Arb. *ḥamm-*, Ugr. *ḥm*, etc. (HALOT 325).
- (2) Akk. (syn.) *šahān-* (v.); Ugr. *šḥn* (v.); Syr. *šḥen*, Urm. *šāḥīn-*; Lbn. *səḥni*; Mlt. *šḥūn*; Jib. *šḥan* // < Sem. *šḥn ‘become warm, warm oneself’: Arb. *šḥn*, Gez. *səḥna, saḥana* (LGz 495) etc., Hbr. *šəḥīn* ‘ulcer, inflamed spot’ (HALOT 1462, 1460).
- Cf. S.Omot.: Ongota *šōni* ‘warm’ (ADB).
- (3) Mec. *ḥārr* // < Sem. *ḥrr ‘be hot’: Hbr. *ḥrr, ḥry*, Arb. *ḥrr*, Gez. *ḥarra*, Jib. *ḥar* ‘heat’, etc. (LGz. 243), see also No. 12 ‘burn’ (2) in Mil. 2010.

- (4) Gez. *məwuk*; Tna. *məwuk*; Tgr. *moqa*; Amh. *muk*; Arg. *ʔamo*; Gaf. *moḳä*; Sod. *muḳəṇna*; Cha. *m^wäk*; Har. *muk*; Wol. *muḳamuk* // Seems to be Eth. only; cf. Gez. *maḳaḳa* ‘burn, sting (e.g. medicine), burn the throat, cauterize’, Tna *mäḳmäḳ bälä* ‘burn (pepper)’ (LGz 355); cf. also with metathesis Akk. *ḳamû* ‘burn (up)’ (CDA 283).
- Cf. Egyp. (18 Dyn.) *mḳḳ* ‘roasts on a spit’ (secondary -ḳ-?); N.Omot. **muk-* ‘ashes’: Dache, Gidicho, Haruru *muḳo*, Gidiche *muḳa*, etc. (compared in EDE III 186–7).
- (5) Tgr. *ḥəfun* (syn.) // No etymology that I could find.
- Cf. Afras. **faḥ-* ~ **HafVw-* ‘fire; burn’: Egyp. (late) *wfḥ* ‘burn’; Brb. **HafVw* ‘fire’: Ghadames *ufa*, Adghaq *efiw*, Izayan *afa*, Nefusa *tə-faw-t*; Chad. W. **fwa/yaH-*: Diri *fáyá* ‘fry’, Zaar *vwa* ‘burn’, Kulere *fyef* ‘roast’, C. **ʔafwaH-*: Kapsiki *nu-fu* ‘boil’, Gisiga *ʔavo* (< **ʔaf-*), Munjuk *afu*, Logone *fùù*, Musgu *àfú* ‘fire’, E.: Kwang *èpè* ‘burn’ (ADB).
- (6) Hob. *huúb*; Hrs. *hwob* // < Sem. **hawb-*: Mhr. *hūwəb*, Jib. *šəheb* ‘se rechauffer’, Tgr. *hawhaw belä* ‘flamber’ (perhaps an Arabism), Arb. *hawb-* ‘intensité, ardeur du feu’ (DRS 383).
- Cf. Afras.: Cush. N.: (?) Beja *nəbʔa* ‘warm’ (adj.), *tibʔa* ‘burn’ (< **n/tV-Hibʔ-?*), E.: HEC **Hibb-* ~ **Hayb-* ‘warm’ (adj.): Sidamo *iibb-ado*, Hadiya *iibba(lla)*, Burji *ḍayb-aḍ-* (vb.) < *ḍa-ib-aḍ-* (ADB).
- (7) Mhr. *gōna* // < Sem. **gawn-*, var. **gVnʔ-* ‘heat’: Jib. *egnúʔ* ‘to protect from cold’, *gúneʔ* ‘sheltered, warm place’ (JJ 77), Arb. *šawn-* ‘jour; rouge foncé’, *šawnat-* ‘rouge; soleil, disque du soleil, rouge à l’heure où il se couche’ (BK 1 359).
- < Afras. **gawn-* ‘heat; fire’: Chad. **gawn-*: C.: Gude *gúnə* ‘fire; heat’, Fali-Bwagira, Bata *gun* ‘fire’, Makeri *ganun* ‘mobile metal fireplace’, E.: Sokoro *gáun* ‘to cook’ (ADB, St. 2011 #659), cf. also **gVn-* ‘smoke’ (ibid.). Comparison (ibid.) with Egyp. (Sarc.) *ḏnḏn.t* ‘fire’ is valid if the latter is < **gnḡn*.
- (8) Soq. *di-yenáḥa* // < MSA: Mhr. *hə-nḥū*, Jib. *e-nḥé*, Soq. *ə-nḥi*, see No. 12 ‘burn’ (#13) in Mil. 2010.
- ◇ Hrs. *ḥrōḳ* is an Arabism. No term registered in Pho., Bib., Mnd., Qur. (in classical Arabic outside the Qur’an it’s *suḥn-*) and Sab.
- Common Semitic **šḥn* ‘become warm, warm oneself’ (2).

94. WATER

- (1) Akk. *mū*; Ugr. *my* (*mym*, *mh*); Hbr. *mayim*; Pal. *may(y)ʾn*, *myy*; Syr. *mayyē*; Mnd. *mi-*; Urm. *məy-*; Qur. *māʔ-*; Lbn. *may*; Mec. *mayya*; Mlt. *ilma* (*il-ma*, with the article); Sab. *maw*, *mwy*; Gez. *may*; Tna. *may*; Tgr. *may*; Har. *mīy*; Wol. *mäy*; Hob. *ḥmóh*; Hrs. *ḥemyoh*; Mhr. *ḥemūh*; Jib. *mih* // Sem. **maʔ-* with **may-* and **mah-* variants
- < Afras. **maʔ-*: Egyp. (OK) *maw* ‘water’; Chad. W.: Guruntum *mà* ‘water’ (with various meaning shifts: Dera *màyé* ‘to bathe’, Angas *mwee* ‘sap, juice’, Ankwe *mua* ‘a liquid substance’, etc.), C.: Gude *màʔinə*, Bachama *màʔyin* ‘water’, etc.; Cush. N.: Beja *muʔ-* ‘liquid’, C.: Khamir *maw* ‘fliessig werden’, E.: LEC: Somali *māy-ad* ‘tidal waters’ (cf. *māh-* ‘eau courante’), S.: Iraqw, Alagwa, Burunge *maʔay*, Asa *maʔa*, Qwadza *maʔaya*, Maʔa *maʔi*, Dahalo *maʔa* ‘water’ (ADB; cf. EDE I, 120)²⁶.
- (2) Arg. *ʔofa* // No etymology that I could find.
- Cf. Afras. **pi(ʔa)ḥ-* ‘rain’: Egyp. (Pyr.) *pʔḥ.t* ‘sky waters’; Chad. C.: Musgu *pīáḥ*, Higi-Hye *gə-fiyà* ‘rainy season’ (cf. also Tera *viya*, Bura *viyà*, Hildi *vwiya*, etc. id.; according to Stolbova, irregular reflexation of the labial is due to a prefix), E.: Kera *pépé* ‘God, rain’ (ADB).
- (3) Soq. *riho* // < Sem. **rVwVy-* ‘abundant water; watering, irrigating’: Soq. *re* ‘drink’, Hbr. *rwy* ‘to drink one’s fill’, (pi.) ‘to give to drink abundantly, water thoroughly (e.g. the furrows)’,

²⁶ Brb **Ham-an* (pl.) ‘water’ belongs to a different Afras. root **ham-*, cognate with Arb. *ḥmw*, *ḥmy* ‘to pour’; Chad. W. **ham-* ‘water’, E. **ʔam-* ‘water, rain’ (ADB).

rāwā ‘well watered, irrigated’, Syr. *rəwū* ‘be satisfied with drink, be drunk, wet, flooded’, Arb. *rwy* ‘to drink one’s fill, be watered’, *ri/ayy-* ‘irrigation, network’, *riyy-* ‘quenching of one’s thirst’ (HALOT 1194–5), Sab. *ryh-twiy* ‘provide a water-supply’, *rwym* ‘tank, cistern’, *mrw*, *mryt* ‘irrigation system’ (SD 119–20), Gez. *rawaya* ‘to drink one’s fill, be watered, irrigated’.

- < Afras. **riway/?- ~ *?Vraw/y-* ‘abundant water; watering; rain; river’: Egyp. (MK) *īwy* to water (field), pour out (liquid) (if < **rwy*); Chad. W.: Hausa *rúwá* ‘water, rain’, Galambu *rwaà*, Geruma *ará*, C.: Mbara *rè*, Musgu *eré*, *ree* ‘river’, Munjuk *aray* ‘canal’, E.: Mokilko *ǎró* ‘to drink’; Cush. C.: Aungi *ári* ‘rain’ (according to App CDA, < Omot.), E.: LEC: Arbore *ǎriy*, Dasenech *ǎir*, Dullay: Gawwada *irrawo*, Harso *irrawo*, etc. ‘rain’, S.: Alagwa, Burunge *raʔu* ‘dew’, Maʔa *re* ‘rain’ (according to Fleming), ‘mare’ (according to Ehret); N.Omot.: Wolaita *ǎira*, Yemsa *irō*, Chara *ira*, Dizi *iru* ‘rain’ (ADB).
 - ◇ Amh. *wəha*, Gaf. *ägä*, Sod. *yiga*, Cha. *ǎka* are considered as Cushitisms: cf., on one hand, Cush. E.: HEC (a source of borrowing for Amh.?): Sidamo *waa*, Hadiya *wóʔo*, on the other, C. (a source of borrowing for Gaf., Sod. and Cha.?): Bilin *aqu*, Khamta *āqw*, Aungi *ayhu*. No term registered in Pho. and Bib.
- Common Semitic **maʔ-* (1).

95. WE

- (1) Akk. *nīnu*; Ugr. *nḥn*; Hbr. *anaḥnū*; Pho. *ʔnḥn*; Bib. *ʔānaḥnāʔ*; Pal. *ʔānan*; Syr. *ḥənan*; Mnd. *anin*; Urm. *ʔaḥənan*; Qur. *naḥnu*; Lbn. *nəḥn(i)*; Mec. *ʔiḥna*; Mlt. *aḥna*; Gez. *nəḥnā*; Tna. *nəḥna*; Tgr. *ḥəna*; Amh. *ʔəñña*; Arg. *ənnā*; Gaf. *ənni*; Sod. *əñña*; Har. *əñña*; Wol. *iñña*; Cha. *yəna*; Hob. *nḥáh*; Hrs. *neḥa*; Mhr. *nḥah*; Jib. *nḥan*; Soq. *ḥanhen* // Sem. **(ʔa)naḥnu ~ *(ʔa)ḥan(an)* ‘we’.
 - < Afras.: Egyp. (Pyr.) *n* (dependent), (NE) *inn* (independent, Coptic *anon*); Brb. **nukni*: Ghadames *nekkənīn*, Ayr *nəkkəni*, Zenaga *nəkni*, Ntifa *nukni*, Qabyle *nəkʷni*, etc.; Cush. N.: Beja *hanín* (< **ḥanin*), C.: Bilin *yin*, Khamir *inaa*, Khamta *yīn* (according to Fleming), *iná* (according to Conti Rossini), Qemant *anniw*, Qwara *anän*, Aungi *ənnó-ǎi* (according to Hetzron), Kunfal *nʷa*, E.: Saho, Afar *nanu*, LEC: Somali *a/inna-ga*, Oromo *nu*, *nuy*, Baiso *no*, Arbore *ʔonó(-lo)*, Elmolo *ínó*, Dasenech *ḥiini* (excl.), Konso *ino*, HEC: Sidamo *ninke* (note suffixal *-k* as in Brb., though in a different position), Kambatta *noʔo*, Hadiya *nees*, Burji *naanuu*, Dullay: Gawwata, Tsamai *ine*, Yaaku *nīniʔ*, S.: Maʔa *nine*, Dahalo *nyányi*; N.Omot.: Male, Gofa, Oyda, etc. *nuni*, Zala *nu*, Zaise *nuú*, Chara *núni*, Gimirra (Bench) *nu(-na)*, Yemsa *inno*, Bworo *noo*, Kafa *no*, Dizi (Sheko) *inu*, Dizi (Nao) *natoknu* (note infixal *-k* as in Brb. in the same position), Mao (Hozo) *nunga* (ADB)²⁷.
 - ◇ No term registered in Sab.
- Common Semitic **(ʔa)naḥnu ~ *(ʔa)ḥan(an)* (1).

96. WHAT?

- (1) Ugr. *mh*; Hbr. *mā*; Pho. *m*; Bib. *mā*; Pal. *mh*; Syr. *mā*; Mnd. *mahu*; Urm. *mū(-dī)*; Qur. *mā*; Sab. *m*; Tgr. *mi/ə* // Sem. **ma(-h)*, variant **mi* (Tgr.)

²⁷ Of interest is a peculiar phenomenon with a more or less similar pattern, obviously distributed across and independently from the genetic grouping: 1) to Sem. **(ʔa)naḥnu* (Akk., Ugr., Hbr., Pho., Bib., Qur., Lbn., Gez., Tna., Hrs., Mhr., Jib.): Afar *nanú*, Dasenech *ḥiini*, Burji *naanuu*, Yaaku *nīniʔ*, Maʔa *nine*, Dahalo *nyányi*, Male, etc. *nuni*, Zaise *nuú*, Chara *núni*, Mao (Hozo) *nunga*; 2) to Sem. **(ʔa)ḥan(an)* (Syr., Mnd., Urm., Mec., Mlt. Tgr., Soq.): Beja *hanín*, Qemant *anniw*, Qwara *anän* (< **Han(a)n* where *H* may represent either **ʔ-* or **h-*), the other Agaw forms may be from **ḥin-* or **ʔin(n)-* like in Egyp. (NE); 3) to Brb. **nukni*: Sidamo *ninke*, Dizi (Nao) *natoknu*. Does this distribution reflect any system or at least its remnants or is it merely accidental?

- < Afras. **mali* or **may* ‘what?’: Egyp. (OK) *m*; Brb. **may(t)* ‘what?’ (sometimes ‘who?’): Fodjaha *metta*, Ghadames *mi*, Ahaggar, Adghaq, Zenaga, Baamrani, Izayan, Senhadja, Mzab, Nefusa, etc. *ma*, Izdeg *may*, Rif *mi*, Shenwa, Snus *matta*, etc.; Chad. **may*: W.: Hausa *mè*, Angas *mai*, Karekare, Ngamo *miyà*, Kirfi *màamí*, Geruma *màama*, Miya *mà*, Fyer *mi*, etc., C.: Tera *nə-ma*, Boka, Bura *mi*, Wamdiu *miya*, Mbara *ʔámá*, Mandara, Mofu *me*, Munjuk *ma*, Sukun *ma...wa*, Gude *mya*, Bata *muno*, Bachama *munə*, Buduma *meni*, Zime *má*, etc., E.: Lele *me*, West Dangla *man*, East Dangla *ma*, Migama *ʔini-mée*, Bidiya *ma*, Birgit *mintu*, Mokilko *moo-me*; Cush. E.: Afar *ma-ħa*, LEC.: Somali *ma-ħ*, Oromo *ma-al*; Dasenech *métè*, Arbore *me*, *meh* (< **mV-ħ*), Elmolo *máa*, Baiso *me*, HEC: Hadiya *maha* (< **ma-ħ*); Sidamo *ma(a)*, *mai*, Burji *miya*, etc. Dullay: Gawwata *moʔo*, Tsamai *mō*, etc., S.: Iraqw *mi-lá*, Qwadza *homi* (< **ho-mi*), Dahalo *má-ka*; N.Omot.: Zayse *ʔáal-ma* (cf. *ʔála* id. in closely related Ganjule), Mocha *ámo* (ADB). For a more detailed review, see EDE III 9–13.
- (2) Akk. *mīnu*; Gez. *mənt*; Tna. *məntay*; Amh. *mən*; Arg. *mən*; Gaf. *mən*, *məni*; Sod. *mən*; Har. *min*; Wol. *mən*; Cha. *mər* // < Sem. **min-*.
- Though probably accounted for by a secondary parallel development, the following forms in the **mali* entry (#1) deserve mention here: Chad. C.: Bata *muno*, Buduma *meni*, Bachama *munə*, E.: West Dangla *man*, Birgit *mintu*; E.Cush.: LEC: Konso *maana* (see more examples of **m-n* in EDE III 9–13).
- (3) Lbn. *šū*; Mec. *ʔayš*; Mlt. *še*; Hrs. *hašen*; Mhr. *hēšən*²⁸ // < Sem. **šayʔ-* ‘thing’²⁹: Hrs. *šī*, *šay*, Mhr. *šī*, Jib. *šé*, Soq. *šī*, *šīʔ* ‘thing’ (JM 150), Arb. *šayʔ-* ‘thing; what?’
- < Afras. **čayʔ-* ~ **ʔič-* ‘thing’: Egyp. (Old) *iš.t* ‘thing (Sache, Habe, Besitz)’, *išs.t* (Pyr.-MK; -šs probably conveys the affricate *č*) ‘what?’; C. Chad.: Masa *šé* ‘thing’ (can reflect both **č* and **š*: see St. 2005 #32); cf. S. Omot.: Ongota *saay* ‘who?’ (ADB).
- (4) Hob. *inīh*; Jib. *ʔinē*; Soq. *inim* (**ʔini-maʔ*) // < MSA **ʔin-* < Sem. **ʔay(y)-Vn-*: Arb. *ʔayn-* ‘ou?’ (BK 1 75), Hbr. *ʔayin* (always with *min*) ‘whence?’ (HALOT 42), Syr. *ʔaynā* ‘what?’, Gez. *ʔaynu* ‘what? which?’, Tna *ʔayyän* ‘which?’ (LGz 49).
- Cf. Afras. ‘what?’: Egyp.: Coptic (Fayumic) *ʔun*; Chad. W.: Kiir *wun*, C. **ʔawVn-*: Dghwede *awe/wini*, Glavda *ʔəwninà*, Zulgo *wana*, Logone *ɣwani*; Cush. N: Beja *nān* (redupl.), E.: Yaaku (*di*)*nyoh* (< **-nVy-Vh*); Omot. N.: Basketo *ɔɔ*, Anfillo *aanto* (unless < **am-t-* with assimilation), S.: Ongota *neeni* (redupl.)³⁰.
- Common North and West Semitic **min-* (2).

97. WHITE

- (1) Akk. *pešū* // < Sem. **pVšVh-* ‘be bright’: Hbr. *pšh* ‘be cheerful, happy’ (HALOT 953), Syr. *pəšīh-* ‘splendens; laetus, hilaris’ (Brock. 587), Mnd. *pših-* ‘bright, clear; cheerful’ (DM 376), Arb. *fšh* ‘apparaître à quelqu’un dans tout son éclat’ (BK 2 599)³¹.
- Most likely related through metathesis with Chad. W. **puHaç-* ‘sun’: Sura, Angas *pūs*, Montol *pəɣəs*, C. **paç-* ‘sun’: Mandara *vačia*, Glavda *fáčiy*, Guduf *fáčiyá*, Mofu *pás*, Gisiga *pas*, Matakam *pâac* (ADB). Both < Afras. **pVçVh-* ~ **puħaç-* ‘bright sun?’

²⁸ With *ʔay* (< *ʔayyu šayʔ-* ‘which/what thing?’) in Mec. and the preposition *h-* ‘to, for’ and (*ə*)*n* “an obsolete marker of oblique case” (JM 150) in Hrs. and Mhr.

²⁹ While the idea of rendering ‘what?’ in Mhr./Hrs. by means of the noun ‘thing’ is likely to be accounted for by the influence of Arabic, the MSA noun itself does not seem to be an Arabism (even an early one, in view of *š* and attestation in all MSA languages).

³⁰ Cf. also Omot. **ʔay/w-ni* ‘who?’: Cancha *ʔooni*, Gimirra (She) *onə*, Yemsa *oon*, Banna *ayne*, etc. alongside **Hay/w-* id.: Basketo *ʔo*, Mao (Hozo) *haya*, Ari (Bako) *ayi*, Galila *háyi*, etc.

³¹ A different and erroneous etymology suggested in SED I #43 had to be corrected for this traditional one after a convincing argumentation had been published by M.Bulakh in Bul. 2003 4.

- (2) Ugr. *lbn*, *labanu*; Hbr. *lābān*; Pho. *lbn*; Mnd. *lbina*; Hob. *lboón*; Hrs. *elbōn*; Mhr. *əwbōn* (pl. *lēbān*); Jib. *lūn*; Soq. *líbehon* // < Sem. **lbn* ‘be white’: Mnd. *lbyn?* ‘white’, Arb. *laban-* ‘milk’ (HALOT 517).
- < Afras. **li/ab(-an)-*: Egyp. (Pyr.) *nb* ‘gold’ (if < **lb*); Chad. W.: Hausa *liibaa* ‘light-colored earth used for glazing’, C.: Affade *elboh* ‘chalk’, (?) Mada *élbésè* ‘mil blanc sp.’; E.: West Dangla *lambany* ‘variety of white sorgho’, Bidiya *lambàny*, Migama *làmbàny* ‘sorgho blanc sp.’ (St. 2005 #77; perhaps also **lVbV* ‘cloud’ ibid. #78 and **lVbV* ‘cotton’ ibid. #79). ADB.
- (3) Bib. *hiwār*; Pal. *hwwr*, *hywr*; Syr. *hewwār-*; Mnd. (syn.) *hiuar-*; Urm. *həwār-* // < Sem. **hw* ‘be black and white’: Hbr. *hw* ‘to grow pale’, Arb. *ʔaḥwariyy-* ‘blanc, qui a la peau blanc’, *ʔaḥwar-* ‘qui a les yeux grands, d’un beau noir, que le blanc qui les entoure...’, Mhr. *höwar* ‘black’, Soq. *hohar* ‘black’, *heyhor* ‘cloud’, etc. (ADB; cf. also Bulakh 2004 273–4).
- (4) Mnd. *shuhar* ‘white’ (syn.) // < Sem. *sh* ‘be light, bright, yellowish, reddish’: Hbr. *šāhōr* ‘white, yellowish-red, reddish grey (exact shade uncertain)’, Syr. *šəḥar* ‘to blush’, Arb. *ʔaṣḥar-* ‘yellowish-red’, etc. (HALOT 1019). A secondary formation with the root extension *-r* (see Mil. RE) < Sem. **šVh-* (see 7 below) can be presumed.
- (5) Qur. *ʔabyaḍ-*; Lbn. *ʔabyaḍ*; Mec. *ʔabyaḍ*; Mlt. *abyat* // < Sem. **bayṣ-* ‘egg’: Hbr. *bēšā*, Syr. *bīṣt-*, Arb. *bayḍat-*, etc. (cf. SED I #43)
- < Afras. **bayṣ-*: W.Chad. *(*m-*)*bwiṣ-* ‘egg’: Polchi *mbwâs*, Geji *mbúsi*, Mangas *mbóš*, Zaar *buùš*, Sayanchi *mbúš*, Buli *mbiš*, Tule *mbòòšə*, Wangday *mboš*, Tala *mbeš*, Zul *mbúše* (ADB); N.Omot. **bawṣ-* ‘white’.
- (6) Gez. *šāḏdā*, *šaḏadā*, *ḏaḏadā*; Tna. *šaḏda*; Tgr. *šaḏda* // No etymology.
- (7) Amh. *nāčč*; Gaf. *našwä*; Sod. *nāččä*; Har. *nāčih*; Cha. *nāčä* // < Sem. **nšh* ‘be pure, clean, brilliant’: Gez. *našha* ‘be pure, clean, etc.’, Arb. *nšh* ‘be pure, sincere’, Syr. *našah* ‘be brilliant, shine’, etc. (LGz 405), with the prefixed root extension *n-* < **šVh-* ‘clear, bright (day, weather)’: Hbr. *šhh* ‘be white’, *šāh* ‘clear, bright’, Syr. *šāḥih-* ‘clear, bright’, Arab. *šḥw/y* ‘become clear, serene (day)’, Gz. *šəḥwa* ‘be serene, bright (sky)’, Tgr. *šāḥa* ‘become clear, bright’, Tna. *šāḥe* ‘clear up (weather)’, Soq. *šhy* ‘be clear (weather)’ (LGz 553).
- < Afras. **čaḥ-* ‘be clear, bright, white’: (?) Egyp. (Pyr.) *ḥd(y)* ‘white’ (< **ḥš* and metathetic?); Chad. W.: Paʔa *càà/cú/câu* ‘get light (of town when light comes)’, C.: Fali-Kiria *cəhu* ‘clear’, Mafa *cáyaʔa*, *céyeʔe* ‘brilliant, shining’; S.Omot.: Ari *caa-mi* ‘white’ (ADB).
- (8) Arg. *zah*, *zahi* // < Sem. **zhy* ‘be brilliant’³²: Jud. *zihā* ‘reflected light, reflexion’ (Ja. 391), Syr. *zəhe*, *zahyā* ‘splendidus’ (Brock. 189), Mnd. *zha* ‘be bright, radiant, shining’ (DM 162), Arb. *ʔazhā* ‘brillant; fauve, rouge (se dit des dattes, quand elles mûrissent)’ (BK I 1024).
- (9) Arg. *zellāha* (syn.) // < Sem. **zlh* ~ **zhl* ‘be bright, clear, clean’: Tgr. *žəḥallāy* ‘stream with turbid water changing colour; übertr. von einem Heer; von der Haut des Elefanten’ (LH 548), Jud. *zəḥal* ‘to be bright, brighten up’ (Ja. 390), Mnd. *zhl*, *zlh*, *zla* ‘to pour down, pour out, cleanse, wash, sprinkle’ (DM 163).
- (10) Wol. *gumärä* // South Eth.: Wolane *gumärä*, Zway *gōmārä* id. Compared in LGur 279 with Amh. *gʷämärä* ‘become white, pale, mature (grain)’; not compared with Ennemor *gāmār*, Gieto *gāmār* ‘grey’ (cattle) (ibid. 278), although these terms seem quite compatible. Perhaps a meaning shift ‘white’ (and ‘grey?’) < ‘smoke colored’ < ‘smoulder, embers, firebrand’, cf. DRS 144: Akk. (NA) *gumār-* ‘bois carbonisé’, pB Hbr. *gimmēr* ‘fumer, fumiger’, Syr. *gāmūrt-*, Mnd. *gumr-* ‘charbon ardent’, Arb. *žamrat-* id., *žumrat-* ‘braise, tison’, *žmr* II ‘fumer, fumiger’, Soq. *migméreh* ‘casserole à encens’, etc.
- (11) Cha. *gʷad* // Probably < Eth.: Gez. *gahada*, *gəhda* ‘to be manifest, be said openly, be revealed, become visible’ (LGz. 185), Tgr. *gahād* ‘open, manifest’ (LH 559), Tna. *gāhadä* ‘to be

³² Otherwise, same as *zellāha* (< **zVllh-*, with *-lli* > *-y/i*, cf. LArg 7); see below (#9).

clear, plain, manifest, obvious' (KT 2200), Amh. *gadä* 'to be revealed, become visible' (K 2035), compared by M. Bulakh (in her doctoral dissertation) with Arb. *ʔaʒhada* 'paraître, se montrer' (BK I 342) < Sem. **ghd* 'be clear, visible'.

◇ No term registered in Sab.

→ Common West and South Semitic **la/ibān-* (2).

98. WHO?

(1) Akk. *mannu*; Bib. *man*; Pal. *man*, *mʔn*; Syr. *man(-nu)*; Mnd. *man*; Urm. *man*; Qur. *man*; Lbn. *meyn*; Mec. *miyn*; Mlt. *mīn*; Sab. *mn*; Gez. *mannu*; Tna. *mān*; Tgr. *mān*; Amh. *man*; Arg. *man*; Gaf. *man*; Sod. *man*; Har. *mān*; Cha. *m^wan*; Hob. *moón*; Hrs. *mōn*; Mhr. *mon*; Jib. *mun*; Soq. *mon* // Sem. **man-* (ADB; see also. EDE III 9–13)

□ < Afras.: Brb.: Rif *man*, Shawiya *mani*, Mzab *manayn*, etc. 'who?' (see also No. 96 'what?' #2).

(2) Ugr. *my*; Hbr. *mī*; Pho. *mī*; Arg. (syn.) *ma*; Sod. (syn.) *ma*; Wol. *ma* // < Sem. **mi/a* (< **may-*?). See also EDE III 9–13.

□ < Afras.: Cush. E.: LEC 'who?': Somali *-uma*, Arbore *ma(h)*, Elmolo *máa* (also 'what?'), Dasenech *maya*, S. 'who?': Iraqw *heemá* (*he* 'person'), Alagwa *miyyaa*, Burunge *miya* (see also No. 96 'what?' #1).

→ Common Semitic **man-* (1).

99. WOMAN

(1) Akk. *sinništ-* // One of the most enigmatic Semitic words. No etymology.

(2) Ugr. *ʔatt*; Hbr. *ʔiššā*; Pho. *ʔšt*; Pal. *ʔä(y)ttā*, *ʔth*; Syr. *ʔi(n)tət-*; Mnd. *ant-*; Sab. *ʔtt* (pl. *ʔnt*); Gez. *ʔanast*; Tgr. *ʔassit*; Arg. *anāšča*; Gaf. *ansatā* // < Sem. **ʔa/int-at-*.

□ Probably to be related with S. Omot. **Hinč-*: Dime *nic*, *nit^h*, Galila *yinč*, Ari *yinc-i* 'child', Ongota *ʔinta*, *hinta* 'man' (ADB).

(3) Urm. *niḳva* // < Hbr.-Arm. **naḳb-at-* 'woman' (< Sem. **nḳb* 'to drill, bore through', Akk. 'to deflower, rape'): Syr. *naḳbat-*, Hbr. *nəḳēbā*, etc. (HALOT 719).

(4) Qur. *ʔimraʔat-*, *marʔat-*; Lbn. *marā*; Mlt. *mara*; Sab. (syn.) *mrʔ-t* // fem. < Sem. **marʔ-* 'male human' (ADB): Akk. *māru* 'son, boy', Arab. *marʔ-*, *murʔ-* 'man, husband', Qatabanian *mrʔ* 'child', Sab. *mrʔ* 'man; male child, lord; social superior', Mhr. (pl.) *ḥə-mrō* 'children' (ADB).

□ < Afras. **marʔ-* 'man/woman, human being': Egyp (OK) *mr.w* 'people'; Chad. W.: Dera *murà* 'woman, sister', Dwot *meer* 'family; woman', Wangday *mar* 'person', Kulere, Fyer *mor* 'woman' Fyer *mará* 'male', *mor*, Kulere *mor* 'woman', Daffo-Butura *máàr* 'boy', *māra* 'masculine', C.: Bura *mara* 'co-wife', Gudu *már*, Nzangi *mùre*, Fali-Muchella *muru*, Bachama *mùré*, Bata *muré* 'man', E.: Dangla *ùmìrnè* 'person', Kera *mar* 'wife', Migama *máar-tù* 'my uncle', Mawa *mər* 'clan, family'; Cush. E.: LEC: Dasenech *mār-ti* 'girl', HEC: Sidamo *mure* 'child, baby'; Omot. N: Mao (Hozo) *mēri*, (Sezo) *méri* 'boy, child, son', S.: Ongota *mara* 'baby' (ADB; on possible broader connections see EDE III 357–358, 380–382).

(5) Mec. *ḥurma* // < Arb. *ḥurmat-* 'chose sacrée; famille d'un homme, surtout sa femme ou sa fille' (BK 1 415) < Sem. **ḥrm* 'forbid, dedicate, declare sacred' (see HALOT 353).

(6) Gez. (syn.) *bəʔəsīt* // fem. < *bəʔəs-i* 'man, male, husband, someone', Tgr. *bəʔəs* 'husband' (see No. 51 'man' (#7) in Mil. 2011)

□ < Afras. **bVʔVs-* 'son, boy, coeval': Chad. W.: Siri *beʔsi* 'son, boy', C.: Tera *bəsoŋka* (< **bəso-n-k?*) 'boys'; Cush. C.: Kemant *baas* 'beget, become father', S.: Qwadza *boʔos-iko* 'friend', Dahalo *ḥosóʔo* (met.) 'friend of same age, coeval'; N. Omot.: Dawro *biša* 'brother', Malo *biša* 'child', Yemsa *buʔsa* (pl.) 'boys', Kafa *bušō* 'boy, son', Mocha *buš(o)* 'child', Bworo (Shinasha) *bušoo* 'son' (ADB).

- (7) Tna. *sābäy-ti*; Amh. *set* // fem. < Tna. *sābʔay*; Amh. *sāb, säwu* < **sabʔ-*, see No. 51 ‘man’ (#9) in Mil. 2011.
- (8) Sod. *məšt*; Wol. *məšt*; Cha. *məšt* // No Sem. cognates that I could find. For the Afras. parallels see No. 51 ‘man’ (#11) in Mil. 2011.
- (9) Har. *idōč, indōč* // (-oč like in *abōč* ‘male’ < **ʔab-* ‘father’?) < Sem. **ʔVd-at-*: Ugr. *ʔdt*; Pho. *ʔdt*, Palm *ʔdt* ‘lady’³³.
- < Afras. **ʔa(n)d-* ‘elder female relative’: Chad. C. **ʔadi* ~ **dada* ‘grandmother’: Mbara *dādā*, Munjuk *adi*, Musgu *adā*, E.: Kera *ādīdā* ‘mother-in-law, grandmother, daughter-in-law’; Cush, N.: Beja *enda*, C.: Bilin *adē* ‘mother’, E.: LEC: Somali *eddo* ‘paternal aunt’, Oromo *adada* ‘aunt’, HEC: Sidamo *adaada*, Darasa *adaada*, Hadiya *adda*, Kambatta *adda* ‘mother’, Dullay: Gawwata *a-atitē* ‘elder sister’; N.Omot. **ʔind-* ‘mother’: Malo *indo*, Haruro *endo*, Yemsa *intoo*, Kafa *indo*, Mocha *ʔinde* (ADB).
- (10) Hob. *tētē*; Hrs. *teṭ*; Mhr. *tiṭ*; Jib. *teṭ* // No etymology, unless < **t-iṭ* < **tV-ʔiṭ-*, reflecting some relict form of **ʔa/inṭ-at-* (see above in #2) with a prefixed, not suffixed, feminine formant.
- (11) Soq. *ʔāže* // fem. < *ʔayg* ‘man’ (with palatalization of -g), see No. 51 ‘man’ (#12) in Mil. 2011.
- ◇ No term registered in Bib. Urm. *baḥṭa* (syn.) is from Kurdish.
- West Semitic **ʔa/inṭ-at-* (or West and South Semitic, if the hypothetical etymology in #10 holds water).

100. YELLOW

- (1) Akk. (*w*)*arḫ-* (also ‘green’); Ugr. *yrḫ* (greenish yellow); Pal. *yrḫ* (also ‘green’); Syr. *yūrāḫ-* (also ‘green’); Mnd. *yuraq-* (also ‘green, pallid’) // < Sem. **warḫ-* ‘yellow, green’ (v. in DUL 982; Bulakh 2003 8–10).
- < Afras. **wVraḫ-* ‘green, yellow’: Egyp (MK) *iḫ* ‘onion, greenery, green stuff’; Brb. **wVraḫ* ~ **rVwaḫ*: Ghadames *arəḡ* ‘be green’, Siwa *a-wray* ‘green’, Zenaga *yərəḡ* ‘yellow’, Ayr *irway*, Ahaggar *irway*, Semlal, Izdeg, Shawiya *awray*, Qabyle (Ait Mangellat) *awray* ‘be yellow’.
- (2) Hbr. *šāhōb* // The color is uncertain: HALOT 1007 gives ‘bright red (of hair)’, Ja. 1264 gives ‘yellow’ for pB Hbr.; in any case < Sem. **šhb* ‘to be light, yellow, light brown’: Arb. *ʔašhab-* ‘alezan (cheval); roux, fauve (chameau); blond, aux cheveux blonds’ (BK I 1379), Mhr. *šahwēw* ‘cream-coloured, creamy-white, fawn’ (JM 360), Jib. *šahbōb* ‘fawn, light brown’ (JJ 237), Soq. *šāḥbab* (with unexpected -ḥ- instead of *-h-) ‘blanc’ (LS 354).
- (3) Qur. *šafṛāʔ-* (f.); Lbn. *ʔašfar*; Mec. *šfar*; Mlt. *isfar*. Mhr. *šāfər* (‘yellow, green; brass’), Jib. *šəfrór* // Soq. *šəfḥər* ‘brass’ (JMhr 359). The MSA words are likely, but not certainly, Arabisms.
- (4) Gez. *beša, bašā*; Tna. *biča*; Amh. *bi/əča*; Arg. *bəča*; Wol. *bičä* // < Sem. **bišš-* ‘bright, brilliant’: Tgr. *bäyyäsä* ‘to be bright, brilliant’, Arb. *bšš* ‘briller, luire’, Soq. *bišiš* ‘blanc’³⁴.
- (5) Tgr. *šāgray* // also ‘greyish brown, light-coloured’ (LH 229), Tna. (Hamasen) *šagəray, šag^wəray* ‘dull gray, dark gray (animal’s coat), cow that has a white and light brown coat’ (KT 871). Either < Sem. **šag^wVr-*: Arb. *ʔasžar-* ‘qui a des taches ou des veines rouges au blanc des yeux (homme); oeil dont le blanc est injecté d’une teinte rouge; lion; étang dont l’eau est limpide’, *sažar-* ‘teinte rouge de sang dans le blanc des yeux’ (BK I 1053) or, according to Leslau 1990 178, a borrowing from Arb. *škr* ‘être de couleur alezan’ (BK I 1254) via Sudanic Arabic, where **ḳ* > *g* and the corresponding form is *ʔašgar* ‘gelbbraun, goldbraun’ (Reichmuth 1981 61)³⁵.

³³ Another possibility is a loan from Cushomotic **ʔind-* ‘mother’ (cf. LHar 19, LGur 18). Tna. *addē* ‘mother’ is either from Sem. **ʔVd-at-* or from Cushomotic **ʔind-*.

³⁴ A correct etymology, published in Bulakh 2003:4, which has convinced the authors that their comparison in SED I #43 of the Eth. terms for ‘yellow’ with Sem. **bayš-* ‘egg’ was wrong.

³⁵ I am indebted to Dr. M. Bulakh for this comment (see footnote 31).

- (6) Sod. *weṭa*; Cha. *weṭa* // Arb. *waḍaha* ‘être manifeste, évident, clair’, *wāḍih-* ‘clair, évident, manifeste; brillant, éclatant de blancheur ou de lumière’, *waḍah-* ‘tout ce qui paraît, brille et se laisse voir distinctement; éclat de l’aurore ou de la lune qui paraît; tache blanche au front ou aux pieds d’un cheval; marqui imprimée sur la peau d’un cheval; canitie, cheveux blancs; lait; lèpre’ (BK 2 1553), *mutawaḍḍih* ‘blanchâtre (chameaux)’ (ibid. 1554).
- (7) Hrs. *hežôr*; Mhr. *hažáwr*; Jib. *šažrór* // Soq. *šézar* ‘green’ (Kog. MS) < MSA *šVžar-. It is tempting to relate this word to Arb. *ṣaḥḍar-* (see No. 35 ‘green’ 3), but MSA *š- (> Hrs. and Mhr. *h-*) does not correspond to Arb. *ḥ*, and there is nothing to do about it. One of the “toughest” cases of Semitic etymology. Contamination may be suspected, but I have so far failed to find any MSA forms by analogy with which *ḥ- could have turned into *š, or any Arabic form by analogy with which *-s- could have turned into -ḥ-.
- ◇ Urm. *zārdā* is an Iranism; Soq. *kirkam* (and *kérkham* ‘safran’ LS 225) is considered a borrowing from Arb. *kurkum-* ‘safran indien, curcuma; henna’ (BK II 888), the same must be true of Hob. *karkmí*. No term registered in Pho., Bib., Sab., Gaf. and Har.
- Common North and West Semitic **wark-* (1).

Literature

- ADB — *Afrasian Data Base*: <http://starling.rinet.ru>.
- AHw — SODEN, W. von., 1965–1981. *Akkadisches Handwörterbuch*. Wiesbaden.
- App CDA — APPELBYRD, D., 2006. *A Comparative Dictionary of the Agaw Languages*. Köln.
- Baet. — BAETEMAN, J., 1929. *Dictionnaire amarigna — français suivi d’un vocabulaire français — amarigna*. Dire-Daoua (Ethiopie).
- BK — BIBERSTEIN-KAZIMIRSKI, A. de., 1860. *Dictionnaire arabe-français*. Paris.
- Bla. Om. — BLAŽEK, V., 2008. Lexicostatistical comparison of Omotic languages. In *Hot Pursuit of Language in Prehistory*. Ed. by John D. Bengtson. Amsterdam-Philadelphia, 57–148.
- Brock. — BROCKELMANN, C., 1928. *Lexicon Syriacum*. Halle.
- Bulakh 2003 — BULAKH, M., 2003. Etymological Notes on the Akkadian Colour Terms. *Studia Semitica* (FS A. Militarev). Moscow, 3–17.
- Bulakh 2004 — BULAKH, M., 2004. Color Terms of the Modern South Arabian Languages: A Diachronic Approach. *Babel und Bibel I*, 269–282.
- CAD — OPPENHEIM, L., E. REINER & M.T. ROTH (ed.), 1956–2010. *The Assyrian Dictionary of the Oriental Institute, the University of Chicago*. Chicago.
- DM — DROWER, E.S. & R. MACUCH, 1963. *A Mandaic Dictionary*. Oxford.
- DRS — COHEN, D., 1970–. *Dictionnaire des racines sémitiques ou attestées dans les langues sémitiques*. La Haye.
- DUL — OLMO LETE, G. & J. SANMARTÍN, 2003. *A Dictionary of the Ugaritic Language in the Alphabetic Tradition*. Leiden / Boston.
- EDE I — TAKÁCS, G., 1999. *Etymological Dictionary of Egyptian, Volume One: A Phonological Introduction*. Leiden / Boston / Köln.
- EDE II — TAKÁCS, G., 2001. *Etymological Dictionary of Egyptian, Volume Two: b-, p-, f-*. Leiden / Boston / Köln.
- EDE III — TAKÁCS, G., 2008. *Etymological Dictionary of Egyptian, Volume Three: m-*. Leiden / Boston.
- EG — ERMAN, A. & H. GRAPOW, 1957–71. *Wörterbuch der ägyptischen Sprache*, I–VII. Berlin.
- HALOT — KOEHLER, L. & W. BAUMGARTNER, 1994–1996, 1999–2000. *The Hebrew and Aramaic Lexicon of the Old Testament I–III*. Leiden, New York & Köln. IV–V. Leiden, Boston & Köln.
- HJ — HOFIJZER, J. & K. JONGELING, 1995. *Dictionary of the North-West Semitic Inscriptions*. Leiden / New York / Köln.
- HRSC — EHRET, C., 1980. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary*. Berlin.
- Ja. — JASTROW, M., 1996. *A Dictionary of the Targumim, the Talmud Babli and Yerushalmi, and the Midrashic Literature*. New York.
- JH — JOHNSTONE, T. M., 1977. *Harsūsi Lexicon*. New York / Toronto.

- JJ — JOHNSTONE, T. M., 1981. *Jibbāli Lexicon*. New York (NY).
- JM — JOHNSTONE, T. M., 1987. *Mehri Lexicon*. London.
- Kane A — KANE, T. L., 1990. *Amharic-English Dictionary*. Wiesbaden.
- Kane T — KANE, T. L., 2000. *Tigrinya-English Dictionary*. Vol. I–II. Springfield.
- Kog. DD — KOGAN, L., 2006. On Proto-Semitic Deverbal Derivation. *Paper read at II Workshop on Comparative Semitic* (Sitges, Spain).
- Kog. Eth. — KOGAN, L., 2005. Common Origin of Ethiopian Semitic: the Lexical Dimension. *Scrinium*. T. 1: *Varia Aethiopica*. In *Memory of Sevir B. Chernetsov (1943–2005)*.
- Kog. MS — KOGAN, L. 100-wordlist of Soqotri (manuscript).
- Kog. Ug. 1 — KOGAN, L., 2006. Lexical Evidence and the Genealogical Position of Ugaritic. *Babel and Bibel* 3. Winona Lake, Indiana, 1–60.
- Kog. Ug. 2 — KOGAN, L., 2010. Genealogical Position of Ugaritic: the Lexical Dimension. Lexical Isoglosses between Ugaritic and other Semitic Languages. Conclusions. *Sefarad*, vol. 70:2.
- LArg — LESLAU, W., 1997. *Ethiopic Documents: Argobba*. Wiesbaden.
- LGaf — LESLAU, W., 1956. *Étude descriptive et comparative du Gafat (Éthiopien méridional)*. Paris.
- LGur — LESLAU, W., 1979. *Etymological Dictionary of Gurage (Ethiopic)*. Vol. III. Wiesbaden.
- LGz — LESLAU, W., 1987. *Comparative Dictionary of Geʿez (Classical Ethiopic)*. Wiesbaden.
- LH — LITTMANN, E. and M. HÖFNER, 1956. *Wörterbuch der Tigre-Sprache. Tigre-deutsch-englisch*. Wiesbaden.
- LHar — LESLAU, W., 1963. *Etymological Dictionary of Harari*. Berkeley and Los Angeles.
- LS — LESLAU, W., 1938. *Lexique Soqotri (Sudarabique moderne) avec comparaisons et explications étymologiques*. Paris.
- Mil. RE — MILITAREV, A., 2005. Root extension and root formation in Semitic and Afrasian. *Proceedings of the Barcelona Symposium on comparative Semitic, 19–20/11/2004, Aula Orientalis* 23/1–2, 83–130.
- Mil. 2000 — MILITAREV, A., 2000. Towards the chronology of Afrasian (Afroasiatic) and its daughter families. *Time Depth in Historical Linguistics*. Volume 1. Ed. by C. Renfrew, A. McMahon & L. Trask. The McDonald Institute for Archaeological Research. Cambridge, 267–307.
- Mil. 2004 — MILITAREV, A., 2004. Another Step towards the Chronology of Afrasian (I). *Babel and Bibel* 1. Moscow, 282–333.
- Mil. 2007 — MILITAREV, A., 2007. Toward a Complete Etymology-Based Hundred Word List of Semitic. Items 1–34 (First Third). *Proceedings of the 7th International Semito-Hamitic Congress (Berlin, Sept.13–15, 2004)*. Aachen, 71–102.
- Mil. 2008 — MILITAREV, A., 2008. Toward a Complete Etymology-based One Hundred Wordlist of Semitic: Items 34–66 (Second Third). *FS H. Jungraithmayr and A. Dolgopolsky*. Ed. by G. Takacs, 194–222.
- Mil. 2010 — MILITAREV, A., 2010. A Complete Etymology-based Hundred Wordlist of Semitic Updated: Items 1–34. *Journal of Language Relationship* 3, 2010, 69–95.
- Mil. 2011 — MILITAREV, A., 2011. A Complete Etymology-based Hundred Wordlist of Semitic Updated: Items 35–54. *Journal of Language Relationship* 5, 2011, 69–95.
- Mil. 2012 — MILITAREV, A., 2012. A Complete Etymology-based Hundred Wordlist of Semitic Updated: Items 55–74. *Journal of Language Relationship* 7, 2012, 71–103.
- ND — DOLGOPOLSKY, A. *Nostratic Dictionary* (unpublished).
- Nak. 1986 — NAKANO, A. 1986. *Comparative Vocabulary of Southern Arabic*. Tokyo.
- Nak. 2013 — NAKANO, A. 2013. *Hobyōt (Oman) Vocabulary with example texts*. Ed. by Robert Ratcliffe. Tokyo.
- Pen. — PENRICE, J., 1873. *A Dictionary and Glossary of the Koran*. London.
- Sarg. — САРГИС (архимандрит), 1909. *Русско-сирский лексиконъ*. Урмия. [SARGIS, archimandrite, 1909. *Russian-Syriac Lexicon*. Urmia.]
- Sat. — SATTERTHWIT, A. C., 1960. Rate of Morphemic Decay in Meccan Arabic. *International Journal of American Linguistics* 26/3, 254–261.
- SD — BEESTON, A. F. L., M. A. GHUL, W. W. MÜLLER & J. RYCKMANS, 1982. *Sabaic Dictionary (English-French-Arabic)*. Louvain-la-Neuve.
- SED I — MILITAREV, A. & L. KOGAN, 2000. *Semitic Etymological Dictionary*. Vol. 1: *Anatomy of Man and Animals*. Münster.
- SED II — MILITAREV, A. & L. KOGAN, 2005. *Semitic Etymological Dictionary*. Vol. II: *Animal Names*. Münster.
- Sok. JP — SOKOLOFF, M., 1990. *A Dictionary of Jewish Palestinian Aramaic*. Jerusalem.
- Sok. Syr. — SOKOLOFF, M., 2009. *A Syriac Lexicon*. Winona Lake, Indiana — Piscataway, New Jersey.

- Starostin 2000 — STAROSTIN, S., 2000. Comparative-historical linguistics and lexicostatistics, in *Time Depth in Historical Linguistics*, vol. 1, eds. C. Renfrew, A. McMahon & L. Trask. (Papers in the Prehistory of Languages.) Cambridge: The McDonald Institute for Archaeological Research, 223–265.
- St. 2005 — STOLBOVA, O., 2005. *Chadic lexical database. Issue I (L, N, Ny, R)*. Kaluga.
- St. 2007 — STOLBOVA, O., 2007. *Chadic lexical database. Issue II (lateral fricatives)*. Moscow-Kaluga.
- St. 2009 — STOLBOVA, O., 2009. *Chadic lexical database. Issue III (sibilants and sibilant affricates)*. Moscow.
- St. 2011 — STOLBOVA, O., 2011. *Chadic lexical database. Issue IV (velars)*. Moscow.
- Stein — STEIN, Peter, 2010. *Die altsüdarabischen Minuskelschriften auf Holzstäbchen aus der Bayer Staatsbibliothek in München*. Band 1. *Die Inschriften der mittel- und spätsabäischen Periode*. 2. Teil; Verzeichnisse und Tafeln. Epigraphische Forschungen auf der Arabischen Halbinsel. Band 5. Berlin.
- Swadesh 1952 — SWADESH, M., 1952. Lexico-statistical dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96, 452–463.
- Swadesh 1955 — SWADESH, M., 1955. Toward greater accuracy in lexicostatistical dating. *International Journal of American Linguistics* 21, 121–137.
- Tak. *h — TAKÁCS, G., 1997. The Common Afrasian Nominal Class Marker *h. *Studia Etymologica Cracoviensia* 2, 241–273.
- Tak. 2011 — TAKÁCS, G., 2011. *Studies in Afro-Asiatic Comparative Phonology: Consonatism*. Berlin.
- Tomb. — TOMBACK, R. S., 1978. *A Comparative Semitic Lexicon of the Phoenician and Punic Languages*. Ann Arbor.
- Tser. — ЦЕРТЕЛИ, К. Г., 1958. *Хрестоматия современного ассирийского языка со словарем*. Тбилиси. [TSERETELI, K. G., 1958. *A Neo-Assyrian Anthology (with glossary)*. Tbilisi.]

А. Ю. МИЛИТАРЕВ. Новый вариант подробной этимологизации стословного списка семитских языков: элементы 75–100.

Статья представляет собой четвертую часть этимологического анализа списка Сводеша, составленного автором для семитских языков (предыдущие три части были опубликованы ранее в № 3, 5 и 7 настоящего журнала). В ней обсуждаются последние двадцать шесть элементов списка; для каждого элемента предлагается прасемитская реконструкция. Помимо этого, особый упор сделан на поиск внешних (афразийских) параллелей к каждому из разбираемых этимонов.

Ключевые слова: семитские языки, афразийские языки, этимология, глоттохронология, лексикостатистика.

О пороге родства и индексе стабильности в базисной лексике при массовом сравнении: атлантические языки

В статье рассматриваются некоторые ключевые проблемы установления дальнего родства языков в условиях, когда отсутствует классическая ступенчатая реконструкция, а именно: 1) проблема установления порога случайных сходств в стословном списке Сводеша (при различных алгоритмах их фиксации); 2) проблема ранжирования базисной лексики по ее устойчивости; 3) проблема многочисленных изолированных корней, не находящих параллелей в других языках семьи. Эти общие сюжеты рассматриваются на основе конкретной проблематики построения генетической классификации атлантических языков. Атлантические языки традиционно считаются одной из ветвей нигеро-конголезской макросемьи, но при этом лексикостатистика говорит о том, что дата их разделения составляет не менее 8000 лет. Публикуемые два критических отзыва на статью в определенном смысле являются ее продолжением, развивая сложную проблематику установления дальнего родства языков; авторский ответ на эти отзывы следует также рассматривать как важный фрагмент публикуемой статьи.

Ключевые слова: праязыковая реконструкция, базисная лексика, лексикостатистика, этимология, атлантические языки, макрокомпаративистика.

В статье рассматривается ряд ключевых проблем в русле давней дискуссии о критериях установления дальнего родства языков. Для семей старше индоевропейской (ВІЕ) такая дискуссия по-прежнему актуальна, несмотря на публикацию нескольких фундаментальных работ, посвященных этому сюжету¹. Речь пойдет о трудных случаях классификации (например, проблемы классификации языков семьи нигер-конго), когда «нельзя» опереться на классические методы (в отсутствие ступенчатой реконструкции), но «очень хочется» получить предварительные результаты, т. е. нас будут интересовать проблемы построения классификации не на основе регулярных соответствий, а на основе сходств в базисной лексике. Мы сосредоточимся в первую очередь не на общих рассуждениях на эту тему (они в изобилии присутствуют в литературе), а на конкретных критериях установления дальнего родства и границ их реального применения — прежде всего для таких трудных случаев, какими являются, в частности, атлантические языки, которыми я

¹ Так, в 2013 г. вышла в свет исключительно интересная монография Г. С. Старостина [Г. Старостин 2013], с которой во многом перекликается данная статья. В этой монографии представлен прекрасный обзор литературы по установлению критериев дальнего родства, в том числе и статистических. Публикация [Г. Старостин 2013] избавляет автора от необходимости системного представления всех идей и экспериментов в этой обширной области. Здесь же будут рассмотрены только такие аспекты проблемы, которые, как представляется, не привлекали специального внимания компаративистов.

занимаюсь. Если принять терминологию, предлагаемую Г. С. Старостиным («группа» — таксоны, показывающие от 40% и выше совпадений, «семья» — от 15 до 40%, «макросемья» — менее 15%, [Г. Старостин 2013: 140]), то атлантические языки — это в лучшем случае макросемья (по подсчетам Д. Сэпира, языки волоф-темне дают 3% (!) сходств [Sapir 1971]).

В [Г. Старостин 2013] постулируются и аргументируются три основных принципа, которые можно использовать в качестве критериев отделения родственных языков от неродственных в случаях предполагаемого дальнего родства: 1) принцип конкурирующих гипотез; 2) принцип регулярности соответствий и, наконец, 3) принцип «динамической градации базисной лексики». Рассматривая эти принципы на заведомо нелепом примере доказательства «родства» между центральносуданским языком лого и корейским языком, которые дают 10% сходств в стословнике Сводеша, Старостин показывает, что минимум 2 критерия из трех (1 и 3) не позволяют нам постулировать родство между этими языками. Эти же принципы, как показал Старостин, вполне позволяют «разобраться» с примером [Campbell & Poser 2008: 382—384], в котором рассматриваются сближения английского и маори (также 10 созвучных слов а стословнике).

Не будем здесь останавливаться на критерии «конкурирующих гипотез». Он основан на вполне разумном рассуждении (действительно, с учетом накопленных знаний легче допустить родство английского и немецкого, чем английского и маори), но практически неприменим для установления дальнего родства в Африке. Дело в том, что в большинстве трудных случаев у нас нет «конкурирующих гипотез», а точнее, у нас есть определенный набор равновероятных конкурирующих гипотез: классифицируя язык лимба, традиционно относимый к атлантическим, мы, учитывая крайне низкий процент его сходств с другими языками, можем с равным успехом допустить, что он относится, например, к бенуэ-конголезским языкам, или представляет собой отдельную нигер-конголезскую ветвь, или же вообще не является нигер-конголезским языком.

Остановимся подробнее на других критериях, главным из которых, конечно же, является процент сходств в базисной лексике.

1. О пороге дальнего родства

Существует ли некое минимальное количество сходств в списке Сводеша по двум языкам, которое позволило бы нам подозревать, что эти языки родственны? По мнению Г. С. Старостина, «наличие 10% совпадений в пределах стословного списка в принципе достаточно для постулирования глубокого родства между сравниваемыми языками — разумеется, при условии того, что совпадения эти не являются случайными, а отражают систему регулярных соответствий» [Г. Старостин 2013: 33]. Действительно, разнообразные эмпирические данные по макросемьям дают нам многочисленные примеры родственных языков, показывающих не более 10% сходств по Сводешу. Вопрос в том, сколько сходств должны давать случайные совпадения в неродственных языках? Чтобы определить порог случайности, необходимо: а) сформулировать, что мы будем считать фонетическим сходством, б) знать инвентарь фонем и словарные частоты фонем в сравниваемых языках.

Что считать фонетическим сходством? Предположим, что мы имеем дело только со структурами CVC (это основная структура в атлантических языках) и допускаем любые сближения согласных в рамках одного локального ряда (например, любых губных или любых палатальных). Тот же критерий мы применяем к сравнению гласных — с тем ис-

ключением, что центральные гласные мы имеем право сближать как с передними, так и с задними гласными. В среднестатистическом языке это дает нам 4 ряда у согласных и 2 ряда у гласных. При равных частотах для каждого ряда в начальной и в конечной позиции мы должны получить 3 случайных сходства в стословнике Сводеша по неродственным языкам. Четверть слов (т. е. 25) совпадет по рядности начального согласного. Четверть из них (т. е. 6) одновременно совпадет по рядности конечного согласного. Половина из них (т. е. 3) покажет сопоставимый гласный. Англ. *meat* ~ русск. *m'as-o* сходны по принятому критерию, поскольку они родственны, а их сходство с манджак *maf* 'мясо' случайно, но удивительно, как и сходство англ. *mouth* и бедик *m'isy* или англ. *liver* и бапен *dembar*.

Мы посчитали эту вероятность для случая равных частот согласных в каждой из двух позиций и в каждом из четырех рядов. Допустим теперь, что в конечной позиции у нас в обоих сравниваемых языках встречаются только согласные двух рядов — 50% дентальных и 50% велярных. В этом случае число случайных сходств в стословном списке = 6, то есть число ожидаемых случайных сходств увеличилось в два раза по сравнению с предыдущим примером. В целом отметим, что чем больше разброс частот в сравниваемых языках, тем больше оказывается вероятность случайных совпадений. Это, в частности, означает, что упрощенные прикидки порога случайности, исходящие из равенства частот всех согласных, позволяют определить **минимум** случайных сходств — их число, полученное по реальным частотам, всегда будет больше.

Подставим в наш пример реальные частоты. Средние частоты согласных, сгруппированных по рядам, в 40 атлантических языках представлены в табл. 1.

Таблица 1.

ATL (%)	Начальные	Конечные
LAB	31,3	15,0
DENT	22,5	46,9
PAL	18,0	14,6
VEL	28,0	24,0

Мы видим, что в конечной позиции ряды согласных представлены неравномерно: половина согласных в словарях атлантических языков относится к дентальному ряду. Для двух языков, близких к среднестатистическому атлантическому языку, эти частоты предполагают в стословнике 10 случайных совпадений начального и одновременно конечного согласного по ряду. И это без такой «роскоши», как допуск банальной палатализации, предполагающей сближение велярных и палатальных или дентальных и палатальных². Учитывая, что качество гласных при хорошем совпадении согласных практически не принимается во внимание при установлении сближений в африканских языках³, приходится констатировать, что грань между случайным сходством и сходством при дальнем родстве практически стерта. Это становится тем более ясным, если учесть, что при пороге 10 число сходств от 7 до 13 не может считаться существенным отклонением, оставаясь в пределах трех сигм.

Как было показано выше, на порог случайных сходств существенно влияют частоты согласных. А это означает, что лексикостатистика на основе сходств в атлантических язы-

² То есть такие сближения, как диола-гусилай *teeη* с англ. *tau*, не будут считаться сходством.

³ Гласные, как известно, как правило менее стабильны, чем согласные. К тому же избранный принцип (сближение только внутри ряда) для гласных существенно более уязвим, чем для согласных.

ках, показывающих от 5 до 20% сходств для языков разных ветвей, практически неприменима для построения атлантического дерева, поскольку небольшие различия коэффициентов могут определяться особенностями частотного распределения в конкретных парах языков.

Приведем два фактора, существенно повышающих и без того высокий порог случайных сходств.

Первый фактор отчетливо проявляется в языках с включенными в корни архаичными префиксами именных классов, то есть в большинстве нигеро-конголезских языков. Проиллюстрируем его на примере атлантических языков. Когда мы вычисляем процент односложных слов (суммарно для структур V, VC, CV, CVC) отдельно для существительных и глаголов в больших словарях атлантических языков, мы обнаруживаем, что исключительно дивергентные атлантические языки едины в одном: процент моносиллабов в существительных никогда не превышает их процент в глаголах, а в подавляющем большинстве языков он существенно выше, чем в глаголах (в среднем по атлантическим языкам в 2 раза [23% моносиллабов в существительных и 47% — в глаголах]), о чем наглядно говорят данные табл. 2 по 24 атлантическим языкам.

Таблица 2.

%	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	СРЕДН
N	3	9	50	49	42	2	1	67	32	69	3	75	42	45	7	4	38	32	38	14	48	6	4	4	23
V	60	88	61	62	47	2	1	83	49	91	63	92	59	79	7	80	67	94	87	39	49	6	32	16	47

Атлантические существительные существенно длиннее, чем глаголы. Отмечу, что для многих языков подсчеты производились по словарям корней, а не слов. Это может означать только одно: в атлантических существительных системно присутствуют утратившие продуктивность префиксы именных классов, которые мы вынуждены сегодня воспринимать как корневые фонемы. Следовательно, в наших сближениях по сходству мы в реальности часто сближаем не корневые начальные согласные, а согласные архаичных префиксов, что неизбежно повышает процент иллюзорных сходств — уже потому, что согласных, входящих в префиксы классов, существенно меньше, чем корневых согласных.

Приведенная выше таблица хорошо иллюстрирует и еще один фактор, резко повышающий процент случайных сходств в таких языках, как атлантические. В многочисленных моносиллабичных корнях, многие из которых имеют структуру CV, для сходства достаточно совпадения лишь одного согласного. К тому же для пары родственных языков, в одном из которых наблюдается повышенное число корней CV, а в другом — CVC, логично предположить соответствие этих структур (возникшее, например, в результате падения некоторых конечных согласных в одном из языков). Но включение пар вроде *bow* ~ *bo* в потенциально возможные сближения также существенно повышает процент ожидаемых случайных сходств.

Конечно же, формализация критериев сходства неоднократно становилась предметом обсуждения в литературе, посвященной дальнему родству. В монографии Г. С. Старостина подробно рассматриваются различные критерии: от наиболее строгих («согласные и гласные одинаковы или различаются только по одному признаку» [Г. Старостин 2013: 61, 62]⁴ до существенно более гибких — введение «фонетически изоморф-

⁴ Этот критерий использует, в частности, Д. Сэпир в своей лексикостатистике по атлантическим языкам [Sapir 1971].

ных параллелей» как «некоторого компромиссного варианта между „классическими регулярными соответствиями“ и аморфным „фонетическим сходством“ в рамках теории „массового сравнения“» [Г. Старостин 2013: 55]. К компромиссным попыткам формализации консонантных сходств относится и выделение 10 непересекающихся консонантных классов А. Б. Долгопольского и 12 классов С. А. Старостина [Г. Старостин 2013: 185—186].

Для интересующего нас сюжета важно то, что каждый из этих подходов предполагает очень разное число ожидаемых случайных сходств. Максимально строгий подход сводит их число практически к нулю, однако и число наблюдаемых сходств сокращается настолько, что коэффициенты бессмысленно сравнивать. При строгом критерии консонантных сближений для русского и английского мы получаем 10 сходств, и это не учитывая расхождения гласных: *cold* — *xolodnyj*, *full* — *polnyj*, *meat* — *m'aso*, *new* — *novyj*, *no* — *ne*, *nose* — *nos*, *sit* — *sid'et'*, *sun* — *solntse*, *two* — *dva*, *water* — *voda*. Отмечаем, что критерий позволяет выявить приблизительно треть фактических соответствий. Для более древнего родства можно в лучшем случае рассчитывать на несколько сходств. В атлантических языках, например между языками волоф и лимба, мы получим только 2 подходящих сближения, а именно: 'птица' (волоф *picc* ~ лимба *beci*, где *be-*, возможно, является префиксом именного класса!) и 'сидеть' (волоф *toog* ~ лимба *dɔŋɔ*, если считать, что **-g** и **-ŋ** различаются одним признаком, а не двумя). Можно еще добавить сюда 'колени' (волоф *oot* ~ лимба *hibiŋ*), но и здесь, строго говоря, мы выходим за рамки различия по одному признаку между **-m** и **-b**, учитывая наличие **-b** в системе. Но и порог случайности, исчисленный для этой пары языков на основе частот их согласных, дает 2 случайных совпадения первого и второго согласных. Несмотря на то что «строгий» критерий, в силу указанных причин, непригоден для классификации, он может быть использован для предварительного определения статуса языков, которые мы считаем изолированными. С большой вероятностью можно утверждать, что языки, дающие по строгому критерию 5 сходств, являются родственными⁵.

Выделение десятка непересекающихся классов не улучшает дела. Проблемы, связанные с применением классов и построенных на их использовании формальных алгоритмов сближений, прекрасно рассмотрены в [Г. Старостин 2013: 185—190]. Абсолютно справедливо замечание Г. С. Старостина о принципиальной ложности исходного посыла: «... фонетические изменения происходят в пределах одного и того же консонантного класса» [Г. Старостин 2013: 188]. Совершенно прав Г. С. Старостин и в том, что реально мы имеем дело с «пересекающимися подмножествами общего консонантного инвентаря» [Г. Старостин 2013, 189]. Заметим, что критерий консонантных классов является столь же рестриктивным, что и критерий полного сходства (или отличия по одному признаку внутри локального ряда). Исчисленный по обоим критериям на основе реальных частот порог случайных совпадений для пары языков волоф и гола дает одинаковый результат: 2 случайных совпадения по согласным. К тому же возникает простой методологический вопрос: как быть, когда, например, класс R, включающий только вибранты, не представлен в одном из языков? В гола отсутствует фонема /r/. Это означает, что /r/ волоф вообще ничему не может соответствовать в гола.

⁵ Этот вывод подкрепляется, в частности, серией тестов, в которых сопоставлялись списки атлантических языков со списками многочисленных языков за пределами Африки. Применение «строгого» критерия (полное совпадение либо различие по одному признаку в рамках локального ряда) ни в одной паре языков не дает больше двух сходств. 10 «строгих» сходств между русским и английским при глубине расхождения более 6000 лет представляется на этом фоне беспорным свидетельством родства этих языков.

С более «либеральными критериями» дело обстоит сложнее. В поисках сходств имеются три ресурса, позволяющие довести порог случайных совпадений до 100%, а именно: а) вольное обращение со сравниваемыми формами; б) допущение семантических изменений; в) подмена парного сравнения сравнением произвольно выбранных пар языков из двух семей для каждого отдельного значения. Чаще всего эти допущения ничем не оправданы и служат для «научного обоснования» националистических теорий: приблизительно по одному алгоритму строятся доказательства происхождения волоф от языка египетских пирамид или патриотические этимологии Михаила Задорнова. С Задорновым спорить бессмысленно — отметим лишь, что мы имеем дело практически со стопроцентной вероятностью обнаружения случайных внешних сходств для словаря любого языка. Более интересны, естественно, алгоритмы поиска сходств, которые используются профессиональными лингвистами.

Приходится признать, что в лексикостатистике мы допускаем фонетические изменения, но не признаем изменений семантических не потому, что таковых нет. Просто, как только мы «отпустим» тормоза с семантикой, мы мгновенно потеряем возможность опереться на список значений, а вероятность случайных сходств возрастет на порядок. Поэтому большинство лингвистов, прибегающих к лексикостатистике, семантику «не трогают»⁶.

Редко нарушается и принцип парного сравнения. Если попарное сравнение языков подменяется выбором созвучной формы в одном из многочисленных языков семьи или группы, так, как это сделано у Дж. Гринберга, то вероятность обнаружения случайных сходств приближается к 100%. Так, в первой десятке списка Сводеша, без введения семантических изменений, можно обнаружить следующие сходства между английским и атлантическими языками (табл. 3):

Таблица 3.

1	all	
2	ashes	limba <i>hɔtɔ</i> , temne <i>r-os</i>
3	bark	joola-kasa <i>ka-fongol</i> , temne <i>foki</i>
4	belly	wolof <i>biir</i> , joola <i>far</i> , limba <i>fɔle</i>
5	big	sereer <i>mag-in</i> , joola mlomp <i>bə-məki</i> , manjak / pepel <i>mag</i> , nalu <i>baŋ</i> , <i>gboŋ</i> , бага <i>maduri baki</i>
6	bird	
7	bite	wolof <i>màtt</i>
8	black	bedik <i>-bàlà</i> , biafada <i>baŋ</i> , peul <i>balee</i> , sereer <i>baal</i> , nalu <i>bal</i> , limba <i>boloi</i>
9	blood	jaad <i>poada</i> , buy <i>biler</i> ,
10	bone	joola <i>wul</i> , manjak <i>moɓ</i> , pepel <i>mɔ</i> , nalu <i>wol</i> , бага- <i>maduri</i> / temne <i>ke-bant</i>

⁶ В анонимной рецензии на рукопись статьи содержалось следующее рассуждение: «Автор не прав, строгий ответ иной. Мы „не трогаем“ семантику, потому что семантическое тождество предполагает *по определению*. Без условия тождества значений лексикостатистика превращается в т. н. этимостатистику (фактически не разработанный метод сравнения)». Это возражение я принять не могу, и не только потому, что оно сформулировано не вполне внятно, но и по сути. Если имеется в виду, что семантическое тождество предполагается по определению, то законен вопрос — почему? Представляется, что эта конвенция, предложенная еще Сводешем, продиктована в первую очередь тем, что учет любого семантического изменения автоматически уводит нас от фиксированного списка значений, состоящего ровно из 100 единиц.

Удивительно не то, что для каждого английского слова в семье из 40 языков находятся созвучные слова. Удивительно скорее, что для двух из них не нашлось близких по звучанию слов.

Отдельно подчеркну следующее. Меньше всего автору хотелось бы, чтобы эти строки воспринимались читателем как очередная попытка ревизии уникальной работы, проделанной Дж. Гринбергом. Именно Гринберг сформировал облик современной африканской компаративистики. Г. С. Старостин пишет о «парадоксальной ситуации», когда большинство лингвистов отвергают метод «массового сравнения», но при этом опираются на гринберговскую классификацию, «в силу полного отсутствия альтернативных схем» [Г. Старостин 2013: 7]. Мне представляется, что этот «парадокс» решается просто: у самого Гринберга не было никакой формализованной методики сравнения. Его этимологические сближения следует воспринимать как «точечные» (в силу огромного материала) иллюстрации к результатам, основанным не на конкретной процедуре анализа, а на энциклопедических знаниях и исключительном исследовательском таланте. Два последних качества не всегда имеются у последователей Гринберга, которые используют его публикации для методологического обоснования массового непарного сравнения, которое может гарантировать «доказательство» любого родства.

Оставим общие рассуждения на эту тему и посмотрим, какие именно звуковые сходства допускал Дж. Гринберг. Это позволит нам оценить вероятность случайных сходств, если использовать его разрешения на совместимость в случае попарных языковых сравнений. Обратимся к одной из наиболее представительных его публикаций [Greenberg 1963]. В ней на примере 49 корней и их рефлексов обосновывается родство нигеро-конголезских языков. В своих иллюстрациях Гринберг старается подбирать слова с одинаковыми гласными, но, если это не получается, сближает любые гласные без ограничений. Минимум 22 корня из 49 имеют во всех рефлексах структуру CV, т. е. включают только один согласный. В 27 корнях для многих языков дается второй — неначальный — согласный. Понятно, что одинаковые согласные сближать разрешается. Если говорить о сближениях согласных разного качества или о соответствиях различных согласных нулю, то разных сближений этого типа в списке имеется 111 для начальных согласных и 157 для неначальных. Многие неоднозначные сближения представлены в различных корнях: так, например, сближение $n = \eta$ в начальной позиции отмечено в 10 корнях (корни 1, 11, 16, 19, 23, 24, 32, 34, 36, 46 в нумерации Гринберга).

Выделим 20 наиболее часто встречающихся у Гринберга согласных и представим сближения согласных разного качества у Гринберга в таблице 4.

Отметим только основные особенности таблицы. Лабиальные и дентальные сближаются в основном внутри локального ряда. При этом для дентальных это справедливо как для начальных, так и для неначальных, а для лабиальных — исключительно для начальных: у Гринберга нет ни одного примера сближения в неначальной позиции двух разных лабиальных согласных. Палатальные, а также велярные согласные практически сближаются только с одинаковыми согласными в обеих позициях. Носовые согласные свободно сближаются как по ряду, так и по серии (поэтому в таблице они сгруппированы). В начальной позиции сближения согласных с нулем не допускаются. Напротив, в неначальной позиции они используются Гринбергом чрезвычайно широко — особенно для плавных и носовых.

Определим, сколько случайных совпадений при этом алгоритме мы должны получить, сравнивая список Сводеша в двух любых языках (при равных частотах всех согласных в обеих позициях). Разрешения для начальных согласных составляют 25% (в таблице, включая диагональ, заполнено 58 клеток из 231). Разрешения для неначальных согласных составляют 32% (включая диагональ, заполнено 75 клеток из 231). Вспомним, что

Таблица 4.

	p	f	b	v	w	t	r	d	l	c	s	j	y	k	h	g	m	n	ɲ	η	∅	
p	—	5	7	4													1					
f		—	5	3																		
b			—	7													1					
v				—													1					
w					—																	
t					1	—	1	2		1	6											
r					1	3	—	7	7									1	1			
d					1	1	4	—	8									1	1			
l					1	2	8	6	—									1	1			
c										—	1											
s						1	1				—	1		1		1						1
j												—										
y										1			—									
k					1									—		2						1
h															—							
g					1					1			1	1		—						1
m																	—	2	1	1		
n							2	1	1					1			1	—	10	1		
ɲ					1	1	2	1	1									3	—	1		
η							2	2	2	1			1	1		1	2	6	1	—		
∅					3	3	12	6	10	1	1		2	2	1	2	3	9	4	13		—

Комментарий к таблице: в ячейках указаны количества корней, в которых отмечено то или иное сближение; выше диагонали в таблице фиксируются сближения для начальных согласных, ниже диагонали — для неначальных.

22 корня из 49 у Гринберга имеют структуру CV, то есть для 48% слов достаточно совпадения только начального согласного. Это дает нам в списке Сводеша уже 12 случайно совпавших слов ($100 \cdot 0,25 \cdot 0,48$). Для 52% нам необходимо совпадение начального и неначального согласных, что дает еще 4 слова ($52 \cdot 0,25 \cdot 0,32$). Итак, при алгоритме Гринберга в списке Сводеша случайное сходство должно обнаружиться у 16 слов! А это уже не просто родство — это уже семья, а не макросемья по Г. С. Старостину. К тому же напомним, что допущение о равенстве частот согласных (которого никогда не бывает) дает минимальное число случайных сходств — чем больше различия частот, тем больше окажется случайных сходств. Это значит, что реально по алгоритму Гринберга, не учитывающему различия гласных, число случайных сходств будет больше 16. В таком случае, имея два любых словаря по 1000 слов, мы можем продемонстрировать родство языков, представив убедительный список в 160 корней⁷.

⁷ Привожу это очевидное рассуждение, учитывая явное статистическое недоразумение в [Г. Старостин 2013, 42], касающееся процента случайных совпадений в «слабой» лексике. В приводимом примере, как пишет автор, «ожидалось бы скорее число, близкое к нулю», хотя в действительности ожидание составляет 50%.

Таблица 5.

	p	t	c	k	f	r	s	h	b	d	j	g	w	l	y	∅	m	n	ñ	η
p	1				1			1	1				1			1				
t		1	1			1	1			1				1						
c		1	1	1		1	1	1			1				1					
k			1	1			1	1				1	1			1				
f	1				1	1	1	1	1				1			1				
r		1	1		1	1	1	1			1			1	1	1				
s		1	1	1	1	1	1	1						1	1					
h	1		1	1	1	1	1	1				1	1		1	1				
b	1				1				1			1	1			1	1			
d		1								1	1			1				1		
j			1			1				1	1	1			1				1	
g				1				1	1		1	1	1		1	1				1
w	1			1	1			1	1			1	1		1	1	1			1
l		1				1	1			1				1	1	1		1		
y			1			1	1	1			1	1	1	1	1	1			1	
∅	1			1	1	1		1	1			1	1	1	1	1		1		
m									1				1				1	1		1
n										1				1		1	1	1	1	1
ñ											1				1			1	1	1
η												1	1				1	1	1	1

На практике в африканской макрокомпаративистике, как показывает анализ конкретных публикаций, алгоритмы выделения когнатов достаточно «либеральны». Наиболее типичными являются следующие критерии фонетической совместимости:

- 1) признаются сходными любые два согласных одного места образования (например, любые два лабиальных);
- 2) применительно к велярным и денальным (особенно перед передними гласными) допускается «реконструкция палатализации», т. е. допускается сближение $ti \sim ci$, $ge \sim je$ и им подобные;
- 3) допускается сопоставление слов CV и CVC, особенно в случаях, когда конечный согласный «слабый» — например, носовые сонанты **w, l, y** (а также **n**) или фрикативные — например, **f, h**;
- 4) допускается в тех же случаях совпадение слов VC и CVC с предположительным падением начального согласного или его интерпретацией как префикса именного класса;
- 5) допускаются сближения CCV и CV (особенно если второй согласный кластера — плавный), а также CC и CVC;
- 6) различие гласных часто игнорируется, особенно в тех случаях, когда два согласных в структурах CVC показывают хорошую совместимость;
- 7) если три первые звука предполагаемых корней показывают сходство по указанным признакам, различие последующих звуков объясняется суффиксной морфологией или просто игнорируется.

Впрочем, иногда легче дать список, чем формализовать все правила его построения. Чтобы приблизительно оценить порог случайных сходств при допущении наиболее типичных соответствий в атлантических языках, допустим, что у нас есть два языка с одинаковым инвентарем, включающим 20 согласных в обеих позициях, причем частоты всех согласных равны. Отметим в таблице допускаемые сходства (табл. 5).

В нашей таблице 5 заполнено 158 ячеек из 400, то есть 40%. При равных частотах это дает $100 * 0,4 * 0,4 = 16$ случайных сходств начального и конечного согласных в словнике по двум языкам, а если ввести 2 непересекающихся класса гласных — соответственно 8 сходств. Если у строгого читателя 10 сходств вызывают активный протест, он может исключить их, и в таком случае вероятность случайного сходства согласных уменьшится до 14.

И наконец, рассмотрим реальный пример с реальными частотами. Определим, сколько сходств мы должны случайно получить для языков волоф и гола. Пример интересен тем, что гола в определенном отношении резко отличается от других языков, традиционно относимых к атлантическим, — в конечной позиции 88% слов гола имеют гласный (в волоф — 33%). Приведем частоты конечных в гола (табл. 6).

Таблица 6.

j	g	m	ŋ	n	l	∅
0,1%	0,1%	1,0%	1,6%	4,1%	5,4%	87,8%

При таком распределении, казалось бы, уже практически не важны частоты конечных согласных, как, впрочем, и индивидуальные решения лингвиста, касающиеся допускаемых сближений конечных согласных волоф и гола: количество случайных совпадений будет определяться по существу частотами только начальных согласных. И тем не менее наши индивидуальные решения даже в данном нетипичном случае принципиально меняют порог случайных сходств. Для «чистоты эксперимента» я попросил заполнить матрицу разрешений волоф—гола моего коллегу, специалиста по атлантическим языкам Гийома Сежерера, который исповедует строгий подход к сближениям (на мой взгляд, слишком строгий). На основе его заполнений случайность определяется как 11 слов (без учета различия гласных). Я применял достаточно либеральный подход, и на основе моих заполнений порог случайности оказался в 2 раза выше и составил 22 слова. В свою защиту отмечу интересное обстоятельство. При «строгом» подходе Г. Сежерера слова волоф с конечными глухими смычными (в частности, высокочастотный конечный -t) и тем более конечные геминированные глухие смычные не могут считаться совместимыми ни с одним из конечных звуков языка гола. Но ведь если мы имеем дело с родственными языками, то чему-то они соответствуют, например, нулю?! Этот пример интересен следующим. Он наглядно показывает, что, в противоположность часто высказываемому мнению, опора на внешнее сходство часто дает меньше сближений, чем выявление регулярных соответствий. Никакое самое вольное понимание сходства не позволит выявить соответствие русск. *дым* и фр. *fumée*. В любом случае отметим, что при имеющихся частотах согласных вероятность их случайного сходства в волоф и гола превышает 10, а приблизительно столько реальных сходств атлантические языки разных ветвей и показывают.

Проверим наши подсчеты на реальных примерах. Сопоставим английский список Сводеша со списками атлантических языков. Смысл такого рода сопоставлений не в том, чтобы предоставить читателю еще несколько примеров курьезных сходств в заведомо неродственных языках, а в том, чтобы по мере накопления данных эмпирически нащупать реальный порог случайных сходств при реально используемых алгоритмах сближения слов.

Таблица 7.

Temne	
ashes	r-os
bark	foki
bone	ka-bant
fly	faliɾ ^w
go	kɔ
me	mi
little	lɔl
one	in
sand	a-sənt
seed	ma-sənt, ma-ther
woman	bom

Biafada	
fire	fuuru
fly	fiili
me	mma
mouth	mməsə
nose	nasin
say	cii, re
white	paas
you sg.	yi
black	baŋg?
breast	bəl?
ground	gal?
fish	yesä / bwa-?
little	tiiti?
rain	səna?

Bedik	
black	-bála
breast	ε-bār
burn	o-bód
hand	ε-kònd
hear	o-hér
me	-mé- / mé-
mouth	b̄-más
white	-fèšà / -pèšà
eye(s)	gí-nguùs?
knee	é-dà-kàlà?
liver	i-dám-bél?
meat	ɲás?

Nalu	
black	bal
drink	dig
go	kɔ
star	ma-saar
you sg.	yi
big	wək? -baŋ?
bone	wol?
breast	n-bin?
fire	mət?
good	kuun?
tree	n-ti?
we	biye?

Limba	
belly	fɔle
black	boloi
breast	hu-bili
fish	fee
tooth	hu-titi
you sg.	yi
ashes	hɔtɔ? ~ fɔtɔ?
drink	tie?
eye	hoya ~ foya?
kill	kɔra?

Mankany	
big	-mag
liver	u-deeb
meat	u-ye-maɕ
nail	ka-ñiiri
what	we
white	p-faaθ
breast	p-wil?
hair	ka-weel?
leaf	ka-toh?
nose	b-yiis?

В наш список были включены 41 атлантический язык (включая несколько языков мель), а также один язык другой нигеро-конголезской группы — манде, а именно язык мандинка. Число сходств с английским списком колеблется в них от 3 (языки биджого, суа, шербро) до 14 (язык биафада). В среднем атлантический язык дает 8,5% случайных сходств с английским⁸. 12 языков показывают 10 и более процентов сходств. Приведем созвучные слова в 6 из них⁹ (табл. 7).

⁸ Отметим, что сравнение атлантических слов с английским системно дает существенно больше сходств, чем с французским или с русским. Мне не представить, с чем это могло бы быть связано. Английские слова более «прототипичны», легче модифицируются в соответствии с указанным алгоритмом сравнения?

⁹ Знаком вопроса отмечены менее «убедительные» сближения.

Следует сжиться с фактом, что десяток сходств при реально существующих алгоритмах их выделения — это всего лишь норма случайных созвучий между любыми языками. А это означает, что на глубине ВЕ расхождения в процентах сходств ни о чем не говорят. Иными словами, они не могут быть, например, использованы для установления конфигурации нигеро-конголезского и даже атлантического древа.

В таких случаях в теории макрокомпаративистики предлагается подключать другой критерий, подробно обсуждаемый в [Г. Старостин 2013], а именно регулярность звуковых «соответствий» в наблюдаемых сходствах. Но этот критерий применить очень непросто.

Во-первых, в десятке слов регулярность может не проявиться — слишком беден материал, а привлекая сходства за пределами списка Сводеша, нетрудно подтвердить все что угодно.

Во-вторых, и заведомо неродственные языки могут обнаруживать регулярность сходств. Рассмотрим приведенные в таблице сходства английского и темне. В 11 приведенных примерах мы обнаруживаем 4 «соответствия» звонких в английском — глухим в темне: *bark* — *foki*, *go* — *ko*, *sand* — *sant*, *seed* — *sant*. При этом ни одного обратного примера в списке нет! В одной паре отмечено «соответствие» **b-** ~ **b-**, но при желании компаративист легко может объяснить это отклонение, реконструировав в этом случае, например, ***bh-**, и уточнить диахроническое правило: в английском звонкие простые оглушаются, а придыхательные утрачивают аспирацию.

Рассмотрим регулярность «соответствий» в английском и бедик. Констатируем, что во всех трех случаях, англ. **b-** соответствует преглоттализованному **ɸ-** в бедик и никогда простому **b-**. Отмечаем, что когда в английском имеется кластер со вторым плавным или **-n**, в бедик ему регулярно соответствует последовательность CVC (конечно же V в бедик — эпентетический гласный!): *black* — *bálà*, *breast* — *ɸár*, *knee* — *kàlà*. Конечное **-t** (вне кластеров) дает в бедик фрикативный: *meat* — *ɸás*, *white* — *fèšà*. Трудно ожидать большей регулярности в 12 словах!

Признаюсь, что в отдельных случаях сближения явно неправомерны. Так, *nose* нельзя сближать с биафада *ɸasin*, поскольку **ɸa-** является в данном случае префиксом именного класса, а корень здесь — *sin*. Но дело в том, что именно этот фактор представляет основную проблему и при сравнении атлантических языков между собой, когда мы очень часто не можем провести аргументированное морфемное членение слов и, соответственно, получаем сближения, аналогичные рассмотренному.

Если десяток созвучных слов не является признаком родства, а критерий регулярности соответствий, как и критерий «конкурирующих гипотез», также не вполне эффективен, остается только один критерий, который был впервые четко сформулирован С. А. Старостиным и подробно рассмотрен в [Г. Старостин 2013] — ранг стабильности единиц списка Сводеша. Рассмотрим перспективы его применения к атлантическим языкам.

2. О рангах стабильности

В своей фундаментальной работе «Определение устойчивости базисной лексики» С. А. Старостин заложил основу для сопоставления диахронической стабильности слов в языках мира [С. Старостин 2007]. На огромном материале (737 языков в 14 семьях!) автор эмпирически определил относительную устойчивость слов в списке Сводеша и предложил критерий ее абсолютной оценки. Создана, по существу, новая область диахронической семантики. К сожалению, полученные Старостиным результаты широко не обсуждаются в публикациях по компаративистике — практически единственным исключением являет-

ся монография Г. С. Старостина [Г. Старостин 2013], в которой предложенный С. А. Старостиним критерий системно рассматривается в связи с проблемой установления дальнего генеалогического родства. В своей небольшой и предельно насыщенной статье С. А. Старостин не имел возможности не только останавливаться на крайне любопытных деталях, но и дать разностороннюю оценку обработанных им данных. Прежде чем интегрировать атлантический материал в сравнительные данные Старостина, остановимся на некоторых следствиях, вытекающих из его анализа, которые в указанной статье не эксплицированы.

С. А. Старостин вводит следующий индекс стабильности: «отношение максимально-го числа языков, использующих для данного значения один и тот же корень, к общему числу языков в семье» [С. Старостин 2007: 827]. Так, по Старостину, индекс стабильности в славянских языках для значения «кора» = 1 (корень **(s)korā* представлен в 13 языках из 13), а для значения «живот» = 0,38 (максимально широко представленный корень **brjuхъ* / **brjuхo* отмечен в 5 языках из 13, менее распространенные корни **žīvotъ*, **měхъ*, **trьbuxъ*, а также заимствования не учитываются).

Вычислим по данным Старостина средний индекс стабильности для одного значения списка в 14 семьях¹⁰ (табл. 8).

Таблица 8.

	YEN (енисейск.)	TH (тайск.)	DR (дравид.)	KT (картв.)	UR (урал.)	IE (и.-е.)	NC (сев.-кавк.)
№ языка	5	47	21	7	17	97	33
сред. стаб.	0,67	0,65	0,56	0,54	0,48	0,45	0,43

	KS (койсанск.)	Alt (алтайск.)	ST (сино-тиб.)	AN (австронез.)	AA (австроаз.)	AUS (австрал.)	AF (афразий.)
№ языка	26	54	132	94	99	36	69
сред. стаб.	0,43	0,41	0,41	0,40	0,37	0,29	0,20

Действительно, как отмечал Старостин, во многом «индекс стабильности зависит от времени, прошедшего с момента распада праязыка данной семьи» [С. Старостин 2007, 828]. Этот фактор хорошо виден в таблице, упорядоченной по убыванию индексов стабильности. Наиболее низкий индекс у афразийских языков — 0,20, что интерпретируется следующим образом: в афразийской семье в среднем максимально широко представленный корень в значениях списка Сводеша отмечен в 20% языков, то есть приблизительно в 14 из 69 языков. При этом, возможно, афразийская семья является наиболее древней из перечисленных. На этом фоне тем более интересен соответствующий индекс в атлантических языках (включая мель), где он равен 0,28¹¹.

Для упорядочивания списка по рангам устойчивости С. А. Старостин прибегает к остроумной, но достаточно сложной процедуре: «1) Ранжируем все значения (от 1 до 110) по индексу стабильности в каждой из рассмотренных семей... 2) Из значений, занимающих позицию n (начиная с $n = 1$ и кончая $n = 11(0)$), будем выбирать такие, которые занимают эту или более высокую ($n-1$) позицию как минимум в двух семьях. При этом при одинаковом ранге (позиции n) „удельный вес“ выбранного значения будет увеличи-

¹⁰ Данные получены по стословнику Сводеша — 10 дополнительных значений при подсчетах были исключены.

¹¹ В атлантическом списке значения «green» и «yellow» заменены на «walk» и «you pl.».

ваться в соответствии с числом семей, в котором он эту позицию занимает» [С. Старостин 2007, 833, 838]. Используя эти процедуры, Старостин получает обобщенные ранги стабильности для каждого значения списка в языках мира (табл. 9).

Таблица 9.

	1	2	3	4	5	6	7	8	9	10
1—10	we	two	I	eye	you sg.	who	fire	tongue	stone	name
11—20	hand	what	die	heart	drink	dog	louse	moon	nail	blood
21—30	one	tooth	new	dry	liver	eat	tail	this	hair	water
31—40	nose	no	mouth	full	ear	that	bird	bone	sun	smoke
41—50	stand	tree	ashes	give	rain	star	fish	neck	breast	leaf
51—60	come	kill	foot	sit	root	horn	fly	hear	skin	long
61—70	road	know	say	egg	seed	knee	black	head	sleep	burn
71—80	ground	feather	swim	white	bite	meat	fat	man	person	all
81—90	night	see	go	hot	red	cold	woman	round	lie	cloud
91—100	big	bark	sand	good	many	yellow	mountain	green	belly	little

Этот список, естественно, очень интересен для выявления порядков, но использованный Старостиным сложный алгоритм его получения не вполне удобен: во-первых, он не прозрачен, во-вторых, он не дает возможности количественного сравнения индексов. С. А. Старостин отказался от сопоставления средних индексов, учитывая «зависимость коэффициента стабильности от глубины семьи и числа сравниваемых языков» [С. Старостин 2007: 833]. Однако и алгоритм, использованный Старостиным, не вполне преодолевает указанные факторы. Вместе с тем средние (числовые) коэффициенты конечно же удобнее для сопоставления. Получим по материалам Старостина иерархию значений по средним коэффициентам (табл. 10)¹².

Таблица 10.

	1	2	3	4	5	6	7	8	9	10	
1—10	I	you sg.	two	eye	we	tongue	name	one ++	dog +	nail +	0,71
11—20	die	water +	tooth +	louse	fire -	blood	hand	moon	nose ++	liver -	0,59
21—30	ear -	eat	heart -	sun +	star ++	stone --	new	what -	who --	smoke +	0,54
31—40	bone	full	hair -	give +	rain +	tail -	tree +	fish +	stand -	drink --	0,5
41—50	no -	leaf	foot +	this --	dry --	horn +	head ++	hear +	black ++	come +	0,44
51—60	sleep +	ground ++	egg +	knee +	long	skin	road +	kill	meat ++	sit	0,41
61—70	night ++	root -	mouth ---	breast --	red ++	white +	ashes --	know	bird ---	feather +	0,38
71—80	person	hot +	see +	bite	seed -	fat	go +	that ----	swim	cloud +	0,35
81—90	fly ---	neck ----	good +	man -	say --	woman	mountain +	burn --	lie	yellow +	0,31
91—100	bark	green	big	cold -	little	belly	sand	round -	all --	Many	0,25

¹² Плюсами и минусами обозначены все изменения десятков: так, два плюса для head означает, что его ранг поднялся на 2 десятка: у С. А. Старостина это значение было в седьмом десятке, а в исчислении по средним коэффициентам — в пятом; четыре минуса у neck обозначает, что из пятого десятка у Старостина оно перешло в девятый десяток.

Какой список «лучше»? Московская школа компаративистики использует в лексикостатистике первую половину списка, ранжированного по коэффициентам стабильности, полученным С. А. Старостиным. Изменение критерия привело к тому, что в первой половине списка мы «потеряли» 6 значений и соответственно 6 «приобрели». Вышли за пределы первой половины: *mouth, breast, ashes, bird, that, neck*. Они были заменены следующими значениями: *foot, horn, head, hear, black, come*. Чтобы оценить эти изменения, обратимся к интуиции компаративистов. По мнению Г. С. Старостина, «в первую очередь из-за технических обстоятельств» 9 значений из первой половины списка (*liver, neck, breast, this, that, full, give, stand, fish*) следует заменить на 9 значений из второй половины, а именно: *kill, foot, horn, hear, meat, egg, black, head, night* [Г. Старостин 2013: 141–143]. Оказывается, с интуицией у автора все в порядке. Пять из девяти предложенных Старостиным значений для включения в стабильную часть вошли в первую половину списка *foot, horn, hear, black, head*, значение *egg* вплотную приблизилось к нему, оказавшись в шестом десятке вместо седьмого, а значение *night* повысило свой ранг на два десятка. Как и хотелось автору, значения *neck, breast, that* вышли из первой половины, а значения *liver, this, stand* существенно понизили свой ранг (*this* теперь занимает 44-е место вместо 28-го). Эти результаты однозначно свидетельствуют о том, что критерий средних коэффициентов не только более удобен, но и лучше работает. Так, например, с его помощью легко оценить различия в стабильности первой и второй половин списка в разных семьях.

Исчислим коэффициент стабильности для каждой из двух половин списка (включив в список данные по атлантическим языкам — ATL, 41 язык). Отсортируем семьи по различиям коэффициентов первой и второй половин (третья строка) — табл. 11.

Таблица 11.

	ST	DR	AN	NC	UR	IE	KT	AA	TH	YEN	Alt	KS	AF	AUS	ATL
1. 1–50	0,56	0,71	0,55	0,58	0,62	0,59	0,65	0,47	0,75	0,75	0,48	0,50	0,26	0,35	0,32
2. 51–100	0,25	0,41	0,25	0,28	0,34	0,32	0,42	0,27	0,56	0,60	0,34	0,36	0,14	0,24	0,25
1–2	0,31	0,31	0,30	0,29	0,28	0,27	0,23	0,21	0,19	0,15	0,14	0,14	0,12	0,11	0,07

Прежде всего, все без исключения семьи показывают существенно более высокую стабильность первой половины по сравнению со второй. Даже в атлантических языках, занимающих по этому критерию последнее место, эта разница большая: в среднем 30% языков для основных корней в первой половине («главные» корни заполняют матрицу на треть) и 25% — во второй половине («главные» корни заполняют матрицу на четверть). В большинстве же семей эта разница исключительно велика — в 8 семьях она превышает 20%.

Интересно, что в самом конце списка оказываются все три африканские семьи плюс австралийская. Индексы стабильности значений, полученные в основном на языках Евразии, существенно отличаются за ее пределами. Это означает только, что накопление материалов по стабильности базисной лексики должно быть продолжено.

Приведем атлантический список, отсортированный по убыванию индексов стабильности (табл. 12).

Очевидно, что атлантический список существенно отличается от среднестатистического. Так, например, значение *pame* перешло из первого десятка в последний, значение *fly* — из девятого десятка в первый. Значение *head*, занимающее второй ранг в атлантических языках, ни в одной из семей не входит не только в первую десятку, но и в первую два-

Таблица 12.

	1	2	3	4	5	6	7	8	9	10
1—10	eye 0,83	head 0,76	ear 0,71	tongue 0,61	tooth 0,56	star 0,54	hair 0,51	tree 0,51	white 0,49	fly 0,46
11—20	knee 0,46	liver 0,44	one 0,44	rain 0,44	sun 0,44	little 0,41	water 0,41	foot 0,39	I 0,39	you sg. 0,39
21—30	die 0,37	bark 0,34	bird 0,34	horn 0,34	mouth 0,34	neck 0,34	bite 0,32	black 0,32	breast 0,32	dry 0,32
31—40	feather 0,32	new 0,32	person 0,32	two 0,32	eat 0,29	egg 0,29	go 0,29	swim 0,29	cold 0,27	louse 0,27
41—50	ground 0,27	man 0,27	road 0,27	root 0,27	stone 0,27	walk 0,27	what 0,27	you pl. 0,27	blood 0,24	good 0,24
51—60	leaf 0,24	lie 0,24	nose 0,24	see 0,24	skin 0,24	come 0,22	drink 0,22	fire 0,22	sit 0,22	who 0,22
61—70	belly 0,20	bone 0,20	dog 0,20	give 0,20	hand 0,20	hear 0,20	kill 0,20	meat 0,20	nail 0,20	night 0,20
71—80	red 0,20	sand 0,20	seed 0,20	smoke 0,20	tail 0,20	this 0,20	woman 0,20	all 0,17	fat 0,17	fish 0,17
81—90	full 0,17	hot 0,17	know 0,17	many 0,17	sleep 0,17	stand 0,17	we 0,17	burn 0,15	cloud 0,15	moon 0,15
91—100	mountain 0,15	say 0,15	ashes 0,12	big 0,12	heart 0,12	long 0,12	name 0,12	that 0,12	no 0,10	round 0,07

дцатку стабильных значений. Вместе с тем очень существенными отличиями характеризуются по признаку стабильности и другие семьи. Так, значение name оказывается в последней десятке и в тайских языках. Значение drink в койсанских и австронезийских языках входит в первую десятку, а в алтайских языках занимает последнее 100-е место. Хотелось бы прояснить самые общие вопросы: какие языки наиболее близки к среднестатистическому, а какие наиболее удалены от него и насколько? Одинаковы ли отклонения от среднестатистических индексов в каждой десятке стословника? Если мы делим список на две половины, сколько значений в каждой семье языков оказываются «не на своем месте»?

Эти ответы (по крайней мере, приближения к ним) нетрудно получить. Для того чтобы изложить количественный алгоритм сравнения, воспользуемся иллюстративной таблицей (табл. 13) с рассмотренным значением name.

- 1) Отметим для каждого из 100 значений номер десятки, в которую оно входит в семье языков. Значение name (строка 1 таблицы) входит в большинстве семей в первую десятку, в DR — в десятку 2, в AUS — в десятку 3, в AN — в десятку 5, в AA и Alt — в десятку 6, в TH и Atl (атлантические языки) — в последнюю десятку¹³.
- 2) Вычтем из полученных индексов (1—10) среднестатистический индекс (в случае с name — единицу) — строка 2. Таким образом, мы определяем, на сколько («десятков») индекс в семье отличается от среднестатистического.

¹³ Точно разбить индексы на группы по 10 невозможно. Одинаковый индекс могут иметь, например, значения 8—12, и в таких многочисленных случаях отнесение двух первых по алфавиту значений к десятке 1, а вторых — к десятке 2 безосновательно. Пренебрежем этим фактором, чтобы избежать существенного усложнения алгоритма сравнения. Исключим лишь енисейские языки из сравнения — в семье только 5 языков, а значит, только 5 различных индексов, а в таком случае указанная погрешность слишком велика.

Таблица 13.

'name'		СРЕДН	KT	UR	IE	NC	KS	ST	AF	DR	AUS	AN	AA	Alt	TH	ATL	$\Sigma 1$	$\Sigma 2$
1	10	1	1	1	1	1	1	1	1	2	3	5	6	6	10	10		
2	10		0	0	0	0	0	0	0	1	2	4	5	5	9	9		35
3	10		0	0	0	0	0	0	0	1	1	1	1	1	1	1	7	
4	20	1	1	1	1	1	1	1	1	1	2	3	3	3	5	5		
5	20		0	0	0	0	0	0	0	0	1	2	2	2	4	4		15
6	20		0	0	0	0	0	0	0	0	1	1	1	1	1	1	6	
7	50	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2		
8	50		0	0	0	0	0	0	0	0	0	0	1	1	1	1		4
9	50		0	0	0	0	0	0	0	0	0	0	1	1	1	1	4	

3) Возможен менее детальный подсчет (строка 3) — вне зависимости от того, насколько отличается индекс от среднестатистического, будем приписывать любому отклонению значение 1, а отсутствию отклонений — значение 0. Например, в картвельских — ранг значение паше приблизительно совпадает со средним (оно оказывается в той же десятке), а в австралийских — отличается от него (оно оказывается в другой десятке).

Продедаем те же процедуры, разбив список не на десятки, а на двадцатки (строки 4—6 в таблице). В тайских и в атлантических языках значение паше находится в пятой — последней — двадцатке (строка 4); его ранг на 4 отличается от среднего (строка 5); его ранг отличается от среднего (строка 6).

И наконец, мы можем разбить список на две половины (строки 7—9). В уральских языках значение паше находится в первой половине списка (строка 7). Это не отличает его от среднестатистического списка (в данном случае строки 8 и 9 принципиально одинаковы).

В последних столбцах вычислены суммы по строкам. $\Sigma 1$ принципиально меньше, чем $\Sigma 2$. Суммы в строке 6 и тем более 9 принципиально меньше, чем в строке 3.

Применение указанных алгоритмов позволяет сопоставлять семьи, значения и группы значений (по 10, по 20, по 50) по рангам стабильности. Ограничимся здесь суммарными данными по десяткам (аналог строки 3 в табл. 13), см. — табл. 14.

Вдумаемся в цифры, приведенные в таблице. Начнем с сумм по столбцам. В афразийских языках, показывающих наименьшее отклонение от среднего, 75 значений из 100 входят не в те десятки, которые вычислены для среднего, а в алтайских языках таких отклонений — 88! Подчеркну, речь идет не о том, что значения имеют другой ранг — они имеют существенно другой ранг. Все 10 алтайских значений, составляющих вторую десятку, не входят в нее в среднестатистическом списке. В первой десятке (наиболее стабильных значений) 8 значений в AA, TH, Atl имеют более низкий ранг в усредненном списке. При этом различие между семьями в целом небольшое — от четверти совпадающих по десяткам рангов значений в афразийском до 12 совпадающих рангов из 100 в алтайском. Интересно сравнить суммы по строкам. Мы видим существенно большие совпадения со средним в первой и последней десятках — то есть ранги, отмеченные в усредненном списке, лучше всего работают в качестве эталона для самых стабильных и для самых нестабильных значений.

Таблица 14.

	AF	ST	IE	NC	KS	UR	DR	KT	AA	AUS	AN	TH	Atl	Alt	Σ
1—10	5	4	4	5	5	4	4	4	8	6	6	8	8	7	78
11—20	7	9	8	7	7	7	8	8	8	9	8	8	8	10	112
21—30	9	8	9	8	9	10	9	8	6	8	8	9	10	9	120
31—40	7	9	10	9	8	10	8	9	9	10	9	7	10	8	123
41—50	9	8	8	9	7	9	7	9	9	7	10	9	10	10	121
51—60	6	7	8	10	10	8	9	9	8	8	9	10	8	9	119
61—70	8	9	7	8	8	9	9	9	9	9	9	10	8	10	122
71—80	8	9	9	8	9	9	10	9	9	10	9	8	9	9	125
81—90	9	7	8	10	9	8	8	9	9	9	9	10	8	9	122
91—100	7	6	6	5	7	6	9	7	6	5	7	6	8	7	92
Σ	75	76	77	79	79	80	81	81	81	81	84	85	87	88	

Таблица 15.

	ST	AN	KT	NC	UR	IE	AF	DR	KS	AA	TH	AUS	Alt	Atl	Σ
1—10	9	13	7	6	10	9	12	9	17	24	41	22	20	34	233
11—20	14	11	19	11	18	31	24	18	17	11	13	26	35	27	275
21—30	14	14	18	20	22	19	26	20	31	22	24	24	18	25	297
31—40	14	18	16	22	17	24	17	23	20	20	16	24	22	33	286
41—50	16	28	19	19	22	19	18	15	14	30	25	15	27	23	290
51—60	17	19	17	24	17	16	12	24	24	18	18	22	15	19	262
61—70	17	22	23	17	20	17	22	19	15	28	23	21	15	29	288
71—80	17	17	22	17	24	18	16	27	24	18	16	29	22	24	291
81—90	15	13	16	21	14	15	22	14	29	25	25	34	31	29	303
91—100	15	15	15	15	16	12	21	25	17	14	13	7	27	31	243
Σ	148	170	172	172	180	180	190	194	208	210	214	224	232	274	

Если мы применим другой критерий и будем учитывать не только сам факт отклонения, но и его величину (аналог строки 2 в иллюстративной таблице 13), существенно поменяется относительная степень близости различных семей к среднему (табл. 15).

По-прежнему наиболее сильно отличаются атлантические и алтайские языки, остальные же языки перегруппировались. В атлантических языках значения первой десятки не просто соответствуют второй десятке в среднестатистическом списке — они в среднем отличаются по рангу на три десятка каждое! Неизменным остаются более высокие суммарные индексы стабильности для первой и последней десятков.

Применим более толерантный критерий — будем фиксировать отличия только в том случае, если значение входит в другую двадцатку, и при этом не будем учитывать радикальность этих отличий (см. пример в строке 6 табл. 13) — табл. 16.

Порядок языков опять поменялся (атлантические языки по-прежнему наиболее сильно отличаются). Первая и последняя десятки по-прежнему дают наименьшие отклонения в рангах стабильности, причем следующими идут вторая и предпоследняя де-

Таблица 16.

	ST	NC	KT	AF	TH	KS	AA	DR	AN	IE	AUS	Alt	UR	Atl	Σ
1—10	3	1	3	3	8	4	5	3	4	4	4	5	3	5	55
11—20	5	5	6	5	3	5	4	7	5	8	7	10	6	7	83
21—30	7	6	5	9	7	8	5	7	6	9	7	8	9	7	100
31—40	7	7	8	6	6	6	8	7	9	9	10	5	9	10	107
41—50	5	5	7	7	8	4	8	6	9	7	7	9	9	8	99
51—60	7	8	8	5	9	9	7	9	7	6	8	7	7	6	103
61—70	5	8	8	7	6	6	8	8	6	6	6	5	8	7	94
71—80	8	6	5	8	6	8	6	7	8	8	9	9	9	9	106
81—90	5	8	5	6	6	7	8	4	6	5	7	7	6	7	87
91—100	4	3	3	5	3	5	4	6	4	3	2	4	5	6	57
Σ	56	57	58	61	62	62	63	64	64	65	67	69	71	72	

Таблица 17.

	TH	DR	KT	UR	IE	NC	KS	Alt	ST	AN	AA	AUS	AF	Atl	Σ
1—10	5	0	0	1	0	0	2	1	0	0	2	2	0	4	17
11—20	1	2	1	1	4	0	1	5	1	0	0	3	3	3	25
21—30	4	3	2	2	2	3	5	1	2	1	4	4	4	3	40
31—40	2	4	3	3	5	4	3	3	1	4	2	5	4	7	50
41—50	5	2	6	4	3	7	5	5	7	7	7	2	3	5	68
51—60	6	1	2	4	7	6	6	4	4	5	5	5	2	4	61
61—70	4	3	5	3	3	3	2	1	3	5	5	3	6	6	52
71—80	3	4	3	3	3	2	4	3	1	2	2	3	1	4	38
81—90	3	1	1	1	1	2	3	4	2	0	2	5	3	4	32
91—100	1	2	1	0	0	1	1	3	1	0	1	0	2	4	17
Σ	34	22	24	22	28	28	32	30	22	24	30	32	28	44	

сятки. Варьирование относительно усредненного списка по-прежнему исключительно высоко. Так, ни одно из 10 значений, составляющих в алтайских языках вторую десятку, не входит в первую двадцатку среднестатистического списка. В таблице мы не видим ни одного ноля и видим только одну единицу: в северокавказских языках 9 слов из 10 самых стабильных слов входит в первую двадцатку усредненного списка.

И в заключение приведем наиболее толерантную суммарную таблицу. В ней в качестве отклонений фиксируются только случаи вхождения значения не в ту половину. Данная таблица, по-видимому, наиболее интересна для оценки возможности использования только первой половины списка в лексикостатистике (табл. 17).

Отметим, что суммы по столбцам принципиально четные: если значение пропало из первой половины, то оно по определению заменилось другим. Это значит, что, например, в тайских языках, которые в этом списке наиболее близки к среднему, 34 значения из 100 оказались не в той половине, а поскольку количество отклонений в первой и второй половинах списка принципиально равно, можно заключить, что из 50 наиболее стабильных

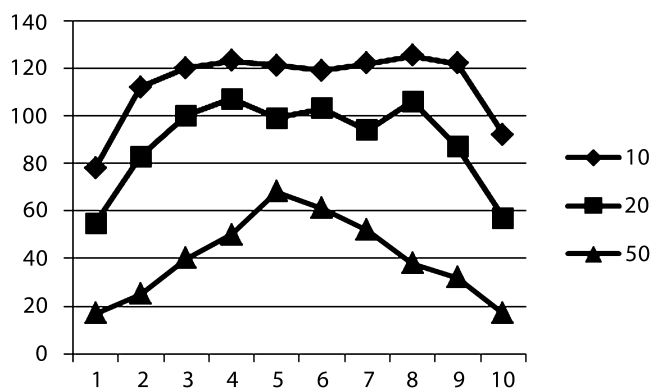


График 1.

слов 18 (приблизительно треть) не являются таковыми в усредненном списке. В атлантических же языках эти два списка не совпадают почти на половину. В таблице наконец появились нули. Например, все значения первой десятки в индоевропейских языках входят в первую половину усредненного списка, а последней десятки — во вторую.

Суммарные данные по строкам очень интересны. Представим на графике (график 1) три кривые: отклонения в пределах десятков, в пределах двадцаток и в пределах двух половин (без учета их радикальности).

График 1 показывает четкую закономерность. Наиболее универсальны начало и конец списка. Наиболее стабильные значения универсального списка являются наиболее стабильными в большинстве языковых семей. Наименее стабильные значения являются таковыми в большинстве семей. Средняя часть «универсального списка» наименее универсальна.

Подведем итог. При таком сильном варьировании индексов стабильности по семьям трудно рассчитывать на эффективное использование усредненных рангов значений для определения языкового родства. Утешает лишь тот факт, что даже в атлантических языках усредненный список показывает несколько большую степень стабильности в первой половине, чем во второй. Возможно, иерархию стабильности значений продуктивно использовать для определения места изолированных языков в семье или макросемье, например для сопоставления языка лимба с различными ветвями атлантических языков и языков мель, однако в этом случае полезнее, вероятно, использовать именно атлантический список, а не «универсальный».

Можно, впрочем, попробовать другие критерии стабильности. Один из альтернативных критериев я использовал для обработки атлантических стословников. Он элементарно прост. Можно, вслед за С. А. Старостиным, брать для каждого значения только максимально представленные в семье корни. А можно просто брать количество корней для каждого значения в семье языков. Умозрительно легко допустить, что эти два критерия дадут разные результаты. Представим 2 теоретические ситуации: (1) в 10 языках 5 корней по 2 языка в каждом — индекс устойчивости 2 по Старостину, индекс 5 — по другой методике; (2) один корень представлен в 6 языках, а в оставшихся 4 языках — 4 разных корня: индекс устойчивости 6 по Старостину (существенно выше) и 5 по сумме корней (тот же индекс). Поэтому неожиданным для меня оказалось то, что эти две разные техники определения рангов дают для атлантических языков абсолютно одинаковые суммарные результаты при исчислении по двадцаткам и по двум половинам, а при исчислении по десяткам оказывается, что вторая методика дает даже большее приближение к рассмотренной выше усредненной последовательности в языках мира, чем первая: если по методике Старостина мы вынуждены были констатировать, что 87 значений ока-

зываются не в тех десятках, что значения усредненного списка, то при ранжировании по количеству разных корней таких значений только 80, что является средним отклонением для языков 14 семей.

Давая те же результаты или даже улучшая их, сумма разных корней как критерий более прозрачна, поскольку учитывает все корни определенного значения, а не один из них (максимальный), и, в частности, учитывает многочисленные изолированные корни, представленные в атлантических языках, то есть корни, отмеченные только в одном из 40 языков. Исследование таких изолированных корней, как представляется, с полным основанием может быть выделено в отдельную область компаративистского исследования. Рассмотрим это положение подробнее.

3. Об изолированных корнях

Буквально в каждом из атлантических языков, в том числе и в тех, которые входят в достаточно «молодые» подгруппы (с совпадением по Сводешу более 50%), наблюдается исключительно высокий процент «уникальной» базисной лексики, которая не находит параллелей **ни в одном** другом атлантическом языке, в том числе и в языках своей подгруппы. Само по себе наличие изолированных слов тривиально для любого языка. Нетривиален высокий процент таких слов. В лимба уникальных корней в стословнике — 47, то есть практически половина списка! Казалось бы, если группа настолько гетерогенна, что само ее существование ставится под сомнение, то наличие исключительно больших пластов уникальной лексики вполне закономерно. Конечно, это так. Но эта ситуация вызывает вопросы, на которые нет вразумительных ответов. В частности, мы имеем все основания задуматься — откуда взялась эта масса уникальных базисных слов?

Первое естественное объяснение — заимствования. Но мы их предварительно исключили. В списках нет ни французских слов, ни арабских. Возможно, там есть еще не опознанные нами древние заимствования из манде, но их не может быть много.

Второе объяснение — лексические дериваты и семантические сдвиги (один из типичных примеров для атлантических языков: «калебаса» > «сердце»). Для большинства наших случаев эти объяснения не годятся — речь идет о непроемных базовых лексических корнях.

Может быть, в этих уникальных корнях отражена древняя нигеро-конголезская лексика, которая сохраняется только в одном из атлантических языков, а в других — исчезает? Отдельные такие случаи вполне вероятны, но предположение о том, что в отдельных атлантических языках, в стословниках Сводеша, отражается 600 нигеро-конголезских корней помимо корней, отраженных в нескольких языках, вряд ли можно рассматривать всерьез.

Тогда что это? Субстрат? Но шесть десятков субстратов — не много ли это для западноафриканского атлантического побережья? С субстратами тоже не получается. Что еще осталось у нас в запасе в качестве возможных объяснений? Может, это до-нигеро-конголезские корни, которые дошли до нас через многие тысячелетия, и в таком случае, выделяя такие изолированные корни, мы получаем ключ к самому древнему лингвистическому наследию человечества? Уже тот факт, что за неимением лучшего мы не можем отбросить эту провокационную космическую гипотезу, говорит о том, что изолированными корнями имеет смысл системно заниматься.

Стратегия компаративистики побуждает нас к поиску соответствий, сходств, сближений, параллелей. Если они не находятся, мы выносим уникальные корни за скобки, отправляя их «в отвал» — и, вероятно, напрасно. Такие корни имеет смысл выделять и

Таблица 18.

10–16	many	full	know	see	woman	sit	night	sand	stand					
9	eat	foot	horn	kill	little	moon	name	neck	root	round	say	sleep	star	water
8	come	feather	fire	nail	rain	red	road	white						
7	blood	dry	good	heart	knee	liver	long	one	seed	skin	stone	walk		
6	big	die	fat	fish	hair	hear	leaf	louse	mountain	swim	two			
5	all	belly	bird	cold	dog	fly	hand	hot	meat	mouth	tree	what	who	
4	ashes	black	breast	burn	cloud	drink	give	lie	nose	smoke	sun	tooth		
3	bite	bone	go	man	person	tail	that	you sg.						
2	egg	new												
1	ear	eye	ground	head	I	this	We	you pl.						
0	bark	no	tongue											

Таблица 19.

Limba	47	Joola-Kwaatay	22
Gola	34	Kisi	22
Bijogo	31	Buy	20
Nalu	29	Konyagi	19
Sua	29	Bayot	19
Wolof	28	Bedik	18
Balant	26	Sereer	17
Sherbro	25	Peul	14
Nyun	23	Biafada	12
Temne	23	Karon	12
Jaad	22	Baga-Maduri	11

публиковать — с тем, чтобы специалисты в смежных областях компаративистики могли провести их экспертизу.

К тому же количество изолированных корней в том или ином значении можно рассматривать как вполне независимый критерий определения ранга нестабильности внутри списка Сводеша. Ранжируем по этому критерию атлантический список — табл. 18 (цифры обозначают количество изолированных корней для каждого значения в 34 языках).

Приведем атлантические языки, в которых количество изолированных корней превышает 10 (табл. 19).

Отметим, что данные, приведенные в табл. 18, получены на представительном, но все же ограниченном списке, включающем 35 языков. Если же учесть все корни во всех атлантических языках и языках мель, количество изолированных корней еще более возрастает, причем возрастает существенно. Приведем один характерный пример — атлантические корни для слова со значением «женщина».

В нашей с Г. Сежерером базе собраны лексические данные по 64 языкам (атлантические и мель)¹⁴. В этих языках выделяется 28 различных корней со значением «женщина».

¹⁴ Число не учитывает многочисленные диалекты и говоры, которые также представлены в нашей базе.

Естественно предположить, что сравнительно-этимологический анализ позволит существенно сократить это число, но два десятка корней в списке останется. Если мы вернемся к первому алгоритму установления сходства, который обсуждался в связи с порогом случайных совпадений (а именно: «разрешается сблизать любые согласные внутри локального ряда»), то теоретически при четырех локальных рядах у нас возможны 16 комбинаций структуры CVC. Добавим к ним 4 комбинации структуры CV. Итого получаем 20 теоретически возможных структур, к которым сводятся слова всех языков мира в рамках выделенных комбинаций. В атлантических корнях со значением «женщина» представлены 16 из 20 возможных структур, причем многие структуры представлены разными корнями¹⁵. Во-первых, это возвращает нас к утверждению, что в этом значении в атлантических языках мы легко можем найти звуковую параллель для английского, русского или практически любого другого языка. Во-вторых, и это для нас главное в данном параграфе, даже если мы не реконструируем корни для праатлантического уровня, это не избавляет нас от необходимости ответить на вопрос: откуда эти корни появились?

Приведем из базы слова со значением «женщина», сгруппировав их по корням в рамках указанных 20 звуковых комбинаций¹⁶.

- PP: бага koba bum, темне bom.
 PT/ПК: balant fag / far.
 PT: темне -bèrà.
 PC: limba basa.
 PK: темне boko / borko.
 PV: sua mfo, nalu fay.
 TP: palor 'be-leb / 'be-lep, ndut 'be-lep, laala leb / 'be-te, safen bi-tip, palor bi-tif ('femme adulte'), noon 'be-te' / bè-dop, peul debb-o / rew-be.
 TP: konyagi -nəm, bapen -nam, bullom -noma, sherbro -nóma (но sherbro no-ma / à-ma – возможно, корень *ma) .
 TP / CP: joola fogny naaba, bayot naw.
 TT: kisi nord lààndò-, kisi sud láno, бага Fore laŋən, бага koba rani, бага sitemu ran, бага maduri / landuma raani.
 TT / (TP): balant niin / ninn / nim, gola -ni(e) — ср. бенуэ-конго *nine / *nene / *nina.
 TT / CT: joola (разные языки) naale / naare, joola banjal -njalena, joola fogny -nalen, biafada naali, bayot -rər.
 TK: bainunk (разные языки) -dik-am, kasanga / kobiana lik-aam, bullom -lak-an, sei lag-an.
 TK / CK: limba nani, joola ejamat nakan, joola kasa nakunta.
 TV: bullom la, bedik lui-m / -nuí-m, limba nda, wolof ndaw 'jeune fille' — ср. Jukunoid *ndá.
 CP: konyagi -səvælá, bedik sʷámبالàŋ / -sámbar, jaad caafe, gola dzàve.
 CP: bapen -suwon.
 CT: gola dzân.
 CC: manjak jije, joola ejamat ?änjä'ijä' / nejäu.
 CC: nalu θay.
 CK: wolof jigéen.

¹⁵ Так, комбинация «дентальный» — «V — лабиальный» отмечена в трех корнях, родство которых далеко не очевидно.

¹⁶ Принятые условные обозначения: P — любой лабиальный, T — любой дентальный, C — палатальный, K — велярный. Например, обозначение PT будет обозначать присутствие начального лабиального и конечного дентального в структуре CVC.

- СК: basari *tóxár* / *sóxár*, bedik *sóyár*, konyagi *ryónkáx*, tanda *zogari*, bainunk-Kas *toxol*.
 СК: joola fogny *seek*, joola ejamat *seh*.
 CV: pepel *šo*.
 КТ: bijogo *-kanto* — ср. Benue-Congo **-kwazi*.
 КС: бага Mboteni *gìsì*.
 КС: joola gusilay *koona*.
 VT: joola (разные языки) *aal* / *aar*, karon *haal*, manjak / mankany *-aats* / *aaf*, manjak churo *aah*, pepel *arş* / *aş*.

Итак, в атлантических языках для значения «женщина» отсутствуют только корни структур ТС, КР и КК, но и в других языках эти структуры корней низкочастотны (подробнее об этом — см. [Pozdniakov & Segerer 2007]).

Вернемся в заключение к исходному вопросу: как быть, когда «нельзя», но «очень хочется»? Из всего вышесказанного вроде бы следует, что все равно «нельзя». На глубине ВЕ у нас нет, к сожалению, надежных инструментов, чтобы строить классификации на основе внешних сходств в базисной лексике, и единственный путь — классическая ступенчатая реконструкция. А пока ее нет, любые разумные критерии — морфологические изоглоссы, фонологическая статистика, данные внешнего сравнения, сопоставимость «слабой» лексики в неконтактирующих языках и многие другие — не являются ни «контролирующими», ни «сопутствующими», а имеют не меньший вес, чем сходства в стословниках.

Д. Р. Лецинер
 Google Inc. (Маунтин-Вью, США)

Заметки к статье К. И. Позднякова «О пороге родства...»

Мне хотелось бы обсудить некоторые выводы статьи, обобщения, сделанные на их основе, методы, приведшие к этим выводам, и границы применимости этих обобщений. Позволю себе кратко перечислить разделы статьи и те ее выводы, которые привлекли мое внимание.

Статья посвящена экспериментальному исследованию (на материале атлантических языков) следующих вопросов:

- существует ли некое минимальное количество сходств в списке Сводеша по двум языкам, которое позволило бы нам подозревать, что эти языки родственны?
- можно ли использовать индексы стабильности лексических значений для процедур определения языкового родства?
- каково происхождение аномально большого количества «уникальной» базисной лексики в отдельных атлантических языках?

Автор, в частности, приходит к следующим выводам:

- на глубине старше индоевропейской (ВЕ) у нас нет надежных инструментов, чтобы строить классификации на основе внешних сходств в базисной лексике, и единственный путь — классическая ступенчатая реконструкция;
- массовое непарное сравнение может гарантировать «доказательство» любого родства;
- при существующем сильном варьировании индексов стабильности (в версии С. А. Старостина) по семьям трудно рассчитывать на эффективное использование усредненных рангов значений для определения языкового родства;
- изолированными корнями имеет смысл системно заниматься. Такие корни имеет смысл выделять и публиковать — с тем, чтобы специалисты в смежных областях компаративистики могли провести их экспертизу.

Представлен подробный анализ варьирования индексов стабильности. Предложена и модификация методики Старостина для расчета индексов стабильности: ранжирование по количеству разных корней для каждого значения в семье языков. Эта модификация представляется весьма интересной. С выводом о нежелательности игнорирования изолированных корней трудно не согласиться. Тот факт, что на вопрос об их природе не находится очевидного ответа, делает эту проблематику еще более интригующей.

Значительно труднее, однако, согласиться с логикой автора в тех пунктах, где он делает выводы о применимости тех или иных эмпирических данных к вопросу определения языкового родства. Сразу должен сказать, впрочем, что основная часть раздела 1, посвященная оценкам вероятности случайного фонетического сходства слов, не вызывает никаких возражений. Напротив — проведен убедительный, аккуратный и впечатляющий анализ, получаемые оценки следует принять¹. Однако, на мой взгляд, из них никоим образом не вытекают окончательные выводы, приведенные выше. Также и вывод о неприменимости индексов стабильности представляется несколько поспешным и недостаточно обоснованным. О индексах стабильности будет ниже сказано несколько слов. Но главным образом я хотел бы обсудить логику, связанную с применимостью процедур массового и попарного сравнения.

Напомню, что исходный вопрос, поставленный автором, следующий: «**Существует ли некое минимальное количество сходств в списке Сводеша по двум языкам, которое позволило бы нам подозревать, что эти языки родственны?**» Однако вопрос этот, в той форме, как он поставлен в статье, имеет значение в первую очередь для случая сравнения **пары изолированных** далеких языков (скажем, шумерского и сонгаи). Именно в этом случае количество сходств при попарном сравнении списков является единственным доступным критерием для принятия решения. И действительно, в этом случае — вряд ли кто против это-

¹ Одним исключением, впрочем, является трактовка локальных рядов согласных при наличии структур VC и CV — нулевой согласный не трактуется автором как самостоятельный ряд. В статье не предлагается никаких критериев сравнения локальных рядов с нулевым рядом. Если допустить его сравнение с любым локальным рядом (такую возможность автор далее упоминает), то уровень случайности действительно зашкалит. Если же не разрешать сравнение с нулевым согласным, тогда проводимые вычисления вероятностей нуждаются в корректировке в сторону понижения.

го возразит — метод поиска фонетических сходств в стословном списке не является надежным инструментом установления даже и гипотетического родства. Однако — именно постольку, поскольку на этот счет среди лингвистов имеется общее согласие — никто из них и не применяет метод в этом варианте (я не имею тут в виду случаи очевидно близкого родства — когда «минимальное количество» действительно «существует»). На практике, область его применения другая — а именно, когда сравниваются:

- две установленные языковые семьи друг с другом;
- несколько установленных языковых семей друг с другом;
- или в крайнем случае (достаточно редко) — один изолированный язык с установленной языковой семьей.

В случае же сравнения языка (или группы языков) с **языковой семьей** — критерий сходства должен быть другим. А именно, допустимыми сближениями должны считаться только сближения с **праязыковой** лексикой семьи (в случае же сравнения двух или более семей — только сближения праязыковой лексики с праязыковой лексикой) — «исконные» сближения. Те сближения, для которых может быть показано, что это не так, должны отбрасываться. Те, для которых их «исконный» (праязыковой) характер сомнителен, должны и учитываться, как сомнительные.

Приведем табл. 7 из статьи (для удобства сопоставления таблиц я переупорядочил их строки в порядке английского списка слов), см. стр. 212. Я выделил серым фоном те значения, которые повторяются хотя бы в паре списков из приведенных шести. Обращает на себя внимание следующее:

- в шести (из 41) «самых похожих» списках сближений 5 значений (black, breast, me, white, you sg.) встречаются чаще чем дважды. Хочу заметить, что я не дал себе труда погрузиться в сравнительное изучение атлантических языков (и никогда ранее ими также не занимался), поэтому любое мое впечатление о сравнении атлантической лексики свободно может оказаться ошибочным. Тем не менее, хотя сближения для black, breast, me, white, you sg. и производят впечатление исконных атлантических корней, я не уверен, относятся ли **banj** для black, **n-bin** для breast к тому же исконному корню, что и прочие формы (если же нет, то их следует исключить). Остается следующее число сближений: black — 3, breast — 4, me — 3, white — 3, you sg. — 3. Итого 16;

Таблица 7а.

Temne	
ashes	r-os
bark	foki
bone	ka-bant
fly	falɪr ^w
go	kɔ
little	lɔl
me	mi
one	in
sand	a-sənt
seed	ma-sənt, ma-ther
woman	bom

Nalu	
black	bal
big	wɔk? -baŋ?
bone	wol?
breast	n-bin?
drink	dig
fire	mɛt?
go	kɔ
good	kuun?
star	ma-saar
tree	n-ti?
we	biye?
you sg.	yi

Biafada	
black	baŋ?
breast	bəl?
fire	fuuru
fish	yesä / bwa-?
fly	fiili
ground	gal?
little	tiiti?
me	mma
mouth	mməsə
nose	nasin
rain	səna?
say	cii, re
white	paas
you sg.	yi

Limba	
ashes	hɔtɔ? ~ fɔtɔ?
belly	fɔle
black	boloi
breast	hu-bili
drink	tie?
eye	hoya ~ foya?
fish	fee
kill	kɔra?
tooth	hu-titi
you sg.	yi

Bedik	
black	-bála
breast	ɛ-bɔr
burn	o-bód
eye(s)	gi-nguùs?
hand	ɛ-kònd
hear	o-hér
knee	é-dɔl-kɔlɔ?
liver	i-dámbél?
me	-mé- / mé-
meat	ɲás?
mouth	bɔ-más
white	-fɛsà / -pɛsà

Mankany	
big	-mag
breast	p-wil?
hair	ka-weel?
leaf	ka-toh?
liver	u-deeb
meat	u-ye-maɕ
nail	ka-ñiiri
nose	b-yiis?
what	we
white	p-faaθ

- из остальных 13 «парных» сближений (а именно: ashes, big, bone, drink, eye, fire, fish, fly, little, liver, meat, mouth, nose) я бы (с той же степенью достоверности) усомнился в по крайней мере в 5 (ashes, bone, eye, fire, little). Подозреваю, что статистически моя оценка близка к правде. Хочу заметить, что отсутствие внутриатлантического родства в данном случае не говорит о необходимости отвергнуть данное сходство с английским целиком — оно только говорит о необходимости отвергнуть **по крайней мере одно из двух** этих сближений. Ито-

го, верхняя оценка числа возможных сближений: $(13 - 5) * 2 + 5 = 21$;

- что же касается остальных 25 «одиночных» сближений — у меня нет материала, чтобы судить об исконности (праатлантической древности) этих корней (однако для большинства из них он, вероятно, есть у автора статьи). Из «общих» соображений я бы не думал, что процент исконных среди них выше, чем достоверно исконных парных сближений. Если это так, получаем верхнюю оценку числа исконных одиночных сближений: $25 * (13 - 5) / 13 \approx 15$;

- итого имеем верхнюю оценку суммарного числа «возможных» исконных сближений на 6 списков: $16 + 21 + 15 = 52$. В среднем же — 8,5% (но это **на шести самых лучших списках** — а не в среднем по семье, как подсчитано в статье). По методике же статьи, в этих списках находится 69 сближений, в среднем — 11,5%. Если мы примем тот же коэффициент завышения и по остальным 35 спискам, то цифра «в среднем 8,5% случайных сходств с английским» превратится в $8,5\% * (8,5\% / 11,5\%) = 6,3\%$.

Казалось бы, 6,3% все еще вполне себе отличная от нуля цифра — хотя уже как-то легче поверить в ее случайность... И действительно, сам автор и приводит один из решающих аргументов против предположения о родстве атлантических языков с английским (в примечании 8): «Отметим, что сравнение атлантических слов с английским системно дает существенно больше сходств, чем с французским или с русским. Мне не представить, с чем это могло бы быть связано». Я тоже затрудняюсь сказать, с чем это связано (вероятнее всего, случайность; если бы мы имели дело с действительным родством, это теоретически могло бы объясняться большим фонетическим или лексическим консерватизмом английского по сравнению с русским и французским — впрочем, мы сейчас даже не говорим о том, что по отношению к конкретной паре английский—русский такое утверждение само по себе вряд ли справедливо), однако очевидно, тем не менее, что наличие подобной ситуации с несомненностью **доказывает случайность** сходства атлантических слов с английским (или, как максимум — наличие «атлантического субстрата» в английском или германских языках; впрочем, насколько я понимаю, о наличии такого субстрата науке на сегодня ничего не известно). Если бы сходжения объяснялись исконным родством, эта близость была бы **воспроизводима** на большинстве индоевропейских языков (а их, как известно, немало, и достаточно разнообразных — а тем самым есть материал для накопления статистики).

В действительности следует усреднить процент сходств не только по подгруппам атлантических языков (именно по подгруппам, а не по отдельным языкам: чтобы не давать неоправданно большого веса ни одной большой совокупности похожих — близкородственных — языков), но и по подгруппам индоевропейской семьи. И далее, если подсчитанный таким образом средний процент будет **статистически достоверно превышать** средний процент при сравнении с **другими** семьями языков (т. е. равновероятными конкурирующими ги-

потезами) — тогда можно говорить, что есть подозрение на наличие глубокого родства (или же древних контактов). Это — а не заранее фиксированный «единый» числовой порог — и будет правильным¹ тестом на родство. При отсутствии же достоверного превышения из подсчетов никакого вывода сделать нельзя. Тогда априорная «равновероятность» гипотез превращается в такую же апостериорную «равновероятность». Однако хочу подчеркнуть, что апостериорная «равновероятность» не является свойством множества гипотез, как такового: она является свойством множества гипотез **и процедуры**, примененной для его анализа.

Фактически аргументы автора статьи сводятся к тезису о большой зашумленности (и тем самым статистической недостоверности) процедуры **парного** сравнения языков на основании фонетического сходства 100-словных списков. С этими аргументами, применительно к обсуждаемой в статье процедуре, следует согласиться². Увы, обсуждаемая процедура плоха не только этим, но и тем, что она отличается от лингвистически оправданной процедуры (см. выше). Но не позволит ли лингвистическая коррекция процедуры исправить и ее статистические недостатки (в том числе и в варианте т. н. массового сравнения)? Рассмотрим этот вопрос.

В этой связи я бы попробовал также уточнить утверждение: «у самого Гринберга не было никакой формализованной методики сравнения». Формально говоря, с этим нельзя не согласиться: формализованной методики сравнения у Гринберга не было. Однако я думаю, что некая **неформализованная** методика у него была — не сводящаяся «лишь» к его энциклопедическим знаниям и исключительному исследовательскому таланту (которые действительно формализовать отнюдь не просто), но в том числе и к более приземленным вещам, которые, вообще говоря, вполне могут поддаваться формализации. Более конкретно, я

¹ Не «единственно правильным» — могут, вообще говоря, существовать и другие разумные тесты (см., в частности, ниже) — а просто «статистически правильным».

² Я в данном случае соглашаюсь не со всеми конкретными обобщениями автора относительно применимости оценок зашумленности сравнения списков к решению вопроса о надежности количественных критериев родства (и в частности, с выводом о полной неработоспособности указанной процедуры «на глубине старше индоевропейской» — должен заметить, что, например, для ностратических языков подобная процедура, при усреднении, работает очень устойчиво), и не с конкретным выбором процедур фонетического сравнения, но лишь с общими оценками вероятности фонетических сходжений (с оговоркой, упомянутой в сн. 1 на стр. 211).

думаю, что именно соображения, подобные изложенным выше, находились в ее основе (и они, собственно говоря, и представляют собой «метод массового сравнения» *per se*). Рассмотрим их суть несколько подробнее.

Пусть нам дано некоторое количество «гипотетических» языковых семей. В отличие от предыдущего рассмотрения, эти семьи не являются «установленными». В данном случае важно не то, что мы, может быть, не полностью уверены в реальности этих семей — пока что можно ограничиться и гипотезой — а то, что у нас нет достоверной информации о том, какие корни в каждой из этих семей являются или не являются исконными. Поэтому, для того чтобы проводить осмысленную процедуру сравнения семей, нам придется заменять эту «внутрисемейную» информацию статистическими гипотезами, основанными на фонетическом сходстве — так же, как и при сравнении семей между собой.

В статье приведены достаточно простые и разумные способы оценки вероятности случайного фонетического сходства — на самом деле не только слов 100-словного списка в **среднем**, но и каждой **данной** пары «сходных» слов. Понятно, что эти числа (как и указано в статье) зависят, в частности, от частотности фонем в данных языках — а также от их позиционного распределения и сочетаемости. Например, вероятность случайного совпадения слов, в которых обнаруживается редкая фонема или сочетание фонем, будет крайне мала, а вероятность случайного совпадения слов типичной структуры — куда больше. (Заметим, что того же нельзя сказать про вероятность совпадения в результате независимого заимствования.) Для нас важно то, что эти числа достаточно несложно получить и воспользоваться ими. И что далее важно: если у нас имеются, скажем, **три** фонетически сходных слова с одинаковым значением в трех языках — что мы можем сказать о вероятности каждого из них быть «исконным» словом в своем языке? (Мы предполагаем, что все три языка родственны друг другу «на одном уровне» — т. е. являются продуктом одновременного распада праязыка на три ветви.) Здесь на самом деле два вопроса: 1) какова вероятность, что по крайней мере два слова из трех восходят к общему корню? 2) если (1) верно — какова вероятность того, что **данное** слово к нему восходит? Я бы затруднился дать точное определение этих величин (не так просто сформулировать полный набор необходимых предположений), но следующие простые оценки будут, видимо, достаточно разумны:

- вероятность (1) = $1 - (1 - p_{12})(1 - p_{13})(1 - p_{23})$
- вероятность (2) = $1 - (1 - p_{12})(1 - p_{13})$
[это для слова в языке 1]

Здесь p_{12} и т. д. — **апостериорная** вероятность родства слов в языках 1 и 2. Она зависит от **априорной** вероятности родства и от вероятности случайного сходства (включая сюда и вероятность общего заимствования!) по следующей формуле:

- $p_{12} = p_{apr} / (p_{apr} + p_{rand}(1 - p_{apr}))$

Здесь мы приняли, что в случае родства вероятность сходства = 1, что, конечно же, неверно. Если же оценить ее как p_0 , тогда будет:

- $p_{12} = p_{apr} * p_0 / (p_{apr} * p_0 + p_{rand}(1 - p_{apr}))$.

Обратим внимание: если $p_{apr} * p_0 \ll p_{rand}$, то $p_{12} \approx p_{apr} * p_0 / p_{rand}$ — и это именно та ситуация, которую мы имеем при достаточно дальнем сравнении. (Если же мы заранее отвергаем всякую возможность родства — т. е. если $p_{apr} = 0$, то и никакое количество сколь угодно убедительных данных нас все равно не убедит; тогда и $p_{12} = 0$.)

Отсюда видно, во-первых, что от оценки p_{apr} зависит очень многое. (И откуда взять ее?) Во-вторых, видно, что если у нас имеется не три, а **много** сходных корней, тогда даже и малые (но не нулевые, конечно!) значения p_{apr} нам не так уже страшны.

А именно, пусть у нас имеется k языков, а вероятность попарного родства для всех пар одинакова (в общем случае это неверно — p_{rand} зависит от конкретных форм слов) и равна ϵ . Тогда вероятность (1) — обобщая формулу на k языков — равна $1 - (1 - \epsilon)^{k(k-1)/2}$. Если ϵ очень мало, это будет приблизительно $\epsilon k(k-1)/2$ — то есть, взяв приблизительно $\epsilon^{-1/2}$ языков, мы уже в неплохой форме. Если, скажем, $\epsilon = 1\%$, десяти языков нам хватит. (Для вероятности (2), к сожалению, значения будут меньше — порядка $\epsilon k \dots$) Это и есть ключ к правильному массовому сравнению — когда примеры по отдельным языкам используются не скопом и без разбора, а проверяя и контролируя друг друга — и только достоверные примеры используются для дальнейшего сравнения.

Но откуда все-таки взять p_{apr} (и тем самым ϵ)? Очень просто. Не забудем, что мы говорили сейчас об установлении исконного родства **внутри** «готовой» гипотетической семьи. Внутри этой семьи мы уже имеем примерные оценки степени родства ее языков — т. е. среднего количества совпадений в 100-словном списке. И кстати, обычно эти оценки не так уж малы — скорее порядка 10% или выше, и уж никак не 1%. Эта оценка, в первом приближении, и есть p_{apr} . Эту оценку можно уточнить в зави-

симости от индекса стабильности значений слов — каким бы методом это не делать это, для высокостабильных корней различие будет очень заметным. Далее, в зависимости от значений p_{rand} получаем ε — которое, вообще говоря, будет заметно выше p_{app} при низких значениях p_{rand} .

В действительности все эти оценки — пока лишь подготовка к главному: к сравнению **между** семьями. Вот тут-то и должна заработать процедура массового сравнения. Необходимые ингредиенты — **высокие уровни уверенности в исконности определенной лексики** внутри отдельных семей (см. расчеты выше) и **наличие большого числа семей для сравнения**, чтобы на основании тех же рассуждений, что и выше, большого числа фонетически сходных форм в разных семьях и высокого уровня уверенности в их исконности прийти к вы-

воду о вероятном исконном родстве этих форм на макроуровне: уровне родства **между** рассматриваемыми семьями. При попарных сравнениях, по очевидным причинам, такой вывод будет заметно затруднен.

На самом деле, как нетрудно заметить, приведенные выше процедуры являются, в определенном смысле, процедурами «ступенчатой реконструкции». Речь не идет, разумеется, о реконструкции исторической фонетики и установлении праформ — объектом «реконструкции» выступает в данном случае гипотетический список форм-потомков в языках-потомках. Это существенно меньше, чем достоверно установленные праформы и исторические соотношения, но с точки зрения статистики почти столь же хорошо для лексикостатистического анализа с целью выявления дальнего родства.

Г. С. Старостин

Российский государственный гуманитарный университет / РАНХиГС (Москва)

Ответ на статью К. И. Позднякова «О пороге родства...»

Статья Константина Позднякова (далее — КП) вновь поднимает ряд вопросов, чрезвычайно актуальных для современного этапа развития компаративистики, и, поскольку автор во многом отталкивается от недавно опубликованной мной монографии («Языки Африки. Опыт лексикостатистической классификации»), большая часть которой посвящена как раз этим вопросам, я с удовольствием использую эту возможность для того, чтобы уточнить отдельные моменты собственной концепции, а заодно и снять некоторые критические возражения автора или предложить возможные пути решения обозначенных им проблем.

В первую очередь, я бы охарактеризовал статью КП как набор любопытных и во многом полезных размышлений, формально связанных друг с другом, но не имеющих единой общей цели (по крайней мере, очевидной). С одной стороны, автор ставит в начале текста вполне конкретный вопрос — «можно ли классифицировать языки на уровне макросемей, не используя ступенчатую реконструкцию на базе сравнительно-исторического метода?» — и в конце текста получает на него столь же конкретный ответ (нельзя). С другой стороны, остается не вполне яс-

ным, к кому обращен этот вопрос. Активные «сплиттеры» (убежденные противники любых гипотез о дальнем языковом родстве) и без статистических выкладок КП уже давно пришли к тем же самым выводам. «Ламперы»-гринбергианцы, оперирующие псевдометодом массового сравнения, принципиально работают не со списком Сводеша (который, очевидным образом, не удовлетворяет критерию «массовости» сравнения), а с полным корпусом лексического и грамматического материала, и для полемики с ними требуется совсем другая статистика. С таким же полным корпусом предпочитают работать и этимологи-традиционалисты, отказывающиеся замечать какие-либо фундаментальные различия между «обычным» и «макросравнением» и относящиеся к лексикостатистике предельно скептически (такие, как, например, покойный А. Б. Долгопольский), так что их из потенциальных адресатов статьи тоже, наверное, можно выкинуть.

Естественно было бы предположить (и какое-то время мне именно так и казалось), что статья обращена в первую очередь на критику некоторых теоретических положений, вынесенных мной во вводную (методологическую) часть вышеупомянутой

монографии, где предлагается некоторая, сама по себе довольно сложная, процедура установления относительного языкового родства для макросемей на «сверхглубоких» уровнях, не предполагающая при этом полной реконструкции праязыков ветвей-потомков. Эта процедура на определенных этапах действительно предполагает сравнение поверхностной фонетической структуры языковых форм, которое может осуществляться в виде автоматического алгоритма, т. е. претендует на некоторую степень объективности, не попадая напрямую в зависимость от интуиции исследователя-компаративиста.

Однако между тем, как предлагается осуществлять сравнение в монографии «Языки Африки», и той процедурой сравнения, которую тестирует на материале атлантических языков в своей статье КП, есть и немало существенных отличий — таких, которые, скорее всего, фундаментально повлияли бы на результаты, полученные автором. Кратко суммирую здесь лишь некоторые из них.

(1) По утверждению КП, «следует сжиться с фактом, что десяток сходств при реально существующих алгоритмах их выделения — это всего лишь норма случайных созвучий между любыми языками». Но при каких алгоритмах? КП иллюстрирует свое утверждение на примере сопоставления слов английского языка и атлантических языков темне (11 сходств), биафада (12 сходств, из них 4 под вопросом), бедик (12 сходств, из них 4 под вопросом), налу (14 сходств, из них 9 под вопросом), лимба (10 сходств, из них 4 под вопросом), манкань (10 сходств, из них 4 под вопросом). При этом, однако:

- автор использует *чрезвычайно* либеральный алгоритм сопоставления; семь его основных принципов, описанных выше в тексте статьи, которые, помимо прочего, допускают объединение в рамках одного класса чересчур большое число согласных, а также игнорирование «слабых» согласных в середине слова. По сравнению с этим алгоритмом применяемый в нашей монографии подход не позволил бы считать «сходством» такие «англо-темне изоглоссы», как *bark : foki*, *little : lol*, *seed : ma-sənt*, и даже *woman : bom*;
- автор бывает непоследователен в вопросах отделения от корня как «застывших», так и вполне продуктивных префиксов. Например, темне *ka-bant* ‘кость’ сравнивается с английским *bone* с отделением префикса, в то время как биафада *nasin* ‘нос’ сравнивается с английским *nose* несмотря на то, что *na-* в этой форме, судя по источникам, — точно такой же

именной префикс. Такого рода непоследовательности, разумеется, ведут к дальнейшему увеличению случайностей;

- наблюдаются и отдельные, возможно случайные, семантические неточности: например, английское *hand* сравнивается с бедик *ε-kənd*, хотя в словаре М.-П. Ферри однозначно указывается, что *ε-kənd* — это ‘bras, de l'épaule au coude’, в то время как *hand (main)* на бедик будет *ε-ɲēt* и с *hand* несопоставимо.

Таким образом, утверждение о *норме* в десять случайных созвучий справедливо далеко не для всех «реально существующих алгоритмов», а лишь для таких, которые основаны на чрезвычайно дозволительных и размытых стандартах поверхностного сравнения. Если же применять алгоритм, аналогичный описанному в «Языках Африки», то, например, число поверхностных сходств между бедик и английским с двенадцати по версии КП падает до четырех, и т. д. Так что, вопреки утверждению КП о том, что «выделение десятка пересекающихся классов не улучшает дела», количество выделяемых классов на самом деле влияет на результат даже при принятии универсального стандарта — и, на мой взгляд, система, изначально предложенная А. Б. Долгопольским, вполне подходит на роль такого стандарта, соблюдая разумный и типологически обоснованный баланс между «чрезмерной дозволительностью» и «чрезмерной строгостью» сравнения.

(2) «Возникает простой методологический вопрос: как быть, когда, например, класс R, включающий только вибранты, не представлен в одном из языков? В гола отсутствует фонема /r/. Это означает, что /r/ волоф вообще ничему не может соответствовать в гола». Вообще говоря, в рамках нашей системы классов в класс R попадает не только вибрант /r/, но и плавный /l/ (что основано на типологически частотном в языках мира колебании между этими двумя фонемами), который, насколько мне известно, в гола наблюдается, так что данный пример не очень удачен.

Допустим, впрочем, что ситуация сложнее, и в сравниваемом языке /r/ изменился не в /l/, а в /d/, т. е. соотв. фонема действительно перешла в другой класс. В этом случае в рамках принятой в нашем исследовании методологии в дело вступает т. н. *метод ручной коррекции* — переоценка результатов подсчетов «поверхностных сходств», основанных на универсальной системе критериев, с учетом «локальной специфики», т. е. легко обнаруживаемых и описываемых особенностей фонологической системы языка, не «отлавливаемых»

универсальным алгоритмом. Грубо говоря, если фонему /r/ языка А «не с чем сравнивать» в языке Б, «ручная коррекция» находит для нее в языке Б ближайший эквивалент, основываясь на данных типологии фонетических изменений (где «ближайшим соседом» класса R оказывается фонема *d* в классе T). При желании «ручную коррекцию» можно формализовать и более жестким способом (например, с помощью статистических подсчетов частотностей фонем или через привязку к т. н. дистанциям Левенштейна), но в целом все эти проблемы носят технический характер (и уж во всяком случае, представляются тривиальными по сравнению, например, с гораздо более фундаментальной проблемой распределения по консонантным классам такой «экзотики», как щелчковые фонемы).

(3) Разумеется, и при более строгих алгоритмах поверхностного сравнения не только не исключены, но и засвидетельствованы отдельные (редкие) случаи, когда число случайных сходств между заведомо неродственными или чрезвычайно дальнородственными языками превышает ожидания и оказывается сопоставимым с числом неслучайных сходств для других пар языков (хороший пример: 10–12 случаев совпадений между английским и омотским, описанный в недавней работе А. Касьяном). Если на это наложить еще и актуальный для некоторых ситуаций ареальный фактор, то, конечно, само по себе поверхностное сходство двух языков, потенциально родственных на уровне очень старой семьи или макросемьи, серьезным аргументом признать нельзя.

Однако ни в работах С. А. Старостина, оперирующих методами «поверхностного» сопоставления языковых форм, ни в моей монографии, собственно говоря, и не предлагается выдвигать даже предварительные гипотезы о языковом родстве на основании прямолинейного сравнения современных языков, связанных потенциальным родством на «макросемейном» уровне. Вообще, очень важно, чтобы любая ситуация родства, так сказать, осознавалась максимально *динамично* — не на основании отдельно взятого анализа сходств между двумя статическими точками, а в рамках некоторого хронологического пространства, где это родство можно проанализировать в том числе и на предмет естественности его эволюции.

Так, в частности, как уже неоднократно в своих работах отмечал С. А. Старостин, если число совпадений между языками при перемещении сравнения с более современного на более древний уровень остается неизменным, совпадения эти, скорее всего, случайны; если же родство настоящее, число

это должно возрастать (т. е. между древнеанглийским и древнерусским должно быть больше сходжений, чем между их современными потомками, а между прагерманским и праславянским оно должно быть еще больше). Любопытно в этом смысле посмотреть на то, как будет вести себя «английско-темне» сопоставление при выходе, например, на уровень «прагерманско-мель» — подозреваю, что общее число сходжений не только не увеличится, но и, хуже того, на место одних сходжений хотя бы частично встанут другие (поскольку прагерманские эквиваленты современных английских слов будут существенно отличаться по своей консонантной структуре).

В монографии «Языки Африки» неоднократно подчеркивается неправомерность «поверхностного сопоставления» по методу консонантных классов для языков, связанных потенциальным родством, более глубоким, чем уровень группы, т. е. в случае атлантических языков такому сравнению должна предшествовать «очистка» материала — выявление наиболее перспективных претендентов на выражение того или иного сводешевского значения на уровне отдельной группы или ветви. Если такая «очистка» приведет к тому, что число сходжений между условным «пра-мель» и прагерманским уменьшится, а возрастет, причем так, что «английско-темне» параллели не будут *заменены* новыми параллелями, а, наоборот, будут ими *дополнены* — значит, языки мель и германские языки, скорее всего, родственны. Однако с большой вероятностью этого не произойдет.

Теперь несколько слов относительно проблематичности критерия динамической градации, который КП подробно рассматривает на примере материалов статьи С. А. Старостина, подключая к ним собственные данные по атлантическим языкам. Хотя сам КП оценивает полученные выводы пессимистично («при таком сильном варьировании индексов стабильности по семьям трудно рассчитывать на эффективное использование усредненных рангов значений для определения языкового родства»), мне, наоборот, кажется, что выводы только подтверждают значимость метода динамической градации.

Дело в том, что выведенный автором график показывает, что *начальная* и *конечная* секции ранжированного 100-словного списка действительно стремятся к универсальной значимости, а это как раз то, что нам и нужно. По существу, на «среднюю» часть списка в такой ситуации можно вообще не обращать внимания — достаточно посмотреть, сколько сходжений, отловленных на этапе

поверхностного сопоставления, относятся, скажем, к первым 10—15 элементам ранжированного списка, а сколько — к последним 10—15. Грубо говоря, если в ваших сравнительных материалах совпадают слова типа ‘маленький’, ‘круглый’ или ‘гора’, но отсутствуют при этом сходжения в личных местоимениях, числительных 1—2, или таких частях тела, как ‘глаз’, ‘ухо’, ‘зуб’ или ‘рука’, или, если они все же присутствуют, но при этом не обнаруживается явного превалирования сходств второго типа над сходствами первого типа, то ситуация должна расцениваться как подозрительная — своего рода «красная лампочка», которая указывает на то, что не все в порядке и следует обстоятельно протестировать идею возможных контактных связей между языками, или нулевую гипотезу (случайные созвучия).

Рассмотрение ранжированного атлантического списка, представленного КП, показывает, что первые 50 позиций в нем на 30 пунктов совпадают со среднестатистическим 50-словным списком, отобранным для «Языков Африки». С одной стороны, кажется, что это не сильно превышает порог случайности (при котором ожидалось бы порядка 25 совпадений). С другой стороны, нельзя не отметить, что «атлантический список» — понятие само по себе довольно смутное, т. к. исследования самого КП и Г. Сежерера вроде бы говорят о том, что «атлантической» семьи как таковой не существует, и не является ли в этом смысле попытка замерить индекс стабильности по «атлантическому списку» аналогом попытки замерить такой индекс, скажем, по «ностратическому списку», в пределах которого мы зачастую просто не можем с уверенностью говорить об этимологических (и, соответственно, лексикостатистических) связях между сравниваемыми формами?

Чрезвычайно подозрителен, например, тот факт, что в «атлантическом списке» столь низкие рейтинги (19-е и 20-е место соответственно) занимают личные местоимения ‘я’ и ‘ты’; в этом они сближаются разве что с тайскими языками, но в тайских нестабильность личных местоимений, как правило, связана с вытеснением старых форм «вежливыми» новообразованиями — явление, типичное для юговостоазиатского ареала, но вроде бы не очень распространенное в Африке. Уже один этот факт заставляет подозревать что-то «неладное» — например, неверную расстановку когнаций, или же просто неправильность подхода к атлантическому списку как к единой сущности. Может быть, следовало бы вместо единого атлантического списка просчитать рейтинги стабильности по твердо установленным группам (мель, бак и т. п.).

Кстати говоря, замечание КП относительно того, что «в самом конце списка [по различиям коэффициентов первой и второй половин. — Г. С.] оказываются все три африканские семьи плюс австралийская», совершенно не обязательно связано с тем, что за пределами Евразии языковые семьи ведут себя как-то по особому в плане среднестатистической устойчивости своих индексов; на самом деле все эти семьи объединяет не столько их «неевразийность», сколько элементарная неразработанность этимологии. Афразийская семья — это макросемья, где определение когнаций на самых верхних уровнях (кушитские и омотские языки + все остальные) чрезвычайно сложно и противоречиво; койсанская семья, как и атлантическая, скорее всего, представляет собой несколько семей, а про австралийскую этимологию и говорить нечего. Так что, вне всякого сомнения, «накопление материалов по стабильности базовой лексики должно быть продолжено», но и с тем, что уже накоплено, как показывает текущий опыт работы с койсанскими и нило-сахарскими языками Африки, работать вполне можно.

Что касается последнего пункта статьи КП, о корнях-изолятах, то я полностью согласен с идеей автора о том, что такие корни следует как минимум инвентаризировать и публиковать. Более того, в рамках моего собственного исследования, объединяющего этимологический подход с лексико-статистическим, ровно это и делается — поскольку в основу анализа кладется не этимологический корпус, а полный список Сводеша, все такие корни оказываются как минимум замеченными, как максимум же иногда проводится попытка их этимологизации, например, как возможные заимствования (не обязательно из древних субстратов, иногда получается и объяснить их как относительно недавние заимствования из соседних семей). Вообще, на мой взгляд, одна из основных проблем современной макрокомпаративистики связана с чрезмерным фокусированием исследователя на «общем» между дальнородственными языками — качественная теория родства требует не только обнаружения между ними системных сходств, но и реалистичного учета и объяснения накопленных расхождений.

Откуда берутся «корни-изоляты»? При условии, что речь идет действительно о родственных языках (что для атлантического массива справедливо на «низких» уровнях дерева, но не столь очевидно на «верхних»), здесь могут быть только два объяснения, и, думаю, что оба должны быть в разной степени верны. Во-первых, до тех пор, пока у нас нет четкой системы фонетических соответствий между

атлантическими языками, не содержащей существенных лагун и наглядно иллюстрированной на материале опубликованного корпуса, вряд ли можно поручиться, что все эти «корни-изоляты» действительно несводимы к тем или иным корням праатлантического или, за неимением такового, нескольких «праатлантических» языков (пра-мель, пра-бак и т. п.).

Во-вторых, в отличие от КП, я был бы готов вполне серьезно отнестись к гипотезе о «шести десятках субстратов» на западноафриканском атлантическом побережье. Не очень понятно, откуда берется число 60 — т. е. понятно, что оно соответствует общему числу атлантических языков, в каждом из которых обнаруживаются «корни-изоляты», но почему мы обязаны считать, что каждый язык обязательно заимствовал из уникального субстрата? Размышляя логически, для этого нам потребовалось бы выполнение следующих условий: (а) некоторые значения (будь то «сводешевские» или любые другие) в *каждом* атлантическом языке выражаются формами, несводимыми друг к другу; (б) ни для одной из этих форм мы вообще не можем предложить никакой внутренней этимологии и вынуждены считать их заимствованиями. Однако речь, насколько я понимаю, идет о другом — в каждом атлантическом языке то и дело встречаются «загадочные» слова с неизвестной этимологией. Это совершенно не эксклюзивно-уникальная особенность атлантических языков: она характерна для подавляющего большинства языков Африки, да, пожалуй, и всего мира.

Единственный тип ситуации, при котором «субстратная» гипотеза была бы априорно маловероятна — это ситуация, при которой территория Западной Африки заселяется носителями атлантических языков «с нуля»; но вроде бы история атлантических народов и археологические данные твердо говорят об обратном, так что в факте смешения атлантических племен с «доатлантическим» населением региона вряд ли приходится сомневаться. Насколько лингвистически разнообразен был этот регион до того, как в нем обосновались разные группы «атлантов», можно только догадываться, в том числе, кстати, привлекая как раз «субстратные» данные, особенно таких богатых на «странные» слова языки, как лимба и т. п.

Можно ли использовать «корни-изоляты» для определения рангов стабильности элементов списка Сводеша? Думаю, что нет или, по крайней мере, такую процедуру следует проводить обособленно от процедуры определения рангов стабильности, описанной в статье С. А. Старостина. Дело в

том, что если в статье Старостина стабильность замеряется в первую очередь на основании *внутренних* замен, то «корни-изоляты» (если только это действительно корни-изоляты, а не нормальные унаследованные основы, для которых не удалось пока найти достоверную этимологию) — это в первую очередь замены *внешние*, т. е. обусловленные специфическими социо-культурными условиями. Ранжирование сводешевских элементов на этой основе, скорее всего, будет как минимум сильно отличаться от ранжирования по методу Старостина, а как максимум, составление для него некоторого общеуниверсального индекса (который учитывал бы и атлантические субстраты, и сонгайско-берберские контакты, и судьбу дравидийского языка брахуи, и китаизмы в тайских языках, и латинизмы в албанском, и т. д. и т. п.) вообще не будет иметь никакого смысла — впрочем, попробовать, конечно, можно.

Резюмируя вышесказанное, хочется в любом случае выразить благодарность автору статьи за поднятые им вопросы и высказанные предложения — хотя бы уже за то, что в статье кратко и наглядно суммируются все основные проблемы, которые ожидают исследователя при попытке применения метода «предварительного лексикостатистического обследования» к материалу столь трудного языкового массива, как атлантический (хотя, на мой взгляд, «трудность» его вполне сопоставима, а во многих отношениях существенно уступает «трудности» анализа, скажем, койсанского языкового массива, к которому этот метод был применен вполне успешно). Будущее покажет, насколько успешно эти проблемы получится преодолеть (в том числе и в рамках продолжения моего собственного проекта по детальному лексикостатистическому обследованию языков Африки).

В определенной степени согласен я и с основным выводом автора («получается, что все равно нельзя»), *при условии*, что мы точно понимаем, чего конкретно нельзя — а именно, формулировать гипотезы языкового родства на глубине «*before Indo-European*» через поверхностное сопоставление базисной лексики современных языков, следующее размытой и предельно обобщенной системе правил и критериев. Там же, где поверхностное сопоставление (а) осуществляется на основе четко сформулированной и типологически обоснованной системы правил, (б) сочетается с дополнительными проверочными фильтрами, такими как ясно оговоренная процедура «ручной коррекции» и тест на динамическую градацию результатов, (в) учитывает принцип ступенчатости, т. е. требует

хотя бы прикидочную реконструкцию *прасписков* для сравнения на глубоких хронологических уровнях, оно, на наш взгляд, представляет собой вполне законный и, может быть, даже оптимальный способ формулирования серьезных гипотез языкового родства на разных уровнях *без* сопутствующего требования полной реконструкции этимологического корпуса (которая, кстати говоря, для многих языковых семей по ряду технических причин вообще не может быть осуществлена).

Что касается альтернативных критериев, защищаемых в последнем абзаце текста КП (морфологические изоглоссы, фонологическая статистика и т. п.), то, разумеется, ни «предварительная», ни «финализирующая» лексикостатистика не отменяют их значимости. Дело здесь скорее в следующем: для любой науки важнейшая задача — правильная расстановка приоритетов, а для лингвиста, занимающегося проблемами языкового родства, важнейший приоритет — выявление в языке максимально устойчивого «ядра». На роль такого «ядра» в качестве *универсального* стандарта, как неоднократно аргументировал в своих работах С. А.

Старостин (и с этим я полностью согласен), в наибольшей степени претендует именно базисная лексика, а если это так, то, разумеется, все остальные критерии по определению могут быть обозначены как «сопутствующие». Другое дело, что бывают ситуации, когда (а) базисная лексика сравниваемых языков известна нам в недостаточной степени — например, из-за неаккуратности описания, или из-за того, что описание фокусируется исключительно на грамматике, или, как это часто бывает с древними письменными языками, известные нам тексты представлены лексикой в основном культурного характера; (б) сравнение базисной лексики осуществляется некорректным образом — это уже вина исследователя, а не искривление закономерного процесса языковой эволюции. И с той, и с другой ситуацией надлежит бороться — в первом случае стремиться к полноте и аккуратности описания, а также к полноценному учитыванию всех доступных источников, во втором — к повышению качества методологии, в чем мы все, разумеется, заинтересованы и чему, хотелось бы надеяться, данная дискуссия тоже будет способствовать.

К. И. Поздняков
INALCO, IUF, LLACAN-CNRS (Paris)

*От дальнего родства позиций — к ближнему
(ответ на отзывы Г. С. Старостина и Д. Р. Лещинера)*

Прежде всего хочу поблагодарить журнал за предложение опубликовать статью в рубрике «Дискуссии». Это идеальный формат для любого автора: он позволяет получить квалифицированные отзывы на публикацию и к тому же ответить на критику. В определенном смысле проще отвечать, когда в отзывах выражается прямое несогласие с авторской позицией. Сложнее реагировать, когда рецензенты обсуждают мотивы написания статьи, цели, которые преследовал автор, круг ее потенциальных читателей и другие вопросы, имеющие лишь косвенное отношение к развиваемым в статье положениям, и при этом выражают согласие с ее основными тезисами. Если оставить в стороне дипломатические обороты, суть отзыва Г. С. Старостина (далее — Г.С.) я понял так: в статье содержатся отдельные интересные наблюдения, но не вполне

ясно, зачем и для кого их нужно публиковать. Г.С. прямо характеризует статью «как набор любопытных и во многом полезных размышлений, формально связанных друг с другом, но не имеющих единой общей цели (по крайней мере, очевидной)». Прежде всего, я охотно принимаю первую часть этой формулировки. Что касается «целей», то они вполне точно сформулированы самим Г.С. в конце его отзыва: «в статье кратко и наглядно суммируются все основные проблемы, которые ожидают исследователя при попытке применения метода „предварительного лексикостатистического обследования“ к материалу столь трудного языкового массива, как атлантический». Надеюсь, что это именно так — для этого статья и написана.

Тогда в чем же основная проблема, которая перевела статью в русло полемики? В отсутствии по-

тенциального читателя, которому эти проблемы могли бы быть интересны? В отзыве Г.С. высказывается мнение, что вопрос, поставленный в первой части статьи¹, не имеет адресата, поскольку он не интересен ни «сплиттерам» (а), «ни ламперам» (б) и при этом статья не «обращена на критику» общих теоретических положений, сформулированных в монографии Г. С. Старостина (с), ибо подменяет важные исходные посылки другими. Охотно принимаю и эти три пункта, хотя не скрою, что выстроенная триада меня несколько озадачила и заставила задуматься, поскольку мне действительно нечего сказать ни сплиттерам, ни ламперам, ни теоретикам глобальной макрокомпаративистики.

Последний пункт для меня по ряду причин особенно важен, и мне хотелось бы, чтобы именно в нем авторская позиция была сформулирована предельно четко. Мне будет искренне жаль, если у читателя сложится впечатление, что в статье делается попытка ревизии основных принципов использования лексикостатистики для установления дальнего родства, разработанных московской школой компаративистики. Эти принципы системно обсуждаются в монографии Г. С. Старостина «Языки Африки», которая на сегодняшний день является, по моему убеждению, лучшей обобщающей публикацией на эту тему. Еще раз подчеркну то, что прямо написано в статье: я не только высоко ценю эту монографию, но и полностью разделяю ее основные положения. В целом я, как и многие компаративисты, с интересом слежу за глобальным проектом «Evolution of Human Languages» (одним из руководителей которого является Г. С. Старостин), нацеленным на создание единой генеалогической классификации языков мира. Отсюда ясно, что предложенная мной статья не полемизирует с изложенными в монографии Г. С. Старостина принципами — она на них опирается. Что же касается сплиттеров и ламперов, то позволю себе напомнить, что параллельно с этими достойными людьми существуют конкретные специалисты в конкретных областях сравнительно-исторического языкознания. Занимаясь малоизученными семьями языков, они вправе задать себе вопрос, который, строго говоря, не имеет отношения к классическому сравнительно-историческому методу²:

можно ли построить генеалогическую классификацию этих языков, и если можно, то как это сделать?

К таким специалистам относит себя и автор, размышления которого адресованы специалистам в смежных областях. Почему бы в таком случае не взять «атлантический фрагмент» мировой лексикостатистической классификации и попросту не опубликовать его с минимальными комментариями? Ответ на этот вопрос и может, вероятно, объяснить суть «полемики».

Года четыре назад с моим коллегой Гийомом Сежерером мы решили пересмотреть генеалогическую классификацию атлантических языков, которыми давно занимаемся, и даже опрометчиво объявили о готовящейся публикации на эту тему. Казалось бы, в чем здесь проблема? Мы заполнили матрицу по стословнику Сводеша и, пользуясь личными связями, обратились к Г. С. Старостину с просьбой обработать наши данные. Через несколько дней (ноябрь 2010) мы получили красивое дерево³. Казалось бы, можно было опубликовать его и считать вопрос закрытым. Но дело в том, что, изучив это дерево, мы увидели, что публиковать его нам совсем не хочется.

Какие результаты мы получили? Праатлантический язык датировался приблизительно 10 000 до н. э. Для тех, кто следит за развитием атлантической компаративистики, это может показаться неудивительным. Накапливается все больше аргументов для того, чтобы считать, что традиционно объединяемые в одну группу языки мель и собственно атлантические языки представляют собой две самостоятельные ветви нигер-конго, а это значит, что указанная дата может относиться к времени распада нигеро-конголезского языка. Но далее оказывается, что ряд языков, традиционно относимых к североатлантическим языкам (волоф, нун, буй), отстоят дальше от остальных северных языков (фула, сереер, тенда и др.), чем, например, языки бак (диола и проч.)! Эта «революция» в атлантическом языкознании прослеживается в период 5000—6000 до н. э., причем очевидно, что по-другому интерпретированные три-четыре слова могут радикально изменить классификацию в ее наиболее важных фрагментах⁴. Понятно, что такое дерево публиковать не только бессмысленно, но и

¹ Г.С. формулирует его так: «Можно ли классифицировать языки на уровне макросемей, не используя ступенчатую реконструкцию на базе сравнительно-исторического метода?»

² О соотношении генеалогической классификации и собственно сравнительно-исторического метода интересно писал В. Я. Порхомовский (Порхомовский 1982).

³ Еще раз хочу поблагодарить Г. С. Старостина и его коллег за оперативную помощь.

⁴ Близкие языки легко объединяются в группы без использования лексикостатистики. Можно сказать иначе: они объединяются в результате применения любой лексикостатистической методики, кроме заведомо абсурдных.

вредно, учитывая, что классификации языков имеют свойство широко тиражироваться и в итоге возвращаться к авторам в виде незыблемых истин.

Снова с помощью Г. С. Старостина мы сделали вторую попытку. Отказавшись от формальных критериев установления сходств, мы заново заполнили матрицу, используя наши (пока еще очень фрагментарные) знания в области атлантической этимологии, и получили новое дерево (март 2011). Естественно, наш праязык значительно помолодел. Разделение собственно атлантических языков и языков мель датируется в новой «классификации» как 6500 до н. э., а разделение собственно атлантических языков (без изолированных языков лимба и гола) на две большие группы — как 5600 до н. э. Исчезли явно нелепые группировки языков, но процент сходств, определяющий отнесение некоторых языков (или подгрупп) к одной из двух групп остался настолько мал, что та или иная интерпретация по-прежнему не может считаться надежной. Так, язык налу объединился с северными языками (при этом индексы его сходств колеблются от 0,10 с волоф до 0,21 с джаад) и, соответственно, отделился от центральных языков, индексы сходств с которыми колеблются у налу от 0,08 с диола керак до 0,18 с балант и манкань. Индексы не только мало различаются в двух группах, они и очень малы абсолютно — настолько, что приближаются к интуитивному порогу случайности. Так, языки пепель и суа показывают 3% сходств, а внутри собственно атлантических биджого-буй дают 8%, как и рассмотренная выше пара налу — диола керак.

Отсюда, в ходе работы над конкретным атлантическим материалом, и возник вопрос — а собственно, сколько случайных сходств между двумя неродственными языками мы должны получить? Оказалось, что этот вопрос изучен явно недостаточно, причем немногочисленные публикации на эту тему ограничиваются несколькими конкретными критериями формального определения сходства и не затрагивают другие алгоритмы (формализованные или интуитивные), которые применяются компаративистами на практике. Так, автору пришлось самому анализировать критерии сходств у Гринберга — их никто, насколько мне известно, не публиковал, хотя ссылок на Гринберга имеются сотни, если не тысячи. Мне показалось интересным поделиться результатом моего анализа с компаративистами, работающими с другими языками. Так, например, Гринберг сближает только **одинаковые** велярные и палатальные и практически никогда не сближает **разные** согласные

внутри этих двух рядов (например, [ç] и [j]). Это значительно более строгий алгоритм, чем допуск различия по одному признаку и тем более допуск сближения в рамках консонантного класса. И тем не менее случайные сходства по критериям Гринберга должны составить не менее 16%, если логика моих подсчетов верна (а ее вроде бы мои оппоненты не оспаривают). Предлагая для публикации результаты такого рода, я надеялся (и продолжаю эту надежду питать), что они могут быть интереснее коллегам, чем очередные общие рассуждения «за» или «против» массового сравнения.

При тестировании количества ожидаемых случайных сходств я использовал очень разные критерии — от самых строгих до наиболее свободных. В отзыве Г.С. выводы первой части статьи ставятся под сомнение в связи с тем, что автор «использует чрезвычайно либеральный алгоритм сопоставления». Но единственная «позитивная» гипотеза статьи, касающаяся минимального числа сходств, ориентирована на самый строгий из всех возможных критериев! В статье на основе исследования вероятностей совпадений высказывается предположение о том, что 5% **полных** сходств убедительно свидетельствует о родстве языков. Если мы будем считать только примеры полной идентичности согласных в двухконсонантных корнях списков французского и русского языков, мы получим 7 сходств, а именно: *tourir* — *умирать* 13⁵, *lune* — *луна* 18, *poiveau* — *новый* 23, *plein* — *полный* 34, *soleil* — *солнце* 39, *blanc* — *белый* 79, *écorce* — *кора* 102. Вероятность случайности этого события настолько низка, что языки можно считать родственными. Конкретно, при 20 согласных в каждом языке и при самых благоприятных условиях для их случайных совпадений (одинаковый инвентарь согласных в двух языках и равные частоты согласных во всех позициях⁶) вероятность полного совпадения одного слова стословного списка (по двум согласным) составляет всего четверть процента ($100 * 0,05 * 0,05$). Подчеркну, что обратная ситуация (отсутствие совпадений) не говорит об отсутствии родства. Так, сопоставление английского и русского списков дает только один пример, удовлетворяющий заданным жестким условиям (если отвлечься от ситуации с гласными): *many* — *много* 106.

⁵ Цифрами обозначены ранги стабильности приведенных значений по С. А. Старостину. Тестирование проводится без учета фактора заимствований.

⁶ Важный в практическом отношении вывод, обоснованный в статье: чем больше варьируют частоты согласных, тем больше вероятность их совпадений / сходств в разных языках.

Вернемся к атлантическим данным. Сходства в стословнике для языка лимба, который традиционно объединяют с атлантическими, дают следующие проценты:

- лимба — североатлантические языки — от 7% до 13%.
- лимба — центральноатлантические языки — от 4% до 14%.
- лимба — мель языки — от 6% до 12%.

Убежден, что, как для меня, так и для моих оппонентов бесспорно следующее: за фрагментом дерева, на котором лимба оказывается на 250 лет ближе к северно- и центральноатлантическим языкам (6250 до н. э.), чем к мель (6500 до н. э.), ровно ничего не стоит. Такая конфигурация определяется одним-единственным сходством. Вместе с тем, очевидно, мы не можем утверждать, что язык лимба неродственен атлантическим языкам и языкам мель, хотя мой «слишком вольный алгоритм» предполагает, что чисто случайные сходства должны давать не меньший процент. То есть про связи лимба мы не можем сказать ничего. Посмотреть, сколько слов из семи имеет высокий ранг семантической стабильности, а сколько — низкий, конечно, можно, но и на этом классификацию не построишь.

Можно объявить язык лимба отдельной ветвью нигер-конго, т. е. приписать языку лимба тот же статус, что и, например, семье бенуэ-конго. Такое решение легко обосновать, и формально оно будет выглядеть безукоризненным. Но все ли возможности мы испробовали, чтобы объединить лимба с атлантическими языками, что соответствовало бы традиционному взгляду на положение лимба? На чем основан (кроме географии) этот интуитивный традиционный взгляд, который разделяется ведущими специалистами? Не стоит ли в этом случае опереться на данные исторической морфологии и, прежде всего, на системы именных классов, которые по ряду важных параметров оказываются более близкими у лимба и у атлантических языков и не похожи в этом плане на системы языков банту? Не имеет ли смысл, в ситуации, когда индексы сходств слишком малы, изучить в первую очередь те 47 слов лимба из стословника Сводеша, которые не находят видимых параллелей ни в одном из десятков атлантических языков? Это далеко не полный круг вопросов, которые должны, на мой взгляд, интересовать специалистов, занимающихся генеалогической классификацией языков Африки, и которые мне представлялось важным обсудить в статье. Если «максимально устойчивое „ядро“» (Г.С.) в языках уже **известно и является универсальным**, то почему же «для лингвиста, занимающего-

ся проблемами языкового родства, важнейший **приоритет — выявление** в языке максимально устойчивого „ядра“» (Г.С.)? Зачем так важно выявлять заново то, что известно заранее? Применительно к атлантическим языкам могу сказать ответственно: если выбирать из двух предложенных формулировок Г.С., следует выбрать вторую, а именно выявить наиболее устойчивый пласт форм, применяя весь комплекс возможных подходов, уже апробированных или совсем новых.

Отмечу, что, на наш взгляд, наиболее показательным для атлантической классификации является традиционный поиск инноваций в лексике и особенно в морфологии и максимальная тщательная верификация этих инноваций с привлечением данных внешнего сравнения. Неожиданно интересные результаты дает и анализ «изолированной» лексики — подход, который рассматривается в третьей части статьи. Я рад, что Г.С. «полностью согласен с идеей автора о том, что такие корни следует как минимум инвентаризировать и публиковать». Более того, резюме Г.С. как нельзя лучше суммирует то, что я пытаюсь донести до читателя — с большим удовольствием приведу формулировку из отзыва Г.С.: «Вообще, на мой взгляд, одна из основных проблем современной макрокомпаративистики связана с чрезмерным фокусированием исследователя на „общем“ между дальнородственными языками — качественная теория родства требует не только обнаружения между ними системных сходств, но и реалистичного учета и объяснения накопленных расхождений». По-моему, лучше не скажешь. Замечу, что далеко не все компаративисты готовы принять эти положения.

Допуская принципиальную ценность изучения изолированных корней, Г.С. выражает сомнение в перспективности их использования для определения рангов стабильности. В этой связи хочу воспользоваться этим пунктом, чтобы высказать некоторые общие соображения о применении статистики в сравнительно-историческом языкознании (помимо традиционной лексикостатистики). Давние поиски в этой области⁷ и многочисленные обсуждения разнообразных методик с коллегами, а также со специалистами в области статистики сформировали у автора твердое убеждение, что «идеальных» формул в этой области не существует.

⁷ Взгляды автора на этот вопрос впервые системно изложены в разделе «Статистика и компаративистика» книги [Поздняков 1993]. Препринт последней общей публикации на эту тему — «Statistique et comparatisme» — доступен на <http://pozdnikov.free.fr>.

Прежде чем раскрыть это «крамольное» положение, позволю себе привести легенду, которую мне довелось слышать в аспирантские годы в Институте языкознания АН СССР (еще на Волхонке). На одной из конференций к трибуне вышел очередной докладчик. Он выдержал эффектную паузу, а затем резко выхватил длинную линейку и, потрясая ей, сообщил притихшей публике: «С этого инструмента должно начинаться любое лингвистическое исследование!» Занимался докладчик полисемией. Он ее мерил... в сантиметрах, сравнивая длину словарных статей. Логика была неувязима: в статье к слову *телефон* — мало сантиметров, а к слову *резать* — много; следовательно, слово *резать* — полисемичнее. Зададим себе вопрос: работает эта статистическая методика или нет? В принципе работает. Если у скептиков найдутся контрпримеры, методика можно будет скорректировать. Возможно, кто-то предложит новую методику, которую проще применить сегодня: можно мерить не в сантиметрах, а в количестве слов словарной статьи. Более того, если поставить себе глобальную задачу — выделить 100 наиболее полисемичных слов в языках мира — возможно, это будет единственная статистическая методика, у которой будет шанс дать любопытные результаты при минимуме временных затрат.

Подчеркну, что, приводя этот лингвистический анекдот, автор не предлагает читателю мерить ранг стабильности слов «в попугаях». Просто одна из ключевых проблем в подборе статистической формулы к сложнейшим языковым массивам, на формирование которых оказывает влияние множество разнообразных факторов, состоит в подборе такой формулы, которая, минимально искажая результат, была бы максимально проста и максимально прозрачна. Под прозрачностью здесь понимается, прежде всего, «физический смысл» формулы, который понятен лингвисту. Так, процентное отклонение от среднего физически понятнее, чем квадратичное отклонение, и при этом для многих массивов применение критерия χ^2 при всех его очевидных преимуществах существенно не меняет получаемые результаты.

В формулу С. А. Старостина, рассмотренную в статье, заложена попытка нейтрализовать очевидную «зависимость коэффициента стабильности от глубины семьи и числа сравниваемых языков». Вместе с тем применение предложенной процедуры подсчета индексов разрушает прозрачность формулы в указанном выше смысле и к тому же предельно затрудняет количественное сопоставление индексов (они задуманы как относительные,

а не абсолютные). Последнее для меня было особенно важно, потому меня интересовало именно сопоставление индексов по каждому отдельному значению. Достаточно неожиданным для меня оказалось то, что применение более простой и физически понятной формулы дает не менее точный результат, а это означает, что влиянием указанного фактора, на который вводился коэффициент, вполне можно пренебречь. Более того, выяснилось, что по ряду важных параметров «примитивная» формула работает лучше: полученные с ее помощью результаты позволяют «узаконить» именно те изменения в первой половине списка Сводеша, которые московская школа и предлагает ввести, обосновывая эти изменения «техническими (?) причинами» (Г.С.). И наконец, оказалось, что очень неплохие результаты дает использование другого простейшего критерия — отношение количества корней в рамках одного значения к количеству языков. Достоинство этого критерия в том, что он учитывает и корни-изоляты, и, таким образом, выявляется, например, особый статус языка лимба при его сопоставлении с атлантическими языками. В связи с этим мне непонятен комментарий Г.С.: «Ранжирование сводешевских элементов на этой основе, скорее всего, будет как минимум сильно отличаться от ранжирования по методу Старостина [...] впрочем, попробовать, конечно, можно». Вообще-то говоря, я попробовал. Результаты такого ранжирования — в тексте статьи, что позволяет рассчитывать на более предметный уровень критики.

В целом приходится признать, что ответственность за целый ряд указанных недоразумений несет автор — вероятно, следовало более четко обозначить задачи статьи в начале и суммировать выводы в заключении. Однако дело не только в этом: суть полемики, как мне показалось, не в столкновении двух разных точек зрения, а в существовании разных углов зрения на одну проблему, — в разных **ракурсах** рассмотрения проблем дальнего родства. Рискну предположить, что у меня будет еще меньше расхождений с Г.С. — автором генеалогической классификации койсанских языков, чем с Г.С., защищающим глобальный проект классификации языков мира.

В этом плане мне был чрезвычайно интересен и третий угол зрения на количественные критерии установления родства и рангов семантической стабильности, представленный в отзыве авторитетного специалиста в области статистики Д. Р. Лещинера. Прежде всего, для меня важно, что предлагаемые в статье количественные критерии в основ-

ном успешно прошли профессиональную экспертизу. Помимо этого, считаю нужным отметить, что отзыв Д. Р. Лещинера, изложенный в виде «Заметок», имеет, на мой взгляд, самостоятельное научное значение безотносительно к обсуждаемой статье. Я имею в виду последнюю часть «Заметок», посвященную возможностям исчисления «апостериорной вероятности родства слов» и подходам к

«ступенчатой» лексикостатистике. Эти новые подходы представляются мне перспективными, тем более, что они заставляют задуматься над рядом других, практически не исследованных вопросов, и, в частности, над вопросом о влиянии конфигурации генеалогического древа на возможности реконструкции. Впрочем, это уже предмет отдельной дискуссии.

Литература

- Поздняков 1993 — К. И. ПОЗДНЯКОВ. *Сравнительная грамматика атлантических языков*. М.: Наука, 1993. [K. I. POZDNYAKOV. *Sravnitel'naya grammatika atlanticheskikh yazykov*. М.: Nauka, 1993.]
- Порхомовский 1982 — В. Я. ПОРХОМОВСКИЙ. Проблемы генетической классификации языков Африки // *Теоретические основы классификации языков мира. Проблемы родства*. М.: Наука, 1982. [V. Ya. PORKHOMOVSKIY. *Problemy geneticheskoy klassifikatsii yazykov Afriki // Teoreticheskie osnovy klassifikatsii yazykov mira. Problemy rodstva*. М.: Nauka, 1982.]
- Г. Старостин 2013 — Г. С. СТАРОСТИН. *Языки Африки. Опыт построения лексикостатистической классификации*. Т. 1: *Методология. Койсанские языки*. М.: ЯСК, 2013. [G. S. STAROSTIN. *Yazyki Afriki. Opyt postroeniya leksikostatisticheskoy klassifikatsii*. Т. 1: *Metodologiya. Kojzanskie yazyki*. М.: YaSK, 2013.]
- С. Старостин 2007 — С. А. СТАРОСТИН. Определение устойчивости базисной лексики // С. А. СТАРОСТИН. *Труды по языкознанию*. М.: ЯСК, 2007. С. 827—839. [S. A. STAROSTIN. *Opreделение ustojchivosti bazisnoj leksiki // S. A. STAROSTIN. Trudy po yazykoznaniiyu*. М.: YaSK, 2007. S. 827—839.]
- Campbell & Poser 2008 — L. CAMPBELL, W. J. POSER. *Language Classification: History and Method*. Cambridge University Press, 2008.
- Greenberg 1963 — J. GREENBERG. *The Languages of Africa*. Bloomington: Indiana University Press, 1963.
- Pozdniakov & Segerer 2007 — K. POZDNYAKOV, G. SEGERER. Similar Place Avoidance: A Statistical Universal // *Linguistic Typology* 11, 2. 2007. P. 307—348
- Sapir 1971 — J. D. SAPIR. West Atlantic: an inventory of the languages, their noun class systems and consonant alternation // T. A. SEBEOK (ed.). *Current trends in linguistics 7: Linguistics in Sub-Saharan Africa*. The Hague & Paris: Mouton & Co, 1971. P. 45—112.

Konstantin POZDNYAKOV. On the threshold of relationship and the “stability index” of basic lexicon in mass comparison: Atlantic languages

The paper deals with certain key aspects of the procedure of establishing long-range genetic relationship between languages in the absence of a classic step-by-step reconstruction, such as: 1) the issue of establishing the randomness threshold for lexical matches on the Swadesh 100-item wordlist, depending on the particularities of the employed algorithm; 2) the issue of stability ranking in the sphere of the basic lexicon; 3) the issue of multiple isolated roots that do not find any parallels in the other languages of the family. These general topics are discussed in the context of specific problems that arise during research on the genetic classification of Atlantic languages. The latter are traditionally considered to constitute a branch of the Niger-Congo macrofamily, but lexicostatistics suggests that their date of separation is no younger than 8000 BP. Two critical responses to the paper present further opinions on the difficult problematics of establishing long-range relationship; the author’s final response is also to be regarded as an important constituent of the published paper.

Keywords: protolanguage reconstruction, basic lexicon, lexicostatistics, etymology, Atlantic languages, macro-comparative studies.