

СТАТЬИ И МАТЕРИАЛЫ

ЛЕКСИКОСТАТИСТИЧЕСКАЯ КЛАССИФИКАЦИЯ АФРАЗИЙСКИХ ЯЗЫКОВ: НОВЫЕ ПОДХОДЫ

А.Г. Козинцев, А.Ю. Милитарев

Александр Григорьевич Козинцев | <https://orcid.org/0000-0002-0165-8109> | alexanderkozintsev@yandex.ru | д. и. н., главный научный сотрудник | Музей антропологии и этнографии им. Петра Великого (Кунсткамера) РАН (Университетская наб. 3, Санкт-Петербург, 199034, Россия)

Александр Юрьевич Милитарев | <http://orcid.org/0000-0002-5684-3053> | amilitarev@gmail.com | д. филол. н., профессор-консультант | Российский государственный гуманитарный университет (Миусская пл. 6, Москва, ГСП-3, 125267, Россия)

Ключевые слова

лексикостатистика, афразийские языки, семитские языки, египетский язык, берберские языки, чадские языки, кушитские языки, омотские языки

Аннотация

Лексикостатистическая классификация 205 афразийских (АА) языков осуществлена по базе данных А.Ю. Милитарева различными методами кластеризации и снижения размерности на основании 100-словных списков. Построены укорененные деревья и сети, применена псевдопространственная модель. Предложенное А.Ю. Милитаревым разделение АА языков на северные (семитские, египетский, берберские, чадские) и южные (кушитские и омотские) подтверждается, хотя монофилия кушитской семьи под вопросом. В пределах семитской семьи арамейские языки с предположительно родственным им сабейским занимают центральное положение, будучи связанными со всеми остальными. Является ли левантийская группировка, включающая, наряду с арамейскими и сабейским языками, ханаанейские и угаритский языки, монофилетическим таксоном или языковым союзом – неясно. Эфиосемитскую и современную южноаравийскую ветви нельзя считать сестринскими. В пределах эфиосемитской ветви северные языки промежуточны между южными и остальными семитскими языками, что, видимо, отражает историю распространения семитской речи от берега Красного моря вглубь континента.

Цель данной работы – применить аналитические методы, не нашедшие еще широкого применения в лексикостатистике, к базе данных по афразийским (АА) языкам, созданной А.Ю. Милитаревым (далее – А.М.) на основании 100-словных списков, представляющих все или репрезентативное большинство ветвей/семей и групп/подсемей АА макросемьи (*Militarev* 2000, 2004, 2005, 2010, 2011, 2012, 2014, 2015)¹. В случаях, когда слова отождествлены на базе более широкой (включающей не только лексику 100-словных списков) этимологии,

Статья поступила 04.04.2022 | Окончательный вариант принят к публикации 23.07.2022
Ссылки для цитирования на кириллице / латинице (*Chicago Manual of Style, Author-Date*):

Козинцев А.Г., Милитарев А.Ю. Лексикостатистическая классификация афразийских языков: новые подходы // *Этнографическое обозрение*. 2022. № 6. С. 132–159. <https://doi.org/10.31857/S0869541522060082> EDN: MRORLQ

Kozintsev, A.G., and A.Y. Militarev. 2022. Leksikostatisticheskaja klassifikatsiia afraziiskikhazykov: novye podkhody [A Lexicostatistical Classification of Afrasian Languages: New Approaches]. *Etnograficheskoe obozrenie* 6: 132–159. <https://doi.org/10.31857/S0869541522060082> EDN: MRORLQ

дается отсылка к этимологической афразийской базе данных, разработанной А.М. Милитаревым совместно с О.В. Столбовой (*Militarev, Stolbova 2007*). Сходная попытка уже была предпринята по отношению к базе данных Г.С. Старостина и его коллег, основанной на 50-словных списках (GLD-1 n.d.; *Козинцев 2021*).

Один из ключевых вопросов АА классификации, от которого зависит, в частности, решение проблемы АА прародины, состоит в том, являются ли омотская и кушитская семьи сестринскими. А.М. дает на него утвердительный ответ (*Militarev 2005: 398*). Он не только относит омотские языки к АА макросемье, что некоторые лингвисты отрицают (*Theil 2012*), но и считает возможным говорить о кушомотском языковом единстве, возникшем еще в Передней Азии. Эту позицию в целом разделяет В. Блажек (*Blažek 2013a, 2013b*). Из данной точки зрения вытекает следующее: своеобразие омотской семьи было преувеличено предшествующими авторами (в частности Г.К. Флемингом); кроме того, оно усугубляется субстратными влияниями, в основном со стороны нилотских языков, а также заимствованиями из равнинно-восточнокушитских (в первую очередь оромо), горно-восточнокушитских (скорее всего, сидама и дараса) и, возможно, центрально-кушитских языков – причем, особенно в последнем случае, определить как сам факт (т.е. отличить заимствование от общей унаследованной лексики), так и направление заимствования очень трудно, и разные авторы часто занимают противоположные позиции и в отношении каждого отдельного случая, и в оценке общей тенденции.

Сторонники же гипотезы африканской прародины АА языков, причисляющие омотские языки к АА, обычно рассматривают омотскую ветвь как наиболее раннюю в АА макросемье (*Fleming 2006: 140; Blench 2006: 148; Ehret 2011: 142*; см. также неопубликованные данные Г.С. Старостина, цит. по: *Blažek 2013b: 126*).

Другим ключевым вопросом является статус почти вымершего языка онгота (юго-запад Эфиопии), который А.М. относит к омотским (см., напр.: *Militarev 2006: 510–511*). Между тем Г.К. Флеминг считал его отдельной ветвью АА макросемьи (*Fleming 2006: 135–139*), а В. Блажек сближает его с нило-сахарской макросемьей (*Blažek 2007*). Г.С. Старостин называет онгота языком-изолятом (*Старостин 2013: 226–227*), а Г. Сава и М. Тоско пишут о нем как о креольском языке, имеющем нилотские, омотские и кушитские черты (*Savà, Tosco 2000*).

В первой части статьи использована общеафразийская база данных (семитская семья представлена в ней 27 языками), во второй – только семитская, представленная 56 языками.

Общеафразийская база

Использованы данные о 176 языках следующих ветвей АА макросемьи (для удобства использования сокращения расположены в алфавитном порядке в пределах семей):

I. Кушитские: AFA – афар, ALA – алагава, ARE – арборе, ASA – аса, AWN – аунги, BAY – байсо, BIL – билин, BON – бони, BRJ – бурджи, BRN, BRP, BJV – беджа (по словарям Л. Райниша, Э. Роупера и М. Ванхоуве соответственно), BUR – бурунге, DAN – дахало, DAS – дасенеч, DIR – дирайта, DRS – дараса, ELM – эльмоло, FAL – фаллаша, GBZ – гобезе, GLG – голанго, GWD – гаввада, HDY – хадийя, HRS – харсо, IRQ – иракв, JID – джидда, KAI – кайлинья, KEM – кемант, KGA – хамтанга, KHA – хамта, KMB – камбатта, KMR – хамир, KNF – кунфал, KON – консо, MAA – маа, MSH – машиле, ORM – оромо, QWA – квадза, QWR – квара, RND – рендилле, SAN – сахо, SID – сидамо, SML – сомали, TSA – цамай, WAG – вааг, YAK – яaku (могогодо).

II. Омотские: ANF – анфилло, ARI – ари, BAN – банна, BSK – баскето, BWO – бворо, CHA – чара, DCH – даче, DIM – диме, DOK – докка, DRZ – дорзе, DZM –

дизи-маджи, DZN – дизи-найи, DZS – дизи-шеко, GFA – гофа, GLL – галила, GMB – гимира-бенчо, GMS – гимира-ше, GMU – гему, GNJ – ганджуле, GNZ – ганза, HMR – хамер, KCH – качама, KFA – кафа, KLO – кулло, KYR – койра, MAO – мао-бамбеши, MCH – моча, MHZ – мао-хозо, MLE – мале, MLO – мало, MSZ – мао-сезо, OYD – ойда, UBA – убамер, WEL – велайта, YEM – йемса, ZAL – зала, ZRG – зергула, ZYS – зайсе.

III. *Онгота* – ONG.

IV. Семитские: АКК – аккадский, АМН – амхарский, АНС – древнеарамейский, ВАВ – вавилонский арамейский, ВИБ – библейский арамейский, ЕАР – египетский арамейский, GEZ – геэз, НАР – харари, НБР – иврит, НОВ – хобьот, HSS – харсуси, ЛИБ – джиббали, МЕС – мекканский арабский, МНР – мехри, MLT – мальтийский арабский, MND – мандейский, OFF – официальный (имперский) арамейский, PAL – палестинский арамейский, PHO – финикийский, QUR – коранический арабский, SAB – сабейский, SOD – соддо, SOQ – сокотри, SYR – сирийский арамейский, TGR – тигре, TNA – тигринья, UGR – угаритский².

V. Египетские: ВОН – бохайрский диалект коптского, EGY – египетский, FAY – фаюмский диалект коптского, SHI – сахидский диалект коптского.

VI. Берберские: ADG – адгхак, AUJ – ауджила, AYR – айр, FJH – фоджаха, GHD – гхадамес, GHT – гхат, HGR – ахаггар, IZD – издег, MZB – мзаб, NFS – нефуса, RIF – риф, SHW – шайя, QBL – кабийский, SIW – сиуа, SKN – сокна, SLL – семлаль, TIM – тимимун, TIQ – тайток, ZNG – зенага.

VII. Чадские: BCH – бачама, BDM – будума, BOL – боле, BRA – бура, DBA – даба, FAK – фали-кириа, FYE – фьер, GDE – гуде, GDR – гидар, GLV – главда, HIN – хиги-нкафа, HSA – хауса, JGU – джегу, KIR – киир, KRA – кера, KRY – карийя, KUJ – куджарге (принадлежность его к чадской семье проблематична), LLE – леле, LOG – логоне, MAS – маса, MDR – мандара, MFA – мафа, MFU – мофу, MGM – мигама, MKK – мокилко, MNJ – мунджук, MSG – мусгу, MUB – муби, MWA – мауа, NDM – ндам, NGZ – нгизим, PAA – паа, PDK – подоко, PVE – певе, SUK – сукур (сакур), SUR – сура, TER – тера, TUM – тумак, UBI – уби, ZAR – зар, ZMB – зиме-батна.

Использованы также данные о языке хадза (HDZ), отношение которого к АА макросемье не вполне ясно. По традиционным представлениям, хадза не принадлежит к АА языкам. Согласно Г.С. Старостину (*Старостин* 2008), его носители изначально говорили на койсанском языке, который затем испытал сильное влияние со стороны АА языков. Впрочем, Г.С. Старостин не отвергает возможность родства хадза с АА макросемьей. Согласно же последним данным А.М., лексические сходжения хадза с АА языками вызваны глубинным родством, а не заимствованиями из отдельных АА ветвей. В некоторых случаях, правда, когнаты трудно отличить от заимствований, но совпадения и в базисной, и в общей лексике слишком многочисленны, чтобы их можно было приписать случайности или заимствованиям из самых разных, даже удаленных, афразийских языков. Что касается специфического родства хадза с отдельными АА ветвями, то здесь результаты довольно противоречивы. Можно говорить о большей близости хадза к южным (кушитским и омотским), чем к северным (семитским, египетскому и берберским) языкам. Большое число сходжений отмечено с кушитской семьей, особенно с беджа, что вполне понятно, но почти столько же – с чадской, что понять труднее. Можно предположить глубинное родство хадза с чадской семьей и некоторое число заимствований, в основном неидентифицированных, из кушитских (не считая нескольких очевидных заимствований из иракв) и особенно североомотских языков. Альтернативный вариант – специфическое родство хадза, например, с беджа или омотской (конкретно североомотской) ветвью и некоторое количество неопознанных заимствований из чадских языков. Это ма-

вероятно ввиду удаленности хадского ареала от Танзании; кроме того, археологические и генетические данные указывают на то, что носители хадза жили в данной местности не менее 50 тыс. лет, возможно, и гораздо дольше (*Shriner et al.* 2018). Еще один вариант, к которому сейчас склоняется А.М.: хадза относится к третьей южной ветви АА макросемьи.

Своеобразие данной работы в методологическом и междисциплинарном аспектах состоит в том, что ее авторы представляют не только разные научные специальности и несколько неожиданные для этих специальностей “побочные” сферы интересов, но и придерживаются разных, в чем-то даже полярных теоретических установок.

Один из них, А.М. Милитарев – лингвист-компаративист, интересы которого выходят далеко за пределы лингвистики и в последние годы сосредоточены на возможностях реконструкции фрагментов древней и особенно древнейшей истории с помощью лингвистических методов (этимологии, реконструкции праязыковой лексики), поэтому в заявленной теме А.М. больше фокусируется не на полученной им традиционным способом уточненной лексикостатистической классификации афразийских языков – и, соответственно, ее более “стереоскопическом”, разностороннем представлении новыми методами кластеризации и снижения размерности, – а на гораздо более сомнительном с точки зрения мейнстрима методе глоттохронологии, единственно дающем надежду на соотнесение лингвистических, исторических, археологических и генетических датировок. Являясь представителем Московской школы компаративистики и достаточно ортодоксальным последователем С.А. Старостина в сильно усовершенствованной последним методике лексикостатистики/глоттохронологии, А.М. твердо стоит на позиции: один предок у каждого естественного языка (особые и сравнительно редкие явления креольских языков и пиджинов здесь не обсуждаются), считая, что ареальные влияния – объект выявления и устранения из лексикостатистических подсчетов (именно поэтому их результат тем более точен, чем лучше разработан языковой таксон в сравнительно-историческом плане)³, а трудно- или вовсе не выявляемые влияния (“ареально-генетические” связи, при которых когнаты сохраняются в сестринских диалектах, а затем и языках благодаря соприкосновению их ареалов, – см. ниже) лишь создают неизбежный “шум”, сравнимый с этимологическими проблемами, возникающими из-за таких трудно поддающихся анализу явлений, как контаминация, корневая вариантность или метатеза.

Другой автор, А.Г. Козинцев (далее – А.К.) – физический и культурный антрополог, интересы которого тоже существенно выходят за рамки соответствующих дисциплин и сосредотачиваются в последнее время на новых методах лексикостатистической классификации языков разных (макро)семей, несколько скептически относится к глоттохронологии. А.К., в отличие от А.М., подвергает сомнению постулат об одном предке у каждого языка, тем самым отрицая универсальную применимость генеалогической модели к задачам классификации, и признает “ареально-генетические связи” в духе волновой теории Й. Шмидта.

Мы надеемся, что это несколько экзотическое “трение” разных и даже полярных позиций двух авторов при полной готовности к совместному поиску научной истины – пусть даже такой частной, как добываемая в данной статье – даст ту самую творческую искру, без которой этот поиск не может привести к так наз. прорыву в науке.

Методы, примененные к материалу А.Ю. Милитарева А.Г. Козинцевым – последнему принадлежит, соответственно, и большинство последующих комментариев, – относятся к категории фенетических (дистантных, т.е. основанных на показателях обобщенной близости – доле слов, признанных совпадающими), а не кладистических (основанных на анализе каждого слова по отдельности).

Многие считают это недостатком, но в некоторых отношениях дистантные методы имеют преимущество. В частности, для моделей, которые здесь используются, применимы лишь они.

Первая модель – генеалогическая – использовалась в двух разновидностях. Одна из них – простое родословное древо, построенное методом BioNJ (Gascuel 1997). Это усовершенствованный метод ближайшего соседа (NJ), который, в свою очередь, выгодно отличается от обычного невзвешенного парно-группового метода построения дендрограмм (UPGMA). Важно, в частности, что NJ и BioNJ не требуют предположения об одинаковых темпах эволюции. Иными словами, длина сестринских ветвей в общем случае различна. Кроме того, эти методы создают неукорененные деревья, тогда как в UPGMA они автоматически укореняются по нижней ветви. В данном же случае укоренение деревьев, как и выбор корневой группы, – результат сознательного решения. Вторая разновидность – сетчатое древо, в котором учитываются не только вертикальные (генеалогические), но и горизонтальные (ареальные) связи⁴. Такое древо заведомо исключает узко-генеалогический подход, согласно которому, как уже было указано выше, у каждого языка может быть всего один предок. Из сказанного следует, что обе разновидности генеалогической модели могут давать сходные результаты, но в случаях, когда условия, предъявляемые узко-генеалогической моделью, не выполняются, результаты могут существенно различаться. Сетчатое древо дает наглядное представление о том, насколько сильно эти условия нарушаются в конкретных ситуациях, составляющих, судя по всему, правило, а не исключение. Речь идет не только о неучтенных заимствованиях, попавших в списки базовой лексики, ведь сам процесс ветвления родословного древа подчинен не только хронологическим, но и географическим закономерностям. Вследствие таких закономерностей возникают “ареально-генетические” связи, при которых когнаты сохраняются в сестринских диалектах, а затем и в языках благодаря соприкосновению их ареалов (Хелимский 1982: 24–25). Сложность этих связей, не поддающихся учету при узко-генеалогическом подходе, и отражают новые модели.

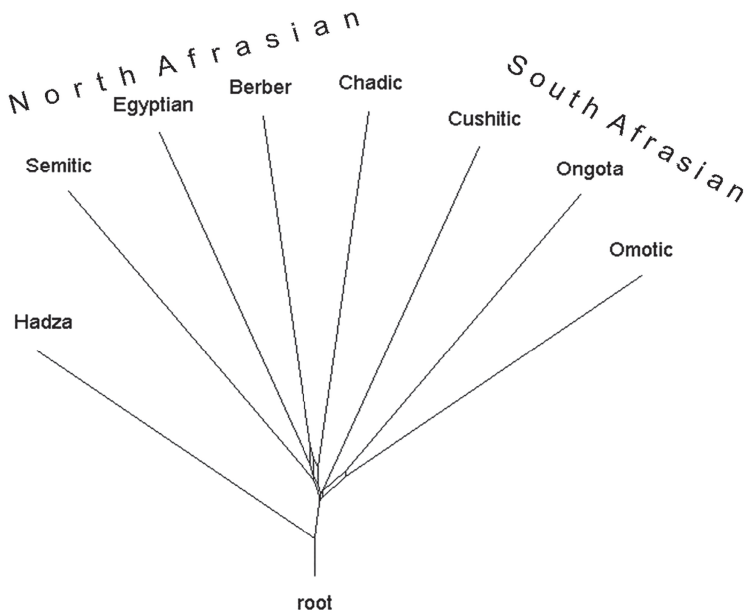


Рис. 1. Сетчатое древо АА семей, укорененное по хадза

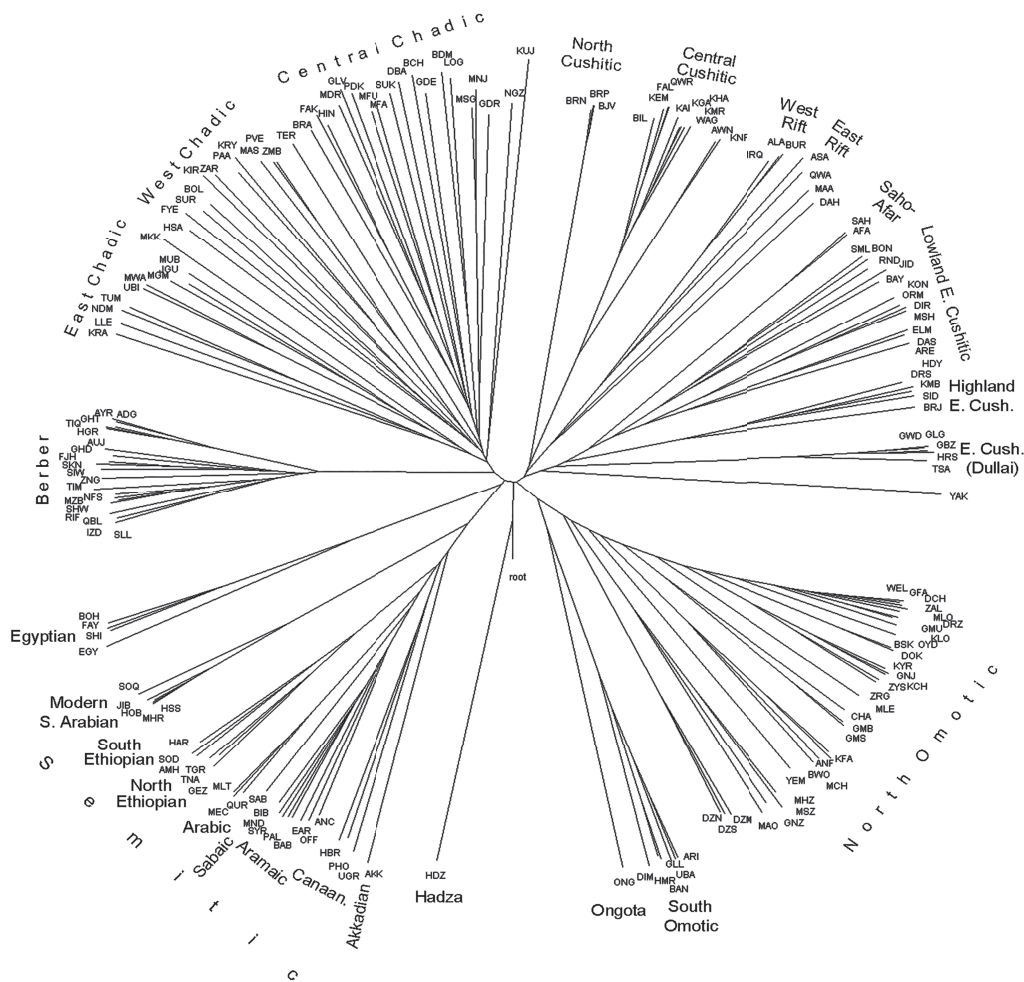


Рис. 2. Родословное дерево АА языков, укорененное по хадза (обозначения см. в тексте; электронную версию дерева см.: РД-2)

Вторая модель – ареальная. В ней языки представлены не в виде ветвей, а в виде точек на плоскости, созданной неметрическим многомерным шкалированием матрицы попарных расстояний (относительного числа несовпадений). Точки соединяются в многомерном пространстве кратчайшим путем – с помощью минимального остовного дерева⁵. Эта модель еще дальше от глоттохронологической, зато открывает широкие перспективы для ареального анализа.

Построим по усредненным данным сетчатое дерево АА семей. Укореним его по формально внешнему (во всяком случае, с наименее ясным местом в классификации) таксону – хадза (см.: Рис. 1). Такой выбор мотивирован тем, что, если пренебречь заимствованиями⁶, сходство той или иной семьи с внешним языком вызвано сохранением черт языка, являвшегося их общим предком.

Структура дерева выражена слабо. Выделяются описанные А.М. группировки: северная (семитско-египетско-берберо-чадская) и южная (кушомотская, к которой принадлежит онгота). Это отличается от классификации Г.С. Старостина и

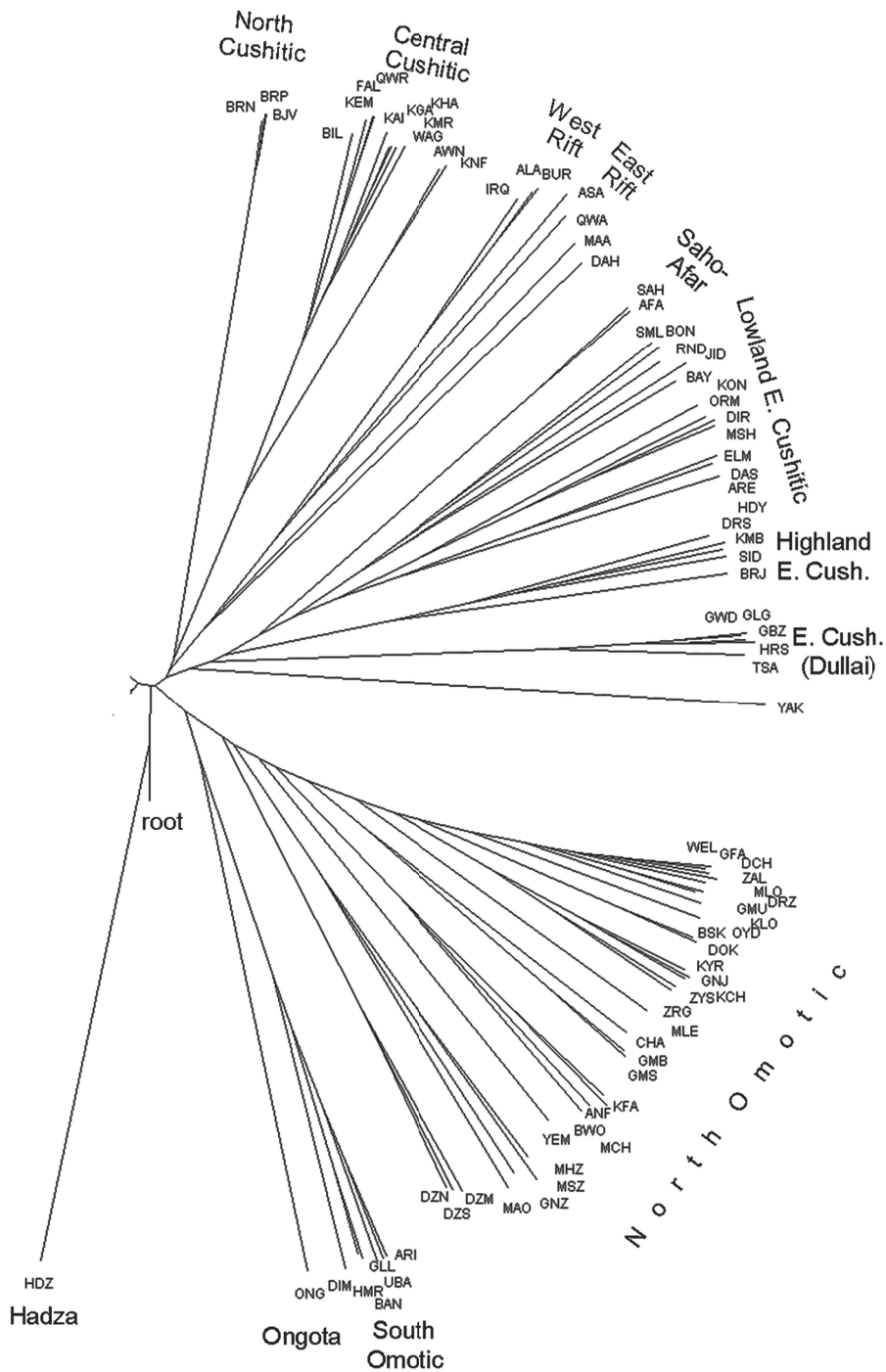


Рис. 3. Родословное дерево южно-АА языков (обозначения см. в тексте; электронную версию дерева см.: РД-3)

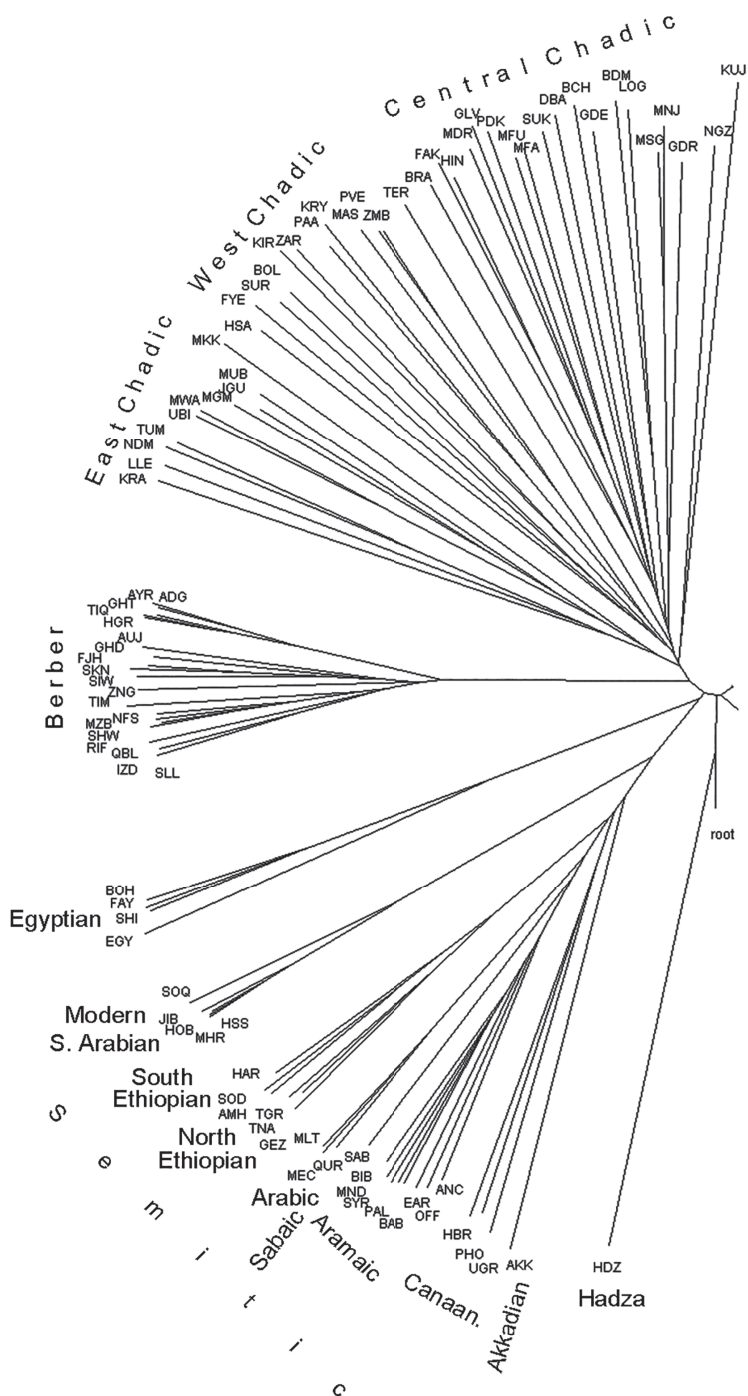


Рис. 4. Родословное дерево северно-АА языков (обозначения см. в тексте; электронную версию дерева см.: РД-4)

коллег, где омотская ветвь оказывается древнейшей в АА семье, а кушитская – сестринской по отношению к северным АА семьям вместе взятым (РД-1).

Обе АА группировки распались, судя по всему, вскоре после их обособления (южная – почти сразу же, северная – вскоре после этого). Коллатераль у основания ветвей показывает, что в пределах южной группировки язык онгота не только родствен омотской семье, но и связан с ней ареальными контактами. При этом онготская ветвь находится в промежутке между омотской и кушитской. Действительно, в онгота есть заимствования из южно-омотского языка хамар (хотя в нескольких случаях можно предположить и противоположное направление заимствований: из онгота в хамар) и восточно-кушитского языка цамай (группа дуллаи). В пределах северной группировки египетская, берберская и чадская ветви особенно близки и генеалогически, и ареально; семитская чуть более обособлена.

Посмотрим на генеалогические отношения отдельных АА языков. Сначала используем простое древо, построенное методом BioNJ. В качестве корневого языка снова выберем хадза (см.: Рис. 2; то же древо, показанное крупным планом отдельно для южной и северной группировок, см.: Рис. 3 и 4 соответственно).

Мы видим здесь, в сущности, то же, что на древе семей (см.: Рис. 1): обе южные семьи разделяются почти сразу же после того, как их общий предок отделяется от предка северно-АА языков (см. Рис. 3). Далее распадается кушитская семья, за ней омотская. Северно-центральнокушитская, а также рифтская и восточнокушитская подсемьи возникают почти одновременно. Ситуация с омотской семьей иная: ее распад выглядит цепью множества последовательных моментов ветвления, причем первой обособляется южноомотская ветвь, к которой принадлежит онгота, а потом одна за другой отделяются северноомотские ветви, начиная с дизоидной. Отчасти сходные результаты были получены В. Блажеком в классификации кушитских – с главным отличием в позиции дахало – и омотских языков (*Blažek 1997, 2008*).

В северной группировке АА языков (см.: Рис. 4) сначала обособляются семитская ветвь с египетской (сестринской по отношению к ней), затем берберская отделяется от чадской. В пределах семитской ветви первым обособившимся языком оказывается общий предок современных южноаравийских языков – праюжносемитский, по А.М. Аккадский язык обособляется вторым, хотя хронологически именно он, наряду с эблаитским, является наиболее древним из известных семитских языков. И наконец возникают (почти одновременно) три западносемитские ветви: ханаанейская (иврит, финикийский, угаритский⁷), арабско-арамейская с южноаравийским эпиграфическим (сабейским) языком и эфиопская. Это несколько отличается от лексикостатистической классификации А.М., где выделяются ветви левантская (угаритская и ханаанейско-арамейская, или южно-levantийская, включающая сабейский язык, что также отличает эту классификацию от большинства принятых в семитологии), арабская и эфиопская.

Если не считать египетской ветви, включающей коптские диалекты⁸, самой консолидированной ветвью АА макросемьи является берберская, распавшаяся сравнительно недавно. Чадская же семья значительно менее однородна, возможно, по причине разных субстратов. Косвенным указанием на это служит то, что чадские популяции генетически чрезвычайно гетерогенны (*Tishkoff et al. 2009; Shriner, Rotimi 2018*). К сожалению, у нас нет достаточных данных по не-АА языкам Африки, позволяющих количественно учесть роль субстратных явлений в дифференциации не только чадской, но и кушитской и омотской семей, а также в своеобразии онгота.

Сетчатое древо АА языков (см.: Рис. 5–7) вносит определенные коррективы, и ветви расположены здесь не так, как на простом древе. Во-первых, выбор хадза в качестве корневого языка может быть подвергнут сомнению. Его носители – соседи носителей западнорифтских языков, в частности языка иракв, который

связан с реконструированным корнем коллатералиями. Масштаб этих коллатералей, впрочем, невелик, так что в данном случае ими, вероятно, можно пренебречь.

Во-вторых, сетчатое дерево ставит под сомнение существование кушитской семьи как монофилетического таксона. Если праомотский язык и здесь заметен в виде отчетливого ствола, соединяющего дочерние омотские ветви, включая онгота, с корневой частью дерева, то пракушитский язык оказывается проблематичным: вместо ствола наблюдается густая сеть коллатералей. Иными словами, кушитская семья сразу возникает в виде довольно аморфного куста древних, связанных между собой ветвей, и нет уверенности, что это монофилетическая, сестринская по отношению к омотской группа, а не языковой союз. Мнение о том, что и вся кушомотская группировка в целом, возможно, не монофилетический таксон, а языковой союз, уже высказывалось В.Э. Орлом и О.В. Столбовой

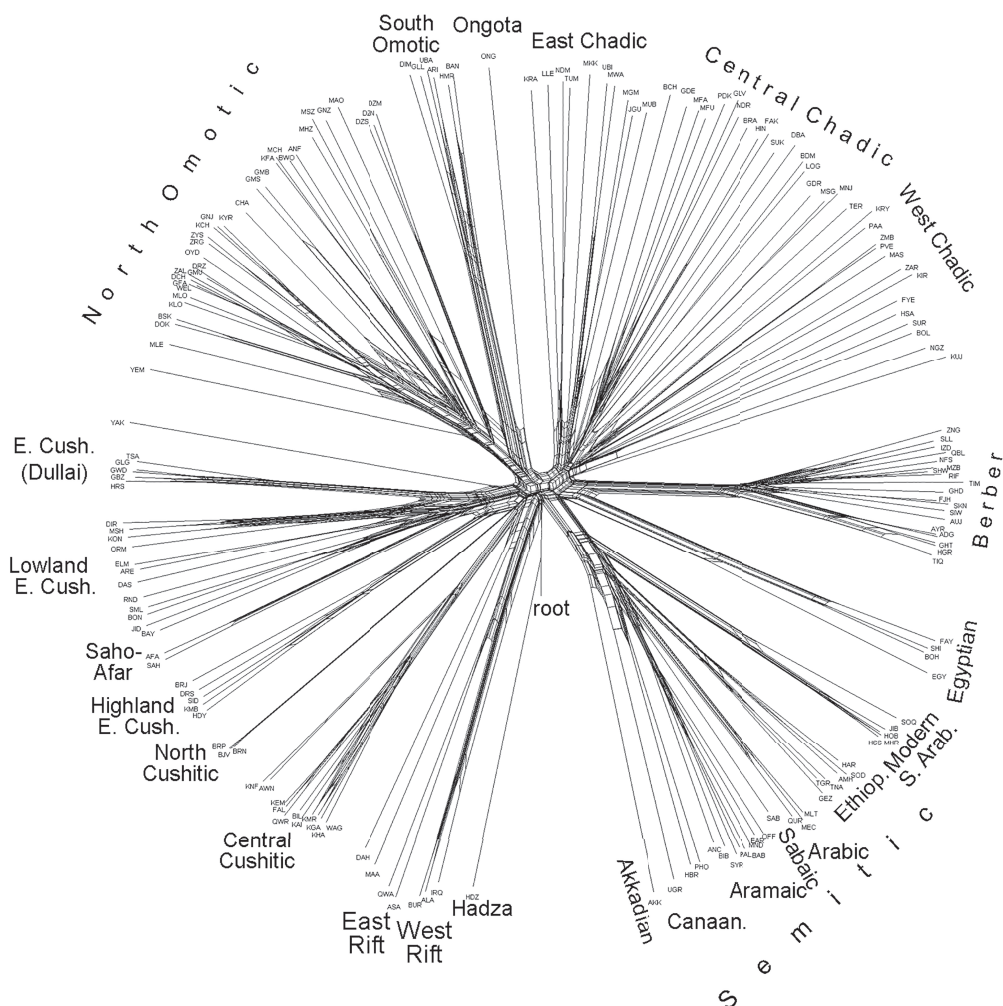


Рис. 5. Сетчатое дерево АА языков, укорененное по хадза (обозначения см. в тексте; электронную версию дерева см.: РД-5)

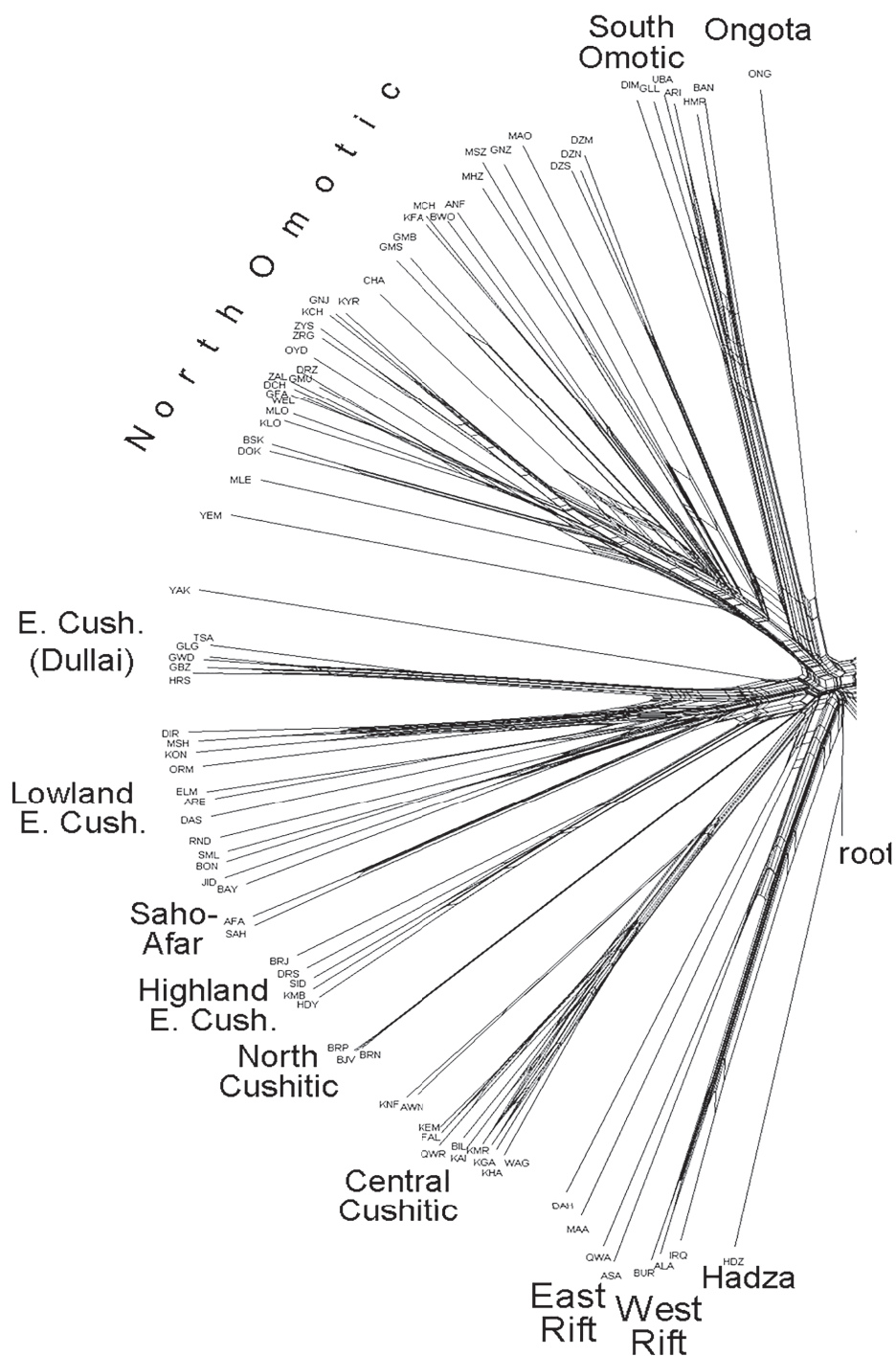


Рис. 6. Сетчатое древо южно-АА языков
(обозначения см. в тексте; электронную версию древа см.: РД-6)

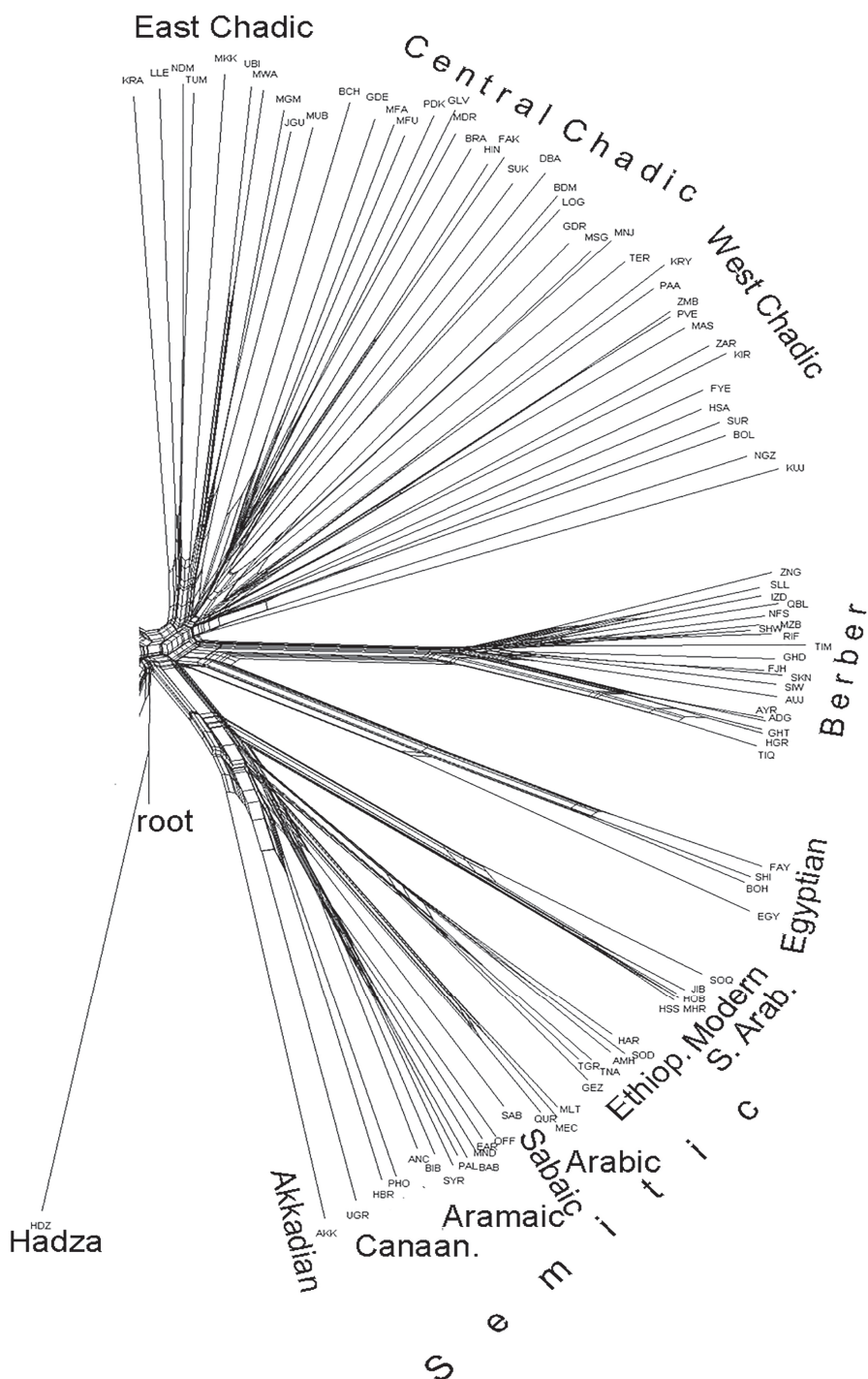


Рис. 7. Сетчатое дерево северно-АА языков (обозначения см. в тексте; электронную версию дерева см.: РД-7)

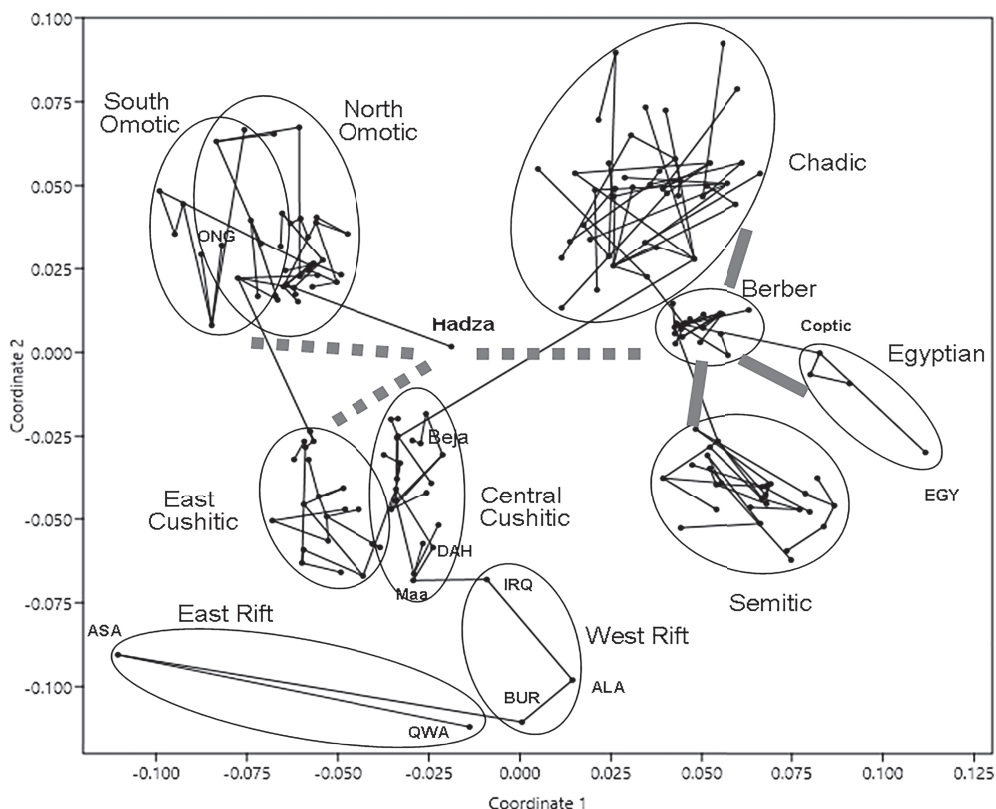


Рис. 8. Взаимоположение языков и семей по результатам неметрического многомерного шкалирования матрицы лексических соответствий. Точки – языки, прямые линии – ребра минимального остовного дерева, показывающего кратчайший путь соединения точек в многомерном пространстве (тонкие линии – по данным об отдельных языках, толстые сплошные – по усредненным данным о семьях и изолированных языках, связанных на уровне 9–12% лексических совпадений, толстые пунктирные – на уровне 7–8% совпадений); овалами обведены семьи (обозначения см. в тексте)

(Orel, Stolbova 1995: X), хотя в последующих работах О.В. Столбовой по чадской этимологии все кушитские языки входят в отдельную ветвь афразийской макросемьи, без объединения с омотской (см., напр.: Столбова 2016). Результаты нашего анализа подтверждают реальность кушомотской макросемьи. Правда, как показывает даже предельно схематизированное обычное дерево (см.: Рис. 2), общий ствол, соединяющий ее с корневой частью, очень короток. Густая сеть ареальных связей на сетчатом дереве (см.: Рис. 5) делает противопоставление северной группировки АА языков южной еще менее отчетливым. Дифференциация североомотской подсемьи здесь уже не кажется цепью последовательных ветвлений, как на Рис. 3, из-за очень интенсивных контактов между омотскими, особенно североомотскими, языками (см.: Рис. 6).

Обратимся к ареальной модели (см.: Рис. 8). Хотя у такого графика с формальной точки зрения в принципе не может быть ни корня, ни центра, усредненные данные о семьях и изолированных языках выявляют два центра.

Южный, подобно корню деревьев (см.: Рис. 1–7), пусть и не локализован точно, должен находиться вблизи языка, либо не принадлежащего к АА макросемье, но близкого к ней, либо, во всяком случае, наиболее обособленного в ее пределах, – хадза. Двумерная проекция, в отличие от сетчатого древа (см.: Рис. 5, 6), не обнаруживает специфической близости хадза к западнорифтской ветви кушитской семьи.

Так как двумерная проекция, в отличие от древа, передает прежде всего общие соотношения, жертвуя частностями, анализ связей между отдельными языками ничего не дает, за некоторыми исключениями. Так, в пределах кушитской семьи бросается в глаза аномально большое различие между двумя восточно-рифтскими языками – квадза и аса. На сетчатом древе (см.: Рис. 6) язык аса промежуточен между квадза и западнорифтским языком бурунге. На двумерной же проекции аса весьма удален и от квадза, и от бурунге, хотя и соединен с ними обоими. Это очень нетипичный случай, требующий особого анализа. Ареальная модель, в отличие от генеалогической, способна выявить такие аномалии.

Ребра минимального остова древа связывают реконструированный южный центр как с обеими южными семьями (омотской и кушитской), так и с одной из северных – берберской. Ввиду позднего распада последней может показаться странным, что именно она оказалась вблизи предполагаемого северного центра дифференциации АА макросемьи, будучи соединена, помимо южного центра, с тремя северно-АА семьями – семитской, египетской и чадской (эти связи особенно сильны). Тот же результат получен по базе данных Г.С. Старостина и коллег (Козинцев 2021). Однако, хотя дочерние берберские языки молоды, праберберский возник вскоре после распада северно-АА группировки (см.: Рис. 4, 7). Поэтому его родство с отдельными северными ветвями установить трудно: на простом древе (см.: Рис. 4) он родствен чадской ветви, тогда как на сетчатом (см.: Рис. 7) и на двумерной проекции (см.: Рис. 8) – как чадской, так и египетской, в особенности коптской.

Псевдоареальная модель позволяет выявить некоторые градиенты, в большинстве случаев географические. Так, восточнокушитские и центральнокушитские языки промежуточные между омотскими и рифтскими, центральнокушитские – между восточнокушитскими и западнорифтскими, берберские – между семитскими и чадскими, а также между египетскими и чадскими. Возможно, это отражает пути расселения северных афразийцев: сначала с востока на запад вдоль северного побережья Африки, затем на юг – в Сахель. Возможны и хронологические градиенты. Заслуживает внимания, например, то, что позднеегипетские (коптские) диалекты обнаруживают сдвиг от более ранней стадии египетского языка в сторону берберских языков. Средняя доля лексических совпадений египетского языка с берберскими равна 8,4%, для коптских диалектов она возрастает до 12,7%. Из 19 берберских языков нет ни одного, который был бы ближе к египетскому языку, чем к коптским диалектам. Даже самый слабый тест (двусторонний знаковый) показывает, что вероятность случайности такого распределения меньше, чем один шанс из 262 тыс., т.е. что на поздней стадии развития египетского языка ареальные его связи с берберской семьей явно усилились⁹. Деревья не могут выявить такие закономерности, а потому генеалогический и ареальный методы дополняют друг друга. Это хорошо видно на примере египетской ветви: сравнение ранней стадии ее развития (собственно египетской) и поздней (коптской), строго говоря, неправомерно в рамках генеалогического подхода, не допускающего сравнение предков с потомками на равных правах¹⁰. Ареальная же модель свободна от такого ограничения и способна выявить диахронические сдвиги в развитии языка (в данном случае – в сторону сближения с берберскими языками).

Семитская база

Использованы данные о следующих 56 семитских идиомах:

I. Аккадский – АКК.

II. Ханаанейские: HBR – иврит, PHO – финикийский.

III. Угаритский – UGR¹¹.

IV. Сабейский – SAB.

V. Арамейские: ANC – древнеарамейский, ASS – современный “ассирийский” арамейский, BAB – вавилонский арамейский, BIB – библейский арамейский, CHR – христианский арамейский, EGP – египетский арамейский, MLL – маалула, MND – мандейский, NAB – набатейский, NRW – нерва, OFF – официальный арамейский, PAL – палестинский арамейский, PLM – пальмирский арамейский, SAM – самаритянский арамейский, SYR – сирийский арамейский, TUR – туройо, URM – урмийский, ZKH – захо.

VI. Арабские: ALG – алжирский арабский, BKH – бухарский арабский, CPR – кипрский арабский, EGY – египетский арабский, HDR – хадрамаутский арабский, IRQ – иракский арабский, JRD – иорданский арабский, KSH – кашкадарьинский арабский, LBN – ливанский арабский, LYB – ливийский арабский, MEC – мекканский арабский, MLT – мальтийский арабский, MRC – марокканский арабский, PLS – палестинский арабский, QUR – коранический арабский, SDN – суданский арабский, SRN – сирийский арабский, YEM – йеменский арабский¹².

VII. Современные южноаравийские: HOB – хобьот, HSS – харсуси, JIB – джиббали, MHR – мехри, SOQ – сокотри.

VIII. Эфиосемитские: южные: AMH – амхарский, ARG – аргобба, CHH – чаха, GAF – гафат, HAR – харари, SOD – соддо, WOL – волане; северные: GEZ – геэз, TGR – тигре, TNA – тигринья.

Сетчатое древо основных подразделений семитской семьи, укорененное по самому древнему из представленных в базе языков – аккадскому, показано на Рис. 9. Следует обратить внимание на два обстоятельства. Во-первых, арамейская и сабейская ветви короче других, что может свидетельствовать либо о замедленном темпе их эволюции (это предположение не разделяется ни одним из авторов данной статьи), либо о значительной роли ареальных связей с иными семьями. Самая же длинная ветвь – современная южноаравийская, что, возможно, отражает ее ускоренное развитие в условиях изоляции (это предположение разделяет лишь А.К.). Во-вторых, ветви древа в пределах кластеров расположены не в произвольном порядке, как на обычном древе, а в том, который определяется ареальными контактами. Последние показаны в виде коллатералей у основания ветвей. Видно, в частности, что современные южноаравийские языки имели контакты с эфиосемитскими, эфиосемитские – с арабскими, а арабские – с сабейским языком. Сабейский, в свою очередь, контактировал с родственными ему арамейскими языками, арамейские языки – с ханаанейскими, а ханаанейские – с угаритским.

Посмотрим на генеалогические отношения отдельных семитских языков. Сначала используем простое неукорененное древо, построенное методом BioNJ (см.: Рис. 10).

Хотя большинство ветвей семитской семьи – арамейская, арабская, современная южноаравийская и эфиопская – выделяются достаточно четко, такая классификация заведомо неудовлетворительна, так как ранг одного из древнейших известных семитских языков – аккадского – явно занижен. Здесь он оказывается не самостоятельной ветвью, а всего лишь веточкой, пусть и наиболее обособленной, в кусте, включающем кроме нее собственно ханаанейские языки (финикийский и иврит), а также угаритский.

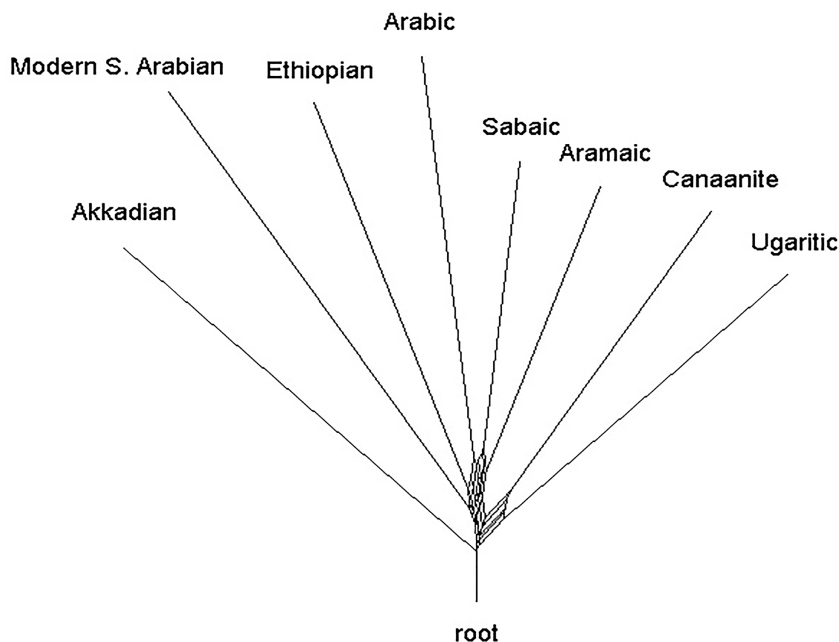


Рис. 9. Сетчатое дерево семитских семей и обособленных языков, укорененное по аккадскому языку

Одна из причин такой аномалии в том, что неукорененные деревья, создаваемые методом NJ/BioNJ, адекватно передают лишь взаимоположение ветвей в верхней части кроны, тогда как порядок ветвления дерева на нижних уровнях остается неизвестным. Для укоренения необходимо выбрать либо внешний язык (в общеафразийской части статьи таким языком был хадза), либо язык, принадлежащий к данной семье, но обособившийся раньше других. Вариантов тут два: либо аккадский, либо предок современных южноаравийских языков. На то, что именно последние, в отличие от аккадского, могут представлять древнейшую ветвь семитской семьи, указывают данные А.М. К сожалению, этот вариант проблему не решает – топология дерева при укоренении его по языку, предковому для современных южноаравийских, остается в точности той же, что и при отсутствии корня.

Остается аккадский язык (см.: Рис. 11). Данный вариант выглядит несколько более адекватно, но аккадский все равно “притягивает” к себе угаритский и ханаанейские языки, которые отделяются от корня вслед за ним. Арамейская ветвь оказывается сестринской по отношению к арабской, а не к ханаанейско-угаритской; к тому же кластеру относится и сабейский язык, который здесь выглядит более родственным арабским, чем арамейским идиомам. Современные южноаравийские и эфиосемитские языки входят в один таксон в качестве сестринских ветвей. Последний результат соответствует выводам ряда авторов (*Bender 1970; Hetzron 1972: 17; Boivin et al. 2009; Kitchen et al. 2009*). Однако исходные данные заставляют в этом усомниться: среднее число соответствий между современными южноаравийскими и эфиосемитскими языками составляет всего 19,4%, тогда как современные южноаравийские языки в среднем ближе

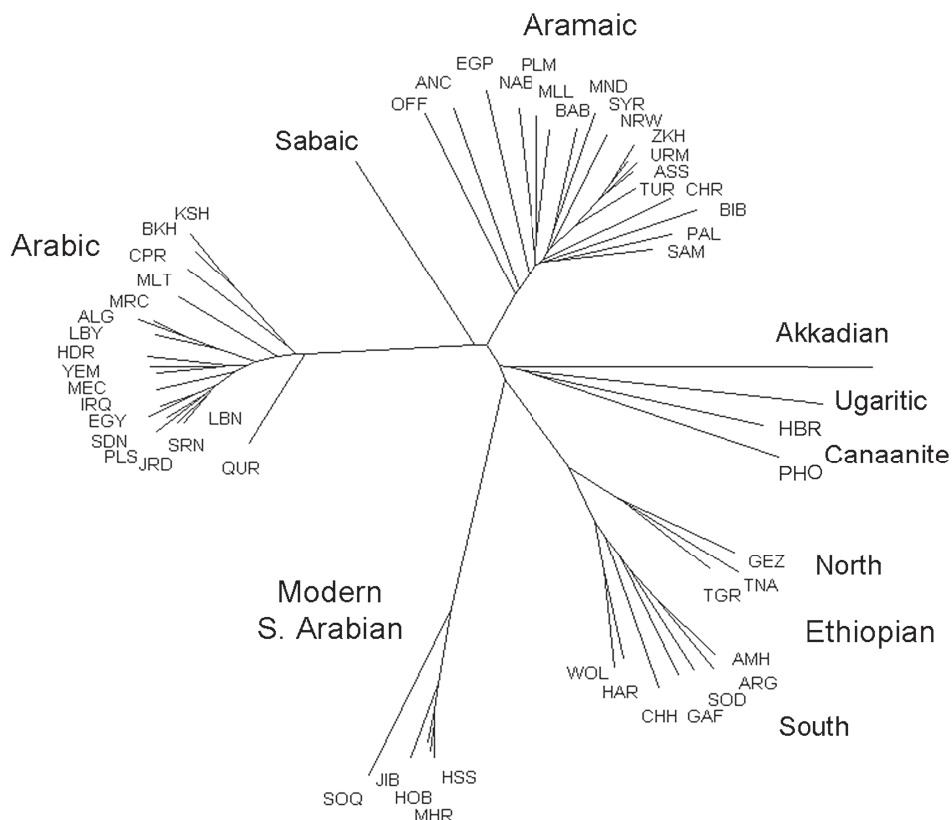


Рис. 10. Неукорененное родословное дерево семитских языков (обозначения см. в тексте)

к арамейским (25,5%) и арабским (21,2%), а эфиосемитские – к арамейским (31,5%), арабским (25,5%) и аккадскому (21,9%). Итак, полученный результат – артефакт кластерного анализа.

Построим теперь сетчатое древо, которое снова укореним по аккадскому (см.: Рис. 12). Такое древо, отражающее не только вертикальные связи, но и горизонтальные, на первый взгляд, легче совместить с идеей А.М. о северозападносемитской (levantийской) ветви, включающий ханаанейские и арамейский языки, а также угаритский (см. также: *Коган, Лёзов* 2009: 205–805). Подобно простым деревьям (см.: Рис. 10, 11), сетчатое древо показывает, что изоглоссы, объединяющие данные языки, имеют в значительной мере ареальное происхождение (ср.: *Kogan* 2015: 600). Сабейский же язык, который на простых деревьях несколько ближе к арабским идиомам, чем к арамейским, здесь промежуточен между теми и другими. Неясна природа коллатералей, соединяющих иврит с официальным арамейским, а сабейский с кораническим арабским (поэтому простые деревья создают иллюзию того, что сабейская ветвь – сестринская по отношению к арабской). Идет ли здесь речь об ареальных связях или о сохранении предковой лексики, сказать трудно. Развитие праязыков – раннеарабского¹³, праарамейского, праэфиосемитского и праханаанейского – отображается на сетчатом древе в виде рядов параллельных линий (диалектов?), что особенно заметно в случае раннеарабского языка. Язык же, предковый по отношению к

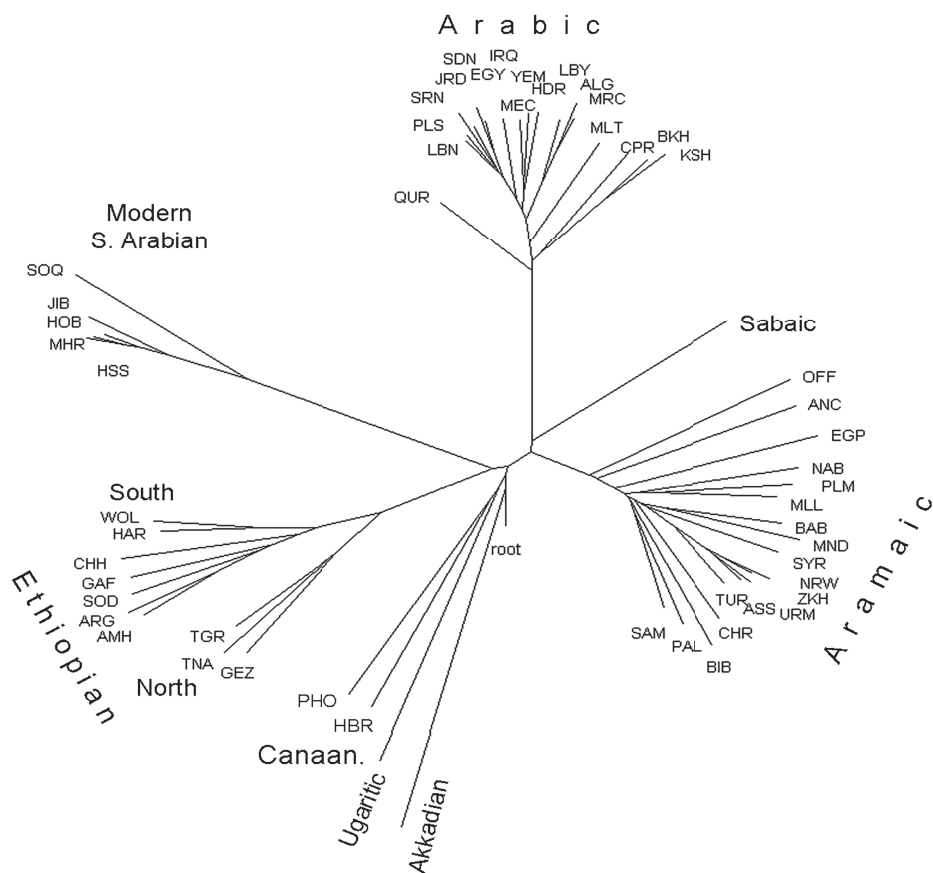


Рис. 11. Родословное дерево семитских языков, укорененное по аккадскому языку (обозначения см. в тексте)

современным южноаравийским, был значительно однороднее. Он и существовал дольше всего, тогда как языки-предки арамейских и особенно ханаанейских языков распались значительно раньше. Все это видно и на простых деревьях (см.: Рис. 9–11).

Эфиосемитская и южноаравийская ветви и на сетчатом древе кажутся сестринскими, хотя перемычка, отделяющая их от корневой части древа, почти отсутствует. Интересно, что если укоренить данное древо по языку, предковому для современных южноаравийских, то не меняется практически ничего: топология древа остается почти той же, что и при выборе аккадского языка в качестве корневого. Это усиливает сомнения в том, что эфиосемитские и современные южноаравийские языки объединены специфическим родством.

Двумерная проекция (см.: Рис. 13) вносит в этот вопрос окончательную ясность: о филогенетической близости эфиосемитских и современных южноаравийских языков не может быть и речи (тот же результат, основанный на иных подходах, см.: *Kogan 2015: 599–600*). Сестринские ветви здесь, как ни удивительно, вообще отсутствуют, за исключением обеих групп эфиосемитских языков (см., впрочем, ниже). Подобная конфигурация не может быть адекватно передана с помощью генеалогической модели, нацеленной именно на

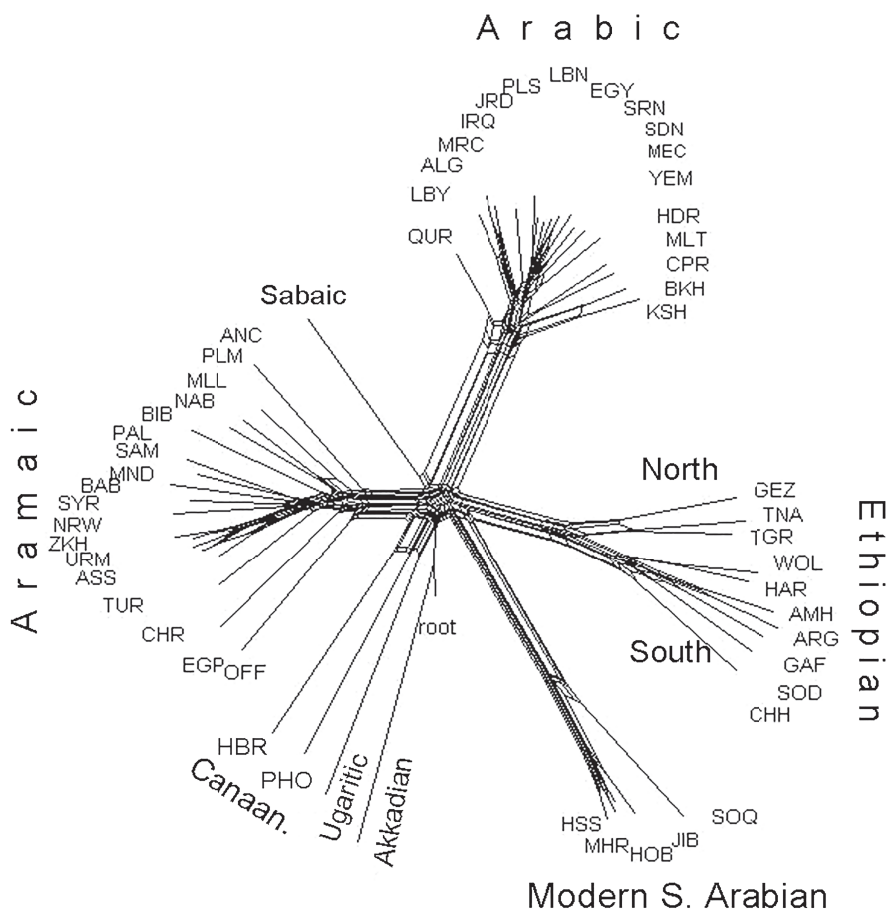


Рис. 12. Сетчатое дерево семитских языков, укорененное по аккадскому языку (обозначения см. в тексте)

выявление сестринских ветвей. Быть может, полученный результат – артефакт пространственной модели? Нет, потому что при использовании ее на более высоком – афразийском – уровне сестринские пары четко выявляются. Таковы, в частности, южноомотская и североомотская, восточнокушитская и центральнокушитская (сближающаяся с беджа), восточнорифтская и западнорифтская ветви (см.: Рис. 8). Но при этом главные афразийские ветви мало различаются по степени обособленности: амплитуда колебания средней доли лексических совпадений с другими ветвями составляет всего 2,1% – от 5,4% у кушитской ветви до 7,5% у берберской. В случае же семитских ветвей размах значительно больше – 14,4% (см. ниже).

Арамейская ветвь с сабейским языком занимает на графике центральное положение. Ребра минимального остонового дерева соединяют ее с тремя другими ветвями, а близкородственный ей сабейский язык – с четырьмя. Между тем ни одна из прочих ветвей не имеет более одной непосредственной связи. По степени лексического своеобразия ветви располагаются так (в порядке возрастания средней доли совпадений с другими): современная южноаравийская – 20,7%; аккадская – 24,0%; эфиопская – 27,0%; арабская – 27,1%; угаритская – 28,9%; ханаанейская – 31,2%; арамейская – 34,7%; сабейская – 35,1%. На обобщенном

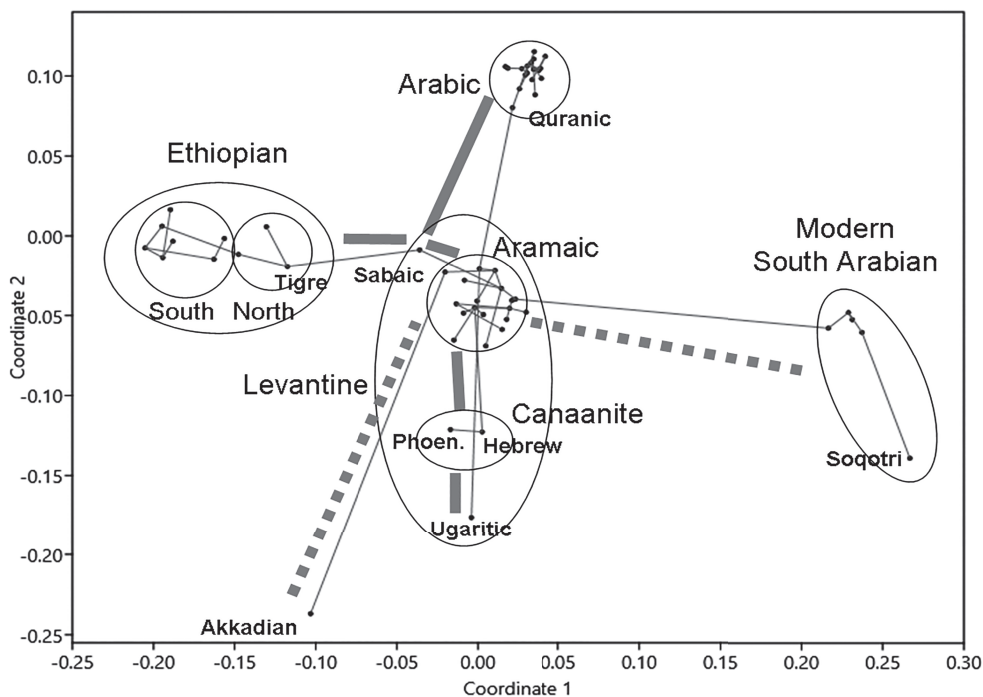


Рис. 13. Взаимоположение семитских языков по результатам неметрического многомерного шкалирования матрицы лексических соответствий. Точки – языки, прямые линии – ребра минимального остовного дерева, показывающего кратчайший путь соединения точек в многомерном пространстве (тонкие линии – по данным об отдельных языках, толстые сплошные – по усредненным данным о семьях и изолированных языках, связанных на уровне 37–47% лексических совпадений, толстые пунктирные – на уровне 26–29% совпадений); кругами и овалами обведены основные ветви (обозначения см. в тексте)

сетчатом древе (см.: Рис. 9) центральное положение арамейских и сабейского языков проявляется в малой длине соответствующих ветвей. В рамках генеалогического подхода это заставило бы рассматривать вопрос о замедленном темпе их эволюции, что маловероятно.

Представляется, что ареальный подход в данном случае более адекватен: центральное положение арамейской группы, вероятно, связано с ролью арамейского как *lingua franca* на всем Ближнем Востоке в течение длительного времени. Поэтому речь, скорее всего, идет не о консервации предковой лексики в арамейских идиомах, а об “ареально-генетических связях” (Хелимский 1982: 24–25), при которых когнаты сохраняются благодаря контактам. Если какая-то ветвь, в отличие от других, играет роль “универсального донора”, то ее своеобразие при прочих равных должно быть минимальным.

Сабейская ветвь наименее удалена от остальных. Она связана в большей степени с арамейской, чем с арабской, как могло бы показаться при использовании генеалогической модели (см.: Рис. 10–12)¹⁴. Усредненная доля лексических совпадений сабейского языка с арамейскими идиомами равна 47,4%, с арабскими – 38,4% (согласно критерию Манна-Уитни, различие в высшей степени значимо: $U=12,5$, $z=4,75$, $p<0,001$). “Арамейское” уклонение коранического арабского языка видно и здесь.

Является ли левантийская (северозападносемитская) группа, включающая, согласно А.М., арамейские и ханаанейские языки, а также угаритский, монофилетическим или же полифилетическим таксоном, т.е. языковым союзом? Иными словами, является ли арамейская группа с сабейским языком сестринской по отношению к ханаанейской группе и угаритскому языку? Простые деревья дают на это отрицательный ответ (см.: Рис. 10, 11), сетчатое дерево – неопределенный (см.: Рис. 12). В двумерной проекции, как мы уже сказали, сестринских групп в строгом смысле вообще нет, причем ветви сильно различаются по степени своеобразия. То, что это своеобразие минимально у левантийских ветвей, особенно у арамейской с сабейским языком, заставляет предполагать, что именно эти ветви находились в средоточии семитского языкового универсума, поддерживая множество контактов, что затрудняет решение вопроса об их монофилии.

Интересно взаимоположение обеих ветвей эффиосемитской группы (см.: Рис. 13). Выясняется, что даже они, несмотря на их обособленность от прочих, не могут считаться сестринскими в строгом смысле слова, что видно и на деревьях (см.: Рис. 10–12): северная ветвь обособляется первой, южная – второй. Северная при этом оказывается промежуточной между южной и прочими семитскими ветвями. Видимо, это отражает распространение семитских языков от побережья Красного моря вглубь Африканского континента путем единой миграционной волны (*Булах, Коган* 2013: 29). В современной южноаравийской группе особое положение занимает сокотри, противостоящий прочим языкам данной группы (см.: Рис. 11–13). Это явный результат его долгой изоляции.

Представляется, что использованные методы открывают интересные перспективы для классификации АА языков и для компаративистики в целом, а значит, и для реконструкции этнокультурной истории на разных этапах. Они применимы также в этнографии, археологии и иных смежных областях.

Благодарности

А.М. пользуется случаем выразить благодарность Л.Р. Городецкому, поделившемуся с ним 100-словными списками ряда современных арабских идиомов, полученными от их носителей.

Мы признательны анонимным рецензентам нашей статьи, чьи замечания помогли ее улучшить, а также всем многочисленным коллегам, дискуссии с которыми были для нас чрезвычайно важны.

Примечания

¹ Последние усовершенствованные А.М. 100-словные списки готовятся к публикации.

² Один из рецензентов – без сомнения, авторитетный семитолог – высказал ряд во многом справедливых претензий к группировке семитских языков в нашей статье. В частности, понятия “египетский арамейский” и “официальный арамейский” перекрываются (то же относится к терминам «современный “ассирийский” арамейский» и “урмийский арамейский”). Авторам это хорошо известно. Однако составленные по соответствующим корпусам “официальных арамейских” и “египетских арамейских” текстов 100-словные списки совпадают, во-первых, не на 100% (расходятся в пунктах “not” и “walk”), а во-вторых, покрывают только часть 100-словника (48 слов + три синонима представлены в текстах на “официальном арамейском”, 52 слова + три синонима – на “египетском арамейском”). Вариант объединения двух 100-словных списков в один представляется паллиативом. Мы сознаем, что перекрывание создает избыточ-

ность, однако, как показывают графики, оно все-таки неполное и дает возможность получить дополнительную информацию.

³ Рецензент пишет: “Огромнейшее большинство афразийских таксонов в сравнительно-историческом плане разработаны *крайне плохо* или *вообще не разработаны*” (курсив наш. – А.К., А.М.) и «В статье встречаются такие (на взгляд рецензента. – А.К., А.М.) малонаучные концепты, как “совпадающие слова”. Что такое “совпадающие слова”? Мне это решительно не понятно. Совпадающие семантически? Точно (“смерть” = “смерть”) или приблизительно (“смерть” = “ветшать”)? Совпадающие фонетически? Но неужели авторам настолько хорошо известны законы афразийского консонантизма, чтобы судить о “совпадении” между отдельными словами живых чадских, семитских или омотских языков?» Комментарий А.М.: В этих суждениях игнорируется многолетняя работа целого ряда афразистов, начиная с И.М. Дьяконова, выработавших систему регулярных консонантных соответствий и этимологизировавших огромные пласты лексики большинства афразийских языков. Замечания рецензента основаны либо на предвзятом отношении к исследованиям, выходящим за рамки общеизвестного, либо на недостаточном внимании к опубликованным работам О.В. Столбовой, В. Блажека, Г. Такача, самого А.М. и многих других авторов, разработавших законы консонантизма в отдельных афразийских ветвях и группах. Под термином “совпадающие слова” в контексте лексикостатистики ничего иного, кроме слов, полностью совпадающих семантически и фонетически на основании регулярных консонантных соответствий, иметься в виду не может. Кроме того, рецензент полагает, что нам следовало уделить внимание глагольной морфологии. Это другая и серьезная проблема, к теме статьи и вообще к лексикостатистике прямого отношения не имеющая.

⁴ Обе разновидности генеалогической модели реализованы с помощью пакета SplitsTree4 Д. Хусона и Д. Брайанта. <https://software-ab.informatik.uni-tuebingen.de/download/splitsree4/welcome.html> (дата обращения: 24.03.2022).

⁵ Ареальная модель реализована с помощью пакета PAST Э. Хаммера, версия 4.05. <https://www.nhm.uio.no/english/research/resources/past> (дата обращения: 24.03.2022).

⁶ В данном случае пренебречь возможностью заимствований не удастся ввиду соседства носителей языка хадза и западнорифтских языков (см. в тексте).

⁷ А.М. не относит угаритский язык к ханаанейским, считая его особой ветвью в пределах левантийской группы. Однако этот язык очень близок к ханаанейским.

⁸ Комментарий А.М.: Неожиданным и, по-видимому, неадекватным результатом является раннее отделение коптских диалектов от “собственно египетского” языка. Теоретически существование такого раннеегипетского диалекта – как бы “пракоптского” – возможно, но обсуждать этот вопрос имеет смысл, только если им заинтересуются египтологи. Комментарий А.К.: Такой результат вызван чисто методической причиной: генеалогическая модель направлена на выявление сестринских ветвей и исключает непосредственное сравнение предков с потомками. Строго говоря, предковая линия перестает существовать в момент, когда она разделяется на дочерние. Это заметно ограничивает применимость данной модели.

⁹ Один из рецензентов отнесся к этому выводу с крайним недоверием, да и нас полученный результат сильно удивил. Обращение к исходным данным не позволяет дать ему культурно-историческую интерпретацию. Но никакого иного объяснения кроме ареального не усматривается, если, конечно, использованный материал не содержит некоего скрытого дефекта.

¹⁰ В глоттохронологии использование разновременных языков возможно бла-

годаря поправкам, разработанным С.А. Старостиным (*Бурлак, Старостин* 2005: 142). Но при этом ни один из непосредственно сравниваемых языков все равно не может быть предковым по отношению к другому. Так как данное исследование не основано на глоттохронологических постулатах, поправки не вносились.

¹¹ Традиционно угаритский язык не причислялся к ханаанейским (*Коган* 2009а, 2009б), с чем согласуются результаты А.М. По новой версии Л.Е. Когана, угарит следует включить в эту группу (*Коган* 2015: 343–349, 600). Наши подсчеты также свидетельствуют о значительной лексической близости угаритского к собственно ханаанейским языкам.

¹² Проблема с арабскими идиомами, отмеченная и одним из рецензентов (в меньшей степени она относится и к современным арамейским), состоит в том, что при записи списков со слов информантов фиксировались, кроме имени информанта и (обычно) его возраста, лишь населенный пункт, чей идиом он представлял, и страна, откуда он прибыл (опросы проводились в Москве, в основном среди арабских студентов). Дополнительного исследования с целью выяснить, к какой группе арабских идиомов относится данный, не проводилось.

¹³ Лексикостатистическое и глоттохронологическое сравнение между собой живых арабских идиомов и коранического арабского дает несколько необычный результат: оно не выявляет “праарабский” язык (в отличие от явно выявляемых “праэфиопского”, “праарамейского”, “праюжносемитского” и др.), вернее, он выглядит очень поздним, близким к рубежу эр, что исторически совершенно невероятно. Возможная интерпретация: все включенные в сравнение арабские идиомы, восходящие к относительно недавно (немногим более двух тысяч лет назад) разделившемуся праязыку – одному из арабских диалектов, наложились на все другие арабские идиомы, существовавшие в Аравии с III тыс. до н.э. и фактически их вытеснили.

¹⁴ Последний результат может быть отчасти вызван тем, что сабейские тексты прочитывались в первую очередь с помощью арабского словаря и в сравнении с арабской лексикой.

Источники и материалы

- РД-1 – Родословное древо языков Евразии и Африки, построенное Г.С. Старостиным с соавторами.
http://starling.rinet.ru/new100/eurasia_short.jpg (дата обращения: 24.03.2022).
- РД-2 – Родословное древо АА языков, укорененное по хадза.
<https://disk.yandex.ru/i/VGfYIVJNoHoNbw> (дата обращения: 24.03.2022).
- РД-3 – Родословное древо южно-АА языков.
<https://disk.yandex.ru/i/PiJHGGEAJrH11A> (дата обращения: 24.03.2022).
- РД-4 – Родословное древо северно-АА языков.
<https://disk.yandex.ru/i/5FkkYTnNKxzW7A> (дата обращения: 24.03.2022).
- РД-5 – Сетчатое древо АА языков, укорененное по хадза.
https://disk.yandex.ru/i/VvaQxV_thpDt4Q (дата обращения: 24.03.2022).
- РД-6 – Сетчатое древо южно-АА языков.
<https://disk.yandex.ru/i/XXtlOsSSg9C6Zw> (дата обращения: 24.03.2022).
- РД-7 – Сетчатое древо северно-АА языков.
<https://disk.yandex.ru/i/j3VCJClxaYPMw> (дата обращения: 24.03.2022).
- GLD-1 n.d. – The Tower of Babel: The Global Lexicostatistical Database.
<http://starling.rinet.ru/new100/trees.htm> (дата обращения: 24.03.2022).
- Militarev, Stolbova* 2007 – *Militarev A., Stolbova O.* Afroasiatic Etymology. 2007.
<https://starlingdb.org/cgi-bin/response.cgi?root=config&morpho=0&basena me=\data\semham\afaset&first=1> (дата обращения: 24.03.2022).

Научная литература

- Булах М.С., Коган Л.Е.* Эфиосемитские языки // Языки мира. Семитские языки. Эфиосемитские языки / Ред. М.С. Булах, Л.Е. Коган, О.И. Романова. М.: Academia, 2013. С. 13–141.
- Бурлак С.А., Старостин С.А.* Сравнительно-историческое языкознание. М.: Academia, 2005.
- Коган Л.Е.* Угаритский язык // Языки мира. Семитские языки. Аккадский язык. Северозападносемитские языки / Ред. Л. Коган, С. Лёзов. М.: Academia, 2009а. С. 205–238.
- Коган Л.Е.* Ханаанейские языки // Языки мира. Семитские языки. Аккадский язык. Северозападносемитские языки / Ред. Л. Коган, С. Лёзов. М.: Academia, 2009б. С. 239–278.
- Козинцев А.Г.* Азия или Африка? О локализации афразийской прародины // Этнографическое обозрение. 2021. № 4. С. 25–42. <https://doi.org/10.31857/S086954150016696-5>
- Старостин Г.С.* К вопросу о генетической принадлежности языка хадза // Африканский сборник 2007 / Отв. редактор В.Ф. Выдрин. СПб.: Наука, 2008. С. 262–278.
- Старостин Г.С.* Языки Африки. Опыт построения лексикостатистической классификации. Т. 1, Методология. Койсанские языки. М.: Языки славянской культуры, 2013.
- Столбова О.В.* Этимологический словарь чадских языков. М.: Институт востоковедения РАН, 2016. <https://book.ivran.ru/f/ilovepdfmerged.pdf>
- Хелимский Е.А.* Древнейшие венгерско-самодийские языковые контакты. М.: Наука, 1982.
- Коган Л.Е., Лёзов С.В.* (ред.) Языки мира. Семитские языки. Аккадский язык. Северозападносемитские языки. М.: Academia, 2009.
- Bender M.L.* A Preliminary Investigation of South Arabia // Proceedings of the Third International Conference of Ethiopian Studies. Addis Ababa: Haile Selassie I University, 1970. P. 26–37.
- Blažek V.* Cushitic Lexicostatistics: The Second Attempt // Afroasiatica Neapolitana: Studi Africanistici, Seria Etiopica. Napoli: Istituto Universitario Orientale, 1997. P. 171–188.
- Blažek V.* Nilo-Saharan Stratum of Ongota // Advances in Nilo-Saharan Linguistics / Eds. M. Reh, D.L. Payne. Köln: Köppe, 2007. P. 1–10.
- Blažek V.* Lexicostatistical Comparison of Omotic Languages // In Hot Pursuit of Language in Prehistory: Essays in the Four Fields of Anthropology in Honor of Harold Crane Fleming / Ed. J.D. Bengtson. Amsterdam: John Benjamins, 2008. P. 57–148.
- Blažek V.* Indo-European Zoonyms in Afroasiatic Perspective // Вопросы языкового родства: Международный научный журнал. 2013а. № 9. P. 37–54.
- Blažek V.* Levant and North Africa: Afroasiatic Linguistic History // The Encyclopedia of Global Human Migration. Vol. 1, Prehistory / Ed. P. Bellwood. L.: Blackwell, 2013b. P. 125–132. <https://doi.org/10.1002/9781444351071.wbeghm815>
- Blench R.* Archaeology, Language, and the African Past. Lanham: Altamira Press, 2006.
- Boivin N. et al.* Archaeological, Linguistic and Historical Sources of Ancient Seafaring: A Multidisciplinary Approach to the Study of Early Maritime Contact and Exchange in the Arabian Peninsula // The Evolution of Human Populations in Arabia / Eds. M.D. Petraglia, J.I. Rose. Berlin: Springer, 2009. P. 251–278.
- Ehret C.* History and the Testimony of Language. Berkeley: University of California Press, 2011.
- Fleming H.C.* Ongota: A Decisive Language in African Prehistory. Wiesbaden: Harrassowitz Verlag, 2006.

- Gascuel O.* BIONJ: An Improved Version of the NJ Algorithm Based on a Simple Model of Sequence Data // *Molecular Biology and Evolution*. 1997. Vol. 14. No. 7. P. 685–695.
- Hetzron R.* *Ethiopian Semitic: Studies in Classification*. Manchester: Manchester University Press, 1972.
- Kitchen A. et al.* Bayesian Phylogenetic Analysis of Semitic Languages Identifies an Early Bronze Age Origin of Semitic in the Near East // *Proceedings of the Royal Society B*. 2009. Vol. 276. No. 1668. P. 2703–2710. <https://doi.org/10.1098/rspb.2009.0408>
- Kogan L.* *Genealogical Classification of Semitic: The Lexical Isoglosses*. Boston: Walter de Gruyter, 2015.
- Militarev A.* Towards the Chronology of Afrasian (Afroasiatic) and Its Daughter Families // *Time Depth in Historical Linguistics*. Vol. 1 / Eds. C. Renfrew, A. McMahon, L. Trask. Cambridge: McDonald Institute for Archaeological Research, 2000. P. 267–307.
- Militarev A.* Another Step towards the Chronology of Afrasian (I) // *Babel und Bibel 1, Ancient Near East, Old Testament and Semitic Studies* / Ed. L. Kogan, N. Koslova, S. Loesov, S. Tishchenko. М.: Изд-во Ипполитова; Waynona Lake: Eisenbrauns, 2004. P. 282–333.
- Militarev A.* Once More about Glottochronology and Comparative Method: The Omotic-Afrasian Case // *Аспекты компаративистики = Aspects of Comparative Linguistics* / Под ред. И.С. Смирнова М.: РГГУ, 2005. С. 339–408.
- Militarev A.* Towards the Genetic Affiliation of Ongota, a Nearly Extinct Language of Ethiopia (II) // *Babel und Bibel 3* / Ed. L. Kogan, N. Koslova, S. Loesov, S. Tishchenko. М.: RGGU; Waynona Lake: Eisenbrauns, 2006. P. 489–512. <https://doi.org/10.1515/9781575065823-018>
- Militarev A.* A Complete Etymology-Based Hundred Wordlist of Semitic Updated: Items 1–34 // *Вопросы языкового родства: Международный научный журнал*. 2010. № 3. P. 43–78.
- Militarev A.* A Complete Etymology-Based Hundred Wordlist of Semitic Updated: Items 35–54 // *Вопросы языкового родства: Международный научный журнал*. 2011. № 5. P. 69–95.
- Militarev A.* A Complete Etymology-Based Hundred Wordlist of Semitic Updated: Items 55–74 // *Вопросы языкового родства: Международный научный журнал*. 2012. № 7. P. 71–104.
- Militarev A.* A Complete Etymology-Based Hundred Wordlist of Semitic Updated: Items 75–100 // *Вопросы языкового родства: Международный научный журнал*. 2014. № 11. P. 159–185.
- Militarev A.* Addenda and Conclusion of an Etymology-Based 100-Item Wordlist for Semitic Languages // *Вопросы языкового родства: Международный научный журнал*. 2015. № 13 (1–2). P. 91–138.
- Orel V., Stolbova O.* *Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction*. Leiden: E.J. Brill, 1995.
- Savà G., Tosco M.* A Sketch of Ongota, a Dying Language of Southwest Ethiopia // *Studies in African Linguistics*. 2000. Vol. 29. No. 2. P. 59–135. <https://doi.org/10.32473/sal.v29i2.107366>
- Shriner D. et al.* Genetic Ancestry of Hadza and Sandawe Peoples Reveals Ancient Population Structure in Africa // *Genome Biology and Evolution*. 2018. Vol. 10. No. 3. P. 875–882.
- Shriner D., Rotimi C.N.* Genetic History of Chad // *American Journal of Physical Anthropology*. 2018. Vol. 167. No. 4. P. 804–812. <https://doi.org/10.1002/ajpa.23711>
- Theil R.* Omotic // *Semitic and Afroasiatic: Challenges and Opportunities* / Ed. L. Edzard. Wiesbaden: Harrassowitz Verlag, 2012. P. 369–384.

Tishkoff S.A. et al. The Genetic Structure and History of Africans and African Americans // Science. 2009. Vol. 324. No. 5930. P. 1035–1044.

Research Article

Kozintsev, A.G., and A.Y. Militarev. A Lexicostatistical Classification of Afrasian Languages: New Approaches [Leksikostatisticheskaia klassifikatsiia afraziiskikh yazykov: novye podkhody]. *Etnograficheskoe obozrenie*, 2022, no. 6, pp. 132–159. <https://doi.org/10.31857/S0869541522060082> EDN: MRORLQ ISSN 0869-5415 © Russian Academy of Sciences © Institute of Ethnology and Anthropology RAS]

Alexander Kozintsev | <https://orcid.org/0000-0002-0165-8109> | alexanderkozintsev@yandex.ru | Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Sciences (3 University Emb., St. Petersburg, 199034, Russia)

Alexander Militarev | <http://orcid.org/0000-0002-5684-3053> | amilitarev@gmail.com | Russian State University for the Humanities (6 Miusskaya sq., Moscow, GSP-3, 125993, Russia)

Keywords

Lexicostatistics, Afroasiatic languages, Semitic languages, Egyptian language, Berber languages, Chadic languages, Cushitic languages, Omotic languages

Abstract

A lexicostatistical classification of 205 Afrasian languages was carried out on Alexander Militarev's database according to 100-item lists using various clustering and dimensionality reduction methods. Rooted trees and networks were constructed, and a quasi-spatial model was applied. The separation of Afrasian languages into two groups, northern (Semitic, Egyptian, Berber, and Chadic) and southern (Cushitic and Omotic) is supported, although the monophyly of the Cushitic family is questionable. Within the Semitic family, Aramaic languages together with the apparently related Sabaic take a central position, being connected with all others. Whether or not the Levantine group, which, apart from the above, includes Canaanite languages and Ugaritic, is a clade or a Sprachbund is not clear. Ethiosemitic and Modern South Arabian are not sister branches. Northern Ethiosemitic languages are intermediate between Southern Ethiosemitic and other Semitic languages, possibly evidencing the spread of Semitic inland from the Red Sea coast.

References

- Boivin, N., et al. 2009. Archaeological, Linguistic and Historical Sources of Ancient Seafaring: A Multidisciplinary Approach to the Study of Early Maritime Contact and Exchange in the Arabian Peninsula. In *The Evolution of Human Populations in Arabia*, edited by M.D. Petraglia and J.I. Rose, 251–278. Berlin: Springer.
- Bender, M.L. 1970. A Preliminary Investigation of South Arabia. In *Proceedings of the Third International Conference of Ethiopian Studies*, 26–37. Addis Ababa: Haile Selassie I University.
- Blažek, V. 1997. Cushitic Lexicostatistics: The Second Attempt. In *Afroasiatica Neapolitana: Studi Africanistici, Seria Etiopica*, 171–188. Napoli: Istituto Universitario Orientale.
- Blažek, V. 2007. Nilo-Saharan Stratum of Ongota. In *Advances in Nilo-Saharan Linguistics*, edited by M. Reh and D.L. Payne, 1–10. Köln: Köppe.

- Blažek, V. 2008. Lexicostatistical Comparison of Omotic Languages. In *In Hot Pursuit of Language in Prehistory: Essays in the Four Fields of Anthropology in Honor of Harold Crane Fleming*, edited by J.D. Bengtson, 57–148. Amsterdam: John Benjamins.
- Blažek, V. 2013. Indo-European Zoonyms in Afroasiatic Perspective. *Voprosy yazykovogo rodstva = Journal of Language Relationship* 9: 37–54.
- Blažek, V. 2013. Levant and North Africa: Afroasiatic Linguistic History. In *The Encyclopedia of Global Human Migration. Vol. 1, Prehistory*, edited by P. Bellwood, 125–132. London: Blackwell. <https://doi.org/10.1002/9781444351071.wbeghm815>
- Blench, R. 2006. *Archaeology, Language, and the African Past*. Lanham: Altamira Press.
- Bulakh, M.S., and L.E. Kogan. 2013. Efiosemitskie yazyki [Ethio-Semitic Languages]. In *Yazyki mira. Semitskie yazyki. Efiosemitskie yazyki* [Languages of the World. Semitic Languages. Ethio-Semitic Languages], edited by M.S. Bulakh, L.E. Kogan, and O.I. Romanova, 13–141. Moscow: Academia.
- Burlak, S.A., and S.A. Starostin. 2005. *Sravnitel'no-istoricheskoe yazykoznanie* [Comparative Historical Linguistics]. Moscow: Akademia.
- Ehret, C. 2011. *History and the Testimony of Language*. Berkeley: University of California Press.
- Fleming, H.C. 2006. *Ongota: A Decisive Language in African Prehistory*. Wiesbaden: Harrassowitz Verlag.
- Gascuel, O. 1997. BIONJ: An Improved Version of the NJ Algorithm Based on a Simple Model of Sequence Data. *Molecular Biology and Evolution* 14 (7): 685–695.
- Hetzron, R. 1972. *Ethiopian Semitic: Studies in Classification*. Manchester: Manchester University Press.
- Kogan, L.E., and S.V. Loesov, eds. 2009. *Yazyki mira. Semitskie yazyki; Akkadskii yazyk; Severozapadnosemitskie yazyki* [Languages of the world. Semitic languages. Akkadian language. Northwest Semitic Languages]. Moscow: Academia.
- Khelimskii, E.A. 1982. *Drevneishie vengersko-samodiiskie yazykovye kontakty* [The Earliest Hungarian-Samoyed Contacts]. Moscow: Nauka.
- Kitchen, A., et al. 2009. Bayesian Phylogenetic Analysis of Semitic Languages Identifies an Early Bronze Age Origin of Semitic in the Near East. *Proceedings of the Royal Society B* 276 (1668): 2703–2710. <https://doi.org/10.1098/rspb.2009.0408>
- Kogan, L.E. 2009. Ugaritskii yazyk [Ugaritic]. In *Yazyki mira. Semitskie yazyki. Akkadskii yazyk. Severozapadnosemitskie yazyki* [Languages of the World. Semitic Languages. Akkadian Language. Northwest Semitic Languages], edited by L. Kogan and S. Loesov, 205–238. Moscow: Academia.
- Kogan, L.E. 2009. Khanaaneiskie yazyki [Canaanite Languages]. In *Yazyki mira. Semitskie yazyki. Akkadskii yazyk. Severozapadnosemitskie yazyki* [Languages of the World. Semitic Languages. Akkadian Language. Northwest Semitic Languages], edited by L. Kogan and S. Loesov, 239–278. Moscow: Academia.
- Kogan, L. 2015. *Genealogical Classification of Semitic: The Lexical Isoglosses*. Boston: Walter de Gruyter.
- Kozintsev, A.G. 2021. Aziia ili Afrika? O lokalizatsii afraziiskoi prarodiny [Asia or Africa? Localizing the Afroasiatic Homeland]. *Etnograficheskoe obozrenie* 4: 25–42. <https://doi.org/10.31857/S086954150016696-5>
- Militarev, A. 2000. Towards the Chronology of Afrasian (Afroasiatic) and its Daughter Families. In *Time Depth in Historical Linguistics*, edited by C. Renfrew, A. McMahon, and L. Trask, 1: 267–307. Cambridge: McDonald Institute for Archaeological Research.
- Militarev, A. 2004. Another Step towards the Chronology of Afrasian (I). In *Babel und Bibel 1: Ancient Near East, Old Testament and Semitic Studies*, edited

- by L. Kogan, N. Koslova, S. Loesov, and S. Tishchenko: 282–333. Moscow: Izdatel'stvo Ippolitova.
- Militarev, A. 2005. Once More About Glottochronology and Comparative Method: The Omotic-Afrasian Case. In *Aspekty komparativistiki = Aspects of Comparative Linguistics* 1, edited by I.S. Smirnov, 339–408. Moscow: RGGU, 2005.
- Militarev, A. 2006. Towards the Genetic Affiliation of Ongota, a Nearly Extinct Language of Ethiopia (II). In *Babel und Bibel* 3, edited by L. Kogan, N. Koslova, S. Loesov, and S. Tishchenko, 489–512. Moscow: RGGU.
- Militarev, A. 2010. A Complete Etymology-Based Hundred Wordlist of Semitic Updated: Items 1–34. *Voprosy yazykovogo rodstva = Journal of Language Relationship* 3: 43–78.
- Militarev, A. 2011. A Complete Etymology-Based Hundred Wordlist of Semitic Updated: Items 35–54. *Voprosy yazykovogo rodstva = Journal of Language Relationship* 5: 69–95.
- Militarev, A. 2012. A Complete Etymology-Based Hundred Wordlist of Semitic Updated: Items 55–74. *Voprosy yazykovogo rodstva = Journal of Language Relationship* 7: 71–104.
- Militarev, A. 2014. A Complete Etymology-Based Hundred Wordlist of Semitic Updated: Items 75–100. *Voprosy yazykovogo rodstva = Journal of Language Relationship* 11: 159–185.
- Militarev, A. 2015. Addenda and Conclusion of an Etymology-Based 100-Item Wordlist for Semitic Languages. *Voprosy yazykovogo rodstva = Journal of Language Relationship* 13 (1–2): 91–138.
- Orel, V., and O. Stolbova. 1995. *Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction*. Leiden: E.J. Brill.
- Savà, G., and M. Tosco. 2000. A Sketch of Ongota, a Dying Language of Southwest Ethiopia. *Studies in African Linguistics* 29 (2): 59–135. <https://doi.org/10.32473/sal.v29i2.107366>
- Shriner, D., et al. 2018. Genetic Ancestry of Hadza and Sandawe Peoples Reveals Ancient Population Structure in Africa. *Genome Biology and Evolution* 10 (3): 875–882.
- Shriner, D., and C.N. Rotimi. 2018. Genetic History of Chad. *American Journal of Physical Anthropology* 167 (4): 804–812. <https://doi.org/10.1002/ajpa.23711>
- Starostin, G.S. 2008. K voprosu o geneticheskoi prinadlezhnosti yazyka khadza [On the Genetic Affiliation of Hadza Language]. In *Afrikanskii sbornik 2007* [African Collection 2007], edited by V.F. Vydrin, 262–278. St. Petersburg: Nauka.
- Starostin, G.S. 2013. *Yazyki Afriki. Opyt postroeniia leksikostatisticheskoi klassifikatsii* [Languages of Africa. An Attempt at Constructing a Lexicostatistical Classification]. Vol. 1, Metodologiya. Koisanskii yazyki [Methodology; Khoisan Languages]. Moscow: Yazyki slavianskoi kul'tury.
- Stolbova, O.V. 2016. *Etimologicheskii slovar' chadskikh yazykov* [An Etymological Dictionary of Chad Languages]. Moscow: Institut vostokovedeniia RAN. <https://book.ivran.ru/f/ilovepdfmerged.pdf>
- Theil, R. 2012. Omotic. In *Semitic and Afroasiatic: Challenges and Opportunities*, edited by L. Edzard, 369–384. Wiesbaden: Harrassowitz Verlag.
- Tishkoff, S.A., et al. 2009. The Genetic Structure and History of Africans and African Americans. *Science* 324 (5930): 1035–1044.