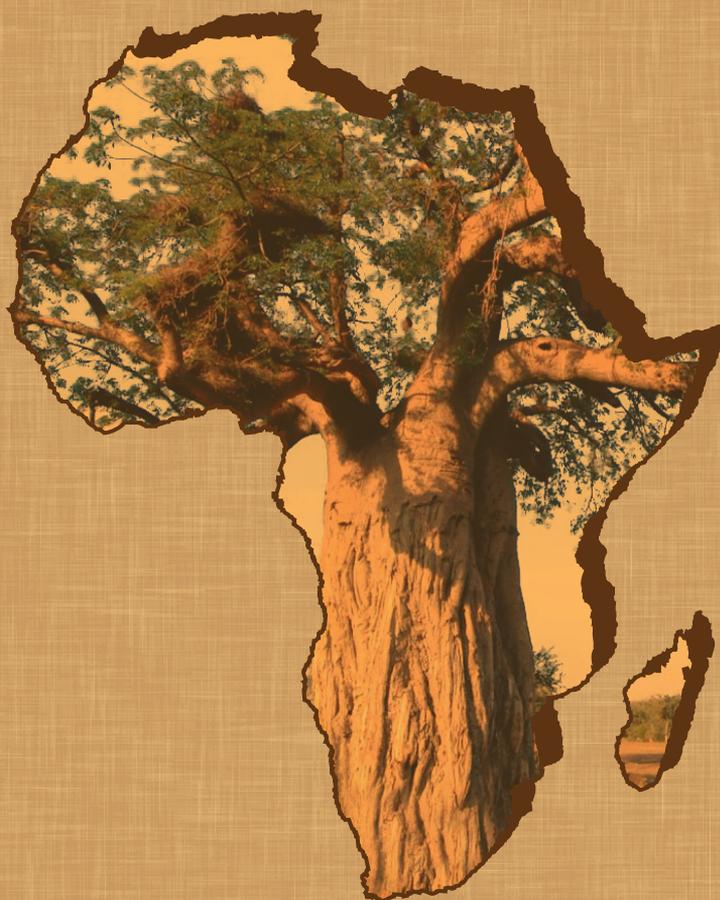


Г. С. СТАРОСТИН

ЯЗЫКИ АФРИКИ

**ОПЫТ ПОСТРОЕНИЯ
ЛЕКСИКОСТАТИСТИЧЕСКОЙ
КЛАССИФИКАЦИИ**



Том I

Методология. Койсанские языки

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ

Г. С. СТАРОСТИН

ЯЗЫКИ АФРИКИ

ОПЫТ ПОСТРОЕНИЯ
ЛЕКСИКОСТАТИСТИЧЕСКОЙ
КЛАССИФИКАЦИИ

Том I

Методология. Койсанские языки



ЯЗЫКИ СЛАВЯНСКОЙ КУЛЬТУРЫ
МОСКВА 2013

УДК 811
ББК 81.2
С 77

Рецензенты

*Доктор филологических наук, зав. отделом этнографии народов Африки
Музея антропологии и этнографии им. Петра Великого*

В. Ф. Выдрин

*Доктор филологических наук, профессор, член-корреспондент РАН,
главный научный сотрудник Института языкознания РАН*

А. В. Дыбо

Старостин Г. С.

С 77 Языки Африки. Опыт построения лексикостатистической классификации. Т. 1: Методология. Койсанские языки. — М.: Языки славянской культуры, 2013. — 510 с.

ISBN 978-5-9551-0621-2

Книга представляет собой первый том масштабного исследования по созданию новой рабочей модели генетической классификации языков и языковых семей африканского континента, которая могла бы представить серьезную альтернативу для т. н. «стандартной модели» классификации африканских языков, разработанной Дж. Гринбергом более чем столетия тому назад и с тех пор неоднократно подвергавшейся критике за недостаточную основательность.

В первый том исследования вошла вводная часть — подробное описание методологии построения классификации, в которой синтезированы элементы классического сравнительно-исторического метода, «многостороннего сравнения» Гринберга и квантитативный подход к языковому материалу; в основе классификации лежит лексикостатистический анализ данных базисной лексики, подкрепленный тщательной этимологической обработкой. Вторая часть первого тома апробирует описанную методику на материале самой малочисленной и, во многих отношениях, «загадочной» из гипотетических макросемей Гринберга — койсанской (бушменско-готтентотской).

Книга предназначена для внимания специалистов по общему, сравнительно-историческому и типологическому языкознанию; африканистов самых различных профилей; и широкого круга читателей, в той или иной степени интересующихся теоретическими, методологическими и практическими аспектами реконструкции лингвистических аспектов предистории человечества.

ББК 81.2

© Г. С. Старостин, 2013

ISBN 978-5-9551-0621-2 © Издательство «Языки славянской культуры», 2013

СО Д Е Р Ж А Н И Е

Предисловие.....	5
------------------	---

Часть I. Методология исследования

1.1. Общее описание задачи. Различные подходы к проблеме обоснования языкового родства	23
1.2. Критерии обоснования языкового родства и проблема «корректирующего фактора»: морфология и лексическая устойчивость	30
1.3. «Слабое» и «сильное» обоснование языкового родства ..	44
1.4. Фонетическая совместимость и фонетический изоморфизм	57
1.5. «Массовое сравнение» и значимость дистрибуционного фактора	65
1.6. Лексикостатистика: к проблеме оптимизации метода ...	79
1.7. Общее описание процедурных стадий исследования	136
1.8. Специфика африканского ареала в плане установления генетического родства между его языками.....	203

Приложение 1. Унифицированная система транскрипции..	254
---	-----

Приложение 2. Сокращения названий языков и другие аббревиатуры	262
---	-----

Приложение 3. Краткий семантический комментарий к элементам 50-словного списка	264
---	-----

Часть II. Койсанские языки.

Общий обзор	301
-------------------	-----

2.1. Севернокойсанская группа	
2.1.1. Общие сведения и источники	310
2.1.2. Историческая характеристика	313
2.1.3. Лексикостатистика	321
50-словный список для севернокойсанских языков	322

2.2. Язык восточный ꠘхоан	
2.2.1. Общие сведения и источники	333
2.2.2. Историческая характеристика	335
50-словный список для языка восточный ꠘхоан	337
2.3. Южнокойсанская группа	
2.3.1. Общие сведения и источники	339
2.3.2. Историческая характеристика	346
2.3.3. Лексикостатистика	353
50-словный список для южнокойсанских языков	355
2.4. Центральнокойсанская группа	
2.4.1. Общие сведения и источники	391
2.4.2. Историческая характеристика	399
2.4.3. Лексикостатистика	406
50-словный список для центральнокойсанских языков ..	409
2.5. Язык квади	
2.5.1. Общие сведения и источники	436
2.5.2. Историческая характеристика	437
50-словный список для языка квади	438
2.6. Язык сандаве	
2.6.1. Общие сведения и источники	440
2.6.2. Историческая характеристика	441
50-словный список для языка сандаве	443
2.7. Язык хадза	
2.7.1. Общие сведения и источники	447
2.7.2. Историческая характеристика	448
50-словный список для языка хадза	450
2.8. Предварительные выводы	456
Приложение 1. Генеалогические деревья	476
Приложение 2. Сопоставительные таблицы	479
Литература	484
Summary	506

Предисловие

Общая идея исследования, предлагаемого вниманию читателя, складывалась у автора постепенно. Все началось примерно с того момента, когда, в качестве участника международного проекта «Эволюция языка»¹, он взял на себя нелегкую обязанность выявить и/или уточнить исторические связи между языками одной из наиболее загадочных языковых общностей планеты — койсанской (бушменско-готтентотской), а также, по возможности, установить наличие или отсутствие таких связей между койсанскими языками и другими языковыми семьями (как на территории самой Африки, так и других континентов).

Работа, начавшаяся с более или менее тривиальных процедур (сбор опубликованных данных, расписывание словарных материалов, составление лексикостатистических списков и т. п.), довольно быстро столкнулась с проблемами гораздо более фундаментального характера. Относительная скудность данных, доступных для изучения; типологическая уникальность фонетики койсанских языков — а вместе с ней и уникальность фонетических изменений, препятствующая уверенному обнаружению регулярных соответствий; сложнейшая система ареальных отношений, в рамках которой почти безнадежно перемешаны друг с другом гене-

¹ Evolution of Human Languages (сокращенно EHL); совместными руководителями проекта на то время (2001 г.) были ныне покойный глава Московской школы компаративистики С. А. Старостин (1953-2005) и лауреат Нобелевской премии по физике, соучредитель международного исследовательского Института Санта-Фе и, по совместительству, глубокий знаток исторического языкознания Мюррей Гелл-Манн. Основная цель проекта — создание единой генеалогической классификации языков мира, базирующейся на совмещении традиционной методологии сравнительно-исторического языкознания, формальных статистических методов и междисциплинарных подходов. По состоянию на 2012 г. руководство проектом на совместных основаниях осуществляют М. Гелл-Манн, Г. С. Старостин и И. И. Пейрос.

тические и контактные сходства соседних языков и языковых групп — все это в конечном итоге дало ясно понять, что единым сконцентрированным «ударом», даже если при этом учитывать весь гигантский опыт этимологических исследований по бесписьменным языкам, накопленный в рамках Московской школы компаративистики, пробить «койсанскую стену» не удастся.

Помимо прочего, стало очевидно, что дать ответ на вопрос «входят ли все т. н. койсанские языки в одну и ту же языковую семью?» принципиально невозможно до тех пор, пока соответствующие языки не будут сопоставлены — на базе единого, строго унифицированного стандарта — со всеми остальными крупными языковыми таксонами африканского континента. Интуитивно любому исследователю, бегло ознакомившемуся с основными типологическими характеристиками и сколь-либо представительной лексической выборкой по языкам Африки, может легко показаться, что на территории этого материка представлено два языковых типа — «койсанский» и «не-койсанский», настолько языки бушменов и готтентотов оказываются непохожими на все остальные. Но действительно ли это впечатление отражает историческую реальность? Подтверждается ли оно возможностью построить конкретный, достоверный, непротиворечивый сценарий развития от «общафриканского» (а, может быть, и «общемирового») языка-предка к современным потомкам? Получить удовлетворительный ответ на эти вопросы без длительной, кропотливой работы над материалом, отобранном по всем без исключения языковым семьям Африки, невозможно.

Единственный на сегодняшний момент опыт такой работы принадлежит одному из самых уникальных лингвистов XX в., едва ли имевшему себе равных по охвату кругозора — Джозефу Гринбергу (1915-2001). Общая классификация языков Африки, разработанная им в 1940-е—1960-е гг., до сих пор пользуется огромной популярностью, даже несмотря на то, что критические замечания в ее адрес со стороны лингвистов-африканистов с каждым годом становятся все громче. По мере того, как наши знания о языках Африки постепенно возрастают благодаря новым полевым исследованиям, а сравнительно-исторический метод начинает применяться к доселе незатронутым мелким языковым семьям континента, все отчетливее выходят на поверхность многочисленные не-

достатки в классификации Гринберга: использование ошибочных и недостоверных данных, размытые критерии фонетического и семантического сходства, недостаточный учет контактных отношений между соседними языками и т. п. На сегодня этого достаточно, чтобы серьезно *усомниться* в истинности модели, предложенной Гринбергом. При этом, однако, никому до сих пор не удалось эксплицитно продемонстрировать *ошибочность* полученных им результатов.

Ситуация оказывается в каком-то смысле парадоксальной: с одной стороны, большинство современных лингвистов категорически отвергает метод «массового сравнения», с помощью которого создавалась африканская и другие классификации Гринберга, с другой — вынуждено, тем не менее, опираться на эту классификацию как на сомнительную, но все же рабочую модель, в силу полного отсутствия альтернативных схем. Единственной альтернативой оказывается негативная — полный отказ от всех классификационных гипотез, за которыми стоят Гринберг и «массовое сравнение», т. е. своего рода «откат» от разработанной им модели четырех африканских макросемей к состоянию, в котором африканское языкознание пребывало в первой половине XX в., когда оно оперировало разрозненным набором из сотни с лишним языковых групп, родственные связи между которыми оставались неопределенными.

Возможно, что эта альтернатива с воодушевлением будет поддержана т. н. «сплиттерами» — специалистами, предпочитающими по возможности оперировать мелкими таксономическими единицами, избегая попыток сведения их к более крупным, не поддержанным строжайшими логическими и статистическими доказательствами. Но даже наиболее последовательные «сплиттеры» вряд ли станут утверждать, что присутствие на территории Африки нескольких десятков языковых семей, *никак* не связанных друг с другом отношениями генетического родства — правдоподобный исторический сценарий. Речь идет скорее о том, что соответствующие отношения могут восходить к такой глубокой древности, что ни «массовое сравнение» Гринберга, ни даже классический сравнительно-исторический метод уже не позволят восстановить их единственно правильным способом.

Основную идею нашего исследования можно вкратце изложить следующим образом. Генетическая классификация, в отличие от типологической, в обязательном порядке предполагает наличие исторической *реконструкции*: либо эксплицитной (как в традиционной компаративистике), либо подразумеваемой (как в «массовом сравнении» Гринберга, которое формально избегает реконструкции, но на практике, разумеется, предполагает существование формы-«предка» для каждого сопоставления). Как известно, любая историческая реконструкция — аппроксимация; реконструкция не может претендовать на отражение абсолютной истины. Однако и аппроксимации могут существенным образом различаться по качеству. Реконструкция, которая целиком и полностью оказывается *внутренне непротиворечивой* (т. е. не предполагает протекания взаимоисключающих исторических изменений), *типологически правдоподобной* (т. е. не постулирует таких конкретных исторических процессов, которые бы открыто противоречили накопленной теоретической и эмпирической базе исторического языкознания) и *системной* (т. е. не представляет собой набора разрозненных, несводимых к единой структуре постулатов), может считаться такой аппроксимацией, «зазор» между которой и истинным историческим процессом ничтожно минимален и, следовательно, пренебрежим. Данными такой реконструкции можно пользоваться примерно с такой же степенью доверия, как и, например, данными письменных исторических источников (в отдельных случаях — даже с бóльшим доверием, т. к. проблему оценки степени фактологической достоверности исторических источников пока что никто не отменял).

Главный недостаток исторического сценария, в свое время предложенного Гринбергом, заключается в том, что он не удовлетворяет ни одному из вышеперечисленных критериев: любой, даже самый поверхностный, анализ выявляет в нем и внутренние противоречия, и неправдоподобность, и бессистемность. При этом, однако, никто из критиков Гринберга не доказал, что все эти проблемы по той или иной причине должны быть имманентно присущи любому макрокомпаративному исследованию (откуда следовало бы, что любые занятия «дальним родством» языков заведомо бесперспективны). Наоборот, скорее можно сказать, что все они вызваны техническими причинами — такими, как скуд-

ность и некачественность конкретных языковых данных, которые находились в распоряжении Гринберга; недостаточная разработанность на тот момент типологии фонетических и семантических изменений; поверхностный подход к сопоставлению материала плохо изученных семей, сознательно игнорировавший сложные вопросы исторической реконструкции на мелких хронологических уровнях и т. п.

Для того, чтобы выйти из сложившейся тупиковой ситуации, на наш взгляд, основной задачей следует сделать построение *оптимального исторического сценария* — такой схемы развития, которая наименее противоречивым, наиболее системно-связанным и наиболее экономным образом объясняла бы те сходства и различия, которые можно наблюдать на языковых уровнях, имеющих первичное значение для генетической классификации (базисная лексика и грамматика). Основное отличие такого подхода от методики Гринберга заключается в том, что оптимальность сценария не может быть удостоверена простым нагромождением сравнительного материала; за каждым сопоставлением в рамках такого сценария предполагается наличие конкретного объяснения *причин* не только фонетических и семантических *сходств*, но и *различий* между членами сопоставления.

Иными словами, для каждой морфемы (лексемы) А из языка L_A , сравниваемой с морфемой (лексемой) В из языка L_B , необходимо иметь возможность не только сказать, что «А и В обладают значительным фонетическим и семантическим сходством и, следовательно, могут быть использованы как аргумент в пользу генетического родства L_A и L_B », но и дополнить это утверждение следующим образом: «Существует такой-то типологически достоверный сценарий — или определенное множество таких-то примерно равновероятных типологически достоверных сценариев — того, каким образом формы А и В сумели развиться из единого источника, и этот сценарий непротиворечивым образом вписывается в общую систему возможных правил развития, постулируемую для ближайшего общего предка языков L_A и L_B ».

Разумеется, невозможно утверждать, что для любой, произвольно взятой, ситуации в том или ином лингвистическом ареале (будь то африканский континент или любая другая зона), всегда удастся установить единую генеалогическую классификацию для

всех ее языков, удовлетворяющую строжайшим критериям сравнительно-исторического языкознания — полный набор регулярных фонетических соответствий; стройная, доказательная система парадигматических изоморфизмов в грамматике и т. п. Но нет ничего безрассудного или предосудительного в утверждении, что для каждой такой ситуации можно по меньшей мере сконструировать определенный выше оптимальный сценарий.

Одним из основных недостатков большого числа работ по сравнительно-историческому языкознанию является своего рода «наивный максимализм» — тенденция к неразличению между правдоподобными, но не строго доказательными гипотезами, и гипотезами заведомо неправдоподобными и недоказуемыми. В рамках этой тенденции, например, ностратическая гипотеза В. М. Иллич-Свитыча и яфетическая «теория» Н. Я. Марра могут упоминаться в одном и том же контексте, несмотря на кардинальные различия в подходах этих исследователей (ностратическая гипотеза как минимум удовлетворяет критерию фальсифицируемости, в отличие от принципиально нефальсифицируемой «яфетологии»).

Идея оптимального сценария, на который можно будет опираться как на рабочую модель, предполагает отказ от такой практики. На самом деле для каждой гипотезы языкового родства возможна *относительная* оценка, представляющая собой подробный ответ на такие вопросы, как: «содержатся ли в данной гипотезе утверждения, противоречащие теоретическому аппарату и/или накопленной эмпирической базе исторического языкознания?»; «существуют ли гипотезы родства, *альтернативные* предложенной и при этом обладающие большей объяснительной силой?»; «будет ли альтернативная интерпретация представленных аргументов — например, как следов языковых контактов или случайных совпадений — более экономной?»; «будет ли она лучше согласовываться с представлениями о том, как, по каким причинам и с какой скоростью протекают языковые изменения?» и т. п.

Важнейшая техническая проблема, стоящая на пути такой оценки — общий *объем* языкового материала. Поскольку два и более языков, в конечном итоге восходящих к общему предку, могут сохранять следы исконного родства на всех языковых уровнях и во всех подсистемах всех инвентарей этих уровней, конкуриру-

ющие гипотезы языкового родства можно предлагать на совершенно разных основаниях. Для одного исследователя предпочтительным материалом сравнения окажутся грамматические морфемы, для другого — порядок слов, для третьего — накопленный этимологический корпус из двух-трех тысяч лексических корней и т. п. В таких условиях проведение сравнительной оценки гипотез невозможно, т. к. подобная оценка может быть дана лишь исходя из унифицированной системы критериев.

Именно поэтому, на наш взгляд, при построении первичной рабочей модели генетической классификации языков того или иного ареала необходимо сознательно ограничить себя определенным «срезом» языкового материала, который, с одной стороны, имеет ключевое значение для определения генетического статуса исследуемых языков (содержит большое количество т. н. «генетических маркеров»), с другой — является *имманентным* всем без исключения языкам ареала (в идеале — всем без исключения языкам *мира*, т. к. «лингвистический ареал» — понятие сугубо расплывчатое и относительное). На дальнейших этапах работы можно и нужно будет расширять аргументацию за счет привлечения дополнительного материала, но это уже будет своего рода «надстройкой», т. е. продвинутым компаративным исследованием в рамках разработанной рабочей модели.

В качестве такого «среза» мы, придерживаясь методики С. А. Старостина, выбираем т. н. «базисную лексику», а в качестве основного квантитативного метода, работающего на этом срезе — лексикостатистический анализ, первоначально разработанный М. Сводешем, в дальнейшем значительно усовершенствованный в трудах С. А. Старостина, и на сегодняшний день прошедший успешное тестирование на материале самых разных языковых семей мира в рамках проекта Московской школы компаративистики «Глобальная лексикостатистическая база данных». Подробное обоснование того, почему основной акцент рекомендуется изначально делать именно на базисной лексике (а не, например, на грамматике), будет дано ниже, в первой (методологической) части исследования. Здесь отметим лишь, что конкретно в исторической африканистике как лексические сопоставления, так и их лексикостатистический анализ до сих пор пользуются большой популярностью при разработке классификационных гипотез, так

что в этом отношении наше исследование не носит принципиально новаторский характер.

Главное отличие стоящей перед нами задачи от предыдущих работ по африканской (и не только) лексикостатистике можно кратко (ниже об этом будет говоритья гораздо подробнее) охарактеризовать следующим образом. Согласно достаточно популярному, и весьма неприятному, заблуждению, *лексикостатистика*, понимаемая как «сопоставление совпадающих по значению и сходных по форме элементов заданного словарного списка», и *сравнительно-исторический метод*, понимаемый как «установление регулярных фонетических соответствий на материале генетически родственных морфем», противопоставлены друг другу как два альтернативных, во многом не связанных друг с другом, метода установления языкового родства; при этом подразумевается, что «серьезный» лингвист в этой ситуации должен отдавать неизменное предпочтение сравнительно-историческому методу как более строгому и доказательному. Более мягкий подход предполагает, что лексикостатистический анализ должен все-таки опираться на результаты сравнительно-исторической обработки материала (т. е. слова в заданном списке должны сопоставляться не на основании фонетического сходства, а на основании ранее установленных фонетических соответствий), но при этом все равно может оставаться «механистическим», т. е., кроме наличия или отсутствия между сравниваемыми словами фонетических и семантических изоморфизмов, не учитывать никаких других деталей — например, информации о дистрибуции форм в сравниваемых языках и т. п.

На самом деле, однако, лексикостатистика и традиционное сравнительно-историческое языкознание никоим образом не являются «антиподами»; напротив, они предельно логично и стройно дополняют друг друга. Лексикостатистика, не опирающаяся на всестороннюю поддержку со стороны теоретическо-эмпирической базы сравнительно-исторического языкознания, будет оставаться не более чем «мертвой» числовой матрицей, которую невозможно достоверно спроецировать на конкретный исторический сценарий. Но, в свою очередь, и многие гипотезы, предложенные (якобы) на базе применения сравнительно-исторического метода, будут оставаться неубедительными «пустышками», если

их серьезность не будет продемонстрирована в рамках единого, унифицированного и наглядного количественного метода — наиболее эффективным и удобным для применения из которых на сегодня и является лексикостатистика. Не секрет, что за последние несколько десятков лет накопилось огромное количество работ по языковому родству (в том числе и африканскому), авторы которых неизменно клянутся в верности сравнительно-историческому методу — но из-за того, что конкретное практическое понимание основных постулатов метода у разных исследователей различается, далеко не все эти работы имеют истинную научную ценность, и лексикостатистика здесь оказывается той самой «лакомусовой бумажкой», которая позволяет отделять перспективные гипотезы от фантазий.

Что касается Африки как условного лингвистического макроареала, на котором можно опробовать предлагаемый подход, то, на наш взгляд, африканский континент — одна из наиболее удобных площадок для тестирования его эффективности, по целому ряду причин, как-то:

1) Африка отличается колоссальным языковым разнообразием, что, впрочем, неудивительно, учитывая, что именно этот континент сегодня уверенно считается прародиной человечества. С другой стороны, для Африки типичными являются и ситуации длительного сосуществования в пределах достаточно замкнутого географического пространства носителей разных языковых семей, что обычно является стимулом к конвергенции. Иными словами, языковые отношения в Африке представляют собой сложнейшую сеть генетических и контактных изоморфизмов. Если выбранная нами методика «этимологической лексикостатистики» сумеет распутать эту сеть, это будет означать, что она, скорее всего, применима и к любому другому языковому ареалу планеты.

2) В статистическом плане, за редкими исключениями, языки Африки — бесписьменные; традиция их изучения лишь в небольшой группе случаев восходит к XIX в. или более ранним периодам. Это означает, что почти все выводы исторического характера можно делать лишь на основании сравнения современных языковых форм, что значительно осложняет работу компаративиста по сравнению, например, с такими крупными языковыми семьями Евразии, как индоевропейская, тюркская, дравидийская, сино-

тибетская и др. Тем важнее, однако, убедиться в том, что «этимологическая лексикостатистика» *не требует* обязательного наличия языков с известной историей для того, чтобы выдавать исторически достоверные результаты. Следует помнить, что ситуации, аналогичные индоевропейской (где для многих современных языков надежно зафиксированы их прямые или «боковые» предки двух-трехтысячелетней давности), в общемировом масштабе представляют собой *исключение*, а не *норму*, и при разработке общих стандартов сравнительно-исторического языкознания в первую очередь нужно учитывать их эффективность применительно к языкам «без истории».

3) С другой стороны, в плане общей изученности ситуация в Африке принципиально отличается, например, от ситуации в таких регионах, как Юго-Восточная Азия (где к систематическому описанию мелких бесписьменных языков лингвисты приступили лишь совсем недавно) и Новая Гвинея (где, судя по недавно проведенной в рамках проекта «Эволюция языка» инвентаризации, накопленные лингвистические данные не только не всеобъемлющи, но даже с большим трудом могут считаться репрезентативными). На сегодняшний день почти для каждой из многочисленных мелких языковых групп Африки доступны по меньшей мере несколько кратких лексических списков; отдельных языков и диалектов, по которым не опубликовано никаких данных, остается еще очень много, но большинство таких языков, согласно предварительной оценке специалистов, уверенно определяются как представители конкретных языковых таксонов мелкого уровня — например, на основании краткого анкетирования носителей, свидетельств соседних племен и т. п. — по *другим* языкам которых опубликованные данные все же имеются. Разумеется, в отдельных случаях такие определения могут оказаться ошибочными, но в целом можно с уверенностью утверждать, что сегодня лексические и грамматические сведения по языкам Африки, находящиеся в нашем распоряжении, достаточно репрезентативны, чтобы полученной на их основании генетической классификации можно было присвоить статус «общеафриканской».

Мы отдаем себе полный отчет в том, что финальная цель, обозначенная в исследовании, предполагает огромный объем работы даже при выполнении тех условий, которые будут подробно опи-

саны в первом, методологическом разделе — определения и обоснования *минимального* объема материала, необходимого и достаточного для фиксации любого языка в конкретной точке генеалогического древа. Для того, чтобы такая работа могла считаться шагом вперед по сравнению со спорными, но, тем не менее, значительными и все еще актуальными результатами Дж. Гринберга, она должна обладать не только и не столько теоретическими, сколько «техническими» преимуществами, такими, как:

— максимально внимательный и аккуратный подход к записи, анализу и интерпретации языкового материала, строго учитывающий в том числе различия между высоко- и малодостоверными источниками;

— максимально детализированная и логически последовательная *экспликация* механизмов сравнения, реконструкции и классификации; необходимо стремиться к тому, чтобы как можно меньше выводов, в том числе «мелких», «промежуточных», подавались как «самоочевидные».

Второе из этих требований стоит, пожалуй, подчеркнуть особо. Мы твердо убеждены в том, что именно такого рода пропуск логических звеньев в цепочках этимологических рассуждений, иногда «доброжелательный» (когда исследователь действительно считает соответствующие звенья самоочевидными), иногда «злонамеренный» (когда отсутствие эксплицитной аргументации призвано замаскировать нежелание или фактическую невозможность справиться с реальной проблемой), является одним из ключевых препятствий к дальнейшему развитию сравнительно-исторического языкознания — не говоря уже о том, что искушение опустить подробную аргументацию часто приводит к губительным недоразумениям, возникающим между исследователем-компаративистом и специалистами «со стороны», на суд которых выносятся результаты исследования. Этот же недостаток способствует и созданию превратного впечатления об этимологии как о всецело субъективной области, где личные пристрастия и индивидуальная интуиция автора якобы играют не меньшую роль, чем строгий логический анализ данных в рамках объективно выработанных методологических стандартов.

Конечно, чаще всего ситуации, в которых читателю приходится самостоятельно «домысливать» отдельные звенья в цепи аргу-

ментации автора, обусловлены техническими причинами — например, соображениями объема; так, произвольно выбранный этимологический словарь размером в 1000 страниц при подробном разборе каждой этимологии легко рискует разрастись до 10,000. Но нужно понимать, что, экономя на объеме, мы рискуем сэкономить и на *степени убедительности* предъявленных этимологий, реконструкций и классификаций. То, что Дж. Гринбергу удалось уместить финальный вариант своей африканской лингвистической модели в 180 страниц текста (вместе с методологическим введением, индексами и библиографией) — достижение, безусловно, уникальное, но на самом деле каждая из страниц его монографии, содержащей конкретный языковой материал, подразумевает имплицитное присутствие подробного комментария объемом не менее чем в дополнительные 10-20 страниц (увы, большая часть этих комментариев так и осталась в голове Гринберга, в лучшем случае — на страницах его до сих пор неопубликованных записных книжек).

Именно поэтому в рамках нашего исследования мы *намеренно* не ставим своей целью краткость и сжатость изложения, особенно в тех его разделах, которые посвящены этимологическому анализу базисной лексики конкретных языковых групп. Это не только предоставит возможность легче и детальнее оценить ход рассуждений и выводы автора и выявить проникшие в них ошибки и неточности, но и облегчит чтение отдельных частей исследования для лингвистов, основной род деятельности которых не связан непосредственно с африканистикой или даже сравнительно-историческим языкознанием (смеем надеяться, что работа вызовет определенный интерес не только у специалистов узкого профиля, но и у значительно более широкого круга читателей).

Оборотная сторона такого решения — невозможность уместить все результаты проделанной работы в пределах одного, даже весьма объемного, тома. Идеальная структура такого исследования видится нам следующим образом. Первые разделы должны быть посвящены «первичному» анализу базисного лексического материала по конкретным макросемьям Африки, выделенным Дж. Гринбергом, в следующей последовательности: (а) сугубо гипотетическая «койсанская» макросемья; (б) не менее гипотетическая «нило-сахарская» макросемья; (в) несколько более строго

обоснованная, но все равно страдающая от нечеткой определенности своих границ, «нигер-кордофанская» макросемья; (г) более или менее общепризнанная, но все же не полностью свободная от проблем таксономического характера афразийская макросемья. Основная цель такого первичного анализа — своего рода «за-чистка» лексического материала, конечный результат которой — реконструкция сравнительных 50-словных списков базисной лексики для всех «мелких» праязыков в пределах соответствующих макросемей (т. е. праформ, привязанных к хронологическому порогу не более чем в 3,000-3,500 лет).

Эта часть исследования — по сути, сводящая все разнообразие из двух с лишним тысяч языков Африки к паре сотен реконструированных лексических списков — является наименее «интересной» с точки зрения макрокомпаративиста, в первую очередь заинтересованного в верификации модели Гринберга, и, по-видимому, одновременно наиболее трудозатратной; однако без тщательного ее выполнения нет никакой возможности перейти ко второй части, т. е. построению единой классификационной модели, которая, в свою очередь, призвана свести полученную пару сотен «генеалогических узлов» к минимально приемлемому числу узлов верхних уровней, подтверждая или опровергая тем самым классификацию Гринберга.

В состав первого тома исследования включены теоретически-методологическая вводная часть (абсолютно необходимая для правильного истолкования общей идеологии исследования) и апробирование предложенной методики на материале одной из гипотетических макросемей Гринберга — койсанской (с которой, собственно говоря, и началась деятельность автора на ниве африканистики). (Параллельно с этим к публикации уже готовится второй том — по «восточносуданской» семье, одному из нескольких крупнейших компонентов, которые были объединены Дж. Гринбергом в «нило-сахарскую» макросемью).

В заключение нельзя не упомянуть, что объемность и сложность поставленной задачи неизбежно повлекут за собой ряд конкретных недостатков. Настолько, насколько это возможно, мы повсюду стараемся избегать противоречивого и двусмысленного употребления терминологии (за исключением специально оговоренных случаев); использования устаревших и недостоверных

материалов (за исключением ситуаций, когда такие материалы являются единственным источником сведений по тем или иным языкам); ошибок в фонологическом и морфологическом анализе лексики; неточностей и опечаток в транслитерации материала и в самом тексте исследования и т. п. Ответственность за все подобного рода огрехи, которые по тем или иным причинам так и не удалось устранить в финальной редакции, несет исключительно автор работы; хочется надеяться, что существенного влияния на корректность основных выводов они иметь не будут.

Наконец, необходимо отметить, что, параллельно с представлениями результатов исследования в форме связного текста, они также постепенно оформляются в виде лексикостатистических баз данных, снабженных подробными синхронными и историческими аннотациями, и выкладываются в открытый доступ на сайте «Глобальная лексикостатистическая база данных» / «Global Lexicostatistical Database» (дочернем проекте упоминавшейся выше программы «Эволюция языка»). В этих базах содержится детальная информация не только по 50-словным лексическим спискам, представляющим собой основной объект исследования в рамках данной работы, но и по полным 100-словникам, включая точные указания на конкретный источник (вплоть до номера страницы) для каждой из использованных форм, которые в текст работы включены не были. В отличие от окончательного текста исследования, базы данных подвергаются постоянной редакции и корректировке, так что все недочеты и исследовательские ошибки, обнаруженные в тексте, будут в первую очередь исправляться в системе баз данных, призванной содержать самые свежие результаты лексикостатистической и этимологической обработки базисной лексики по языкам как Африки, так и других лингвистических ареалов планеты.

Автор приносит глубокую благодарность своим коллегам — африканистам, компаративистам, лингвистам общего профиля — за многочисленные ценные замечания содержательного и технического характера, высказанные ими как в рамках бесед и консультаций, проводившихся в ходе работы над монографией, так и по итогам чтения исходного текста рукописи. Речь идет в первую очередь об отечественных специалистах, имевших возможность детально ознакомиться с текстом (К. В. Бабаев, С. А. Бурлак,

В. Ф. Выдрин, А. В. Дыбо, В. А. Дыбо, А. С. Касьян, А. Ю. Милитарев, О. А. Мудрак, С. Л. Николаев, И. И. Пейрос, В. А. Плунгян, К. И. Поздняков, О. В. Столбова и другие), а также о зарубежных специалистах, оказавших автору незаменимую поддержку в ходе сбора языковых данных и различных консультаций: это в первую очередь Р. Бленч, К. Коллинз, Н. Кроухолл, Б. Сэндс, Р. Фоссен, Х. Хаммарстрём, Г. Хонкен (последнему, к сожалению, благодарность выражается посмертно) и др.

Особо следует упомянуть С. А. Старостина (1953-2005) — отца и учителя, памяти которого в какой-то степени посвящена данная монография, так как без его научного руководства на протяжении десяти лет работы написание ее было бы абсолютно невозможным. Надеемся, что описанная в ней методика историко-лингвистического исследования языкового материала с попыткой выйти на «глубинный» хронологический уровень может рассматриваться как естественное дальнейшее развитие тех идей и методов, которые он сам постоянно разрабатывал и совершенствовал на протяжении более чем тридцати лет работы.

Отдельную благодарность хотелось бы выразить лицам и организациям, уже много лет подряд поддерживающим исследовательские программы «Вавилонская башня» и «Эволюция языка», в рамках которых автором была проделана основная работа над языками Африки. Это в первую очередь Центр компаративистики Института восточных культур и античности РГГУ (директор — В. А. Дыбо); Институт Ближнего Востока (президент — Е. Я. Сатановский); Институт Санта-Фе (соучредитель и соруководитель программы «Эволюция языка» — М. Гелл-Манн); а также Фонд фундаментальных лингвистических исследований, оказавший существенную поддержку в публикации монографии. Кроме того, работа была выполнена при поддержке отечественных фондов РГНФ (проект №12-04-00293) и РФФИ (проект №12-6-640-3) и Программы фундаментальных исследований Секции языка и литературы ОИФН РАН «Язык и литература в контексте культурной динамики».

Часть I

Методология исследования

1.1. Общее описание задачи. Различные подходы к проблеме обоснования языкового родства.

Генеалогическая классификация языков коренного населения африканского континента — задача, неразрывно связанная с общим вопросом научного изучения языков Африки в их историческом аспекте, т. е. анализа имеющихся в нашем распоряжении данных по этим языкам с помощью сравнительно-исторического метода. В идеале такой анализ должен содержать все традиционно предписанные для него стадии, как-то: (а) построение предварительных гипотез о генетическом родстве тех или иных языков на основании обнаруженных между ними фонетических, лексических и грамматических сходжений; (б) установление между потенциально родственными языками системы фонологических соответствий; (в) реконструкция лексического и грамматического инвентаря праязыка; (г) рекурсивное применение предыдущих стадий для обнаружения и доказательства более глубокого родства, где материалом для сравнения служат уже не отдельные языки, а целые группы языков с реконструированными для них прото-состояниями¹.

Вопрос о том, какие из этих стадий работы лингвиста-компаративиста необходимы и достаточны для убедительного обоснования той или иной модели генеалогической классификации, остается открытым. Вынося за рамки данной работы теоретическую проблему того, насколько «древесную» модель классифика-

¹ Общая методология сравнительно-исторического языкознания в пунктах (а-в) в целом остается неизменной с начала «младограмматической» эпохи в компаративистике и вплоть до нынешнего времени; из последних по времени учебных изданий, подробно излагающих основы метода, см. [Campbell 2004; Hock & Joseph 1996]. Последний пункт — значимость промежуточных реконструкций для восстановления еще более древних языковых состояний — особенно важен для макрокомпаративистики; подробное обоснование см. в [Бурлак & С. Старостин 2005]. Конкретные практические рекомендации по стадиям работы компаративиста и их последовательности см. в [Gell-Mann, Peiros, & Starostin 2009].

ции языков вообще можно считать адекватной с точки зрения отображения реально имевших место исторических процессов¹, можно четко обозначить две полярные точки зрения на этот вопрос. Общим для обеих является лишь собственно понимание термина «генеалогическая классификация» как, по сути, набора иерархизированных ответов на вопросы (а) «восходят ли языки $x_1, x_2... x_n$ к единому общему предку?» (вопрос *доказательства языкового родства*) и (б) «при условии верности (а), какие из возможных подмножеств множества языков $x_1, x_2... x_n$ имеют промежуточных общих предков?» (вопрос построения *внутренней классификации языковой семьи*).

Первую точку зрения можно условно назвать «гиперконсервативной». Согласно представлениям ее сторонников, доказательство генетического родства между группой языков и, следовательно, генеалогическая классификация этой группы возможны лишь при условии максимально подробной обработки языковых данных посредством сравнительно-исторического метода, включая каждую из перечисленных выше стадий; без детальной системы фонетических соответствий, этимологического корпуса, на котором определена эта система, а также наличия общей (желательно парадигматической) морфологии родство языков считается недоказанным, а генеалогическая классификация их —

¹ Проблемам достоинств и недостатков традиционных «древесных» классификаций по сравнению с все более популярными сегодня «сетевыми» классификациями («networks»), которые применительно к историческому языкознанию, по сути, представляют собой формализованный и упорядоченный вариант «волновой модели», предложенной И. Шмидтом еще в 1872 г., в частности, посвящена значительная часть коллективной монографии [Renfrew et al. 2006]. С нашей точки зрения, «сетевое» моделирование межъязыковых отношений по определению носит смешанный «генетическо-ареальный» характер и, следовательно, никоим образом не может замещать «древесную» схему, претендующую на отражение исключительно генетических связей; таким образом, построение исторически осмысленных «деревьев» и «сетей» — не взаимоисключающие, а скорее взаимодополняющие задачи. При этом ни для одной «сети» по определению невозможно предложить оптимальную историческую интерпретацию, если она не будет сопровождаться «деревом», обратное же неверно.

несостоятельной¹. Только такой подход в состоянии уберечь исследователя от ошибок, типичных для истории изучения языкового родства, когда за следы такового неправомерно принимаются последствия активных языковых контактов или просто случайные совпадения формы и значения в нескольких языках.

Противоположный подход представлен в первую очередь в работах сторонников т. н. метода «массового» или «многостороннего» сравнения («mass» / «multilateral» comparison), связанного в первую очередь с именем выдающегося американского лингвиста Дж. Гринберга. Согласно теории, разработанной Гринбергом и его последователями, обосновать генетическое родство группы языков и построить для нее многоуровневую классификацию можно, вообще не прибегая к сравнительно-историческому методу; внимательный анализ лексических и грамматических данных сравниваемых языков и статистический учет числа обнаруживаемых сходств должен сам по себе привести исследователя к правильным ответам на эти вопросы. При этом, поскольку разные потомки одного языка-предка, как правило, со временем утрачивают не одни и те же, а разные элементы его грамматики и лексики, степень точности получаемых результатов напрямую зависит от количества привлекаемых к сравнению языков. Таким образом, например, позицию в той или иной генеалогической схеме семьи из 20 языков будет заведомо проще определить, чем место в ней одного языка-изолята, не имеющего ближайших родственников².

Жесткая полемика, развернувшаяся в последние двадцать — тридцать лет в западной лингвистике между «консерваторами» и «гринбергианцами» в конечном итоге завершилась почти полным вытеснением последних из академических кругов. Однако на самом деле борьба эта во многом вызвана скорее «околонаучными» или даже «вненаучными» причинами, чем «антинаучной» сущностью того или другого направления. Можно даже утвер-

¹ Подробное теоретическое обоснование этой позиции, частично основанное на детальном историческом анализе ряда гипотез о языковом родстве (с точки зрения их «успешности»), см. в [Campbell & Poser 2008].

² Основные методологические постулаты «массового сравнения» изложены в многочисленных статьях Дж. Гринберга на эту тему (см., например, [Greenberg 2001]), а также в предисловии к его монографии, посвященной вопросам классификации языков Америки [Greenberg 1987].

ждать, что основная причина, по которой полемика носит столь острый характер — неоправданная категоричность полемизирующих сторон, каждая из которых неизменно склонна рассуждать о языковом родстве в терминах «доказательства» наличия или отсутствия такового, т. е. получения, с помощью собственной (и ничьей другой) методологии заведомо однозначных ответов на поставленные вопросы с последующим имплицитным возведением их в статус догмы, не подлежащей дальнейшим сомнениям¹.

Между тем вряд ли подлежит сомнению тот факт, что любая гипотеза, формулируемая в виде «языки $x_1, x_2... x_n$ генетически родственны друг другу», является прежде всего гипотезой *вероятностной*, особенно в тех случаях, когда праязык для этих языков оказывается исторически незасвидетельствованным и подлежит реконструкции. Чем больше аргументов в пользу подобной гипотезы, тем выше ее вероятность, однако на сегодня остается формально недоказанным, может ли эта вероятность составлять сто процентов даже в самых «очевидных» случаях родства².

По-видимому, существует определенный «предел» аргументации, при преодолении которого (например, в случае индоевропейской семьи) сомнения относительно родства тех или иных языков могут оставаться лишь у тех, кто вообще по тем или иным соображениям не приемлет традиционную концепцию языкового родства (предпочитая ей, например, идею «языкового скрещивания» или какие-либо новейшие конвергентные модели). Фор-

¹ Ср.: «It is tempting to speculate that the litres of ink spilt over mass comparison would have reduced to mere blots if Greenberg had not attempted to present it as an independent method in its own right» [McMahon & McMahon 2005: 20].

² Отметим, что данный факт неоднократно подчеркивался и в работах самого Дж. Гринберга, ср., например: «I believe there is at least a dim realization that in all empirical sciences, as against logic and mathematics, in which truths flow infallibly and tautologically from definitions, 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]. К сожалению, несмотря на то, что в теоретическом плане с этим утверждением, скорее всего, согласятся как убежденные сторонники, так и противники Гринберга, на практике оно часто успешно забывается и теми, и другими.

мального определения такого «предела», однако, не существует. В практических целях в качестве одного из возможных критериев здесь, вероятно, можно привлекать интуицию носителей соответствующих языков, обязательно помноженную на мнение специалистов (в противном случае ложное ощущение «родства» с языком-донором может возникнуть у носителей языка-акцептора, заимствовавшего большое количество лексических элементов). Но на определенном этапе такое интуитивное «ощущение родства» заканчивается — и здесь уже маловероятно, а иногда и просто невозможно достичь единого мнения как относительно того конкретного момента, на котором оно заканчивается, так и относительно критериев обоснования языкового родства за пределами «порога аргументации».

Даже при условии согласия с консервативным подходом, когда родство считается доказанным (или «обоснованным») только при условии соблюдения всех требований сравнительно-исторического метода, нередко оказывается, что исследователи расходятся в конкретном практическом понимании этих требований.

Так, например, согласно утверждениям сторонников ностратической гипотезы (к которым причисляет себя и автор данной работы), реконструированный в работах В. М. Иллич-Свитыча и А. Б. Долгопольского праностратический язык восстановлен в полном соответствии с классической исторической методологией, а родство отдельных семей, составляющих ностратическую макросемью, обосновано едва ли не в той же степени, что и родство между собой языков, составляющих эти семьи — настолько, что данными ностратики можно уверенно пользоваться для прояснения спорных вопросов исторической фонетики и грамматики семей более низкого уровня (см., в частности, «оптимистический» отчет о текущем состоянии ностратики в [Dybo 1989]).

С ними, однако, не согласны критики ностратики, утверждающие, что даже в лучших трудах по этой дисциплине приверженность сравнительно-историческому методу заявляется лишь голословно, в то время как тщательный анализ данных обнаруживает многочисленные нарушения и отступления от этого метода, достаточные для того, чтобы поставить выводы под сомнение или даже вовсе обесценить их с научной точки зрения (см., например, [Campbell 1998], где об этом говорится на основании анализа

автором индоевропейско-уральских параллелей в материалах В. М. Иллич-Свитыча).

Разрешить спорность данной ситуации, как казалось бы на первый взгляд, могли бы помочь формальные алгоритмы проверки существующих гипотез, основанные на грамотном применении статистических и вероятностных методов. Большинство таких алгоритмов, разработанных западными специалистами¹, представляют собой относительно простые компьютеризированные процедуры, задача которых — протестировать данные, потенциально значимые для обнаружения языкового родства, на предмет их статистической значимости, т. е. определить, превышают ли отмеченные сходства тот или иной «порог случайности», преодоление которого необходимо для констатации родства. Несмотря на то, что все такие алгоритмы являются вероятностными, т. е. результаты их применения не обладают «доказательной силой» в классическом понимании этого термина, при условии корректности построения они могли бы систематически повышать (или понижать) доверие специалистов к тестируемым с их помощью гипотезам и вынуждать их принимать (или отвергать) эти гипотезы в качестве рабочих моделей.

¹ Подробный обзор литературы, посвященной применению математических методов к сравнительным языковым данным, в рамках нашего исследования занял бы слишком много места. В частности, в последнее десятилетие значительно возрос интерес к историческому языкознанию со стороны представителей различного рода смежных наук (антропологии, социологии, биологии и даже физики), в связи с чем общий вал публикаций такого рода начал возрастать едва ли не в геометрической прогрессии; к сожалению, многие из них демонстрируют низкий уровень владения собственно лингвистическим материалом, что обесценивает их результативность даже при условии формальной безупречности математического аппарата. Наибольший интерес в этом жанре до сих пор представляют работы, написанные профессиональными лингвистами, одновременно владеющими математическими навыками: индоевропеистом Д. Ринджем [Ringe 1992], синологом У. Бэкстером [Baxter 1995; Baxter & Manaster-Ramer 2000], лингвистом общего профиля Б. Кесслером [Kessler 2001] и т. п. Достаточно информативную (хотя на сегодняшний день уже слегка устаревшую) сводку данных о состоянии всего направления на текущий момент можно найти в монографии [McMahon & McMahon 2005].

К сожалению, общий недостаток большинства алгоритмов такого рода — их неспособность учитывать собственно *исторический фактор*. Исходным материалом обычно служат данные современных языков, из которых автоматическая процедура не умеет должным образом извлечь релевантную историческую информацию — начиная с установления регулярных фонетических соответствий (алгоритмы обычно основываются на поверхностном фонетическом сходстве сопоставляемых лексем) и заканчивая способностью правильно оперировать таксономическими единицами разных уровней (например, определять, что «генетический вес» одного-единственного языка А, составляющего отдельную ветвь семьи АВ, может быть сопоставим с соответствующим «весом» любого количества языков $B_1, B_2 \dots B_n$, составляющих вторую ветвь этой же семьи).

В тех немногочисленных случаях, когда тестирование все же проводится на материале реконструкций, внутри которых алгоритм пробует установить ряд соответствий, превышающих случайные ожидания (например, в работе [Ringe 1998]), полученные негативные результаты могут свидетельствовать не только об отсутствии родства, но и об элементарной недостаточности данных, используемых для тестирования¹. Так, если тестирование проводится на материале списков Сводеша, внутри которых истинно родственны между сравниваемыми языками (реконструкциями) не более 12-15% элементов, вряд ли уместно ожидать даже от идеально построенного алгоритма корректного обнаружения когнатов, формально доказывающих родство сравниваемых языков.

В итоге, как правило, результаты, полученные с помощью таких алгоритмов, согласуются с результатами, полученными в ходе «ручной» обработки материала компаративистами, только в тех случаях, когда речь идет о сравнительно «молодых» праязыках,

¹ Отметим, что результаты вероятностного теста, проведенного в [Ringe 1998] на материале праиндоевропейского и прафинно-угорского стословных списков, все же свидетельствуют скорее в пользу наличия, чем отсутствия генетического родства между этими семьями, что вынужденно признает и сам автор работы (стр. 187) — при этом все равно оговариваясь, что строго доказательными эти результаты считать нельзя.

потомки которых еще не успели измениться до взаимной неузнаваемости в плане фонетики и лексики.

Построение формального алгоритма нового типа, призванного в определенной степени автоматизировать ручную работу лингвиста-компаративиста — задача исключительно сложная, и правильное ее решение во многом зависит от того, удастся ли в будущем наладить серьезную кооперацию между компаративистами и специалистами в области статистических методов¹. На текущем этапе исследований более гибким и надежным все же представляется традиционный «ручной» способ обоснования гипотез языкового родства, с той оговоркой, что при его применении следует по возможности минимизировать число субъективно-интуитивных факторов, влияющих на принимаемые решения.

1.2. Критерии обоснования языкового родства и проблема «корректирующего фактора»: морфология и лексическая устойчивость.

Чтобы предложить максимально удобный и методологически корректный «ручной» способ, необходимо сначала вспомнить о

¹ Из новейших разработок такого рода следует особо выделить статью [Steiner et al. 2011], посвященную подробному описанию формального алгоритма, цель которого (едва ли не впервые в западной квантитативной исторической лингвистике) определена не просто как обнаружение неслучайных совпадений звучания в фиксированных списках, а как частичная симуляция работы компаративиста. Такое развитие необходимо всячески приветствовать, несмотря на то, что конкретные результаты тестирования, описанные в статье, пока что не дают возможности понять, будет ли соответствующий алгоритм в состоянии успешно справляться с «нетривиальными» ситуациями (далее языковое родство, умение различать между генетически общими когнатами и следами ареальных когнатов и т. п.). В частности, из двух протестированных семей алгоритм выдает позитивные результаты для цезских языков (близкородственных в рамках одной ветви нахско-дагестанской семьи), но неопределенные для языков матако-гуайкуру (гипотетическая семья в Южной Америке, относительно генетического единства которой не существует единой точки зрения), т. е., по сути, застревает на тех же сложностях, на которых дает сбой ручная процедура, осуществляемая компаративистом на «субъективных» основаниях.

том, что, собственно, называется «обоснованием языкового родства» (во избежание нежелательных коннотаций мы будем сознательно избегать сильного термина «доказательство»). В работе [С. Старостин 1999] предложено следующее определение:

«Языки В и С родственны друг другу, если: (а) все фонемы языков В и С регулярно соответствуют друг другу; (б) эти фонетические соответствия действуют на множестве базисной лексики языков В и С; (в) доля общей лексики возрастает, если берется выборка из более устойчивой лексики» (стр. 785-786).

Данное определение во многом пересекается с несколько менее формалистичным утверждением, обнаруживаемым в недавней работе американских лингвистов Л. Кэмпбелла и У. Позера, посвященной методологическим вопросам генетической классификации языков:

«...throughout the history of linguistics the criteria employed in both pronouncements about method and in actual practice for establishing language families consistently included evidence from three sources: **basic vocabulary**, **grammatical evidence** (especially morphological), and **sound correspondences**» [Campbell & Poser 2008: 4].

Два из предлагаемых трех критериев можно на самом деле объединить в один, который мы назовем *лексико-фонетическим*: обнаружение системы регулярных фонетических соответствий, истинной для подмножества базисной лексики сравниваемых языков. Дополнительный критерий, который в определениях С. А. Старостина и Кэмпбелла-Позера различается, можно назвать «контролирующим». Необходимость его введения мотивируется двумя соображениями:

а) в редких, но возможных, случаях базисная лексика языка А, обычно устойчивая к массовым заимствованиям, все же содержит большой процент заимствований из языка Б, и получающаяся в результате система соответствий будет отражать ареальные, а не генетические связи¹;

¹ С. А. Старостин на эмпирических основаниях предполагал, что в рамках 100-словного списка не может быть представлено более 15%

б) более серьезной опасностью можно считать появление «псевдо-систем» соответствий, когда морфемы объявляются родственными якобы на основании того, что составляющие их фонемы соответствуют друг другу, в реальности же оказывается, что эти «соответствия» работают на крохотной группе примеров, которые к тому же являются неудовлетворительными с точки зрения семантики и/или репрезентативности в языках или языковых подгруппах сравниваемых семей.

Как правило, слабость «псевдосистем» состоит именно в том, что они не выдерживают лексикостатистического тестирования¹; можно, однако, по крайней мере теоретически представить и такую ситуацию, при которой «псевдосистему» можно разработать, отталкиваясь от единичных случаев фонетического сходства элементов 100-словного списка, регулярность которых затем обосновывается дополнительными примерами из других областей лексики, ни один из которых сам по себе не является достаточно убедительным.

В качестве заведомо абсурдного примера рассмотрим произвольно выбранную «гипотезу» о языковом родстве между языком лого, относящимся к ветви мору-мади большой центральносуданской семьи, и корейским языком. В рамках 100-словного списка

заимствований [S. Starostin 1995a: 395]. Сегодня, однако, надежно зафиксирована небольшая группа случаев, когда число заимствований превышает 25%, как в контексте активных контактов одновременно с несколькими языками-донорами (дравидийский язык брахуи), так и на начальной стадии впоследствии прерванной «креолизации», где донором-лексикализатором является только один язык (северно-сонгайские языки) [Starostin 2010: 111].

¹ Конкретный пример — система «соответствий», якобы установленная Д. Макальпином между прадравидийским и эламским языками [McAlpin 1981] на материале 80 лексических сопоставлений; в критической работе [Starostin 2002] показано, что она практически неприменима к базисной лексике этих языков и наглядно свидетельствует в пользу отсутствия, а не наличия близкого родства между ними. В рамках африканистики всеми симптомами «псевдо-системности» обладают реконструкции глубокого уровня, выполненные К. Эретою для нило-сахарской [Ehret 2001] и койсанской [Ehret 2003a, 2003b] макросемей.

между ними обнаруживается целый ряд интересных схождений, ср., например¹:

‘кора’ — кор. *k̄ar-čil* : лого *kobo*; ‘кость’ — кор. *p̄jə* : лого *fa*; ‘грудь’ — кор. *čət* : лого *cici*; ‘собака’ — кор. *kä* : лого *kəkʰi*; ‘ухо’ — кор. *kwi* : лого *bi*; ‘нога’ — кор. *pal* : лого *pa*; ‘лист’ — кор. *ip* : лого *bi*; ‘мясо’ — кор. *sal* : лого *za*; ‘имя’ — кор. *iri-m* : лого *ru*; ‘дождь’ — кор. *pi* : лого *bu*.

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. Для того, чтобы продемонстрировать ее регулярность, нам необходимо привлекать дополнительный сравнительный материал, где, однако, мы уже не обязаны придерживаться столь же строгих семантических критериев.

Предположим, что кор. *pal*, лого *pa* ‘лист’ ← пракорейско-лого **pal*, а кор. *sal*, лого *za* ‘мясо’ ← **zal*, в результате действия двух фонетических законов: (а) ауслатное **-l* выпадает в лого; (б) **z-* оглушается в корейском.

Правило (а) уже в пределах 100-словного списка подтверждается на двух примерах, но можно привлечь и дополнительный материал. Ср.: кор. *tul* ‘2’ = лого *tu* ‘вместе’ (← ‘двоем’); кор. *mul* ‘вода’ = лого *mvi* ‘пить’; кор. *sul* ‘вино’ = лого *su* ‘жидкость’; кор. *adil* ‘сын’ = лого *adu* ‘сын’ и т. д. К правилу (б) ср.: кор. *sā-* ‘рассветать’ = лого *za* ‘рассветать’; кор. *sā-* ‘исчезать’ = лого *ze* ‘испражняться’. Наконец, к обоим правилам одновременно ср. кор. *sil-* ‘исчезать’ = лого *zi id*.

Далее, «нетривиальное» (но в типологическом плане вполне естественное) соответствие «кор. *kw-* : лого *b-*» (‘ухо’) подтверждается дополнительной корреляцией кор. *hā-kwi* ‘рассвет’ (где первый слог действительно морфологически отделим) = лого *bi-ŋi*

¹ Корейские формы цитируются по сравнительному словарю [S. Starostin et al. 2003] в транслитерации авторов; «параллели» в лого взяты из сравнительных списков в [Tucker 1940] и частично выверены по словарю [Vallaеys 1986].

'утро', а также, возможно, кор. *kwē-* 'любить' = лого *be* 'следовать (за)', '(быть) вместе с'. Немногочисленность таких корреляций не должна в данном случае быть поводом для смущения, т. к. и начальный *b-* в лого, и начальные сочетания *kw-*, *kʷ-* в корейском в принципе встречаются довольно редко.

Похожего рода «правила» можно предложить и для других «соответствий», предложенных выше. При этом как немногочисленность подтверждающих примеров, так и их семантическая размытость легко могут быть оправданы хронологической глубиной постулируемого родства¹; с точки же зрения типологической правдоподобности предлагаемые фонетические законы и семантические развития безупречны.

Применительно к данной конкретной ситуации контролирующим фактором будет в первую очередь являться наличие *конкурирующих гипотез*. Действительно, абсурдность «корейско-лого» родства для нас очевидна постольку, поскольку сопоставление лексического материала этих языков с материалом *других* языков убедительно показывает, что аргументов в пользу наличия у корейского ближайших родственников на Дальнем Востоке (японский), а у лого — в Африке (центральносуданские языки) намного больше, чем аргументов в пользу того, что ближайшим

¹ Аналогичные «абсурдные» группы сопоставлений, иллюстрирующие ненадежность постулирования генетического родства на основании анализа базисной лексики, в лингвистической литературе встречаются довольно часто; ср., например, англо-баскские сопоставления Л. Траска [Trask 1996: 114-115] или англо-маори сопоставления Л. Кэмпбелла и У. Позера [Campbell & Poser 2008: 382-384], о которых речь еще пойдет ниже. Однако подобного рода сопоставления, как правило, не идут далее общего перечня случайных сходств; их авторы не обсуждают возможность построения на их основе «псевдосистемы» соответствий. Классический же пример построения такой системы — работа Е. А. Хелимского, в которой автор на материале мнимых, но «поверхностно убедительных» «соответствий» между индоевропейским и семитским материалом показывает несостоятельность этимологической методики ностратиста А. Бомхарда [Хелимский 1989]. «Этимологии» Хелимского можно, правда, опровергнуть с помощью лексикостатистического критерия; но, как видно на примере «корейско-лого» сравнения, даже лексикостатистика может оказаться «фиктивной».

родственником корейского является лого, и наоборот. Это не сводит к нулю вероятность родства между лого и корейским, но делает ее «опосредованной», а корейско-лого сопоставление — методологически неправомерным; при корректном применении сравнительно-исторического метода место этого сопоставления должно, в лучшем случае, занять сопоставление реконструированного праалтайского языка с реконструированным працентральносуданским.

Однако нет гарантии, что аналогичную «псевдосистему» нельзя будет предложить и для «алтайско-суданского»; на место одной группы «выбывших из строя» случайных совпадений всего-навсего встанут другие. В очень большом числе ситуаций метод конкурирующих гипотез оказывается принципиально неприменим просто постольку, поскольку ни одна из этих гипотез сама по себе не будет обладать достаточной убедительной силой. Так, если между языками А и В, обнаруживающими 8% сходств в 100-словном списке, построить на этой основе «псевдо-систему» соответствий, вряд ли гипотезу об их родстве можно рассматривать серьезно лишь потому, что, например, между языками А и С такую же «псевдо-систему» удалось заставить работать только на 6% слов в списке, а между В и D — только на 5%!

Таким образом, введение надежного контролирующего фактора, который был бы способен «распознавать» ложные когнаты, базирующиеся на псевдосистемах соответствий, полностью оправдано. Для Л. Кэмпбелла и У. Позера, как уже указывалось выше, в роли такого фактора выступают морфологические параллели: поскольку грамматика по умолчанию считается более устойчивой к языковым изменениям, чем лексика (во всяком случае, такое представление является довольно типичным в диахронистической среде), схождения, наблюдаемые между грамматическими морфемами, в целом надежнее в качестве аргумента генетического родства, чем схождения в области лексики. При этом, однако, пытаться обобщить «морфологический критерий» в качестве универсального показателя степени языкового родства было бы невозможно хотя бы потому, что хорошо известны языки и целые языковые семьи, вообще лишенные парадигматической морфологии (например, в южноафриканском или юго-восточноазиатском регионах).

Более того, само по себе представление о какой-то особой устойчивости морфологического уровня языка по сравнению с остальными, насколько нам известно, не имеет строгого обоснования. По-видимому, оно базируется на типичном для структурализма интуитивном понимании грамматики как «языкового скелета», на который в дальнейшем «нанизывается» в значительной степени преходящая лексика. С точки зрения чисто синхронной лингвистики такую модель языка действительно можно считать наиболее естественной и удобной. В историческом плане, однако, подавляющее большинство подтверждающих ее конкретных примеров относится к сфере *близкого*, легко демонстрируемого языкового родства типа индоевропейского (и то, скорее, при учете в первую очередь древних языков, резко сокращающих хронологическую дистанцию от эмпирических данных до момента существования праязыка).

В качестве одного из аргументов, подрывающих абсолютизацию морфологического критерия, приведем цитату из работы [Jungraithmayr & Ibrizimow 1994], в которой обосновывается особая важность составления сравнительных лексических списков для обоснования и построения внутренней классификации чадской семьи языков (подчеркнем, что оба автора являются крупнейшими в своей области специалистами):

«...Chadic morphology is so diversified that its reconstruction will still take many years to come. Thus, the principal and, for the time being, sole criterion for grouping together some 130 languages spoken in the Central Sudan region... in one genealogical unit, lies in the existence of a hitherto quite restricted number of common lexical items, assumed to have been inherited from a common proto-language or proto-linguistic unit» (p. XI).

Показательно, что речь идет о чадской семье — генетическом объединении из почти 200 (так по данным каталога «Этнолог»; 130 — по-видимому, заниженное число) языков, составляющем одну ветвь в рамках афразийской макросемьи, общепризнанность единства которой, как это часто считается, в первую очередь держится как раз на данных сравнительной *грамматики*. Данный пример как нельзя лучше показывает, что абсолютизация морфологического критерия при обосновании генетического родства — не более чем искусственное ограничение, накладываемое на соот-

ветствующую процедуру рядом лингвистов под влиянием, с одной стороны, традиций исторического языкознания, выросшего из классической индоевропеистики, с другой — их личной интуиции.

Объективный анализ ситуаций, имеющих место в различных областях сравнительно-исторического языкознания, наглядно показывает, что *нет ни одной общепризнанной языковой семьи, генетическое родство членов которой можно обосновать исключительно на основе данных сравнительной морфологии и, наоборот, невозможно обосновать на основе лексических данных*¹. Обратное, между тем, неверно: существуют такие языковые группы, родство которых убедительно показывается на материале лексики и только лексики (в первую очередь это касается, разумеется, языков изолирующих — например, тай-кадайских). Известна даже ситуация, в которой лексика гораздо лучше, чем грамматика, демонстрирует преемственность между древней и современной стадиями одного и того же языка: речь идет о переходе от древнекитайского (I тыс. до н. э.) к современному состоянию, в ходе которого язык почти полностью обновил всю систему служебных слов (функционально соответствующих словоизменительным морфемам флективных языков), сохранив, тем не менее, значительный процент базисной лексики.

Лишь в отдельных случаях «сверхустойчивость» морфологической структуры языка или, по крайней мере, отдельных ее парадигматических блоков способствует признанию той или иной гипотезы о макрородстве языков на хронологическом уровне, превышающем 6 — 8 тысяч лет; классический пример — упоминавшаяся только что афразийская макросемья, ср.: «It is a family of very great age, undoubtedly the world's oldest demon-

¹ Неудачной попыткой привести такой пример является разбор А. Вовиным ситуации в скуликском диалекте атаяльского языка, якобы утрачивающим едва ли не всю основную австронезийскую базисную лексику и остающегося «очевидно австронезийским» лишь в плане морфологии [Vovin 2002: 159]. Как показано в подробном разборе материала в [Dybo & Starostin 2007: 131-135], данное мнение целиком основано на предвзятом выборочном подходе к материалу и легко опровергается при проведении детального историко-фонетического анализа 100-словного списка скуликского и других атаяльских языков.

strated family... though relatedness of the whole family can be demonstrated on paradigmatic and other grammatical evidence... lexical evidence is lacking» [Nichols 2010: 363].

Данное утверждение, на наш взгляд, представляет собой преувеличение, обусловленное поверхностным знакомством автора с достижениями сравнительно-исторической лексикологии афразийских языков. Действительно, между большинством ветвей афразийской макросемьи наблюдаются (особенно в глагольной системе) разительные грамматические изоморфизмы, способные оказать на психологию исследователя не менее сильное впечатление, чем классическая система парадигматических соответствий между, например, древними индоевропейскими языками. Факт этот, однако, (а) сам по себе является скорее исторической случайностью, и неправомерно было бы предъявлять аналогичные требования к любой другой гипотезе о макрородстве; (б) не должен никоим образом принижать роль афразийских лексических соответствий, количество которых, вопреки мнению Дж. Николс, достаточно велико, в том числе и в области базисной лексики (см. хотя бы [Militarev 2000]). Единственной ветвью афразийской макросемьи, базисная лексика которой может вызывать серьезные опасения относительно ее генетического статуса, можно считать омотскую — однако, любопытным образом, именно эта же ветвь вызывает сомнения и в плане нахождения в ней типично «афразийских» грамматических изоглосс, что заставляет исследователей в последнее время все чаще призывать к радикальному пересмотру омотских данных (см., в частности, [Theil 2006]).

Таким образом, нисколько не преуменьшая значимости морфологии в вопросах установления генетического родства, нельзя все же согласиться с использованием ее в качестве аргумента *sine qua non* вне зависимости от конкретных обстоятельств (на чем, в частности, среди современных африканистов настаивает в своих типологических и сравнительно-исторических исследованиях Том Гюльдеманн; см. [Güldemann 2008] и др. публикации того же автора). Утверждение, что отсутствие парадигматических изоглосс в морфологических системах сравниваемых (пра)языков делает невозможным установление между ними родственных связей, в лучшем случае безосновательно, в худшем

же может завести дальнейшие исследования в этой области в методологический тупик или пустить их «по ложному следу».

Вернемся теперь к контролирующему фактору в системе С. А. Старостина: «языки В и С родственны друг другу, если... доля общей лексики возрастает, если берется выборка из более устойчивой лексики» (авторство идеи принадлежит С. Е. Яхонтову, хотя на практике этот принцип был впервые применен именно С. А. Старостиным к лексикостатистическому обоснованию алтайской семьи [С. Старостин 1991]). В истинности данного утверждения трудно усомниться; действительно, если все наблюдаемые сходства между базисной лексикой языков В и С случайны, то каждое подмножество исходного 100-словного списка должно содержать примерно одинаковое количество таких случайных схождений. Неравномерное распределение схождений по подмножествам, выделенным на основании формального критерия (например, «устойчивости»), будет однозначно свидетельствовать в пользу того, что случайность — не единственный фактор, отвечающий за наличие таких схождений.

Разумеется, для того, чтобы признать за данным контролирующим фактором право на существование, необходимо удостовериться в том, что «степень устойчивости» базисной лексики — объективная реальность, а не интуитивная фикция. Работа по выявлению небольшой группы лексических элементов, сверхустойчивых к изменениям, была впервые проведена А. Б. Долгопольским [1964, 1965], выделившим на основании разработанных им статистических алгоритмов список из 15 таких элементов; однако для конкретных лексикостатистических подсчетов более приемлемым в практическом плане оказался 35-словный список С. Е. Яхонтова (см. выше), активно использовавшийся в работах С. А. Старостина и других исследователей.

«Массовая» статистическая проверка степени устойчивости тех или иных элементов списка Сводеша оказалась, тем не менее, возможной только после того, как в рамках деятельности Московской школы компаративистики оказались накоплены и компьютеризированы в значительных количествах 100-словные списки Сводеша по различным языковым семьям Евразии и (в несколько меньшем объеме) Африки, Америки и Австралии, привязанные к этимологическим базам данных по соответствующим

щим языкам. Это, в свою очередь, позволило С. А. Старостину [2004] вывести, на основании статистического анализа всей совокупности данных, т. н. «индекс стабильности» для элементов 100-словного списка, т. е. ранжировать их в порядке максимальной среднестатистической устойчивости. Полученную иерархию (за вычетом 10 добавочных «контрольных» элементов из второй половины 200-словного списка Сводеша) имеет смысл процитировать здесь целиком:

1 мы; 2 два; 3 я; 4 глаз; 5 ты; 6 кто; 7 огонь; 8 язык; 9 камень; 10 имя; 11 рука; 12 что; 13 умирать; 14 сердце; 15 пить; 16 собака; 17 вошь; 18 луна; 19 ноготь; 20 кровь; 21 один; 22 зуб; 23 новый; 24 сухой; 25 печень; 26 есть; 27 хвост; 28 этот; 29 волос; 30 вода; 31 нос; 32 не; 33 рот; 34 полный; 35 ухо; 36 тот; 37 птица; 38 кость; 39 солнце; 40 дым; 41 стоять; 42 дерево; 43 пепел; 44 давать; 45 дождь; 46 звезда; 47 рыба; 48 шея; 49 грудь; 50 лист; 51 приходить; 52 убивать; 53 нога; 54 сидеть; 55 корень; 56 рог; 57 лететь; 58 слышать; 59 кожа; 60 длинный; 61 мясо; 62 дорога; 63 знать; 64 сказать; 65 яйцо; 66 семья; 67 колено; 68 черный; 69 голова; 70 спать; 71 жечь; 72 земля; 73 перо; 74 плавать; 75 белый; 76 кусать; 77 жир; 78 мужчина; 79 человек; 80 весь; 81 ночь; 82 видеть; 83 идти; 84 теплый; 85 красный; 86 холодный; 87 женщина; 88 круглый; 89 желтый; 90 лежать; 91 зеленый; 92 облако; 93 большой; 94 кора; 95 песок; 96 хороший; 97 много; 98 гора; 99 живот; 100 маленький.

Во избежание недоразумений следует подчеркнуть, что данное ранжирование является именно *среднестатистическим*, т. е. совершенно не обязано соблюдаться для каждой отдельно взятой семьи. Отклонения от «стандарта» неизбежны (да и сам он является, очевидно, не единственно возможным, т. к. вполне допустимы и другие алгоритмы ранжирования, отличные от предложенного в системе С. А. Старостина¹), и изучение их возможных

¹ Один из таких алгоритмов, в частности, разработан в рамках проекта ASJP (Automated Similarity Judgment Program), о котором см. также ниже, и описан в работе [Holman et al. 2008]. Несмотря на то, что процедура ранжирования проводилась авторами на значительно меньшем объеме материала, чем в работе С. А. Старостина, а также совершенно не учитывала историко-этимологический компонент, будучи целиком автоматизированной, результаты ранжирования оказались во многом

масштабов в зависимости от ареально-географической и исторической дистрибуции языковых семей может в дальнейшем стать одним из увлекательнейших и наиболее перспективных направлений в исторической и синхронной семантике.

Тем не менее, в целом не подлежит сомнению — и легко верифицируется на практике — что, если ранжированный таким образом список разбить пополам, то пересечение его первой («сильноустойчивой») половины с «сильноустойчивой» половиной 100-словного списка, вычисленной индивидуально для *любой* языковой семьи, вне зависимости от ее размера или возраста, будет значительно превышать 50%. Так, «сильноустойчивая» половина индоевропейского списка имеет 31 пересекающийся элемент с «сильноустойчивой» половиной усредненного ранжированного списка; «сильноустойчивая» половина уральского списка — целых 39 элементов и т. п. Исключений из данного правила, по-видимому, нет (во всяком случае, в числе тех семей, которые уже были подвергнуты лексикостатистической обработке), что неопровержимо доказывает значимость «индекса стабильности» и целесообразность его применения для обоснования языкового родства дальнородственных семей.

Вернемся в этой связи к приводившемуся выше «заведомо абсурдному» примеру родства между лого и корейским и рассмотрим, выдерживает ли соответствующая гипотеза проверку «индексом стабильности». На первый взгляд, ответ должен быть утвердителен: из 10 предполагаемых схождения между этими двумя языками 7 относится к первой половине списка и лишь 3 ко второй, что в целом соответствует ожиданиям — дальнейшее родство такого типа предполагает больше параллелей в «сильно-», нежели «слабоустойчивой» половине списка. Более тщательный анализ, однако, показывает, что на самом деле никакого родства здесь нет, по следующим причинам:

сопоставимы с приведенным нами списком (показательный пример: слово 'маленький' в обоих ранжированных списках последовательно занимает последнее место по устойчивости), хотя отдельные моменты и производят впечатление случайных курьёзов (так, наиболее устойчивым элементом списка согласно данному алгоритму оказывается слово 'вошь').

а) соотношение 7 : 3 для родства между двумя современными языками выглядит не очень реалистично. Так, для сравнения, в аналогичной ситуации между древнееврейским и зенага, двумя представителями афразийской макросемьи, число схождения между которыми (по подсчетам А. Ю. Милитарева) также не превышает 10 единиц, 9 из 10 схождения относятся к «сильноустойчивой» половине списка ('кровь', 'умирать', 'я', 'имя', 'язык', 'два', 'мы', 'кто', 'что'; десятое этимологическое сближение, 'желтый', на самом деле очень сомнительно). В нашем случае доля сохранности «слабой» лексики составляет 30%, хотя ожидалось бы скорее число, близкое к нулю;

б) результаты будут еще плачевнее, если список поделить не на две, а на три или более частей. Так, если принять во внимание только первую «супердесятку» списка, между теми же французским и бенгали родственными в ее составе окажутся 8 из 10 корней (все, кроме 'огня' и 'камня'); между корейским же и лого в ней обнаруживается лишь одно сходство ('имя') — причем, любопытным образом, именно оно легче всего опровергается, даже не прибегая к дополнительным сравнительным или историческим данным (кор. *irim* 'имя' — прозрачное именное производное от глагола *iri-* 'говорить, сообщать'). С 11-го по 30-й элемент списка также обнаруживается только одно совпадение ('собака').

Таким образом, стратификация лексических параллелей между корейским и лого, основанных на элементарном фонетическом сходстве и, иногда, «зачаточных» фонетических соответствиях показывает, что во всех этих случаях речь идет о случайных совпадениях¹ — в том случае, разумеется, если мы принимаем в качестве методологического руководства «корректирующий фактор» С. А. Старостина.

¹ Подчеркнем, что в чисто теоретическом плане по крайней мере некоторые из отмеченных схождения могут быть и неслучайны, т. е. отражать «сверхглубокое» родство между рассматриваемыми языками. Однако на данном этапе сравнения у нас нет никакой возможности отделить случайные созвучия от «сверхдревних» когнатов. Методологически корректным здесь будет только такое сравнение, которое рассматривает корейский и лого материал в исторической перспективе тех семей и, далее, макросемей, к которым относятся эти языки.

С чисто умозрительной точки зрения нельзя, разумеется, исключать, что возможны ситуации, в которых наблюдаемый феномен (для которого можно предложить условное название «динамическая градация лексики») будет отсутствовать, т. е. «сильноустойчивая» лексика по какой-то причине будет замещаться быстрее, чем «слабоустойчивая». Нам, однако, такие случаи неизвестны, и, более того, можно осмелиться предположить, что, будь они реально засвидетельствованы, речь шла бы о каких-то уникальных ситуациях креолизации, для описания которых вообще требуется особый подход¹.

Все вышесказанное позволяет нам в окончательном виде сформулировать следующие методологические постулаты:

1) Универсальным и неизменным условием обоснования языкового родства должна служить *общая базисная лексика* рассматриваемых языков, с определенными на ней *регулярными фонетическими соответствиями*. Морфологические изоглоссы — по крайней мере, в рамках построения общей классификации языков мира или отдельно взятого обширного географического ареала, такого, как африканский континент — имеют второстепенное значение; грамматические морфемы могут, по-видимому, рассматриваться наравне со «среднеустойчивыми» элементами базисной лексики, но лишь в рамках хронологически неглубоких семей — использовать морфологическое сравнение для обоснования глубинного родства, за редкими исключениями, скорее бесперспективно.

2) Любая гипотеза о возможном родстве между языками или языковыми семьями должна в обязательном порядке учитывать

¹ Детальное рассмотрение проблемы креолизации, как вообще, так и применительно конкретно к африканскому континенту, не входит в цели нашего исследования, т. к. современные африканские креолы (такие, как крио и др.) распознаются более или менее однозначно, а гипотезы о существовании языков, устроенных аналогичным образом, в глубокой древности могут пока что существовать исключительно в умозрительной плоскости. Относительно использования лексикостатистики в целях опознания возможных случаев «креолизации» см., в частности, [Бурлак 2007]. Некоторые дополнительные соображения будут высказаны ниже, в рамках дискуссии о соотношении генетически унаследованных и ареальных слоев в лексическом инвентаре языков Африки (раздел 1.8.3.3).

фактор динамической градации базисной лексики рассматриваемых языков; особенно актуален этот фактор для гипотез «дальнего» родства, при котором процент совпадений в базисной лексике опускается ниже определенного порога (скажем, в 25-30%). Без надлежащего учета динамической градации лексики соответствующую «гипотезу» можно выстроить для любой произвольно взятой пары языков или языковых семей, т. к. (а) для *любой* такой пары найдется хотя бы небольшое количество случайных созвучий, равномерно распределенных по всему множеству базисной лексики; при большом желании и небольшом усилии такие созвучия легко будет затем трансформировать в «псевдосистему» фонетических соответствий; (б) для *некоторых* таких пар число «ложных» сходств в базисной лексике будет превышать порог случайности за счет заимствований, процент которых в «слабоустойчивой» половине 100-словного списка будет заведомо выше, чем в «сильноустойчивой».

1.3. «Слабое» и «сильное» обоснование языкового родства.

Выше, на примере заведомо фиктивного «корейско-лого» сравнения, было в общих чертах показано, что три условия обоснования языкового родства, обозначенные в работе [С. Старостин 1999], по-видимому, заведомо *достаточны* для того, чтобы языки, удовлетворяющие этим условиям, были общепризнаны как генетически родственные.

Тем не менее, на практике, особенно при работе с плохо описанными языками и языковыми семьями, перед компаративистом регулярно встает и будет вставать вопрос о том, является ли выполнение этих условий *необходимым в полном объеме*. Речь в первую очередь идет об условии (а) — «все фонемы языков В и С регулярно соответствуют друг другу»¹.

¹ Стоит, вероятно, уточнить, что под «всеми» фонемами уже в исходном определении С. А. Старостина имеются в виду «все фонемы, унаследованные от праязыкового состояния»; во многих случаях (будь то русский спирант *ф*, ряды придыхательных смычных в дравидийских языках или губные взрывные в севернокойсанских языках, см. ниже) удастся легко показать, что та или иная фонема или даже подсистема фонем имеет в том или ином языке контактно-ареальное происхождение, и

В техническом плане построение полной системы фонетических соответствий между потенциально родственными языками эквивалентно выполнению праязыковой реконструкции; оно предполагает подробный анализ всего лексического корпуса рассматриваемых языков и создание этимологического словаря этих языков. Следовательно, абсолютизация данного требования вынуждает нас признать, например, факт наличия родства между сахарскими языками тубу и канури необоснованным, даже несмотря на то, что все специалисты по этим языкам поголовно убеждены в том, что они происходят из общего источника — постольку, поскольку ни исчерпывающего списка фонетических соответствий между этими языками, ни представительного этимологического словаря этих языков до сих пор не существует.

Обычный здравый смысл подсказывает, что такого рода сверхтребовательный подход, по-видимому, чрезмерен, и можно лишь согласиться с вполне рациональной позицией Дж. Гринберга, категорически отвергавшего его во многих своих публикациях¹. Демонстрация исчерпывающей системы фонетических соответствий между языками А и В, подкрепленной этимологическим словарем в 1,000 — 2,000 морфемных единиц, является наглядным доказательством *реконструкции праязыка АВ*², но для того, чтобы убедительно обосновать сам факт родства между собой этих языков, это требование завышено, как с точки зрения интуиции, так и элементарной статистики.

Конечно, с точки зрения психологии исследователя, особенно «традиционного» лингвиста-компаративиста, неполнота установленной системы соответствий и малочисленность представленных

искать ей регулярные соответствия в генетически родственных языках бессмысленно.

¹ См., например, возражения в [Greenberg 1987: 7-8] на «гиперкритическую» позицию венгерского лингвиста И. Фодора, которую тот занимает в ходе полемики с африканской классификацией Гринберга [Fodor 1966]; по мнению Фодора, только построение детальной, регулярной на 100% системы фонетических соответствий способно обладать настоящей доказательной силой.

² Разумеется, при условии, что соответствия установлены корректно, т. е. работают на множестве базисной лексики и удовлетворяют требованию динамической градации.

этимологий будет всегда оставлять определенную почву для сомнения, поскольку «эталоном» компаративистики всегда была и, по-видимому, навсегда останется индоевропейская семья языков, для которой эти проблемы давно преодолены (если закрыть глаза на отдельные нерегулярности и спорные фонетические интерпретации наблюдаемой системы соответствий).

Так, если мы станем утверждать, что нам удалось установить между языками А и В наглядные фонетические соответствия в области консонантизма, но не вокализма (такова на сегодняшний день, например, ситуация с реконструкцией алтайской макросемьи), за этим практически неизбежно последует критика, требующая, для убедительности, предоставить и систему вокалических соответствий — даже несмотря на общеизвестный факт, что вокалические системы в языках мира в целом значительно менее стабильны, чем консонантные, и вытекающее из него следствие: по крайней мере для некоторых языковых семей надежно и однозначно реконструировать систему консонантизма — задача реальная, а систему вокализма — нереальная.

Обоснование языкового родства, сопровождаемое полной реконструкцией праязыковой фонологической системы и основного морфемного инвентаря, будем называть «сильным»; усомниться в факте такого родства можно, лишь отбросив все основные теоретические постулаты и методологические принципы сравнительно-исторического языкознания. Не менее ошибочным, однако, будет и признание сильного обоснования единственно приемлемым и «по-настоящему научным», по следующим, довольно очевидным, причинам:

а) *техническая*: для огромного количества языков мира «сильное» обоснование их таксономической принадлежности просто «физически» невозможно в силу элементарной нехватки описательных данных — если, например, материалы по группе языков ограничены обзорными списками из сотни лексических единиц, то, сколь бы близким и «интуитивно очевидным» ни было родство, его нельзя будет подтвердить доказательной системой соответствий, не содержащей значительных лакун;

б) *историческая*: подавляющее большинство общеизвестных и в целом общепринятых гипотез языкового родства получило массовое признание задолго до установления между ними пол-

ной системы фонетических соответствий (этот аргумент, строго говоря, нельзя считать «научным», но нельзя и игнорировать, т. к. он де-факто говорит о наличии в родственных языках своего рода «диагностических маркеров», распознавание которых позволяет избежать классификационных ошибок уже на ранних этапах сравнительно-исторического исследования);

в) *прагматическая*, связанная с элементарным фактором экономии усилий — если можно показать, что задача «обоснования языкового родства» решается быстрее, чем задача «подробной праязыковой реконструкции», нет никакой целесообразности в том, чтобы настаивать на полном их отождествлении, хотя, разумеется, успешное решение первой ни в коем случае не отменяет необходимость второй.

Значимость фонетических соответствий и праязыковой реконструкции для обоснования языкового родства обычно акцентируется — что вполне естественно — в первую очередь компаративистами «традиционной закалки», как правило, специалистами в области исторического изучения тех или иных конкретных языковых семей. Есть, однако, как минимум две группы исследователей, которые склонны скорее принижать или даже совершенно игнорировать эту значимость:

а) лингвисты-синхронисты, типологи и представители смежных наук (социологии, антропологии, биологии и т. д.), не отягощенные багажом компаративистики и предпочитающие детальной и во многом неизбежно субъективной работе компаративиста формализованные статистические алгоритмы, оперирующие небольшим массивом лексических, грамматических и типологических данных. Традиционная компаративистика среди таких специалистов может не считаться обладающей доказательной силой — «реконструкция» должна осуществляться уже после доказательства родства, базирующегося на статистике;

б) представители школы «массового сравнения». Хорошо известна, в частности, скептическая позиция уже самого Дж. Гринберга по отношению к «соответствиям» и «реконструкции» как условиям *sine qua non* для доказательства языкового родства. Так, в работе [Greenberg 1987: 9-17], перечисляя многочисленные типы ситуаций, в которых даже самые консервативные компаративисты-традиционалисты допускают этимологическое отожде-

ствление морфем, нарушающих фонетические соответствия (аналогия, сандхи, контаминации, диалектизмы и т. п.), он тем самым подводит читателя к выводу, что четкой грани между сравнительным методом и «массовым сравнением» на самом деле не существует, т. к.:

— с одной стороны, формально механистические принципы, якобы лежащие в основе сравнительного метода, на практике работают лишь на определенном проценте этимологизируемого материала;

— с другой стороны, ничто не мешает исследователю обнаруживать отдельные, самые наглядные — и, как правило, самые доказательные — образцы регулярных соответствий непосредственно в ходе первичного обследования материала «на глаз», т. е. «массового сравнения».

Логичность аргументации Гринберга в данном случае вряд ли подлежит сомнению. Более того, следует, по-видимому, согласиться и со следующим утверждением, даже если, на первый взгляд, оно может поставить под сомнение всю область макрокомпаративистики:

«...what many linguists fail to appreciate is that anything approaching a complete and highly convincing reconstruction on the basis of recurrent correspondences is in general possible only with languages so closely related that it is unnecessary anyway» [Greenberg 1987: 33].

Ключевыми в этом утверждении являются слова «complete» и «highly convincing». Неверно, что регулярные соответствия *в принципе* не могут быть установлены для языков, связанных глубоким родством — успехи, достигнутые в этой области, например, в алтаистике или афразистике, свидетельствуют об обратном — но, с другой стороны, сравнение любого из имеющихся на сегодняшний день вариантов афразийской реконструкции, например, с прасемитской реконструкцией, показывает, что по сравнению с прасемитской системой праафразийскую нельзя определить ни как «полную» (хотя бы потому, что даже ведущие афразисты признают наличие определенных лаун в системе соответствий), ни как «в высшей степени убедительную» (хотя бы потому, что общий уровень скепсиса по отношению к работам по афразий-

ской реконструкции намного превышает скепсис по отношению к реконструкции семитской).

Исходя из этих соображений, следует согласиться с Гринбергом, что «полнота» праязыковой реконструкции не должна считаться ключевым условием обоснования языкового родства. Даже наоборот: опыт показывает, что, чем «полнее» реконструкция, тем больше конкретных претензий можно предъявить к отдельным ее элементам, т. е. тем более обесценивается ее, казалось бы, неоспоримое — на начальном этапе — преимущество: механистичность соответствий, исключая вредное влияние субъективной позиции исследователя.

Однако может ли считаться удовлетворительной альтернатива, предлагаемая Гринбергом? Его во многом справедливая критика сравнительно-исторического метода в конечном итоге приводит его к полному отказу от работы в рамках этого метода, аналогично тому, как справедливые критические замечания в адрес глоттохронологии в свое время привели к совершенно незаслуженному отказу лингвистического «мэйнстрима» от соответствующей методики в целом. Отвергая «сильное обоснование родства» как единственно возможное, он противопоставляет его «массовому сравнению» — технологии сопоставления языковых данных, в принципе не снабженной никаким методологическим аппаратом, кроме интуиции и личного опыта исследователя.

Критике «массового сравнения» — занятию, в целом не особенно затратному и поэтому довольно популярному в лингвистической среде — посвящено огромное количество работ, хотя подавляющее большинство оппонентов Гринберга обычно концентрируется не столько на попытках показать несостоятельность его подхода в теоретическом плане, сколько на выявлении конкретных ошибок, допущенных им при цитировании и анализе языкового материала. Думается, однако, что в текущей ситуации более важной задачей, чем показать «непрофессионализм» Дж. Гринберга (стоит расценивать его скорее как побочный эффект колоссальности конкретных задач, которые этот выдающийся ученый не боялся ставить перед собой, нежели как следствие необразованности или ненаучности мышления), нужно считать поиск определенного компромисса между сравнительно-

историческим методом в его «механистической» форме и «массовым сравнением» с его относительной гибкостью.

Попробуем найти этот компромисс на конкретном примере. Выше уже было упомянуто тройное сопоставление базисной лексики английского, хинди и маори языков в [Campbell & Poser 2008: 382-384] — один из наиболее наглядных и впечатляющих образцов критики «массового сравнения» за последнее время. Как показывают авторы, поверхностный анализ материала, основанный на фонетическом сходстве (семантика сравниваемых лексем при этом совпадает, что уже само по себе представляет собой «ужесточение» методики Гринберга), дает 10 потенциальных когнатов между английским и хинди, столько же между хинди и маори, и целых 12 «когнатов» между английским и маори, т. е. оказывается, что «инспективный» метод Гринберга не в состоянии отличить реальное родство между двумя современными индоевропейскими языками от эффектов случайных совпадений.

Предсказуемым ответом на данный пример со стороны гринбергианской школы могло бы быть заявление о методологической некорректности самого факта такого сопоставления: очевидно, что в рамках настоящего «массового сравнения» эти три языка необходимо рассматривать в совокупности с многими другими, включая их настоящих ближайших родственников — что позволит легко определить германскую принадлежность английского, индоарийскую для хинди и полинезийскую для маори, после чего совокупность всех данных по индоарийским языкам неизбежно породнит их с германскими, но не с австронезийскими.

Такая реакция, однако, будет скорее уклонением от реального ответа, чем адекватным решением проблемы; выше, в разделе, посвященном «корейско-лого» сравнению, уже говорилось о том, что подобного рода ситуации легко могут возникнуть и в тех случаях, когда у рассматриваемых языков (или языковых семей, или праязыковых реконструкций) нет «очевидных» ближайших родственников, и тогда «массовость» сравнения не даст ощутимых преимуществ. Критерий «массовости» вообще, по-видимому, эффективно применим лишь в тех ситуациях, когда процессы языковой дивергенции хронологически равномерны, т. е., например, язык А за тысячу лет распадается на языки A_1 и A_2 , эти языки еще через тысячу лет распадаются соответственно на $A_{1.1}$ / $A_{1.2}$ и

$A_{2,1}$ / $A_{2,2}$, и так далее, с одинаковыми интервалами в течение длительного промежутка времени. Если же развитие идет неравномерно, или если большая часть образованных по ходу этого развития языков со времен вымирает, не дав «потомства» — что гораздо более типично для языков мира — мы оказываемся лицом к лицу с «изолятами» (как отдельными языками, например, айнским или кунама, так и небольшими языковыми семьями, например, енисейской или сонгайской), для которых «массовость» «массового сравнения» не имеет никакого значения.

Представим поэтому такую ситуацию, в которой английский и хинди — единственные живые представители индоевропейской семьи, об истории которых нам ничего не известно, а маори — единственный представитель полинезийской (или даже шире, австронезийской) семьи, лексика которого оказалась нам доступна. Проанализируем теперь сходства, которые были отмечены между этими тремя языками в работе Кэмпбелла и Позера (для полноты картины включим сюда также «сомнительные» фонетические схождения, которые авторы помечают знаком вопроса, выделив их курсивом):

Слово	Английский	Хинди	Маори
этот	<i>this</i>		<i>te:nei</i>
тот	<i>that</i>		<i>te:ra; te:na:</i>
что		<i>kyā:</i>	<i>he-aha</i>
кто	who		wai
не	not	nahī	
много		bahut	maha
два	two	do:	<i>rua</i>
большой	big	<i>baṛa:</i>	pi:ki
длинный	long	<i>lamba:</i>	<i>roa</i>
женщина	woman		wa:hine
собака		kutta:	kuri:
семя		<i>bi:ṣ</i>	<i>pua</i>
кора		<i>kha:l</i>	<i>kiri</i>
мясо	meat	mā:s	mi:ti
кость	<i>bone</i>		<i>poroiwi</i>
яйцо	egg		he:ki
рог	horn		haona

Слово	Английский	Хинди	Маори
хвост	tail		te:ra
перо	feather	pa:ɾ	
волос	hair		huruhuru
глаз		ā:kh	kanohi
нос	nose	na:k	
рот	mouth	mūh	ma:ngai
зуб	tooth	dā:t	
нога	foot	pā:v, pair	pu:, waewae
рука	hand	ha:th	
живот	belly	pe:tʃ	
есть		kha:-	kai
кусать		ka:tʃ-	kakati
умирать		mar-	mate
убивать		ma:r-	-mate
ходить	walk		wa:ke
лежать	lie	leʃ-	
давать		de:-	tapae
сказать		kah-	ko:rero
солнце	sun	su:raʒ, su:rya	ra:
луна	moon		marama
звезда	star	ta:r, sita:ra:	
вода	water		wai
камень		patthar	po:hatu, ko:hatu
огонь		a:g	ahi
пепел		ra:kh	pungarehu
гора	mountain		maunga
красный		la:l	whero
зеленый	green		kirini
желтый		pi:la:	punga
полный	full	pu:ra:	puhapuha
новый	new	na:ya:	
имя	name	na:m	

Достаточно беглого взгляда на эту таблицу, чтобы понять, насколько бесплодным будет анализ обнаруженных сходжений, не выходящий за рамки «массового сравнения». Даже учет «слабых» сходжений наравне с «сильными» не сильно меняет

картину: 18 англо-хинди параллелей при 22 англо-маори и 25 хинди-маори.

Попробуем, однако, исследовать эти параллели на предмет их соответствия принципу динамической градации. В пределах «устойчивой» половины 100-словного списка мы обнаруживаем 11 англо-хинди параллелей при 10 англо-маори и 13 хинди-маори — что настораживает, т. к. необходимое условие родства (*более 50% когнатов должно относиться к «устойчивой» половине списка*) соблюдается только для «англо-хинди», но никак не для двух других пар.

Нельзя, конечно, сказать, что соотношения 10 : 12 и 13 : 12 чересчур разительно отличаются от соотношения 11 : 7; для того, чтобы показать значимость этого отличия, необходимо прибегать к более сложным статистическим процедурам. Однако посмотрим теперь, какова доля параллелей в «устойчивой» половине 100-словника, которые сами Кэмпбелл и Позер относят к «сильным» схождениям, т. е. таким, в которых консонантный костяк сравниваемых форм совпадает почти полностью.

Оказывается, что из 11 англо-хинди параллелей «сильными» являются целых 10, в то время как из 10 англо-маори параллелей «сильных» — только 4, а из 13 хинди-маори — только 7. Более того: внимательный анализ «сильных» соответствий позволяет нам, даже не выходя за пределы списка, подметить ряд потенциальных нетривиальных соответствий между английским и хинди, например, англ. *t-* : хинди *d-* (*two* : *do*, *tooth* : *dā:t*), что для других пар практически невозможно¹.

¹ Нельзя не отметить также определенную ангажированность 100-словных списков Кэмпбелла и Позера, по крайней мере в отношении языка маори, для которого авторы, в нарушение собственного принципа, иногда приводят квази-синонимы (присутствие которых в данном списке должно быть исключено), по-видимому, только для того, чтобы максимизировать число «ложных когнатов» в «маори-английском» и «маори-хинди». В частности, основное слово со значением 'большой' в маори — *nui*; 'кость' — *wheua*; 'хвост' — *waero* или *whiore*; 'волос (на голове)' — *takawe* (*huruuru* — 'волосы на теле'); 'полный' — *ki*: [Biggs 1990]. Подстановка правильных эквивалентов на место предвзятых «убивает» как минимум 1 «сильную» хинди-маори и 2 «сильных» англо-маори параллели.

В работе [Baxter & Manaster-Ramer 2000] авторы, на наш взгляд, довольно убедительно показывают, что данных даже 35-словных списков английского и хинди языков вполне достаточно для надежного статистического обоснования («доказательства») их генетического родства. Тем не менее, мы намеренно хотели бы уклониться от попытки «добить» этот вопрос с помощью формальных статистических методов — хотя бы потому, что «идеального», универсального и общепринятого статистического алгоритма доказательства родства, как говорилось выше, не существует — и вместо этого подчеркнуть следующий факт: в рассматриваемой ситуации для нас значимы не абсолютные статистические показатели, а их относительное распределение.

Представленный выше анализ, начатый Кэмпбеллом и Позером и продолженный уже в рамках данной работы, по сути, является не чем иным, как опытом относительной формализации «массового сравнения» — наложением строгих ограничений на используемый в ходе этой процедуры материал и привлечением к анализу отдельных элементов сравнительно-исторического метода. Получается, что «массовое сравнение» как таковое бесильно определить, какой из двух языков — хинди или маори — в общемировой классификации ближе к английскому. Но если оказываются выполнены следующие условия:

- рассматриваемый лексический материал сжимается до «стабильного минимума» — 100-словного списка;

- внутри 100-словного списка потенциальные когнаты распределяются по группам с учетом принципа динамической градации;

- «слабые» фонетические схождения рассматриваются отдельно от «сильных» (правила, формально разграничивающие эти два типа схождений, сформулировать в целом нетрудно; это будет сделано ниже);

- также *желательно* (но не обязательно), чтобы либо внутри самого 100-словного списка, либо с привлечением дополнительных данных, можно было хотя бы для некоторых «слабых» схождений продемонстрировать сколь-либо регулярные фонетические соответствия;

- то этого в большинстве случаев будет достаточно для того, чтобы предложить т. н. «слабое обоснование языкового родства».

Последнее, отталкиваясь от приведенного выше определения С. А. Старостина (хорошо подходящего для *сильного* обоснования), можно определить следующим образом:

— для языков В и С можно в рабочем порядке считать, что они являются ближайшими родственниками, если (а) внутри множества базисной лексики языков В и С¹ обнаруживается некоторое количество фонетически изоморфных параллелей; (б) доля изоморфной лексики возрастает, если берется выборка из более устойчивой лексики; (в) ни для одного языка D не показано, что хотя бы для одной из пар В-D или С-D условия (а) и (б) выполняются в большей степени, чем для пары В-С.

Данное определение требует некоторых пояснений.

1) Выражение «в рабочем порядке» указывает на то, что слабое обоснование родства не следует считать «доказательством» или заслуживающим «догматического» статуса. Окончательно развеять сомнения в этом вопросе может только «сильное обоснование» — сопровождаемое детальной реконструкцией и этимологией.

2) Формулировка «ближайшие родственники» (вместо ожидаемого «родственны друг другу») необходима для того, чтобы, во-первых, подчеркнуть известную аксиому о невозможности доказательства неродства языков, во-вторых, чтобы подчеркнуть относительный и ступенчатый характер обоснования.

3) Фонетический изоморфизм — сложное понятие, заслуживающее отдельного объяснения, которое будет дано ниже; ограничимся пока констатацией того, что речь идет о некотором компромиссном варианте между классическими «регулярными соответствиями» и аморфным «фонетическим сходством» в рамках теории «массового сравнения».

4) Сохранение практически в неизменном виде параметра (б) из определения С. А. Старостина имеет ключевое значение. Поскольку фонетические изоморфизмы даже в пределах 100-словного списка будут неизменно обнаруживаться между любыми произвольно взятыми языками (корейский и лого, английский и маори и т. д.), только разумный учет этого параметра

¹ «Базисная лексика» в данном случае определяется как стандартный 100-словный список Сводеша для языков В и С.

гарантирует, что ложное «слабое обоснование» языкового родства в большинстве таких случаев окажется невозможным.

5) Учет параметра (в) имеет важный практический смысл: он призван уберечь исследователя от формулирования большого количества лишних гипотез. Так, на практике попытка слабого обоснования англо-хинди родства, даже если она и будет в целом успешной (как показано выше), обесмысливается тем, что проще и убедительнее будет сперва обосновать вхождение английского языка в германскую группу, а хинди — в индоарийскую, после чего общегерманский «инвариант», в свою очередь, легко скомпонуется с общеиндоарийским.

В строго теоретическом плане, конечно, данный параметр должен заставить нас при формулировании любой гипотезы о родстве языков В и С тестировать ее на прочность, сравнивая их данные с данными всех остальных языковых семей мира. В обозримом будущем, когда будет проведена компьютерная обработка всех необходимых данных и будут разработаны соответствующие алгоритмы, такое тестирование окажется выполнимым. Пока что, однако, имеет практический смысл ограничиться своеобразной «презумпцией невиновности»: если слабое обоснование родства между В и С представлено, оно будет функционировать как рабочая гипотеза до тех пор, пока не будет представлено статистически более значимое обоснование родства между языками В и D или С и D.

При этом, разумеется, если мы говорим, например, о выяснении таксономического статуса языков какого-либо данного ареала («сахарский регион», «южноафриканский регион» и т. п.), то и в теоретическом, и в практическом плане выгоднее провести полное обследование данного региона, сопоставляя данные всех зафиксированных в нем языков, а не отдельных образцов, для того, чтобы как можно раньше отсеять подавляющее большинство ложных гипотез. Вряд ли, однако, имеет серьезный смысл, говоря, например, о связях нилотских языков с сурмийскими (в Эфиопии), настаивать на том, что слабое обоснование их родства невозможно до тех пор, пока нилотский и сурмийский материалы — по отдельности — не будут сопоставлены также, скажем, с австронезийскими или юто-ацтекскими данными.

На самом деле в концепции «слабого обоснования» как таковой нет ничего кардинально инновативного. По сути, это не более чем попытка частично формализовать и подвести под некий общий стандарт первый, подготовительный этап «сильного обоснования», т. е. реконструкции — селекцию сравнительного материала и определение «на глазок» потенциальных когнатов и типов фонетических соответствий. В целом мы согласны с Дж. Гринбергом и его школой, что именно *этот* этап имеет ключевое значение для определения родства, по крайней мере в чисто практическом плане — любой исследователь обычно формирует гипотезу родства *до* того, как приступает к обнаружению и системному описанию нетривиальных фонетических соответствий. Но тем важнее постараться максимально «десубъективировать» данный этап, исключив из рабочего материала максимальное число шумовых эффектов и поставив остаток на квантифицируемую основу, чтобы не только узкие специалисты в области той или иной языковой семьи, но и гораздо более широкий круг исследователей, обладающих базисным лингвистическим образованием, могли без лишних усилий оценивать степень убедительности различных конкурирующих гипотез.

От чисто теоретических соображений перейдем теперь к описанию ряда конкретных аспектов «слабого обоснования языкового родства».

1.4. Фонетическая совместимость и фонетический изоморфизм.

Ключевым и, вместе с тем, пока что наиболее туманным в предложенном выше определении «слабого родства» является понятие «фонетического изоморфизма» (далее — ФИ), которым было заменено традиционное понятие «регулярного фонетического соответствия» (далее — РФС). Под ФИ мы понимаем более широкий спектр явлений, чем РФС, т. е. РФС — это, по сути, один из видов ФИ. Тем не менее, в рамках «слабого обоснования» (а, как уже говорилось выше, для некоторых вполне разумных и высоко вероятных гипотез родства «слабое обоснование» является единственно возможным) мы не можем позволить себе роскошь использования РФС и *только* РФС в качестве возможных научных аргументов. Следует помнить, что для того, чтобы данное

фонетическое соответствие считалось полностью регулярным, необходимо выполнение как минимум двух условий (а) и (б), и крайне желательно выполнение еще двух условий (в) и (г):

(а) *рекуррентность*: соответствие должно воспроизводиться на материале такого количества лексических пар (при бинарном сопоставлении языков), которое превышает ожидаемый порог случайности (последний, в свою очередь, серьезно варьирует в зависимости от степени выполнения второго условия);

(б) *семантическая неоспоримость*: хотя бы несколько лексических пар, на которых определено соответствие (точное количество не поддается общему определению, т. к. зависит от частотности соответствующих фонем в конкретных языках, а также от общего объема анализируемого материала), должны иметь *полностью совпадающие значения* или, по крайней мере, такие различия в семантике, которые можно считать типологически тривиальными (многократно засвидетельствованные в языках мира, напр. 'дождь': 'вода', 'глаз': 'видеть' и т. п.);

(в) *системность*: регулярность соответствия обычно подтверждается тем, как оно вписывается в общую постулируемую систему, т. к. фонетические изменения чаще всего протекают на уровне отдельных дистинктивных признаков, а не их пучков. Так, регулярность соответствия «*t* : *t*» на фоне наличия регулярных соответствий «*p* : *b*», «*k* : *g*», «*č* : *ž*» и, наоборот, отсутствия регулярных соответствий «*p* : *p*», «*k* : *k*», «*č* : *č*» будет а priori сомнительной (для принятия его придется, скорее всего, требовать повышенного количества примеров). Еще более сомнительной будет регулярность соответствия «*t* : *t*» между языками А и В, если при этом фонемы *p*, *k* и *č* в языке А не обнаруживают вообще никаких параллелей в языке В — ситуация странная, но теоретически допустимая;

(г) *фонетическая интерпретируемость*: любое предлагаемое соответствие, проецируемое на фонологическую систему реконструируемого праязыка, должно иметь не менее одной фонетической интерпретации, не вызывающей серьезных нареканий с точки зрения общей типологии фонетических изменений. Так, если для родственных языков постулируются соответствия «*n* : *b*», «*n* : ?», «*n* : *k*» с реконструкцией на их месте праязыковых «фонем» **n₁*, **n₂*, **n₃* (или **b₁*, **?₁*, **k₁*), реалистичность такой реконструкции

будет сомнительной даже при условии цитирования двух-трех семантически неоспоримых примеров на каждое из этих соответствий (что не следует, впрочем, интерпретировать как принципиальную *невозможность* такого рода соответствий — в исключительных ситуациях иногда обнаруживаются достаточно реалистичные сценарии, убедительно объясняющие их возникновение).

Последние три условия на практике часто игнорируются исследователями, особенно макрокомпаративистами, забывающими о том, что работа с языками, связанными дальним, интуитивно неочевидным родством, требует более детального и формального очерчивания понятия «РФС», чем работа с близкородственными языками, где шансы допустить грубую этимологическую ошибку значительно ниже. Особенно часто игнорируется — в первую очередь из-за того, что оно, как правило, не прописывается эксплицитно в фундаментальной литературе по компаративистике — условие семантической неоспоримости. Тем не менее, вряд ли будет ложным утверждение, что за всю историю компаративистики бесспорно общепринятым не оказалось ни одно постулированное РФС, которое бы грубо нарушало условия (а), (б) или (г).

С другой стороны, при работе с языками, представленными ограниченным объемом материала — будь то ограничение *естественное*, вызванное недостаточной описанностью соответствующих языков, или *искусственное*, возникающее из-за необходимости опоры на небольшой массив стабильной базисной лексики — на практике часто оказывается, что для многих сопоставлений, интуитивно представляющихся вполне приемлемыми, невозможно дать строгое обоснование на базе РФС. Особенно часто такая ситуация возникает, когда речь идет о достаточно дальнем родстве языков; в этом случае непреодолимые препятствия могут возникнуть как с выполнением условия семантической неоспоримости (из-за того, что подавляющее большинство родственных слов успело сменить свое значение), так и рекуррентности (если родственных слов между языками вообще осталось мало).

Поэтому представляется, что на уровне «слабого» обоснования языкового родства в тех случаях, когда этимологическое тождество некоторых морфем не удастся строго обосновать с помощью РФС, допустимо определенное смягчение требований.

Чтобы это смягчение не носило произвольно-субъективный характер, необходимо ввести своеобразную трехступенчатую градацию по «силе» фонетических схождений между морфемами сравниваемых языков:

I. *Фонетический изоморфизм*: основные (в первую очередь консонантные, см. ниже) фонемы сопоставляемых морфем коррелируют друг с другом по определенным фонетическим признакам, причем в типологическом плане звуковые изменения, приводящие к возникновению таких корреляций, вполне естественны. Фонетический изоморфизм может быть *полным*, когда артикуляция сравниваемых фонем совпадает, или *частичным*. Так, в паре «хинди *mar-* : маори *mata*» изоморфизм первого согласного корня — полный, второго — частичный (т. к. *r* и *t* — разные согласные, но примеры перехода *t* в *r* или наоборот в языках мира время от времени действительно обнаруживаются).

II. *Фонетическая совместимость*: разновидность фонетического изоморфизма, для которого можно предложить историческую интерпретацию в рамках по крайней мере отдельного сегмента определенной компаративной модели, причем интерпретация должна носить реалистичный характер, т. е. не противоречить ни общим историко-типологическим закономерностям, ни нашим конкретным знаниям (= сильно аргументированным гипотезам) о соответствующих языках.

III. (*Регулярное*) *фонетическое соответствие*: разновидность фонетической совместимости, которая (а) носит статистически доказательный характер, т. е. можно показать, что количество семантически неоспоримых примеров, иллюстрирующих существование РФС, заведомо превосходит число случайно ожидаемых корреляций; (б) является частью более или менее цельной и завершенной *системы* соответствий, работающей на всей совокупности доступного сравнительного материала.

Насколько нам известно, никто из лингвистов, занимавшихся и занимающихся вопросами обоснования (доказательства) языкового родства на ограниченном корпусе данных, без полной языковой реконструкции, не пытался дать точное определение понятию «фонетической совместимости» (под этим или любым другим именем), промежуточному между «фонетическим сходством» («изоморфизмом») и «фонетическим соответствием» —

даже несмотря на то, что многие из них совершенно эксплицитно оперируют им на практике. Объясняется это, скорее всего, тем, что любое определение такого понятия неизбежно будет намного более сложным, расплывчатым и субъективным, чем определение сущности РФС и даже «фонетического сходства».

Вместо того, чтобы пытаться, прибегнув к формальной терминологии, одной фразой определить «фонетическую совместимость», попробуем описать эту концепцию в рамках своего рода практической «инструкции по применению». Примем за прототипическую структуру корня CVC (где V — любой гласный, C — любой согласный, в том числе нулевой, или сочетание согласных; более сложные структуры, такие, как CVCV, CVCVC и т. п., будем, в рамках допустимого упрощения, считать производными от CVC). Сформулируем теперь условия, при которых корень $C_{1a}V_aC_{2a}$ в языке A может считаться «фонетически совместимым» с корнем $C_{1b}V_bC_{2b}$ в языке B.

1. Вокализм: V_a будет считаться совместимым с V_b в следующих случаях: (а) $V_a = V_b$; (б) $V_a \neq V_b$, но фонетические различия между гласными минимальны, т. е. ограничиваются одним дистинктивным признаком (близлежащие ступени подъема, напр. $e : i, o : u$; упреждение или отсутствие такового, напр. $i : \acute{i}, u : \acute{u}$; +ATR / -ATR, напр. $e : \varepsilon, u : v$); (в) $V_a \neq V_b$, причем различия между гласными существенны, но объяснимы либо в рамках некоторой модели исторического развития, предполагаемой для данных языков (например, развитие сингармонизма или грамматикализация вокализма, т. е. аблаут), либо как позиционные изменения, зависящие от консонантного окружения.

В ситуации (в) как для того, так и для другого объяснения важно не столько наличие подтверждающих их дополнительных примеров, сколько отсутствие очевидных контраргументов. Так, корни $tap_a : top_b$ будут считаться фонетически совместимыми при условии, что параллельно с ними не обнаружены корни $tar_a : tar_b, taŋ_a : taŋ_b$ и т. п. (изменение $a \rightarrow o$ после губного согласного совершенно естественно, но должно носить регулярный характер). Если такие корни все же обнаруживаются, задача обоснования фонетической совместимости усложняется: в этом случае соответствие $tap : top$ может отражать особую фонему праязыка

(например, **тэл*), но здесь уже необходимы дополнительные подтверждающие примеры.

Перечисленные условия означают, что требования к вокализму сравниваемых корней значительно менее строги, чем требования к консонантизму — это объясняется общеизвестной историко-типологической тенденцией вокализма к большей изменчивости. Однако это ни в коем случае не означает, что на вокализм при сравнении можно вообще не обращать внимание. Например, корни *tak_a* и *tik_b* не будут считаться фонетически совместимыми сами по себе; для того, чтобы это стало возможным, необходимо либо (а) продемонстрировать еще хотя бы два-три семантически неоспоримых примера на такое же соответствие; либо (б) предложить для него типологически вероятное объяснение, не противоречащее конкретным языковым фактам — например, возвести обе формы к двусложной структуре **taki* с переогласовкой в языке В и последующей редукцией второго слога в обоих языках (при этом ни в языке А, ни в языке В двусложные структуры типа *taki* встречаться не должны). Идеальная ситуация (приближающая нас к уровню постулирования РФС) предполагает выполнение и того, и другого условия.

2. Консонантизм: C_a будет считаться совместимым с C_b в следующих случаях:

(а) $C_a = C_b$;

(б) $C_a \neq C_b$ с «тривиальными» различиями между согласными, ограничивающимися одним дистинктивным признаком (например, глухость / звонкость, придыхательность / непридыхательность, наличие / отсутствие лениции и т. п.) или одним типологически достоверным фонетическим переходом (например, «ротацизм» $s \rightarrow r$ и т. п.). Рекуррентность такого изоморфизма следует продемонстрировать на материале нескольких семантически неоспоримых примеров (по крайней мере в тех случаях, когда мы имеем дело с достаточно большими и профессионально записанными объемами материала), однако анализ вполне допустимо проводить на уровне дистинктивных признаков. Так, если анализируемый материал содержит всего один пример на соответствие $p : p^h$, но также один на соответствие $t : t^h$ и один на соответствие $k : k^h$, все три пары корней можно считать фонетически совместимыми;

(в) $C_a \neq C_b$ с «нетривиальными» различиями между согласными, не поддающимися рационализации в рамках модели системного перехода на уровне дистинктивных признаков (типичные примеры таких соответствий — $k : ?$, $p : \emptyset$, $t : r$, $l : w$ и т. п.). Для того, чтобы постулировать в этих случаях фонетическую совместимость, необходимы подтверждающие ее семантически неоспоримые примеры (желательно иметь хотя бы два-три случая на каждое такое соответствие, хотя точное число трудноопределимо; по-видимому, оно должно зависеть от общей частотности соответствующих фонем).

В самом крайнем случае — особенно если речь идет о статистически редких фонемах рассматриваемых языков — фонетическую совместимость можно «диагностировать» и для одного примера, но только при условии отсутствия явных противоречий и/или необходимости реконструкции на этом месте «особой» фонемы праязыка.

Иными словами, корреляция типа «А *mal* : В *taw*» может считаться примером фонетической совместимости, если (а) найдены аналогичные корреляции в виде, например, «А *kal* : В *kaw*»; или (б) не найдено ни одной аналогичной корреляции, но при этом столь же эксплицитно отсутствуют корреляции вида «А *-l* : В *-l*», что позволяет предложить внутренне непротиворечивый сценарий о регулярном развитии пра-АВ ауслатного $*-l \rightarrow -w$. Если же, наряду с уникальным примером «А *mal* : В *taw*», обнаружен хотя бы один пример вида «А *-l* : В *-l*», фонетическая совместимость исключается до тех пор, пока рекуррентность корреляции $l : w$ в рассматриваемых языках не подтвердится.

В практическом плане анализ материала на уровне фонетической совместимости отличается от анализа на уровне фонетических соответствий в первую очередь *стратификацией* материала; к анализу привлекается почти исключительно базисная лексика, в первую очередь — список Сводеша или определенная его часть, с возможностью подключения дополнительных параллелей, носящих семантически неоспоримый характер и подтверждающих ту или иную гипотезу фонетической совместимости. Помимо этого, фонетическая совместимость может постулироваться на базе *уникальной* корреляции, в отличие от фонетического соответствия (разумеется, в повседневной компаративистской практике

нередки случаи, когда «регулярное фонетическое соответствие» обосновывается на материале одного-единственного примера, но это не более чем практическая условность — очевидно, что соответствие не может формально называться «регулярным», если регулярность его не показана).

Для языков или языковых групп, связанных близким или «средним» родством, такой анализ представляет собой упрощенную, но действенную процедуру обоснования родства (так, «англо-хинди» родство, даже без привлечения данных по другим германским или индоарийским языкам, с его помощью легко устанавливается). Что касается групп и семей, связанных между собой дальним родством (например, семей, образующих ностратическую, сино-кавказскую и т. п. макросемью), то здесь есть основания утверждать, что для обоснования такого родства «тестирование на фонетическую совместимость» — если не единственно возможный, то, по крайней мере, наиболее естественный механизм обоснования исторической реальности таких макросемей.

Дело здесь в том, что, чем глубже мы уходим в лингвистическую предысторию человечества, тем труднее становится задача обнаружения семантически неоспоримых параллелей между языками. Анализ таких «образцовых» (по состоянию на текущий момент) этимологических работ по макросемьям, как ностратические словари В. М. Иллич-Свитыча и А. Б. Долгопольского или сино-кавказская база данных С. А. Старостина, показывает, что такие параллели находятся в абсолютном меньшинстве по сравнению с этимологиями, семантическое сходство членов которых неочевидно и имеет более редкие типологические прецеденты. Данное обстоятельство часто ставится в вину макрокомпаративистам со стороны скептически настроенных исследователей (см., например, [Campbell 1998: 117]), которые, однако, упускают из виду, что (а) семантически неоспоримых параллелей между языками или праязыками, связанными дальним родством, по определению не может быть много; (б) корректная оценка любой гипотезы дальнего родства должна в первую очередь отделить семантически неоспоримые параллели от оспоримых и проверить их как в количественном отношении, так и в качественном (насколько хорошо они отражают предложенную систему фонетических совместимостей или соответствий).

Установить между дальнородственными языками полноценную систему РФС можно только при условии отказа от критерия семантической неоспоримости. Однако ослабление семантических требований неизбежно влечет за собой повышенную вероятность принятия за когнаты случайных совпадений и, в конечном итоге, построения «псевдосистемы» соответствий. Поэтому на этапе обоснования родства пренебрегать семантической точностью недопустимо; с вероятностной точки зрения разумнее пренебречь стопроцентной регулярностью соответствий и опираться вместо этого на принцип фонетической совместимости.

Подчеркнем еще раз, что анализ потенциальных когнатов на их фонетическую совместимость ни в коем случае не предполагает чисто механистический подход к проблеме, аналогичный тем, которые, например, лежат в основе формальных аналитических алгоритмов Д. Ринджа, Б. Кесслера, С. Вихманна и других исследователей. В этих работах речь идет, как правило, не о фонетической совместимости, а о фонетических изоморфизмах, и они вообще не предполагают компаративной обработки материала, что, разумеется, серьезно облегчает задачу формализации и десубъективизации анализа, но вместе с тем заставляет исследователя смириться с включением в результирующую статистику заведомо недостоверных результатов.

Введение принципа «фонетической совместимости» представляется нам рациональным компромиссом между полным отказом от компаративистического аппарата, лежащим в основе как метода «массового сравнения», так и автоматических алгоритмов поиска фонетического сходства, и, наоборот, работой строго в рамках традиционной сравнительно-исторической модели со всеми ее достоинствами и недостатками. Насколько объективным и независимым от интуиции и личных предпочтений исследователя может быть этот принцип, и призвана, в частности, выяснить данная работа, во многом носящая экспериментальный характер.

1.5. «Массовое сравнение» и значимость дистрибуционного фактора.

В предыдущих параграфах обсуждалась в основном бинарная модель языкового родства, т. е. вопрос обоснования родства

между отдельно взятой парой языков или праязыков. С точки зрения формальной кладистики такая модель — единственно возможная: для каждого объекта допустим один и только один «ближайший родственник» (по крайней мере, в тех случаях, где количество признаков / параметров, по которым сопоставляются «родственники», настолько велико, что вероятность того, что три объекта будут отличаться друг от друга по одинаковому числу равновесных параметров, пренебрежимо мала).

На практике, однако, языки часто объединяются в группы из трех и более «примерно равноудаленных» членов, особенно в ситуациях, когда выбор небольшого числа параметров не дает явного перевеса для выделения того или иного бинарного узла (например, если лексикостатистические подсчеты показывают одинаковый процент для трех различных языков), а увеличение числа параметров либо затруднительно в техническом плане, либо сомнительно в историческом.

Еще более остро эта проблема встает там, где речь идет о дальнем языковом родстве. При том, насколько затруднительной оказывается задача опознания бинарных узлов в рамках даже такой общепризнанной семьи, как индоевропейская (где отдельные бесспорные узлы типа «балто-славянского» или «индо-иранского» сосуществуют с дискуссионными типа «итало-кельтского», «греко-армянского» и многих других), разговор о бинарных узлах для, например, ностратической макросемьи на данном этапе наших представлений об этой общности почти бессмысленен; в большинстве этимологических работ по ностратике в каждой отдельной этимологии можно встретить прауральскую, праиндоевропейскую, праалтайскую и т. п. формы, но едва ли где-либо можно обнаружить наравне с этим попытку постулирования «праурало-алтайской» или «праиндо-уральской» и т. п. формы — даже если исследователь-ностратист и имеет собственное представление о внутренней классификации ностратической макросемьи, он обычно не пытается трансформировать его в часть формального описания этой семьи.

Сущностная причина данной проблемы очевидна. С исторической точки зрения одновременное разделение одного языка на три или более (если «разделение языка» понимать как исторический сценарий, при котором одна часть носителей в ходе

миграции, дислокации, природного катаклизма и т. п. оказывается географически обособленной от другой части, что создает условия для начала самостоятельного языкового развития) — явление достаточно редкое, но только в том случае, если слово «одновременное» воспринимается полностью формально, т. е. «с точностью до конкретного дня». При некоторой — в практическом плане неизбежной — степени огрубления «одновременной» будет и такая языковая дивергенция, при которой, например, с одного и того же «насиженного» места в течение трех-четырех поколений в разные стороны мигрирует около десятка индивидуальных групп (семей, кланов и т. п.). Если представить, что для конкретного сценария такого рода мы владеем всей необходимой информацией о том, какими путями и в какие годы проходили миграции, изобразить надлежащее языковое древо, состоящее исключительно из бинарных узлов, будет несложно. Однако информация такого рода нам, как правило, недоступна.

Таким образом, в стандартном случае компаративист, как в силу самой природы устройства изучаемого объекта (включающей «волновые» эффекты диффузии языковых изменений, искажающие первичную модель дивергенции), так и из-за обстоятельств технического характера (недостаток информации) не может ограничиваться бинарными сопоставлениями. Из этого также логически вытекает, что он совершенно не обязан для любой тройки языков А, В, С предоставлять уверенный ответ на вопрос «является ли ближайшим родственником языка А язык В или язык С?», т. к. любая попытка дать такой ответ опять-таки сводит ситуацию к системе бинарных узлов.

Можно ли считать такую «несводимость к бинарности» преимуществом исторического языкознания или, наоборот, его серьезной проблемой? С точки зрения Дж. Гринберга, это очевидное преимущество, на котором, собственно, и зиждется сформулированный им принцип «массового сравнения»: поскольку *разные* языки-потомки одного праязыка неизбежно утрачивают *разные* элементы этого праязыка, привлечение к сравнению *большого* числа языков-потомков позволит обнаружить *большее* число неслучайных сходств, т. е. праязыковых черт. С точки зрения оппонентов Гринберга, это скорее проблема (за исключением разве что случаев близкого родства): чем больше языков привлекается к

сравнению, тем выше вероятность искажения истинной исторической картины за счет многочисленных «шумовых» эффектов (ведь очевидно, что найти, скажем, сто фонетически созвучных и семантически близких лексических пар на множестве 10 корпусов неродственных языков намного легче, чем на аналогичном множестве всего из двух корпусов).

Поясним ситуацию на конкретном примере. Перечисляя фонетические изоморфизмы, якобы подтверждающие, на его взгляд, существование «нило-сахарской» макросемьи, Дж. Гринберг приводит следующую группу форм: гао сонгай *taba*, канури (сахарская семья) *tambu*, динка (нилотская семья) *thyep* 'пробовать (на вкус)' [Greenberg 1966: 146]. При этом nilотская семья, согласно его классификации, составляет лишь одну из подгрупп огромной шари-нильской семьи, в которую в общей совокупности входит порядка двухсот единиц. Логично задать вопрос: насколько вероятно, что, взяв произвольную языковую семью, состоящую из двухсот языков, мы не сможем обнаружить ни в одном из них корень типа **TVP ~ *TVM* со значением 'пробовать'? Сравним для японское *tamesu*, корякское *tawan*, и даже древнекитайское *thep* id.: вряд ли из данного обстоятельства должно вытекать, что и эти языки также должны быть включены в nilо-сахарскую макросемью. (Не подлежит сомнению, что и для многих других nilо-сахарских сопоставлений Гринберга в этих языках можно найти похожие формы; ср. в этой связи не лишнее остроумия шуточное «доказательство» Л. Кэмпбелла америндской принадлежности финского языка при условии следования критериям Гринберга [Campbell 1988]).

Попробуем, однако, подойти к этой проблеме в плане ответа не на вопрос «насколько доказательным/убедительным является данное сопоставление, как само по себе, так и в совокупности с другими сопоставлениями данного рода?», а на вопрос «каким образом мы можем конструктивно верифицировать данное сопоставление, исходя из опыта, накопленного в рамках сравнительно-исторического языкознания?» Скорее всего, ответственный компаративист традиционной закалки не будет ни отвергать, ни принимать данное сопоставление «с порога». Вместо этого он будет пытаться найти ответ на следующие вопросы:

1) являются ли приведенные Гринбергом формы (а) изолированными в рамках указанных языков/диалектов, или же (б) наличие лишь небольшая выборка, ограниченная техническими причинами (нехватка материала или незнакомство с материалом, сокращения объема монографии и т. п.), на самом же деле можно показать, что все три формы восходят к соответствующим формам с таким же значением в прасонгай, прасакхарском и пранилотском?;

2) если формы действительно изолированы, то (а) существует ли вообще возможность спроецировать их на соответствующий прауровень, или же (б) можно наглядно показать их инновативный характер (например, то, что они на самом деле заимствованы из близлежащих языков, или представляют собой относительно недавние семантические развития из слов с совершенно иными значениями)?;

в) если перечисленные формы в том или ином виде все-таки проецируются на прауровень, то можно ли подтвердить неслучайность фонетического сходства между соответствующими формами через (хотя бы частично установленную) систему РФС?

На первые два вопроса, как правило, ответ дать достаточно легко (и тем удивительнее выглядит категорическое нежелание большинства сторонников парадигмы «массового сравнения» обращать на них внимание). Так, в нашем случае можно показать, что гао сонгай *taba* имеет надежные параллели и в других сонгайских языках (зарма *tàbâ*, денди *tá:bà*, койра чини *taba* и др.) и легко проецируется на прасонгайский уровень как основной глагольный корень со значением 'пробовать'¹; что ближайшая параллель к канури *tambu* — тубу *tam id.*, т. е. корень восстановим на празападносахарском уровне, но не имеет явных параллелей в восточносахарских языках; наконец, что динка *thyep* 'пробовать' — слово, не имеющее явных, семантически неоспоримых параллелей ни внутри «ближайших родственников» этого языка (напр., нуэр), ни в других ветвях большой нилотской семьи.

¹ Справедливости ради следует отметить, что на момент написания основополагающих трудов по классификации языков Африки единственным существенным материалом по сонгай, доступным Дж. Гринбергу, была монография А. Проста по диалекту гао [Prost 1956].

Таким образом, на первом этапе мы, скорее всего, поступим правильно, выбросив из сопоставления форму в динка; при общем размере нилотской семьи (порядка 60 языков) вероятность того, что лишь в динка — языке, который даже не представляет сам по себе одну из первичных ветвей пранилотского, а входит в маленькую подгруппу динка-нуэр западнонилотской ветви — сохранился в малоизменившемся виде древний «нило-сахарский» архаизм **TVP* 'пробовать', не может быть значимой.

Единственный серьезный шанс для этого корня претендовать на пранилотский уровень — ситуация, при которой *каждая* мелкая группа уровня динка-нуэр имела бы свой индивидуальный корень со значением 'пробовать'. Из этого следовало бы, что любой из этих корней имеет примерно одинаковый шанс на архаичность, и, таким образом, вопрос не имеет решения без попытки выхода на более высокий, внешний уровень — каковой попыткой и является сопоставление Гринберга. В данном случае, однако, показать это заведомо невозможно — даже не выходя за пределы западнонилотской семьи, очевидно, что глагольное значение 'пробовать' в праязыке этой семьи выражалось не тем же корнем, что динка *thyer*, а другим (ср. хотя бы ланго *bil'o* = луо *bi:lo* = нуэр *bi:lɛ* id.).

Следующий этап анализа — сопоставление прасонгайской формы **taba* и канури-тубу формы **tam*, имеющих одинаковое значение. Фонетическое сходство их опять-таки несомненно в силу совпадения места артикуляции C_2 , однако для того, чтобы сопоставление носило убедительный характер, ему необходимо дать системное обоснование, т. е. (а) подкрепить его дополнительными, в такой же степени семантически неоспоримыми примерами и (б) предложить для него историческую интерпретацию в рамках разрабатываемой модели.

Сравнительные материалы Гринберга таких дополнительных примеров не содержат; единственное сопоставление, в котором это «соответствие» якобы подтверждается — сонгай *kaba* 'борода': канури *ɲgimi* 'жевать' (автор подключает сюда также шаринильский корень со значением 'челюсть', восстанавливая общую семантику 'подбородок') — семантически оспоримо. Исторической или структурной интерпретации данного соответствия также не предлагается. Следовательно, сравнение — по крайней мере, в

контексте материала, приводимого Гринбергом — не может быть использовано в целях обоснования исторической реальности «нило-сахарской» макросемьи.

Разбор данного примера (аналогичному анализу можно подвергнуть и любой другой пример из африканистических работ Гринберга), на наш взгляд, обнажает главный недостаток как всей идеологии «массового сравнения», так и обращенного против нее критического аппарата, разработчики которого, хотя и добились на практике поставленной цели — вытеснения «массового сравнения» на обочину компаративистики — тем не менее, по сути приняли те же самые правила игры.

Недостаток этот уже был вкратце упомянут выше: речь идет о практически полном игнорировании в ходе «массового сравнения» *исторического компонента* как такового. Классификационная методика Гринберга, конечно, признает идею различных степеней родства, и применять ее предлагается для построения сложной, многоуровневой таксономии, а не для того, чтобы «доказать», что все языки мира родственны друг другу. Но сама классификация при этом строится через усреднение результатов лексических сопоставлений, ни одно из которых само по себе не оценивается на степень исторической достоверности. При рассмотрении материала в том виде, в котором он обычно представлен в работах по «массовому сравнению», для большинства «этимологий» невозможно как следует оценить ни дистрибуцию рефлексов «пракорней» по языкам-потомкам, ни систему фонетических законов, ни основные направления семантических переходов. Налицо лишь чистая статистика, без единой попытки хоть как-то минимизировать неизбежные погрешности.

Разумеется, в такой ситуации основной упрек к «массовости» «массового сравнения» — чем больше языков привлекается к сравнению, тем выше вероятность случайных совпадений — полностью оправдан. Гринберг в свое время пытался выйти из положения, подчеркивая вероятностную значимость в такой ситуации совпадений между тремя и более языками [Greenberg 1989: 112]. При этом, однако, в своих исследованиях он активно пользуется и бинарными совпадениями, а, учитывая слабость и размытость требований, предъявляемых им как к фонетике, так и к семантике сопоставляемых единиц, вопрос о реальной степени значимости

таких совпадений на самом деле не имеет простого и очевидного вероятностного решения.

Если, однако, поставить «массовое сравнение» под определенный «контроль» ряда базисных постулатов классического сравнительно-исторического языкознания, это даст возможность более или менее объективно оценивать каждое из предложенных сопоставлений с точки зрения исторической реалистичности — и, тем самым, значительно повысить общий уровень доверия к методу.

Предположим, что мы имеем дело с двумя семьями языков: семьей А, в которую входят языки А₁, А₂, А₃, А₄, А₅, и семьей В, в которую входят языки В₁, В₂, В₃, В₄, В₅. Генетическое родство языков в пределах каждой из этих двух семей очевидно и давно продемонстрировано с помощью модели РФС и этимологической реконструкции; родство семей А и В при этом, если существует вообще, является гораздо более дальним и требует первичного обоснования, не подкрепленного строгой системой РФС. При сопоставлении слов, имеющих в разных языках одно и то же значение *n* (ограничимся пока что стопроцентными семантическими совпадениями), мы получаем следующую картину:

Язык	А ₁	А ₂	А ₃	А ₄	А ₅	В ₁	В ₂	В ₃	В ₄	В ₅
Структура <i>n</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>e</i>

где *a*, *b*, *c*, *d*, *e* — пять различных «фонетических костяков», несводимых тривиальным образом друг к другу, но «фонетически совместимых» внутри себя (т. е., например, в языках А₁ и А₂ формы, выражающие значение *n*, регулярно соответствуют друг другу, а форма *n* в языке В₁ в плане фонетики похожа на них, удовлетворяя основным критериям «фонетической совместимости», перечисленным выше).

В рамках обычного «массового сравнения» — скажем, в любой из работ Дж. Гринберга или М. Рулена — «значимость» этой картины была бы выражена следующим образом:

n: А₁ *a*, А₂ *a*, В₁ *a*.

На констатации данного факта сходства этап «массового сравнения», собственно говоря, заканчивается; оно засчитывается

как статистически значимый аргумент в пользу генетического родства А и В, а шансы того, что сходство это вызвано случайным совпадением или контактно-ареальной диффузией, не оцениваются никак.

Если, однако, подключить к анализу прописные истины компаративистики, нам необходимо констатировать, что в историческом плане ближайшими потенциальными родственниками в этой схеме будут не живые языки А₁, В₁ и т. д., а их праязыки — *А и *В¹.

¹ Здесь, по-видимому, стоит вкратце остановиться на понимании термина «праязык» в рамках нашего исследования. Во-первых, на наш взгляд, историческая реальность «праязыков», корректно обоснованных в ходе компаративных исследований, не подлежит сомнению, будучи наиболее вероятным и системным объяснением совокупности наблюдаемых изоморфизмов — хотя каждый из этих изоморфизмов сам по себе может быть в той или иной степени сомнительным, т. е. любая реконструкция «праязыка» есть, разумеется, аппроксимация.

Во-вторых, «праязык» на протяжении всего нашего исследования будет пониматься как *единый* язык, диалектные различия внутри которого, если и существовали вообще, то были не столь значительны, чтобы серьезно влиять на взаимопонимание или затрагивать базисные структуры языка. Это довольно принципиальный момент, т. к. многие исследователи, склонные к переоценке роли конвергентных процессов в складывании диалектных континуумов, предпочитают заранее «закладываться» на *изначальную* диалектную раздробленность любого праязыка; к сожалению, на практике это чаще всего оборачивается тем, что на «диалектизмы» в праязыке списывается либо субъективная неспособность исследователя, либо объективная невозможность, связанная с нехваткой информации, корректно осуществить праязыковую реконструкцию, в результате чего многочисленные фонетические и грамматические нерегулярности, нагромождения немотивированных синонимов и т. п. остаются необъясненными, а у неспециалиста может сложиться иллюзия праязыка как «диалектной каши».

В плане отражения исторической реальности представляется гораздо более естественной такая модель, при которой, скажем, «праиндоевропейский» язык на некотором этапе своего развития — по мере первичного распространения его носителей по территории Евразии — мог делиться на «диалектные зоны», каждая из которых впоследствии явилась фундаментом для образования таких ветвей, как балтославянская,

Поскольку в каждой из обеих семей значение *n* выражается тремя различными способами, из этого можно заключить, что (а) *либо* в соответствующих праязыках значение *n* выражалось одной лексемой (**a*, **b* или **c* в праязыке А; **a*, **d* или **e* в праязыке В), сохранившейся в одной части языков и заменившейся в другой; (б) *либо* оно могло выражаться двумя или даже тремя лексемами, впоследствии распределившимися по отдельным языкам. (Теоретически можно предположить и вариант (в), при котором значение *n* в праязыках *А и *В выражалось четвертой, неизвестной нам лексемой, исчезнувшей во всех языках-потомках; вероятность такого сценария, однако, в обычном случае ничтожно мала, а принятие его в качестве серьезной, заслуживающей внимания альтернативы бессмыслило бы существование компаративистики как таковой).

Вариант (б), часто рассматриваемый лингвистами в практическом плане как допустимый или даже естественно ожидаемый,

индоиранская, германская, итальянская и т. п., но на *начальном* этапе развития все же должен был существовать как единый диалект относительно небольшой, компактной группы носителей. Единственная альтернатива данной модели — предположить, что уже на этапе первичного отделения от своего непосредственного предка (будь то «ностратический», «индоуральский» или какой-либо еще праязык более высокого уровня), т. е. в тот момент, когда произошло изначальное обособление индоевропейского племени (племен) и его язык начал независимую эволюцию, он *уже* был разделен на диалекты. Однако в этом случае мы, по сути, имели бы развитие не столько от «ностратического» к «индоевропейскому», сколько от «ностратического диалекта А», «ностратического диалекта В» и т. п. к «индоевропейскому диалекту А», «индоевропейскому диалекту В» и т. п.; учитывая же, что временной интервал данного развития (судя как по данным глоттохронологии, так и ряду общих соображений) должен был бы занять не менее нескольких тысяч лет, «индоевропейский А», «В» и т. п. разошлись бы за это время настолько, что о возможности единой «праиндоевропейской» реконструкции не могло бы быть и речи. Таким образом, альтернатива «монолитной» модели праязыка в виде «извечной» диалектной раздробленности представляется нам не только вредной в практическом плане (поскольку размытость предлагаемой схемы идет вразрез с необходимостью большей формализации и строгости в компаративистике), но и противоречащей здравому смыслу в чисто теоретическом отношении.

на самом деле почти невероятен. Подчеркнем, что он предполагает не «частичную» или «квази-», а *полную* праязыковую синонимию — существование в праязыке двух или более *полностью взаимозаменяемых* синонимов. Полная синонимия такого рода, когда два слова могут свободно варьировать друг с другом, не вызывая даже мелких стилистических изменений в смысле, в живых языках встречается крайне редко и, как правило, носит динамический характер, т. е. наблюдается на переходном этапе развития, когда новое слово-замена активно конкурирует со старым (см. раздел 1.6.3.2, где это явление предлагается именовать «транзитной» синонимией). Но даже в этом случае, во-первых, набор синонимов ограничивается максимум двумя лексическими единицами (ср. русск. *око* и *глаз*, англ. *hound* и *dog* 'собака', др.-китайск. *shǒu* и *tóu* 'голова'), во-вторых, сама по себе характеристика этого набора как динамического подразумевает четкое разграничение между *старым* словом (*око*, *hound*, *shǒu*) и *новым* (*глаз*, *dog*, *tóu*), т. е. неизбежность снятия наблюдаемой синонимии на более древнем этапе существования рассматриваемого языка.

Таким образом, естественной ситуацией по умолчанию необходимо считать вариант (а) — значение *n* выразилось в праязыке *А лексемой **a* или **b* или **c*, в праязыке *В лексемой **a* или **d* или **e*. Но если так, то очевидно, что значимость сопоставления «*A* *a*, *A*₂ *a* : *B*₁ *a*» будет серьезным образом зависеть от того, удастся ли показать, что **A*(*n*) = *a* и **B*(*n*) = *a*. Если окажется, что такой вариант маловероятен или невозможен, оперировать данным сопоставлением как доказательным будет нельзя.

Поясним вышесказанное на конкретном примере. Латышское *mēļns* 'черный' и др.-греческое μέλας *id.* — две лексемы, обладающие одним и тем же значением (точнее, многозначные, но пересекающиеся друг с другом в интересующем нас «стословном» значении 'черный') и, безусловно, удовлетворяющие критерию «фонетической совместимости» даже в гипотетической ситуации отсутствия твердо установленной системы РФС для индоевропейских языков. При этом, однако, более широкий, «массовый» анализ потенциально родственных форм структуры *MVL-* в других языках, привлекаемых к сравнению, показывает, что архаичность значения 'черный' для этого корня маловероятна. Так, в литовском, ближайшем родственнике латышского, однокоренные

лексемы *mélas*, *mélynas* имеют значение 'синий', а в др.-индийском та же основа *malíná-* означает 'грязный' и, более того, формально производна от *mála-* 'грязь'.

Если бы из всех индоевропейских языков были известны только латышский и греческий, для их общего предка можно было бы постулировать корень вида **mel-* с базисным значением 'черный' и засчитать данное сопоставление как серьезно значимый аргумент в пользу «греко-латышского» родства. Привлечение материала других языков уберегает нас в данном случае от антиисторической ошибки: совокупность перечисленных выше фактов говорит о том, что корень **mel-* в общем языке-предке означал не 'черный', а скорее 'грязный'; в литовском и греческом имел место вторичный семантический переход (для того, чтобы определить, протекал ли он в этих двух языках независимо друг от друга, нужны дополнительные данные по внутренней таксономии).

Приведенный пример наглядно демонстрирует важнейшую функцию «массового» принципа сравнения, обычно упускаемую из вида в работах школы Гринберга. Последний, как правило, апеллирует к «массовости» в первую очередь для того, чтобы максимально *увеличить* количество *позитивных* данных — фонетико-семантических параллелей между сравниваемыми языковыми семьями. Но «массовость» может столь же эффективно работать и на получение *негативных* данных — на оценку степени исторической надежности этих параллелей и приемлемости их как серьезных аргументов в пользу той или иной классификации.

Рассмотрим еще один, несколько более сложный тип ситуации, довольно часто встречающийся не только в работах по «массовому сравнению», но и во вполне традиционно ориентированных исследованиях по компаративистике:

Язык	A ₁	A ₂	A ₃	A ₄	A ₅	B ₁	B ₂	B ₃	B ₄	B ₅
Структура <i>n</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>c</i>

Дистрибуцию форм вида *a* и *c* по языкам семей А и В, т. е. ситуацию, когда внутри и той, и другой семьи мы обнаруживаем не менее двух совпадающих фонетических структур с общим значением *n*, мы назовем *лексическим пересечением*. Сторонники

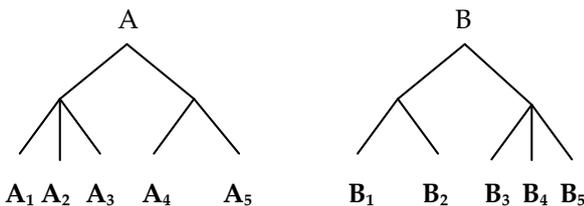
метода «массового сравнения», скорее всего, будут считать каждую из таких ситуаций благоприятной для аргументируемой классификации, представляя их в виде:

- (1) n_1 : $A_1 a, A_2 a, B_1 a$;
 (2) n_2 : $A_5 c, B_5 c$,

т. е., на первый взгляд, налицо не менее двух лексико-фонетических аргументов в пользу генетического родства семей А и В.

При этом, однако, упускается из виду одно простое обстоятельство: если для того, чтобы обосновать наш *исторический* сценарий, мы будем считать действительными *оба* аргумента, то окажемся лицом к лицу с очевидным историческим *противоречием*, будучи не в состоянии ответить на элементарный вопрос: какой же конкретно корень — **a* или **c* — выражал в языке **АВ* значение *n*?

Предположим, что в ходе изучения внутреннего устройства «бесспорных» таксономических единиц А и В мы определили для них, на основании целой совокупности данных (лексикостатистика, эксклюзивные фонетические и морфологические изоглоссы и др.), в качестве наиболее вероятных такие структуры:



В этом случае наиболее вероятным «претендентом» на выражение значения *n* в **А* будет **b*, т. к. такое решение будет являться максимально экономным: **b* сохраняется в старом значении в обоих промежуточных языках **А₁₋₃* и **А₄₋₅* и далее в современных языках *А₃* и *А₄* (по одному языку из обеих основных ветвей). Выбор в качестве первичной формы основы **a* или **c* предполагает заведомо большее количество семантических переходов; для того, чтобы его обосновать, необходимо привести серьезные исторические аргументы в пользу того, что замена старой основы на *b* в

языках A_3 и A_4 могла произойти независимо друг от друга. Аналогичным образом в стандартном случае, без наличия каких-то особых обстоятельств, для праязыка $*B$ в значении n будет восстанавливаться $*d$.

Из этого, в частности, следует, что два аргумента в пользу родства семей A и B , выдвинутых в ходе анализа их лексических данных с помощью «массового сравнения», остаются без надежного исторического обоснования — сопоставление форм структуры a и структуры c проводится на основании их семантической объединенности значением n , но если в праязыках $*A$ и $*B$ этим значением они не обладали, то сопоставление автоматически лишается семантического стержня и становится необубедительным.

Это не означает (как, возможно, поспешно заявили бы отдельные оппоненты «гринбергианского» направления в историческом языкознании), что оба сопоставления следует немедленно откинуть как ложные. Во-первых, всегда остается небольшая, но вероятность того, что наши реконструкции $*A(n) = *b$ и $*B(n) = *d$ все же не отражают историческую реальность, даже несмотря на методологическую безукоризненность их получения (даже максимально экономное и учитывающее все релевантные факты решение может оказаться неистинным). Во-вторых, даже если наше решение исторически верно, формы a и c в дальнородственных семьях A и B все равно *могут* восходить к одному и тому же корню праязыка $*AB$, с той лишь особенностью, что значение n появилось у них в этих семьях независимо друг от друга (эта проблема будет разобрана подробнее в следующем разделе).

Тем не менее, из этого все же следует, что соответствующие сопоставления — ни то, ни другое — нельзя рассматривать как *сильную* аргументацию в поддержку исторической реальности семьи AB . Здесь опять же проявляется *корректирующая* роль массового сравнения: рассмотрев значение n во *всех* языках исследуемых семей, а не в избранных, мы получаем для него диахроническую перспективу, которая позволяет оценить вероятность случайных совпадений между A и B не через абстрактную статистическую модель, а через возможность или невозможность достоверно «вписать» родство форм вида a или форм вида c в конкретный исторический сценарий.

Переход от бинарного к многостороннему сравнению, таким образом, не только полезен, но и практически неизбежен, если нашей целью действительно является максимальное приближение к исторической реальности, стоящей за предлагаемыми нами классификациями. Проблема реконструкции лингвистической предыстории для того или иного ареала не может быть успешно решена за счет отдельных бинарных сопоставлений, своего рода «семплирования», аналогичного процедурам, применяемым для создания типологических классификаций языков. Но и принцип «массовости», в противовес позиции Гринберга, не может и не должен применяться на чисто поверхностно-статистической основе, когда ни экспертное сообщество, на суд которого выносятся исследование, ни даже сам исследователь не имеет ни малейшего представления о возможных исторических сценариях, стоящих за каждым из его сопоставлений. Поэтому наш анализ, заимствуя некоторые из методологических положений Гринберга, будет в первую очередь стремиться к тому, чтобы поставить их на более прочную, фальсифицируемую основу, в рамках своего рода синтеза элементов «массового сравнения» со стандартной методологией, принятой в компаративистике.

1.6. Лексикостатистика: к проблеме оптимизации метода.

На сегодняшний день, по-видимому, можно считать общим местом высокую значимость базисной лексики в обосновании языкового родства; разногласия между исследователями сводятся, в основном, к вопросам о том, должна ли базисная лексика считаться «главным» или «достаточным» критерием, а также о том, что конкретно следует понимать под этим термином (включая, например, вопрос об универсальности базисной лексики, вопрос об идеальном составе 100- или 200-словного списка и т. д.). Гораздо сложнее обстоит дело со статистическими методами анализа базисной лексики.

Как собственно лексикостатистика (метод построения относительной классификации языков одной семьи на основании анализа процентов этимологических сходжений между ними в пределах заданного списка базисной лексики), так и ее усложненный вариант, глоттохронология (метод построения абсолютной клас-

сификации языков, определяющий конкретные даты языкового распада) основаны на принципиально недоказуемой презумпции неизменной (в классической версии Сводеша) или закономерно меняющейся (в ряде модификаций) скорости замещения лексических единиц в пределах списка¹. Соответственно, все специалисты в области исторического языкознания, так или иначе обозначившие свое отношение к данной проблематике в своих работах, подразделяются на три категории:

а) те, кто более или менее категорично отвергает как глоттохронологию, так и лексикостатистику вообще, не видя между ними принципиальных различий;

б) те, кто признает ценность лексикостатистической классификации для построения языкового древа, но отказывается привязывать ее к конкретным хронологическим датам;

в) те, кто убежден в перспективности глоттохронологии как надежного и объективного инструмента лингвистического датирования (хотя убежденных сторонников глоттохронологии в ее первоначальном, «сводешевском» варианте сегодня, по-видимому, уже не осталось).

Первую позицию занимает, в частности, Л. Кэмпбелл, трактующий оба термина как синонимы [Campbell 2004: 201] и подвергшийся за это справедливой и конструктивной критике, например, в работе [Sidwell 1999] (критика относится к первому изданию учебника Кэмпбелла, но осталась неуцтенной и во втором). Тем

¹ Подробное изложение как сути лексикостатистического метода, так и исторического развития всей соответствующей области выходит далеко за рамки нашего исследования. Отметим лишь, что ключевыми работами по этой теме до сих пор остаются первичные труды М. Сводеша [Swadesh 1952, 1955]; большая статья С. А. Старостина, в которой впервые вводятся значительные поправки к «классической» лексикостатистике, лежащие в основе всей дальнейшей деятельности Московской школы компаративистики [С. Старостин 1989]; а также двухтомник [Renfrew et al. 2000], с представительной подборкой статей как сторонников, так и противников глоттохронологического анализа. Детальное описание всей истории развития лексикостатистики до середины 1980-х гг. с обширной библиографией можно найти в работе [Embleton 1986]; новейшие тенденции и достижения обсуждаются в коллективной монографии [Renfrew et al. 2006].

не менее, нельзя все же не признать, вслед за Кэмпбеллом, что даже в отрыве от глоттохронологических формул лексикостатистика, претендующая на построение только относительной, а не абсолютно-хронологической, классификации языков, все равно имеет серьезный практический смысл только при том условии, что исследователь признает закономерный, а не хаотический, характер скорости изменений базисного лексического состава.

Так, если между языками А и В насчитывается 50% совпадений в 100-словном списке Сводеша, а между парами [А, С] и [В, С] — порядка 35%, это в принципе позволяет постулировать языковую семью из трех членов со вполне определенной историей разделения ([1] *АВС → *АВ и С; [2] *АВ → А и В) и при этом отказаться датировать эту историю, ссылаясь на недостоверность имеющихся глоттохронологических моделей — постольку, поскольку мы не доверяем идее постоянной скорости «распада» 100-словного списка. Но если такого доверия нет, то, вообще говоря, ничто не мешает предположить, что, например, все три языка А, В и С отделились друг от друга абсолютно одновременно; что постулированный на основании лексикостатистики промежуточный узел *АВ абсолютно фиктивен; и что заниженный процент совпадений между парами [А, С] и [В, С] (35%) объясняется всего лишь тем, что в какой-то момент скорость изменения базисного лексического фонда в языке С серьезно увеличилась по сравнению с языками А и В (например, за счет каких-то неизвестных нам социальных потрясений).

Таким образом, лексикостатистика без глоттохронологической поддержки может существовать в лучшем случае как один из возможных классификационных критериев, удобный за счет своего формализованно-унифицированного характера, но безусловно вспомогательный по отношению к «нетривиальной» методике поиска общих инноваций на фонологическом и морфологическом уровнях. Устранение идеи «глотточасов» (англ. *glotto-clock*), неприемлемой для многих исследователей скорее исходя из их теоретических представлений о природе языковых изменений, чем из эмпирических данных, в целом скорее обесмысливает лексикостатистический метод, чем совершенствует его.

В рамках нашего исследования, для которого лексикостатистика играет центральную роль, мы будем поэтому, если и не

использовать оба термина как полные синонимы, но по умолчанию все же считать, что генеалогическое древо любой языковой семьи, построенное на лексикостатистической основе, имеет полное право не только отображать дистанцию между своими узлами через различия в их длине (относительная классификация), но и сопровождать каждый из узлов конкретной глоттохронологической датировкой (абсолютная классификация).

Тем не менее, для того, чтобы с уверенностью пользоваться лексикостатистическим критерием, возлагая на него основные надежды в плане построения исторически корректного генеалогического древа, следует вкратце обсудить хотя бы основные критические замечания в его адрес, в свое время затормозившие общественный интерес к этому методу и с тех пор регулярно «перекочевывающие» из одной обзорной работы в другую. (Многочисленные конструктивные ответы на эти замечания, публиковавшиеся в работах С. А. Старостина, А. Ю. Милитарева, И. И. Пейроса, П. Сидвелла и других убежденных «лексикостатистов», к сожалению, в работах их оппонентов чаще всего игнорируются¹).

Большинство таких замечаний можно, по-видимому, разделить на три типа: (а) *сущностно-теоретические*, подвергающие сомнению базисные принципы лексикостатистики в силу того, что их существование невозможно объяснить теоретическими причинами; (б) *сущностно-эмпирические*, пытающиеся опровергнуть универсальность постулатов Сводеша и его последователей на конкретных языковых примерах; (в) *технические*, в рамках которых всячески подчеркиваются трудности практического

¹ В качестве конкретного примера можно указать критическую работу [Bergsland & Vogt 1962], якобы продемонстрировавшую несостоятельность теоретических постулатов М. Сводеша на «классическом» примере скандинавских языков, скорость изменения базисной лексики в которых варьирует от 4% за тысячу лет (в исландском) до 20% (в литературном норвежском). Несмотря на то, что аргумент этот был удовлетворительно дезавуирован С. А. Старостиным еще в 1989 г. (см. ниже), ссылки на работу Бергсланда и Фогта до сих пор фигурируют почти в каждой исследовательской работе, содержащей критику глоттохронологического метода, так, как будто она до сих пор представляет собой последнее слово в соответствующей дискуссии.

характера, связанные с составлением конкретных списков и критериями их анализа. Разберем каждую из этих групп отдельно.

1.6.1. *Теоретические проблемы лексикостатистики.* В наибольшей степени недоверие к глоттохронологии, по-видимому, подпитывается соображениями чисто теоретическими — недоумением относительно возможных причин такого явления, как пресловутые «глотточасы». Ср. у Л. Кэмпбелла: «There is nothing inherent in the nature of vocabulary (or in the organisation of the lexicon) which would lead us to suspect any sort of regular pattern to lexical change, certainly not that basic vocabulary should be replaced everywhere at the same rate» [Campbell 2004: 207]. Подобного рода «механистический» подход к языковым изменениям регулярно становится объектом жесточайшей критики со стороны исследователей — ср., например, у К. Титера:

«...one can get lexical items by simple eliciting, without the use of inference; for the lexicon is nothing but the outward face a language turns to its associated culture. Words occur in nature, as grammars do not... words are on the surface of language, and words may be freely added or dropped as the culture changes. Lexical similarities and dissimilarities do not come about in any one simple way, and any mechanical method of counting lexical similarities cannot separate those due to chance, universals, diffusion, and common origin. Lexical change is the result of many factors, and all are scrambled together in the final result» [Teeter 1963: 641].

Здесь в первую очередь стоило бы отметить, что историческое языкознание — наука, основанная прежде всего на учете и анализе эмпирических данных, предоставленных в распоряжение исследователя языковым разнообразием планеты, и подобного рода умозрительные рассуждения, сколь бы они ни казались естественными с точки зрения «общего здравого смысла», не могут считаться приоритетными на фоне конкретных статистических результатов; никакая теория языковых изменений не может строиться, исходя из принципа «тем хуже для фактов».

С другой стороны, нельзя отрицать и то, что теорию стабильного изменения базисной лексики нельзя в полном смысле этого слова называть «теорией», если она обладает исключительно описательной, но не объяснительной силой. До тех пор, пока идея «глотточасов» сама по себе заставляет недоумевать

относительно ее истоков, глоттохронология вряд ли может претендовать на всеобщее признание. Это, по-видимому, отчетливо осознавал и сам М. Сводеш, поскольку уже в самых первых своих публикациях на тему лексикостатистики он пытался найти резонный ответ на этот вопрос. Ср., например:

«Язык является чрезвычайно сложной системой символов, выполняющих жизненно важную коммуникативную функцию в обществе. Эти символы подвержены изменениям благодаря влиянию многих обстоятельств, однако они не могут изменяться слишком быстро, не нарушая общепонятности языка... Причины, обуславливающие языковые изменения, без сомнения, разнообразны... в то время как язык подвержен разнородным импульсам, побуждающим его к изменению, он должен все же сохранять значительное единообразие... У старшего и младшего поколений часто наблюдаются различия в лексике и в употреблении слов, но эти различия никогда не достигают таких размеров, которые явились бы причиной для нарушения взаимной понимаемости языка. Это обстоятельство максимально ограничивает скорость изменения языка» [Сводеш 1960: 44-45].

Таким образом, по мнению М. Сводеша, лексические изменения, в том числе внутри базисной лексики, с одной стороны (а) неизбежны, т. е. обеспечивают наличие *нижнего предела скорости* (НПС) лексических изменений, с другой — (б) ограничены необходимостью сохранения взаимопонимания между представителями старшего и младшего поколений, т. е. подразумевают наличие аналогичного *верхнего предела скорости* (ВПС). Вряд ли оба эти постулата можно оспорить, однако идея «глотточасов» сама по себе не вытекает автоматически из их принятия. Здесь возникают дополнительные вопросы:

(1) почему об НПС/ВПС уместно говорить только по отношению к лексике (точнее — к базисной лексике), а не к другим языковым уровням (фонетике, морфологии и т. д.)?;

(2) для того, чтобы лексикостатистика давала исторически приемлемые результаты, необходимо, чтобы интервал между НПС и ВПС был минимальным; в противном случае погрешности, накапливаемые за промежуток в тысячу (две, три, пять и т. д. тысяч) лет, обесмысливают применение метода как такового.

Откуда берется убежденность в том, что такой интервал действительно минимален, т. е. близок к нулю¹;

(3) как относиться к проблеме массовых заимствований, в ходе которых за небольшое время значительно преобразуется как культурный, так и базисный лексические фонды (ср. такие случаи, как албанский язык в индоевропейской семье, брахуи в дравидийской, отдельные представители тай-кадайской семьи и т. п.)? Не доказывают ли такие случаи абсурдность постулирования самой идеи ВПС?

Проще всего, по-видимому, ответить на первый вопрос. Множество лексических элементов языка — для простоты можно даже понимать под «лексическими элементами» множество корневых морфем, имеющих знаменательные значения — отличается от множества фонем или грамматических морфем языка не только значительно бóльшим размером, но и значительно меньшей степенью упорядоченности и систематизированности. Для фонетики и морфологии не обязательны, но типичны *системные* изменения, затрагивающие сразу много элементов (например:

¹ К. Эрет в специальной работе, посвященной глоттохронологии, утверждает, что защищать этот метод следует исходя не из постулата о постоянной скорости распада списка (т. е. «НПС = ВПС»), а из того, что лексикостатистика определяет систематические тенденции накопления случайных изменений («the patterned accumulation of individually random change among quanta of like properties» [Ehret 2000: 373]), что подразумевает сколь угодно большие расхождения между НПС и ВПС в рамках одного или нескольких поколений, но сглаживание общей кривой на значительных временных интервалах.

Эта позиция в целом справедлива, резонна и гораздо лучше согласуется с реальными данными, чем утверждение о «незыблемой» скорости распада без каких-либо погрешностей, но при этом она (а) скорее просто эксплицитно выражает то, что и так было имплицитно очевидно большинству сторонников метода, нежели представляет собой теоретический прорыв; (б) все равно не отвечает на вопрос, каких размеров может достигать ВПС — ведь чем выше скорость распада на небольшом временном интервале, тем больше потребуется времени для того, чтобы «выровнять кривую», и если допустить, что в каком-то отдельно взятом случае ВПС достигнет, например, 15-20 элементов списка за сто лет, вместо ожидаемых 1-2, любые совместные подсчеты по этому языку в контексте даже его дальних «родственников» будут выдавать ошибочные результаты.

палатализация *всех* дентальных или велярных согласных; редукция и отпадение *всех* конечных гласных в структуре *CVCV; исчезновение *всей* системы падежного склонения; развитие *полностью* новой системы глагольного спряжения в ходе грамматикализации местоимений и т. п.). Напротив, для базисной лексики, элементы которой не связаны друг с другом парадигматическими отношениями, такие массовые изменения маловероятны — в большинстве случаев «реакция» не будет носить «цепной» характер из-за высокой степени автономности заменяемых элементов¹.

Следует также отметить, что существенные структурные изменения в фонетике и грамматике, которые могут произойти от одного поколения к другому, не будут оказывать столь же пагубное влияние на взаимопонимание между поколениями, как аналогичный массивный сдвиг в области лексики; «молодой» язык, заменяющий один тип артикуляции той или иной группы фонем на другой, или утративший ту или иную именную или глагольную категорию, останется более понятным для носителей «старого» языка, чем «молодой» язык, заменивший повышенное количество базисной лексики.

Ускоренный «крах» фонетической и морфологической системы языка естественно ожидать также в ситуациях, когда престижный язык одной группы населения распространяется среди других, ранее иноязычных групп. Лексика престижного языка в таких случаях обычно перенимается легче и в более полном объеме, чем фонетика и грамматика (ср. типичный процесс «пиджинизации», хотя в данном случае подразумеваются скорее структурно-типологические изменения, например, латинского или китайского языков под влиянием перехода на них изначально нелатино- и некитаеязычного населения). Разумеется, во

¹ Данное утверждение, разумеется, не означает, что то или иное лексическое изменение всегда будет единичным; хорошо известны случаи параллельного семантического развития внутри небольших групп слов, объединенных общими семантическими компонентами, напр. 'рука (выше кисти)' → 'кисть (руки)' одновременно с 'нога (выше стопы)' → 'стопа (ноги)', или 'видеть' → 'глаз' одновременно с 'слышать' → 'ухо'. Однако такие случаи являются скорее исключениями, чем правилами, и почти никогда не затрагивают группу более чем из двух элементов (по крайней мере, в пределах 100-словного списка).

многих подобных случаях «новый» язык оказывается все же отягощенным «субстратным» лексическим наследием — однако соответствующие элементы можно вычлениить с помощью этимологического анализа и вычесть из статистики как «внешние замены» (см. ниже).

Таким образом, нет ровным счетом ничего удивительного в том, что из всех уровней языка на роль «глотточасов» лучше всего подходит именно базисная лексика — и вряд ли случаен тот факт, что никому на сегодняшний день не удалось построить успешную модель, аналогичную лексикостатистической, путем калибровки скорости фонетических или морфологических изменений¹.

Гораздо труднее дать убедительный ответ на вопрос (2), который, по-видимому, еще надолго останется основным препятствием на пути теоретического обоснования глоттохронологии.

Как известно, М. Сводеш в своих ранних работах определил коэффициент λ (константу скорости распада списка) как 0.14, т. е. постулировал замену в среднем 14 слов из 100 за период в 1000 лет. С. А. Старостин, установив завышенность этого числа как следствие неразличения Сводешем «контактных» и «автономных» замен (т. е. иноязычных заимствований и замещений из собственного лексического фонда), снизил его до 0.05, что в целом лучше согласуется с калибровочными данными.

Приняв за условное «поколение» период в 30 с небольшим лет, будем считать, что тысячелетнему периоду непрерывного развития языка соответствует примерно 30 поколений; в этом случае коэффициент Сводеша предполагает утрату одного элемен-

¹ Из известных нам попыток такого моделирования можно отметить работу П. Хеггарти, попытавшегося — разумеется, безуспешно — замечать скорость фонетических изменений на материале романских языков [Heggarty 2000]. Крайне любопытную и во многом инновативную «морфохронологию» для языков тюркской семьи, в рамках которой 100-словный список Сводеша заменен на «вопросник» из 83 фонетических и морфологических критериев, предлагает в недавней работе О. А. Мудрак [Мудрак 2009]; к сожалению, откалибровать этот вопросник на материале других семей невозможно, т. к. он ориентирован на историческое развитие именно тюркских языков, а попытка универсализировать его неизбежно привела бы к подмене сегментно-фонетическо-морфологического анализа типологическим, что неприемлемо.

та списка не менее чем за два поколения, а коэффициент Старостина — не менее чем за шесть поколений. Трудно представить себе, что коммуникативная связь между поколениями оказалась бы нарушена, если бы, например, с каждым новым поколением замещалось всего одно слово из списка. Но любая ситуация такого рода губительна для лексикостатистического метода: установив для НПС значение «1 слово за 1 поколение», а для ВПС — старостинское значение «1 слово за 6 поколений», мы можем уже за первую тысячу лет независимого развития получить до 25-30% замен в одном списке по сравнению всего с 5% замен в другом.

Таким образом, теоретическая опора на абстрактный принцип «сохранения удобства коммуникации» сама по себе недостаточно обосновывает лексикостатистику как надежный метод классификации. Интуитивно кажется очевидным, что скорость изменения 100-словного списка, равная «10 словам за 1 поколение», относится к области фантастики, но совершенно не обязательно утрировать ситуацию — даже относительно незначительные, интуитивно «безобидные» скоростные колебания между родственными языками вполне достаточны для дискредитации лексикостатистического анализа, при условии, что они сохраняются на протяжении достаточно длительного периода.

По-видимому, постоянность скорости базисно-лексических изменений все же обусловлена не столько преемственностью между поколениями, сколько самой природой этих изменений. Конкретные причины выпадения тех или иных слов из списка и замены их на новые могут различаться очень сильно, от табуирования «неблагоприятных» элементов до тривиального снятия омонимии, но в качестве основной причины «по умолчанию» все же следует, наверно, видеть связанные факторы *полисемии* (в нашем случае — ситуация, когда элемент 100-словного списка имеет одно или несколько значений помимо 100-словного) и *синонимии*, точнее — *квази-синонимии* (когда существует одно или несколько слов со значением, близким к значению элемента 100-словного списка)¹.

¹ Недопустимо, на наш взгляд, рассматривать в качестве основной причины эволюции базисной лексики «социально-исторический»

Как полисемические, так и квази-синонимические отношения между отдельно взятой лексемой X и, с одной стороны, ее переносными значениями $X_1, X_2... X_n$, с другой — ее квази-синонимами $Y_1, Y_2... Y_n$, постепенно накапливаются в языке от поколения к поколению, причем каждое образование такого нового значения или квази-синонима повышает вероятность замены X на одну из лексем ряда Y . «Средняя устойчивость» того или иного элемента X 100-словного списка может, таким образом, определяться как средняя скорость, с которой этот элемент «обрастает» переносными значениями и квази-синонимами.

Так, например, личные местоимения 'я', 'ты', 'мы' обычно отличаются высокой стабильностью, поскольку для них нехарактерны ни появление переносных значений, ни квази-синонимия; исключение — языки с категорией вежливости (Юго-Восточная Азия и ряд других регионов), где для этих местоимений часто встречаются богатые синонимические ряды, заметно снижающие их общий индекс стабильности. Для таких частей тела, как 'глаз'

фактор, как это делает, например, Г. Хольм [Holm 2007] и ряд других исследователей; согласно этой позиции, состояние базисной лексики, как и других слоев языка, зависит в первую очередь от конкретных и, как правило, непредсказуемых изменений в жизни общества, откуда естественным образом следует невозможность постулирования каких-либо универсальных констант. На самом деле эта теория является демонстративно верной только для лексических замен внешнего происхождения (см. ниже); для подавляющего большинства ситуаций внутренних замен обосновать ее невозможно. Какими историческими потрясениями или социальными переменами могла быть обусловлена, например, замена праиндоевропейского корня **ous-* 'ухо' на *klots* (от и.-е. **kley-* 'слышать') в тохарском А, или древнеанглийского *fugel* 'птица' (сохранилось в совр. англ. как *fowl* в узком значении 'домашняя птица') на совр. *bird* (исначально 'потомство' → 'птенец')?

В определенных случаях можно, действительно, предполагать наличие фактора «ареального калькирования», когда семантические механизмы деривации, естественные для одних языков, переносятся и на их географических соседей (ср. развитие 'ухо' → 'лист' = 'ухо дерева', типичное для североафриканского региона). Однако практика показывает, что такие ситуации находятся в явном меньшинстве по отношению к индивидуальному «поведению» языков или языковых групп, необъяснимому за счет воздействия каких-либо внешних стимулов.

или 'ухо', обычно характерно обилие переносных значений (в основном через метафорические развития: 'глазок' как 'отверстие' или 'нарост, пятнышко' по отношению к самым разным объектам, 'ушко' примерно с такой же семантикой и т. п.), но квази-синонимия для этих слов обычно довольно ограничена, и заменяются они, как правило, на сильно маркированные жаргонизмы. Напротив, для цветообозначений квази-синонимия чрезвычайно распространена в самых разных (возможно, во всех) языках мира — это связано с возможностью сложной и разнообразной детализации цветовой палитры, в ходе которой частотные, «базовые» цветообозначения постоянно находятся под угрозой замещения одним из многочисленных «второстепенных» терминов.

Изучение такого рода парадигматических связей базисной лексики показывает, что в целом полисемические и квази-синонимические ряды, образуемые этими лексемами, склонны очень сильно пересекаться или даже полностью совпадать в самых разных языковых ареалах. Это наглядно следует из того факта, что для каждого значения X , представленного в рамках 100-словника, можно выделить «типовые» смежные значения $X_1, X_2... X_n$, которые по отношению к значению X могут быть производящими (т. е. слово, ныне обладающее значением X , до этого обладало одним из значений $X_1, X_2... X_n$) или производными (т. е. слово, обладавшее значением X , приобрело одно из переносных значений $X_1, X_2... X_n$, причем последнее либо вытеснило значение X , либо продолжает сосуществовать с ним); перечень некоторых таких «типовых» значений для 50-словной части 100-словного списка будет приведен ниже.

Не вдаваясь в сложную (хотя и чрезвычайно любопытную и полезную) дискуссию относительно того, отражает ли наличие такого рода типовых семантических связей какое-то «глубинно-универсальное» членение семантического пространства, отметим лишь следующее: если само развитие базисной лексики по языкам мира обнаруживает в целом сходные тенденции, вполне естественно предположить, что и *скорость* этого развития по языкам мира также будет сходной. В самом деле, если скорость зависит от накопления парадигматических семантических связей, а связи эти во многом универсальны, то и накапливаться они будут более или менее одинаковым образом.

Разумеется, в каждом отдельно взятом языковом регионе будет наблюдаться своего рода «люфт», т. е. отдельные элементы базисной лексики, ведущие себя неожиданным образом — в частности, обнаруживающие непредсказуемые семантические связи или, наоборот, полностью лишённые самых обычных связей такого рода. Однако эти ситуации все же будут исключительными: опыт работы с базисной лексикой самых разных семей в рамках Московской школы компаративистики пока что не выявил ни одного языка или языковой семьи, базисная лексика которых систематически, а не в исключительных случаях эволюционировала бы по каким-то уникальным, не имеющим аналогов закономерностям.

Данная гипотеза носит предварительный характер и нуждается в обстоятельном тестировании, как теоретически-формальном, так и эмпирическом; к сожалению, подробные исследования, посвященные эволюции базисной лексики в языках, история которых известна нам в рамках хотя бы двух-трех тысяч лет, пока отсутствуют, но хотелось бы надеяться, что это всего лишь вопрос времени. В любом случае вряд ли подлежит сомнению тот факт, что теоретическая дискуссия о скорости изменения базисной лексики не может проводиться в отрыве от анализа непосредственных механизмов этого изменения¹.

Остается ответить на третий вопрос: как соотносится неоспоримый факт возможности массовых заимствований в области базисной лексики и даже непосредственно в 100-словный список Сводеша с идеей предсказуемо закономерной скорости лексических изменений?

¹ Представление о «нарастающей полисемии» как об основном, или, по крайней мере, одном из наиболее значительных «двигателях» лексических изменений (хотя и не как об основной причине постоянной и универсальной скорости этих изменений), подробно обсуждается в работе С. А. Старостина [2005], описывающей автоматический алгоритм симуляции этого процесса. То, что получающиеся в ходе симуляции результаты оказываются почти неотличимыми от результатов обычных глоттохронологических подсчетов, косвенно подтверждает справедливость высказанной выше идеи: скорость распада списка напрямую зависит от количества побочных семантических развитий у его элементов, развития же в разных языках обусловлены схожими механизмами.

Условимся разделять все наблюдаемые и гипотетические замены в 100-словном списке на *внутренние*, т. е. не связанные с межъязыковыми контактами, и *внешние*, т. е. заимствования. Тогда все высказанные выше соображения относительно механизмов и скорости языковых изменений по умолчанию относятся только к внутренним заменам; для внешних, очевидным образом, требуется вводить специальные коррективы — так, согласно методике, предложенной С. А. Старостиним, опознанные заимствования необходимо вообще исключать из статистических подсчетов во избежание заведомо искаженных результатов¹.

Опыт сопоставительного изучения лексики языков мира показывает, что массовые заимствования в рамках 100-слового списка (т. е. превышающие, скажем, 10% от общего числа элементов) — явление редкое, не затрагивающее даже такие хрестоматийные случаи языковых контактов, как английский с французским (2 заимствования: *mountain* 'гора' и *round* 'круглый'), японский с китайским (3 заимствования: *shinzō* 'сердце', *kanzō* 'печень', *niku* 'мясо') и т. п. Тем не менее, случаи массовых заимствований все же обнаруживаются в самых разных ареалах — ср. многочисленные китаизмы в чжуан-тайских языках, латинизмы в албанском, индоаризмы и арабизмы в брахуи и т. п.; аналогичные ситуации, как будет показано ниже, встречаются и в языках Африки.

Внешние замены такого рода уже не могут быть объяснены «естественными» причинами, такими, как развитие полисемии и

¹ На самом деле открытым остается вопрос, необходимо ли элиминировать *все* без исключения заимствования или поступать подобным образом только в ситуациях, когда налицо массивированные заимствования в 100-словник. Так, нередко встречаются случаи, когда слово заимствуется в язык в «нестословном» значении, какое-то время «обживается» в нем, и только потом меняет свое значение на стословное (ср., например, название 'печени' в романских языках: итальянск. *fegato*, испанск. *higado*, французск. *foie* и т. п. ← народн. латинск. **fīcātum* 'печень', заимствованное из греческ. σικωτόν в сочетании ἥπαρ σικωτόν 'печень, начиненная фигами' [Meyer-Lübke 1972: 699]; вряд ли будет корректным считать замену классического латинского *iesur* 'печень' на **fīcātum* внешней, т. к. заимствовано это слово было в «нестословном» значении '/приготовленная/ печень животного'). К сожалению, далеко не всегда мы располагаем подобного рода информацией.

синонимии. Их обусловленность носит социальный характер, т. к. они вызваны определенным рода психологическим давлением со стороны языка-донора. В этих условиях допустимо и легко объяснимо нарушение условия «преемственности поколений», т. к. ситуации подобного рода по большей части сводятся к следующим двум типам:

(а) *Субстратный*. Язык-донор в социальной среде вытесняется языком-акцептором с сохранением части своего лексического фонда — например, при переходе на новый доминантный язык завоеванного или культурно «подавленного» народа (в целом), или женской части этноса, в массе своей выходящей замуж за этнически и лингвистически отличающихся представителей соседних племен. В этой ситуации «преемственность поколений» не имеет значения, т. к. массовый переход на чужой язык происходит в первом поколении, не имеющем прецедентов, и здесь налицо все условия для серьезной «языковой ломки» — создания нового языка с элементами смешанности. Субстратные элементы, как в фонетике и грамматике, так и в базисной лексике в этом случае имеют своей целью «облегчить» для новоявленного носителя языка процесс овладения последним. Типичным примером субстратной ситуации такого рода можно считать язык бай в китайской провинции Юньнань, предки современных носителей которого когда-то, по-видимому, говорили на языке, представлявшем собой отдельную ветвь тибето-бирманской семьи, но уже в начале I тыс. н. э. массово перешли на китайский язык, сохранив, однако, небольшой «сверхдревний» слой лексики некитайского происхождения¹.

(б) *Суперстратный*. Элементы языка-донора, в том числе и относящиеся к базисной лексике, активно заимствуются языком-акцептором как «престижные», «культурные», «модные» и т. п., но не приводят при этом к полной смене языковой парадигмы.

¹ Подробнее о ситуации в бай см. монографию [Wang 2006]; в статье [Yeon-Ju, Sagart 2008] предпринята попытка переопределить таксономический статус этого языка как тибето-бирманский, исходя из наличия 12 элементов «сверхдревнего» субстратного слоя в 100-словном списке, однако в статистическом плане уместнее все же квалифицировать эту ситуацию как смену языка (тем более, что и для указанных 12 элементов авторы не в состоянии предложить единый источник происхождения).

Необычный, но вряд ли уникальный, пример в этом отношении представляет дравидийский язык брахуи, лексика которого по разным подсчетам содержит от 50% до 80% слов, заимствованных из иранских (белуджский, персидский, пушту) и индоарийских (синдхи, ленда, панджаби, урду) языков [Андронов 1971: 33-35]; в пределах стословного списка брахуи обнаруживает сильно меньший, но все же значительный процент заимствований — ок. 25%. При этом, однако, о смене языка в этой ситуации говорить не приходится: и базисная лексика, и грамматика брахуи как минимум на три четверти сохраняют дравидийское происхождение, а заимствования из основных языков-доноров распределены настолько равномерно, что перспектива постепенной «трансформации» брахуи в какой-то один из них маловероятна (гораздо более вероятно в не столь отдаленном будущем вымирание брахуи как такового и переход его носителей на разные престижные языки тех ареалов, в которых они проживают).

При таком обилии заимствований «преемственность поколений» неизбежно будет нарушена, причем нарушение это может носить сознательный характер: молодое поколение намеренно дистанцируется, в той или иной степени, от языка своих родителей. Подчеркнем, что при внутренних заменах такое дистанцирование вряд ли возможно — для него требуется наличие активного стимула, каковым и является соседство «престижного» языка.

Таким образом, принципиален здесь тот факт, что языковые механизмы, ответственные за внешние лексические замены (заимствования), носят кардинально иной характер по сравнению с механизмами внутренних замен и в общем случае не должны быть задействованы в ходе постепенного, «внеконтактного» изменения 100-словного списка. Следовательно, при условии грамотного отделения внешних от внутренних изменений в ходе анализа материала и использования только данных внутренних изменений для статистических подсчетов можно рассчитывать на довольно высокую надежность полученных таксономических и хронологических выводов. О конкретных критериях такого отделения мы будем говорить ниже.

Отдельно стоит остановиться на мнении ряда специалистов, согласно которому сама концепция «базисной лексики» не выдерживает критики постольку, поскольку не существует формальных

оснований, по которым можно было бы отделить «базисные» элементы словарного инвентаря от «культурных». Чаще всего при этом в качестве главного «претендента» на такое формальное основание называется способность / неспособность слова быть заимствованным, вслед за чем на материале специально подобранных примеров провозглашается, что незаимствуемых слов на самом деле нет. Из известных нам работ наиболее последовательно эта позиция озвучена в монографии [Haarmann 1990a], ср.:

«The 'basicness' of certain sectors of the vocabulary in natural languages cannot be substantiated because the notion is itself a misconception. An effort has to be made to eradicate the idea of a so-called precultural strata of concepts which have been assumed for setting up the list of lexical items on which lexicostatistics is based... The most crucial factor in glottochronology, arguably, is the assumption that there are sectors of the vocabulary in any language which are more resistant to borrowing than other sectors. It has never been made explicit why certain sectors of the lexicon should be, predominantly or exclusively, comprised of cognates. Among the diffuse implications of the notion of a 'basic' vocabulary is the idea that denominations for concepts which are elements of universal human experience are not easily replaced in a given speech community. Borrowings in the assumed 'basic' sector of the lexicon are considered by glottochronologists as rare exceptions, whereas their occurrence is in fact a much more common phenomenon in processes of language contacts than is acknowledged even by very critical opponents of lexicostatistics» [pp. 150-151].

За этим введением следует большой перечень конкретных примеров заимствований в разных языках лексических элементов, относящихся к «базисной» сфере: частей тела, природных явлений, числительных, местоимений и т. п., призванный неопровержимым образом подтвердить правоту автора. При внимательном анализе его теоретических рассуждений трудно, однако, не усмотреть в них ряд элементов казуистики, раскрытие которых делает его позицию не столько уязвимой, сколько нерелевантной для решения «глотнохронологической проблемы».

Так, очевидно, что Г. Хаарманн в целом негативно относится к идее того, что одни лексические группы в языках мира более устойчивы к заимствованиям, чем другие — в частности, из-за

отсутствия ее надежного теоретического обоснования. Вместе с тем ни здесь, ни ниже по тексту мы не видим эксплицитного утверждения о том, что эта идея является *ложной*. Вместо этого предлагается просто обратить внимание на то, что «базисная» лексика заимствуется *чаще*, чем это (якобы) обычно предполагается сторонниками (и даже отдельными противниками) глоттохронологии. Закономерный вопрос — *насколько* чаще? Г. Хаарманн не приводит никакой статистики, хотя было бы интересно, например, подсчитать количество документально засвидетельствованных или надежно реконструированных случаев заимствования таких слов, как 'глаз' или 'вода' по сравнению с числом аналогичных случаев для слов 'очки' или 'запруда'. Возможно, создание такой статистики хотя бы для небольшого подмножества элементов 100-словного списка помогло бы раз и навсегда поставить точку в вопросе о наличии или отсутствии противопоставления между «базисной» и «культурной» лексикой.

При этом, разумеется, никто не отрицает, что заимствования в 100-словный список, в том числе и массированные, действительно бывают¹, и что заимствованным теоретически может

¹ Нельзя, впрочем, обойти молчанием тот факт, что конкретные списки заимствованной базисной лексики, представленные Г. Хаарманном, зачастую искажают саму суть лексикостатистического метода, а кое-где содержат просто грубые ошибки или грешат нарочитым субъективизмом. Так, индонезийское *lailah* 'ночь', действительно заимствованное из арабского (стр. 186), никак не является в этом языке базисным словом, а употребляется только в «церемониальном» значении (основное слово — *malam* — имеет исконное происхождение). Аналогичным образом «персидские» вопросительные местоимения-арабизмы *man* и *ma* 'кто', 'что' даже не упоминаются в стандартных словарях этого языка, в отличие от обычных *kī*, *čī* исконно индоевропейского происхождения (*ibid.*). На той же странице обнаруживается финское *moni* 'много', якобы заимствованное из «Indo-European» (*sic!*), при том, что слово имеет хорошую общеуральскую этимологию и никаких надежных аргументов в пользу его заимствования, кроме общего фонетического сходства, не существует. Огромное количество примеров приводится автором из малайско-полинезийского языка ибан, якобы заимствовавшего чуть ли не весь свой базисный словарный состав из малайского — при том, что общепризнанным является близкое генетическое родство ибан и малайского, и по каким критериям автор определяет происхождение того или иного

оказаться любой из элементов этого списка — *любой из элементов*, но не *любое количество* элементов. При определенных условиях, по-видимому, даже возможно «окультуривание» отдельных базисных лексем. Так, например, в одних языковых ареалах слово 'звезда' может не нести на себе никакого груза религиозных коннотаций, в других же может, наоборот, оказаться прочно втянутым в локальную систему мифологем, переходя в семантическую категорию названий божеств, духов и т. п., заведомо менее устойчивую, чем категория немаркированных природных объектов.

Теоретическое решение этой проблемы на самом деле тривиально; все, что требуется — это более или менее строгое разграничение между внутренними и внешними заменами. Если, несколько огрубляя ситуацию, мы станем расценивать каждое замещение какого-либо элемента в 100-словном списке на заимствование как результат «окультуривания» базисного слова, то основной постулат лексикостатистики переформулируется следующим образом: «Базисная лексика изменяется с постоянной скоростью, за исключением тех случаев, когда она в ситуации языковых контактов переходит в разряд культурной».

Разумеется, это ставит нас перед новой проблемой — необходимостью уметь корректно отличать внешние замены от внутренних, в том числе для языков с плохо разработанными (или вообще не разработанными) этимологиями. Но эта задача носит уже технический, а не сущностный, характер, и будет разбираться отдельно.

В целом, завершая раздел, хотелось бы отметить, что в лингвистической терминологии понятия «базисная лексика» и «культурная лексика» находятся друг с другом примерно в таких же отношениях, как понятия «язык» и «диалект»: обе оппозиции чрезвычайно полезны и удобны на практике, но ни одна из них так и не получила формального и, главное, общепризнанного определения. В обоих случаях существует некоторое интуитивное и универсальное понимание того, что *некоторые* объекты,

ибанского слова, остается непонятным, и т. п. Все эти случаи не только серьезным образом подрывают аргументацию автора, но и тормозят создание *настоящего* типологического инвентаря заимствованной базисной лексики в языках мира.

подпадающие под эти определения, заведомо относятся к множеству А («французский», «суахили», «глаз», «я», «вода»), другие же — к множеству Б («канадский французский», «киунгуджа суахили», «очки», «слуга», «испарения»), но существует при этом и колоссальных размеров пограничная территория, где все зависит от текущей конъюнктуры и личных предпочтений исследователя («молдавский язык/диалект», «китайские языки/диалекты», «зрачок», «вождь», «водоворот»).

Для того, чтобы не попасть здесь в терминологическую ловушку (как это, с нашей точки зрения, случилось в работе Г. Хаарманна), необходимо лишний раз подчеркнуть, что для конкретных целей, преследуемых лексикостатистическим методом, разница между «базисной» и «культурной» лексикой на самом деле *не является принципиальной*. Конкретные слова в 100-словном списке Сводеша во многом представляют собой произвольную выборку, обусловленную не только и не столько «базисностью», «универсальностью», «культурной независимостью» соответствующих значений, сколько элементарным *удобством исследователя*.

В самом деле, если заменить текущий список на другой, состоящий в основном из «культурной» лексики, мы просто-напросто рискуем попасть в ситуацию, когда подавляющее большинство или даже все слова из этого списка в исследуемом языке окажутся заимствованными — и список, таким образом, будет непригодным к использованию в качестве диагностического. Каковы бы ни были конкретные претензии по поводу «базисности» или «незаимствуемости» тех или иных элементов списка Сводеша (некоторые из них будут рассмотрены ниже), самое главное — это то, что за 50 с лишним лет, прошедших со времени его составления, он в целом зарекомендовал себя как удобный и практичный инструмент, позволяющий на регулярной основе применительно к самым разным языковым семьям получать исторически правдоподобные результаты.

Возможно, дискуссионного пыла было бы меньше, если бы участники дискуссии договорились отказаться от употребления условных терминов «базисный» и «культурный», заменив их на формализованную «шкалу среднестатистической устойчивости», аналогичную «индексу стабильности», разработанному С. А. Старостиным. В рамках такой замены на «устойчивость» можно

было бы протестировать и другие лексические элементы, и тем самым перейти от ситуации произвольного выбора, основанного на интуиции Сводеша, к более объективному обоснованию. Скорее всего, дальнейшее развитие лексикостатистики будет подразумевать решение этой задачи, но в цели текущего исследования оно не входит.

1.6.2. *Эмпирические проблемы лексикостатистики.* Под эту рубрику попадают случаи, в которых идея постоянной скорости распада диагностического списка входит в явное противоречие либо с историческими фактами (там, где история языка или языковой группы/семьи прослеживается по источникам), либо с той или иной версией праязыковой лексической реконструкции, в ходе которой обнаруживается, что некоторые ветви в семье якобы подвергаются «лексической эрозии» заметно быстрее, чем другие ветви той же семьи. Понятно, что обнаружение даже одной-двух таких ситуаций в лучшем случае серьезно снижает уровень доверия к постулатам лексикостатистики, а в худшем — приводит к отрицанию ее научной ценности вообще.

Важно, однако, сразу же подчеркнуть, что там, где речь идет об исторически верифицируемых датах распада, абсолютное большинство претензий со стороны «глотто-скептиков» относится к ситуациям значительного увеличения скорости распада списка в ходе языковых контактов — т. е. грешит неумением или нежеланием отличать *внутренние* замены от *внешних*.

«Хрестоматийным» примером такого рода до сих пор остается работа [Bergsland & Vogt 1962], в которой для исландского языка на базе неоспоримых фактов устанавливается скорость изменения списка, равная 0.04% за тысячу лет, а для близкородственного риксмала — 0.2% за тысячу лет. В работе [С. Старостин 1989] было, однако, показано, что из двадцати слов, заменившихся в базисной лексике риксмала, по меньшей мере шестнадцать следует квалифицировать как заимствования, так что количество *внутренних* замен, имевших место за последнее тысячелетие, в исландском и риксмале примерно одинаково¹.

¹ В свете этого и других подобных случаев уместно вообще поставить вопрос о корректности применения глоттохронологического метода к т.н. «литературным» языкам, во многом носящим искусственный характер и

Аналогично выглядит и ситуация с албанским: несмотря на то, что общее количество лексических замен в нем действительно чрезвычайно диспропорционально по отношению даже к соседним языкам, таким, как греческий [Haarmann 1990b; Holm 2009], подавляющее большинство этих замен — заимствования из латинского суперстрата, исключение которых из подсчетов дает исторически приемлемые результаты¹.

С другой стороны, вряд ли можно игнорировать, например, полученные параллельно друг с другом результаты применения формулы, откалиброванной по методу С. А. Старостина, к столь разным языковым ситуациям, как, с одной стороны, современные китайские «диалекты» северной и миньских групп [S. Starostin 1995b], с другой — древнеегипетский язык Текстов пирамид и классический коптский [Militarev 2000]; в первом случае дата расхождения устанавливается как III в. н. э., во втором — сравниваемые языки оказываются разделены тридцатью тремя столетиями дивергенциями, т. е. язык Текстов пирамид по сравнению с коптским датируется XXVIII в. до н. э. Оба результата хорошо согласуются с историческими данными и вряд ли могут представлять собой случайное совпадение.

Конкретные примеры ситуаций, в которых данные лексико-статистики входили бы в резкое, труднообъяснимое противоречие с результатами «традиционной» классификации (полученной

зачастую представленным текстами, намеренно архаизированными (сохраняющими, в частности, старые, «высокостильные» элементы лексики, давно вышедшие из употребления в живом языке) или ориентированными на языковые нормы «престижного соседа», как риксмол, сознательно инкорпорирующий большое количество лексики из датского языка. Особенно важно учитывать эту особенность при составлении 100-словных списков по мертвым письменным языкам, для текстов которых часто можно заподозрить неточное отражение живой разговорной речи того периода, в который они были написаны.

¹ Корректное определение позиции албанского на индоевропейском древе, впрочем, серьезно затрудняется большим количеством плохо этимологизируемых вхождений, про многие из которых невозможно достоверно определить, продолжают ли они старые индоевропейские корни или заимствованы из неизвестного источника.

с помощью «метода»¹ совместных инноваций) языков, соотносимой с полученной праязыковой реконструкцией, весьма немногочисленны; собственно говоря, нам известно всего два таких случая:

(а) статистические подсчеты Р. Бласта [Blust 2000], сопоставившего реконструированный им для праавстронезийского 200-словный список Сводеша с такими же списками для современных языков; согласно выводам Бласта, сохранность праавстронезийского базисного фонда варьирует в языках-потомках от 50 и выше процентов (малайский: 58%, ибан: 54% и т. п.) до, наоборот, 10 и ниже (асумбоа: 7,2%, каулонг: 5,2% и т. п.), что очевидным образом дискредитирует метод;

(б) противоречие между лексикостатистической классификацией семитских языков, представленной, например, в [Militarev 2000: 303], согласно которой т. н. современные южноаравийские языки (джиббали, мехри, сокотри) первыми отделяются от общесемитского ствола на грани V и IV тыс. до н. э., и более традиционными представлениями об этой семье, согласно которым

¹ Слово «метод» помещается здесь в кавычки для того, чтобы подчеркнуть, что классификация родственных языков на основании совместных инноваций не является *методом* в строгом смысле слова, т. е. объективной, формализованной процедурой, не зависящей в очень большой степени от личных предпочтений исследователя. Нам не известна ни одна работа по какой-либо языковой семье, в которой предпринималась бы попытка представить *полное* исчисление инноваций (пусть хотя бы только фонологических и грамматических, без учета лексики), имевших место в языках-потомках, сопровождаемое убедительным обоснованием их инновативного статуса и исчерпывающим статистическим анализом. Отчасти благодаря этому для подавляющего большинства языковых семей, даже хорошо изученных, сегодня существует как минимум несколько классификационных моделей, сторонники которых в принципе неспособны прийти к единому мнению ввиду отсутствия универсальных критериев для их сравнения. Неопределенность и размытость «метода совместных инноваций» идеально суммировал сино-тибетолог Дж. Мэтисофф (скептическое отношение которого по отношению к возможности формализации универсального метода классификации языков хорошо известно): «...the point is that there is no mechanical way to assign relative weight to conflicting patterns of innovative rule-sharing. The analyst must make such decisions *on the basis of hard-earned intuitions* [курсив мой — Г. С.] as to what is critical and what is of lesser importance» [Matisoff 2000: 356].

южноаравийские языки обладают важными совместными изоглоссами с западносемитской ветвью и никак не могли отделиться от прасемитского раньше, чем, например, аккадский (подробнее см., например, в [Hetzron 1972]).

Нельзя не обратить внимание на тот факт, что в обоих случаях «лексикостатистические аномалии» обязаны тем представителям в целом хорошо исследованных семей, историческое изучение которых либо находится на низком уровне, либо сталкивается с рядом специфических проблем. Так, вряд ли случаен тот факт, что наибольший процент лексики, унаследованной от праавстронезийского, Р. Бласт обнаруживает в малайском (языке с богатой и хорошо известной историей), в то время как наименьший обычно представлен в «западноокеанийских» языках — одной из самых спорных подгрупп австронезийской семьи, многие из языков которой до сих пор не встроены как следует в общевосточноазиатскую систему соответствий, а некоторые, возможно, и вообще не следует считать австронезийскими (более подробную критику методики Блоста см. в работе [Peiros 2000]).

Что касается южноаравийских языков, то и в этом случае нельзя однозначно утверждать, что лексикостатистический метод приводит к заведомо неверной классификации. Так, А. Ю. Милитарев считает, что морфологические изоглоссы, объединяющие южноаравийские языки с западносемитскими, но не с аккадским, не столь значительны, чтобы «перевешивать» данные лексикостатистики [Militarev & Kogan 2000: XLII]. Возможно также, что глоттохронология действительно дает неверный результат, но не из-за некорректности метода, а из-за неверных этимологических суждений: если в будущем удастся показать, что в формировании южноаравийских языков значительную роль играл субстратный фактор (на что вроде бы имеются отдельные указания), то «вычет» плохо этимологизируемой южноаравийской лексики из общесемитских подсчетов может значительно изменить ситуацию.

Опыт работы со 100-словными списками разных семей убедительно показывает, что эволюция базисной лексики довольно редко происходит «строго по плану»; даже учитывая все поправки к классической формуле М. Сводеша, предложенные в работах С. А. Старостина, «идеальные» лексикостатистические матрицы представляют скорее исключения. Подавляющее большинство

непредсказуемых расхождений с ожиданиями, тем не менее, лежит в рамках небольшой и вполне допустимой статистической погрешности. Дать этой погрешности твердое математическое определение — задача ближайшего будущего; пока что констатируем лишь тот очевидный факт, что для получения корректных результатов размер ее должен находиться в прямой зависимости от процента совпадений в списке (чем больше между языками А и В общей лексики, тем больше допустимая погрешность, и наоборот).

Для того, чтобы наглядно продемонстрировать бессилие лексикостатистики как надежного метода для классификации языков и датирования языкового распада, нам необходим хотя бы *один* случай, при котором (а) число лексических расхождений между языковыми парами А : В и А : С, в которых А, В и С, согласно данным, полученным другими путями, хронологически равноудалены друг от друга, заведомо превышает возможную погрешность (скажем, А : В = 90%, В : С = 70%); (б) можно убедительно показать, что все предполагаемые замены в списках относятся к числу *внутренних*, а не суб- или суперстратных. Без преувеличения можно утверждать, что ни в одной из многочисленных работ, критически настроенных по отношению к глоттохронологии, ни одного подобного примера до сих пор не было приведено: большинство из них элементарно не проводит границ между внутренними и внешними заменами, что полностью обесценивает аргументацию.

1.6.3. *Технические проблемы лексикостатистики.* Как это ни странно, реальным «врагом номер один» успешного применения лексикостатистической методики на сегодня является вовсе не невозможность ее теоретического осмысления и интерпретации, а сугубо практические трудности, возникающие как при составлении 100-словных списков, так и при их последующей аналитической обработке. Большинство этих трудностей так или иначе упоминается в литературе (особенно критической), но в ходе конкретного рабочего процесса их принято либо замалчивать, либо решать *ad hoc* — какого-либо единого «глоттохронологического стандарта» среди лексикостатистов не существует, и ожидать его появления в ближайшее время вряд ли имеет смысл.

Отметим, что слова «практический» и «технический» лишь подчеркивают здесь тот факт, что без адекватного решения этих вопросов применение лексикостатистического метода оказывается невозможным (в отличие от, например, вопроса о реальности или иллюзорности постоянства скорости распада списка; лексикостатистикой можно заниматься даже при условии отказа от ответа на него). При этом, конечно же, на самом деле все эти трудности имеют и свою сущностную интерпретацию, а различные стратегии их разрешения будут серьезным образом влиять на получаемые результаты, так что выбор этих стратегий также должен быть обдуманным, а не случайным.

К числу важнейших «технических» вопросов относятся следующие:

(а) 100-словный список изначально составлен не на метасемантическом языке, а на английском; т. е. элементы его представлены не значениями, а словами естественного языка. Если эти слова полисемичны, какое значение следует выбирать?

(б) Как поступать с проблемой синонимии, когда одному и тому же значению в словаре или текстовом корпусе языка соответствует два или более эквивалента?

(в) Как быть в тех случаях, когда эквивалентов не обнаруживается вообще, будь то по техническим причинам (недостаток данных) или сущностным (значение действительно отсутствует в языке)?

(г) При проставлении помет, определяющих когнацию, как следует поступать с композитными основами, состоящими из двух или более лексических морфем, а также с теми случаями, когда слова, объединенные общим этимологическим корнем, имеют при этом различное морфологическое оформление?

(д) Если этимология слова неизвестна, каким образом определить, заимствовано ли оно (чтобы исключить его из подсчетов) или является исконным (чтобы подсчитать его как внутреннюю замену)?

(е) Правильно ли мы поступаем, не проводя никаких формальных разграничений между *полной* утратой лексической единицы (исчезновением ее из языка) и *частичной*, когда она остается в языке либо с изменившимся значением, либо в качестве

архаизма? Не приводит ли это к существенной потере информации и искажению результатов?

В рамках нашего исследования мы не претендуем на то, чтобы дать исчерпывающий ответ на все эти вопросы, тем более что по мере конкретной работы со списками перед исследователем регулярно встают все новые и новые их разновидности, и предусмотреть все возможные «подводные камни» заранее вряд ли возможно. Можно и нужно, однако, предложить некую единую рабочую модель, которая, с одной стороны, учитывала бы опыт и практику конкретных лексикостатистических исследований, уже проводившихся в рамках хотя бы Московской школы, с другой — оптимальным образом подходила бы как для целей и задач, поставленных в нашем исследовании, так и в дальнейших работах по лексикостатистическому анализу языков мира.

1.6.3.1. *Значения элементов 100-словного списка.* Проблема множественности возможных эквивалентов для отдельных единиц из 100-словного списка связана с тем, что в исходных работах М. Сводеша как 200-, так и 100-словные списки задавались простым перечнем соответствующих английских слов; лишь в отдельных случаях, опираясь на собственный опыт работы с многозначными лексемами, автор сопровождал их минимальным комментарием, уточняющим искомое значение (например: 'теплый (о погоде)', 'кожа (человека)' и т. п.); дальнейшая практика работы показала, что комментарий этот явно недостаточен, т. к. многие неоткомментированные элементы списка (равно как и ряд откомментированных) все равно остались представлены английскими словами, передающими два или более смежных значения, которые во многих языках последовательно различаются лексически, и далеко не всегда очевидно, какое из них следует считать более «базисным». Так, для слова *hair* непонятно, о каких волосах идет речь — на голове или на теле (а также о том, имеются ли в виду собирательные 'волосы' или индивидуальный 'волосок'); для слова *breast* 'грудь' неясно, имеется ли в виду женская или мужская грудь; слово *eat* 'есть' в некоторых языках регулярно имеет два эквивалента — 'есть жесткую пищу' (мясо) и 'есть мягкую пищу' (растительную) и т. п.

Разумеется, сам М. Сводеш при составлении списка стремился по возможности избегать попадания в него «вызывающе»

полисемичных значений или слов с трудноопределимой, расплывчатой семантикой; то, что с этой задачей в целом ему скорее удалось справиться, чем не удалось, доказано хотя бы тем, как успешно итоговый список, и по сей день активно используемый в самых разных лингвистических целях, прошел проверку временем. Однако уже в самой первой работе, посвященной критическому разбору нового стословника [Hoijer 1956], на материале всего одного языка навахо было показано, что и этот список значений оказывается слишком неточным, по крайней мере, при попытке строгой апробации его на материале неиндоевропейских языков. В дальнейшем список имеющихся претензий только расширялся.

В некотором смысле данная проблема, по-видимому, неразрешима принципиально. Исходя из того, что сегодняшние достижения лингвистики в изучении лексической картины мира на материале самых разных языков значительно превосходят уровень наших знаний 50 или 100 лет назад, в чисто *техническом* плане ситуация, когда каждый из элементов списка Сводеша имеет строго делимитированное толкование, исключающее двусмысленность в выборе эквивалента, вполне вообразима. Но каковы будут при этом критерии выбора — например, как определить, что более «базисным», т. е. допускающим включение в список, является значение 'волосы на голове', а не 'волосы на теле' (или наоборот)? Подобная постановка вопроса автоматически возвращает нас и к дискуссии о релевантности противопоставления «базисного» и «культурного» слоев, и к перспективе объективного статистического ранжирования лексики по степени устойчивости, т. е. к задачам, не имеющим на сегодняшний день возможности решения.

На самом деле, однако, для сугобо практических целей лексикостатистики проблема эта оказывается несущественной. В рамках соответствующей процедуры ключевое значение имеет не *абсолютная* устойчивость того или иного значения, а *относительная*. Вне зависимости от того, какое значение будет включено в стословный список — 'волосы на голове' или 'волосы на теле' — принципиально для нас не понимание того, какие из этих 'волос' более базисны, чем другие, а осознание того факта, что выбранные нами 'волосы' (предположим, что это 'волосы на голове')

имеют меньшую среднюю устойчивость по языкам мира, чем слова 'имя' или 'два', но бóльшую, чем слова 'круглый' или 'маленький'.

При выборе того конкретного узкого значения, которое мы хотим видеть за словом 'hair', 'breast' и т. п., допустимо, таким образом, позволить себе руководствоваться не только интуитивными или эмпирически обоснованными представлениями о степени устойчивости, но и разного рода «техническими» соображениями — например, традицией (если какое-то узкое значение давно и активно используется при сборе данных и компиляции списков, ему можно отдать приоритет) или практическим удобством (скажем, для значения 'волосы на голове' обычно бывает проще получить корректный эквивалент, чем для значения 'волосы на теле', где неизбежны сложные семантические отношения с 'шерстью', 'мехом' и т. п.).

Единственное условие, которое при этом необходимо соблюдать максимально строгим образом — *последовательное* использование в статистическом анализе одного и того же набора значений для всех привлекаемых к сравнению языков. Недопустимым, например, было бы сопоставление 'волос на теле' в языке А с 'волосами на голове' в языке Б (разумеется, только в том случае, если эти значения выражаются в обоих этих языках с помощью этимологически разных корней).

Если исследование замкнуто в пределах отдельной языковой семьи или одного лингво-географического ареала, то в принципе допустимо оперировать «индивидуальным» списком, составленным с учетом ареальной языковой специфики. Исследования, отталкивающиеся от исходного списка Сводеша, но при этом заменяющие «неудобные» значения в этом списке на «удобные» (например, такие, для которых легче находить эквиваленты, или которые для данного региона оказываются более устойчивыми), встречаются очень часто. Так, например, в известном «списке Бендера», разработанном специально для полевых исследований в Эфиопии [Bender 1971: 169], из списка Сводеша исключены такие слова, как 'круглый', 'зеленый', 'желтый' (заменяются на 'трава', 'мокрый', 'три'), слово *fly* 'летать' замещается на омоним *fly* 'муха', а вместо *feather* 'перо' используется *father* 'отец' (sic!). Еще дальше идет Б. Сэндс в своем лексикостатистическом анализе

койсанских языков: в ее 100-словный список включены такие ареальные слова, как 'буйвол', 'лев', 'леопард', 'слон', 'жираф' и т. п. [Sands 1998a].

Помимо того, что такого рода варьирование элементов списка серьезно затрудняет глоттохронологическую интерпретацию результатов (для этого надо было бы как минимум высчитать среднестатистический индекс устойчивости соответствующих слов по языкам Африки), они к тому же оказываются несопоставимыми друг с другом — при том, что в условиях, когда генетическая классификация языков заданного макроареала, по большому счету, остается неизвестной, чрезвычайно важно располагать для всего этого макроареала универсальной методологической основой. На данном этапе наших знаний и гипотез о глубинном родстве между языками Африки совершенно не исключено, например, что какие-то из т. н. «койсанских» групп на самом деле генетически ближе к каким-либо из т. н. «нило-сахарских» групп, чем к остальным «койсанским». При наличии разных списков для «койсанских» и «эфиопских» языков, однако, такая гипотеза окажется принципиально нетестируемой на лексическом материале.

Наша задача, таким образом, сводится к выполнению двух условий: (а) максимально конкретному определению допустимой области значений элементов списка Сводеша; (б) последовательному использованию именно этих, и никаких других¹, значений для всех языков, данные по которым мы хотим считать сопоставимыми и объединить в рамках цельной системы. Первое из этих условий частично выполнено в коллективной публикации [Kassian et al. 2010] по уточнению семантики стословных элементов через набор синтаксических контекстов и семантических

¹ Разумеется, следует отдавать себе отчет в том, что лексические данные, доступные нам по тем или иным языкам, в разных источниках фиксируются с разной степенью детализации; к тому же ни одно описание не гарантировано от переводческих ошибок, иногда грубых. Это, однако, не должно быть достаточным основанием для полного отказа от использования «сомнительных» данных; важно лишь, чтобы по каждой анализируемой языковой группе в наличии было хотя бы некоторое количество высоко надежных источников, сравнение которых с «сомнительными» материалами может служить корректирующим фактором.

комментариев; работа очевидным образом не снимает *все* возможные неопределенности, возникающие при составлении стословных списков, но, на наш взгляд, показывает, что проблема однозначного определения семантики базисных лексем не является настолько принципиально неразрешимой, насколько ее часто описывают в критических работах. Второе условие в той степени, в которой это возможно, выполнено в настоящем исследовании.

1.6.3.2. *Проблема синонимии и возможные пути ее разрешения.* Ужесточение семантических требований к стословному списку во многом разрешает традиционную проблему выбора того или иного синонима при составлении стословного списка, но все же не исчерпывает ее.

Само по себе понятие «синонимии» не имеет единого толкования и может как расширяться, так и сужаться в зависимости от личных предпочтений или практических целей исследователя. Известны, например, три условия, предложенные Ю. Д. Апресяном как достаточные для того, чтобы считать две лексические единицы синонимичными: «... (1) чтобы они имели полностью совпадающее толкование... (2) чтобы они имели одинаковое число активных семантических валентностей... (3) чтобы они принадлежали к одной и той же (глубинной) части речи». При этом автор добавляет: «...это определение не требует от синонимов совпадения или хотя бы частичного сходства их сочетаемости или конструкций, в которых они употребляются, а также совпадения их стилистических свойств» [Апресян 1995: 223].

Такое понимание синонимии имеет понятное практическое применение — в рамках этой концепции, отличающей «синонимию» от «квази-синонимии» в первую очередь по признаку полного или частичного совпадения толкований, удобно анализировать различные типы парадигматических отношений между лексическими единицами языка. Лексикостатистика, однако, преследует совершенно иную цель — из массы «синонимов» выбрать, по возможности, один-единственный, опираясь на объективные внутриязыковые критерии. Поэтому для нас разница между такими двумя парами, как, с одной стороны, *есть* и *жрать* («синонимы» в понимании Ю. Д. Апресяна, различающиеся только стилистически) или *высокий* и *длинный* («квазисинонимы»,

имеющие лишь частично совпадающие толкования) оказывается несущественной: и в том, и в другом случае отобрано должно быть только одно из двух слов.

В связи с этим мы будем здесь и ниже понимать «синонимию» в максимально узком смысле этого термина — как полную взаимозаменяемость двух лексических единиц, употребленных в одном и том же значении, во всех синтаксических и стилистических контекстах, в которых они могут встречаться¹.

Интуитивно кажется очевидным, что «абсолютная» лексическая синонимия такого рода должна встречаться в языке редко, а в области базисной лексики и, уже, 100-словного списка, как подтверждает опыт работы с хорошо исследованными языками — в исключительных случаях. В подавляющем большинстве ситуаций «синонимы» при тщательном их рассмотрении обнаруживают грамматическое, семантическое или стилистическое распределение. Для наших целей все такие ситуации можно считать «квази-синонимическими».

Там, где общие правила распределения известны, обычно удается предложить наиболее естественное разрешение ситуации, опираясь на следующие положения:

(а) выбор наиболее стилистически нейтрального эквивалента, не несущего дополнительных субъективных оттенков значения — пейоративного, мелиоративного и т. п.;

(б) выбор наиболее «свободного» варианта, вступающего в синтагматические отношения с максимальным разнообразием лексических элементов, т. е. отказ от использования идиоматически закрепленных квази-синонимов;

(в) при наличии текстового корпуса или подробно иллюстрированного словаря — опора на частотность употребления; подходящим «базовым» квази-синонимом, как правило, оказывается слово, чаще всего обнаруживаемое в требуемом «сводешевском» значении;

¹ Это в целом совпадает с классическим «максималистским» определением синонимии, сформулированным в свое время С. Ульманом: «only those words can be described as synonymous which can replace each other in any given context without the slightest change either in cognitive or emotive import» [Ullmann 1951: 108-109] («тотальная синонимия», по терминологии Дж. Лайонза).

(г) при наличии супплетивных основ, т. е. грамматического распределения, следует по возможности выбирать наипростейшую форму (например, ед. ч. вместо мн. ч., за исключением языков, где грамматически немаркированной является именно форма мн. ч.; положительную, а не сравнительную или превосходную степени сравнения прилагательных и т. п.), за исключением специально оговариваемых случаев (супплетивная парадигма у местоимений, см. ниже).

При последовательном применении такого подхода удастся снять многочисленные «технические» проблемы, которые на первый взгляд могут казаться непреодолимыми. С другой стороны, легко понять, что корректное разрешение этих проблем напрямую зависит от качества и количества доступного материала. Языки, в том числе и Африки, для которых мы располагаем подробными словарями, большими текстовыми корпусами и возможностью непосредственной работы с информантами, чрезвычайно немногочисленны по сравнению с языками, для которых в наличии оказываются, в лучшем случае, небольшие и во многом «сырые» вокабулярии, а в худшем — краткие лексические списки, часто составленные с большим количеством ошибок, или две-три статьи, посвященные отдельным аспектам грамматики или фонетики, из которых при желании можно выбрать лексический материал, за точность перевода и толкования которого, однако, никто не отвечает.

Можно, конечно, занять такую позицию, согласно которой ужесточение требований к выбору квази-синонима предполагает принципиальный отказ от использования «сомнительных» источников данных и опору исключительно на подробные описания (желательно апробированные на конкретных носителях). Однако в этом случае из рассмотрения придется исключить не просто отдельные языки, но даже целые семьи языков, по которым подробные описания до сих пор не составлены; для тех же языков, которые уже являются вымершими или вымирающими, такой подход вообще исключает возможность их классифицирования на основе лексикостатистики.

Разумнее, на наш взгляд, придерживаться прямо противоположного подхода: включать в рассмотрение данные любого происхождения и качества, с единственным условием — чтобы по

каждому отдельно взятому языку или диалекту этих данных было достаточно для составления 100-словного списка или хотя бы заполнения в нем по меньшей мере 60-70 позиций (эмпирически выявлено, что 60-70 лексических единиц — тот примерный порог, при преодолении которого обычно удается избежать грубых классификационных ошибок).

Конечно, при этом практически неизбежными будут искажения классификации на «микро»-уровнях: неверное заполнение даже 5-6 позиций списка в совокупности с «естественной» лексикостатистической погрешностью может привести к построению неверного генеалогического древа для близкородственных языков на глубине от одной до двух тысяч лет. Однако для целей более глубокого сравнения «массовость» избранного подхода, напротив, имеет принципиальное значение, т. к. чем больше близкородственных языков будет вовлечено в лексикостатистический анализ, вне зависимости от качества данных, тем с большей уверенностью можно будет восстановить правильный элемент сводешевского списка для их общего праязыка.

Как это ни парадоксально, в ситуациях, когда дело приходится иметь с плохо описанной группой языков, каждым из которых в свое время занимался всего один исследователь, *реконструкция пра-лексемы*, т. е. парного сочетания «звучание / значение», нередко оказывается более достоверной, чем выражение соответствующего значения в каждом отдельно взятом языке-потомке. Поясним это утверждение на конкретном примере.

В т. н. экойдную группу языков (район Огоджа в Нигерии; близкородственны языкам банту) входит порядка 15-20 единиц, лишь по двум-трем из которых существуют подробные грамматические, словарные и текстовые источники. Единственным источником лексических сопоставлений по всей группе на сегодня является небольшая монография [Crabb 1965], словарные списки в которой были собраны непосредственно автором. Таким образом, степень достоверности нашего знания того факта, что, например, в языке экпарабонг базисное значение 'пить' соответствует фонетическому облику *у́*, отражает то, как это слово в ходе полевой работы услышал и записал один исследователь; у нас нет ни твердых гарантий того, что слово было записано абсолютно правильно (например, что в нем действительно представлен -ATR-

гласный или высокий тон), ни, что еще более значимо для лексикостатистических целей, что лексема *yó* на самом деле записывает именно базисное значение 'пить', а не, например, 'глотать', 'сосать', или какие-нибудь сложные стилистически маркированные значения ('потягивать', 'выпивать /спиртное/' и т. п.). Более или менее твердые гарантии такого рода может предоставить только подробный словарь, проиллюстрированный многочисленными примерами синтаксических контекстов, а еще лучше — корпус текстов, на котором значение можно будет верифицировать.

То же самое касается и любого другого из 14 экоидных языков, учтенных в монографии Крэбба. Но когда выясняется, что для всех без исключения 14 языков в списке Крэбба даются рефлексии одного и того же корня (балеп *ǰǰ*, бендеге, этунг, нта *wǰ*, ннам, экаджук *wó* и т. п.), вероятность того, что именно эта основа в праэкоидном языке выражала значение 'пить', оказывается *превышающей* достоверность соответствующей информации для каждого из языков-потомков. (Предположим, что вероятность допущения ошибки для одного языка равна 0,2; тогда для двух языков она оказывается равна 0,04; для семи языков, т. е. половины, при которой реконструкция станет сомнительной — 0,0000128!).

Ровно по этой причине в нашем исследовании методологически допустимым и даже рекомендуемым является совмещение двух подходов, на первый взгляд, противоречащих друг другу: с одной стороны — формальное ужесточение, настолько, насколько это позволяют обстоятельства, требований к заполнению конкретных позиций внутри 100-словного списка, с другой — принятие к рассмотрению лексических данных самой разной степени достоверности, начиная от фонетически и семантически сомнительных и принципиально невыверяемых сведений по мертвым языкам, записанным исследователями-«любителями» еще в додескриптивистскую эпоху, и заканчивая выполненными на современном уровне подробными и неоднократно проверенными описаниями.

Как поступать с ситуациями, когда источник приводит для одного и того же значения из списка Сводеша два или более эквивалента без указания на семантические, стилистические или сочетаемостные различия (условимся называть такие эквиваленты «техническими синонимами»)? В этом случае принципиальное

значение будет иметь внешний фактор, учет которого позволяет сформулировать такие правила:

(1) если ни для одного из двух или более «технических синонимов» X_1, X_2 , выражающих значение X в языке A , в потенциально родственных ему языках $A_1... A_n$ не обнаружено этимологически тождественного эквивалента с тем же значением, позицию в списке Сводеша можно заполнить любым из приводимых слов или обоими, т. к. это не имеет принципиального значения для результатов подсчетов (лексикостатистическая матрица останется неизменной при замене X_1 на X_2 и наоборот);

(2) если из двух «технических синонимов» точные внешние соответствия в языках $A_1... A_n$ имеет только X_1 , позиция в списке им и заполняется. Так, в севернокойсанском языке жу|хоан для значения 'вошь' главный источник данных дает эквиваленты $s'i$ и $\check{f}a$, из которых первый имеет параллели с тем же значением во многих родственных языках/диалектах, а второй — ни в одном. Эту ситуацию, при отсутствии эксплицитных пояснений, можно толковать трояко: (а) $\check{f}a$ на самом деле обозначает не 'головную вошь', а какую-то отдельную, более узко-специфическую разновидность вши, для которой нет места в 100-словном списке; (б) $\check{f}a$ — «новая» 'вошь', слово, которое относительно недавно начало вытеснять из живого зууса старое $s'i$, но еще не довело процесс до конца; (в) $s'i$ — «старая» 'вошь', т. е. архаизм типа 'ока', сохраняющийся лишь в специфически маркированных контекстах (идиомы, пословицы, старые песни, заклинания и т. п.). При отсутствии дополнительных аргументов в пользу того или иного из перечисленных сценариев, совокупная вероятность (а) и (б) заведомо превышает вероятность (в), что и позволяет принять соответствующее решение;

(3) если из двух «технических синонимов» X_1 имеет точное соответствие в родственных языках $A_1 ... A_n$, а X_2 имеет точное соответствие в родственных языках $B_1 ... B_n$, решение должно приниматься, исходя из количественной и дистрибуционной характеристики всех учитываемых языков. Здесь уже задействовано столько различных факторов, что выстроить для них единую предварительную модель довольно трудно. К счастью, ситуации подобного рода на практике встречаются редко; отдельные случаи будут подробно обсуждаться и аргументироваться ниже, в

ходе анализа конкретного лексического материала по языкам Африки.

Важно подчеркнуть, что гипотетическая ситуация, описанная в части (б) правила (2), по сути представляет собой единственно допустимую ситуацию *реальной* (не условной или «технической») синонимии в рамках 100-словного списка, которую мы будем обозначать как «*транзитную синонимию*». Дело в том, что элементарный анализ как диахронического развития языков, письменная история которых нам известна, так и многочисленных бесписьменных языков, по которым существуют подробные лексикографические описания и корпуса текстов, показывает, что лексическое замещение — процесс скорее постепенный, чем внезапный: «новое» слово какое-то время сосуществует со «старым», прежде чем окончательно превратить последнее в стилистически окрашенный архаизм или вообще вывести его из употребления.

В том случае, когда лексикограф застаёт язык именно в его «переходном» состоянии, соответствующие два (но ни в коем случае не более чем два!) эквивалента удобнее всего трактовать как синонимы; единственное альтернативное решение — разбивать описываемый язык на два «идиолекта» и уточнять, что работа ведется по «архаичному» или, наоборот, «инновативному» варианту языка. Избранная для данного исследования методика рекомендует все же последовательно ориентироваться на «архаичный» вариант, хотя здесь следует еще раз специально обратить внимание на то, что «архаичный вариант» элемента стословника и «архаизм» — принципиально разные понятия; «архаичный вариант» — это слово, которое представители старших поколений используют в речевом общении как единственный немаркированный эквивалент соответствующего значения, а «архаизм» — слово, которое могут одинаковым образом использовать носители любого поколения, но только в специально маркированных контекстах, типы которых перечислены выше. «Архаичные варианты» поэтому включать в стословник не только допустимо, но и в высшей степени желательно, а «архаизмы», напротив, везде, где они могут быть заподозрены с достаточными на то основаниями, из стословника должны исключаться.

Резюмируем вышесказанное: а) *лексикостатистическая синонимия* в рамках 100-словного списка понимается нами как полная

взаимозаменяемость двух или более лексических эквивалентов в любых семантических, стилистических и синтаксических контекстах; б) единственный реальный случай такой синонимии, типологически естественный для языков мира — *транзитная синонимия*, которая предполагает использование разных эквивалентов в зависимости от «генеалекта», т. е. принадлежности идиолекта говорящего к группе идиолектов, свойственной для пожилых поколений или для молодых (при этом сам говорящий может иметь любой возраст — все зависит от того, чьи языковые нормы он перенимает); в) все остальные случаи синонимии — это либо *лексикостатистическая квази-синонимия*, предполагающая возможность выбора эквивалента, более точно удовлетворяющего априорным требованиям к значению и сочетаемости лексемы, либо *техническая синонимия*, обусловленная неполнотой имеющейся информации и предполагающая обращение к внешним данным сравнения при проставлении когнации.

1.6.3.3. *Проблема отсутствия эквивалента*. В процессе составления 100-словных списков нередко случаются случаи, когда для того или иного значения в списке Сводеша, даже при условии его строгого семантического определения, в рассматриваемом языке вообще невозможно подобрать необходимый эквивалент. Чаще всего это также объясняется технической причиной — нехваткой материала; гораздо более редкими, но более значимыми для теории лексикостатистики являются случаи, когда в языке отсутствует само значение требуемого элемента. Разберем эти ситуации отдельно.

На практике отсутствие одного или нескольких элементов в словарных списках или текстовых корпусах анализируемого языка само по себе не лишает исследователя возможности использовать неполный список. Незаполненность n позиций для списка X просто означает, что 1 процент совпадений между X и родственным ему языком Y будет эквивалентен не 1 случаю когнации, а $(100 - n)/100$ случаям. Очевидно, что чем больше число лагун, тем менее достоверными окажутся результаты подсчетов, причем чем хронологически ближе друг к другу оказываются языки X и Y , тем более точными и полными должны быть составляемые списки, чтобы итоговая классификация оказалась надежной.

Поясним сказанное на тривиальном примере. Лексикостатистические подсчеты между литовским и латышским языками

дают 70 случаев когнации = 70% совпадений в 100-словном списке, что, согласно формуле С. А. Старостина, соответствует $\approx 2,000$ годам независимого развития. Сократив список наполовину, т. е. проводя подсчеты только по первым 50 единицам, получаем 66%; до 30 единиц — 60%; до 10 — 50%. Это означает, что для языковых групп порядка балтийской (точнее, «восточнобалтийской») сокращение списка даже наполовину не приведет к существенным искажениям результатов. При дальнейшем сокращении, однако, такие искажения все же возникают; следовательно, должен существовать определенный «порог насыщения» списка, преодоление которого обязательно для того, чтобы результаты подсчетов можно было рассматривать всерьез.

Задача математического выведения формулы такого «порога» весьма нетривиальна, т. к. она должна учитывать не только примерную хронологическую дистанцию между языками, но и зависимость надежности сокращенного списка от индексов стабильности исключенных из него единиц. Для наших целей мы поэтому ограничимся интуитивно-эмпирическим правилом, согласно которому язык можно привлечь к лексикостатистическим подсчетам, если число известных нам единиц 100-словника превышает хотя бы половину, а 50-словника — хотя бы порог в 30 единиц. Если, однако, при этом оказывается, что рассматриваемый язык не имеет близких родственников, т. е. по традиционной терминологии оказывается «изолятом», то для рассмотрения его возможного статуса как изолированного члена какой-нибудь «макросемьи», конечно, желательно значительное превышение 50-процентного порога заполненности списка, причем в первую очередь требуется, чтобы в нем присутствовало подавляющее большинство единиц с высоким индексом стабильности.

Второй, гораздо более редкий, но при этом более интересный в теоретическом отношении тип лакуны — это ситуация, когда строго очерченное значение того или иного элемента из списка Сводеша в языке X *отсутствует*, т. е. не сочетается ни с одним означающим. Эти случаи также делятся на две группы:

1) Значение отсутствует в языке как следствие отсутствия референта, с которым оно могло бы соотноситься. Хрестоматийный пример — отсутствие в океанийских языках слова 'рог' по причине незнакомства носителей с рогатыми животными (по крайней

мере, вплоть до XVIII в., когда их начали завозить на тихоокеанские острова европейцы) [Беликов 2006: 25]. Для некоторых языков трудно подобрать подходящее слово для выражения значения 'круглый' (что может быть связано с отсутствием колесного транспорта, гончарного круга и т. п. объектов, гарантирующих релевантность данного признака, даже несмотря на то, что круглые объекты сами по себе встречаются в природе вне зависимости от уровня технологического развития локальных языковых носителей) и т. п.

Обнаружение такого рода фактов, хотя и сказывается негативно на представлении о «базисности» 100-словного списка, в практическом отношении серьезных последствий не имеет. Незаполнимые позиции для каждого конкретного языка вряд ли будут превышать одну-две единицы, и обходиться с ними следует так же, как и с заимствованной или незасвидетельствованной лексикой, т. е. исключать из подсчетов.

2) Значение отсутствует в языке из-за того, что соотносимый с ним класс референтов в сознании носителей делится на два или более непересекающихся друг с другом класса, например, 'большая птица' и 'маленькая птица' вм. 'птица' (вообще); 'есть жесткую (мясную) пищу' и 'есть мягкую (растительную) пищу' вм. 'есть' (вообще) и т. п. Эта проблема, по сути, возвращает нас к дискуссии о синонимии; задача лексикостатиста в этом случае — либо постараться максимально сузить требуемое значение *a priori*, либо, там, где это невозможно (скажем, классы референтов типа 'большая птица' : 'малая птица' могут различаться в зависимости от конкретного языка), ориентироваться на «степень базисности» той или иной лексемы, причем в первом приближении «базисность» можно понимать хотя бы как элементарную частотность употребления; частные пояснения применительно к конкретным элементам списка (50-словного) будут приведены ниже.

1.6.3.4. *Проблема полиморфных основ.* Одно из ключевых предписаний лексикостатистики гласит, что когнация постулируется в тех случаях, когда сравниваемые слова имеют общий лексический корень. Нередки, однако, ситуации, в которых то или иное базисное значение выражается словом, состоящим из двух (или, очень редко, даже более чем двух) корней. Отдельные случаи такого рода встречаются и в индоевропейских языках, хотя наиболее

«болезненный» характер эта проблема имеет для языковых семей Юго-Восточной Азии и Северной Америки, где корнесложение является одним из основных способов словообразования. Языки Африки в этом плане занимают скорее промежуточное положение, но и в них композитные основы — не редкость.

Предположим, что некое «стословное» значение в языке X_1 выражено сложной основой $A+B$, где A и B — лексические корни; в родственном ему языке X_2 — простой основой A ; наконец, в родственном им обоим языке X_3 — сложной основой $A+C$, где C это третий лексический корень, этимологически отличный от B . С исторической точки зрения самое вероятное объяснение данной ситуации таково: искомое значение в праязыке X выражалось простой основой A , которая в двух из трех потомков независимо друг от друга была затем расширена за счет дополнительного корня. Как в таком случае следует маркировать когнацию? Возможны два пути рассуждения:

(а) исконный корень A сохраняется во всех трех языках, следовательно, отношение когнации во всех случаях положительно;

(б) расширение основы A за счет дополнительного корня можно рассматривать как случай лексического замещения, поскольку структура означающего изменилась не только за счет фонетических переходов (которые лексикостатистика не учитывает), но и за счет существенного изменения морфологической структуры. Следовательно, отношение когнации во всех случаях должно иметь отрицательное значение.

Отметим, что последовательное применение принципа (б) должно в конечном итоге привести к тому, что помечаться как неродственные будут не только композитные основы, состоящие из двух или более лексических корней, но и сложные основы, различающиеся словообразовательными суффиксами. Иными словами, такие ситуации, как, например, абхазск. $a=\acute{s}a-p\acute{r}\acute{o}$ 'нога' : адыгейск. $\acute{t}a:-q:\acute{w}a$ 'нога' (где $\acute{s}a = \acute{t}a$: ← празападнокавказск. $*\acute{\lambda}a$, а вторые корни в композите имеют различное происхождение [Nikolayev & Starostin 1994: 760]), должны рассматриваться наравне с др.-индийск. $s\acute{u}v-ar-$: английск. $su-n$: французск. $so-l-eil$: русск. $co-l-ни́це$, отражающими различные суффиксальные варианты индоевропейской основы $*sarw-(el)- \sim *szw-(el)- \sim *szw-(en)-$ [Pokorny 1958: 881].

Интуитивно кажется, что стратегия (б) представляет собой чрезмерное ужесточение требований. Однако для того, чтобы оправдать это интуитивное представление, можно прибегнуть к следующему объективному аргументу: в основе процессов лексического замещения, когда значение перестает выражаться морфемой X и начинает выражаться морфемой Y, и лексического наращения, когда значение вместо простой морфемы X начинает выражаться конкатенацией X+Y, как правило (исключения допустимы и неизбежны, но речь идет именно о типичной, а не универсальной ситуации), лежат принципиально разные механизмы.

Лексическое замещение — процесс в первую очередь семантический, связанный с развитием переносных значений, метафоризацией и т. п. Что же касается наращений, особенно корнесложения, то их появление в базисной лексике чаще всего связано с таким «техническим» вопросом, как снятие омонимии в языках с моносиллабической структурой корня; не случайно, что в наибольшей степени эта проблема актуальна для таких языковых семей, как абхазо-адыгская, сино-тибетская, австроазиатская, ряд североамериканских и т. п., где корнесложение вынужденно является одной из самых продуктивных моделей словообразования.

В такого рода ситуациях некорректно было бы говорить, что новая сложная основа «вытесняет» старую простую; речь идет скорее об уточняющей функции добавочной морфемы, почти вынужденной необходимостью ее употребления для того, чтобы «подтвердить» значение основного корня, отделяя его тем самым от фонетически сходных или омонимичных¹. Следовательно, нет

¹ В ряде работ иногда встречается идея, согласно которой фактор «снятия омонимии» в языке с упростились фонетическими структурами может стимулировать в том числе и процесс лексических замещений. Эту идею, однако, очень трудно обосновать убедительно: наглядных примеров в языках с хорошо исследованной историей очень мало, и даже там, где они есть, находятся контраргументы. Так, А. Б. Долгопольский, обосновывая тезис, что структурные изменения в языке могут непредсказуемым образом ускорить изменения лексические, приводит в качестве примера латинское *eō / ī-re* 'идти', заменившееся во французском языке на (*je*) *vais / aller* якобы из-за того, что в противном случае оно бы фонетически совпало с *edo* 'есть' [Dolgorolsky 2000: 403]; при этом не упоминается, однако, что замена эта на самом деле произошла в народной

никаких оснований для того, чтобы маркировать такие основы как неродственные (что было бы абсурдно и с этимологической точки зрения).

Рассмотрим теперь более сложный случай, когда полиморфность распространяется как бы «по цепочке». Идеальным примером здесь могут служить слова или идиоматические сочетания из нескольких слов, выражающие в языках австронезийской семьи значение 'солнце'. В пределах лексической выборки из 80 языков, опубликованной в [Tryon 1995: 68-69], чаще всего встречаются следующие эквиваленты (семантическая точность данных для нас на этом этапе не имеет большого значения):

а) рефлексы праавстронез. **waRi*: атаяльск. *wayɪʔ*, рукой *vai* и др.; эта же основа сама по себе во многих других языках имеет значение 'день';

б) словосложение **mata + *waRi*, где **mata* = 'глаз', т. е. букв. 'глаз дня': индонезийск. *mata hari*, мадурск. *mata ari* и др.;

в) само по себе слово **mata*: иснаг *mata*, северн. танна *mət* и др.;

г) словосложение **mata +* другая основа, этимологически не тождественная **waRi*: зап.-фиджи *mata-ni-hiŋa*.

Процедура, допускающая наличие в языках абсолютной (нетранзитной) синонимии, хотя бы для подобного рода случаев корнесложения, требует пометить основу **waRi* номером 1, основу **mata* номером 2, а основу *hiŋa* в фиджи номером 3. Соответственно, все формы группы (б) в рамках 100-словного списка будут иметь по два номера (1, 2), и, таким образом, индонезийское *mata hari* будет одновременно находиться в отношении когнации к иснаг *mata* и танна *mət* (по первому корню) и к рукой *vai* (по второму) — при том, что иснаг *mata* и рукой *vai* когнатами считаться не будут.

В результате образуется парадоксальное нарушение транзитивности, когда $A = B$ и $A = C$, но $B \neq C$. Строго говоря, отношение лексикостатистической когнации и отношение тождественности

латыни еще задолго до радикальных фонетических изменений в французском языке: ср. итальянск. *vado* 'иду' = французск. *vais*. Даже если такой психологический механизм замещения все-таки существует, он, несомненно, носит маргинальный характер по сравнению с хорошо известным и изученным механизмом снятия омонимии за счет расширения омонимичных основ, а не их искоренения из языка.

не обязаны обладать в точности одними и теми же свойствами, но дело здесь не столько в том, что такой подход логически противоречив на уровне абстракции, сколько в том, что он плохо коррелирует с наиболее вероятным из исторических сценариев развития наблюдаемой языковой ситуации.

Анализ дистрибуции перечисленных вариантов по языкам и подгруппам австронезийской семьи показывает, что наиболее частотными оказываются простая основа (а) и сложная (б); при этом (б) полностью отсутствует в формозской ветви (или ветвях), противопоставленной (-ых) малайско-полинезийской ветви. На прауровне в этой ситуации естественнее всего реконструировать основу **waRi* с полисемией 'день / солнце', к которой на прамалайско-полинезийском уровне подключается «уточняющий» вариант **mata-waRi*, первоначально использовавшийся носителями в ситуациях, требовавших максимальной однозначности, а в дальнейшем обобщившийся на все контексты. Тогда соответствие «рукай *vai* : индонезийск. *mata-hari*» определяется как когнативное, поскольку лексического замещения на этом этапе не происходит. Однако в отдельных, сравнительно немногочисленных случаях прамалайско-полинезийское **mata-waRi* 'солнце' = 'глаз-дня' «усекается» до первой морфемы. Поскольку причина этого усечения носит психический, а не фонетический характер, здесь уже можно постулировать лексическое замещение: 'солнце' эксплицитно становится 'глазом'. Следовательно, несмотря на то, что индонезийская и иснагская формы объединены общим корнем 'глаз', они не должны считаться лексикостатистическими когнатами, т. к. в первой из них еще сохраняется старая основа, а во второй — нет.

Сложнее всего принять строгое решение относительно ситуации в фиджи. Здесь, по-видимому, имело место частичное замещение старой сложной основы **mata-waRi* на новую *mata-ni-hiŋa*, и все зависит от того, какой из двух компонентов считать основным. В целом развитие **waRi* → **mata-waRi* → *mata-ni-hiŋa* можно было бы рассматривать как движение по той же «траектории», что и **waRi* → **mata-waRi* → **mata*: постепенное смещение лексического значения 'солнце' со старого корня **waRi* на новый **mata*. В этом случае «смыслообразующей» морфемой в фиджи было бы именно *mata*. Однако подключение к анализируемому сопоставлению

коррелята в восточном фиджи (бау) — *siŋa* — показывает, что на самом деле и в этой подгруппе *mata*- имеет скорее «уточняющую» функцию.

Таким образом, наиболее верным решением на данном этапе было бы считать когнатами (а) и (б) (по основе **waŋi*), отделяя от них (в) (основа **mata*) и (г) (основа **siŋa*). Фактор везения заключается здесь в доступности для рассмотрения данных по очень большому количеству языков, относящихся к семье с продвинутой (хотя и далекой от идеала) степенью исторической изученности (австронезийской); далеко не во всех случаях этих данных хватает для того, чтобы на их основании построить цельный сценарий развития. Это не означает, однако, что для этого не следует прилагать все возможные усилия: опыт показывает, что, если семья состоит хотя бы из нескольких близкородственных языков, разобраться в структуре и предыстории ее композитных основ — задача в целом решаемая. В исключительных случаях допустимо, наверное, прибегать и к «нетранзитивным» решениям, но лишь после того, как будет показано, что единого (т. е. в большей степени согласующегося со всеми фактами, чем все остальные) сценария поведения анализируемых сложных основ не существует.

Сразу же отметим, что описанный механизм лексического замещения на самом деле может быть гораздо более частотным в языках мира, чем заметно на поверхности. Многие лексические замещения метонимического характера могут начинаться с образования композита (например, 'дождь' → 'вода-неба'; 'дерево' → 'стоящее-дерево') и завершаться утратой некогда главной основы (→ 'небо', 'стоящее'); очень вероятно, что случаи, когда в языке удастся реально зафиксировать такие композиты, находятся в меньшинстве по сравнению с «переходными» ситуациями, когда соответствующий композит удерживается в активном употреблении недолгое время, после чего упрощается до одной основы.

Для языков Африки проблема композитных основ не столь актуальна, как для языков Юго-Восточной Азии или Америки, где они распространены повсеместно. Однако и здесь словосложения в пределах стословного списка — не редкость; хорошо известны такие типичные развития, как 'облако' = 'волосы-неба', 'лист' = 'ухо-дерева' и др. Особо выделяется проблема устройства указательных и вопросительных местоимений, часто состоящих

из двух или более морфем ('кто' ≈ 'какой-человек', 'что' ≈ 'какая-вещь', 'этот' ≈ 'здесь-такой', 'тот' ≈ 'там-такой' и т. п.); общее решение этой проблемы для вопросительных местоимений см. ниже.

1.6.3.5. *Проблема идентификации заимствований.* Этот вопрос в «классической» глоттохронологии Сводеша, где внутренние и внешние замены трактовались единообразно, не был актуален. Гораздо большее значение он имеет в усовершенствованной модели С. А. Старостина, которая предполагает непосредственную зависимость точности результатов от корректной идентификации и исключения из статистических калькуляций всех случаев внешних замен.

Очевидно, что единого формального критерия, по которому можно было бы разграничивать внешние и внутренние замены, существовать не может. В стословных списках, подготовленных и обработанных в рамках деятельности Московской школы компаративистики, заимствования чаще всего помечаются на основании сравнительно-исторического анализа данных: чем детальнее изучена та или иная семья, тем надежнее можно аргументировать принятие соответствующего решения. При этом:

а) даже для таких всесторонне обследованных семей, как индоевропейская, нередки ситуации, когда слово либо не имеет этимологии вообще — ни внешней, ни внутренней (др.-английск. *docga*, совр. *dog* 'собака'), либо имеет очень слабую и плохо аргументированную этимологизацию (ср. др.-греческ. φθεΐϛ 'вошь', обычно объясняемое как производное от глагола φθεΐϛω 'уничтожать' [Роконну 1958: 487-488]; типологические параллели для такого семантического развития неизвестны, так что высока вероятность народной этимологии, вторично сблизившей исконно греческую глагольную основу с субстратным элементом);

б) в плохо исследованных или представленных скудными лексическими материалами семьях количество спорных случаев может возрастать буквально в разы, как будет видно из конкретных анализов 50-словных списков по африканским языкам;

в) наконец, в ситуации, когда лексикостатистическое исследование является *предварительным*, т. е. отталкивается от интуиции исследователя или носит обзорно-ареальный характер, разграничение внутренних и внешних замен невозможно в принципе.

Последний тип ситуации будет обсуждаться ниже, в рамках общего методологического описания рабочей процедуры, принятой нами для обследования языковых семей Африки. Что касается ситуаций, в которых все же присутствует определенный «исторический ресурс» для обоснования лексикостатистических решений, то здесь нужно обратить внимание на следующие обстоятельства:

— во-первых, ошибочное зачисление *единичных* внешних замен в ранг внутренних не обязательно должно привести к серьезному искажению результатов; см. п. 1.6.1 о возможности «ассимиляции» заимствования в языке в нестословном значении, после чего оно уже замещает старый элемент 100-словного списка на равных основаниях с внутренними заменами. Существенные искажения могут иметь место лишь в тех случаях, когда неидентифицированным остается субстрат или адстрат, из которого в список было одновременно перенесено сразу несколько элементов (или, как в случае брахуи, целое множество хронологически последовательных или одновременных адстратов);

— в тех случаях, когда стословный список языка X действительно содержит большое количество заимствований из неизвестного источника (условно — от 5% и выше), объективная вероятность этого тем выше, чем больше в этом языке обнаруживается слов, не имеющих *этимологических* (не лексикостатистических!) параллелей в близкородственных ему языках. Именно таким образом, т. е. через попытки хотя бы примерной этимологизации базисной лексики, удастся сформулировать гипотезы субстратного влияния на такие «нило-сахарские» языки, как канури, нобин, квегу и другие (эти ситуации будут подробно разобраны во втором томе исследования).

Разумеется, если в стословном списке языка X обнаружено 10 или 20 слов, не этимологизируемых на основе сопоставлений с лексическим составом родственных ему языков $X_1 \dots X_n$, это не означает, что *каждое* из этих слов непременно имеет субстратное происхождение и должно исключаться из подсчетов. Здесь очень важно уметь правильно использовать преимущество лексикостатистического метода, который сам по себе (иногда даже без этимологического анализа) позволяет сформулировать гипотезу о наличии в том или ином языке «скрытого субстрата». Рассмотрим

для примера ситуацию в сурмийской группе языков, сделав для нее репрезентативную выборку из пяти языков южносурмийской ветви (полная матрица по сурмийским языкам будет опубликована во втором томе исследования):

	Теннет	Чай	Ме'ен	Квегу
Дидинга	0.88	0.43	0.39	0.26
Теннет		0.41	0.39	0.27
Чай			0.65	0.46
Ме'ен				0.47

Данные матрицы наглядно показывают, что языки дидинга и теннет образуют тесное единство («югозападносурмийская» группа), противопоставленное несколько более далеким друг от друга, но все же имеющим общего промежуточного предка чай и ме'ен («юговосточносурмийская» группа). Что касается квегу, то этот язык на дереве должен объединяться с юго-восточной ветвью, т. к. имеет с каждым из двух ее языков по $\approx 45\%$ совпадений. Итоговая классификация имеет вид:

1. Юго-западные языки: дидинга, теннет;
2. Юго-восточные языки: 2.1. Квегу; 2.2. Чай, ме'ен.

Такое членение в целом совпадает с большинством моделей, предложенных в свое время исследователями на основании «метода совместных инноваций», и может считаться убедительным. Налицо, однако, по крайней мере одна серьезная аномалия. Будучи представителями одной из двух ветвей семьи, квегу, чай и ме'ен должны были бы показывать примерно один и тот же процент совпадений с любым из юго-западных языков. Между тем сходные цифры наблюдаются только в случае чай и ме'ен (от 39 до 43% совпадений с дидинга и теннет); квегу же дает значительно более низкий процент (26-27%) — расхождение более чем в 10% явным образом превышает ожидаемую погрешность, каким бы образом последняя ни высчитывалась.

Матрица, таким образом, однозначно указывает на *завышенную* скорость лексических изменений в квегу; критики лексико-статистического метода наверняка воспользовались бы этим фактом, чтобы еще раз указать на его несостоятельность — однако

для реального опровержения базисного постулата глоттохронологии в данном случае требовалось бы показать, что соответствующее ускорение вызвано большим количеством «лишних» *внутренних*, а не внешних замен. Между тем следующий этап анализа матрицы — этимологический — показывает, что в очень большом числе случаев, когда в квегу наблюдается «особый» корень, для него вообще не удается найти этимологических параллелей, т. е. доказать «внутреннее» происхождение замен оказывается невозможным.

В качестве эксперимента попробуем теперь исключить из подсчетов *все* этимоны в квегу, исконно сурмийское происхождение которых сомнительно в силу отсутствия этимологий (в предыдущей матрице из подсчетов были исключены только *очевидные* заимствования, в основном из близлежащего омотского языка кара, донорский статус которого для квегу хорошо известен). Новая матрица будет иметь следующий вид:

	Теннет	Чай	Ме'ен	Квегу
Дидинга	0.88	0.43	0.39	0.38
Теннет		0.41	0.39	0.39
Чай			0.65	0.63
Ме'ен				0.65

В этом варианте какие-либо аномалии, как видно, отсутствуют: как чай и ме'ен, так и квегу демонстрируют одинаковые или очень близкие проценты совпадений с югозападносурмийскими языками (максимум — 43%, минимум — 38%). Более того, внутри юго-восточной группы квегу ощутимо сближается с прочими языками: теперь уже можно говорить не о бинарном (чай-ме'ен / квегу), а скорее об изначально тернарном членении этой ветви (чай / ме'ен / квегу); именно эта схема классификации, кстати, является лидирующей среди африканистов-«сурмистов».

Некоторое беспокойство вызывает тот факт, что, если мы принимаем данную модель как историческую, число заимствований в 100-словном списке квегу будет колебаться в пределах от 35 до 40 элементов, т. е. побьет даже рекорды брахуи и северносонгайских диалектов. Однако в данном случае мы действительно,

скорее всего, имеем дело с редкой ситуацией, которую можно охарактеризовать как своеобразную «субстратно-адстратную деконструкцию» языка: сурмийский язык квегу мог первоначально быть перенят несурмийско-язычным племенем, перенесшим в него большой пласт своей собственной лексики, после чего уже в относительно недавнем прошлом вновь оказался под влиянием, на сей раз идущим «сверху» — от культурно-доминантных омотских языков.

Отметим, что согласно критерию динамической градации квегу, вне всякого сомнения, в генетическом отношении продолжает оставаться сурмийским языком: так, из первых двадцати элементов в «индексе стабильности» С. А. Старостина 17 имеют бесспорно сурмийскую этимологию и только три ('имя', 'собака', 'ноготь') объясняются как заимствования (одно субстратное и два адстратных).

Важнейшим условием корректного вычисления заимствований является применение *рекурсивного* принципа при расстановке когнаций в тех случаях, когда у нас нет априорных оснований исходить из уже установленного родства между анализируемыми языками.

Предположим, что тот же квегу — недавно открытый язык, единственной доступной информацией по которому является собранный в ходе ареального диалектного опроса 100-словный список. В этом случае на первом этапе сопоставления с потенциально родственными языками из подсчетов нельзя будет исключить ни одного слова, т. к. у нас нет твердой уверенности, идет ли речь о языке сурмийской, омотской или какой-то третьей, доселе неизвестной семьи. На втором этапе мы обнаруживаем, что максимальные связи квегу обнаруживает с сурмийскими (и даже конкретнее — с юговосточносурмийскими) языками; на втором месте по численности — лексика, не поддающаяся этимологизации; на третьем — омотская лексика. Подключив к этому принцип динамической градации, устанавливаем, что на основании текущих данных квегу должен генетически характеризоваться как сурмийский язык; следовательно, «омотские» слова в нем — уверенные заимствования, а слова, не поддающиеся этимологизации — потенциальные. Это дает возможность сделать «второй проход» по списку и исключить из подсчетов по крайней мере

омотизмы, а при наличии соответствующих оговорок и комментариев — также и слова из неизвестного субстрата.

Единственная ситуация, в которой данная процедура может оказаться неприменимой — классификация языка-изолята, ближайший родственник которого отстоит от него по меньшей мере на несколько тысячелетий. Если базисная лексика рассматриваемого изолята при этом еще и подверглась разрушительному воздействию одного или нескольких суб- или адстратов, любая позиция, которую такой изолят займет на общем генеалогическом древе, будет заведомо нестабильной. В такого рода случаях следует, по-видимому, воздерживаться от однозначных классификационных суждений, по крайней мере, до тех пор, пока к ним не будет подключен подробный этимологический компонент, выходящий далеко за пределы стословного списка.

1.6.3.6. *Проблема утраты информации.* В отдельных критических работах, посвященных лексикостатистике и глоттохронологии, можно обнаружить тревогу по поводу того, что при построении классификации метод формально требует отсеивания и игнорирования большого количества доступной (и иногда весьма наглядной) информации, традиционно считающейся не менее важной для обоснования классификационных аргументов. Утрата информации происходит даже при непосредственном переходе от размеченных списков к лексикостатистической матрице, ср.: «...the conversion of character-state data to percentage similarity scores between languages results in a loss of information and hence a reduction in the power of the method to reconstruct tree topology and branch lengths accurately» ([Atkinson et al. 2005: 195], со ссылкой на работу [Steel et al. 1988]).

Данный аргумент мог бы носить угрожающий характер в том случае, если бы удалось показать, что степень точности корреляции «генетическая классификация языков семьи X : реальная история разделения языков семьи X» *значительно и неоспоримо* увеличивается при подключении дополнительной информации, отсеянной на этапе лексикостатистических подсчетов — такой, как степень фонетического сходства между сравниваемыми лексемами, степень регулярности / частотности соответствий, наличие или отсутствие этимологических когнатов и т. п. К тому же для каждого типа такой информации требовалось бы разработать

собственную формализованную, объективную процедуру оценки (что, безусловно, входит в задачи сравнительно-исторического языкознания, но на данный момент технически невыполнимо).

Насколько нам известно, ни одной практической демонстрации такого рода проведено не было, и аргумент остается чисто умозрительным. Разумеется, в тех ситуациях, когда мы имеем дело с близкородственными диалектами (имеющими, например, от 95% до 99% совпадений), лексикостатистические данные оказываются недостаточными для объективной классификации ввиду неотличимости их показаний от статистической погрешности. В этих случаях приходится прибегать к дополнительным данным фонетических, грамматических и лексических изоглосс (если про такие ситуации вообще можно утверждать, что для них существует острая необходимость построения более чем одноуровневой генетической классификации). Но для семьи, скажем, индоевропейского масштаба «утрата информации» на стадии лексикостатистической обработки не будет иметь катастрофических последствий как минимум до тех пор, пока не будет твердо установлено, *какие* конкретные виды утраченной информации заведомо исказили полученные классификационные результаты.

С другой стороны, нельзя не признать, что фактор утраты информации принимает более серьезные очертания по мере углубления в «хронологические дебри». С. А. Старостиным было в свое время показано, на примере разбора алтайского материала, что значимость лексикостатистики для обоснования глубинного родства существенно возрастает при переходе от использования данных современных языков к данным промежуточных реконструкций [С. Старостин 1999: 788-790]; так, при сопоставлении русского и эвенкийского языков налицо 11 совпадений в 100-словном списке (на самом деле — не более 9, т. к. русские слова *звезда* и *пепел*, сравниваемые с эвенкийскими, не отражают соответствующие праиндоевропейские корни), но при сопоставлении праиндоевропейского и праалтайского — 27 (на самом деле — порядка 22-24, с рядом этимологически спорных ситуаций).

Такое соотношение, действительно, вряд ли может отражать случайность (при условии объективности выбора соответствующих праформ; корректная методика будет обсуждаться ниже). Однако теоретически мыслима (и, как будет показано на мате-

риале языков Африки, реально зафиксирована) и ситуация, при которой даже на прауровнях количество допустимых когнаций, определяемых по принципу фонетической совместимости, угрожающе приближается к «порогу необедительности» — 15, 10, 5% и т. п. Как быть в такого рода ситуациях?

Естественным искушением было бы обратиться напрямую к разработанному С. А. Старостиным альтернативному методу «этимостатистики», или «корневой глоттохронологии», в рамках которого слово A_1 в языке L_1 считается формально совпадающим со словом B_2 в языке L_2 в том случае, если в L_2 существует также слово A_2 , этимологически тождественное A_1 , но отличающееся от него своей семантикой (т. е. немецкое *Hund* 'собака' будет считаться формально-тождественным неродственному англ. слову *dog*, поскольку в англ. языке сохраняется родственное слово *hound* с узко-специализированным значением 'охотничья собака')¹.

Переход от «лексической глоттохронологии» к «корневой» неизменно увеличивает количество когнатов и во многом нейтрализует проблему «утраты информации». Однако этимостатистика, в отличие от обычной лексикостатистики, по определению может быть применена *исключительно* к семьям, сравнительно-историческая фонология и лексикология которых уже находятся на очень высокой стадии развития. Главная отличительная черта этимостатистики — отказ от требования однозначных семантических соответствий — увеличивает объем «позитивного» материала, но лишь за счет неконтролируемого роста субъективно принимаемых решений. Даже в процессе этимостатистической обработки двух индоевропейских языков, относящихся к разным ветвям, регулярно возникают ситуации, когда невозможно твердо сказать, есть ли у разбираемого слова этимологические параллели в сравниваемом языке или нет².

¹ Данное определение характеризует конкретный случай применения этимостатистики к 100-словному списку, хотя в целом метод был разработан с перспективой возможности его применения к произвольным текстам или словарным спискам сравниваемых языков. Подробнее см. [С. Старостин 1989].

² Например, для литовского и немецкого этот вопрос актуален уже при разборе первого слова в стословнике — вопрос «родственно ли немецкое *alles* 'все' литовскому *aliái* 'целый, полный'» в индоевропеистике не

Тем не менее, нет ничего принципиально запретного в том, чтобы, отказавшись от применения этимостатистики *в целом*, заимствовать из этого метода такие элементы, которые можно поставить под более или менее объективный контроль. Точно так же, как принцип фонетического соответствия можно на определенном этапе подменить более дозволительным, но при этом формально определяемым принципом фонетической совместности, принцип однозначного семантического совпадения, вероятно, можно с некоторыми оговорками подменить принципом *типовой семантической корреляции*, который мы определим так:

— слово X в языке A₁, имеющее основное значение¹ S₁, и слово Y в потенциально (или «доказанно») родственном ему языке A₂, имеющее основное значение S₂, находятся в отношении *типовой семантической корреляции*, если (а) значение S₁ выводится из значения S₂ с помощью не более чем *одного* метонимического или метафорического сдвига и (б) в языках мира документально зафиксированы *многочисленные* случаи полисемии S₁ / S₂ или хотя бы морфологической деривации S₁ от S₂.

Пункт (б) имеет при этом подчеркнуто важное значение, т. к. семантическая связь между потенциально родственными словами

имеет однозначного разрешения (см. [Fraenkel 1962: 7], где *aliái* выводится из польского заимствования *alè*).

¹ Следует внести важное терминологическое уточнение: здесь и ниже под «значением» S слова или корня X мы будем понимать именно «основное» значение, т. е. не весь комплекс возможных семантических толкований слова, включающий переносные (как правило, стилистически маркированные) метонимические и метафорические употребления, а только то (или те) употребления, которые являются для него наиболее частотными и стилистически нейтральными. Соответственно, ситуация, при которой переносное значение слова становится основным, будет рассматриваться как изменение (основного) значения. Так, про слово *bucca* в классической латыни мы будем говорить, что его «значение» — ‘нижняя часть щеки’, и при переходе от латыни к романским языкам имеет место изменение значения → ‘рот’, несмотря на то, что *переносное* (стилистически «грубое») значение ‘рот’ для этого слова встречается и в латинских памятниках (точный хронологический момент, в который *bucca* в народной латыни становится основным разговорным эквивалентом значения ‘рот’, определить трудно, но не подлежит сомнению тот факт, что *bucca* — инновация, со временем вытеснившая старую основу *ōs*).

очень часто устанавливается исследователем исключительно на основании собственной «семантической интуиции», не подкрепленной реальными данными семантической типологии, при том, что интуиция эта, разумеется, не только серьезнейшим образом различается от одного специалиста к другому, но и нередко бывает обманчива. Так, например, обзор данных показывает, что семантическое развитие 'видеть' → 'глаз' (как правило, через морфологическую деривацию, т. е. 'глаз' как *'виделка') является типовой корреляцией, в то время как обратное 'глаз' → 'видеть' встречается чрезвычайно редко (хотя переход к *направленному активному* действию, т. е. 'смотреть', 'рассматривать', 'наблюдать', наоборот, вполне частотен, ср. русск. 'глазеть' или англ. *to eye*); сомнительно, чтобы этот факт был интуитивно очевиден для произвольно взятого этимолога.

К пункту (б) необходимо дать еще два комментария. Во-первых, теоретически допустимо заменить документальную фиксацию полисемии в конкретном языке на документальную фиксацию исторического развития от S_2 к S_1 в языках с длительной письменной историей — проблема лишь в том, что таких языков в нашем распоряжении значительно меньше, чем живых языков без соответствующей истории. Сложнее обстоит дело с реконструируемым семантическим развитием, пусть даже для очевидно близкородственных языков, т. к. здесь мы вступаем в зону относительности: понятно, что, например, греческ. πῦρ 'огонь' и чешск. *průž* 'раскаленная зола', скорее всего, родственны друг другу и что сопоставление подтверждает возможность развития от 'огня' к 'золе', но полную уверенность здесь все же придаст лишь обнаружение соответствующей полисемии или хотя бы деривации в других языках мира, не обязательно индоевропейских (к счастью, таких случаев достаточно).

Во-вторых, определенного уточнения требует и слово «многочисленные». Так, например, в отдельных бушменских языках отмечены такие случаи полисемии, как 'звезда / ёж' или 'звезда / цесарка', в ряде языков северных народов (уральских, енисейских) — 'луна / дед', в семитских и некоторых других афразийских — 'глаз / родник, источник' и т. п. Общим для всех этих случаев является то, что они, как правило, ограничены определенным лингвогеографическим ареалом, для которого иногда удается

даже выдвинуть гипотезу об отправной точке «семантической диффузии». В этих случаях лучше говорить о *локальной* (или «микротиповой») семантической корреляции на заданном языковом пространстве; вне зависимости от того, сколь много конкретных языков, ограниченных этим пространством, будут ее обнаруживать, вряд ли, например, единичную попытку возвести слово 'луна' в каком-нибудь африканском языке к слову со значением 'дед' можно будет серьезно оправдать апелляцией к данным северных языков. Напротив, семантическая корреляция 'глаз' / 'видеть' или 'луна' / 'белый' обнаруживается в разных, географически не связанных друг с другом ареалах; такие корреляции можно назвать *универсальными*.

Для локальной семантической корреляции разумно предположить в качестве исходной гипотезы одноразовое возникновение (которое могло быть связано как с культурной индивидуальностью этноса, так и с чисто «техническими» факторами — например, омонимией, получившей статус полисемии) с последующей контактной диффузией; это, по сути, ставит локальные корреляции в один ряд с «редкими» или «уникальными» семантическими развитиями, которые не сумели получить широкое распространение в своем ареале. Напротив, универсальные корреляции, по-видимому, способны время от времени вызывать одинаковые сдвиги значения в разных ареалах независимо друг от друга, отражая те принципы членения семантического пространства, которые могут быть присущи языку в целом.

Выявление и инвентаризация универсальных семантических корреляций в сфере базисной лексики — важнейшая задача для дальнейшего развития лексикостатистики, т. к. ее решение позволит к статистической информации по «когнации 1-го уровня» (полное совпадение значений сравниваемых слов) приплюсовать информацию по «когнации 2-го уровня» (значения сравниваемых слов, связанные универсальной семантической корреляцией). Вполне вероятно, что и для «когнации 2-го уровня», т. е. того, что мы называем «тривиальными семантическими изменениями», можно будет вывести определенный коэффициент скорости изменений, значительно расширив тем самым объект применения лексикостатистики. Однако даже и без детального алгоритмирования учет «когнации 2-го уровня» для дополнительного

подкрепления лексикостатистической гипотезы, сформированной на основе анализа «когнации 1-го уровня», будет чрезвычайно полезен.

Полная инвентаризация универсальных корреляций не может входить в цели нашего исследования, т. к. предполагает написание по меньшей мере отдельной монографии на базе детального исследования базисной лексики по всем языковым семьям мира. Частичный перечень таких корреляций на основании уже проанализированных списков, однако, вполне возможен и будет приведен ниже, в специальном приложении, посвященном подробному описанию значений 50-словного списка.

Представляется, что объективный и по возможности формализованный учет семантических корреляций во многом компенсирует фактор «утраты информации», особенно на уровне глубинных сравнений, когда такого рода компенсация будет жизненно необходима для верификации результатов «обычной» лексикостатистики. В дальнейшем, возможно, удастся разработать и другие компенсаторные механизмы (например, учитывающие и «нетривиальные» семантические изменения), но очень важно следить за тем, чтобы каждый из таких механизмов применялся только к объекту строго определенного типа, причем объекты должны быть стратифицированы от «стабильных», «универсальных», «частотных» и т. п. к «нестабильным», «локальным», «редким» и т. п. и к анализу привлекаться в строгой последовательности от первых к вторым. Тем не менее, до тех пор, пока мы не научимся корректно работать с универсальными семантическими корреляциями, конструировать процедуры еще более сложного уровня, наверное, преждевременно.

1.6.4. *Заключение.* Все вышеперечисленные «технические» проблемы лексикостатистики были рассмотрены нами лишь вкратце, и по большинству из них пока удалось лишь бегло очертить механизмы их преодоления; детали этих механизмов должны прорабатываться уже в ходе решения практических задач по классификации языков конкретных семей (в нашем случае — африканских).

Тем не менее, на наш взгляд, можно показать, что ни одна из этих проблем не является фундаментально непреодолимой, и для каждой из них можно предложить такое решение, которое было

бы одновременно формализуемым и при этом касалось бы сущности вопроса, а не представляло собой произвольно-условный выход из положения. Более того, как кажется, решение этих задач вполне может заинтересовать даже «закоренелых скептиков» в отношении лексикостатистического метода, т. к. оно непосредственно связано с всесторонним изучением таких проблем, как синхронные и диахронические аспекты синонимии; универсальность и ареальность языковых значений; типология семантических переходов; выявление неизвестных субстратных элементов и т. п.

Подводя окончательные итоги раздела 1.6, остается лишь еще раз подчеркнуть, что претензии, в разное время предъявленные к лексикостатистике и глоттохронологии, оказались неспособны дискредитировать самую суть метода — однако при этом надлежит отдать должное даже «деструктивной» критике, всесторонне способствующей его дальнейшему развитию и постепенному устранению недостатков. Хотелось бы надеяться, что и наше исследование внесет в это развитие свою посильную лепту.

1.7. Общее описание процедурных стадий исследования.

От обсуждения общетеоретических («идеологических») основ исследования перейдем к изложению конкретного плана его устройства. Здесь можно выделить следующие стадии:

а) *фиксация инвентаря тестовых значений*, т. е. составление неизменного по количеству элементов, общему составу и методике формального анализа лексикостатистического списка;

б) *первичный отбор конститuentов генетической классификации*, т. е. определение того, каким должен быть минимальный и может быть максимальный состав «языкового таксона 1-го уровня», тестируемого на родство более высокого уровня с помощью лексикостатистики;

в) *лексикостатистический анализ 1-го уровня*, т. е. составление и анализ списков по языковым таксонам 1-го уровня, создание на их основе генетического древа и сверка его с традиционными классификациями;

г) *подготовка прасписков 1-го уровня*, т. е. этимологическая обработка лексического материала по всем таксонам 1-го уровня с целью выведения на его основе максимально вероятного списка

праязыковых форм, допустимых к сравнению на более высоких уровнях;

д) *предварительный анализ прасписков 1-го уровня*, сам по себе включающий три этапа: (д1) приблизительная простановка когнаций между прасписками «вручную» на основании критерия фонетического сходства и (лишь отчасти) фонетической совместимости; (д2) автоматизированная простановка когнаций между прасписками на основании применения простейших аналитических алгоритмов; (д3) сопоставление результатов ручной и автоматической обработки данных и выработка на их основе начальных гипотез о родстве «2-го уровня», т. е. устанавливаемого между праязыками (или праязыками и языками-изолятами);

е) *этимологическая проверка гипотез глубокого родства* — верификация результатов стадии (д) через попытку реконструкции 50-словных списков для праязыков 2-го уровня, основанную как минимум — на систематическом применении критерия фонетической совместимости, как максимум — на обнаружении наглядно демонстрируемых фонетических соответствий. Родство «2-го уровня», прошедшее через этап (е), будет считаться «обоснованным» и приниматься в качестве рабочей модели;

ж) *рекурсивное применение (д) и (е) для выдвижения гипотез родства «n-уровня»*, т. е. последовательное, «ступенчатое» обоснование таксономических единиц более высокого уровня с помощью аналогичных процедур до тех пор, пока такое обоснование остается принципиально возможным. Этап, на котором оказывается невозможным показать ни фонетическую совместимость, ни успешную работу принципа динамической градации, определяется как *терминальный*.

Перейдем теперь к более подробному рассмотрению каждой из перечисленных стадий исследования.

1.7.1. *Фиксация инвентаря тестовых значений*. Решение этой задачи, на первый взгляд, очевидно и сводится к формальному подтверждению того, что в качестве основы для классификации мы выбираем 100-словный список Сводеша — не столько потому, что отобранный Сводешем инвентарь сам по себе «идеален», сколько потому, что на активном использовании именно этого списка завязана огромная традиция, отказываться от достижений которой было бы чрезвычайно неудобно в практическом плане.

Технические проблемы, связанные с использованием стословного инвентаря Сводеша, во многом снимаются или упрощаются при условии выбора уточненных значений, так, как это сделано в работе [Kassian et al. 2010] (более подробная информация для первой половины списка будет приведена в разделе 1.8).

Есть, однако, основания усомниться в том, что на *всех* этапах исследования позитивные результаты могут быть достигнуты только при условии опоры на *полный* стословный список Сводеша. На самом деле количество элементов списка, необходимое для получения сколь-либо достоверной классификации, напрямую зависит от генетического возраста анализируемой семьи. Так, для языков (диалектов), распавшихся около тысячи лет тому назад, разумеется, необходим полный список: согласно глоттохронологической формуле С. А. Старостина, за этот период два языка должны разойтись друг с другом на «дистанцию» не более чем в 8-10 слов из стословного списка (учитывая только внутренние замены), так что значительное усечение списка может свести расхождения к числу, неотличимому от обычной погрешности.

Для языковых семей «среднего» уровня, таких, как индоевропейская, обязательное использование всего стословного списка уже можно заподозрить на избыточность. Вне зависимости от того, полагается ли исследователь при подсчетах на объем в 100, 80, или даже 50 элементов, германские, романские, славянские и индоарийские языки на индоевропейском дереве будут неизменно разноситься по разным ветвям. Для того, чтобы, например, процент сходства между литовским и хинди значительно превысил стандартные 25-26%, список придется урезать до «неприличного» минимума — но даже в этом случае лексическое сходство между литовским и хинди невозможно нивелировать до уровня сходства между близкородственными литовским и латышским.

Обозначим через N общее количество элементов лексикостатистического списка, ранжированных в порядке убывания «индекса стабильности» (т. е. чем больше N , тем больше в его составе «неустойчивых» значений), через P_1 — процент совпадения между двумя дальнородственными индоевропейскими языками (литовский и хинди), через P_2 и P_3 — проценты совпадения между близкородственными индоевропейскими языками (литовский/латышский и хинди/бенгали соответственно). Тогда зависимость P

от N наглядно демонстрируется эмпирически в виде следующей таблицы:

N	P ₁	P ₂	P ₃
100	26%	70%	81%
80	27%	72%	79%
50	39%	76%	81%
30	48%	70%	80%
10	50%	80%	80%

Анализ как этой, так и многих других аналогичных ситуаций, подводит нас к двум важнейшим выводам:

1) Для языков, связанных достаточно дальним родством (определяемым примерно в три и более тысячи лет независимого существования), общий размер лексикостатистического списка *не имеет существенной значимости* — по крайней мере, до тех пор, пока количество опознаваемых когнатов не станет неотличимым от погрешности (т. е. пользоваться списком в 10 единиц, из которых между литовским и хинди совпадает 5, все же не рекомендуется). Это число будет *всегда* значительно отличаться от соответствующего процента совпадений между близкими языками.

2) При сокращении списка за счет последовательного исключения из него «нестабильных» элементов общий процент совпадений будет *значительно* увеличиваться для дальнородственных языков (ср. подскок от 26% до 50% для литовского/хинди) и, наоборот, оставаться в общем и в целом неизменным для близкородственных (ср. явное отсутствие какой-либо зависимости P от N в паре хинди/бенгали и возможную, но в лучшем случае спорную и очень слабую зависимость P от N в паре литовский/латышский, связанной чуть более дальним родством).

Что это означает в практическом плане? То, что для определения родства между двумя и более языками (или праязыками), хронологическая дистанция между которыми примерно превышает две тысячи лет, полный стословный список действительно избыточен. Более того, во многих случаях использование его может оказаться не просто избыточным, но, в самом прямом смысле этого слова, *вредным*. Речь идет в первую очередь о ситуациях дальнего родства типа ностратической, когда набор

потенциальных «кандидатов» на то или иное значение в нестабильной части списка достигает размеров, с трудом охватываемых исследователем. Установить когнацию ностратического уровня между словами 'глаз' или 'пить' в целом несложно, т. к. и для того, и для другого мы уверенно представляем себе праиндоевропейскую (**ok^w*-, **rōy*-), прауральскую (**šilmä*, **yuyē*), прадрavidийскую (**kaŋ*, **uŋ*-) и т. п. реконструкции; дойти же до однозначной лексической реконструкции таких этимонов, как 'желтый', 'маленький', 'сказать' и т. п. обычно бывает намного труднее, и здесь может быть особенно силен соблазн закрыть глаза на ступенчатый принцип реконструкции и перейти непосредственно к сопоставлению на уровне мелких подгрупп или даже отдельных языков без прояснения общей дистрибуционной картины.

Исходя из этих соображений, для текущего исследования принимаются следующие методологические установки:

А. Родственные отношения внутри «мелких» таксономических единиц (которые мы будем по возможности называть «группами»¹), этимологическая общность базисной лексики которых превышает 40-50%, будут подсчитываться *по полному стословному списку*. Результаты этих подсчетов, в виде лексикостатистической матрицы и генеалогического древа, будут опубликованы в соответствующих разделах исследования, но полностью все списки и их подробный этимологический разбор приводиться не будут, из

¹ Учитывая традиционную относительность таких таксономических терминов, как «группа», «ветвь», «семья», в рамках нашего исследования мы не собираемся претендовать на их строгую формализацию (тем более что добиться однозначного понимания и употребления терминологии такого рода в лингвистическом сообществе — задача, скорее всего, невозможная). Тем не менее, чтобы минимизировать уровень путаницы хотя бы для данной работы, по возможности хотелось бы закрепить термин «группа» за таксонами низкого уровня (от 40% и выше совпадений в 100-словном списке), «семья» — среднего уровня (≈ от 15% до 40%), «макросемья» — высокого уровня (менее 15% совпадений между живыми потомками). Термин «ветвь» будет в этом случае сугубо относителен (так, каждая из «семей», входящих в «макросемью», является одной из «ветвей» этой макросемьи и т. д.); для обозначения «сверхмелких» таксономических единиц можно употреблять также термины «подгруппа» или «диалектный пучок/кластер».

соображений и без того немалого объема работы, главная цель которой — попытка разобраться в общей историко-лингвистической панораме африканского континента, а не в отдельных тонкостях конкретных взаимоотношений языков и диалектов очевидно близких групп того же уровня, что и славянские или тюркские¹. Значимость подсчетов по полному стословному списку связана в первую очередь с тем, что это дает возможность точнее определить глоттохронологическую дату распада праязыка каждой группы, что, в свою очередь, чрезвычайно важно для сопоставлений на более высоких уровнях.

Б. Родственные отношения внутри «крупных» таксономических единиц (семей и макросемей) будут определяться на основании *сокращенного наполовину* списка Сводеша, из которого отсеиваются в первую очередь среднестатистически неустойчивые элементы, этимологическая обработка которых по группам сопряжена с повышенной технической сложностью и при этом обладает в целом низким «коэффициентом отдачи».

Простейший способ отсеять такие элементы — воспользоваться «индексом стабильности» С. А. Старостина, т. е. включить в состав 50-словного списка первую половину 100-словника, ранжированного от наиболее устойчивых значений к наименее устойчивым (от слова 'мышь' до слова 'лист'). Однако в работе [Starostin 2010] уже было указано на то, что удобнее воспользоваться компромиссным вариантом, согласно которому 9 слов из первой половины 100-словника будут все же заменены на 9 слов с несколько более низким коэффициентом стабильности, в первую очередь из-за технических обстоятельств. Это следующие элементы:

— 'печень': несмотря на общую устойчивость, в языках, представленных диагностическими списками, а не подробными словарями, это слово очень часто отсутствует;

— 'шея', 'грудь': эти слова занимают в «индексе стабильности» номера 48 и 49 соответственно, т. е. сами по себе довольно

¹ Составление и публикация полных 100-словных списков по языкам Африки с детальным этимологическим разбором является частью общего Интернет-проекта «Глобальная лексикостатистическая база данных» (<http://starling.rinet.ru/new100>), запущенного в 2011 г. в рамках работы Московской школы компаративистики и международной программы «Эволюция языка».

неустойчивы, причем 'шея' в описаниях часто путается с 'горлом', а при слове 'грудь' часто не указывается, о какой груди идет речь (мужской или женской);

— 'этой', 'той': 50-словный список и без этих элементов оказывается вполне репрезентативным в плане местоименных основ ('я', 'ты', 'мы', 'кто', 'что'), а указательные основы из всех этих местоимений наиболее проблематичны в плане семантического описания, особенно в языках с несколькими степенями ближнего и дальнего дейксиса;

— 'полный': также трудное слово, ввиду того, что во многих лексикографических описаниях разграничиваются (иногда даже на корневом уровне) глагольное и адъективное значения, и без детальной проработки текстов не удастся понять, какое из них является основным;

— 'давать', 'стоять': здесь также очень частотны случаи семантического смещения в описаниях, составленных без надлежащей степени тщательности ('давать' может различаться для 1-го, 2-го, 3-го лица и императива; 'стоять' часто путается с 'вставать', что особенно плохо в тех случаях, когда эти значения выражаются разными корнями);

— 'рыба': это слово, во-первых, не является универсальным (для многих языков, носители которых живут в пустынных ареалах, например, койсанских, оно вообще не зафиксировано или представляет собой недавнее заимствование), во-вторых, так же, как и 'печень', часто оказывается отсутствующим в диагностических списках по малым языкам.

Вместо этих девяти слов было предложено включить в состав сокращенного списка следующие элементы: 'убивать', 'нога', 'рог', 'слышать', 'мясо', 'яйцо', 'черный', 'голова', 'ночь'. Выбор именно этих эквивалентов в первую очередь определялся конкретным опытом составления списков, показывающем, что для этих элементов, как правило, не представляет особого труда определить однозначный семантический эквивалент. Тот факт, что три последних слова имеют индекс стабильности в диапазоне от 71 до 86, вряд ли способен существенно исказить результаты; напротив, есть даже основания считать, что в состав «сверхустойчивого» списка вполне имеет смысл включить два-три «средне-неустойчивых» элемента в специальных диагностических целях.

Приведем теперь окончательный список в том виде, в котором он будет фигурировать в основной части нашего исследования (из двух нижних индексов первый отражает место элемента в нашем 50-словнике, второй — в исходном индексе стабильности С. А. Старостина)¹:

- 1) *пепел*_{38/43} (*ashes*); 2) *птица*_{33/37} (*bird*); 3) *черный*_{48/71} (*black*); 4) *кровь*_{20/20} (*blood*); 5) *кость*_{34/38} (*bone*); 6) *ноготь*_{19/19} (*claw /nail/*); 7) *умирает*_{13/13} (*die*); 8) *собака*_{16/16} (*dog*); 9) *пить*_{15/15} (*drink*); 10) *сухой*_{24/24} (*dry*); 11) *ух*_{32/35} (*ear*); 12) *ест*_{25/26} (*eat*); 13) *яйцо*_{47/68} (*egg*); 14) *глаз*_{4/4} (*eye*); 15) *огонь*_{7/7} (*fire*); 16) *нога*_{43/53} (*foot*); 17) *волос*_{127/29} (*hair*); 18) *рука*_{11/11} (*hand*); 19) *голова*_{49/72} (*head*); 20) *слышат*_{45/59} (*hear*); 21) *сердце*_{14/14} (*heart*); 22) *рог*_{244/57} (*horn*); 23) *я*_{3/3} (*I*); 24) *убивает*_{42/52} (*kill*); 25) *лист*_{41/50} (*leaf*); 26) *вошь*_{17/17} (*louse*); 27) *мясо*_{46/63} (*meat*); 28) *луна*_{18/18} (*moon*); 29) *рот*_{31/33} (*mouth*); 30) *имя*_{10/10} (*name*); 31) *новый*_{23/23} (*new*); 32) *ночь*_{50/86} (*night*); 33) *нос*_{29/31} (*nose*); 34) *не*_{30/32} (*not*); 35) *один*_{21/21} (*one*); 36) *дождь*_{39/45} (*rain*); 37) *дым*_{36/40} (*smoke*); 38) *звезда*_{40/46} (*star*); 39) *камень*_{9/9} (*stone*); 40) *солнце*_{35/39} (*sun*); 41) *хвост*_{26/27} (*tail*); 42) *ты*_{15/15} (*thou*); 43) *язык*_{8/8} (*tongue*); 44) *зуб*_{22/22} (*tooth*); 45) *дерево*_{37/42} (*tree*); 46) *два*_{2/2} (*two*); 47) *вода*_{28/30} (*water*); 48) *мы*_{1/1} (*we*); 49) *что*_{12/12} (*what*); 50) *кто*_{6/6} (*who*).

Сразу же отметим, что стандартная лексикостатистическая формула С. А. Старостина, исходящая из значения коэффициента $\lambda = 0,05$, для 50-словника неприменима; если для полного объема 100-словного списка калибровка показывает среднюю скорость «5% внутренних замен за 1000 лет», то для сверхустойчивой половины этого списка скорость должна быть ниже. Для большей надежности все подсчеты по 50-словному списку производятся по алгоритму С. А. Старостина, интегрированному в компьютерную лингвистическую среду StarLing, в ходе которого для каждой сравниваемой пары языков вычисляется индивидуальный коэффициент, зависящий от среднестатистической стабильности тех элементов списка, которые в этой паре совпадают или различаются. Сопоставительный анализ классификационных схем и глоттохронологических датировок по большинству языковых

¹ Элементы стословника как в данном варианте списка, так и в составе конкретных разделов, посвященных разбору материала отдельных групп, по техническим причинам приводятся в алфавитном порядке их англоязычных эквивалентов (т. е. *ashes, bird, black* и т. п.).

семей Евразии, доступных для лексикостатистической обработки, показывает, что в целом результаты обсчитывания 50-словника по индивидуальным коэффициентам не сильно отличаются от аналогичных результатов с использованием более традиционной формулы С. А. Старостина, но с заменой значения $\lambda = 0,05$ на $\lambda = 0,04$. Это наглядно подтверждает высказанное выше интуитивное соображение и исторически интерпретируется как то, что для 50-словника в среднем следует ожидать выпадения в ходе внутренних замен не более чем двух слов за одну тысячу лет.

1.7.2. *Первичный отбор конституентов генетической классификации.* Следующий этап, после того, как определен общий состав списка — разбиение всего множества языков, к которому мы хотим применить лексикостатистические алгоритмы, на непересекающиеся «подмножества 1-го порядка», т. е. задать некие критерии, по которым языки будут разбиваться на множества $\{A_1, A_2 \dots A_n\}$, $\{B_1, B_2 \dots B_n\}$ так, что каждое из этих множеств может быть объединено в пределах одной базы данных и конвертировано в единое генеалогическое древо на основании единой лексикостатистической матрицы, но объединение *обоих* этих множеств при этом будет допустимо уже только на уровне *празыков*, т. е. должно будет иметь вид $\{*A, *B\}$.

Поясним, что конкретно имеется в виду. По умолчанию для такой важной задачи, как построение генетической классификации языков африканского континента (пусть даже предварительной) на базе унифицированного сравнительного инвентаря и единой формальной методологии, ожидалось бы уже на самом первом этапе применение именно этой и никакой другой методологии. Т. е., условно говоря, нельзя однозначно утверждать, что, например, языки ари и диме относятся к южноомотской группе, а языки менде и лоома — к юго-западным манде до тех пор, пока не будут обсчитаны все бинарные корреляции между этими языками. По сути, «чистое» применение метода предполагает, что первоначально стословные списки будут составлены для *всех* языков континента, и только после этого можно запускать ручные и/или автоматические алгоритмы составления лексикостатистической матрицы.

Процедуру такого рода, однако, чрезвычайно трудно осуществить в чисто техническом плане; к тому же нет серьезных

оснований считать, что строгое ее выполнение приведет к результатам, существенно отличным от тех, которые уже получены лингвистами в ходе детального (или даже поверхностного) применения обычного сравнительно-исторического метода к «мелким» языковым таксонам африканского континента. В частности, что касается только что приведенного конкретного примера, нет особого смысла тратить время на «первичное» лексикостатистическое обоснование единства менде-лоома, противопоставленного единству ари-диме, т. к. это будет лишь еще одним обоснованием того, что уже было неоднократно и ко всеобщему удовлетворению обосновано в различных работах по африканистике.

Введем понятие «языкового таксона 1-го уровня», с которого собственно и будет начинаться этимологический и статистический анализ, т. е. единицу, которую можно в рамках нашего исследования считать своеобразным «конституентом». Представляется, что такой таксон:

а) должен быть *интуитивно мыслимым*, т. е. все языки, входящие в него, в своей базисной лексике (и, факультативно, в инвентаре грамматических морфем) должны демонстрировать высокий процент фонетикосемантического сходства (этот неформальный критерий мы заимствуем из «массового сравнения»);

б) должен быть *общепринятым*, т. е. среди специалистов по языкам соответствующего ареала необходимо наличие консенсуса относительно того, что все без исключения языки данного таксона действительно восходят к общему предку;

в) должен, в нашей терминологии, являться *группой*, т. е. демонстрировать в пределах стословного списка не менее (желательно более) 40% лексических совпадений, устанавливаемых на базе фонетических соответствий (если к группе уже применялся сравнительно-исторический анализ) или согласно критерию фонетической совместимости.

Из этого определения вытекает и конкретная процедура составления исходного списка таксонов 1-го уровня. Отгалкиваться необходимо от уже существующих материалов: классификационных схем, сопоставительных анализов, праязыковых реконструкций и т. п., предложенных специалистами и являющихся рабочими моделями для современных исторических исследований по языкам Африки. Таксону, «подозреваемому» на принадлежность

к 1-му уровню на основании критериев (а) и (б), ставится в соответствие автономная лексикостатистическая база данных, анализ которой, согласно критерию (в), должен окончательно подтвердить (или — в исключительных случаях! — опровергнуть) его генетическую реальность.

Таким образом, можно обойтись без предварительного сопоставления языков ари и диме с языками менде и лоома (равно как и с подавляющим большинством других языков Африки) хотя бы уже из-за того, что объединение их в рамках «таксона 1-го уровня» не проходит по критерию (б): консенсус среди специалистов по поводу возможности близкого родства между этими языками отсутствует, более того, насколько известно, никто вообще никогда не выдвигал подобной гипотезы. Даже не глядя на конкретные материалы по этим языкам, логично предположить, что отсутствие таких гипотез напрямую связано с нарушением критерия (а) — никаких интуитивно очевидных схождений, достаточных для постулирования ближнего родства, между этими языками не наблюдается, в противном случае такие гипотезы были бы уже давно выдвинуты и консенсус, скорее всего, был бы достигнут.

Предположим, впрочем, что мы *ошибаемся*, и близкое родство между ари-диме и менде-лоома на самом деле существует, а незамеченным оно оставалось, например, из-за значительной географической отдаленности этих языков, в результате чего исследователям просто не приходило в голову сопоставлять их напрямую (такие ситуации даже в современном сравнительно-историческом языкознании в принципе не исключены). В этом случае вместо одного таксона 1-го уровня мы ошибочно постулируем два. Однако «смертельной» такого рода ошибка не будет — она непременно будет исправлена при переходе с 1-го на 2-й уровень сравнения, поскольку реконструированные пятидесятисловные инвентари условных «пра-ари-диме» и «пра-менде-лоома» будут в значительной степени совпадать (как будут в значительной степени совпадать, например, лексические варианты «праславяно-германского» и «праитало-индийского», реконструированные независимо друг от друга).

В любом случае саму вероятность допущения такой ошибки можно считать а priori ничтожной: корректно идентифицировать таксон 1-го уровня, или «группу», при наличии достаточного

количества данных по составляющим его языкам несложно, и последовательный анализ языков Африки, в той или иной степени уже попадавших в поле зрения исследователей, показывает, что подавляющее большинство этих таксонов установлено достоверно. Обратная ошибка, т. е. ложное объединение дальнородственных или неродственных языков в близкородственный таксон, едва ли возможна в принципе, а если что-либо подобное и случается, то речь обычно идет лишь о конкретном случае терминологической неопределенности.

Так, для Дж. Гринберга в его работе 1963 г. *семьей (family)* является и «нило-сахарский» таксон, и входящий в его состав «шари-нильский», и входящий в состав «шари-нильского» «восточносуданский», и входящий в состав «восточносуданского» «нилотский» — при этом, однако, даже nilотские языки, как в этом нетрудно убедиться в ходе предварительных лексикостатистических подсчетов, не составляют таксона 1-го уровня; таковыми можно считать лишь три отдельных «ветви» nilотского — западную, южную и восточную, генетическое родство которых, в отличие от всех узлов более высокого уровня, интуитивно очевидно, общепризнано и легко демонстрируемо в рамках стандартного применения сравнительно-исторического метода.

Итак, для выделения языковых «групп» как начальных объектов сопоставления, на наш взгляд, вполне достаточно опоры на классификации, полученные специалистами как в ходе строгого сравнительно-исторического исследования, так и «на глазок», в ходе массового сравнения. Допущенные при этом ошибки будут неизбежно скорректированы на дальнейших этапах работы.

Отдельно стоит коснуться вопроса о языках-изолятах. Как известно, «изолят» — понятие относительное и означает не обязательно отсутствие родства «изолята» с каким-либо другим языком вообще, а лишь отсутствие его в рамках таксона уровня x ; бывают «изоляты уровня группы» (армянский в рамках индоевропейского, туцзя в рамках сино-тибетского и т. п.), «изоляты уровня семьи» (бурушаски в рамках сино-кавказского, египетский в рамках афразийского и т. п.) и, наконец, наиболее загадочные из всех «изоляты уровня макросемьи», действительно не обнаруживающие прочных связей ни с какими другими языковыми объединениями планеты (такие, как шумерский или нихали).

По-видимому, корректное применение методики требует составления отдельных списков для всех «изолятов уровня группы», каковых на африканском континенте обнаруживается довольно много, и сопоставления их с праязыковыми списками «многочленных» групп уже на следующем этапе сравнения. Так, даже несмотря на то, что, например, язык-изолят западный ꜥхоан, скорее всего, объединяется в общий таксон с северно-койсанской (жу) группой языков, а язык-изолят нара — с нубийской группой, лексикостатистические (равно как и грамматические) различия между этими изолятами и их ближайшими родственниками на порядок выше, чем различия внутри языков этих групп. Для того, чтобы правильно оценить степень родства, необходимо *сначала* тщательно рассмотреть лексический материал этих таксонов по отдельности, реконструировать необходимые прасписки (разумеется, только для «многочленных» групп — для изолятов реконструкция, за исключением рискованных попыток внутренней этимологизации, невозможна), и только затем переходить к следующему этапу.

Впрочем, в отдельных, очень редких, случаях перескочить через один этап (точнее, через «пол-этапа») кажется допустимым. Речь идет о ситуациях строго бинарного ветвления, когда язык-изолят представляет собой одну из двух ветвей таксона, промежуточного между семьей и группой (можно назвать его «макрогруппой» или «микросемьей»), причем обнаруживаемые между ним и «многочленной» ветвью макрогруппы лексические и грамматические изоглоссы настолько многочисленны и наглядны, что родство является общепризнанным, даже несмотря на то, что проценты совпадений в стословном списке не доходят до требуемых 40%, и родство может не быть интуитивно очевидным.

Таков, например, случай языка маджанг, составляющего северную ветвь сурмийской макрогруппы, противопоставленную южной ветви, в которую входят примерно 10 гораздо более близких друг к другу языков. Родство маджанг с южносурмийской общностью было опознано не сразу, и некоторое время этот язык изучался просто в общем ареальном контексте языков Эфиопии, сопоставляясь и с нилотскими, и даже с афразийскими (кушитскими, омотскими) языками. В настоящее время, однако, систематизация сравнительных данных по региону убедительно

показывает, что с южносурмийскими его связывают намного более тесные лексико-грамматические изоглоссы, чем с любыми другими языками ареала, и статус его как наиболее раннего обособления от общесурмийского ствола не оспаривается ни одним специалистом.

В подобного рода случаях, скорее из соображений экономии места, чем общетеоретических, можно позволить себе сразу составить единый 100-словный список по такого рода таксону «1½ уровня», не опасаясь последствий; опять-таки подчеркнем, что *ошибочное* определение такого таксона (предположим, что у маджанг в Африке есть более близкие родственники, чем южносурмийские языки) будет непременно скорректировано на более высоком этапе сравнения (маджанг будет в этом случае стабильно показывать более высокий процент совпадения со своими «истинными» ближайшими родственниками, чем с пра-южносурмийским).

Общий вывод заключается в том, что членение анализируемого языкового массива на «таксоны 1-го уровня» — процесс, не лишенный элемента субъективности, и в силу этого не застрахованный на сто процентов от получения ошибочных результатов; однако дальнейшие этапы анализа, основанные на гораздо более формализованном подходе, гарантируют коррекцию любых ошибок, допущенных в ходе первичного «нарезания пирога» на минимальные сегменты.

1.7.3. *Лексикостатистический анализ 1-го уровня.* На этом этапе решается несколько важнейших задач, подробное документирование которых в нашем исследовании опущено из соображений объема, но во многом отражено в структуре и комментариях к компьютерным базам данных, постепенно публикуемым на сайте «Глобальная лексикостатистическая база данных». Это:

а) составление 100-словных списков по таксонам 1-го уровня (одному таксону соответствует одна база данных). Списки составляются для *всех без исключения* языков таксона, по которым в нашем распоряжении имеются опубликованные¹ лексические

¹ Предпочтение неизменно отдается именно опубликованным данным: словарям, словарным спискам, фонетическим и грамматическим описаниям, подкрепленным значительным количеством примеров,

данные, достаточные для включения языка в общую систему статистических подсчетов; из ста элементов списка требуется заполнение не менее половины позиций (для языков, имеющих близких родственников, представленных полными или почти полными списками), в отдельных случаях — не менее 70-80% списка (в первую очередь — для языков-изолятов; здесь низкая степень заполненности списка может легко исказить достоверность классификационных результатов даже на неглубоких уровнях сравнения).

Подчеркнем, что в рамках принятой методики исключительную важность имеет опора на массовость сравнения (см. раздел 1.5); «выборочное» составление списков, хотя и сокращает общий объем работы, может иметь крайне негативное влияние на результаты следующего этапа (реконструкция прасписков), не говоря уже о том, что практика работы со стословными списками почти неизменно демонстрирует повышение точности результатов в зависимости от полноты охвата языковой картины анализируемой семьи.

Отдельная проблема — степень достоверности используемых данных, особенно актуальная там, где источник легко заподозрить в неточности как фонетической записи, так и семантических эквивалентов. По возможности следует стремиться к опоре на максимально современные описательные работы (хотя и здесь иногда случается встретиться с ярко выраженной «непрофессиональной» записью материала); «старые» источники в этом случае можно не учитывать при составлении собственно списков,

корпусам текстов и т. д. Лишь в отдельных случаях используются рукописные, официально не попавшие в печать, материалы (отчасти из-за того, что доступ к таким материалам, как правило, ограничен, отчасти из-за «сырости» данных). Материалы, опубликованные в электронной версии в сети Интернет, приравниваются к печатным в тех случаях, когда речь идет о «законченных» трудах, прошедших первичную редакторскую обработку и имеющих определенный формальный статус (сведения об авторе, дата публикации и т. п.). «Анонимные» данные без адекватной верификации используются лишь в исключительных ситуациях (например, там, где они являются единственным источником информации по языку-изоляту, имеющему большое значение для общей классификации языков соответствующего ареала) и специально оговариваются.

но допустимо и рекомендуемо пользоваться их данными как контрольными на этапе этимологического анализа.

Если, кроме «сомнительных» источников, лексические данные по тому или иному языку достать не удастся, решение о возможности включения этих данных в списки принимается в зависимости от ситуации. Так, например, для южнокойсанской группы языков, которая в XIX — первой половине XX в. еще была представлена довольно значительным числом сильно отличающихся друг от друга наречий, современные лексикографические описания доступны только для языка !хонг (и, отчасти, для недавно «переоткрытого» н|уки); все остальные языки вымерли и известны нам исключительно по некачественным (как правило) словарным и грамматическим описаниям, отказ от которых автоматически означал бы исключение южнокойсанского таксона из выстраиваемой классификационной модели, что недопустимо.

Тем не менее, стандарты разрабатываемой модели представляются достаточно гибкими, чтобы учитывать возможность использования такого рода «некачественных» записей. Во-первых, упор на базисную лексику во многом упрощает проблему, особенно в отношении семантики (ошибочная запись стословных значений в целом встречается реже, чем запись более сложных и менее устойчивых в семантическом плане лексем). Во-вторых, замена критерия *фонетической регулярности* на критерий *фонетической совместимости*, который требует от исследователя не обнаружения строгих «механистических» корреляций, а выстраивания непротиворечивых и достоверных исторических сценариев, означает, что многочисленные мелкие ошибки в транскрипционной записи могут оказаться несущественными. Поскольку речь идет не о реконструкции, полностью соответствующей классическим требованиям сравнительно-исторического метода, а о своеобразной «прикидке», правдоподобной в типологическом отношении и эксплицитно неопровержимой, нет ничего страшного в том, чтобы опираться, например, на данные источника, игнорирующего те или иные фонологические оппозиции, некорректно записывающего (или вообще не записывающего) просодические (тональные) характеристики анализируемого языка и т. п. Условно говоря, если, например, низкое качество данных по южнокойсанскому языку |хам все же позволяет однозначно отнести его

именно к южнокойсанской, а не какой-либо другой, группе, то нет никаких априорных причин считать, что определить его дальнейшие, более глубокие, родственные отношения это же качество уже не позволит;

б) первичная лексикостатистическая обработка списков таксонов 1-го уровня, т. е. сопоставление каждого слова с *индексом когнации*, единым для этимологически родственных слов и различающимся для слов, имеющих разное происхождение.

Когнации проставляются «вручную» (несмотря на то, что разработаны и простейшие алгоритмы автоматического опознавания когнатов, о которых будет говориться ниже) в соответствии с критерием фонетической регулярности (если для анализируемого таксона имеется список регулярных соответствий, выработанный в ходе применения к нему сравнительно-исторического метода) или фонетической совместимости (если детальный исторический анализ таксона отсутствует). Общие правила расстановки когнации в нетривиальных ситуациях в целом описаны выше (п. 1.6.3).

Конечным результатом этого этапа является построение лексикостатистической матрицы, содержащей информацию о процентном совпадении базисной лексики между каждой парой языков таксона, и, на основании матрицы, его генеалогического дерева (как матрица, так и дерево строятся автоматически на базе компьютерной среды StarLing, разработанной С. А. Старостиным). Одновременно с этим верифицируется и предположительный «возраст» таксона (по глоттохронологической формуле Сводеша — Старостина); для большинства таксонов 1-го уровня он не должен превышать 3,000-4,000 лет (предельный возраст для «группы», за которым лексикостатистику уже рекомендуется проводить на уровне сопоставляемых праязыков).

Полученное классификационное дерево имеет смысл уже на данном этапе сопоставить с классификационными моделями, предлагавшимися для данного таксона ранее, как на основании альтернативных лексикостатистических подсчетов, так и более традиционных, субъективных методов, связанных с оценкой «общих инноваций». При обнаружении значительных расхождений между «старой» и «новой» классификациями (особенно в тех случаях, если «старая» классификация основана на не зависящих друг от друга исследованиях более чем одного специалиста)

ситуация требует серьезного дополнительного анализа, включающего как повторную проверку корректности составления списков и определения когнаций, так и, наоборот, тестирование на достоверность аргументации, лежащей в основании «старой» классификации. Предположение о том, что такие ситуации должны возникать исключительно редко, и что тщательный их анализ должен и будет приводить не к дискредитации лексикостатистического метода, а к его усовершенствованию (см. раздел 1.6.2 о лексикостатистических аномалиях), на материале языков Африки в целом выдерживается.

1.7.4. *Подготовка прасписков 1-го уровня.* Для того, чтобы от построения классификаций на уровне групп перейти к более глубокой хронологии, необходимо максимально строго определить тот конкретный инвентарь форм, который может быть вынесен на «второй уровень» лексикостатистики — сопоставление в рамках *семьи*. На этом этапе происходит переход от 100-словных списков к 50-словным: «нестабильная» половина списка отбрасывается как избыточная (что совершенно не исключает возможности вернуться к ней на каком-либо будущем этапе или учитывать ее данные как «вспомогательные»), а для «стабильной» требуется предварительная реконструкция прасписка.

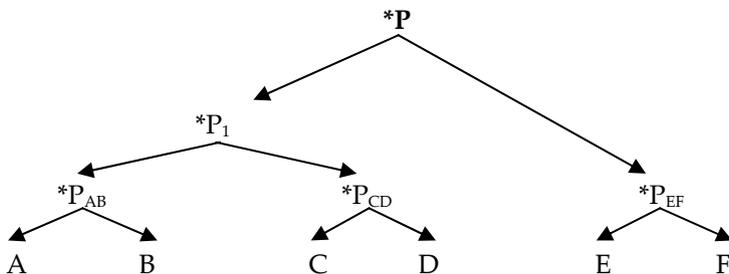
Подчеркнем, что основной целью реконструкции прасписка в нашем исследовании является не доскональное описание фонетических соответствий, включая редкие и нетривиальные, а определение наиболее вероятного множества *этимонов*¹, являвшихся в праязыке основными носителями значений элементов 50-словника. Формы под звездочкой, полученные на выходе, должны удовлетворять в первую очередь критерию фонетической совместимости в том виде, в котором он был описан в разделе 1.4. Это тем более важно, что для многих языков мира вообще и Африки в частности количество и качество лексических данных просто не позволяет провести «полноценную» реконструкцию.

¹ Под *этимон*ом в общем случае понимается скоррелированная пара элементов «звучание (X) : значение (Y)», восстанавливаемая для праязыка на основании рефлексов, сохранившихся в языках-потомках. От *корня* или *основы* этот термин отличается лишь тем, что предполагает не только однозначное фонетическое, но и семантическое содержание; на практике, впрочем, эти понятия во многих контекстах взаимозаменяемы.

Гораздо большее значение на данном этапе имеет корректный учет *дистрибуции* разных корней с одинаковой базисной семантикой по языкам таксона 1-й группы. Точно так же, как на этапе составления стословного списка важнейшей (и, как правило, решаемой) задачей является минимизация (по возможности — полная ликвидация) синонимии, при реконструкции прасписка требуется во что бы то ни стало минимизировать разнообразие вариантов, исходя из той теоретической предпосылки, что абсолютная синонимия на праязыковом уровне должна была носить столь же исключительно редкий характер, какой она имеет на уровне эмпирически наблюдаемых живых языков.

Насколько нам известно, подробных формальных попыток описания «ручного алгоритма» выбора оптимального праязыкового этимона в истории компаративистики до сих пор не предпринималось; отчасти это объясняется скептическим отношением к лексикостатистике, формальные возможности которой явно недооценены в среде компаративистов, и традиционной непопулярностью строгой семантической реконструкции. Важнейшее исключение из общей тенденции — работа Л. Когана [Kogan 2006], в которой такая попытка предпринята впервые и апробирована на материале семитских языков; тем не менее, предложенный Л. Коганом алгоритм не исчерпывает весь потенциал сравнительно-исторического метода и поэтому нуждается в ряде дополнений, частично перечисленных в статье [Starostin 2010].

Чтобы описание процедуры реконструкции прасписка имело более наглядный характер, определим ее на конкретизированном объекте: таксон 1-го уровня (т. е. «группа»), лексикостатистическая классификация которой соответствует следующей древесной структуре:



Вопрос о хронологической составляющей длины узлов в данной ситуации не является принципиальным, равно как и вопрос о количестве «уровней»; очевидно, что для таксона 1-го уровня, возраст которого, по нашему определению, не должен превышать трех-четырёх тысяч лет, многоярусная система ветвления маловероятна, но даже если анализируемый таксон показывает более сложное ветвление, описываемые ниже правила будут применимы для любой степени ветвления на рекурсивной основе (т. е. этимон, реконструированный для уровня * R_1 , проецируется «вверх» как один из возможных кандидатов на статус * R и т. д.). Аналогичным образом описываемые правила (точнее, необходимое их подмножество) вполне пригодны и для менее сложных ситуаций, чем та, которая представлена данным деревом.

При условии (более или менее) корректного составления 100-словных списков и проставленности индексов когнации механизм реконструкции прасписка будет определяться двумя наборами правил: *дистрибуционными* правилами, принимающими во внимание только распределение потенциальных рефлексов праязыковых этимонов по конкретным языкам, и *экстра-дистрибуционными* правилами, способными прояснить ситуацию в тех случаях, где не удастся принять однозначного решения исключительно на основе дистрибуционных правил.

Перечислим основные *дистрибуционные* правила реконструкции для праязыка P этимона * E_P , имеющего 100-словное значение S :

Д.1. Если значение S во всех языках таксона выражено одним и тем же корнем (т. е. $E_A = E_B \dots = E_F$), этот же корень в значении S реконструируется и на праязыковом уровне, т. е. $E_A = *E_P$.

Пример: праславянское **golva* 'голова' однозначно реконструируется на основании того, что более или менее регулярные рефлексы этой основы сохраняются во всех без исключения славянских языках. (Правило практически совпадает с формулировкой в [Kogan 2006: 465]: «If a PS (Proto-Semitic — Г. С.) root functions with the same basic meaning in all Semitic languages, there is hardly any reason to doubt that it did so also in the proto-language»).

Д.2. Если (а) значение S выражено одним и тем же корнем *хотя бы* в одном из языков *каждого* из минимальных узлов таксона и (б) для остальных языков этимологических совпадений между минимальными узлами не обнаруживается (т. е., например, $E_A =$

$E_C = E_E$, но $E_A \neq E_B \neq E_D \neq E_F$), общий для всех узлов корень в значении S реконструируется и на праязыковом уровне, т. е. $E_A = E_C = E_E = *E_P$. Отметим, что на уровне *групп* такие ситуации, тем более в пределах «сверхстойчивого» 50-словного списка, на практике встречаются очень редко, однако при переходе на более глубокий хронологический уровень вероятность встречи с ними значительно повышается.

Пример: при объединении бинарных узлов «диалект Токио / диалект Кагосима» и «диалект Сюри / диалект Хатерума» на праяпонском уровне (эти четыре диалекта выбраны как показательные; полная классификация японских диалектов, разумеется, гораздо более сложная) изоглосса 'язык' объединяет дальнородственные диалекты Токио (*shita*) и Хатерума (*sita*), при отдельном корне для Кагосима (*bero*) и отдельном для Сюри (*šiba*). На праяпонском уровне при этом закономерно восстанавливается **sita* [С. Старостин 1991: 263, 268], несмотря на то, что надежной этимологизации ни для *bero*, ни для *šiba* не обнаружено; т. е. в принципе допустимо, что на праяпонском уровне основным 'языком' было одно из них, но это допущение предполагает неэкономный исторический сценарий (независимый семантический переход **sita* '?' → 'язык' в нескольких диалектных пучках), дополнительные аргументы для поддержки которого отсутствуют, и поэтому не может получить статус рабочей гипотезы.

Д.3. Если (а) значение S выражено одним и тем же корнем *хотя бы в двух языках*, относящихся к двум *разным первичным* узлам таксона (т. е. узлам, отражающим первое хронологическое разделение праязыка *P) и (б) подобное совпадение *единично*, т. е. ни одной другой межузловой изоглоссы такого же рода не обнаружено (например: $E_A = E_E$, но $E_B \neq E_E$, $E_C \neq E_E$, $E_D \neq E_E$, $E_B \neq E_F$, $E_C \neq E_F$, $E_D \neq E_F$), на праязыковом уровне в требуемом значении S восстанавливается именно этот общий корень (т. е. $E_A = E_E = *E_P$). На уровне *группы* это правило является чисто теоретическим: ситуации такого рода, когда «старый» корень сохраняется лишь на периферии языкового пространства группы, но при этом не менее чем в двух разных точках этой периферии, практически отсутствуют. Однако при переходе от таксонов 1-го уровня к уровням более высоким их встречаемость опять-таки значительно повышается.

Применение дистрибуционных правил для реконструкции праэтимона достаточно в *неконкурентных ситуациях*, условия которых изложены в подпунктах (б) правил Д.2 и Д.3: межъязыковые изоглоссы либо ограничены единственной параллелью вообще (Д.1, Д.2), либо единственной межузловой параллелью (Д.3; при наличии изоглосс (1) $E_A \neq E_B = E_E \neq E_F$ и (2) $E_C = E_D$ на праязыковой статус может претендовать только (1), т. к. альтернативный сценарий будет менее экономен).

Нередкими даже на уровне групп, а тем более — на уровне семей и макросемей, однако, являются *конкурентные ситуации*, когда дистрибуция корней по языкам не позволяет сама по себе сделать однозначный выбор. Эти ситуации, в свою очередь, можно поделить на *непересеченно-конкурентные* и *пересеченно-конкурентные*.

Более естественными и частотными являются, как правило, *непересеченно-конкурентные* ситуации, в которых n корней, выражающих одно и то же базисное значение S , оказываются распределены по n первичным узлам дерева (чаще всего $n = 2$, но нередки и случаи первичного разделения праязыка на три и более ветви, в каждой из которых требуемое значение может иметь собственный звуковой эквивалент). *Пример*: значение 'мясо' из трех первичных узлов германской группы имеет эквивалент (1) *timz* в готской ветви, (2) **flaiskaz* в восточной ветви, (3) **kjot* в северной ветви; очевидно, что правила Д.1-Д.3 не дают возможности отобрать для прагерманского уровня наиболее вероятную кандидатуру.

Простейший и в целом оправданный выход из положения — признать невозможность адекватного заполнения соответствующей позиции в прасписке («прочерк»). Такое решение, однако, приемлемо лишь в том случае, если мы не собираемся использовать полученную информацию о базисной лексике праязыка на более глубоком уровне сопоставления. В этом случае можно просто удовлетвориться констатацией факта, что для праязыка *P, применив дистрибуционные правила Д.1-Д.3, удалось восстановить n элементов 100- или 50-словного списка, и на этом остановиться (такое решение, в частности, принимает Л. Коган для семитского материала, в конечном итоге постулируя возможность фонетико-семантической реконструкции лишь для 52 этимонов прасемитского списка [Kogan 2006: 483]).

Если, однако, реконструкция прасписка на уровне *a* группы (семьи) представляет собой промежуточный этап на пути к реконструкции прасписка на уровне *a+1* семьи (макросемьи), ограничиваться лишь теми этимонами, которые реконструируются заведомо однозначно, на основании правил Д.1-Д.3, было бы рискованно, а в худшем случае — просто бесполезно. Самоограничение такого рода может, на первый взгляд, выглядеть демонстрацией здорового научного скепсиса, но на самом деле является скорее классическим образцом гиперскептицизма, препятствующего не только перетеканию хода мыслей исследователя в область субъективных фантазий, но также и, наоборот, построению вполне реалистичных и высоковероятных историко-лингвистических сценариев.

На самом деле в ситуациях, аналогичных случаю со словом 'мясо' в германских языках, ключевым утверждением должно быть не негативное «для прагерманского 100-словного списка невозможно заполнение конкретным этимоном позиции 'мясо'», а позитивное «позиция 'мясо' в прагерманском 100-словном списке может быть с *примерно равной* вероятностью занято одним из трех корней, представленных в трех первичных ветвях этой группы»¹. Первое утверждение сознательно или подсознательно подталкивает к тому, чтобы *вообще* отказаться от учета данных по этой позиции в списке, как на уровне внутренней, так и внешней реконструкции; второе, напротив, позволяет при соблюдении определенных мер предосторожности надежно расширить ту материальную базу, на которой будут выстраиваться дальнейшие уровни классификации.

Итак, в ситуации, когда $(E_A = E_B) \neq (E_C = E_D) \neq (E_E = E_F)$, вместо единого E_P следует постулировать *три* гипотетических сущности: $E_A = E_B = *E_{1P}$, $E_C = E_D = *E_{2P}$, $E_E = E_F = *E_{3P}$. Про $*E_{1P}$, $*E_{2P}$, $*E_{3P}$ можно сказать, что и та, и другая, и третья реконструкции могли быть реальными корнями праязыка $*P$, но *только один* из них мог

¹ Возможен (в отдельных случаях даже доказуем) и еще один вариант, когда в праязыке группы требуемый этимон был на самом деле представлен корнем, *отличным* от рефлексов во всех языках-потомках. На практике, однако, «бритава Оккама» рекомендует игнорировать эту вероятность во всех тех случаях, когда для такой версии не удастся представить ни одного конкретного аргумента.

представлять собой этимон, являющийся элементом множества «100-» или «50-словный список»¹. Для того, чтобы оценить примерную относительную вероятность каждой из «кандидатур» на праязыковой этимон, во многих случаях можно прибегнуть к применению *экстра-дистрибуционных* правил, требующих дополнительной исторической проработки материала. Полный список таких правил еще предстоит установить, но уже сегодня накоплен определенный опыт анализа и реконструкции, позволяющий предложить хотя бы в общем виде некоторый предварительный инвентарь.

Ключевым подспорьем в этом вопросе может быть успешная *взаимная этимологизация* конкурирующих основ, т. е. в нашем случае — обнаружение этимологических (но не лексикостатистических) параллелей к основе праязыка *P_{AB} в праязыках *P_{CD} / *P_{EF}

¹ Вероятность «транзитной синонимии» (см. выше) в данной ситуации должна быть близка к нулевой, т. к. сама идея транзитной синонимии предполагает, что вытеснение старого этимона новым *неизбежно* и является лишь вопросом времени (нескольких поколений). Реконструировать для праязыка «транзитную синонимию» можно только в том случае, если старый этимон в языках-потомках сохраняется с измененной семантикой, ограниченной (связанной) контекстной дистрибуцией или в качестве архаизма.

Так, например, общеиндоевропейская основа **ōis-* 'ухо' в праиранском языке в целом вытесняется основой **gauša-*, производной от индоиранского глагольного корня **ghauš-* 'слышать(ся), звучать' [Расторгуева & Эдельман 2007: 247], но в авестийском языке, наряду с «новой» лексемой **gaōša-*, при этом еще сохраняется старая форма двойственного числа *ušī*, в основном применительно к животным или в переносном значении 'слух, восприятие' [Bartholomae 1961: 414]. Это означает, что на определенном этапе развития праиранского языка в нем, скорее всего, существовала транзитная синонимия (**uš-* → **gaōša-*), что хорошо согласуется с общеиндоевропейской перспективой. Тем не менее, нет никаких эксплицитных аргументов в пользу того, что носители праиранского языка «накануне» своего первого разделения на ветви активно пользовались основой **uš-* как базовой; напротив, все данные иранских языков свидетельствуют скорее в пользу обратного. Поэтому при заполнении соответствующей позиции в праиранском 100- или 50-словном списке учитываться будет только основа **gaōša-*; связь же ее с семантически близкой основой **uš-* может и должна определяться как типовая семантическая корреляция.

и наоборот. На уровне группы такую этимологизацию, как правило, удается обнаружить довольно легко, при условиях, что (а) требуемые параллели действительно *есть*, т. е. «базисный» этимон в той ветви, в которой он перестал быть «базисным», не был ликвидирован полностью, а сохранился в видоизмененном значении; и (б) в нашем распоряжении достаточно словарных данных, чтобы суметь доказать (а).

Если взаимные этимологические параллели отсутствуют, еще один способ расположить «кандидатуры на базисность» в порядке возрастания или убывания вероятности их представленности в 100-/50-словном списке праязыка — ограниченное применение *внутренней реконструкции*, которая в ряде случаев может подсказать первичность или вторичность происхождения той или иной основы.

Наконец, третий возможный источник снятия/ограничения «конкуренции» — тщательный учет *ареальных связей* рассматриваемой группы. Несмотря на то, что начинать проводить его нужно уже на предыдущем этапе, когда из собранных списков вычищаются наиболее очевидные заимствования, при разрешении конкурентных ситуаций часто бывает полезно вернуться к этому вопросу еще раз, т. к. далеко не все случаи заимствования (особенно если речь идет о языковых группах, лишь относительно недавно ставших предметом исторического рассмотрения) удается корректно идентифицировать до начала собственно сравнительно-исторического анализа, имеющего целью реконструкцию.

Перечислим теперь основной набор экстра-дистрибуционных правил, применимых к ситуации *бинарного* противопоставления, т. е., согласно общей схеме, к случаю, когда, например, $E_A \neq E_B$ и, следовательно, вопрос о сущности $*E_{(PAB)}$ не может быть решен на основании только дистрибуционных правил. Для наглядности проиллюстрируем этот набор на материале этимологически исследованных бинарных групп (т. е. либо реально включающих всего два языка, либо таких, для которых надежные данные по базисной лексике существуют всего для двух языков); в дальнейшем соответствующие правила можно применять рекурсивно.

Э.1. Если для E_A зафиксирована полисемия, т. е. «стословное» значение S_1 + близкое значение S_2 , а для E_B — только значение S_1 , то при обнаружении в языке В этимологического соответствия E_A

со значением S_2 естественно реконструировать на праязыковом уровне для E_A значение S_2 , а для E_B — «стословное» значение S_1 .

Пример: в коттском языке слово *halči:g* 'S₁ = ноготь'; 'S₂ = копыто' соответствует кетскому *qólés* 'S₂ = копыто'; значение 'S₁ = ноготь' в кетском передается неродственным корнем *iñ* [С. Старостин 1995: 195, 304]. Теоретически возможны две ситуации: (а) присутствие на праенисейском уровне основы $*\chi\lambda V[\check{c}]iG$ (енисейские реконструкции здесь и ниже следуют системе С. А. Старостина) с полисемией 'ноготь' / 'копыто', с последующим замещением в кетском на основу *iñ* неясного происхождения; (б) присутствие на праенисейском уровне основы $*\chi\lambda V[\check{c}]iG$ в значении 'копыто' и основы $*?i:\acute{i}-$ в значении 'ноготь', с последующим слиянием обоих значений в одной основе в коттском. При условии отсутствия дальнейших аргументов сценарий (б) предпочтительнее, т. к. обладает большей объяснительной силой (не требуется предполагать лишних неизвестных).

Подчеркнем, что реконструкцию праенисейского ногтя в виде $*?i:\acute{i}-$, а не $*\chi\lambda V[\check{c}]iG$ ни в коем случае нельзя считать *доказанной* (это же уточнение будет справедливым и для большинства остальных экстра-дистрибуционных правил, излагаемых ниже): речь идет исключительно о выстраивании гипотез в «оптимальном» порядке. При переходе к более высокому уровню сравнения этот порядок вполне может оказаться «перебитым» (см. об этом ниже). Однако чем лучше данные внешнего сравнения будут соотноситься с тем порядком оптимальности гипотез, который складывается на уровне сравнения внутреннего, тем более убедительный характер будет носить результирующая классификация.

Правило Э.1 можно распространить и на те случаи, когда корни E_A и E_B находятся в дополнительной дистрибуции, т. е. E_A вообще не имеет этимологических параллелей в B , и наоборот, хотя на практике такие ситуации встречаются предельно редко.

Э.2. Если для E_A зафиксировано «стословное» значение S_1 , а в языке B этот же корень имеет отличное от «стословного» значение S_2 , решение о первичности или вторичности значения S_1 в A может зависеть от данных семантической типологии. Если маловероятным или невозможным является семантическое развитие $S_1 \rightarrow S_2$ для языка A следует предполагать семантическую инновацию, и наоборот.

Пример: мегрельскому *dudi* 'голова' ← пракартвельск. **dud-* противостоит лазское *ti id.* ← **(s₁)taw-* [Климов 1964: 75, 175]. При этом лазскому *ti* в мегрельском регулярно соответствует основа *ti-*, имеющая целый ряд значений, выводимых из общей семантики 'начальник', 'глава семьи'. Метафорическое развитие 'голова' → 'начальник' для языков мира частотно и естественно (ср. русск. *глава*, французск. *chef* и т. д.), в то время как обратное развитие, даже если в каком-либо исключительном случае его и удастся обосновать, все равно будет уникальным.

Можно, таким образом, даже без привлечения дополнительных данных в качестве основной рабочей гипотезы считать более архаичным в значении 'голова' занский (лазско-мегрельский) корень **ti-*, а мегрельское *dudi* рассматривать как инновацию. В лазском языке основа *dud-* имеет значение 'кончик', что и разумно принять за первичную семантику этого корня, с опять-таки типологически естественным развитием 'кончик' → 'верхушка', 'голова'. Отметим, однако, что *это* семантическое развитие, в отличие от 'головы' / 'начальника', в языках мира бывает разнонаправленным (ср. *головка* какого-л. предмета), что не позволяет использовать его в качестве аргумента для правила Э.2.

Э.3. Если для E_A в стословном значении S зафиксирована *производная* основа, такая, что производность ее может быть продемонстрирована либо на синхронном уровне языка A , либо на диахронном уровне праязыка $*AB$, а для E_B в том же значении зафиксирована *простая* основа, это можно считать серьезным аргументом в пользу выведения на прауровень основы E_B .

Пример: слово 'кровь' в нильгирийской подгруппе южно-дравидийской группы выражается основой *netr* в языке ката и *po:x* в языке тода [Burgow & Emeneau 1984: 335, 482]. Первая из них на собственно нильгирийском уровне неразложима на составные компоненты; только дальнейшее внешнее сравнение показывает, что она на самом деле восходит к архаичному словосложению **ney-* 'жир' + **to:r* 'течь', но для пранильгирийской реконструкции эта информация нерелевантна. Вторая, напротив, даже на синхронном уровне тода может рассматриваться как продуктивное образование (через конверсию) от глагольной основы *po:x* 'вытекать', т. е. разумно предположить, что *po:x* первоначально имело узкое значение 'вытекающая (спущенная) кровь' с последующим

обобщением. Для пранильгирийского, таким образом, *первичной* кандидатурой на значение 'кровь (вообще)' будет основа, дающая в кота рефлекс *netr*.

Э4. Если для E_A в стословном значении S зафиксирована основа, для которой нельзя предложить конкретных аргументов в пользу ее заимствованного происхождения, а для E_B такие аргументы существуют, предпочтение при реконструкции должно отдаваться E_A .

Это правило нуждается в очень существенных пояснениях и ограничениях, т. к. при большом желании потенциальным «заимствованием» может быть назван (и, как показывает практика, часто называется) любой этимон, который в той или иной степени не укладывается в рамки исторической модели, предлагаемой исследователем — например, «нарушает» постулируемую им систему регулярных соответствий при том, что сама по себе достоверность конкретно этой системы никем не была строго верифицирована, и т. п.

Для того, чтобы иметь право «заподозрить» для E_B заимствованное происхождение, требуется, чтобы ситуация удовлетворяла хотя бы одной из нижеследующих комбинаций условий:

(1) язык B расположен в том же географическом ареале, что и язык-донор C (желательно, чтобы они граничили друг с другом); лексика языка B содержит множество слов $\{E\}$, фонетически совпадающих или регулярно соответствующих словам языка C , но близкого генетического родства с ним не демонстрирует, что видно из процентных соотношений в области базисной лексики; этимон E_B относится к множеству $\{E\}$. Это — идеальная ситуация для постулирования заимствования и исключения E_B из дальнейшей процедуры сравнения;

(2) то же, за исключением того, что языки B и C *не* имеют общих границ распространения или вообще располагаются в удаленных друг от друга ареалах. Эта ситуация очень редкая, на уровне групп — практически беспрецедентная, т. к. подразумевает, что период тесных контактов между B и C был прерван миграционными процессами (либо переселение носителей B , либо носителей C , либо «вклинивание» в пограничную между ними территорию носителей третьего языка D), которые обычно занимают достаточно много времени. На более глубоких уровнях

сравнения, однако, такой расклад обстоятельств вполне обычен (ср., например, тюркизмы в венгерском, закономерно отсутствующие в близкородственных венгерскому обско-угорских языках, или, на еще более высоком уровне — индоиранизмы в графинно-угорском);

(3) в базисной лексике языка В содержится множество слов {E}, фонетически изоморфных словам языка или языковой группы С, но не обнаруживающих регулярных соответствий с этой группой. Если при этом ни одно из слов этого множества не обнаруживает этимологических параллелей в языках, близкородственных В, речь может идти о «скрытом» субстратном влиянии языков группы С, в том числе и таких, которые не сохранились сами по себе, но оставили следы в виде соответствующих субстратных явлений (отсюда трудности в установлении соответствий контактного характера между живыми языками группы С и В);

(4) в базисной лексике языка В содержится множество слов {E}, ни одно из которых не имеет *прозрачной* этимологизации в близкородственных языках (т. е. такой, которая не предполагает ни нерегулярных соответствий, ни уникальных семантических развитий) и не списывается на контактное влияние со стороны какого-либо из эксплицитно зафиксированных языков, не состоящих в близком родстве с В. Такая ситуация, как и (3), может предполагать скрытый субстрат, с тем отличием, что последний вымер, не оставив не только «потомков», но и «близких родственников». Возможный пример такого «фантомного» субстрата обсуждался выше, в п. 1.6.3.5, на материале языка квегу.

Ситуации (1) и (2), при условии должного охвата материала, легко опознаются еще на этапе составления списков: такие очевидные контактные явления, как берберизмы в сонгай, эфио-семитизмы в восточносуданских языках, бантуизмы в койсанских языках и т. п. выявляются без серьезных методологических затруднений, хотя бы уже на основании того, что они, как правило, существенно нарушают принцип динамической градации. Сложнее обстоит вопрос с ситуациями (3) и (4), в которых вся аргументация носит сугубо косвенный характер.

На этапе расстановки индексов когнации решение вопроса о заимствованном или исконном характере «подозрительной» лексики не обязано быть однозначным. Допустимо техническое

решение, при котором этимон *ne* помечается как заимствование, если для него невозможно указать конкретный источник заимствования; при этом, однако, все такие случаи тщательно инвентаризируются, и если общая совокупность их превышает критическую массу — например, при исключении их из подсчетов изменяется место языка/праязыка в классификации — для рассматриваемой ветви/семьи строятся *два* альтернативных дерева, верифицируемые в дальнейшем уже на внешнем уровне. (Такой подход был бы продуктивен, например, для разрешения противоречий в классификации семитских языков, см. 6.2).

На этапе реконструкции 100- / 50-словных списков для *праязыков*, тем не менее, однозначное решение желательно как часть общей стратегии по минимизации псевдо-синонимии на прауровнях, ведущей к построению единого оптимального исторического сценария. Из этого следует последнее и, в определенном смысле, самое важное из экстра-дистрибуционных правил:

Э.5. Если для E_A в стословном значении S зафиксирована основа, имеющая этимологическую параллель (со значением S_1) в B , а для E_B в стословном значении S этимологических параллелей в A нет, предпочтение при реконструкции должно отдаваться E_A , поскольку для E_B более высока вероятность оказаться заимствованием из неизвестного источника.

Пример: при выборе между колымским юкагирским *роупə* 'белый' и тундренным юкагирским *я:we-* id. [Nikolayeva 2006: 291, 355] предпочтение при реконструкции будет отдано колымскому слову, т. к. *роупə* имеет этимологическую параллель в тундренном *роупе-* 'белесый, выцветший' (уточненное значение взято из словаря [Курилов 2001: 376]), а тундренное *я:we-* никаких следов в колымском не обнаруживает (и к тому же подозрительно как потенциальное заимствование из эвенского, хотя однозначно установить направление невозможно).

Хотя внешне правило Э.5 выглядит очень просто, при его применении возможны конфликты с правилами Э.1 и Э.2, для оптимального разрешения которых могут помочь данные типологии семантических изменений. Разберем следующую гипотетическую ситуацию, когда в языках A и B оказываются представлены такие два этимона:

	А	В
E ₁	'кровь'	'жир'
E ₂	—	'кровь'

Согласно правилу Э.5, при выборе наиболее вероятного эквивалента для праязыка *АВ мы должны отдать предпочтение этимону E₁, который по крайней мере в плане выражения однозначно проецируется на прауровень. Однако семантическая реконструкция 'кровь' для E₁ сталкивается с серьезной типологической проблемой: развитие 'жир' → 'кровь' для языков мира хорошо известно (ср. хотя бы упоминавшийся выше пример дравидийского **ney-to:r* 'кровь' ← 'жиро-течение'), обратное же развитие, по крайней мере, на данный момент времени вроде бы не зафиксировано ни в одном из языковых ареалов планеты. Следовательно, в данной ситуации оптимальным решением будет реконструкция E₁ в значении 'жир' и E₂ в значении 'кровь'.

Надежность такой реконструкции будет максимальной, если контрольный анализ обнаружит в рассматриваемых языках еще один этимон, например, с такими значениями:

	А	В
E ₃	'жир'	'мазать'

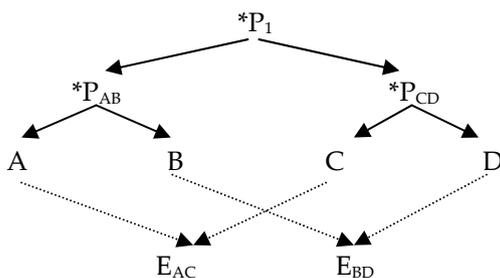
В этом случае исторический сценарий получает естественное логическое завершение: язык В сохраняет во всех трех случаях праязыковую семантику, в то время как в языке А имеют место параллельные развития: метонимическое 'мазать' ('мазь, мазилка') → 'жир' и «эвфемистическое» 'жир' → 'кровь' (можно даже говорить о возможной относительной хронологии этих изменений: поскольку чистая полисемия 'жир' / 'кровь' неизбежно приведет к серьезным коммуникативным неудобствам, логично предположить, что для того, чтобы стал возможным переход 'жир' → 'кровь', язык должен *сначала* обеспечить появление нового синонима в значении 'жир', на который ассоциация с 'кровью' не распространяется).

Таким образом, общее значение правила Э.5, сколь бы удобным оно ни было при реконструкции прасписки, не следует переоценивать; применять его следует только *после* внимательного

анализа семантики рассматриваемых этимонов и их коррелятов в близкородственных языках.

Перейдем теперь к рассмотрению наиболее сложной из возможных ситуаций: *пересеченно-конкурентной*. Вновь обратившись к дереву на стр. 151, отметим, что до сих пор речь шла только о случаях конкуренции между членами бинарных оппозиций, связанных ближайшим родством: например, о различных корнях, представляющих одно и то же значение в языках А и В, что затрудняет реконструкцию единого праэтимона для $*P_{AB}$; или о различных реконструкциях на уровнях $*P_{AB}$ и $*P_{CD}$, что осложняет дальнейшую реконструкцию на уровне $*P_1$.

Допустим, что $E_A \neq E_B$ и $E_C \neq E_D$, но $E_A = E_C$. Такая ситуация легко объяснима на основании простых дистрибуционных правил: старый этимон сохранил свою семантику в одном из двух языков в каждой ветви, а в других членах пары ее утратил. Но как быть, если при этом $E_B = E_D$?



На уровне групп такого рода «семантические пересечения» встречаются исключительно редко, но количество их неуклонно возрастает при переходе на более высокие уровни сопоставления, так что закрывать глаза на эту проблему недопустимо, и уже на первом этапе сравнения необходимо предложить для таких ситуаций как теоретическое объяснение, так и практический «рецепт». С чисто умозрительной точки зрения они могут быть вызваны следующими обстоятельствами:

(1) Неверно построенное дерево. Каждый обнаруженный случай «семантического пересечения» в 100- / 50-словном списке — хороший повод дополнительно перепроверить как корректность составления списков, так и первичную расстановку когнаций. Тем

не менее, бессмысленно утверждать, что правильно построенное дерево, т. е. такое, которое претендует на отображение реального исторического процесса, должно быть полностью свободно от «пересечений»: любому специалисту по сравнительно-исторической лексикологии хотя бы одной крупной языковой семье хорошо известно, что «идеальных» деревьев не бывает.

(2) Синонимия: E_{AC} и E_{BD} выражали требуемое значение в качестве свободно заменимых синонимов как на уровне общего для всех четырех языков праязыка $*P$, так и на уровне промежуточных праязыков $*P_{AB}$ и $*P_{CD}$; только в современных живых языках (причем во *всех* четырех!) произошло «урезание» синонимии — каждый из языков «отказался» от одного из синонимов.

С общетеоретической точки зрения такой сценарий кажется невероятным. Во-первых, он предполагает для всех праязыковых уровней эмпирически нереальную ситуацию полной синонимии в пределах базисной лексики (подчеркнем, что речь может идти только о синонимии абсолютной; «квази-синонимия» в рамках подхода, требующего максимальной точности сводешевских значений, все равно подразумевает существование на уровне $*P$ лишь одного «главного» синонима, а «транзитная синонимия» невозможна постольку, поскольку, если бы ситуация транзитной синонимии была представлена уже в $*P$, «старый» синоним по определению оказался бы лишенным своего исходного значения уже на уровнях $*P_1$ и $*P_2$, и все четыре языка A, B, C, D согласованно пользовались бы исключительно «новым» синонимом). Во-вторых, он крайне неэкономичен, т. к. требует постулировать четыре независимых семантических изменения во всех четырех языках.

Как это ни парадоксально, однако, на практике (например, при создании этимологических словарей) именно этот подход является превалирующим, что вызвано исключительно недооценкой роли семантических сопоставлений в праязыковой реконструкции. Если, например, в словаре Ю. Покорного приводится пять или шесть «праиндоевропейских» корней-входов, каждый из которых сопоставлен с одним и тем же значением, это ни в коем случае нельзя понимать как утверждение, что в пра-и.-е. *действительно* существовало пять семантически тождественных корней. В большинстве случаев это не более чем результат «синонимического» разрешения ситуации семантического пересечения —

которое на самом деле ничего не разрешает, а лишь уводит исследователя от необходимости потратить дополнительное время и силы на уточнение реконструкции.

Так, значение 'видеть' в словаре Покорного соотносится не менее чем с десятком различных корней. При этом анализ материала показывает, что корень **uel-* на самом деле имеет значение 'видеть' только в британских языках (валлийск. *gweled* и т. п.) и, возможно, в др.-английском *wli-tan* (хотя большой словарь [Bosworth & Toller 1898: 1259] все же дает значение 'to look, gaze', а не 'to see'); что корень **sek^w-* представлен значением 'видеть' только в германских языках и, возможно, в хеттском (*sakuḫai-*, где ситуация осложняется формальной производностью этой глагольной основы от *sakuḫa-* 'глаз') и т. п. Механистическая проекция всех этих пересечений на уровень праязыковой семантики создает иллюзию безудержной синонимии, что, в свою очередь, чревато далеко идущими выводами (например, о каких-то принципиальных отличиях языковых структур народов раннего и среднего неолита от современных и т. п.).

Исходя из высказанных соображений, объяснение (2) для ситуации семантического пересечения следовало бы отвергнуть как одновременно ничтожно маловероятное и эксплицитно вредное для дальнейшего развития сравнительно-исторического языкознания. Постулирование на основании семантических пересечений абсолютной праязыковой синонимии должно допускаться тогда и только тогда, когда хотя бы один из современных языков-потомков (A, B, C, D) дает для этого эксплицитные основания, т. е. сам демонстрирует такую синонимию. При этом стоит сразу указать, что на материале 50-словных списков по языковым семьям Евразии и Африки такая ситуация не была надежно зафиксирована ни для одного из проанализированных языков (со второй, менее стабильной, половиной списка положение более сложное, т. к. в него входит немалое количество понятий с «расплывчатой» семантикой).

(3) Не идентифицированные вовремя контактно-обусловленные (ареальные) изменения. Допустим, что языки B и D, хотя и относятся к разным ветвям группы/семьи, в ходе миграций оказались на смежной территории, и к старым «вертикальным» связям между собой добавили новые, «горизонтальные». В этом

случае их общий лексический элемент E_{BD} может являться либо непосредственным заимствованием (из B в D или наоборот), либо совместной ареальной семантической инновацией: некая праязыковая основа, не имевшая ни в $*P_{AB}$, ни в $*P_{CD}$ стословного значения, сначала получила его в B , после чего и в D со временем также начала употребляться в стословном значении под влиянием тесных контактов носителей обеих наречий.

(4) Наконец, последняя и, на наш взгляд, чрезвычайно недооцененная в существующих работах по сравнительно-историческому языкознанию возможность — *независимые однонаправленные семантические инновации*, т. е. ситуации, когда этимон E с праязыковым значением S изменяет это значение на S_1 в двух (или более) ветвях одной группы или семьи вследствие работы тех или иных «импульсов», не связанных взаимным влиянием, а возникающих независимо.

Чисто психологический фактор воздействия семантического сходства или тождества, как правило, обуславливает то, что возможность таких инноваций игнорируется в пользу трактовки их как «изоглосс», напр., «германско-анатолийских», «кельтско-германских» (см. примеры выше), «албанско-балтийских» и т. д. и т. п.; каждая такая изоглосса интуитивно воспринимается как неслучайная, а это, в свою очередь, становится важнейшим источником недоверия к «древесной» модели классификации, в рамках которой неслучайные изоглоссы такого рода необъяснимы. Отсюда, в частности, берут начало и неоправданная популярность «теории волн», преувеличение роли диалектных континуумов, размывание понятия праязыка и т. д.

Между тем наличие и даже неизбежность независимых однонаправленных семантических инноваций очень легко обосновать, если принять во внимание тривиальное эмпирическое наблюдение: типологический инвентарь надежно зафиксированных или убедительно реконструированных семантических изменений *чрезвычайно скуден* по сравнению с теоретическими пределами возможностей таких изменений. Грубо говоря, подавляющее большинство лексических элементов языка (по крайней мере, базисных) при условной «необходимости» изменить свое значение чаще всего подчиняется типовым семантическим корреляциям, число которых можно навскидку оценить как не превышающее

одного десятка для каждого отдельного значения (обычно даже меньше). Разумеется, в каждом конкретном случае возможны и «экзотические», «нетривиальные» сценарии развития, но в общемировом масштабе доля таких сценариев ничтожна и статистически пренебрежима.

Предположим, что для слова 'глаз' тремя наиболее вероятными источниками происхождения являются (а) глагольная основа 'видеть'; (б) глагольная основа 'сверкать, сиять'; (в) именная основа 'дырка, отверстие'; все остальные сценарии развития (включая русский и др.) будем считать типологически маловероятными. В этом случае вероятность *независимого* развития значения 'глаз' из значения (а), (б) или (в) в разных языках одной и той же группы/семьи будет находиться в прямой зависимости от количества языков в этой семье: если для одного языка вероятность семантического перехода 'видеть' → 'глаз' составляет, допустим, 0.1 (точный расчет такой вероятности не входит в непосредственные задачи нашего исследования, но должен, разумеется, учитывать индекс стабильности слова 'глаз' и количество типовых семантических корреляций), то для двух языков вероятность *независимого* семантического перехода 'видеть' → 'глаз' хотя бы в одном из них будет составлять уже 0.19 и т. п.

Эта чисто математическая корреляция наглядно объясняет впечатление, создающееся при тщательном рассмотрении этимологических словарей разных семей: чем глубже уровень сравнения, тем больше «синонимичных» семантических дефиниций выносятся во вхождения словарных статей. Грубейшей ошибкой было бы считать все эти дефиниции результатом семантической реконструкции: в лучшем случае это результат интуитивной попытки «на глазок» вывести для множества значений, представленных в языках-потомках, наиболее естественный, с субъективной точки зрения конкретного исследователя, общий семантический инвариант. При этом возможность независимой однонаправленной инновации при такого рода прикидке чаще всего не рассматривается вообще.

В рамках нашего исследования, исключающего «тотальную синонимию» в праязыке как явление, теоретически невозможное (при условии, что реконструируемые праязыки в структурном отношении не имели принципиальных отличий от современных),

любое семантическое пересечение будет по умолчанию трактоваться как вариант либо независимой (случай 4), либо контактно-обусловленной (случай 3) однонаправленной семантической инновации (принципиальные различия между этими двумя ситуациями отсутствуют).

Что касается конкретной методики *выбора* между двумя или более вариантами (т. е., возвращаясь к исходной схеме, ответа на вопрос, чему равен E_p : E_{AC} или E_{BD} ?), то она, по большому счету, ничем не отличается от методики выбора в непересеченно-конкурентных ситуациях, т. е. учитываться должны все те же самые экстра-дистрибуционные факторы (полисемия, морфологическая производность, взаимная этимологизация, возможность заимствованного происхождения и т. п.).

В качестве примера (правда, в хронологических рамках семьи, а не группы, но это не имеет принципиальной важности) возьмем ситуацию с праиндоевропейским эквивалентом значения 'зуб'. Для него традиционно, в т. ч. и в словаре Покорного, приводятся две основы: **dont-* ~ **dnt-* и **ĝombho-*, причем обе находятся в отношении семантического пересечения даже вне зависимости от того, какой конкретно классификационной схемы придерживаться в отношении праиндоевропейского, ср.:

а) **d(o)nt-*: др.-инд. *dan* ~ *danta-*, иранск. **dantan-*, греч. *ὀδών*, балтийск. **danti-* (литовск. *dantìs*, прусск. *dantis*), германск. **tanþu-*, латинск. *dēn-s*, кельтск. **danto-*;

б) **ĝombho-*: тохарск. **keme*, (?) др.-инд. *jambha-*, албанск. *dhëmb*, славянск. **zqbъ*, балтийск. **žamb-* (латышск. *zùobs*).

Очевиднее всего факт пересечения ощущается на материале балто-славянских языков; дистрибуционные правила подсказывают для прабалтийского вариант **danti-* (изоглосса между литовским и прусским, представляющим кардинально различные ветви), но латышский при этом имеет изоглоссу не с остальными балтийскими, а со славянскими.

Проще всего было бы «списать» эту ситуацию на праязыковую синонимию, но нет ни одного живого индоевропейского языка, в котором такая синонимия имела бы непосредственное продолжение. Вряд ли случаен тот факт, что отмечена она только для др.-индийского, где 'зуб' — обычное словарное толкование как для слова *dan*, так и для слова *jambha-*; это объясняется тем, что

существующие контексты не дают возможности установить тонкие семантические различия. При этом, однако: (а) как средневековые пракрыты, так и современные новоиндийские языки в общем значении 'зуб' продолжают только основу *dan(ta)-*; (б) *только* для основы *jambha-*, помимо общего значения 'зуб', для ряда контекстов установлены также более узкие значения 'глазной зуб', 'клык' (ср. также производную основу *jambhya-* 'резец'); (в) *только* основа *jambha-* может формально анализироваться как именное производное от глагольного корня *jabh-* 'огрызаться; лязгать (челюстями)' [Turner 1966: 282, 283, 352]. Все это недвусмысленно говорит о том, что и в др.-индийских памятниках соответствующая синонимия, скорее всего, не являлась абсолютной, а была ограничена лишь определенными смысловыми контекстами, в которых противопоставление между *dan(ta)-* и *jambha-* могло нейтрализовываться без ущерба для общего смысла текста.

Если абсолютная синонимия как вариант разрешения конкурентной ситуации неприемлема, остается лишь констатировать, что один из двух анализируемых корней должен был вытеснить другой в нескольких ветвях индоевропейской семьи независимо друг от друга. Согласно хотя бы критерию экономности, инновацией следовало бы считать **ǵombho-*; к счастью, еще более убедительным аргументом является наличие полисемии — **d(o)nt-* во всех индоевропейских языках имеет основное значение 'зуб', несводимое к другой семантике¹, в то время как **ǵombh(o)-* регулярно встречается либо как глагольная основа, либо в общем значении 'острый предмет' (ср.: греческ. ὑόμφος 'колышек';

¹ Согласно довольно расхожему представлению, и.е. **d(o)nt-* можно проэтимологизировать как старую форму активного причастия от глагола **ed-* 'есть' (что особенно удобно с ларингалистических позиций, т. к. реконструкция **Hed-* → **Hd-(o)nt-* якобы позволяет объяснить вокалическую протезу в греческом ὀ-δών). Но, во-первых, эту внутреннюю этимологию нельзя считать убедительной, так как деривация 'есть' → 'зуб' ('едящий', 'едок') не имеет типологических параллелей (значение 'зуб' обычно связывается с семантикой 'кусать', 'жевать', 'грызть'); во-вторых, даже при ее принятии она все равно остается *внутренней* этимологией, т. е. спекуляцией на тему развития от «до-праиндоевропейского» к праи.е. состоянию. На синхронном уровне ни в одном древнем и.е. языке продуктивная связь между лексемами 'зуб' и 'есть' не просматривается.

прагерманск. **kamba-z* 'гребень'; литовск. *žaĩbas* 'острый предмет', 'грань', 'мыс').

Учитывая семантику глагольного корня в др.-индийском, естественно было бы предложить первичность глагольной основы **ḡembh-* 'огрызаться, кусать (резко, с силой)', откуда далее произведено имя **ḡombho-s* '*кусалка' → 'клык / резец', перен. 'острый предмет', противопоставленное основе **d(o)nt-* 'зуб' (любой, как родовое понятие). При такого рода соотношении значений на пра-и.-е. уровне нет ничего удивительного в том, что **ḡombho-* (возможно, в рамках определенной «жаргонизации» лексического состава) вытеснило старую основу **d(o)nt-* в трех ветвях семьи независимо друг от друга. Отметим, что если, например, для албанского еще можно думать о семантическом влиянии со стороны славянского, то для тохарского вариант контактно-обусловленного изменения исключен.

В ряде работ по сравнительно-историческому языкознанию встречается концепция противопоставления «центра» и «периферии» как особо значимого для изучения динамики языковых изменений; согласно этой концепции, на периферии ареала, занимаемого языковой семьей, следует ожидать большей степени сохранности архаизмов, т. к. для инновации, исходящей из «центра», естественно ожидать постепенное проникновение в близлежащие диалекты/языки, но при этом вполне вероятно, что до дальних, «периферийных» областей распространения она не дойдет. С этой точки зрения можно было бы поспорить с реконструированной выше ситуацией: **ḡombh(o)-*, как «редкая» основа, представленная скорее на «периферии» индоевропейского мира (тохарский, албанский, балто-славянский) противопоставлена центробежной инновации **d(o)nt-* и должна рассматриваться как глубокий архаизм.

К сожалению, хотя такая модель развития и выглядит достаточно элегантно, универсализация ее на практике вредна ничуть не менее, чем некритическое принятие гипотезы о районе максимального разнообразия языков одной семьи как наиболее вероятной прародины этой семьи (еще один компаративистский «миф», основанный на неправомерном обобщении отдельных наблюдаемых фактов). Во-первых, «центробежные инновации» возможны только на самом раннем этапе существования языковой

семьи, когда между «свежеразделившимися» потомками праязыка еще существует тесное общение с элементами взаимопонимания; во-вторых, прозрачной внутренней этимологии естественно ожидать именно для таких «центробежных инноваций» (т. к. они зарождаются уже после распада праязыка на диалектные зоны / языковые ветви), а не для периферийных «архаизмов», в нашем же случае внутреннюю этимологию заведомо легче предложить для **ḡombho-*, чем для **d(o)nt-*.

Отдельным, и довольно существенным, методологическим вопросом является возможность *лексикостатистической коррекции* индексов когнации, расставленных на предыдущем этапе, по результатам исследования семантических пересечений. Согласно представленной выше схеме, E_A и E_C должны иметь один индекс когнации (например, '1'), E_B и E_D — другой (например, '2'). Допустим, что в ходе этимологической интерпретации материала удастся аргументированно обосновать, что E_A и E_C — праязыковые «ретенции», т. е. отражают исходное 100-словное значение, а E_B и E_D — независимые однонаправленные инновации. Следует ли в этом случае заново присвоить E_B и E_D *различные* индексы когнации ('2' и '3') и в свете этого перестроить матрицу и дерево, или оставить все как есть?

Однозначный ответ на этот вопрос, удовлетворяющий всем подтипам ситуаций, дать трудно. По сути, зависит он в первую очередь от того, что на самом деле понимается под индексом когнации. Если правило гласит, что «одинаковые индексы когнации присваиваются словам (корневым морфемам), имеющим общее этимологическое происхождение», менять ничего не следует. Но, как показывает наличие однонаправленных инноваций, «иметь общее этимологическое происхождение» и «являться общими рефлексами праязыкового эквивалента одного из элементов 100-словного списка» — не одно и то же.

Так, тохарское *klots* и ирландское *cluas* 'ухо' имеют общее этимологическое происхождение, являясь именными новообразованиями от и.-е. глагольной основы **kley-* 'слышать'; но ни то, ни другое не отражают базисный пра-и.-е. эквивалент 'уха', именную основу **ous-*. Генеалогическое дерево, построенное на основании первичного анализа, в ходе которого предвзятые представления о классификации не учитываются, не дает никаких оснований для

выделения тохарских и кельтских языков в единый узел, так что речь о совместной инновации в «пратохаро-кельтском» идти не может (более того, в др.-ирландском еще хорошо сохраняется старая основа *ai* ← **ous-*). В этой ситуации естественно считать, что лексические замещения в тохарских и в кельтских языках имели место независимо друг от друга. Правомерно ли будет, приняв этот сценарий в качестве рабочей гипотезы, оставить их в списке под общим номером?

С «сущностной» точки зрения, безусловно, нет. Единые индексы когнации должны отмечать общие архаизмы или совместные, т. е. восходящие к промежуточному праязыку, инновации; в противном случае матрица получает неоправданный «подскок» процентных соотношений для языков / групп, не имеющих непосредственного общего предка. Поэтому реконструкция общего сценария лексико-семантического развития того или иного сводешевского значения от праязыка к его потомкам, в ходе которой удастся выявить независимые однонаправленные изменения, не только может, но и *должна* приводить к пересмотру исходных индексов когнации (в частности, тохарское и кельтское 'уши' должны получить разные индексы). Применение этого принципа к конкретному материалу языковых семей показывает, что относительная классификация при этом обычно не изменяется (что неудивительно), но глоттохронологические датировки несколько удревняются (что также предсказуемо, т. к. коррекции такого рода могут только снизить, но не увеличить процентные совпадения между сравниваемыми языками).

В теоретическом плане такой подход кажется едва ли не самоочевидным; на практике, однако, ситуации, аналогичные развитию и.-е. 'уха', встречаются не чаще (а, возможно, даже намного реже), чем ситуации, в которых предельно ясного выбора между этимонами, находящимися в отношении семантического пересечения, сделать нельзя. Даже в случае с и.-е. 'зубом', разобранным выше, не все остается понятным. Действительно ли, например, отталкиваясь от балтийских форм, восходящих к и.-е. первичному этимону **d(o)nt-*, общий индекс когнации между русским *зуб* и латышским *zuobs* следует исправить на два разных (объясняя, например, развитие в латышском влиянием «новой» семантики этого слова в соседних славянских языках)? Или же

здесь следует пойти на определенный компромисс с теорией волн и представлениями о диалектных континуумах, и согласиться с определением «балто-славянской» общности как одного из таких континуумов, где уже на самом раннем этапе развития могли иметь место как «общебалтийские», так и «латышско-славянские» семантические сдвиги?

На этот вопрос также нельзя ответить однозначно, по крайней мере, на текущем этапе наших представлений о конкретных механизмах исторических изменений внутри диалектных континуумов. В связи с этим не имеет смысла, по-видимому, радикально настаивать на «драконовских» пересмотрах индексов когнации по итогам проверки материала на наличие однонаправленных инноваций — по крайней мере, в пределах одной языковой группы или даже одной семьи, все члены которой находятся в географически сопредельных ареалах. Тем не менее, определенный контроль все же необходим, и здесь требуется каждый конкретный случай семантического пересечения рассматривать в общем контексте. Если языки А и С входят в разные подгруппы одной большой группы, отношения между ними могут носить следующий характер:

(1) А и С не имеют географических точек соприкосновения, не являются членами единого диалектного континуума и, по-видимому, никогда ими не являлись, о чем свидетельствует почти полное отсутствие между ними общих инновативных изоглосс;

(2) А и С имеют общие точки соприкосновения, между носителями имеют место «соседские» отношения, общие инновативные изоглоссы присутствуют, но в количествах, сильно уступающих инновативным изоглоссам между ними и их «настоящими» родственниками (т. е. на дереве уверенно выделяются группы АВ и CD, но не выделяется группа АС);

(3) А является «доминантным» языком в своем ареале и оказывает сильное влияние на С, в результате чего в С регулярно проникают лексические инновации из А.

Установить, в каком из этих трех типов отношений находятся анализируемые языки, иногда бывает непросто, однако это скорее технический вопрос, ответ на который зависит от общего количества данных по тому или иному региону. Пересмотр индексов когнации обязателен только в том случае, если для А и

С естественно предполагать ситуацию (1) — это и есть, например, «тохаро-кельтская» ситуация, в которой более вероятно независимое развитие 'слышать' → 'ухо', чем совместная инновация в пределах «тохаро-кельтского континуума» (что подтверждается еще и тем, что в пределах стословного списка больше не отмечено ни одной «тохаро-кельтской» инновативной изоглоссы).

В ситуации (2) можно, действительно, говорить о некотором «волновом» типе изменения, *особенно* в том случае, если инновация представляет собой семантический сдвиг лексемы, восстанавливаемой на уровне праязыка. Можно утверждать, например, что латышское *zùobs* 'зуб' — результат семантического сдвига из родового значения 'острие, заостренный объект', имевшего место *под влиянием* соседних славянских языков, которые, в свою очередь, унаследовали этот сдвиг от праславянского. В определенном смысле это можно было бы считать и заимствованием, но *zùobs* регулярно соответствует литовскому *žait̃bas*, и нет оснований считать, что латышский язык сначала утратил рефлекс общебалтийской основы, после чего «восстановил» его, но уже не из балтийского, а из славянского фонда. Таким образом, даже несмотря на то, что ближайший родственник латышского — литовский, а не славянские языки, и эта ситуация надежно фиксируется на любом из возможных вариантов лексикостатистического дерева, постулирование для русского *зуб* и латышского *zùobs* единого когнационного индекса не лишено смысла: оно отражает возможность совместных ареальных инноваций, имеющих место между относительно близкородственными языками¹.

¹ В ситуации, когда степень родства чрезвычайно высока (т. е. начало языковой дивергенции имело место совсем недавно), такие совместные ареальные инновации могут даже привести к переходу языка из одной языковой группы (или, скорее, *подгруппы*) в другую: такого рода процесс описан в работах А. В. Дыбо для ряда тюркских языков. Подчеркнем, однако, что такие переходы возможны исключительно на хронологическом «микроуровне»: можно предполагать, что если между языками А и В пролегает временная граница хотя бы в 500-600 лет более или менее независимого развития, дивергентный процесс становится необратимым, и если в *такой* ситуации все же оказывается, что «язык А» в какой-то момент перешел в языковую группу «языка В», то речь, скорее всего, идет не о смене языком А своей генетической характеристики, а об

Наконец, в ситуации (3) методологически верным подходом было бы отмечать все случаи семантических пересечений между А и С, противоречащих «естественной» таксономии этих языков, как заимствования из А в С. Для этого либо необходимо на основании анализа общей дистрибуции показать, что соответствующие слова имели инновативный характер в А, после чего были переданы в С; либо, если языки хорошо изучены в историческом плане, показать, что фонетические соответствия между А и С не укладываются в общую схему соответствий, предложенных для группы в целом, и должны объясняться как отражающие конвергентные, а не дивергентные процессы. Если ни того, ни другого сделать не удастся, индексы когнации лучше не пересматривать, т. к. гипотеза о заимствованиях в пределах базисной лексики между А и С не получила достаточного подтверждения.

Подводя итоги несколько затянувшегося раздела 7.4, следует остановиться еще на одном важнейшем моменте. Реконструкция прасписки — сложная задача, предполагающая выбор *оптимального* варианта на основании разветвленной системы правил «фильтров», с подробной и ясной аргументацией каждого случая. Практика показывает, что такой вариант, как правило, можно обосновать: даже там, где, на первый взгляд, аргументы между двумя «кандидатами» распределены в соотношении 1 : 1, всегда можно найти какие-то дополнительные соображения, пусть даже сугубо косвенные (как-то: «странности» в фонологическом или фонотактическом облике слова; «ареально-бродячий» характер этимона; абсолютное количество языков, сохраняющих этимон), которые могут сместить предпочтения в пользу одного из них.

Однако вне зависимости от того, насколько доказательны дистрибуционные или экстра-дистрибуционные аргументы, *оптимальность* выбранного нами варианта ни в коем случае не должна приравниваться к *элиминации* всех прочих. Любая реконструкция по сути является аппроксимацией: даже про праславянскую реконструкцию **golva*, рефлексы которой представлены во всех

элементарном отказе носителей от языка А и переходе их на один из диалектов языка В (в котором, правда, в качестве «субстратного» слоя могут удержаться и какие-то элементы, некогда специфические для языка А).

языках-потомках без исключения, нельзя с полной уверенностью утверждать, что именно это слово на праславянском уровне было базисным в значении 'голова' — теоретически допустим и сценарий, при котором **golva* становится основной 'головой' уже на уровне отдельных праязыков для ряда раннеславянских диалектных групп, в каких-то случаях — как однонаправленная независимая семантическая инновация, в каких-то — через ареальное влияние. Можно лишь говорить о том, что «вещественные» аргументы в пользу единой праславянской формы **golva* 'голова' значительно превосходят (в данном случае — на все 100%) все остальные; это разрешает нам оценивать вероятность альтернативного, чисто «умозрительного» сценария как пренебрежимую, и оперировать реконструкцией **golva* как элементом, почти столь же «реальным», как и его рефлексy в языках-потомках.

В конкурентных ситуациях аргументы распределяются иным образом, и практика показывает, что допустить историческую ошибку при их анализе относительно несложно. Следовательно, даже выбрав оптимальный вариант, необходимо сохранять возможность «переигрывания» ситуации в тех случаях, когда внешние (сравнение на более высоком таксономическом уровне) или внутренние (привлечение к сравнению новых, доселе недоступных, языковых материалов) данные вступают в неразрешимое противоречие с выводами, сделанными ранее.

Иными словами, при необходимости сделать «праязыковой выбор» между этимонами $E_1, E_2 \dots E_n$, речь на самом деле идет не столько о том, чтобы «выбрать» один элемент множества и навсегда забыть об остальных, сколько о том, чтобы упорядочить все элементы по степени аргументированности их возможной архаичности в «сводешевском» значении. «Оптимальный вариант» должен постулироваться в качестве рабочей гипотезы, и оставаться таковой вплоть до того момента, когда окажется, что новые данные, способные в той или иной степени «поколебать» его праязыковой статус, принципиально необнаружимы (что мыслимо в теории, но на практике, по-видимому, недостижимо до тех пор, пока не будет создана единая, всеобъемлющая и абсолютно непротиворечивая классификация всех без исключения языков и диалектов мира).

Так, возвращаясь к уже упомянутому выше примеру трудности однозначного выбора прагерманского эквивалента 'мяса' из готского *timz*, западногерманского **flaiskaz* и скандинавского **kjot*, можно заметить, что последовательное применение к этому случаю перечисленных правил логически приводит к выбору **flaiskaz* как «оптимального» кандидата — на том основании, что из всех трех основ только **flaiskaz* убедительно реконструируется на общегерманском уровне (к западногерманской праформе со значением 'мясо' ср. далее др.-исландск. *flesk* 'свинина, бекон' [Orel 2003: 104; Vries 1962: 130]), в то время как *timz* и **kjot* вообще не имеют никаких параллелей в германских языках других ветвей; семантическое сужение 'мясо' → 'свинина' при этом ничуть не менее естественно, чем, наоборот, расширение 'свинина' → 'мясо'.

Если на этом основании будет принято решение рассматривать **flaiskaz* и только **flaiskaz* в роли прагерманского базисного этимона 'мясо', мы неизбежно допустим как лексикостатистическую, так и просто историческую ошибку, поскольку на следующем этапе реконструкции станет очевидным, что основной индоевропейский этимон для 'мяса' — **mēns-*, имеющий «законное» продолжение в готском *timz*; как общий принцип экономности, так и эмпирически наблюдаемый принцип необратимости семантических изменений (по крайней мере, на относительно кратких временных отрезках) говорят о том, что именно готский этимон должен продолжать прагерманский, который, в свою очередь, продолжает индоевропейский (неслучайно В. Орел с полной уверенностью реконструирует прагерманское **metzan* 'мясо' [Orel 2003: 267] на основании одной-единственной готской формы — исключительно потому, что она представляет собой регулярное отражение индоевропейской основы).

Таким образом, и на этом этапе исследования, как на этапе первичных статистических подсчетов, обязательно должна быть заложена возможность коррекции праязыковой реконструкции по итогам дальнейших этапов анализа: «кандидаты» должны быть упорядочены, но не элиминированы полностью. Общие результаты этапа реконструкции прасписка 1-го уровня должны выглядеть следующим образом:

1. Для каждого из анализируемых элементов списка Сводеша: «ранжированный» список «кандидатов на праэтимон», располо-

женных в порядке от наиболее к наименее оптимальному. «Список» не обязательно должен пониматься в строго формализованном смысле слова (т. е., например, как заключенное в фигурные скобки упорядоченное множество), т. к. в рамках нашего исследования не предполагается автоматическая обработка такого рода множеств вероятностными алгоритмами; гораздо более важной задачей видится детализированная *экспликация* упорядочивания, с максимально подробным изложением дистрибуционных и экстра-дистрибуционных аргументов.

2. 100- или 50-словный (в рамках нашего исследования — 50-словный) список для реконструируемого праязыка, формально представляющий собой поле в составе базы данных, в которую в дальнейшем будут добавляться аналогичные списки для других праязыков. Каждая запись такого поля (за отдельными исключениями, которые будут специально оговорены ниже) соответствует *одному* «сводешевскому значению» и содержит *один* праязыковой этимон, отбираемый по следующим правилам:

а) в случаях, когда «оптимальность» кандидата может быть установлена *дистрибуционным* образом, позиция в списке заполняется этим и только этим кандидатом;

б) в случаях, когда «оптимальность» кандидата устанавливается *экстра-дистрибуционным* образом, позиция в списке заполняется «оптимальным» кандидатом с обязательным занесением остальных в специальный комментарий к базе;

в) в случаях, когда единый «оптимальный» кандидат не устанавливается (как дистрибуционные, так и экстра-дистрибуционные правила оказываются неприменимыми или нерезультативными), позиция в списке заполняется *произвольно отобранным* кандидатом, с занесением остальных в комментарий.

Последний пункт может вызвать определенные нарекания, но ниже мы постараемся показать, что произвольность выбора в случаях типа (в) в общем случае не будет оказывать существенного влияния на построение классификации на более высоких уровнях; к тому же, оставляя за собой возможность дальнейшей коррекции праязыковых списков, мы сохраняем и возможность верификации нашего выбора на следующих этапах сравнения.

Праязыковые этимоны, полученные в ситуациях а) и б), мы будем называть *мотивированными*; полученные в ситуации в) —

немотивированными или *фиктивными*. Немотивированные реконструкции особенно характерны для «нестабильных» элементов списка, когда, например, каждый язык или каждая подгруппа в составе рассматриваемого таксона передает соответствующее значение через собственный этимон. В этом случае относительно высока вероятность того, что реальный праэтимон был на самом деле утрачен во *всех* языках-потомках (будучи, например, замещен различными другими транзитными синонимами уже на ранних этапах ветвления дерева), т. е. может оказаться и так, что наш «немотивированный этимон» вообще не имеет отношения к праязыковому состоянию, вне зависимости от того, каким из «кандидатов» оказалась произвольно заполнена соответствующая позиция. Ниже (в разделе 1.7.6) будет отдельно обсуждаться вопрос о том, как трактовать такую ситуацию с формально-лексико-статистической точки зрения.

1.7.5. *Предварительный анализ прасписков 1-го уровня.* После того, как для каждого таксона 1-го уровня, обследованного в пределах заданного ареала, построен праязыковой список и все списки оказываются включены в состав единой базы данных, можно переходить к следующему этапу: построению предварительных классификационных гипотез относительно таксонов 2-го уровня — или лексикостатистической верификации таковых, если они были уже предложены ранее в рамках сравнительно-исторической работы над языками рассматриваемого ареала или в рамках «массового сравнения» этих языков.

Здесь впервые имеет смысл прибегнуть к разделению анализа на две части — *ручную* и *автоматическую*, с возможностью взаимной коррекции. Дело в том, что, работая с языками Африки, при переходе от таксонов уровня «групп» (\approx 2-3 тысячи лет дивергенции) к уровню «семей» (от 4 до 6-7 тысяч лет дивергенции) мы, как правило, оказываемся на территории, либо вообще не затронутой применением сравнительно-исторического метода, либо, в лучшем случае, прошедшей лишь «начальный этап» его применения, на уровне скорее общих прикидок и установления самых очевидных соответствий, чем систематической реконструкции, тем более ступенчатой (даже для столь, казалось бы, подробно исследованной семьи, как банту, ступенчатые реконструкции по большей части отсутствуют). Следовательно, для того, чтобы

сформулировать на основании унифицированных критериев, пусть даже на предварительном уровне, оптимальный сценарий исторического соотношения выделенных нами таксонов 1-го уровня, опираться можно только на фонетические изоморфизмы. Только построив сначала своего рода «псевдо-дерево»¹, отталкиваясь исключительно от звукового сходства, можно в дальнейшем перейти к более строгой проверке отдельных узлов-«семей» этого дерева на предмет исторической реальности, применяя элементы этимологического анализа и проверяя их на наличие фонетической совместимости между группами, составляющими «семьи».

Важнейшим различием между отношением фонетического изоморфизма и фонетической совместимости является то, что автоматически установить на материале 100- (и тем более 50-) словного списка можно только первое из них, но никак не второе. Постулировать наличие или отсутствие фонетического изоморфизма между морфемой X_{L_1} языка L_1 и морфемой X_{L_2} языка L_2 можно без привлечения дополнительных данных: так, хинди *kha:* 'есть' изоморфно маори *kai* (см. раздел 3) вне зависимости от конкретных причин такого изоморфизма (в данном случае — историческая случайность). Ни один подобного рода изоморфизм сам по себе не может считаться аргументом в пользу генетического родства, но совокупность таких изоморфизмов между L_1 и L_2 , статистически превосходящая совокупность изоморфизмов между L_1 и любым другим из языков, привлеченных к сравнению, вполне может стать основанием для первичной гипотезы о том, что L_1 и L_2 восходят к общему предку.

В отличие от тестирования на «изоморфность», тестирование на «совместимость» в автоматическом режиме провести почти невозможно. Для того, чтобы определить морфемы X_{L_1} и X_{L_2} как «совместимые», особенно в том случае, если сегментный состав их не является идентичным, требуется подтверждающий материал,

¹ Точнее, множество «псевдо-деревьев» для «псевдо-семей»: на основании фонетических сходств между реконструкциями таксонов 1-го уровня можно легко построить единое дерево для всех таксонов (как это сделано, например, в рамках проекта Automated Similarity Judgement Program), но нас пока что будут интересовать только «псевдо-узлы», разделяющиеся на хронологическом уровне не глубже 4-5 тыс. до н. э. (согласно формуле С. А. Старостина).

т. е. дополнительные пары морфем $Y_{L1} : Y_{L2}$, $Z_{L1} : Z_{L2}$ и т. д., демонстрирующие аналогичные фонемные соотношения (например, хинди $kh : maori k$). Сам по себе поиск такого рода дополнительных пар можно автоматизировать (попытку построения простейшего алгоритма такого рода см. в [Starostin 2008b]), но на ограниченном материале 100-, а тем более 50-словного списка большинство из них, скорее всего, просто не будет обнаружено. Автоматизация может дать более надежные результаты, если списки удастся расширить, но, во-первых, это связано с колоссальным увеличением технической работы (сопоставительные словарные списки придется составлять уже не по 100, а по тысячам элементов — при том, что для многих языков столь подробные лексические данные просто отсутствуют), во-вторых, следует помнить о том, что на данном этапе уже не требуется полное совпадение значений сравниваемых слов, а, следовательно, алгоритм требуется «обучить» типологии семантических изменений — задаче, которой еще только предстоит быть решенной в отдаленном будущем.

Именно поэтому к автоматическому анализу данных мы будем прибегать только как к средству объективного построения первичных, сугубо *предварительных* гипотез относительно родства высокого уровня, а деревья, полученные с помощью такого анализа, будем называть «псевдо-деревьями», чтобы подчеркнуть даже не ненадежность, а *невозможность* их использования для дальнейших исторических разысканий до тех пор, пока соответствующие классификационные схемы не будут подтверждены или опровергнуты на следующем этапе.

Для того, чтобы получить «псевдо-дерево», мы, согласно уже оформившейся традиции, воспользуемся идеей т. н. «консонантных классов», введенной в обиход А. Б. Долгопольским [1964] и с тех пор неоднократно апробированной в различных исследованиях; ср., в частности, [Baxter & Manaster-Ramer 2000] и др. Согласно методу Долгопольского, весь консонантный инвентарь языков мира (точнее, Евразии, но принцип поддается обобщению и на общемировом уровне) разбивается на 10 непересекающихся классов; в основе разбиения лежит эмпирическое наблюдение того, что фонетические изменения консонантизма по языкам мира чаще связаны со сменой *способа* образования

согласного, чем *места*, причем под сменой способа образования подразумеваются в первую очередь колебания ларингальных признаков (глухость/звонкость и т. п.), но не противопоставление по шумности/сонорности.

Концепция «консонантных классов» была, в частности, применена в компьютерной лингвистической среде StarLing, где на ее основе действует функция Sound(), разработанная С. А. Старостиным. Эта функция для каждой неразрывной цепочки символов выдает ее «консонантный костяк», т. е. регулярно соответствующую исходной цепочку символов, в которой гласные устранены, а каждый согласный конвертируется в условное обозначение того класса, к которому он принадлежит. При этом 10 исходных «классов Долгопольского» расширены до 12:

- 1) класс P: губные взрывные и спиранты (*p, b, p^h, f, v* и др.);
- 2) класс M: губные носовые (*m* и др.);
- 3) класс W: губные глайды (*w*);
- 4) класс T: дентальные/альвеолярные взрывные и межзубные спиранты (*t, d, θ, ð*);
- 5) класс C: свистящие и шипящие аффрикаты (*c, z, č, ž* и др.);
- 6) класс S: свистящие и шипящие спиранты (*s, z, š, ž* и др.; классы 5 и 6 в системе Долгопольского не разделяются);
- 7) класс N: негубные носовые сонанты (сюда входят одновременно и дентальный *n*, и палатальный *ɲ*, и велярный *ŋ* и др.);
- 8) класс R: вибранты (*r* и близкие звуки);
- 9) класс J: среднеязычный глайд (*y*);
- 10) класс L: латеральные сонанты, аффрикаты и спиранты (т.е. весь диапазон звуков от *l* до *ʎ, Ł, ł*; в системе Долгопольского *l* объединено с *r*, а латеральные аффрикаты и спиранты не учитываются);
- 11) класс K: велярные и увулярные взрывные и спиранты (*k, g, q, ɢ, x, χ, ʁ* и др.);
- 12) класс H: ларингальные согласные и «нуль» в начальной позиции (*h, h, ʔ* и т. п.). При этом в неначальной позиции противопоставление между классом H и т. н. «слабыми» классами W и J снимается, что отражает универсальную тенденцию срединных и конечных глайдов к вокализации (т. е. такие формы, как *ta, taŋa, taŋa, taŋa* будут сводиться к единому «костяку» M).

Таким образом, например, отдельные элементы русского 50-словника будут обработаны функцией Sound() следующим образом: *пепел* = PPL, *птица* = PTC, *черный* = CRN, *огонь* = HKN, *имя* = HM и т. д.¹

Первоначально функция Sound() была разработана С. А. Старостинным для того, чтобы облегчить поиск фонетически изоморфных корней внутри двух и более баз данных по группам или семьям, связанным между собой либо гипотетическим, либо «доказанным» отношением генетического родства. Очевидно, однако, что с таким же успехом ее можно использовать и для автоматической простановки индексов когнации на материале 100- или 50-словных списков; этот алгоритм был технически реализован Ф. С. Крыловым. В исходном виде он выглядит очень просто:

а) пользователь задает максимальную длину «консонантного костяка»: по умолчанию сравнение идет по первым двум согласным (т. е. *черный* = CR, *огонь* = НК и т. д.), но в сопоставительных целях допустимо также сравнение по 1 или 3 согласным;

б) пользователю разрешено предварить автоматический анализ морфологической разметкой материала, при наличии которой «костяк» будет строиться только по корневым согласным (так, русск. *умирать* = HM, но при наличии разметки *у=мир-ать* = MR);

в) для каждого элемента списка формам с полностью совпадающим консонантным костяком присваивается общий индекс когнации; формам, отличающимся в одной и той же позиции хотя бы на один символ — разный. Таким образом, например, польск. *pies* 'собака' = чешск. *pes* id. («костяк» PS), но *pies* ≠ русск. *собака* («костяк» SP).

Практическое применение соответствующего алгоритма к материалу, например, индоевропейских или уральских языков показывает, что он в целом неплохо справляется с задачей построения классификации на «среднем» уровне, т. е. достаточно

¹ При необходимости количество классов можно увеличивать или уменьшать: тестировались, в частности, и такие варианты, которые не разграничивают классы C/S и R/L, и, наоборот, более детализированная система (с разграничением велярных и увулярных, билабиальных и лабиодентальных и т. п.). По-видимому, оптимальной является система из 10-12 классов, последовательно и унифицированно апробированная на всем материале.

аккуратно выделяет группы в пределах семьи; так, в составе индоевропейского дерева корректно определяются все без исключения основные «ветви», вплоть до подтверждения таких хронологически глубоких единиц, как балто-славянская или индоиранская. Более проблематичной оказывается классификация на мелких уровнях, где, как нетрудно предсказать, в первую очередь группируются друг с другом фонетически архаичные языки, противопоставленные языкам, претерпевшим значительные фонетические изменения; особенно пагубно этот фактор может сказаться на классификации, в которую одновременно включены древние и современные языки — так, внутри германской ветви современный английский язык оказывается первым звеном, отделяющимся от западногерманской ветви, задолго до древнеанглийского, а внутри индоиранского санскрит и авестийский в определенных условиях могут выделиться в единую ветвь, противопоставленную, с одной стороны, современным индийским, с другой — современным иранским языкам.

Помимо этого, от такого рода «автоматизированного» дерева предсказуемо следует ожидать опознания значительно меньшего числа истинных когнатов, чем аналогичного числа, определяемого на основании детального этимологического анализа; так, даже для пары «русский — польский» автоматизированный алгоритм выдает $\approx 60\%$ совпадений в пределах 100-словного списка, в то время как «ручная» обработка материала убедительно демонстрирует не менее 80% совпадений. При сопоставлении таксонов 1-го уровня это, как правило, не приводит к существенным искажениям относительной классификации; бессмысленно в этих условиях лишь применение глоттохронологических формул, катастрофически завышающих возраст таксонов. Однако уже для таксонов 2-го уровня (вне зависимости от того, представлены ли они в составе базы данных живыми языками или реконструкциями по таксонам 1-го уровня) заниженное число опознаваемых когнатов является серьезной проблемой во всех отношениях.

Неспособность алгоритма, ориентированного на консонантный анализ, опознать те или иные «истинные» когнаты в первую очередь связано с ложностью исходного посыла — «фонетические изменения происходят в пределах одного и того же консонантного класса». Во-первых, справедливость этого утверждения

находится в прямой зависимости от хронологического отрезка, разделяющего сравниваемые языки: если, например, фонетические переходы $k \rightarrow x$ и $t \rightarrow d$ алгоритмом учитываются, то «двухшаговые» переходы $k \rightarrow x \rightarrow \emptyset$ и $t \rightarrow d \rightarrow r$ останутся непонятыми. Во-вторых и в-главных, существует на самом деле значительное число фонетических переходов, хорошо знакомых специалистам по исторической типологии, в которых задействованы элементы разных фонологических классов (те же самые развития $x \rightarrow \emptyset$, $d \rightarrow r$ и многие другие). Упрощение автоматической процедуры анализа за счет игнорирования таких развитий можно объяснить стремлением найти простейшее из возможных объективных решений; непонятно, однако, почему на этом следует останавливаться и не предпринимать дальнейших попыток слегка усложнить алгоритм (на вполне рациональной и объективной основе), если такое усложнение почти наверняка гарантирует получение более точных результатов.

В работе [Starostin 2008b], развивающей идеи С. А. Старостина по автоматическому анализу синхронной и диахронической фонологии, частично реализованные им в компьютерной среде StarLing, был предложен алгоритм установления регулярных фонетических соответствий между 100-словными списками множества сравниваемых языков или праязыковых реконструкций (программная реализация алгоритма выполнена Ф. С. Крыловым) на основании в целом тривиального анализа частотностей. Для целей данного исследования, однако, этот алгоритм в целом непригоден, т. к. установить регулярные соответствия между реконструкциями праязыков таксонов 1-го уровня на материале 50-словного списка, не привлекая дополнительных лексических данных, в обычном случае невозможно; на автоматическом этапе обработки ориентироваться можно только на сам *факт* сходства, но никак не на его *регулярность*.

Представляется, что в нашем случае гораздо перспективнее было бы ограничиться методом обычного консонантного анализа, но с возможностью в случае необходимости дополнить идею замкнутых, не имеющих общих точек соприкосновения консонантных классов на представление о консонантных классах как *пересекающихся* подмножествах общего консонантного инвентаря, точки пересечения которых определяются неоспоримыми

данными типологии фонетических изменений («сомнительные», т. е. редкие, спорные, неизвестные для таксонов 1-го уровня и т. п. фонетические корреляции во внимание приниматься не будут).

Так, например, в консонантный класс К входят как велярные согласные k, g , так и увулярные $q, ɠ$. Для первых в языках мира постоянно и, в некоторых случаях, даже предсказуемо фиксируется развитие палатализации: $k \rightarrow \check{c}/c, g \rightarrow \check{ʒ}/ʒ$, т. е. в определенном смысле оправдано помещение этих согласных не только в класс К, но и в «аффрикатный» класс С. Напротив, для увулярных согласных такое развитие, по крайней мере, как «одношаговое», невозможно: артикуляторное расстояние между увулярной и палатальной областью столь велико, что в общем случае увулярные согласные могут палатализироваться только через велярную стадию, т. е. нормально развитие $q \rightarrow k \rightarrow \check{c}/c$, но не $*q \rightarrow \check{c}/c$. Аналогичным образом, если мы помещаем латеральный сонант l и латеральную аффрикату λ в один класс L, а вибрانت R — в другой, глупо было бы игнорировать прописную истину, согласно которой развитие $l \leftrightarrow r$ в языках мира частотно, а развитие $\lambda \leftrightarrow r$, наоборот, практически не встречается.

Таким образом, для более гибкого, но ничуть не менее объективного, функционирования алгоритма анализ по консонантным классам имеет смысл дополнить унифицированным списком возможных типовых фонетических переходов элементов одного класса в другой. При сопоставлении, например, фонетических структур $\{k-n\}$, $\{k-\eta\}$ и $\{k-k\}$, первые две из которых будут сводиться к костяку KN, а третья — к костяку KK, алгоритм должен учитывать также естественность соотношения $\eta : k$ (два заднеязычных согласных, противопоставленных по признаку наличия / отсутствия сонорной артикуляции).

В практическом плане учет «естественных» пересечений между консонантными классами, однако, не может быть трансформирован в элементарную простановку индексов когнации между всеми «пересекающимися» этимонами. Так, в разбираемом нами гипотетическом случае три консонантных костяка образуют три бинарные пары: (1) $[k-n : k-\eta]$, (2) $[k-n : k-k]$, (3) $[k-\eta : k-k]$, из которых пара (1) должна получить единый индекс когнации из-за совпадения по классам ($k-n = k-\eta = KN$), а пара (3) — из-за пересечения элементов η (класс N) и k (класс K). Пара (2) сама

по себе не может иметь один и тот же индекс когнации, т. к. элементы n и k относятся к непересекающимся подмножествам классов N и K соответственно, но встает вопрос о возможном применении к данной ситуации принципа транзитивности: если (а) $k-n = k-n$ и (б) $k-n = k-k$, не следует ли отсюда, что (в) $k-n = k-k$?

Резонность такого подхода очевидна в том случае, если мы подразумеваем, что автоматическая лексикостатистика по принципу фонетического сходства претендует на построение более или менее непротиворечивой исторической картины. Действительно, с одной стороны, нет ничего невероятного в том, чтобы, например, исходная лексема типа **kaŋa* сохранилась без изменений в одном языке, деназализировалась в **kaka* в другом и «упереднила» артикуляцию носового согласного в третьем → **kana*. С другой стороны, сама процедура присвоения индексов когнации подразумевает транзитивность: если $A = B$ и $B = C$, то в стандартной лексикостатистической процедуре неизбежно $A = C$.

На самом деле, однако, обобщение принципа транзитивности может привести к сугубо неудовлетворительным результатам — особенно в том случае, если алгоритм применяется к большому массиву языков или реконструкций, не находящихся друг с другом в отношении сверхблизкого родства. Добавим к нашему инвентарю пересечений еще два «естественных» (типологически частотных) фонетических перехода: деназализация $n \rightarrow d$ и палатализация $k \rightarrow c/\check{c}$. Теперь допустим, что анализируемый нами этимон, помимо структур $k-n$, $k-n$, $k-k$, еще в двух языках представлен консонантными структурами $\check{c}-n$ и $\check{c}-d$ соответственно. Принцип транзитивности будет в этом случае предполагать, что все пять структур должны получить одинаковый индекс когнации — в том числе и такие фонетически далекие структуры, как $k-k$ и $\check{c}-d$, через гипотетические (но, до некоторой степени, вероятные!) сценарии развития (а) $*k-n \rightarrow k-k$, (б) $*k-n \rightarrow *k-n \rightarrow *c/\check{c}-n \rightarrow \check{c}-d$. В конечном итоге это может привести к тому, что автоматически расставленные индексы когнации вообще потеряют какой-либо смысл: при включении в подсчеты достаточно большого числа языков «транзитивные цепочки» можно будет расставить между подавляющим большинством этимонов. Помимо этого, встанет также вопрос о том, не следует ли алгоритму «реконструировать» недостающие звенья цепочки — правомерно ли присваивать

единый индекс когнации консонантным структурам $k-k$ и $\check{c}-d$ при наличии «промежуточных звеньев» $k-n$ и $k-\eta$ и не присваивать его при их (возможно, случайном) отсутствии? Учитывая, что пересекающиеся элементы классов в конечном итоге обеспечивают возможность перехода любого класса в любой другой, обобщение принципа транзитивности логически приводит к утрате алгоритмом значимости, т. к. при неограниченном количестве «шагов» любая фонетическая структура рано или поздно может перейти в любую другую.

Теперь видно, что, на первый взгляд, логичная идея учета возможности естественных переходов фонем из одного класса в другой не принималась во внимание в процедурах, разработанных А. Б. Долгопольским, С. А. Старостиним, У. Бэкстером и др. не столько по недосмотру, сколько из-за серьезных методологических проблем, связанных с воплощением ее на практике. Тем не менее, на наш взгляд, следует все же не отказываться от этой идеи совсем, а лишь ограничить ее определенными фильтрами. Итоговый алгоритм должен выглядеть так:

1) на первом этапе индексы когнации проставляются автоматически с опорой исключительно на совпадение по консонантным классам, т. е. в соответствии с «классической» процедурой. Лексикостатистическая матрица и классификационная схема, полученные на этом этапе, не игнорируются, а рассматриваются в качестве «исходной версии»;

2) на втором этапе алгоритм для каждого из анализируемых этимонов, помимо уже определенного на первом этапе «главного» костяка, строит множество «вторичных» костяков, опираясь на информацию о возможных межклассовых переходах представленных в них согласных. При этом, во избежание нагромождения случайных совпадений, вторичные костяки должны быть ограничены не более чем *одним* межклассовым переходом в каждом из них. Таким образом, вместо одного исходного костяка $[C_1C_2]$ мы получаем для одного стословного элемента языка L_1 множество $\{C_{1.1}C_{2.1}\} \dots \{C_{1.n}C_{2.1}\} \dots \{C_{1.1}C_{2.m}\}$, где n — количество консонантных классов, совместимых с исходным через допустимые «межклассовые» переходы, для 1-го согласного, а m — для 2-го.

Пример: при заданности межклассовых фонемных переходов $[k \leftrightarrow \check{c}]$, $[k \leftrightarrow \eta]$ структуре **kaŋa* на этапе 1 соответствует только

костяк KN, на этапе 2 — множество костяков KN, CN, NN, KK. Однако даже на этапе 2 ей *не* соответствует костяк СК, подразумевающий одновременно два межклассовых перехода;

3) на последнем этапе анализа алгоритм сопоставляет *каждый* костяк из получившегося множества с *базовым* костяком анализируемого стословного элемента во всех остальных языках. При обнаружении хотя бы одного совпадения анализируемый стословный элемент языка L_1 *дублируется* в списке (формально это будет выглядеть как «синонимия»), и соответствующий «дубль» снабжается отдельным индексом когнации.

Пример: допустим, что некий стословный элемент в языке (или праязыковой реконструкции) L_1 имеет вид *kaŋa* (базовый костяк KN), в языке L_2 — *sana* (базовый костяк CN), в языке L_3 — *čada* (базовый костяк СТ). Добавим к этому правила допустимых межклассовых переходов: (1) $k \leftrightarrow c$, $k \leftrightarrow \check{c}$ (палатализация), (2) $n \leftrightarrow d$ (/де/назализация). Тогда:

— на этапе 1 *kaŋa* получает индекс когнации 1, *sana* — индекс когнации 2, *čada* — индекс когнации 3 (все базовые костяки различны);

— на этапе 2.1 *kaŋa* получает вторичный костяк CN, *sana* получает вторичные костяки KN, СТ, *čada* получает вторичные костяки КТ, CN;

— на этапе 2.2 (а) *kaŋa* получает продублированную общую когнацию с *sana* (через вторичный костяк CN, совпадающий с базовым костяком в *sana*), но не получает общей когнации с *čada* (т. к. не имеет вторичного костяка, общего с базовым костяком *čada*); (б) *sana* получает продублированную общую когнацию с *čada* (через вторичный костяк СТ, совпадающий с базовым костяком в *čada*).

Для большей наглядности укажем, что на первом («стандартном») этапе анализа соответствующее вхождение в лексикостатистической базе данных будет иметь следующий вид:

Слово	Язык 1	Инд. 1	Язык 2	Инд. 2	Язык 3	Инд. 3
слово X	<i>kaŋa</i>	1	<i>sana</i>	2	<i>čada</i>	3

На втором этапе эта единая запись будет преобразована в три «квази-синонимических» записи:

Слово	Язык 1	Инд. 1	Язык 2	Инд. 2	Язык 3	Инд. 3
слово X	<i>kaŋa</i>	1	<i>saŋa</i>	2	<i>čada</i>	3
слово X	<i>kaŋa</i>	4	<i>saŋa</i>	4		
слово X			<i>saŋa</i>	5	<i>čada</i>	5

Подчеркнем, что для каждого случая «успешного» сопоставления по межклассовым переходам слова должны дублироваться с *отдельными* индексами когнации (т. е. *kaŋa*, как представляющее костяк [KN], должно иметь индекс когнации 1, но *kaŋa*, как представляющее видоизмененный костяк [CN], должно иметь индекс 4, и соответствующее ему *saŋa* также должно быть продублировано с индексом 4). Только в этом случае нам удастся избежать «ловушки транзитивности»: получить одно когнационное соответствие между языками 1 и 2, другое между языками 2 и 3, но не получить при этом соответствия между языками 1 и 3.

Оба этапа алгоритма имеет смысл применять и анализировать в плане результатов *последовательно*, т. к. значимыми для дальнейшего исследования могут быть выводы, полученные и на том, и на другом из них. Этап 1 — сопоставление только по «базовым» консонантным костякам, без учета возможности межклассовых переходов — дает более грубые, общие результаты, но при этом отражает заведомо реалистичную картину (существует хотя бы минимальная теоретическая вероятность того, что алгоритм выявит истинные этимологические связи). Этап 2, учитывающий межклассовые переходы, «отловит» заведомо большее количество объективных фонетических сходств, но при этом будет а priori «анти-историчным», т. к. отказ от принципа транзитивности будет приводить к присвоению одной и той же морфеме нескольких разных индексов когнации, что лишает нас возможности исторической интерпретации лексико-статистических результатов.

К статистическим матрицам и классификационным схемам, полученным на обоих этапах, имеет смысл добавить матрицу / схему, полученную в результате *ручной* обработки материала (или ручной корректировки результатов автоматического анализа этапа 1). Несмотря на то, что здесь мы вступаем в более субъективную область, существует целый ряд ситуаций, в которых

описанный выше алгоритм все же оказывается беспомощным, а опытный специалист, напротив, будет в состоянии провести грамотный, логически и фактологически обоснованный анализ материала. Это, например:

1) Ситуации, в которых та или иная реконструкция для таксона 1-го уровня однозначно опознаваема как заимствование. Это особенно актуально в тех случаях, когда таксон представлен языком-изолятом или пучком близкородственных диалектов, подверженных очевидному ареальному влиянию (например, албанский или брахуи в Евразии; эфиосемитские или северно-сонгайские языки в Африке и т. п.). В этих случаях регулярный компаративный анализ, как правило, приводит к неоспоримым выводам, в то время как автоматический алгоритм, описанный выше, не будет в состоянии самостоятельно вычленить заимствования (для этого придется по меньшей мере обучить его на следующем этапе распознавать «лексикостатистические аномалии»).

2) Ситуации, когда фонетический изоморфизм налицо, но сравниваемые слова различаются структурой корня: например, один таксон представлен преимущественно структурами CVC(V), другой — только CV. Здесь, в отличие от предыдущего случая, алгоритм теоретически мог бы быть «обучаемым», но в целом ситуации такого рода встречаются редко, а диапазон допустимых типов соотношений, наоборот, довольно велик: между одной парой таксонов возможна корреляция $C_1VC_2V : C_1V$ (с отпадением второго слога), между другой — $C_1VC_2V : C_2V$ (с отпадением первого), в третьей исходная структура может давать C_1C_2V с дальнейшими нетривиальными упрощениями кластера, в четвертой корреляции между двусложной и односложной структурами могут быть ограничены лишь определенным типом основ и т. п. Предусматривать все эти ситуации на автоматическом, сутобо предварительном этапе анализа — чересчур трудоемкая задача; здесь экономнее будет осуществлять тщательный, подробный аргументированный ручной контроль.

3) В некоторых случаях лингвист, осуществляющий ручной контроль, обладает определенными эксклюзивными сведениями относительно фонологической системы того или иного таксона, недоступными автоматическому алгоритму, но позволяющими уточнить расстановку приблизительных индексов когнации еще

до этапа подключения этимологического анализа. Так, например, если обнаружено, что на праязыковом уровне в начальной позиции отсутствует фонема **n-*, но присутствует **d-*, в который эта фонема когда-то перешла, то автоматический алгоритм на втором этапе способен «отловить» такие параллели, как, например, *dVrV : nVrV, dVkV : nVkV* и т. п. (из-за наличия правила «межклассового перехода» $d \leftrightarrow n$), но не сумеет опознать пары *dVrV : nVdV, dVkV : nVηV* и др., связанные уже двумя межклассовыми переходами. В обычной ситуации это скорее преимущество алгоритма, не допускающего чересчур большого фонетического расхождения, но в той конкретной ситуации, когда **d-* в праязыке X является однозначным и единственно возможным эквивалентом **n-* в праязыке Y, это скорее недостаток, препятствующий учету информации, которая может иметь ценность для принятия дальнейших этимологических решений.

Таким образом, основная (и единственная) задача ручной верификации результатов автоматической обработки данных — учет неоспоримых индивидуальных особенностей реконструкций таксонов 1-го уровня. В целом *нежелательна* ситуация, при которой классификационная схема, полученная на автоматическом этапе, сколь-либо существенно отличается от полученной в результате ручной верификации, т. к. это может привести к подозрениям в субъективности и даже предвзятости исследователя. С другой стороны, следует помнить, что любая классификация, основанная исключительно на фонетических изоморфизмах, вне зависимости от того, получена ли она строго алгоритмическим образом или же с вмешательством «человеческого фактора», в любом случае будет являться лишь очередным промежуточным этапом исследования. Вполне допустимо, что на следующем этапе (анализ на предмет фонетической совместимости) она окажется опровергнута.

Как при автоматической, так и при «ручной» расстановке когнаций полученное генеалогическое древо неизменно будет *единым*: даже если при сопоставлении двух праязыковых реконструкций 1-го уровня процент совпадений будет минимальным (например, 1 слово из 50 = 2%), в чисто теоретическом плане это значение все равно может быть конвертировано как в относительную степень близости (длину соответствующего узла в рамках

общего древа), так и в абсолютную дату глоттохронологического распада (так, 2% схождения примерно соответствуют периоду дивергенции в 18,000 лет).

Разумеется, такого рода результаты никакой значимости иметь не будут (кроме негативной): столь низкие проценты находятся в зоне статистической погрешности. В нашу задачу на данном этапе, однако, не входит формальное математическое обоснование процентно-хронологического порога, разграничивающего классификационные узлы, которые заслуживают доверия, от заведомо бессмысленных узлов. Такой порог задан изначально в определении «семьи» как таксона, состоящего из нескольких групп, так, что языки из разных групп имеют между собой от 15% до 40% (для *п्राязыков* этих групп желателен слегка более высокий процент — от 20% до 50%, т. к. количество общей лексики при переходе от родственных живых языков семьи к родственным праязыкам неизменно должно возрастать).

Следовательно, главный результат описываемого этапа — не та единая классификационная схема, которая появляется в ходе предварительной лексикостатистической обработки, а только отдельные ее части — «*семейные узлы*». Именно они будут служить главным объектом анализа на следующих ступенях исследования.

В качестве примера можно использовать, например, результаты подсчетов, проведенных на материале праязыковых реконструкций «мелких» групп индоевропейской и уральской семей, уточненных и согласованных в рамках заседаний Ностратического семинара в Центре компаративистики РГГУ. (Данный эксперимент, правда, нельзя назвать чистым, т. к. индексы когнаций составлялись с учетом известной этимологической информации, но для текущих иллюстрационных целей это непринципиально). Полученная матрица показывает между отдельными индоевропейскими группами проценты совпадений от 30% и выше, между отдельными уральскими — примерно тот же результат, т. е. подтверждает не только реальность обоих таксонов, но и их статус «2-го уровня», т. е. «семьи». Напротив, потенциальные когнаты между отдельными ветвями индоевропейского и уральского дают принципиально более низкий результат — в диапазоне 13-16%.

Это говорит о том, что, вне зависимости от интерпретации этих результатов (случайные или неслучайные, контактные или

генетические), «индоуральский» таксон в любом случае не является «семьей», и, хотя полученные результаты и можно взять на заметку, вернуться к «индоуральскому» сопоставлению следует уже на новом этапе, после тщательной этимологической реконструкции 50-словного списка для праиндоевропейского и прауральского языков соответственно. Такой же принцип должен применяться и к языковым группам/семьям любого другого лингвистического ареала.

1.7.6. *«Глубинные» этапы анализа.* Все последующие стадии, вплоть до терминальной, уже не заслуживают столь детального методологического комментария, как предыдущие, т. к., по сути, являются рекурсивными: получив гипотетический список таксонов 2-го уровня, мы должны обрабатывать лексический материал этих таксонов (представленный праязыковыми реконструкциями на уровне групп) в полном соответствии с методами и правилами, описанными в разделах 7.3 и 7.4, т. е. тестировать полученные «псевдо-когнаты» на предмет фонетической совместимости (или фонетических соответствий), подключая там, где это необходимо, дополнительный этимологический материал.

В ходе этого тестирования можно будет подтвердить или, наоборот, опровергнуть корректность полученной на предыдущем этапе классификации (на уровне семьи). Если окажется, что классификация была все же установлена неверно (что, на самом деле, при требуемых процентах довольно маловероятно), необходим «откат» к предыдущему уровню, с ручной коррекцией тех индексов когнации, ложность которых была выявлена тестированием на совместимость. В противном случае для семьи, на основании все тех же дистрибуционных и экстра-дистрибуционных правил, вырабатывается собственный оптимальный прасписок.

На этом этапе особенно остро встает вопрос о статусе немотивированных праязыковых этимонов. Рассмотрим его подробнее на гипотетическом (но вполне реалистичном) примере.

Допустим, что для таксона 1-го уровня А реконструирован прасписок из 30 мотивированных и 20 немотивированных этимонов; для таксона 1-го уровня В — аналогичный список с таким же соотношением (разумеется, мотивированные и немотивированные этимоны А и В не будут однозначно коррелировать друг с другом). На этапе ручной/автоматической первичной обработки

удалось получить «сигнал родства», после чего более тщательная этимологическая обработка данных привела к следующему результату:

Таксон А	Таксон В	
	30 м.	20 н.
30 м.	+10 / -10	+5 / -5
20 н.	+6 / -7	+3 / -4

Самой «прочной» (точнее, самой убедительной в интуитивном плане) частью аргументации здесь будут 10 когнатов между «мотивированными» секциями прасписков, однако не менее важным для реконструкции и классификации будет наличие дополнительных 14 когнатов, включающих «немотивированные» этимоны — этот факт одновременно (а) подтверждает родство между А и В (сама суть глоттохронологического метода требует, чтобы между родственными праязыками устанавливались когнации и в «мотивированной», и в «немотивированной» частях прасписков) и (б) позволяет снять «немотивированность» реконструкций 1-го уровня, т. е. верифицировать их, через внешнее сравнение. Но означает ли это, что в 50-словных списках языков пра-А и пра-В на момент, непосредственно предшествовавший их трансформации в «группу А» и «группу В», действительно было 24 общих слова, т. е. 50% соответствий? Вряд ли; скорее всего, таких совпадений было больше.

Выше уже отмечалось, что понятие «немотивированный праязыковой этимон» не эквивалентно определению «любой произвольно взятый элемент из этимонов $X_1, X_2 \dots X_n$, представленных в ветвях/языках таксона X », т. к. помимо этих элементов имеется также существенная вероятность того, что в пра- X на этом месте находился праэтимон $*Y$, этимологически не тождественный ни одному из этимонов $X_1, X_2 \dots X_n$, т. е. замещенный во всех языках-потомках. Вероятность эта тем выше, чем менее «ветвистой» является внутренняя классификация X^1 .

¹ Классические примеры таких ситуаций — бинарное членение, например, в индо-хеттской (индоевропейский + анатолийские) или уральской (финно-угорский + самодийский) семьях.

Из приведенной выше таблицы можно с высокой степенью уверенности вывести лишь следующее: (а) 24 «плюса» отражают 24 праэтимона 2-го уровня, т. е. 24 элемента списка Сводеша для языка «пра-АВ»¹; (б) в 10 других случаях мотивированность этимологически нетождественных реконструкций пра-А и пра-В позволяет утверждать, что либо одна, либо другая, либо обе ветви имели здесь инновацию.

Однако во *всех* тех случаях, где эксплицитная когнация между А и В отсутствует, но либо А, либо В, либо обе группы при этом представлены немотивированными реконструкциями (т. е. $5 + 7 + 4 = 16$ позиций), поручиться за реальное отсутствие когнации на уровне «свежеразделившихся» праязыков А и В невозможно. «Немотивированность», по сути, почти эквивалентна «неизвестности»: в любой из 16 позиций на праязыковом уровне могла на самом деле иметь место когнация, не оставившая следов.

Таким образом, в данной ситуации 24 — это не максимальное, а *минимальное опознаваемое* число когнатов между А и В. На самом деле в 50-словных списках этих праязыков могло иметь место до 40 совпадений. Неучет этого обстоятельства может в отдельных случаях иметь негативные последствия для относительной лексикостатистической классификации, но в первую очередь самым негативным образом он будет отражаться на глотто-хронологической датировке: хронологическая разница между 0.48% и 0.80% совпадений колоссальна.

Возможным выходом из ситуации было бы математическое усреднение результата, однако, на наш взгляд, разумнее было бы в каждом случае такого рода просто проводить два финальных подсчета: (а) *включающие* данные немотивированных реконструкций и (б) *исключающие* эти данные (т. е. немотивированные реконструкции в первом случае отмечаются положительными индексами, во втором — отрицательными). Сопоставление результатов обоих подсчетов, во-первых, сможет показать степень прочности

¹ Впрочем, в отдельных случаях это могут быть и однонаправленные независимые семантические инновации; прояснить этот вопрос поможет только «массовость» сравнения (с привлечением других групп, входящих в тот же таксон 2-го уровня, и анализом дистрибутивных особенностей когнатов) или же выход на реконструкции таксонов 3-го уровня.

полученной относительной классификации, во-вторых, предоставит конкретный интервал, очерчивающий возможные хронологические грани распада анализируемого таксона (что в целом даже предпочтительнее строгого самоограничения одной-единственной глоттохронологической датой).

Гипертрофированной версией такого подхода было бы *полное* исключение немотивированных реконструкций из сопоставления — на том основании, что, поскольку серьезных внутренних аргументов для выбора праформы 1-го уровня в этих случаях нет, недопустимо «подгонять» устраивающий нас вариант под внешние данные (т. е., например, «пользоваться» готским *timz* 'мясо' только потому, что оно этимологически коррелирует с когнатами в других индоевропейских ветвях, нельзя) во избежание порочного круга. Такие меры предосторожности, по-видимому, все же чрезмерны, т. к. заставляют без необходимости отказываться от конструирования вполне реалистичных исторических сценариев. Несомненен тот факт, что степень доказательности классификации во многом зависит от пропорционального соотношения использованных в ней мотивированных и немотивированных реконструкций; конкретная практика, однако, показывает, что практически не бывает ситуаций, в которых следы исходного родства сохранялись бы исключительно в «ненадежных» этимонах и, наоборот, бесследно исчезали бы из «надежных». (Если такая ситуация обнаруживается, это, как правило, приводит к переинтерпретации обнаруженных сходств как контактных).

Следует отметить, что максимально детализированный подход требовал бы реконструкции собственного прасписки на *каждом* из узлов получаемого дерева, т. е., например, при переходе от отдельных уральских групп к прауральскому имело бы смысл посвятить отдельный раздел 50-словнику для угорского узла (обско-угорский + венгерский), северо-западного узла (прибалтийско-финский + саамский), общефинно-угорского узла и т. п. В той или иной степени оперировать такими уровнями в любом случае придется при анализе дистрибуции, семантических пересечений и т. п.; на практике, однако, допустимо (точно так же, как это допускалось на уровне групп, многие из которых также состоят из нескольких узлов) обсуждать праязыковую лексику одновременно для всех ветвей одной семьи, при условии

постоянного учета ее внутренней структуры для решения спорных вопросов.

Таким образом, в целом наше исследование предусматривает два этапа применения рекурсивного принципа — на уровне «семьи» и, после реконструкции базисной лексики для праязыков всех выделяемых семей, на уровне «макросемьи», т. е. сверхглубоких объединений, которые, в случае Африки, могут совпасть или не совпасть с макросемьями, постулируемыми в классификациях Дж. Гринберга и других африканистов. Задача анализа материала на еще более глубоком уровне («макро-макро-семьи», «мегафилы» и т. п.) в рамках нашего исследования не ставится вообще; во-первых, такой анализ, вне всякого сомнения, должен будет выходить уже далеко за пределы африканского континента, во-вторых, есть серьезные основания сомневаться в том, что он вообще возможен в рамках стандартной лексикостатистики; возможно, здесь могла бы как-то помочь «корневая» глоттохронология (этимостатистика), но для того, чтобы последняя корректно работала на временных глубинах в 15 и более тысяч лет, необходимы принципиальные усовершенствования этого метода, пока что никем не предложенные.

Подчеркнем еще раз, что использование в исследовании элементов реального сравнительно-исторического подхода, включающего поиск системных фонетических закономерностей, а также верификации результатов на основе типологии фонетических и семантических изменений, *не предполагает*, что неизбежным его результатом обязательно станет выявление нами на территории Африки небольшой группы макросемей, аналогичной гринберговской классификации или даже совпадающей с ней. К такому результату (причем, скорее всего, неверному) может привести полная его автоматизация (так, как это произошло, например, с «древом лексических сходств» между языками мира, полученными в ходе работы проекта ASJP); но этимологический контроль, которому обязаны подвергаться все выявленные «на глазок» или в ходе применения автоматических алгоритмов параллели, неизбежно устранил все случайные связи, оставив только такие, которые не будут взаимно противоречивыми или невероятными в рамках общего исторического сценария для языков африканского (или любого другого) ареала.

1.8. Специфика африканского ареала в плане установления генетического родства между его языками.

1.8.1. *Общие замечания.* Несмотря на то, что описание методологии нашего исследования вынужденно растянулось почти на двести страниц, оно все равно не может претендовать на статус исчерпывающего. Заранее предусмотреть *все* возможные технические и теоретические препятствия, возникающие на пути решения столь объемной задачи, как создание единой генетической классификации языков столь колоссального лингвогеографического ареала, как Африка, в принципе невозможно. Очертив общие контуры метода, мы при этом оставляем до некоторой степени размытыми те его положения, которые регламентируют определение наличия или отсутствия фонетической совместимости; возможности или невозможности семантической связи; определение того или иного таксона как «хронологически глубокой группы» или «хронологически неглубокой семьи» и т. п.

С нашей точки зрения, это оправдано: попытка полной и абсолютной формализации метода применительно к любым возможным обстоятельствам неизбежно завлечет исследователя в своего рода «ловушку теоретизирования», внутри которой, вместо того, чтобы применять на практике метод, успешно работающий на 90% материала, он будет вынужден тратить время и силы на нахождение одновременно объективных и теоретически обоснованных решений для 10% материала, в рамках которых предложенный метод работает плохо.

Наглядный пример тому — отказ ряда лингвистов от использования лексикостатистики на основании трудностей, связанных с выбором подходящего синонима. Конкретная практика показывает, что трудности эти в *подавляющем* большинстве случаев ограничены лишь небольшой частью списка, и *подавляющее* большинство их легко снимается при условии уточнения семантических определений элементов, входящих в эту часть. Опрямительным и даже в каком-то смысле заносчивым было бы утверждать, что можно полностью объективным, непротиворечивым и «сущностно обоснованным» образом прийти к *стопроцентной* точности в выборе синонима для любого произвольно взятого языка мира. Но для достижения поставленной цели — создания

единой, непротиворечивой и оптимальной из всех возможных генетической классификации — такой абсолютной, механистической точности и не требуется, тем более что результаты ее не претендуют на математическую доказательность.

Разбору вопросов, возникающих при применении описанной методики к конкретным группам языков Африки, будет посвящена значительная часть нашего исследования — настолько значительная, что при ознакомлении с этимологическими разборами 50-словных списков может возникнуть впечатление, что непредвиденных трудностей на самом деле гораздо больше, чем десять процентов, и что метод на самом деле работает так слабо, что на каждом шагу приходится применять то или иное решение *ad hoc*. Это на самом деле не так: просто для разбора трудных, неоднозначных ситуаций, естественным образом, приходится находить большее количество данных и аргументов, чем на разбор простых, где иногда достаточно просто привести лексический материал, выводы из которого самоочевидны.

Тем не менее, прежде чем перейти к рассмотрению базисной лексики конкретных языковых групп, имеет смысл попытаться дать самое общее представление о *типовых* проблемах, связанных с историко-лингвистическими исследованиями лексического материала языков Африки. Речь идет не об общей синхронно- и диахронически-типологической характеристике языков этого континента/ареала (все желающие могут без труда ознакомиться с соответствующими характеристиками в таких компетентных обзорных трудах, как [Heine & Nurse 2000]; [Childs 2003] и другие)¹, а исключительно о том, чтобы заранее указать те «специфически африканские» трудности, которые будут особенно часто выходить на поверхность в ходе нашего анализа.

1.8.2. «Стандартная модель» классификации языков Африки. Африканский континент заметно выделяется на фоне прочих

¹ Совсем недавно в свет вышла также обзорно-методологическая монография [Dimmendaal 2011], ценная тем, что основной упор в ней делается именно на сравнительно-историческое, а не типологическое изучение языков Африки; полноценно учесть этот труд с пользой для нашего собственного исследования уже не удалось, что не мешает рекомендовать его в качестве отдельного ценного введения в соответствующую проблематику.

лингвистических ареалов планеты не только за счет колоссального множества и типологического разнообразия представленных на нем языков, но и за счет того обстоятельства, что на сегодняшний день это *единственный* крупный ареал, в котором дискуссия об исторических связях между его языками регулярно ведется на хронологическом уровне *макросемей*. В то время как термины «ностратический», «сино-кавказский», «австрический» (для Евразии), «америндский» (для Америки), «индо-тихоокеанский» (для языков Новой Гвинеи и Австралии), хотя и обладают чрезвычайно различной степенью аргументированности, неизменно вызывают подчеркнуто скептическую реакцию среди большинства представителей современного лингвистического «мэйн-стрима», макроклассификация языков Африки, разработанная Дж. Гринбергом, до сих пор, несмотря на многочисленные оговорки, на практике является рабочей моделью для средне-статистического лингвиста-африканиста, а вслед за ним — как для специалистов широкого профиля, так и для популяризаторов¹.

«Стандартная модель» африканской классификации была построена Гринбергом в 1940-е — 1950-е гг.; отдельные ее положения сначала выходили в свет в виде последовательной серии статей (публиковались с 1949 по 1954 гг. в *Southwestern Journal of Anthropology*, впоследствии перепечатаны в едином издании как [Greenberg 1955]), затем были суммированы в итоговой монографии [Greenberg 1966] (первое издание — в 1963 г.). Как известно, Гринберг во многом опирался на результаты своих предшественников-классификаторов (К. Майнгоф, Д. Вестерманн, К. Доук, В. и Д. Блик и многие другие), но только ему принадлежит честь создания первой всеохватывающей классификационной модели, в рамках которой все (или почти все) языки Африки сводятся к минимальному исходному набору «макро»-таксонов.

Нет никаких сомнений в том, что популярности «стандартной модели» способствовало повышенное внимание Гринберга к методологической стороне ее построения, к тем конкретным критериям, на основании которых аргументируются те или иные

¹ Ср., например, перечисление основных африканских макросемей без каких-либо критических замечаний в одном из наиболее популярных описаний культурной истории Африки [Reader 1997: 109-110].

классификационные гипотезы. Три основных методологических постулата, лежащих в основе исследования Гринберга, формулируются им на первой же странице итоговой монографии [Greenberg 1966: 1]; по крайней мере два из них — (а) недопустимость использования чисто типологических аргументов при построении классификации и (б) нерелевантность экстралингвистической (антропологической, социально-культурной и т. п.) информации — хотя и могут показаться сегодняшнему специалисту едва ли не самоочевидными, на практике часто нарушались исследователями в «догринберговский» период (например, в рамках т. н. «нило-хамитской» теории К. Майнгофа, достаточно популярной среди африканистов первой половины XX в.). В определенном смысле эти нарушения были неизбежны, поскольку использование таких данных хоть как-то компенсировало громадные пробелы в сопоставительно-историческом изучении известных на тот момент африканских языков. Но по мере того, как пробелы эти начинали постепенно восполняться, становилось все более очевидным, что генетическая классификация языков континента нуждается в тщательной «фильтрации», каковая, в наиболее эксплицитном виде, и была проведена именно в работах Гринберга, что само по себе является выдающимся достижением.

Третий постулат — необходимость перехода от *изолированного* сравнения отдельных языков к *массовому* сравнению больших языковых множеств, что, по мнению Гринберга, неизбежно приводит к отсечению большинства случайных сходств — является в его методологии наиболее спорным и логически уязвимым. Как сильные, так и слабые стороны этого подхода уже обсуждались выше (раздел 1.5); здесь отметим лишь, что на момент публикации «Языков Африки» эффект от его применения был все же скорее позитивным, так как во многих случаях это позволило надежно отличить контактные связи между отдельными представителями крупных языковых групп/семей от их генетических связей.

Четыре «макро-таксона», предложенных Гринбергом для Африки, в порядке убывания количества входящих в них языков выглядят следующим образом: 1) *нигер-кордофанская* макросемья (название для этого таксона сконструировал лично Гринберг);

2) *афразийская*, или *афро-азиатская* макросемья (более старое, но постепенно выходящее из употребления название — *семитохамитская*; общий языковой состав определен задолго до работ Гринберга, хотя определенный вклад в уточнение классификации им все же был внесен); 3) *нило-сахарская* макросемья (название и общий состав таксона определены Гринбергом); 4) *койсанская* макросемья (название принадлежит Й. Шапера, но в лингвистический обиход было введено Гринбергом; он же впервые эксплицитно утвердил представление о генетической общности всех койсанских языков).

Несмотря на то, что «стандартная модель» действительно закрепилась в мировой африканистике намного прочнее, чем аналогичные модели, впоследствии разработанные Гринбергом для других ареалов планеты, нельзя сказать, что она была изначально принята «на ура» даже простым большинством специалистов. На самом деле ситуация здесь в некотором смысле парадоксальна: почти все профессионально-лингвистические отклики на классификационные труды Гринберга в лучшем случае выражали скепсис, в худшем — категорическое неприятие¹. Безоговорочную поддержку основных постулатов «стандартной модели» продемонстрировала, пожалуй, лишь очень небольшая группа африканистов, в целом тяготеющих к макрокомпаративным исследованиям (Л. Бендер, К. Эрет, Г. Флеминг); общая же тенденция к дискредитации результатов Гринберга со временем только увеличивается (так, за последнее десятилетие в африканистике имел место фактический «развал» койсанской макросемьи, хотя, справедливости ради, стоит отметить, что койсанский материал в трудах Гринберга был изначально очень ненадежен).

И, тем не менее, «стандартная модель» во многом остается в силе: именно она неизменно оказывается в центре любого общего

¹ Ср. следующие цитаты из рецензий на работу [Greenberg 1966]: «Most students of African languages have paid due attention to this work, few have accepted its conclusions» [Meeussen 1963: 170]; «...it would be possible to place much more confidence in this whole work if the author had shown an awareness of the inconclusive nature of a great part of the evidence» [Guthrie 1964: 135]; «The implications of Greenberg's study are certainly stimulating and deserve careful study, although we may doubt the validity of some of his genetic language families» [Westphal 1964: 1446].

изложения основ исторической африканистики. Отметим, в частности, что в чрезвычайно популярном на сегодняшний день социолингвистическом каталоге языков мира «Этнолог» [Ethnologue 2009] четыре макросемьи Гринберга подаются как действительные таксономические единицы — в отличие, например, от «америндской» макросемьи, которая в каталоге не упоминается вообще (вместо нее там фигурирует несколько десятков разрозненных мелких семей), не говоря уже об «индотихоокеанской» или о таких макрокомпаративистических гипотезах, как ностратическая или сино-кавказская, не связанных непосредственно с именем Гринберга (хотя в определенной степени поддержанных им в своих публикациях).

В задачи данного раздела не входит ни описание истории изучения перечисленных макросемей, ни описание их состава и основных исторических и типологических характеристик; и ту, и другую информацию уместнее, разбив на части, представлять как введение к соответствующим разделам, посвященным конкретному этимологическому и лексикостатистическому анализу данных по отдельным макросемьям. Общий перечень макросемей приведен здесь лишь постольку, поскольку любая попытка создать *новую* классификацию языков Африки, в том числе и предпринимаемая в рамках нашего исследования, должна отталкиваться от *старой*. Сколь бы спорной и методологически сомнительной ни была «стандартная модель» Гринберга, она подразумевает цельный, непротиворечивый и реалистичный исторический сценарий, в основе которого лежит сравнительный анализ языковых фактов, проведенный, вне всякого сомнения, одним из наиболее выдающихся лингвистов XX века.

Своеобразной данью классификации Гринберга в нашем исследовании будет выбор последовательности разделов, посвященных конкретным таксонам. При составлении и этимологическом анализе 50-словных списков таксонов 1-го уровня мы будем постепенно продвигаться от одной макросемьи к другой, начиная с койсанской и заканчивая афразийской; *внутри* соответствующих макросемей очевидные и неоспоримые таксоны 1-го уровня будут разбираться в том порядке, в котором они перечисляются в итоговой монографии Гринберга [Greenberg 1966] (разумеется, за вынужденным исключением тех таксонов,

которые в первоначальную классификацию войти не успели, из-за неизвестности или неизученности их составляющих на момент подведения итогов).

Что касается верификации макрогипотез Гринберга, то здесь мы выбираем своего рода компромиссный вариант. С одной стороны, как уже говорилось выше, при вводе в обиход новой методики исследования было бы некорректным апробировать ее на макротаксонах Гринберга, не будучи заранее убежденным в их истинности. Иными словами, для того, чтобы протестировать «на прочность», например, нило-сахарскую гипотезу, надежно установленные таксоны 1-го и 2-го уровней, составляющих нило-сахарский макротаксон, необходимо сопоставлять не только друг с другом, но и с максимально большим числом «нило-сахарских» таксонов — ведь совершенно не исключено, что какие-то компоненты «нило-сахарского» на самом деле окажутся более близкими каким-то компонентам «нигер-кордофанского» или даже «койсанского» и т. п. По сути, требуется вернуться к тому же, с чего в свое время начинал и сам Гринберг — «массовому сравнению», хотя и поставленному на принципиально иную основу.

С другой стороны, полный лексикостатистический анализ таксонов 1-го уровня всего африканского континента — задача, рассчитанная на много лет работы, и в этой связи имеет смысл время от времени давать хотя бы промежуточную оценку гринберговских макросемей по мере того, как входящие в них мелкие группы проходят «зачистку» по методике, описанной в предыдущем разделе. В частности, реконструировав на предварительном уровне прасписки для всех таксонов 1-го уровня, входящих в одну из макросемей, будет несложно как применить к ним процедуру автоматического установления «псевдо-когнаций», так и расставить соответствующие псевдо-когнации вручную (и то, и другое согласно параметрам, описанным в п. 1.7.5).

Вне зависимости от того, будут ли соответствующие результаты позитивными (т. е. подтверждающими предложенные ранее гипотезы относительно состава конкретных языковых семей и макросемей) или негативными (т. е. либо отражающими иную классификацию, либо основанными на минимальных процентных схождениях, лежащих в области статистической погрешности), перейти к более тщательному этимологическому анализу

лексикостатистических данных и верификации полученной древесной модели, по существу, все равно можно будет только по завершении первичной обработки всех данных по Африке. Тем не менее, промежуточные результаты, даже замкнутые в рамках одной конкретной (но сугубо гипотетической) макросемьи, могут уже на таком поверхностном уровне анализа показать хотя бы относительную степень надежности тех или иных макро-гипотез: если, например, общая предварительная глоттохронологическая оценка макросемьи А дает дату изначального расхождения в 8 тыс. лет, а макросемьи Б — в 18 тыс. лет, понятно, что первой гипотезе следует в дальнейшем уделять более пристальное внимание, чем второй.

Вышеперечисленные соображения в значительной степени мотивируют общую структуру первых разделов исследования. Каждый такой раздел посвящен лексикостатистической обработке одной из макросемей Дж. Гринберга, краткий обзор которой (история гипотезы, отношение со стороны ведущих специалистов, общий состав и основные историко-типологические характеристики помещаемых в нее таксонов) выносится в начало части. Вслед за этим по отдельности рассматривается каждый из таксонов 1-го уровня: состав, источники данных и обзор наиболее релевантной литературы, краткая историческая характеристика (включая оценку праязыковой реконструкции, если таковая кем-либо уже была предложена), лексикостатистическая матрица, полученная на основании подсчетов по полному 100-словному списку и, наконец, главный раздел: обстоятельный разбор 50-словных списков с реконструкцией (хотя бы приблизительной) общего прасписка.

В заключении к каждой части приводятся результаты первичной автоматической и прикидочной ручной обработки списков, сопровождаемые (там, где это возможно) выводами относительно степени достоверности предложенных ранее классификаций (на данном этапе — степень эта может значительно возрасти или, наоборот, снизиться в зависимости от результатов дальнейшего анализа). Эти результаты должны быть сведены вместе в единую схему в финальной части исследования, посвященной выделению таксонов 2-го и 3-го уровней, а, точнее, верификации автоматизированной и скорректированной вручную лексикостатистики

с помощью критериев фонетической совместимости и, где это возможно, строгого сравнительно-исторического метода.

1.8.3. *Основные препятствия на пути создания единой классификации языков Африки.* В задачи данного подраздела не входит конкретная критика методологии и практики построения Гринбергом «стандартной модели»; общие критические замечания в адрес «массового сравнения» уже были суммированы выше (см. раздел 1.5), а специфическая критика классификационных результатов, полученных Гринбергом, и так представлена большим количеством работ и не нуждается в дополнениях. Поскольку мы принимаем «стандартную модель» лишь в качестве технически удобной исходной точки, а в методику нашего исследования с самого начала заложены механизмы, способные как подтвердить, так и полностью перестроить «стандартную модель» в зависимости от результатов анализа данных, отдельный разбор аргументации Гринберга не имеет в данной ситуации ключевого значения; по ходу исследования читатель сможет без особого труда самостоятельно разобраться в том, какие из базисно-лексических сопоставлений Гринберга выдерживают «проверку на историчность», а от каких, наоборот, лучше отказаться из-за невозможности встроить их в конкретный сравнительно-исторический сценарий.

Само собой разумеется, тем не менее, что сопоставительное изучение языков Африки сталкивается с целым рядом проблем объективного характера, связанных не столько с методологией исследования, сколько с природой, количеством и качеством собственно языковых данных. Проблемы эти можно поделить на пять основных типов: (1) отсутствие достаточного количества языковых данных и низкое качество многих из имеющихся данных; (2) «переизбыток» общего числа языков и языковых семей, часто выливающийся в «методологические срывы» исследователя; (3) тесные языковые контакты между языками разных групп и семей, затрудняющие корректное установление генетических связей; (4) нестандартные особенности синхронных фонологических систем и их диахронического развития; (5) в очень многих случаях — неизбежность внутренней морфологической реконструкции и одновременно недопустимость переоценки ее возможностей.

Со всеми пятью типами можно столкнуться практически в любом лингвистическом ареале планеты; наша задача в данном

разделе — попытаться описать их специфику конкретно для африканского континента.

1.8.3.1. *Проблема недостатка информации.* По данным [Ethnologue 2009], на территории Африки в настоящее время локализовано чуть более двух тысяч языков; из них около полутора тысяч включается в нигер-кордофанскую макросемью, более трехсот — в афразийскую, более двухсот — в нило-сахарскую, и лишь около тридцати — в койсанскую¹. Из этой огромной массы *полностью удовлетворительными* языковыми описаниями (т. е. включающими детальную грамматику, подробный словарь по меньшей мере на две-три тысячи лексических единиц и хотя бы небольшой корпус записанных текстов) обладают в лучшем случае двести-триста языков (полная статистика пока что отсутствует). Это — те языки, по которым можно не только составить надежные стословные списки, но и заниматься продуктивным этимологическим анализом. Особый интерес к ним со стороны исследователей чаще всего объясняется большим количеством носителей и повышенной социально-культурной значимостью для того или иного региона (суахили, йоруба, фула, нама, хауса, канури и т. п.), но бывают и исключения, обычно связанные с повышенным интересом и высокой степенью самоотдачи (доходящей иногда в буквальном смысле до самопожертвования) отдельных исследователей по отношению к описываемым «мелким» языкам (например, работа П. Дикенса и Э. Трэйлла над койсанскими языками жу|хоан и !хонг, или «образцово-показательные» труды Дж. Хита по ряду диалектов сонгайской группы).

К сожалению, задача построения общеафриканской классификации исключительно на «близком к идеальному» материале принципиально нерешаема, и не только по той очевидной

¹ В целях аккуратности, конечно, стоит отметить также наличие на территории Африки языков других семей — австронезийской (мальгашский), индоевропейской (английский, французский, африкаанс и др.) и т.п., однако мы все же будем ограничиваться языками, «родными» для африканского континента, а не появившимися на нем в ходе относительно недавних иммиграционных процессов (что касается о-ва Мадагаскар, то с лингвогеографической точки зрения его вообще можно исключить из «африканского ареала» и считать частью «малайско-полинезийского» языкового пространства).

причине, что из нее придется исключить подавляющее большинство африканских языков, но и потому, что для надежной генетической привязки того или иного языка необходим принцип «массовости» сравнения (см. раздел 1.5), а опора исключительно на «полностью удовлетворительные» описания скорее сопоставима с принципами типологических исследований (строго ограниченная выборка из ряда «опорных точек»), не годящимися для целей исторического языкознания. Привлечение «неидеальных» источников значительно расширяет кругозор, позволяя в разы увеличить число учитываемых языков — с двух-трех сотен по меньшей мере до полутора тысяч, т. к. на начало XXI в. африканских языков, по которым африканистам неизвестно ничего, кроме названия, осталось очень немного.

«Неидеальные» источники можно расклассифицировать по двум параметрам: (А) общий тип работы и (Б) качество данных. По параметру (А) выделяются следующие категории:

А.1. *Словарный список*. В отличие от подробного словаря, словарные списки строго ограничены в количественном отношении (чаще всего содержат от 50 до 500 лексем) и, как правило, не содержат ни парадигматической, ни синтагматической информации. Скудность данных объясняется либо техническими причинами (недостаток времени для работы с информантом; публикация в виде журнальной статьи, ограничивающей объем публикации), либо характером исследования: чаще всего словарные списки появляются в рамках языковых «обследований» (surveys) отдельных лингвогеографических ареалов. В последнем случае список, как правило, составлен по заранее заданному лексическому инвентарю из определенного количества элементов (в редких случаях используется стандартный список Сводеша, но чаще всего каждый исследователь или исследующая организация имеет собственный инвентарь).

Традиция составления и заполнения обзорных лексических списков по языкам Африки восходит еще к работам XIX века; классическим образцом такого обзора является труд Сигизмунда Кёлле «Polyglotta Africana» [Koelle 1854], в котором на основе единой унифицированной системы транскрипции были сведены вместе трехсотсловные списки по ста с лишним языкам Западной Африки. Из более современных обзорных работ можно в первую

очередь отметить коллективные труды [WALDS I, II] по Западной Африке и [BCCW] по огромной семье бенуэ-конго. В отдельных случаях обзорные исследования целого ареала могут проводиться и на индивидуальной основе, ср. классический труд Л. Бендера по сопоставительной лексикостатистике языков Эфиопии [Bender 1971] или 400-словные списки Ч. Крафта по чадским языкам [Kraft 1981]. Особенно активно сбором подобного рода списков в последние два-три десятилетия занимаются сотрудники Летнего института лингвистики, отчеты которых (к сожалению, не всегда содержащие сами данные) регулярно появляются на веб-сайте института (<http://sil.org>).

Словарные списки, особенно те, которые пересекаются со 100-словным списком Сводеша хотя бы на 60-70% — второй по значимости источник лексической информации после собственно словарей, а если подходить к вопросу с чисто количественной стороны — первый, т. к. бо́льшая часть языков Африки на сегодняшний день представлена именно словарными списками, а не каким-либо другим типом данных. Однако типичным и практически неизбежным недостатком словарного списка является его «поверхностность»: поскольку заполняется он чаще всего поспешно, без выяснения деталей, для каждой отдельной позиции высока вероятность ошибки из-за возможности недопонимания со стороны информанта или, наоборот, исследователя. Нередки случаи, когда два списка для одного и того же языка, составленные двумя разными специалистами, расходятся по целому ряду позиций — не потому, что специалисты работали с разными диалектами, а потому, что, например, для того или иного слова европейского языка не был строго очерчен круг значений, и в итоге для него были выбраны два разных квази-синонима.

По этой причине словарные списки имеют наибольшую ценность в таких ситуациях, когда они составлены одновременно в большом количестве для целой группы близкородственных языков или диалектов; это часто приводит к тому, что праязыковая реконструкция того или иного этимона для соответствующей группы оказывается более надежной, чем любая произвольно взятая форма (о чем уже говорилось выше, в п. 1.6.3.2, на примере списков Крэбба по экойдным языкам).

А.2. *Грамматика*. Поскольку интерес исследователей часто фокусируется исключительно на языковой структуре, данные по многим языкам Африки доступны лишь в виде грамматического описания. К некоторым из них могут прилагаться краткие словарные списки (как, например, [Reh 1985] для кронго, [Jakobi & Crass 2004] для загава и другие), но, как правило, лексический материал приходится извлекать непосредственно из иллюстрирующего материала (например, для языков фур по [Jakobi 1990], балесе по [Vorbichler 1965] и др.).

Если грамматическое описание оказывается достаточно подробным, извлечь из него необходимый лексикостатистический минимум, как правило, удается; бывает даже, что грамматика оказывается полезнее словаря, т. к. слово извлекается непосредственно из синтаксического контекста, подтверждающего его значение. Для отдельных элементов списка Сводеша (личные, указательные, вопросительные местоимения; отрицание 'не') хорошее грамматическое описание также имеет бóльшую значимость, чем словарь. К сожалению, более типичным в этой области все же является жанр краткого очерка, чем подробного описания, так что в большинстве случаев 100-словный список, составленный по грамматике, имеет не менее 30-40% лакун.

А.3. *Работы по отдельным аспектам*. Во многих случаях полевые исследования проводятся не столько с целью детально задокументировать структурные характеристики и лексический инвентарь языка, сколько с целью анализа и оценки этих характеристик в рамках той или иной модели или теории, составляющих основной предмет интересов исследователя. В этих ситуациях сами по себе языковые данные, как правило, не публикуются; вместо этого опубликованной оказывается статья или подборка статей, посвященных соответствующим аспектам. Типичные примеры — «койсанский» язык западный \ddot{x} хоан, большинство данных по которому до сих пор извлекается из морфосинтаксических работ Дж. Грубера и К. Коллинза, или многочисленные статьи Т. Андерсена по целому ряду западнонилотских языков.

Если подборка статей оказывается достаточно большой и, главное, разнообразной (например, не ограничивается исключительно областью глагольной морфологии), составить по ней представительный лексический список — задача не сильно более

сложная, чем составить такой список по общему грамматическому описанию. Однако на полноту такого списка рассчитывать невозможно.

Сборники текстов в отдельный тип источников не выделяются: тексты, как правило, не публикуются в отрыве от словарей (т. е. либо содержат словарь в приложении, либо сами представляют собой приложение к вышедшему ранее словарю), хотя наличие текстового корпуса почти всегда серьезно облегчает задачу корректного выбора синонима.

По второму параметру — качество данных (имеется в виду в первую очередь степень фонологической и фонетической точности транскрипции) — известные нам работы можно расклассифицировать примерно на три типа:

А. *Недостоверные*. Это работы, в которых могут смешиваться и неправильно опознаваться сегменты, имеющие лишь самое поверхностное фонетическое сходство (в нашей модели — «фонетически несовместимые»). Фонетические системы языков Африки в целом не отличаются повышенной сложностью, что позволяет африканистам до сих пор активно пользоваться, хотя и с оговорками, материалами, собранными еще в XIX в. такими исследователями, как Кёлле, Барт, Нахтигаль, Райниш и др., несмотря на отсутствие у их авторов современных уровней «фонетического тренинга». Существенное исключение — работы XIX-го и первой половины XX-го века по «койсанским» языкам (В. и Д. Блик, Л. Ллойд и многие другие), в которых, как показывает сопоставление с данными более поздних исследований, на регулярной основе смешивались разные типы щелчковых звуков, в том числе такие, которые явно должны были различаться по месту артикуляции.

«Недостоверные» данные допустимо привлекать к лексико-статистическому исследованию (для многих койсанских языков данные более высокого качества отсутствуют в принципе), но только в том случае, если они доступны для значительного количества близкородственных языков. В этой ситуации сопоставительный анализ материала обычно позволяет сделать корректные этимологические выводы и даже, в отдельных случаях, реконструировать «правильную» фонологию записанного слова. Идеальная ситуация — группа из нескольких «недостоверно» записанных

языков, включающая хотя бы один «достоверный» язык, который можно использовать как своего рода «якорь» (такова, например, сегодня роль языков н|у и !хонг для южнокойсанской группы).

Б. *Проблемные*. «Проблемными» в том или ином отношении являются почти все полевые записи, собранные при исключительно пассивном участии носителей (т. е. без верификации данных самими носителями, прошедшими специальную лингвистическую тренировку). В случае с языками Африки среди наиболее типичных проблем следует отметить:

(а) некорректную запись просодии (в худшем случае — полное неразличение тональных оппозиций или их сознательное игнорирование, в лучшем — элементарные ошибки в определении тона): практика показывает, что лексические материалы двух исследователей, независимо друг от друга описывающих один и тот же язык (или даже диалект), будут непременно различаться в плане записи тонов хотя бы в 10-15% случаев (иногда и намного чаще);

(б) неточности в записи вокализма, особенно такие, которые имеют отношение к различению т. н. «оппозиции по +/-ATR». Такие противопоставления, как $e : \epsilon$, $o : \text{ɔ}$, $u : v$ и др., чрезвычайно широко распространены по всему африканскому континенту, но в одних языках они имеют фонологический статус, в других же являются позиционными вариантами. Соответственно, теоретически возможны и нередко встречаются на самом деле две крайности: с одной стороны, выделение огромного количества позиционных вариантов как самостоятельных «гласных» (сопровожаемое почти неизбежным нагромождением транскрипционных ошибок), с другой — наоборот, полное игнорирование богатых вокалических оппозиций со сведением всего разнообразия гласных к стандартной пятичленной «трапеции»;

(в) неточная запись консонантизма, в первую очередь — т. н. ларингальных признаков, особенно там, где речь идет об оппозициях, нетипичных для «европейского уха», таких, как противопоставление по имплозивности/эксплозивности.

В рамках нашего исследования сознательно не ставится цель максимально точно разрешить соответствующие проблемы, т. к. многие из них заслуживают написания отдельных монографий, включающих полную роспись сравнительного материала по

разным источникам данных. Отметим, что все эти гипотетические ошибки и неточности в целом укладываются в концепцию «фонетической совместимости», допускающую постулирование (хотя бы в рабочем порядке) когнации *без* установления сто-процентно работающей системы соответствий.

Так, например, анализируя рефлексы празападносурмийской основы 'огонь' в языках-потомках, мы сталкиваемся с более или менее произвольным варьированием звонкой и импловивной артикуляции в начальной позиции: дидинга *gʷiɔ* и мурле *gʷ* противопоставлены нарим *giɔ* и баале *gʷo* [Yigezu 2002: 369]. Этот ряд соответствий не является регулярным; в «стандартных» случаях как звонкий *g-, так и импловивный *gʷ- должны сохраняться во всех языках-потомках. Следовательно, наблюдаемая вариация может объясняться одним из трех способов:

а) различия в артикуляции начального согласного *реальны* и отражают специфическое позиционное развитие прафонемы (например, перед губными гласными?) в двух из четырех языков;

б) различия в артикуляции начального согласного *фиктивны* и отражают ошибочную транскрипцию (или даже элементарные опечатки) в работе исследователя (М. Игезу);

в) природа различий носит *смешанный* характер — например, в одном из языков действительно имело место позиционное развитие, в другом же имела место ошибка (опечатка, плохо слышанный сегмент, работа с конкретным «нетипичным» носителем, имеющим тенденцию «подстраиваться» под артикуляцию соседнего языка и т. п.).

Конкретный опыт работы показывает, что подавляющее большинство таких «проблемных» данных в принципе не может получить убедительной интерпретации без выхода на детально-профессиональный уровень (системная работа с носителями всех известных диалектов и говоров, целевым образом направленная на снятие всех фонетических и фонологических неоднозначностей). Единственное, что можно предпринять в данной ситуации — стараться максимально четко разграничивать между «надежными» и «проблемными» зонами внутри доступного для анализа языкового материала.

В частности, для языка мурле фонологическое различие между звонким *g* и импловивным *gʷ* отмечает только М. Игезу; ни

в одном из предшествующих описаний этого языка (Лит, Такер, Арсенен) такой оппозиции не отмечено (имплозивный *g* не постулируется даже в качестве фонетического варианта). Напротив, для баале касательно вопроса о противопоставлении звонкого и имплозивного велярных согласных существует консенсус хотя бы между двумя главными исследователями этого языка (см. [Yigezu & Dimmendaal 1998]), и, судя по сопоставительным данным, при разграничении старых звонких и имплозивных согласных в целом следует опираться скорее на баале, чем на мурле или дидинга.

Это не означает, что проблему реконструкции начального согласного в западносумийской лексеме 'огонь' можно считать разрешенной. Опираясь в первую очередь на данные баале, мы предполагаем в качестве более вероятного праязыковой вариант **gno*, но не даем при этом конкретного ответа на вопрос, откуда в мурле и дидинга (в записях Игезу) появляется имплозивный *g*. Вместо этого мы *допускаем*, что (а) все перечисленные формы связаны генетическим родством, т. е. восходят к общему западносумийскому источнику; (б) смена артикуляции со звонкой на имплозивную (или, что менее вероятно, с имплозивной на звонкую) вызвана каким-то неизвестным фактором, который мог быть как объективным (позиционное изменение), так и субъективным (неточность в записи). Мы *не* разделяем *g*-группу и *g*-группу форм на два неродственных друг другу набора когнатов, поскольку образуемые ими корреляции (а) связаны, если и не рекуррентным фонетическим соответствием, то, по крайней мере, высокой степенью фонетического сходства; (б) входят в «проблемную» зону, для которой детальное выяснение соответствий на данный момент невозможно по техническим причинам.

Аналогичным образом в области *вокализма* сопоставляемых языков серьезной, требующей обстоятельного анализа, проблемой можно считать значительные вокалические различия, например, по ряду или подъему (корреляции типа *a : o*, *a : i*, *i : u* и т. п.¹), но не различия по +/-ATR (*ε : e*, *l : i* и т. п.), которые иногда

¹ Практика показывает, что в языках Африки, особенно Северной, такие различия, если они имеют место в реальных когнатах, чаще всего вызваны общей тенденцией к гармонии гласных («умялутной» переогласовкой корневого вокализма под влиянием суффиксального или наоборот).

отражают просто степень фонетической детализации транскрипции исследователя, а иногда вообще фиктивны. Разумеется, для многих языков противопоставление по +/-ATR строго фонологично, но это не означает автоматически, что в каждом случае, когда в языке А записано ϵ , а в родственном ему когнате в языке Б записано e , этому факту необходимо дать строго обоснованное историческое объяснение, чтобы формы могли считаться «настоящими» когнатами; прежде чем отправляться на поиски такого объяснения, неплохо хотя бы убедиться (например, по данным двух независимых друг от друга описаний, если таковые имеются) в том, что речь не идет о простой ошибке.

В. *Достоверные*. Это — работы, выполненные на наиболее высокопрофессиональном уровне, желательно (хотя и не обязательно) с активным участием самих носителей языка. Вплоть до 1970х—1980х гг. такие работы выполнялись в основном лишь по крупнейшим языкам континента, но за последние десятилетия общий уровень исследовательской работы по языкам Африки, включая мелкие и вымирающие языки, значительно вырос.

Тем не менее, как уже было указано выше, до сих пор типична ситуация, когда из нескольких языков, составляющих единый таксон, по-настоящему достоверное и подробное описание существует не более чем для одного или двух. Это неизбежно заставляет использовать «достоверные» описания как опорные точки, но при этом исследователю-компаративисту необходимо следить и за тем, чтобы избежать соблазна автоматически переносить подробно и достоверно описанные особенности опорных языков на прауровень анализируемого таксона без тщательного сопоставления их с данными по проблемным и даже «недостоверным» источникам.

Скажем, тот факт, что из всех южнокойсанских языков к концу XX в. в живых остался только !хонг, вряд ли мог быть исторически обусловлен тем, что именно !хонг удалось сохранить наиболее архаичную фонетическую структуру — следовательно, сложнейший звуковой инвентарь этого языка, описанный Э. Трэйллом, нельзя априорно проецировать на праюжнокойсанский уровень. Разумеется, вероятность того, что исследователи южнокойсанских языков, работавшие с ними в довоенный период (В. и Д. Блик, Л. Ллойд и др.), упустили из виду значительное

число фонологических оппозиций, весьма высока, но, с другой стороны, трудно поверить и в то, что сверхсложная система консонантизма !хонг сохранилась в неизменном виде с момента первичного распада южнокойсанского праязыка (глотохронологически датируемого по меньшей мере концом 2-го тыс. до н. э.). Разрешить эти сомнения поможет лишь тщательный компаративный анализ материала.

Подводя итоги раздела, подчеркнем еще раз, что при разработке лексикостатистической классификации общая релевантность источника зависит в первую очередь от того, содержит ли он уникальную информацию по лексике (как базисной, так и культурной) соответствующего языка, а не от того, насколько «достоверно» записана эта информация. Разумеется, каждый источник должен получить определенную, пусть даже очень приблизительную, оценку по представленным параметрам, но игнорировать, т. е. исключать из рассмотрения, источники только из-за их «недостоверности» категорически недопустимо. Если исходить из презумпции, что «недостоверность» тех или иных источников отражает лишь случайные ошибки исследователя, а не злонамеренную фальсификацию материала, то исторические сценарии, реконструируемые для эволюции лексических систем, должны в идеальной ситуации работать на *всем* объеме известного материала — именно в этом, а не в отбрасывании скопом «плохих» данных, и заключается настоящая научная скрупулезность.

1.8.3.2. *Проблема языкового разнообразия.* Недостаток описательных данных по подавляющему большинству языков Африки исследователи-африканисты нередко пытаются компенсировать «глобальным охватом» своих описаний или сопоставлений. В силу того, что на африканском континенте сконцентрировано от одной четверти до, возможно, одной трети всех языков планеты, это неизбежно подталкивает компаративиста к тому, чтобы с самого начала оперировать крупными таксономическими единицами — чему способствует и неоспоримый факт действительного присутствия на территории Африки по крайней мере нескольких таких единиц, являющихся самоочевидными.

Так, вряд ли можно даже интуитивно усомниться в том, что единую генетическую общность составляют все языки банту

(более 500) или, по меньшей мере, «центральные» банту (более 300). Легко определяется генетическое родство внутри трех основных ветвей чадской семьи (от 40 до 70 языков в каждой), трех ветвей нилотской семьи (в среднем 20-30 языков в каждой), семьи манде (ок. 70 языков) и ряда других. В подобного рода ситуациях у исследователя, заинтересованного в построении общей модели, предсказуемо может возникнуть соблазн «перескочить» от таксонов 1-го уровня непосредственно к таксонам 2-го или даже 3-го уровней, что на практике выливается в приравнивание данных *живых* языков к данным реконструкций низких уровней. К сожалению, такая практика в лучшем случае оказывается методологически неудовлетворительной, в худшем — приводит к получению некорректных результатов.

В качестве примера возьмем второе по порядку вхождение в этимологическом словаре нило-сахарских языков К. Эрета [Ehret 2001: 257]: **bā* 'мокнуть, промокать' (первое вхождение, **bā* ~ **bā*: 'часть', содержит гораздо меньшее количество материала). Этимология, предлагаемая на уровне макросемьи (таксон 3-го уровня в нашем понимании), представлена следующими данными¹:

(а) працентральносуданск. **ba* 'дождь'. Центральносуданские языки — таксономическая единица как минимум 2-го уровня, возможно, даже 3-го; реконструкция принадлежит автору и отсылает, вместо конкретных языковых данных, к неопубликованной рукописи. Без соответствующих данных трудно понять, что имеется в виду, но стоит подчеркнуть, что из всех известных нам центральносуданских языков только в языке аджа (небольшая и очень проблематичная группа креш) слово 'дождь' звучит как *ba*, т. е. выводимо (как редупликация?) из праформы **ba*;

(б) фур *bà:n* 'ручей', *bà:ù* 'лужа'. Фур — язык-изолят, так что дистрибуция, очевидно, удовлетворительна, хотя никаких аргументов в пользу того, что эти два слова действительно восходят к одному корню, автор не приводит, просто отделяя финальные

¹ Сравнительный материал более или менее удовлетворяет фонетическим соответствиям, которые автор постулирует во вводящей части; с семантической точки зрения многие из сопоставлений вызывают серьезные нарекания, но для нас в данный момент значение имеет в первую очередь дистрибуция когнатов.

сегменты как архаичные суффиксы, восходящие якобы к пра-нило-сахарскому уровню;

(в) донголави *ba*: 'участок орошаемой земли, окопанный по краям для удержания воды'. Донголави — один из двух основных представителей диалектного пучка кенузи-донголави, входящего в нильско-нубийскую подгруппу нубийской семьи. «Реконструируемость» столь специфического культурного термина на общенубийском уровне чрезвычайно сомнительна, если ни в одном другом нубийском языке или диалекте не обнаружено фонетически и семантически надежных параллелей (судя по всему, таковые действительно отсутствуют);

(г) гаам *bāi* 'болото'. Архаичность этого слова было бы неплохо подтвердить параллелями из ближайших родственников гаам — восточноджебельских языков ака, кело и моло (что, впрочем, затруднительно из-за скудности лексических данных);

(д) западнодаджу **ba(y)-* 'плавать' → шатт *baʉa*, лигури *ba*. В восточнодаджу подгруппе языков даджу, представленной большим количеством языков, эта основа отсутствует. Прежде чем привлекать **ba(y)-* к внешнему сравнению, следует по крайней мере постараться определить, нет ли каких-либо существенных аргументов в пользу инновативного происхождения основы 'плавать' в той или иной из двух основных ветвей даджу, что, по-видимому, сделано не было;

(е) мурле *ba:yið-* 'плавать'. Фонетическое и семантическое сходство с даджу **ba(y)-*, на первый взгляд, разительное, однако тщательный анализ данных мурле в контексте его ближайшего генетического окружения полностью разрушает эту иллюзию. Мурле — один из наиболее хорошо описанных представителей «наддиалектного» пучка дидинга-мурле, входящего далее в юго-западносурмийскую ветвь южносурмийской группы. В конкретных источниках данных по мурле форму, цитируемую у Эрета, обнаружить не удалось; значение 'плавать' в этом языке, согласно одному источнику, выражается основой *tiri* [Yigezu 2002: 388], согласно другому — основой *rit* [Lyth 1971: 92]. По-видимому, цитируемое *ba:yið-* 'плавать' — это на самом деле *bai-ð-* 'переходить (реку)' [Lyth 1971: 5], которое следует далее сопоставлять с дидинга *bagi* 'переходить', 'переправляться' [Driberg 1931: 157]. Таким образом, более тщательное внешнее сравнение на «мелком»

уровне помогает определить, что речь здесь идет не о корне **ba(y)*- 'плавать', а о корне **bag-* или **bag-* 'переходить', сравнивать который с центральносуданским **ba* 'дождь' и др. неуместно;

(ж) очоло (= шиллук) *bay* 'перетекать через край' (ср. также именное производное *bay-o* 'наводнение'). Шиллук — один из приблизительно 30 языков западнонилотской группы, не менее десяти из которых представлены среднего или крупного размера словарными описаниями. То, что рефлекс предполагаемого нило-сахарского этимона обнаружен исключительно в шиллук, означает, что основа либо не может быть реконструирована на пра-западнонилотском уровне, либо просто не была выверена по другим словарям (впрочем, беглый их анализ показывает, что параллелей в нуэр, динка, ануа, луо и др. языках действительно нет);

(з) тесо = *baun* 'промокать'. Ситуация аналогична предыдущей: тесо — *один* из не менее чем 15-20 языков восточнонилотской группы, из которых по меньшей мере для пяти-шести существуют достойные словарные описания (масаи, бари, туркана и др.). Без попытки вывести эту основу на правосточнонилотский уровень любые дальнейшие внешние сравнения будут сомнительны.

Подчеркнем также, что «общенилотское» родство, если оно действительно существует (что весьма вероятно), все же не может оправдывать изоглоссу типа «шиллук-тесо»: три основных ветви нилотской семьи отделены друг от друга очень значительным временным интервалом, возможно, даже превышающим дистанцию между ветвями индоевропейской семьи, и для того, чтобы с достаточной уверенностью говорить о «пранилотском» статусе той или иной основы, необходимо продемонстрировать ее «реконструируемость» на уровне *каждой* из ветвей, в которых она якобы оказывается представленной;

(з) ик *ɪ=áb* 'быть холодным'. Ик — один из трех известных языков кулякской («руб» в терминологии Эрета) группы, образующий отдельную ветвь; в двух близкородственных ему языках, тепес и ньянги, параллелей к этой основе не обнаруживается. Следует признать, что дистрибуционный аргумент в *данном* случае не несет такой же силы, что и, скажем, в ситуации с нилотскими языками: общее число кулякских наречий сравнительно невелико, и на момент публикации труда Эрета только по ик существовал более или менее подробный словарь (сегодня

доступно также большое лексикографическое описание по тепес, или со). Таким образом, подменять общекулякскую реконструкцию синхронной основой в ик в принципе допустимо. При этом, однако, бросается в глаза разительное отличие от большинства перечисленных выше форм: семантическое (значение 'холодный' — единственное из всей группы, не имеющее непосредственного отношения к 'воде') и фонетическое (даже отделив потенциальный глагольный префикс t= , мы остаемся с неясным образом «метатезированной» формой =áʔ).

Общий анализ ситуации неизбежно приводит к следующему выводу: построить данную «этимологию» оказалось возможным исключительно благодаря тому, что две сотни известных на сегодняшний день «нило-сахарских» языков предоставляют богатый лексический материал, избирательное использование которого позволяет создать иллюзию надежно реконструируемого праязыкового корня, сохраняющегося в большом количестве дочерних ветвей. Спроецировав ситуацию на область, например, ностратической этимологии, мы увидим, что сопоставление Эрета примерно эквивалентно попытке обосновать праностратическую основу на материале, скажем, (а) польского, (б) праобско-угорского, (в) азербайджанского, (г) сванского, (д) северно-дравидийского (еще и со значительным семантическим разбросом рефлексов) — при том, что в серьезных работах по ностратике (начиная со словаря В. М. Иллич-Свитыча) такого рода «этимологии» не допускаются принципиально.

Впрочем, справедливости ради стоит отметить, что ностратическая реконструкция без них легко обходится, т. к. для всех больших семей, к которым относятся эти языки — индоевропейской, уральской, тюркской и т. п. — существуют подробные праязыковые реконструкции и этимологические словари на несколько тысяч реконструированных корней и основ; для языковых семей Африки такие словари исчисляются единицами (несколько более выгодное по сравнению с прочими положение здесь занимают семьи, входящие в афразийскую макросемью — возможно, вследствие активного влияния на специалистов по этим семьям со стороны классической семитологической традиции). Но отсутствие таких словарей никак не может служить оправданием непосредственного «перепрыгивания» от уровня

живых современных языков сразу же на уровень макросемьи десяти тысячелетней (или более) давности.

Колоссальное лексическое разнообразие языков Африки, помноженное на некритическое восприятие основных классификационных постулатов Дж. Гринберга, часто мешает воспринимать как убедительные работы даже намного более узкого характера, чем этимологический словарь Эрета — например, исследования, посвященные обоснованию генетической принадлежности того или иного конкретного языка-изолята или небольшой языковой группы.

Не отрываясь от «нило-сахарской» тематики, возьмем для примера работу [Blažek 2007], посвященную обоснованию нило-сахарского происхождения онгота — языка, относительно генетического статуса которого до сих пор не существует общего согласия, хотя большинство исследователей, специально занимавшихся проблемой онгота, склонны относить его либо к кушитской, либо к омотской ветви афразийской макросемьи (К. Эрет, М. Тоско, Г. Сава), либо вообще выделять в отдельную, самостоятельную афразийскую ветвь (Г. Флеминг). Поскольку более или менее очевидным является наличие в онгота нескольких лексических слоев, из которых только один должен оказаться генетически значимым (при условии, что мы либо отвергаем для онгота классический сценарий «креолизации», либо требуем его максимальной экспликации, т. е. четкого разграничения лексических слоев с выделением «языка-донора» и «языка-рецептора»), проблема в целом сводится к стратификации и квантификации данных — требуется определить, какой слой в онгота представлен бóльшим количеством элементов и в каком из них налицо бóльший процент базисной, а не культурной лексики.

Согласно концепции В. Блажка, «первичный» лексический слой в онгота должен определяться как нило-сахарский, а не афразийский. В поддержку этого утверждения он приводит 30 эксклюзивных лексических изоглосс между онгота и нило-сахарской макросемьей и еще 14 изоглосс, где нило-сахарские параллели можно привести в дополнение к уже известным афразийским (т. е. «оспариваемые» случаи).

Вне зависимости от того, насколько само по себе число 44 (или, допустим, 30) может считаться достаточным для надежного

определения генетического статуса такого языка-изолята, как онгота (при том, что в списке содержится как базисная, так и культурная лексика), конкретные выводы, к которым приходит автор, не всегда удачно согласуются с той методикой, с помощью которой эти выводы получены. Оптимальный сценарий представляется В. Блажку следующим образом:

«Ongota is the last relic of a Nilo-Saharan language or a group of languages, maybe representing an independent branch within East Sudanic or even Nilo-Saharan, of the type Nara, Fur, Gumuz, Kuliak, etc. (classified by Bender and Ehret differently as branches of the East Sudanic = Kir-Abbaian superbranch or as independent members of the Nilo-Saharan phylum)» [Blažek 2007: 8].

Из *двух* взаимоисключающих гипотез, высказанных в этом пассаже (онгота — *либо* одна из ветвей восточносуданской ветви нило-сахарской макросемьи, *либо* одна из первичных ветвей нило-сахарской макросемьи), автор реально тестирует только вторую, т. к. лексический материал берется им из всего разнообразия нило-сахарских языков, и специфическая близость именно к восточносуданским не продемонстрирована статистически; разброс материала позволяет говорить об онгота лишь как о «нило-сахарском» языке *вообще*. Однако это автоматически ставит новый вопрос: *какие* генетические маркеры (лексические, грамматические, фонетические) мы можем считать «показательно нило-сахарскими», и верно ли, что именно они, хотя бы в минимальном количестве, обнаружены в онгота? Если речь идет о лексике, верно ли, что приводимые сопоставления по-настоящему *реконструируемы* для нило-сахарского, или это не более чем набор фонетически сходных форм, которые удалось собрать вместе благодаря богатому выбору языков?

К сожалению, тщательный анализ материала скорее склоняет к второму варианту ответа. Так, например, для онгота *i:ʔa* 'рука', 'палец' в качестве возможных «нило-сахарских» параллелей приводятся такие формы, как ныманг *áiyi ~ áízi* 'рука' и тама *kuá* 'ляжка' (!). Очевидно, что первая из них стоит ближе к онгота по семантике (но требует экспликации «метатезы» вокализма), вторая — по фонетике (но требует постулирования семантического перехода или развития из родового значения 'конечность').

Ситуация, однако, осложняется тем, что и ньяманг, и тама во всех классификациях нило-сахарской макросемьи, предложенных на настоящий момент (кроме классификации К. Эрета, доверие к которой минимально), включаются в восточносуданскую ветвь, а К. Рильи [Rilly 2009] даже приводит аргументы, согласно которым ньяманг, тама, нара и нубийские языки связаны особой близостью в рамках «северной» ветви восточносуданского (в «южную» входят сурмийские, нилотские и др. языки).

Если это действительно так (от собственных выводов мы воздержимся вплоть до завершения первичного анализа лексико-статистического материала, но аргументация Рильи выглядит очень серьезно), можно усомниться в том, что вообще возможна какая-либо этимологическая связь между ньяманг *áiyi* 'рука' и тама *ɫya* 'ляжка'. Скорее следовало бы попытаться установить такую связь между ньяманг *áiyi* и тама *àwí* 'рука' (что и делает в своей монографии Рильи), а для тама *ɫya* искать другие параллели. Установив эту связь, необходимо далее заняться вопросом реконструкции общевосточносуданской основы 'рука', для которой, как показывает сопоставительный анализ материала, на самом деле очень высока вероятность наличия на прауровне переднеязычного взрывного или спиранта¹. Если окажется, что эта основа действительно должна восстанавливаться в виде **aT-(i)*, идея генетического родства с онгота *iʔa* оказывается более чем сомнительной — разве что удастся продемонстрировать фонетическую совместимость интервокальных согласных.

Еще более проблематичным оказывается сравнение в случае с другим «стабильным» элементом списка Сводеша — 'вода'. Здесь онгота *ɕaʔawa* 'вода', 'река' (слово, скорее всего, производно от глагольной основы *ɕaʔaw-* 'пить') сопоставляется с:

— рядом форм в близкородственных восточноджебельских языках (ака *ɕí*: 'вода' и др.), с трудом выводимых на общеджебельский уровень;

¹ Это видно даже по тому ограниченному материалу, который приводит сам В. Блажек: сравнивая онгота *iʔa* с баале (югозападносурмийская группа) *ayí* 'рука', он упоминает о существовании в дидинга (та же группа) формы *adít id.*; общеюжносурмийская основа однозначно восстанавливается именно как **adí*, а не как *ayí*.

— опо *č'i* ~ *zi'i* 'вода' (команская семья). Сопоставление этого слова с данными остальных команских языков [Bender 1983: 274] однозначно показывает, что на пракоманском уровне была представлена форма **yi'i* с вторичной аффрикативизацией в опо;

— масалит (семья маба) *sá* 'вода' или (!) маба *anži*: id.: два несомненно *разных* корня, с одинаковой вероятностью привлекаемых к сравнению.

Если речь идет об элементарном «моноконсонантном» сравнении, без учета таксономической структуры сравниваемых единиц, к подобного рода сравнению легко подключить в дальнейшем и афризийские «параллели» (например, в омотских языках: бенч, ше *so?* 'вода'; кафа *ač'o* id.), и даже нигер-кордофанские (много похожих форм в языках ква, манде, атлантических и др.).

На данном примере (*далеко* не единственном в той части африканистики, которая ориентирована на глубинное сравнение) можно наглядно убедиться в том, что *только* «ступенчатое» сравнение, с постепенным, достоверно обоснованным и подробно эксплицированным сведением языкового разнообразия Африки к значительно меньшему количеству праязыковых реконструкций может если и не полностью ликвидировать, то, во всяком случае, существенно ослабить угрозу принятия случайных совпадений за отражение генетических общностей.

Отсюда вытекает методологическое требование, которого мы будем стараться придерживаться на протяжении всего исследования: *ни одно слово ни одного языка Африки* не должно и не может привлекаться к сопоставлению на «глубинных» уровнях до тех пор, пока мы не суммировали наши знания о его исторических связях на уровне «поверхностном». Недопустимо сравнивать, например, слово со значением X в группе боле-тангале западночадской семьи со словом с тем же значением X в ираквской группе южнокушитской семьи — даже при условии, что мы считаем родство этих двух групп «доказанным» в пределах афризийской макросемьи — до тех пор, пока не будет показана «реконструируемость» этих этимонов в том же значении X сначала на западночадском и южнокушитском, а затем и на общечадском и общекушитском уровнях. Только такой подход обеспечит возможность детального и эмпирически достоверного *моделирования* процесса исторического развития от праафризий-

ского языка к его современным потомкам — то, к чему, собственно говоря, и должна стремиться любая компаративистика, как традиционная, так и «макро-».

Для африканского же континента, с его головокружительным языковым разнообразием и, скорее всего, глубочайшей древностью большинства представленных на нем макросемей, такой подход особенно актуален. Тем важнее на начальных этапах исследования вообще «забыть» о макросемьях — и вернуться к вопросу их выделения и обоснования только после того, как лексикостатистический и этимологический анализ материала позволит с максимально возможной тщательностью вырезать все промежуточные ступеньки лестницы, по которой можно будет затем спуститься в более глубокое прошлое.

1.8.3.3. *Проблема языковых контактов.* Вопрос разграничения между сходствами, отражающими «вертикальную» и «горизонтальную» трансмиссии (праязыковое наследие и ареальные контакты), актуален для всех без исключения языковых семей мира, и говорить о какой-либо «эксклюзивно-африканской» специфике здесь неуместно. Тем не менее, любой специалист, хорошо знакомый с проявлениями языкового разнообразия на территории Евразии, скорее всего, увидит ряд принципиальных отличий общего характера от соответствующей ситуации в Африке.

Если в самом общем виде воспользоваться спорной, но популярной классификацией лингвогеографических ареалов на «зоны сохранности» («residual zones») и «зоны распространения» («spread zones»), в свое время введенной Дж. Николс [Nichols 1992], то Евразию в целом скорее можно охарактеризовать как spread zone, Африку же скорее — как residual zone. Разумеется, внутри этих макроареалов классификация намного более детальна: в Евразии нередки примеры таких «мини-зон», как Кавказ (residual), в Африке же типичным примером spread zone является почти вся территория к югу от экватора, охваченная относительно недавней экспансией языков банту. Тем не менее, за исключением ситуации с банту, остальная территория Африки (Сахара, Судан, Эфиопия, Западная Африка и т. д.), на которой как раз и сконцентрировано основное языковое разнообразие континента, в значительно меньшей степени охватывалась последовательными волнами языковой унификации, чем Евразия. Здесь для

дальнородственных и неродственных семей типичен скорее сценарий «вклинивания» представителей одной семьи на территорию, ранее занятую другой (другими), которое либо вообще не приводит к языковой ассимиляции, либо растягивает этот процесс на многие тысячелетия.

Логическим следствием такого расклада является наличие на территории Африки большого количества мелких и крупных «языковых союзов», в рамках каждого из которых генетически не связанные друг с другом языки вторично дрейфуют по направлению друг к другу, заимствуя в ходе сосуществования на одной территории как лексические элементы, так и грамматические морфемы и даже типологические черты. Сегодня значение ареального фактора для решения актуальных вопросов генетической классификации языков Африки однозначно признается всеми африканистами, ср., например, утверждение в [Heine & Nurse 2008: 2]: «...genetic relationship does not provide the only parameter for diachronic language classification in Africa; rather, there is reason to maintain that the African continent can equally well be classified in terms of areally defined groupings». Выделение и обоснование таких «ареальных группировок», проводимое на стыке исторического языкознания и общей типологии, постепенно завоевывает популярность как отдельная область африканистики; ср., в частности, предложенную недавно Т. Гюльдемманном ареально-типологическую классификацию языков Африки на пять «макроразнов»: (1) сахарскую, (2) чадско-эфиопскую, (3) макро-суданскую, (4) банту, (5) калахарскую [Güldemann 2010].

Поскольку в основу данного исследования положена лексико-статистика, очевидно, что главным теоретическим вопросом, завязанным на ареальном факторе, для нас будет следующий: насколько попадание того или иного языка или языковой группы в гущу «языкового союза» способно вытеснить релевантные для установления их генетического статуса лексические маркеры? Иными словами, может ли один африканский язык заимствовать из другого столько базисной лексики, чтобы происхождение его оказалось невозможно установить, опираясь исключительно на лексическую информацию?

Как и в других подобных случаях, дать сколь-либо четкий ответ на этот вопрос можно только в рамках униформистской

модели: если ситуация, изложенная в заданном вопросе, не обнаруживается в современных языках, нет никаких оснований предполагать, что она могла иметь место и в отдаленном прошлом (по крайней мере, вплоть до уровня выделения макросемей). При этом конкретный опыт работы с языками на «мелких» хронологических уровнях показывает, что научиться аккуратно отделять контактные лексические связи от унаследованных на самом деле нетрудно; все зависит в первую очередь от количества и качества доступных лексических данных.

Сопоставительное исследование базисной лексики языков в рамках произвольно взятого таксона 1-го уровня показывает, что в *большинстве* случаев базисная лексика в языках Африки ведет себя точно так же, как и во всех остальных, т. е. либо не заимствуется совсем, либо в исключительных случаях — особенно в пределах 50-словной («стабильной») части списка, где явные заимствования, как правило, исчисляются единицами. Такую ситуацию для Африки следует считать естественной *нормой* — но, разумеется, это не означает, что из соответствующей нормы принципиально невозможны многочисленные (в абсолютном, но никак не в относительном плане) исключения.

Во-первых, в Африке, как и в других лингвогеографических ареалах планеты, определенное распространение получили пиджины и креольские языки, где в роли лексификатора, как правило, выступают либо европейские языки (португальский, французский, английский и др.), либо арабский (подробный список см., например, в: [Childs 2003: 203-216]). Вопрос о том, являются ли «креолы» специфическим порождением Нового времени, обусловленным определенным витком экономического и культурного развития человечества, или же креолизация в том или ином виде была возможна всегда (или хотя бы, например, с начала неолитической революции и т. п.), носит скорее историко-философский характер, т. к. до сих пор ни для одного из реконструированных праязыков сколь-либо глубокого уровня не удалось, на базе сопоставления его с другими праязыками, показать его «креольскую» природу.

Впрочем, с точки зрения лексикостатистического анализа этой проблемы вообще не существует: «классический» креол однозначно определяется как генетический потомок того из

языков-лексификаторов, который «поставляет» для итогового продукта наибольший процент базисной лексики¹. В той редкой ситуации, когда нам удастся наблюдать очевидное «скрещивание» лексики одного языка с грамматикой другого, можно, при большом желании, говорить о «смешанном» происхождении этого языка, но «смешанность» можно определять и просто как повышенную (по сравнению со среднестатистической нормой) роль конвергентных процессов в эволюции исследуемого языка.

В качестве конкретного примера упомянем одну из «хрестоматийных» проблем генетической классификации — язык ма'а или мбуту в Танзании, накладывающий базисную лексику южнокушитского происхождения на очевидно бантускую грамматику. Консенсус относительно исторического статуса этого языка недостижим, поскольку зависит в первую очередь от личных предпочтений исследователя: так, К. Эрет [Ehret 1980] однозначно определяет ма'а как южнокушитский, а А. Б. Долгопольский [Dolgo-polsky 2008: 64] — как язык банту (что, впрочем, не мешает ему пользоваться лексическим инвентарем ма'а для уточнения южнокушитских этимологий). На тему определения исторического статуса ма'а и раскрытия возможных причин появления на свет столь странного языка написано несколько статей ([Goodman 1971; Tucker & Bryan 1974; Thomason 1983] и др.) и даже целая монография ([Mous 2003]), но «стабильного» места в общей классификации языков Африки ма'а занять так и не удалось.

На наш взгляд, однако, в подобного рода ситуациях гораздо важнее уметь правильно *описать* структуру и состав языка, вскрыть диахроническое происхождение большинства его элементов, чем обеспечить единую и непоколебимую *интерпретацию* соответствующего синхронно-диахронического описания. Заведомо позитивным моментом в ситуации с ма'а является то, что можно уверенно утверждать, что грамматика его уходит корнями в сферу

¹ Для того, чтобы при этом получались приемлемые классификационные результаты, те относительно немногочисленные элементы базисной лексики, которые сохраняются от «исходного» языка после его столкновения с лексификатором, следует формально трактовать как заимствования (хотя и это не всегда решает проблему, ср. аномальную ситуацию с лексикостатистической матрицей по креольским языкам токписин и бислама в [Беликов 1998: 32]).

банту, а лексика — в южнокушитскую; дальнейшая же классификация будет зависеть от того, какой уровень языка мы склонны считать более «генетически значимым» — лексический или морфологический.

Поэтому, вместо того, чтобы углубляться в принципиально неразрешимый спор о том, существуют ли «смешанные» языки и т. п., остановимся подробнее на вопросе, гораздо более актуальном для исторической африканистики: бывает ли вообще базисная лексика в этих языках, с точки зрения ее происхождения, устроена «необычным» образом?

Выше (раздел 1.6.3.5), на примере сурмийского языка квегу, было показано, как выявление существенной аномалии в лексико-статистической матрице, включающей этот язык, позволяет сформулировать гипотезу о том, что большой процент базисной лексики этого языка был в свое время заимствован из неизвестного субстрата. Если бы подобного рода ситуации всегда «реконструировались» на основании исключительно косвенных данных, могли бы возникнуть сомнения в их реалистичности. К счастью, имеются и гораздо более наглядные примеры таких ситуаций — хотя следует еще раз подчеркнуть, что все они носят исключительный характер.

Одним из таких примеров является бросающееся в глаза различие между лексическим составом двух ветвей сонгайской семьи. В южносонгайских языках (зарма, денди, каадо и др.) заимствования в базисной лексике носят скорее окказиональный характер, ограничиваясь единичными арабизмами и, возможно, мандеизмами (иногда труднодоказуемыми); число их обычно не превышает 3-5% от общего состава 100-словного списка. Напротив, в северносонгайской ветви (тадаксахак, табарог и др.) число заимствований из берберских (туарегских) языков столь велико, что легко достигает 20 и более процентов от общего состава; в гипотетической ситуации, при которой исследователь знаком только с северной, но не с южной ветвью, он легко мог бы ошибочно причислить соответствующие языки к берберским или, по крайней мере, счесть их наиболее рано отделившейся ветвью «пара-берберской» общности. Впрочем, даже в этом случае на выручку приходит принцип динамической лексической градации: так, например, в языке тадаксахак из первой десятки элементов,

ранжированных в соответствии с «индексом стабильности», берберизмом оказывается одно слово ('язык'), из второй — три ('сердце', 'луна', 'ноготь'), из третьей — четыре ('зуб', 'новый', 'печень', 'хвост'); дальнейшая градация уже не столь показательна, но соотношения в пределах первой двадцатки лексем уже достаточно надежно показывают, что речь наверняка идет о позднейшей конвергенции.

Чем объясняются столь разительные расхождения между северной и южной ветвями сонгай? Как северные, так и южные языки все последние две тысячи лет (распад сонгайской общности датируется примерно I-II вв. н. э.) существовали в тесном окружении неродственных им афразийских и нигер-конголезских языков, но только для северной ветви социальные условия сложились таким образом, что лексические системы ее языков оказались «свободно открытыми» для проникновения иноязычных элементов. Современная социалингвистика пока что не в состоянии ни уверенно объяснить причины таких различий, ни, тем более, предсказать вероятную степень интенсивности языковых контактов, исходя из таких сведений, как общая численность носителей языка-донора и языка-рецептора, образ их жизни, характер взаимоотношений и т. п.

Естественно было бы предположить, что массивные заимствования из соседнего языка в базисную лексику собственного — закономерный первый шаг на пути к полному вымиранию языка, т. е. процесс, логическим завершением которого должен стать полный переход на язык соседа. Но, во-первых, смена языка далеко не всегда происходит именно таким образом — скорее наоборот: известные нам данные по лингвистической истории Европы и Азии показывают, что переход с одного языка на другой может быть постепенным, т. е. растягиваться на несколько поколений, но обычно характеризуется не существенным смешением кодов, а скорее «переходным» периодом двуязычия. Во-вторых, далеко не все северносонгайские языки можно назвать вымирающими — ср., например, нижеследующую оценку статуса диалекта тасавак:

«The Ingalkoyyu have a very positive attitude towards Tasawaq. Their language is one of the main things which distinguishes them as a people

and they are proud of it. Tasawaq seems likely to remain the primary language of the Ingalkoyyu as long as the town of Ingal survives, and the prospects for this seem good since the Ingalkoyyu appear to be a happy, healthy, and industrious people» [Rueck & Christiansen 2001: 13].

Таким образом, исторически допустима ситуация, при которой язык-рецептор внедряет в свою базисную лексику аномально большой процент заимствований из языка-донора, что, тем не менее, не лишает язык-рецептор самостоятельной жизнеспособности и не превращает его в новый диалект языка-донора (т. е. не приводит к смене генетического статуса). Такие языки можно метафорически назвать «языками с пониженным лексическим иммунитетом», т. к. они легко склонны к замене большей части своего лексического кода чужеродным (подчеркнем, однако, что речь идет именно о стабильных, четко фиксированных заимствованиях, а не о ситуации «переключения кода», при которой в дискурсе могут сосуществовать, в свободном или прагматически обусловленном варьировании, лексические элементы обоих языков).

Специфика африканского континента по сравнению с Евразией заключается в том, что столкновение двух языков, из которых один оказывается «доминирующим», а другой — «доминируемым», в целом чаще приводит к понижению у «доминируемого языка» лексического иммунитета, чем это бывает в Евразии, где типичным результатом такого столкновения в подавляющем большинстве случаев является полная языковая ассимиляция (возможное исключение — юго-восточноазиатский регион, за колоссальным языковым разнообразием которого могут скрываться многочисленные ситуации, аналогичные африканским).

В целом, по-видимому, ожидать повышенной концентрации языков с «пониженным иммунитетом» следует в тех географических областях, где общая языковая картина является наиболее пестрой, т. е. налицо множество языков из разных семей, зажатых в пределах сильно ограниченной территории и в буквальном смысле вынужденных к тесным контактными отношениям. Для Африки такие области — это прежде всего район Южного Судана и Эфиопии (особенно выделяются здесь Кордофан и горы Нуба), кенийско-танзанийские территории к востоку от озера Виктория и атлантическое побережье Западной Африки, но «ми-

ни-зоны» с повышенной вероятностью встречаемости таких языков могут быть обнаружены практически в любой области континента.

«Пониженный иммунитет» к заимствованиям может проявляться и в таких ситуациях, особенно трудных для анализа, когда язык-рецептор заимствует базисную лексику из нескольких окружающих его языков-доноров, в том числе относящихся к разным группам. В случае с языком квегу такие «вливания» имели место на разных хронологических этапах его развития (старый этап — из неизвестного, скорее всего, вымершего субстрата, возможно, первоначального языка, на котором говорили предки современных носителей квегу; новый этап — из доминирующих сегодня в районе обитания квегу омотских языков). В других случаях они могут, по-видимому, носить и одновременный характер — например, в базисной лексике «койсанского» языка хадза встречаются заимствования из банту, южнокушитских и южнонилотских языков, ни одно из которых не производит впечатления особой архаичности (впрочем, сам по себе язык хадза трудно причислить к языкам с «низким иммунитетом» — по сравнению с тем же северносонгайским его словарный состав намного более устойчив).

Фактор массовых заимствований в базисную лексику ни в коем случае не следует переоценивать, как это иногда происходит в чрезвычайно модном сегодня направлении «ареальной» лингвистики. Тем не менее, не следует забывать, что для любого африканского языка, как мелкого, так и крупного, всегда существует вероятность *субстратного* происхождения — историческое возникновение его как результат перехода изначально небольшой группы населения со старого языка на новый, с удержанием при этом определенной части старого лексического инвентаря, т. е. своего рода «мини-пиджинизация». Такая вероятность, по-видимому, даже выше, чем вероятность обратного процесса — «адстратизации», имевшей место в северносонгайских диалектах.

Так, например, субстратное происхождение можно с высокой степенью уверенности предположить для центральнокойсанской (кхой) семьи, включающей «готтентотские» языки Намибии и ряд «бушменских» языков Ботсваны и Зимбабве. Поверхностный анализ их базисной лексики выявляет сходства не только с

другими «койсанскими» семьями, но также и с языками банту и даже афразийскими, особенно чадскими (параллельно с лексическими изоморфизмами наблюдаются и определенные типологические и сегментные сходства со всеми тремя перечисленными языковыми общностями и в области грамматики, отмеченные в свое время еще К. Майнгофом в рамках уже упоминавшейся выше «нило-хамитской» гипотезы). Субстратное происхождение может характеризовать и такие африканские языки-изоляты, как онгота (см. выше критику гипотезы В. Блажка), шабо, лаал и другие; во всех этих случаях первоочередная задача исследователя — максимально четко разграничить лексические слои, коррелирующие с разными таксонами, и, с учетом фактора лексической градации, на основе исторического и статистического анализа выявить «языкообразующий» слой, отделив его от суб- и адстратных слоев¹.

1.8.3.4. *Типичные проблемы исторической фонологии и морфонологии языков Африки.* Большинство исторических процессов, которые удастся проследить и реконструировать для африканских таксонов 1-го уровня (т. е. с высокой степенью уверенности), имеет определенные типологические аналогии и в других лингвистических ареалах. Такие фонетические изменения, как: колебания по ларингальным признакам (оглушение / озвончение); палатализация велярных согласных под влиянием гласных переднего ряда; ослабление и выпадение интервокальных смычных; редукция срединных гласных в многосложных структурах, приводящая к образованию консонантных кластеров; гармония гласных, чаще всего проявляющаяся как влияние суффиксального вокализма на корневой — все эти развития, будучи, по-видимому, естественными (и, в отдельных случаях, даже предсказуемыми) для человеческого языка вообще, представлены в той или иной пропорции практически во всех макро- и микро-лингвоареалах африканского континента.

¹ Громоздкое определение «языкообразующий» употреблено здесь вместо естественного термина «первичный» постольку, поскольку «первичным» можно было бы в определенном смысле считать как раз субстратный слой — лексические (и какие-либо другие) следы того языка, на котором первоначально говорили люди, сознательно перешедшие на язык своих соседей.

«Специфически африканские» диахронические процессы, по-видимому, можно постулировать лишь в той мере, в которой они являются историческими преобразованиями «специфически африканских» типологических черт; грубо говоря, «странные» развития характерны только для «странных» фонем и т. п. Подробная сравнительно-типологическая характеристика языков Африки не входит в цели нашего исследования; серьезный, хотя и во многом поверхностный, типологический обзор Африки как единого лингвоареала (а также отдельных макро- и микроареалов африканского континента) легко доступен сегодня в виде коллективной монографии [Heine & Nurse 2008]; см. особенно статью [Heine & Leyew 2008], авторы которой задаются вопросом, корректно ли считать Африку отдельным лингвистическим ареалом (linguistic area), противопоставленным всей остальной планете, и в конечном итоге приходят к утвердительному ответу, выделяя шесть типологических признаков, которые вообще не встречаются в языках мира за пределами Африки, и десять дополнительных признаков, представленных в языках Африки намного чаще, чем в любых других ареалах планеты.

Если исключить из этого набора ряд морфосинтаксических признаков, не актуальных для этимологического и статистического анализа 100- и 50-словных списков, оставшиеся девять «специфически африканских» черт можно поделить на три группы:

А. «Необычные» фонетические единицы.

Сюда относятся:

А.1: щелчковый звук, или «кликсы», не встречающиеся в функции фонем естественного языка за пределами Африки. Являются своего рода визитной карточкой «койсанской» макросемьи (по Гринбергу), однако в результате ареальной диффузии проникли в ряд языков банту (зоны R и S по классификации М. Гасри), а также в южнокушитский язык дахало;

А.2: губные одноударные согласные («лабиальные флэпы»); за пределами Африки не встречаются (возможное исключение — австронезийский язык сика на о. Флорес). Район встречаемости ограничен Центральной Африкой (ЦАР, Камерун и ряд прилегающих территорий), но не имеет при этом четко очерченных генетических границ, т. к. лабиальные флэпы зафиксированы как в языках адамава-убанги (нигер-конголезская макросемья), так и в

центральносуданских (нило-сахарская макросемья) и даже чадских (афразийская макросемья);

А.3: лабиально-велярные согласные *kp*, *gb*, *ŋt* (не путать с лабиовелярными, т. е. огубленными велярными *k^w*, *g^w*, *ŋ^w*). Помимо Африки, представлены и в отдельных других ареалах (в первую очередь — в языках Новой Гвинеи), но в целом являются все же редким типом артикуляции. Наличие лабиально-велярных согласных — в первую очередь дистинктивная черта языков макросемьи нигер-конго, но встречается и в нило-сахарской макросемье (почти во всех центральносуданских языках), и в ряде чадских языков, т. е. на самом деле носит такой же ареальный характер, как и наличие лабиальных флэпов;

А.4: имплозивные (глоттализованно-ингрессивные) согласные, иногда определяемые как аллофоны звонких смычных, но чаще всего являющиеся самостоятельными фонемами (например, в составе тройной оппозиции «глухой : звонкий : имплозивный»); чаще всего встречаются в губном (*b*) и дентальном/альвеолярном (*d*) рядах, несколько реже встречаются велярный имплозивный *g* и палатальная имплозивная аффриката *ʒ*. Имплозивная артикуляция также не уникальна для Африки (встречается и в языках Юго-Восточной Азии и Южной / Центральной Америки), но максимальная концентрация языков, содержащих имплозивные фонемы, приходится именно на этот континент;

А.5: преназализованные согласные (*mb*, *nd*, *ɲz* и т. д.) в начальной позиции: встречаются в основном в нигер-конголезских языках, но с «заходом» на нило-сахарскую и афразийскую территорию (опять-таки прежде всего в центральносуданских и в чадских языках). Аналогичная особенность также характерна для ряда языков Юго-Восточной Азии, Океании и Австралии, но в несколько меньшей степени.

Подчеркнем еще раз, что важность фонетических характеристик А.3-А.5 для исторического языкознания заключается не столько в их *уникальности* для африканского континента, сколько в их географической *изолированности*: ближайшие типологические параллели обнаруживаются не на смежных с Африкой территориях (Европа, Ближний Восток и т. п.), а в значительном отдалении — как правило, в Юго-Восточной Азии и в Америке. На данный момент невозможно предложить конкретный ответ на

вопрос, следует ли этот факт рассматривать как результат исторической случайности, или же он отражает какие-то глубинные доисторические связи между языковыми макросемьями этих регионов; возможно, что в ходе дальнейших исследований эта ситуация когда-нибудь прояснится. Пока что соответствующую фонетическую специфику языков Африки имеет смысл рассматривать именно как *дистинктивную* особенность африканского континента, и искать ей объяснение, стараясь не выходить за пределы африканского ареала.

Б. «Необычные» фонетические процессы.

Сюда относятся:

Б.1-2: несколько типов уникальной или редкой по сравнению с другими языками мира гармонии гласных, из которых два вида уподоблений встречаются исключительно в языках банту (Б.1; см. [Clements & Rialland 2008: 53-55]), а один (Б.2) — гармония по степени напряженности, или, в современной западной терминологии, по +/-ATR (advanced tongue root) — характерен, напротив, почти для всей территории «макросуданского» и прилегающих ареалов, т. е. является важной типологической изоглоссой, охватывающей многочисленные языки и нило-сахарской, и нигер-кордофанской, и афразийской макросемей.

Сам по себе дистинктивный признак «напряженности», определяемый на множестве вокалических фонем языка, далеко не уникален для Африки, что и побуждает исследователей в качестве «экзотического» явления выделять скорее *поведение* этих гласных, чем сам факт их наличия. При этом, однако, нельзя упускать из виду то обстоятельство, что только на территории Африки фонологическая оппозиция по +/-ATR принимает столь внушительные масштабы. Для многих языков макросуданского и других ареалов типичными являются 9- или 10-членные фонологические системы, в которых +ATR гласным [i], [u], [e], [o] противопоставлены -ATR гласные [ɪ], [ʊ], [ɛ], [ɔ] (гласный нижнего подъема [a] чаще остается без пары, но иногда и здесь налицо оппозиция между -ATR [a] и +ATR [ʌ]). В *некоторых* языках, правда, оппозиция по +/-ATR нефонологична, т. е. соответствующие пары оказываются позиционно распределенными аллофонами; существующие описания в таких случаях часто противоречат

друг другу, и требуется немало усилий, чтобы определить, какое из них ближе к истине¹.

Вероятность того, что в языках с «развитым» фонологическим противопоставлением по +/-ATR (т. е. таких, где этот дистинктивный признак определен на всей системе вокализма или на ее большей части) неизбежно будет представлена и гармония гласных по +/-ATR (т. е., грубо говоря, в пределах одного фонетического слова можно встретить либо только +ATR, либо только -ATR), довольно высока — аналогичные случаи описаны в литературе для индийских языков (ассамский, бенгальский), халха-монгольского и других языков Евразии. Конкретные механизмы, отвечающие за фонетическую реализацию вокалических элементов в слове, могут при этом существенно отличаться: неудивительно, что гармония по +/-ATR — практически неисчерпаемая тема для различных исследований как для современной африканистики, так и для общей теории фонетики и фонологии. Насколько, однако, учет этих различий жизненно важен для определения генетической принадлежности конкретных языков и языковых групп, остается неясным.

Б.3: использование *исключительно* просодических (тональных) оппозиций для маркирования падежных отношений. Эта любопытная типологическая особенность характеризует очень небольшую группу языков Восточной Африки, к тому же еще и применяющих стратегию маркированного номинатива [König 2006]; учет ее для лексической реконструкции и лексикостатистического анализа необязателен (в историческом плане такого рода явления обычно бывают обусловлены грамматикализацией разнотонированных служебных частиц с последующим отпадением их в ауслутной позиции).

В. «Необычные» морфологические характеристики. К этой группе относится всего один признак — системы именных

¹ Так, например, почти во всех «старых» фонетических и словарных описаниях северно- и южнокойсанских языков, опубликованных в первой половине XX в., можно обнаружить противопоставление «открытых» (т. е., по сути, -ATR) гласных [ɛ], [ɔ] и «закрытых» (т. е. +ATR) [e], [o]. Ни в одном из более современных описаний тех из них, которые еще остались в живых, о подобного рода *фонологических* оппозициях не упоминается.

классов, не уникальные для Африки, но в наиболее массовом порядке распространенные именно на этом континенте.

Как известно, наличие в том или ином виде (префиксы, суффиксы) системы именных классов, маркируемых непосредственно внутри именной словоформы, или хотя бы застывших следов некогда продуктивной системы такого рода, характеризует почти все языковые семьи, которые были включены Дж. Гринбергом в состав нигер-кордофанской макросемьи. При этом Гринберг, строго исповедовавший принцип отделения типологических сходств от сегментных, в качестве одного из главных признаков нигер-кордофанского единства рассматривал не наличие именных классов *вообще*, а наличие *конкретной* и во многом поддающейся реконструкции системы клитик / аффиксов.

Сами по себе именные классы отмечены и за пределами нигер-кордофанской семьи — например, в «койсанских» языках (!хонг, жу|хоан и др., хотя в этих языках классовую принадлежность, как правило, нельзя предсказать по морфологическому оформлению имени — она проявляется эксплицитно только в глагольном согласовании). Для «нило-сахарских» языков продуктивные именные классы несвойственны, но в многих из них без труда выделяются «окаменевшие» или «псевдо-тематические» именные префиксы и суффиксы, рационально объяснить происхождение которых можно лишь исходя из гипотезы, что когда-то они были вполне продуктивными маркерами конкретных семантических категорий.

Из общей канвы при этом выбиваются афразийские языки, общей чертой которых являются скорее *родовые*, чем классовые противопоставления, т. е. оппозиция между (чаще всего) мужским и женским родом — которую, впрочем, саму по себе можно рассматривать как сильно упрощенный вариант системы классов. Любопытно, что система, во многом аналогичная афразийской (правда, с добавлением еще и третьего, «общего» рода), обнаруживается также на крайнем юге Африки — в центральнокойсанской семье языков.

Значимость классовых систем как потенциальных маркеров языкового родства, безусловно, высока, но переоценивать ее не следует. Даже если исключить из рассмотрения такие крайности, как ситуацию в уже упоминавшемся выше языке ма'а — с точки

зрения лексически ориентированного подхода, южнокушитского языка, заимствовавшего всю систему именных классов из банту — хронологический и ареальный факторы легко могут со временем привести к «естественной» перестройке старой системы или даже к полной ее замене. Неслучаен тот факт, что отдельные исследователи (например, Р. Бленч в ряде недавних публикаций) время от времени склоняются к мысли, что развитую систему именных классов следует считать скорее общей инновацией лишь для одного из «языковых блоков» нигер-конго (пусть даже и включающего в себя подавляющее большинство современных языков), в то время как пранигер-конголезская ситуация могла быть более сходной с наблюдаемой в современных языках догон или манде, где именные классы отсутствуют.

Нас, однако, в данной ситуации — на том этапе исследования, когда мы еще не готовы вплотную заниматься вопросами макросравнения — интересует скорее чисто практический аспект, а именно, (а) насколько важно и (б) насколько затруднительно использовать морфологические данные, в первую очередь данные по именовым основам, в ходе этимологического и лексикостатистического анализа материалов по языковым таксонам «низких» уровней. Здесь возможны два типа ситуаций:

а) В языках с *продуктивными*, «живыми» системами именных классов сам по себе морфологический анализ основы, как правило, не вызывает затруднений. В отдельных случаях можно (и даже нужно) ожидать морфологических переразложений (аналогичных классическому примеру переразложения в суахили арабизма *kitabu* 'книга' на ед. ч. *ki=tabu*, мн. ч. *vi=tabu*), опознать которые можно только путем внешнего сравнения. Сложнее всего приходится при работе с языками, в которых классные показатели (чаще всего префиксальные) демонстрируют нетривиальное фонетическое взаимодействие с корнем; хороший пример — языки «атлантической» семьи, где грамматические морфемы сливаются с корневыми, приводя к образованию грамматикализованных рядов консонантных чередований. Тем не менее, сколь бы запутанной ни оказалась конечная система, сама по себе ее продуктивность служит надежной гарантией того, что грамматические морфемы в конечном итоге можно будет уверенно «отрезать» от корневых.

Подчеркнем, что принципиальным моментом в нашем анализе является реконструкция как на низких, так и на высоких уровнях именно *корня*, а не полной *основы* лексемы (т. е. корня + классифицирующей грамматической морфемы или морфем), по той причине, что для обычной лексикостатистики значимость имеет только этимологическое совпадение сравниваемых корней: так, слову 'пепел' в экойдных языках эфутоп ($\dot{n}=t\dot{u}ŋ$), нде ($\dot{n}=t\dot{s}ŋ$) и этунг ($\dot{a}=t\dot{s}ŋ$) будет присвоен один и тот же индекс когнации, несмотря на заведомо бóльшую близость между полными основами этого слова в эфутоп и нде и бóльшую отдаленность от них варианта в этунг, что подразумевает переход слова из одного грамматического класса в другой либо в эфутоп-нде, либо в этунг.

При этом, однако, допустимо и во многих случаях даже желательно по возможности *учитывать* классифицирующие парадигматические характеристики анализируемых лексем — по той причине, что реконструируемость, пусть даже на очень «мелком» уровне, полной основы иногда можно использовать как существенный аргумент при разрешении спорных этимологических вопросов, от которого, в свою очередь, зависит и лексико-статистический результат.

б) В языках с *непродуктивными* «остатками» некогда продуктивной системы именной или глагольной аффиксации ситуация намного сложнее. *Заподозрить* следы такой системы, как правило, нетрудно — главным условием здесь также будет наличие у соответствующего языка или языковой группы достаточно близких «родственников», в соответствующих словах которых можно увидеть отличия, потенциально трактуемые как некогда продуктивные расхождения в грамматической маркировке корня (ср. в языках сара-бонго-багирми центральносуданской семьи: 'нос' = бака *sa-tə*, кара *ku-mi*, кенга *o-to*; 'птица' = бака *su-lu*, кара *si-li*, кенга *ye-le*; 'собака' = бака *i-si*, кара *bi-si* и др.). *Доказать* (хотя бы в «слабом» смысле слова), что речь в *каждом* из таких конкретных случаев идет не о случайных совпадениях, намного труднее.

Главная опасность, подстерегающая на этом пути исследователя — неоправданная *экстраполяция* относительно надежных результатов, полученных на определенном *срезе* материала, на *весь* материал. Поясним это на конкретном примере, хорошо известном большинству африканистов — т. н. «подвижном *k*»

или «*k-mobile*» («movable *k*» в терминологии Дж. Гринберга) в «нило-сахарских» языках.

Данная морфема, существование которой Гринберг уверенно постулирует на пранило-сахарском уровне и присутствие следов которой в отдельных языковых группах, согласно его критериям, является одним из диагностических признаков их нило-сахарского происхождения, самим Гринбергом определяется следующим образом: «This is a prefix which occurs, basically on nominals, in an apparently capricious manner» [Greenberg 1981: 105], и интерпретируется как т. н. «артикуль 3-й стадии» («stage III article»), т. е. действительный элемент, который на определенном этапе имел как конкретное значение (напр., «определенности»), так и способность более или менее свободно сочетаться с любыми именными основами, но впоследствии утерял либо продуктивность, либо первоначальную семантику и в одних языках остался в виде застывшего префикса, в других приобрел значение показателя сингулятива и т. п. При этом чаще всего «*k-mobile*» иллюстрируется на материале нилотских языков, где оно выделяется наиболее прозрачным образом (ср. один из примеров Гринберга: масаи *ki=ɾaŋ* 'крокодил' = нуэр *noŋ* id.).

Разумеется, в том, что в *каких-то* ветвях гипотетической нило-сахарской макросемьи в том или ином виде (продуктивном или застывшем) присутствуют «*k-образные*» элементы, сомневаться не приходится. Но при этом ни в специально посвященной «*k-mobile*» работе 1981 г., ни в каких-либо других работах Гринберга не был детально описан сценарий развития этого «артикуля 3-й стадии» в отдельных ветвях нило-сахарской семьи. Вместо этого была, по сути, предложена неправомерная экстраполяция: исходя из того, что в *некоторых* нило-сахарских языках наблюдается «хаотичное» поведение *k*- (например, в центральносуданской группе мору-мади, где, действительно, разноразной может иметь место между близкородственными языками), возможность такого поведения молчаливо предполагается для *любых* нило-сахарских языков. В целом это означает, что в любом слове любого нило-сахарского языка, начинающегося на *k*-, этот велярный согласный можно при желании отбросить и сравнивать с данными других языков любой из двух вариантов — «+*k*» или «-*k*».

Справедливости ради стоит подчеркнуть, что сам Гринберг соответствующим «правом» в своих нило-сахарских сопоставлениях пользовался относительно редко, хотя отдельные разительные примеры время от времени встречаются (например, сравнение нара *kele* 'голова' с нубийским *ur* и гаам *ol id*. [Greenberg 1966: 101], очевидным образом исходящее из презумпции, что *k-* в нара — застывший префикс, несмотря на то, что никаких внутренних оснований для этой презумпции нет). Однако в этимологическом словаре К. Эрета [Ehret 2001] «*k-mobile*» отрезается от того или иного корня уже едва ли не на каждой странице, что выглядит как откровенное злоупотребление концепцией «артикла 3-й стадии».

Отметим, что на самом деле нет даже твердой уверенности в том, что все до единого *реальные* случаи «*k-mobile*» (т. е. ситуации, в которых вариантность форм с и без *k-* можно показать достаточно наглядно) отражают одну и ту же морфему пранило-сахарского языка. В конкретных ветвях «нило-сахарских» языков префикс $k(V)=$ может выполнять абсолютно разные функции: например, помимо «сингулятивной», отмеченной Гринбергом (как в канури *at* 'люди' : $k=at$ 'человек'), в языках группы темейн он, наоборот, образует *множественное* число, в языках группы ныманг — адъективные основы и т. п. С точки зрения Гринберга, «сингулятивное» и «множественное» $k=$, конечно, считаются разными морфемами — но в этом случае непонятно и стремление во что бы то ни стало возвести все без исключения случаи «окаменевшего *k-mobile*» к отражениям единого пранило-сахарского «артикла». Это не говоря уже о том, что префикс $k(V)=$ хорошо известен также для чадских языков, заведомо не связанных с «нило-сахарскими» никакими отношениями, кроме ареальных — что наводит на мысль не столько о возможности заимствования соответствующей морфемы, сколько о параллельных тенденциях развития, при которых префикс $k=$ может возникнуть в нескольких дальнородственных или вообще не родственных семьях одного ареала из разных источников (в одних — из указательного местоимения, в других — из притяжательного, в третьих — из грамматикализованного существительного-классификатора и т. п.).

Задача исследователя в ситуации подобного рода чрезвычайно сложна. Здесь необходимо найти «золотую середину» между

двумя крайностями: с одной стороны, полным отказом от диахронического сегментирования основ, несегментируемых на синхронном уровне (что абсурдно, т. к. застывшие префиксы в африканских языках — сам по себе неоспоримый факт), с другой — произвольным сегментированием, главной целью которого будет являться наращивание (любой ценой!) сравнительного материала.

Представляется, что правильный выход из положения может быть обеспечен только максимальной *систематизацией* имеющегося материала и обнаружением регулярных (рекуррентных) корреляций, для которых можно предположить конкретную и типологически достоверную интерпретацию. Так, сам по себе факт сопоставления нара *kele* 'голова' с гаам *ol* 'голова' не может считаться убедительным свидетельством родства этих этимонов лишь потому, что для «нило-сахарских» языков *вообще* известны случаи «k-mobile». Но если, в силу совокупных аргументов более убедительного характера, с высокой вероятностью окажется, что нара и гаам — языки, состоящие в достаточно близком родстве, к данной квази-этимологии можно будет вернуться, при условии, что (а) будут обнаружены и другие случаи такой же корреляции, т. е. можно будет показать, что процесс «фоссилизации» $k=$ в нара не был ограничен одним словом; (б) для «пра-нара-гаам» k -можно будет хотя бы примерно предположить какую-то праязыковую семантику (например, классный префикс частей тела).

Иными словами, отношение к потенциально «окаменевшим» элементам в сравниваемых языках должно быть примерно таким же, как и к «нетривиальным» фонетическим соответствиям, т. е. их «окаменелость» должна подтверждаться статистикой и интерпретацией. Если для языков А и В обнаруживаются два ряда соответствий: (1) $k- : k-$; (2) $k- : \emptyset-$, то, в соответствии с классическим сравнительно-историческим методом, мы в первом случае, скорее всего, будем восстанавливать $*k-$, во втором — $*k_1-$. Было ли это $*k_1-$ отдельной фонемой (например, увулярным $q-$) или отдельной морфемой, пропавшей в языке В по морфологическим, а не фонетическим причинам — особый вопрос, ответить на который уже вряд ли удастся без выхода на внешний уровень сравнения.

Но самое главное не это, а то, что в *обоих* случаях соответствие (2) должно быть настоящим *соответствием*, т. е. быть определенным на *однородном* множестве, состоящем из более чем одной

пары примеров (конкретное количество должно определяться примерно, исходя из общего объема материала). Под однородностью множества подразумевается возможность единого исторического объяснения всех его членов — в случае с нара *kele*, например, это означает, что другие примеры «окаменевшего *k*» должны в этом языке также обнаруживаться в составе, по меньшей мере, существительных (а не, скажем, прилагательных или, еще хуже, глаголов), при этом желательно — в первичных именных основах (а не, скажем, в отглагольных образованиях), а еще лучше — в составе частей тела. Если такое единое объяснение отсутствует, рассматриваемый пример, будь он даже истинен в историческом плане, по сути неотличим от наблюдения случайного созвучия.

Таким образом, и в плане морфологической специфики языков Африки (особенно «макро-суданского» ареала) мы приходим к тем же выводам, к которым пришли выше в ходе дискуссии о лексических сравнениях: *массовость* сравнения ни в коем случае нельзя понимать как возможность произвольно-случайной, «удобной» для предвзятого исследователя *выборки* из «массы» языков — ее истинной целью должно быть построение *оптимального исторического сценария*, удовлетворительно объясняющего *максимальное* и оставляющее без достоверного объяснения *минимальное* количество языковых фактов. Концепция «артикла 3-й стадии» сама по себе не производит впечатления фантастической, но никоим образом не является и оптимальным сценарием, т. к. оставляет без ответа огромное количество конкретных вопросов — и на данный момент неизвестно, что произойдет, когда хотя бы на некоторые из них ответ все же будет найден: не исключено, что от гипотезы «единого» «подвижного *k*» в пранило-сахарском вообще придется отказаться.

1.8.5. *Выводы*. Перечисленные и кратко описанные выше проблемы, ожидающие исследователя при погружении в сферу африканского исторического языкознания, не являются исчерпывающими; затронуты были лишь наиболее общие и важные, на наш взгляд, аспекты, не упомянуть о которых перед тем, как приступить к конкретному анализу материала, было бы неправильно. Но даже из столь беглого рассмотрения можно извлечь некоторые выводы теоретического и методологического

характера, составляющие часть того фундамента, на котором будет в дальнейшем конструироваться здание африканской генетической классификации.

1) Никаких «общеафриканских» типологических черт или законов исторического развития, более или менее равномерно распределенных по всему континенту, по-видимому, не существует. На территории Африки правомерно выделять отдельные типологические макроареалы (не обязательно совпадающие с классификацией Т. Гюльдемманна), но дистанция, отделяющая, скажем, среднестатистический «койсанский» язык от среднестатистического «восточносуданского», оказывается ничуть не меньшей, чем дистанция между «восточносуданским» и любой другой языковой семьей мира, а, возможно, даже и большей.

2) С другой стороны, типологическое сходство между отдельными группами языков, сосуществующими в пределах одного макроареала, ни в коем случае не должно рассматриваться как аргумент в пользу их генетического родства: даже поверхностный анализ «горизонтальных» и «вертикальных» связей в пределах, скажем, «макро-суданского» ареала убедительно показывает, что в ситуации длительного, относительно стабильного сосуществования разных групп на общей территории вторичное сближение языков по их типологическим характеристикам — явление, естественное для всего африканского континента.

3) Колоссальное разнообразие языков Африки неизбежно влечет за собой желание подходить к решению вопроса об их классификации на основании *выборочного* принципа — например, опираясь исключительно на те языки, для которых существуют подробные современные описания и т. п. Однако опора на выборочный принцип почти неизбежно ведет к грубым ошибкам — таким, как неучет фактора независимых однонаправленных семантических инноваций, смешение контактных и генетических связей, и принятие фонетически инновативных форм за фонетически архаичные. Таким образом, выборочному принципу мы должны предпочесть *тотальный* учет данных, даже если это в несколько раз увеличивает объем необходимой работы.

4) Принимая в качестве рабочей модели ту или иную гипотезу об исторических связях отдельных групп или семей Африки, необходимо в первую очередь опираться на *системные*

данные. В «африканских условиях» это относится не только к фонетике сравниваемых лексических материалов, но и, в очень большой степени, к их морфологической структуре, произвольного сегментирования которой следует стараться по возможности избегать. Выделяя «окаменевшие» морфологические показатели, необходимо хотя бы как-то стараться очертить возможные причины и механизмы их «фоссилизации», затрагивающие отдельную *группу* примеров, а не изолированные формы, несводимые к единой модели. В отдаленной перспективе такой подход мог бы привести к реконструкции четких сценариев исторического развития классифицирующей морфологии языков Африки.

5) Наконец, учитывая поистине грандиозный объем сопоставительного материала, который подлежит обхвату *даже* при условии самоограничения в рамках базисной лексики (напомним, что, хотя непосредственным объектом сравнения служат только 100- или 50-словные списки Сводеша, для решения многочисленных спорных вопросов невозможно обойтись без подключения дополнительных этимологических данных), жизненно важно уметь отличить *первоочередные* для решения вопросы от *второстепенных*, ответы на которые могут быть получены уже после построения общей классификационной модели. Для языков африканского континента такими «второстепенными» по отношению к основной цели нашего исследования вопросами могут оказаться следующие:

а) подробная этимологическая проработка *всех* данных 100- или 50-словных списков. В частности, если дистрибуционный анализ материала показывает, что такое-то слово в таком-то языке, скорее всего, является инновацией (в «сводешевском значении»), это не означает, что на нас лежит обязательство *во что бы то ни стало* найти для этого слова внутреннюю (т. е. «генетическую») или внешнюю (т. е. контактную) этимологию. Конкретный опыт работы показывает, что если такая этимология не находится «легко и быстро» в рамках имеющихся данных, время, затраченное на дальнейшие поиски, вряд ли будет адекватно пропорционально убедительности «сомнительной» этимологизации;

б) тщательная реконструкция всех фонологических деталей каждого из этимонов, проецируемых на тот или иной уровень. В частности, не требует доказательств тот факт, что

вокалические и просодические характеристики основ во многих языках Африки гораздо менее устойчивы, чем их консонантный состав. Особенно ненадежными могут быть такие признаки, как +/-ATR и тональная артикуляция в системах с противопоставлением трех (высокий, средний, низкий) и более регистров — здесь можно столкнуться не только с повышенной сложностью соответствий, но и, что намного хуже, с низким качеством записи материала (даже в описаниях, выполненных на более или менее современном уровне). На наш взгляд, «недопроработанность» такого рода аспектов в предварительных реконструкциях не может являться серьезной претензией к выводам, полученным на основании таких реконструкций;

в) детальное исследование морфосемантики как синхронных, так и реконструированных классифицирующих грамматических элементов. Даже несмотря на то, что, как говорилось выше, для каждого восстанавливаемого форманта *желательно* хотя бы примерное представление о выполнявшейся им функции, предсказуемым и неизбежным будет присутствие в праязыковых системах и таких морфем, чья семантика окажется «пустой» — аналогично соответствующим морфемам, надежно засвидетельствованным и в современных языках.

Останавливаться на каждом из этих технических аспектов подробно в рамках данного раздела вряд ли имеет смысл, т. к. все они в дальнейшем будут возникать уже в ходе решения конкретных задач анализа лексического материала. Подчеркнем, пожалуй, лишь тот момент, что стопроцентно универсальных решений, приемлемых в любой ситуации, наше исследование предложить пока что не в состоянии. Вполне допустимо, что какой-то аспект, нерелевантный или сугубо второстепенный для одной или даже большинства разбираемых языковых групп, окажется, напротив, решающим хотя бы в минимальной группе ситуаций. Например, в общем и в целом скорее верно, что тональная просодия не играет столь же ключевую роль в процессе составления этимологического корпуса, как консонантный инвентарь. Но в отдельных случаях — например, в языках с преимущественно открыто-моносиллабической структурой корня (CV) и относительно простым консонантизмом — на тональные оппозиции, несомненно, следует обращать гораздо большее

внимание, чем на аналогичные оппозиции в языках со структурой корня SVC, т. к. в этой ситуации активно развивается «сегментная омонимия», помочь различить которую могут только просодические признаки.

Тем не менее, намного более важной, чем нахождение универсального решения для каждой из разбираемых проблем, нам представляется максимально детальная и аргументированная *экспликация* принимаемых решений. Простое перечисление материала, сопровождаемое предварительными реконструкциями, позволяет сэкономить на объемах выполняемой работы (и печатного текста), но очень часто не позволяет постигнуть внутреннюю логику исследователя и оценить общее качество аргументации. В ситуации, когда упор делается не на «подавление» читателя колоссальным объемом материала, а на разбор относительно небольшого числа «базисных» этимологий, подробный комментарий к каждой из них должен являться неотъемлемой частью работы. Этого принципа мы, по мере сил, будем стараться придерживаться на всем протяжении исследования.

Приложение 1. Унифицированная система транскрипции.

В рамках исследования «Лексикостатистическая классификация языков Африки» приходится иметь дело с большим количеством работ дескриптивного характера, чрезвычайно разнообразных как по времени написания, так и по степени адекватного соответствия принятых в них систем фонетической нотации реальному языковому материалу. Соответственно, при цитировании материала можно придерживаться одного из двух подходов: (а) консервативного — все цитируемые формы подаются ровно в том виде, в котором они приводятся в оригинальном источнике, с приведением необходимых транскрипционных пояснений для каждого из этих источников; (б) унифицирующего — строгой транслитерации всех материалов согласно единой системе транскрипции, пригодной для всех анализируемых языков.

Консервативный подход, по здравом размышлении, приходится в данной ситуации признать неприемлемым. Даже для близкородственных языков и диалектов, фонетические различия между которыми незначительны, часто приходится сводить вместе данные из источников, использующих различные и (нередко) идиосинкратические системы транскрипции. Такие системы мотивируются иногда личными творческими инициативами, не получившими в дальнейшем широкого признания (например, сложная система записи щелчковых согласных в работе [Doke 1925]), иногда — стремлением внедрить их в качестве официальной письменности для записываемого языка, в результате чего «сложные» фонемы записываются непредсказуемыми комбинациями латинских букв во избежание «нестандартных» символов.

В связи с этим точная передача всех исходных транскрипционных систем, так или иначе задействованных в данном исследовании, во-первых, привела бы к чрезмерному разрастанию и без того объемного труда (по существу, пришлось бы разработать своего рода «транскрипционный конкорданс» — задача, безусловно, полезная, но не имеющая непосредственного отношения к целям исследования); во-вторых, причинила бы серьезные технические неудобства тем читателям, которые не обладают

специализированными знаниями одновременно по нескольким языковым семьям Африки, не говоря уже о не-африканистах, для которых данная работа тоже может представлять теоретический интерес.

Таким образом, для подачи материала в нашем исследовании используется унифицирующий принцип: транскрипции и орфографии разных источников настолько строго, насколько это возможно, транслитерированы в единую систему. Отдельные исключения сделаны для тех старых записей, которые не сопровождаются подробными фонетическими комментариями: в таких источниках точно определить фонетическую характеристику той или иной буквы, диграфа или диакритики часто бывает невозможно. Однако, поскольку данные из таких источников редко носят определяющий характер при разработке предварительных фонетических реконструкций праязыковых корней (их учет скорее имеет значение для выяснения особенностей дистрибуции того или иного этимона в рамках языковой семьи), в этих случаях позволительно приводить оригинальную транслитерацию без специальных пояснений относительно фонетической интерпретации знаков, чтобы не превращать все исследование в детальный филологический комментарий к старым описательным работам.

Принятая в работе унифицированная система транскрипции (УСТ) в целом совпадает с соответствующей системой, разработанной автором совместно с А. С. Касьяном и М. А. Живловым для Интернет-проекта «Глобальная лексикостатистическая база данных», т. е. призвана обслуживать нужды не только языков Африки, но и всего языкового разнообразия планеты. В ее основе лежит широко распространенная в мировой практике система Международного фонетического алфавита, но с внесением отдельных технических изменений. В частности, практикуется: (а) отказ от использования диграфов для записи аффрикат (т. е. *c*, *z*, *č*, *ž* вм. *ts*, *dz*, *tʃ*, *dʒ* и т. п.); (б) симметричный подход к записи рядов аффрикат и спирантов (включая использование единого «хвоста» для обозначения палатальной или альвео-палатальной артикуляции: *ç*, *ʒ*, *ʃ*, *ʒ̥* и т. п.); (в) отказ от двусмысленного знака *j* в пользу записи палатального глайда как *y*; (г) использование традиционных «умянутных» букв *ï*, *ö*, *ï* вместо МФА *æ*, *ø*, *y* соответственно.

При унификации транскрипции мы не ставили задачу последовательно придерживаться либо строго фонологического, либо строго фонетического принципа записи, т. к. достижение абсолютной точности по этому параметру было бы, во-первых, невозможным (во многих использованных источниках отсутствуют адекватные фонетические описания), во-вторых, излишним с точки зрения общей цели исследования. В целом наша транскрипция, вслед за большинством систем, разработанных африканистами, тяготеет скорее к фонетическому принципу, но степень ее детализации напрямую зависит от источника.

Ниже приводится полный перечень унифицированных транскрипционных знаков, используемых в работе «Лексикостатистическая классификация языков Африки», сопровождаемый краткими комментариями.

А. Согласные

«Обычные» (т. е. нещелчковые):

b	звонкий лабиальный взрывной
b^h	звонкий придыхательный лабиальный взрывной
β	имплозивный лабиальный взрывной
c	глухая свистящая (альвеолярная) аффриката
cʔ	(пре)глоттализованная свистящая аффриката
c^h	придыхательная свистящая аффриката
č	глухая шипящая (пост-альвеолярная) аффриката
čʔ	(пре)глоттализованная шипящая аффриката
č^h	придыхательная шипящая аффриката
ɕ	глухая (альвео-)палатальная аффриката
ɕʔ	глоттализованная палатальная аффриката
ɕ^h	придыхательная (альвео-)палатальная аффриката
d	звонкий денальный/альвеолярный взрывной
d^h	звонкий придыхательный денальный взрывной
ɖ	имплозивный денальный/альвеолярный взрывной
ɗ	звонкий ретрофлексный взрывной
ɗ	звонкий денальный взрывной
ɗ	имплозивный денальный взрывной
ð	звонкий интерденальный спиронт

f	глухой лабиодентальный спирант
g	звонкий велярный взрывной
g ^h	звонкий придыхательный велярный взрывной
ǧ	имплозивный велярный взрывной
ɠ	звонкий увулярный взрывной
ɣ	звонкий велярный спирант
h	глухой ларингальный спирант
k	глухой велярный взрывной
k ^h	глухой велярный придыхательный взрывной
kʷ	глоттализированный велярный взрывной
k ^w	глухой лабиовелярный взрывной
l	латеральный сонорный
ɮ	глухой латеральный фрикативный
ɬ	глухая латеральная аффриката
ɬʷ	глоттализированная латеральная аффриката
ɬ ^h	придыхательная латеральная аффриката
ɮ	звонкая латеральная аффриката
m	лабиальный носовой сонорный
n	дентальный носовой сонорный
ɲ	палатальный носовой сонорный
ɳ	велярный носовой сонорный
p	глухой лабиальный взрывной
pʷ	глоттализированный лабиальный взрывной
p ^h	глухой лабиальный придыхательный взрывной
q	глухой увулярный взрывной
qʷ	глоттализированный увулярный взрывной
q ^h	глухой увулярный придыхательный взрывной
r	альвеолярный вибрант
ɾ	альвеолярный одноударный
ɽ	ретрофлексный одноударный
s	глухой свистящий спирант
s ^h	придыхательный свистящий спирант
ʃ	глухой шипящий спирант
ʂ	глухой (альвео-)палатальный спирант
t	глухой дентальный/альвеолярный взрывной ¹

¹ Конкретная артикуляция согласных *t*, *d* в африканских языках обычно варьирует от дентальной до альвеолярной, но фонологическое

t ^h	глухой дентальный придыхательный взрывной
tʰ	глоттализированный дентальный взрывной
t̥	глухой ретрофлексный взрывной
t̄	глухой дентальный взрывной
θ	глухой интердентальный спирант
v	звонкий дентолабиальный спирант
w	лабиальный глайд
x	глухой велярный спирант
χ	глухой увулярный спирант
y	среднеязычный глайд
z	звонкий свистящий (альвеолярный) спирант
ʒ	звонкий шипящий (пост-альвеолярный) спирант
ʒ	звонкая свистящая (альвеолярная) аффриката
ʒ ^h	звонкая свистящая придыхательная аффриката
ʒʰ	свистящая глоттализированная аффриката с предозвончением
ʒ̣	звонкая шипящая (пост-альвеолярная) аффриката
ʒ̣ ^h	звонкая шипящая придыхательная аффриката
ʒ̣ʰ	шипящая глоттализированная аффриката с предозвончением
ʒ̣	звонкая (альвео-)палатальная аффриката
ʒ̣̣	импловивная (альвео-)палатальная аффриката
ʔ	глухой ларингальный взрыв (гортанная смычка)

противопоставление этих двух рядов встречается очень редко и к тому же во многих случаях может рассматриваться не столько как оппозиция по месту образования, сколько по способу (альвеолярные взрывные *t, d* : (интер)дентальные спиранты *θ, ð*). В тех немногочисленных ситуациях, когда фонологическую оппозицию действительно образуют альвеолярные и дентальные взрывные (нилотские и сурмийские языки), мы даем для первых транскрипцию *t, d*, для вторых — *t̄, ð̄*. Там, где подобная оппозиция отсутствует, во избежание нагромождения лишних диакритик мы всегда записываем *t, d*, вне зависимости от их фонетической характеристики.

Щелчковые («кликсы»)¹:

²	дентальный щелчковый (без доп. признаков)
ǀ	палатальный щелчковый (без доп. признаков)
!	альвеолярный щелчковый (без доп. признаков)
	латеральный щелчковый (без доп. признаков)
θ	лабиальный щелчковый (без доп. признаков)
!!	ретрофлексный щелчковый (без доп. признаков)
, ǀ...	щелчковый с озвончением
ǁ, ǂ...	назализованный щелчковый
ʔ, ǀʔ...	щелчковый с исходом на гортанный взрыв
^h , ǀ ^h ...	щелчковый с аспирацией
^h , ǀ ^h ...	щелчковый с озвончением и аспирацией
ǁ ^h , ǂ ^h ...	назализованный щелчковый с аспирацией
ʔ ^h , ǀʔ ^h ...	щелчковый с исходом на гортанный взрыв и аспирацией
x, ǀx...	щелчковый с исходом на велярный фрикативный
x, ǀx...	щелчковый с озвончением и исходом на вел. фрикативный
ǀkx, ǀkx...	щелчковый с исходом на велярную аффрикату
q, ǀq...	щелчковый с исходом на увулярный смычный

¹ Фонетическая терминология, принятая в нашей работе для обозначения типов щелчковых согласных, на сегодняшний день считается общепринятой (за исключением разве что обозначения особого артикуляторного типа, известного только для северо-койсанской группы, как «ретрофлексного»). При этом в работах по койсанистике, написанных до 1980-х — 1990-х гг., нередко встречаются терминологии, существенно отличающиеся не только от современной, но и от одного исследователя к другому. Неспециалисту при сопоставлении современных работ (включая наше исследование) со старыми источниками рекомендуется ориентироваться на конверсионную таблицу в [Ladefoged & Traill 1994: 35].

² В старых работах по койсанским языкам кликсы с исходом на обычную велярную смычку принято записывать как сочетания /k, ǀk и т. д., а с исходом на гортанный взрыв — как просто /, ǀ и т. д. В более новых работах за немаркированный исход принимается, наоборот, именно велярная смычка, так что в фонологически ориентированных системах транскрипции для этих кликсов принята запись /, ǀ, а в фонетически ориентированных — /k, ǀk; кликсы с исходом на гортанный взрыв во всех системах записываются как /ʔ, ǀʔ. Мы пользуемся фонологически ориентированной системой нотации, унифицированно заменяя транскрипцию /k, ǀk в примерах из старых работ на /, ǀ.

q ^h , t̪q ^h ...	шелчковый с исходом на увулярный и аспирацией
qʔ, t̪qʔ...	шелчковый с исходом на глоттализованный увулярный

Б. Гласные

л	неогубленный среднего ряда нижнего подъема, [+ ATR]
а	неогубленный среднего ряда нижнего подъема, [- ATR]
е	неогубленный заднего ряда нижнего подъема
е	неогубленный переднего ряда среднего подъема, [+ ATR]
ε	неогубленный переднего ряда среднего подъема, [- ATR]
і	неогубленный переднего ряда верхнего подъема, [+ ATR]
г	неогубленный переднего ряда верхнего подъема, [- ATR]
о	огубленный заднего ряда среднего подъема, [+ ATR]
ö	огубленный переднего ряда среднего подъема
ɔ	огубленный заднего ряда среднего подъема, [- ATR]
u	огубленный заднего ряда верхнего подъема, [+ ATR]
ü	огубленный переднего ряда верхнего подъема
u	огубленный заднего ряда верхнего подъема, [- ATR]
ш	неогубленный заднего ряда верхнего подъема
ʁ	неогубленный заднего ряда среднего подъема

В. Диакритические знаки

~	фарингализованная артикуляция гласного
–	слоговой (ударный) сонант (<i>и, т</i> и т. п.)
~	назализация (гласного, глайда)
˝	сверхвысокий тон
˘	высокий тон
ˉ	средний тон
˘	низкий тон
˝	сверхнизкий тон
˘	нисходяще-восходящий (контурный) тон
˘	восходяще-нисходящий (контурный) тон
h	придыхательная артикуляция гласного
:	долгота (гласного)

Г. Прочие условные обозначения

- Отделяет корневую морфему от суффиксальной
- = Отделяет корневую морфему от префиксальной
- «развивается в...», «дает рефлекс...»
- ← «развивается из...», «является рефлексом...»
- ~ Отделяет фонетические или морфологические варианты
- || Отделяет комментарий от перечисления форм

**Приложение 2. Сокращения названий языков и другие
аббревиатуры.**

<u>анд</u>	г анда [центральнокойсанская группа]
<u>ани</u>	ани [центральнокойсанская группа]
<u>ауе</u>	ау ен [севернокойсанская группа]
<u>аун</u>	ауни [южнокойсанская группа]
<u>буг</u>	буга [центральнокойсанская группа]
<u>ган</u>	г ана [центральнокойсанская группа]
<u>гви</u>	г ви [центральнокойсанская группа]
<u>дан</u>	даниси [центральнокойсанская группа]
<u>дет</u>	дети [центральнокойсанская группа]
<u>жцх</u>	жу хоан [севернокойсанская группа]
<u>квд</u>	квади [язык-изолят]
<u>кнг</u>	!кунг [севернокойсанская группа]
<u>кор</u>	!ора ~ корана [центральнокойсанская группа]
<u>куа</u>	куа [центральнокойсанская группа]
<u>кхо</u>	кхве [центральнокойсанская группа]
<u>мас</u>	масарва [южнокойсанская группа]
<u>нам</u>	нама [центральнокойсанская группа]
<u>нар</u>	наро [центральнокойсанская группа]
<u>нгк</u>	нг!ке [южнокойсанская группа]
<u>нуе</u>	н у ен [южнокойсанская группа]
<u>нху</u>	н уки [южнокойсанская группа]
<u>окн</u>	!о!кунг [севернокойсанская группа]
<u>сан</u>	сандаве [язык-изолят]
<u>тсх</u>	тсиха [центральнокойсанская группа]
<u>хаб</u>	‡хаба [центральнокойсанская группа]
<u>хад</u>	хадза [язык-изолят]
<u>хай</u>	хайсе [центральнокойсанская группа]
<u>хам</u>	хам [южнокойсанская группа]
<u>хас</u>	хаси [южнокойсанская группа]
<u>хег</u>	хегви [южнокойсанская группа]
<u>хнг</u>	!хонг [южнокойсанская группа]
<u>хоа</u>	‡хоан [язык-изолят]
<u>цуа</u>	цуа [центральнокойсанская группа]

<u>чар</u>	чара [центральнокойсанская группа]
<u>чуа</u>	чуа [центральнокойсанская группа]
<u>эко</u>	экока-!кунг [севернокойсанская группа]

Аббревиатуры имен собственных (в 50-словных списках)

Alm.	Almeida, Antonio de [Güldemann & Elderkin 2003]
B.	Bleek, Dorothea / Wilhelm [1929, 1956]
Col.	Collins, Chris [1998, 2001a, 2001b, 2001c; Bell & Collins 2001]
Dm.	Dempwolff, Otto [1916]
Ea.	Eaton, Helen [2002]
	Eaton, H. & Hunziker, D. & E. [2007, 2008]
Eld.	Elderkin, Derek [1983, 1998]
Gr.	Gruber, Jeffrey [1973, 1975];
	[также в Honken 1977, Heine & Honken 2010]
Ka.	Kagaya, Ryohei [1993]
LH	Lanham, Leonard & Hallows, D. P. [1956a, 1956b]
Sd.	Sands, Bonny [1998a, 1998b]
T.	Traill, Anthony [1974, 1994a]
Ta.	Tanaka, Jiro [1978]
TBW	Tucker, Bryan, & Woodburn [1977]
Vi.	Visser, Hessel [2001]
W.	Westphal, E. O. J. [1965, 1971]
Wu.	Wuras, C. F. [1920b]
Zr.	Ziervogel, D. [1955]

Приложение 3. Семантический комментарий к 50-словному списку.

Данное приложение задумано как своеобразный методический справочник, цель которого — дать, насколько это вообще возможно, максимально точную информацию о семантических границах и свойствах тех значений, которые составляют 50-словное подмножество 100-словного списка Сводеша (напомним, что отобрано оно было в первую очередь в соответствии с эмпирическим критерием «максимальной стабильности», хотя отдельные элементы все же пришлось заменить на менее стабильные в целях обеспечения большего разнообразия значений и большего «технического» удобства, см. 1.7.1).

В целом «справочник» основан на разработках, выполненных коллективно в рамках ряда заседаний Ностратического семинара и опубликованных в виде коллективной статьи [Kassian et al. 2010]. Тем не менее, ввиду особой важности этой темы для нашего исследования, а также для удобства читателя, представляется необходимым включить соответствующую информацию в том числе и в состав данной монографии, тем более, что по сравнению с опубликованной статьей некоторые комментарии будут изложены более подробно. Еще одна существенная инновация — подключение диахронических семантических сведений, т. е. информации о том, какие семантические развития для соответствующих элементов списка можно было бы считать «типичными».

Структура каждого из 50 вхождений «справочника» единообразна и состоит из следующих частей:

1. «краткое / общее значение слова», представляемое в виде лексического эквивалента на английском и русском языках¹;

¹ В связи с тем, что основным языком интерфейса компьютерной среды СтарЛинг является английский, все 100- и 50-словные списки в базах данных СтарЛинг индексируются по английскому алфавиту, и в этой же последовательности их элементы обрабатываются и в нашем исследовании. Этим объясняется кажущаяся хаотичность элементов в 50-словных списках, подробно анализируемых в основной части работы. (Альтернативный вариант — оставить в каждом вхождении исходный английский эквивалент слова — в русскоязычной монографии смотрелся бы довольно странно).

2. *уточнение* значения для тех случаев, когда ни русский, ни английский языки вне синтаксического контекста не позволяют понять слово однозначно. Подчеркнем, что в наши задачи *не входит* построение для каждого элемента списка детального семантического толкования (или выбор из ранее предложенных толкований). Это не только исключительно тонкая и сложная работа, но и чреватая полным уходом в проблемы, не имеющие первоочередной релевантности для нашего исследования — весь опыт многолетней работы со 100-словными списками приводит нас к твердой убежденности в том, что в самом общем случае «огрубленной» семантики, основанной на анализе лексической пересекаемости между разными языками, оказывается вполне достаточно для лексикостатистических целей;

3. *два-три синтаксических контекста* (по-русски), наглядно и диагностично иллюстрирующих требуемую семантику (такие контексты могут быть чрезвычайно полезны как для полевой работы с информантом, так и при сопоставлении их с текстовыми корпусами или словарными примерами, в целях выбора оптимального квази-синонима);

4. *синхронный комментарий* — информация о том, с какими трудностями можно столкнуться при заполнении данной позиции в списке и о возможных путях их преодоления, а также о типичных «квази-синонимах», которые подлежат отсеиванию;

5. *диахронический комментарий* — информация о типовых семантических коннотациях, характерных для данного элемента списка и способных быть по отношению к его значению *семантическими предками* (например, значение 'нога' по отношению к значению 'стоять'), *семантическими потомками* (например, значение 'быть' по отношению к значению 'стоять'), или просто *семантическими родственниками* (когда изменение значения может идти в обоих направлениях — например, 'голова' и 'верхушка').

Диахроническая информация извлекается в основном из 100-словных списков по языкам Евразии и Африки (в сильно меньшей степени — Америки и Тихоокеанского региона), а также этимологических баз данных, составленных участниками международного проекта «Эволюция языка» (Москва — Санта-Фе) в 2001-2011 гг., но учитывает при этом и данные различных этимологических словарей по языкам мира (индоевропейским,

уральским, алтайским, дравидийским, афразийским и т. д.), опубликованных признанными специалистами в этих областях.

Подробное обоснование того, почему тот или иной семантический переход может быть признан «типовым», само по себе было бы отдельной монографией, т. к. должно сопровождаться большим иллюстративным материалом. К тому же приводимая ниже информация ни в коей мере не может считаться исчерпывающей; ее скорее следует воспринимать как предварительные наброски к построению строго формальной системы. Мы ограничиваемся перечислением лишь таких переходов, которые в наибольшей степени бросаются в глаза при сопоставительном изучении исторической семантики базисной лексики.

1.1. **ashes** / **пепел**. 1.2. 'Пепел' как в первую очередь продукт сгорания наиболее «базовых» материалов — дерева, сухой травы и т. п. — но также, по возможности, и любых других. 1.3. *Ветер развеял пепел по окрестностям; взять горсть пепла из потухшего костра*. 1.4. Следует отличать от 'углей' (*кусков* сгоревшего материала, в отличие от пепла как порошкообразной субстанции), а также от специфически маркированных видов пепла, напр., 'горячего пепла', 'пепла, образованного от сжигания одного конкретного объекта (травы, навоза и т. п.)'¹. 1.5. Тип. сем. предки: именные основы 'пыль', 'грязь'; адъективная или глагольная основа 'серый'; реже встречаются производные образования от слова 'огонь' (хотя непосредственное развитие 'огонь' → 'пепел' маловероятно).

2.1. **bird** / **птица**. 2.2. 'Птица' как родовой термин, по возможности противопоставленный 'зверю' (т. е. 'класс летающих животных', в который вполне могут попадать также, например, летучие мыши, противопоставленный 'классу бегающих живот-

¹ Для русского языка проблематичным является выбор между 'пеплом' и 'золой', имеющими чрезвычайно близкие, но все же не полностью тождественные, значения. *Contra* [Kassian et al. 2010], где в качестве первичного эквивалента для русского языка предложена 'зола', мы все же склонны считать последнее слово более узко специализированным (употребляется применительно *только* к продукту сгорания дерева, в отличие от 'пепла', который может употребляться применительно и к дереву, и ко всему остальному: ср. *восстать из пепла* = англ. *rise from the ashes*, но никак не **восстать из золы*).

ных). 2.3. *Птицы летают по небу, звери бегают по лесу; высоко в небе летит какая-то птица.* 2.4. При наличии нескольких родовых понятий для птицы — чаще всего встречается 'большая птица' vs. 'маленькая птица' — следует по возможности ориентироваться на более статистически частотное (во избежание попадания в список чего-л. вроде русского 'пичужки') и более явно противопоставленное 'зверю', но если данных не хватает, допустимо использовать оба термина в качестве синонимов. Разумеется, категорически недопустимо использовать вм. 'птицы (вообще)' термины, обозначающие конкретных птиц (за исключением случаев полисемии). 2.5. Тип. сем. предки: (а) названия отдельных конкретных птиц, «олицетворяющих» весь класс птиц (так сказать, «птица *par excellence*») — например, 'орел', 'стервятник', реже 'воробей' и т. п.; (б) значения, связанные с типичными признаками или функциями птицы — чаще всего 'перо' ('птица' ← 'нечто пернатое') или 'летать' ('птица' ← 'нечто летающее'). Тип. сем. потомки: встречается и обратное сужение общеродового значения до обозначения конкретных видов птиц (см. выше).

3.1. **black / черный.** 3.2. Цвет RGB 0:0:0; типичен для таких объектов, как 'уголь', 'ночь', раскраска или оперение отдельных животных и птиц ('ворон', 'пантера' и т. п.); желательное участие в идиоматических противопоставлениях, основанных на оппозиции 'черный' : 'белый'. 3.3. *Это ворон, у него черные перья; снег белый (молоко белое), а уголь черный.* 3.4. Следует отличать от различных специально маркированных «оттенков» черного (обычно гораздо менее частотных в речевом употреблении). Отдельная проблема — существование во многих языках особой адъективной (или глагольной) основы 'темный' ('dark'), которая во многих контекстах взаимозаменяема с 'черным' и из-за этого иногда может ошибочно переводиться как 'черный'. При извлечении материала из ранее составленных кратких списков (а не подробных словарей) следует всегда учитывать возможность такой ошибки. 3.5. Тип. сем. предки: (а) «типично черные» объекты — 'уголь'; 'ночь'; 'черная птица'; реже 'земля', 'грязь'; (б) другие цветовые обозначения, относящиеся к «темному» спектру — прежде всего, собственно 'темный', но также 'серый', намного реже '(темно)-синий', '(темно)-зеленый' и т. п. Тип. сем. потомки: все эти семантические изменения, судя по всему, могут идти в обоих направлениях, т. е. от

адъективной основы 'черный' может образовываться именная основа 'ночь' и т. п.

4.1. **blood** / **кровь**. 4.2. Как и в других случаях, касающихся анатомических терминов, имеется в виду *кровь человека* (актуально для языков, регулярно противопоставляющих человеческую и животную кровь). 4.3. *Если порезаться, потечет кровь; кровь имеет красный цвет*. 4.4. Во многих языках семантическое поле 'кровь' чрезвычайно разнообразно: имеются отдельные лексемы для 'свернувшейся крови', 'менструальной крови', 'артериальной' или 'венозной крови', 'сгустка крови' и т. п.; тем не менее, нам не известен ни один язык, в котором наряду со всеми этими специальными терминами не существовало бы и общеродового (как правило, наиболее статистически частотного) понятия; базисным является оно и только оно. 4.5. Тип. сем. предки: (а) адъективная основа 'красный' (как основной характеризующий признак крови); (б) намного реже — названия отдельных жидкостей, с метафорическим развитием ('жир (жидкий)'; 'сок (дерева)' и т. п.); (в) метонимическое развитие из '(кровяного) сырого мяса'. Тип. сем. потомки: 'красный' (← 'цвета крови'); абстрактные именные основы со значением 'душа', 'жизненная сила' и т. п. (также метонимия: 'кровь' как «разносчик жизни» по всему организму).

5.1. **bone** / **кость**. 5.2. Общий термин для обозначения (по возможности) любой конкретной кости как человека, так и животного. 5.3. *Подавиться костью; бросить собаке кость; отделить мясо от костей*. 5.4. Следует строго отличать от обозначений конкретных костей (для многих языков характерен богатый инвентарь этимологически разных основ, соответствующих разным типам костей), а также от коллективного термина 'множество костей', т. е. 'скелет'. 5.5. Тип. сем. предки: виды конкретных костей ('локтевая', 'берцовая' и т. п.), а также общие термины для обозначения конечностей ('рука-arm', 'нога-leg'). Тип. сем. потомки: те же значения (развитие может идти в обоих направлениях).

6.1. **claw (nail)** / **ноготь**. 6.2. Англ. термин *claw* 'коготь' сохраняется лишь в силу традиции, т. к. именно это слово фигурирует в исходном списке Сводеша; на практике эта позиция почти всегда заполняется термином для обозначения ногтей на руках и ногах *человека*, т. е. англ. *nail (fingernail, toenail)*. 6.3. *На каждом пальце у человека есть ноготь; сломать ноготь на*

большом пальце. 6.4. Во многих языках мира 'коготь' (животного или птицы) и 'ноготь' (человека) лексически не различаются (иногда этот же термин употребляется и для выражения значения 'копыто'), но там, где лексическое различие есть, всегда следует выбирать именно 'ноготь' (руки). 6.5. Тип. сем. предки: слово может являться именным производным от глагольных основ 'царапать', 'скрести' и т. п. Прочие семантические связи, как правило, либо спорны, либо редки.

7.1. **die** / **умирать**. 7.2. Нейтральный, немаркированный термин (т. е. именно 'умирать', а не 'скончаться', 'преставиться' или, наоборот, 'подохнуть'). 7.3. *Он умер год назад; человек живет и умирает*. 7.4. В случае супплетивного противопоставления основ ед. и мн. ч. ('умирать (об одном человеке)' / 'умирать (о многих)') по умолчанию выбирается сингулятивная форма. Отметим, что в языках с развитой категории вежливости при выборе стилистически нейтрального синонима не всегда помогает опора на статистическую частотность — здесь следует всегда обращать внимание на прагматический контекст. 7.5. Тип. сем. предки: значение 'умирать' чаще всего развивается «эвфемистическим» путем из таких значений, как 'исчезать, пропадать', 'не быть, не существовать', 'ухудшаться, портиться, болеть'. Тип. сем. потомки: нередки ситуации, когда глагольная основа сама по себе исчезает из языка, оставляя при этом морфологические производные — 'смерть', 'покойник', 'труп'. Широко распространено также морфологическое производное 'убивать' (обычно — в языках с продуктивными каузативными моделями), см. ниже.

8.1. **dog** / **собака**. 8.2. Наиболее общий и нейтральный термин для обозначения домашней собаки. 8.3. *Собака живет с человеком; не забудь покормить собаку перед уходом*. 8.4. В языках, где существует отдельный общий термин для 'охотничьей собаки' (наподобие англ. *hound*), все равно требуется выбирать более общий эквивалент, обозначающий собаку как вид животного (т. е. и охотничью, и сторожевую, и собаку как «pet»). 8.5. Тип. сем. предки: практически отсутствуют. Иногда встречается перенос значения с более узкоспециализированных терминов, таких, как 'щенок', 'сука'. Довольно частотны случаи отсутствия этимологизации термина, что говорит о его «бродячем» характере в некоторых лингвогеографических ареалах. Во многих ситуациях

бросается в глаза звукоподражательный характер основы, хотя эксплицитно зафиксировать случай недавнего образования стилистически нейтрального термина 'собака' на фоносимволической основе пока не удастся. Тип. сем. потомки: как правило, речь может опять-таки идти об узкой специализации — сохранении старой основы в значениях 'охотничья собака'; 'сука'.

9.1. **drink** / **пить**. 9.2. Стилистически нейтральный глагол, обозначающий поглощение жидкости (или жидкой пищи; в качестве типичных объектов, помимо обязательно приемлемых 'воды', 'молока', 'вина' и т. п., допустимы также 'суп', 'каша'). 9.3. Он *пьет воду; чтобы жить, нужно есть и пить*. 9.4. Как обычно, следует отличать от более сложных значений с дополнительными семантическими компонентами — 'глотать', 'хлебать', 'пить мелкими / большими глотками' и т. п. 9.5. Тип. сем. предки: 'глотать', 'сосать' и др. «специализированные» способы принятия внутрь жидкости. Реже встречаются ситуации, в которых для значений 'есть' (твердую и/или жидкую пищу) и 'пить' (жидкость) существует единый глагол, который впоследствии может получить узкую специализацию исключительно в значении 'пить'. Вопреки популярного среди «глобальных этимологов» мнения об очевидности семантической связи между глаголом 'пить' и существительным 'вода', такая связь не является статистически частой на общемировом уровне, но может быть типичной для отдельных ареалов (например, «арктического», где она наблюдается и в чукотско-камчатских, и в эскимосско-алеутских языках). Тип. сем. потомки: 'напиваться, быть пьяным'; помимо этого, некоторые из перечисленных выше семантических переходов могут идти в обоих направлениях (по крайней мере, 'пить' → 'глотать').

10.1. **dry** / **сухой**. 10.2. Это — либо чисто адъективная, либо адъективно-глагольная основа (т. е. в тех случаях, когда лексемы 'быть сухим' и 'сухой' выражаются разными корнями, выбирается вторая, хотя на самом деле такие ситуации чрезвычайно редки). Значение 'сухой' может иметь различные смысловые оттенки и коннотации: требуемый вариант можно истолковать скорее как 'не содержащий *излишней* влаги', т. е. применительно, например, к одежде, волосам, коже и т. п. (после просушки или, наоборот, перед намоканием), но не как 'лишенный *положенного* количества влаги', т. е. не 'засохший', 'иссушенный', 'увядший' и т. п. 10.3.

Здесь земля сухая, а там (у реки / болота) она влажная; эта одежда уже сухая, а та еще мокрая. 10.5. Тип. сем. предки/потомки: вышеупомянутые значения ('сухой' и 'засохший') легко переходят друг в друга. Помимо этого, значение 'сухой' тесно связано с такими качествами, как 'твердый', 'жесткий'; 'пустой' (очевидное пересечение там, где определяемым объектом является сосуд для жидкости); 'жаждущий' или, реже, 'голодный' (метафорический перенос значения).

11.1. **ear** / **ухо**. 11.2. 'Ухо' как выступающая часть головы (т. е. следует отличать от 'внутреннего уха', а также абстрактного 'уха' = 'слуха', как в англ. *to have a good ear for smth.*). 11.3. *Отрезать кому-л. ухо; у человека маленькие уши, а у зайца (слона и т. п.) — большие.* 11.4. При наличии супплетивизма 'ухо' / 'уши' (что бывает иногда свойственно парным частям тела, хотя конкретные примеры для 'уха' нам не известны¹) следует ориентироваться на форму ед. ч. для поддержания принципа унификации. 11.5. Тип. сем. предки: 'ухо' чаще всего бывает морфологическим производным от глагола 'слышать'. Гораздо реже встречаются метонимические или метафорические переносы значения с других частей тела (напр., 'висок', 'щека' и т. п.) или похожих объектов ('раковина' и т. п.) — ни одного из таких потенциальных предков нельзя признать типовым. Тип. сем. потомки: (а) абстрактные значения — 'слух', 'понимание' и т. п., см. выше; (б) метафорические переносы — 'ушко' (чего-л.) и различные «уховидные» объекты. Для языков Африки типичнейшим производным от слова 'ухо', например, является 'лист' (как «ухо дерева»).

12.1. **eat** / **есть**. 12.2. Стилистически нейтральный, максимально частотный глагол, обозначающий процесс поглощения человеком твердой пищи; желательна возможность образования устойчивой идиоматической пары с глаголом 'пить'. 12.3. *Что ты сегодня ел?; чтобы жить, нужно есть и пить.* 12.4. Не эксклюзивной, но очень характерной для языков Африки типологической чертой являет-

¹ Ср., возможно, в авестийском языке: *gaōša-* 'ухо', но *uši* (дв. ч.) 'уши' (вторая основа более архаична) [Bartholomae 1961: 414, 486]. Этот пример, однако, нельзя считать хорошей иллюстрацией, т. к. для основы *gaōša-* также засвидетельствованы формы дв. и мн. ч., а *uši* можно скорее рассматривать как архаизм, сохраняющийся не столько в прямом, сколько в переносном значении ('ухо' → 'слух', 'понимание', 'интеллект' и т. п.).

ся возможность наличия большого числа различных глагольных основ, объединяемых общей семантикой 'есть', но сочетающихся с различными типами объектов: 'есть (мясо)', 'есть (фрукты)', 'есть (корешки)' и т. п. Все эти «мини-синонимы», тем не менее, на практике применяются относительно редко, и пока что нам не известно ни одного случая, когда бы в языке, наряду с этими словами, не существовало общеродовой основы 'есть (что-л.)'. Единственное допустимое исключение из этого правила — бинарное противопоставление 'есть (твердую пищу, в первую очередь мясную)' : 'есть (мягкую пищу, т. е. растительную)', хорошо известное, например, в чадских и центральнокойсанских языках. Рассмотрение отдельных конкретных ситуаций, тем не менее, приводит к выводу, что более «базовым» глаголом в них является тот, который сочетается с *мягкой* (растительной) пищей: он, как правило, обладает более гибкой сочетаемостью (в частности, участвует в образовании большего количества идиоматических выражений) и большей продуктивностью, а также, по видимому, более стабилен: семантика 'есть (жесткую пищу)' часто оказывается очевидно вторичной, т. е. производной от таких значений, как 'кусать', 'грызть' или (в центральнокойсанских языках) от именной основы 'мясо'. Таким образом, для лексикостатистики по всем параметрам рекомендуется учитывать как более «базисное» значение 'есть (растительную пищу)'. 12.5. Тип. сем. предки: как и в случае с 'пить', глагол 'есть' типично образуется из-за расширения таких значений, как 'кусать'; 'грызть'; 'жевать', а также стилистически маркированных 'жрать' (т. е. 'грубо есть' или 'есть /о животном/'). Тип. сем. потомки: обычное морфологическое производное ('еда') иногда оказывается устойчивее, чем сама производящая основа, при этом за ним нередко закрепляется более узкое значение, отсылающее к «базовому» пищевому продукту для той или иной культуры ('мясо', 'рыба' и т. п.).

13.1. **egg / яйцо**. 13.2. В первую очередь *птичье* яйцо (как родовой термин), хотя желательна применимость термина и к другим классам животных (пресмыкающимся, насекомым и т. п.). 13.3. *Птицы несут яйца, а звери — нет; он взял в руку яйцо*. 13.4. Следует, разумеется, отличать от узкоспецифических терминов ('икра', т. е. 'рыбьи яйца'; 'яйцо конкретной птицы', напр. 'страусиное яйцо'), если речь не идет о полисемии, когда один и

тот же термин применяется одновременно к понятию яйца вообще и к яйцу конкретной птицы — такие ситуации иногда встречаются, причем они полностью аналогичны ситуации с самим словом 'птица', см. выше). 13.5. Тип. сем. предки: 'камень' ('камушек', 'кругляшок'); 'круглый' (слово может быть морфологическим производным от адъективной или глагольной основы). В некоторых языках встречается отдельная глагольная основа 'нести яйца' (сама по себе производная от глаголов давания, движения и т. д.), от которой, в свою очередь, может быть образовано именное производное 'яйцо'. Тип. сем. потомки: 'яйцеобразные объекты', чаще всего — 'яички' (*testiculi*). (Есть основания предполагать, что возможно и обратное развитие, от 'яичек' к 'яйцу', хотя претендовать на «типовой» статус оно едва ли способно).

14.1. *еуе* / *глаз*. 14.2. 'глаз' человека как орган зрения. 14.3. *Выколоть кому-л. глаз; у всех людей два глаза, а у этого человека только один глаз*. 14.4. При наличии супплетивных основ в целях унификации подхода выбирается основа ед. ч.¹ Как и в случае с 'ухом', 'глаз' следует отличать от более узких ('зрачок' и т. п.) и от абстрактных ('зрение') терминов. 14.5. Тип. сем. предки: чаще всего 'глаз' оказывается морфологическим производным от глагольной основы 'видеть' (обратное развитие встречается очень редко и не может быть признано типовым). Прочие развития (метафорические — от 'камушка', 'кругляшка'; метонимические — от 'зрачка' и т. п.) по сравнению с развитием 'видеть' → 'глаз' встречаются намного реже и вряд ли могут считаться типовыми. Впрочем, для языков Африки и некоторых других ареалов характерна частая семантическая связь между понятиями 'глаз' и 'лицо', причем развитие здесь может, по-видимому, идти в обоих направлениях (ср. относительную взаимозаменяемость таких контекстов, как *я взглянул ему в лицо* и *я взглянул ему в глаза*). Тип. сем.

¹ Отметим, что с исторической точки зрения такие случаи синхронного супплетивизма, как французск. ед. ч. *œil* : мн. ч. *yeux*, супплетивными не считаются, т. к. отражают один и тот же исходный корень. Аналогичные ситуации «псевдо-супплетивизма» для этого и других слов встречаются и в языках Африки, что еще раз подчеркивает важность тщательного историко-этимологического анализа сопоставляемых лексических элементов на «неглубоком» уровне перед тем, как им будут присвоены индексы когнации.

потомки: помимо 'лица', слово 'глаз' может участвовать в образовании глагольных производных, как правило, обозначающих *направленное* действие, т. е. не *'видеть', а 'смотреть', 'рассматривать' (ср. русск. *глазеть*, англ. *to eye smbd. / smth.* и т. п.).

15.1. **fire / огонь**. 15.2. Наиболее общее и стилистически нейтральное обозначение 'огня' как субстанции (например, противопоставленной 'воде'). 15.3. *Огонь можно потушить водой; огонь обжигает кожу*. 15.4. Следует строго отличать от таких терминов, как 'пламя' (= 'языки огня', 'плазма'), 'костер' ('специально разведенный под открытым небом огонь'), 'очаг' и т. п. 15.5. Тип. сем. предки: в первую очередь — такие глагольные основы, как 'гореть', 'жечь' и 'сверкать', 'сиять'; реже — перечисленные выше именные основы, прошедшие через этап расширения значения ('пламя', 'костер', 'очаг'). Во втором случае семантическое развитие свободно идет в обоих направлениях (т. е. возможны и сужения: 'огонь' → 'очаг' и т. д.), но развитие от именной к глагольной основе ('огонь' → 'гореть') встречается редко и вряд ли может считаться типовым.

16.1. **foot / нога**. 16.2. 'Нога' ниже лодыжки (т. е. 'стопа', хотя для русского языка корректным эквивалентом является именно 'нога', см. ниже); в функциональном плане — 'нога' как инструмент ходьбы. 16.3. *Ноги нужны для того, чтобы ходить; у меня замерзли руки и ноги; ботинок (сандалию) нужно надеть на ногу*. 16.4. Как и в случае с 'рукой', все языки мира по большому счету делятся на те, которые лексически различают 'ногу выше лодыжки' (*leg*) и 'ногу ниже лодыжки' (*foot*), и те, которые используют для этих двух значений единый термин (точнее — в которых просто отсутствует соответствующая семантическая оппозиция). При составлении списков Сводеша исследователи иногда задаются конкретной целью во что бы то ни стало найти различие между значениями 'leg' и 'foot' даже в тех языках, где оно на «бытовом» уровне отсутствует — что приводит к появлению в ячейке 'foot' различных идиоматических выражений (таких, как 'конец ноги', 'голова ноги' и т. п.), на самом деле не употребляемых в повседневном общении за исключением тех редких случаев, когда в речевом контексте требуется целенаправленно подчеркнуть, что речь идет о 'foot', а не о 'leg' (напр., *'какая часть ноги у тебя болит?' — 'стопа' и т. п.*). Такой подход пред-

ставляется излишне педантичным и даже напрямую ошибочным, поскольку он заставляет маркировать специальную, а не базисную терминологию. Аналогичным образом стоит избегать маркирования таких специфических терминов, как русск. 'стопа', употребляемых скорее в «анатомически-ориентированном» дискурсе, нежели в повседневной речи. 16.5. Тип. сем. предки: широко распространены новообразования от глагольных основ — 'идти', 'ходить' (возможность развития из основ 'стоять', 'останавливаться' сомнительна, хотя и допускается в ряде этимологических источников). Что касается именных основ, то здесь наиболее типичны «жаргонные» переносы значения с частей тела животных — 'лапа', 'копыто', 'коготь' (ср. русск. *ноготь* и *нога*, где с точки зрения исторической семантики первична семантика 'ногтя'). Тип. сем. потомки: 'нога' регулярно развивает такие переносные значения, как (а) 'шаг', 'шагать' и (б) 'след', 'оставлять следы'. Помимо этого, возможны двунаправленные изменения 'foot' → 'leg' и 'leg' → 'foot'.

17.1. **hair** / **волосы**. 17.2. 'Волосы' (собирательный термин — ед. ч. 'волос' допустим только в том случае, если сингулятив представлен тем же корнем, что и форма мн. ч.), в первую очередь *на голове человека*. (На данный момент трудно установить, какое из двух значений — 'волосы на голове' или 'волосы на теле' = 'волосистой покров', 'шерсть', 'мех' и т. п. — является в средне-статистическом плане более устойчивым; мы останавливаемся на конкретизации 'волосы на голове' в первую очередь по техническим причинам, т. к. в большинстве собранных 100-словных списков традиционно большей популярностью пользуется именно эта семантика). 17.3. *Дернуть кого-л. за волосы; на голове растут волосы, на руках — ногти*. 17.4. Помимо вышеуказанных семантических различий, значение 'волосы' следует также строго отличать от многочисленных специфических типов волос ('длинные волосы', 'волосы на висках', 'чуб' и др.) или причесок ('шиньон', 'коса' и др.); данное семантическое поле, как правило, очень богато ввиду важного социокультурного статуса «волос» в разных обществах. 17.5. Тип. сем. предки: для значения 'волосы на голове' чаще всего встречается развитие из 'шерсти', 'волос на теле' (т. е. сужение первоначального значения, когда одна и та же основа могла выражать идею любого волосистого покрова); чуть менее частотен, но вполне естественен переход 'голова' → 'волосы'

(вероятно, берущий начало от таких пересечений, как *мыть голову* = *мыть волосы* и т. п.). Возможны также расширения исходных значений, перечисленных выше ('коса' и т. п. → 'волосы' вообще). Тип. сем. потомки: переходы, *обратные* вышеуказанным, в целом встречаются нечасто, т. е., например, специализация общего термина 'волосы' в значении 'конкретный вид волос или прически' никак не может быть признана типовой. Тем не менее, 'волосы' часто допускают такие метонимические или метафорические переносы, как, например, → 'перо' (= 'волосы птицы') или 'облако' (= 'волосы неба'); впрочем, для этих целей более «удобным» обычно оказывается термин 'волосы на теле', чем 'на голове'.

18.1. **hand** / **рука**. 18.2. 'Рука' от запястья до пальцев (т. е. 'кисть'); в функциональном плане — 'рука' как инструмент хватания; см. замечания к №16 'нога' (все они применимы и к данному случаю). 18.3. *У человека на руке пять пальцев; отрубить кому-л. руки; взять камень рукой*. 18.4. См. также комментарии к №16. 18.5. Тип. сем. предки: ситуация и здесь аналогична 'ноге', т. к. двумя наиболее типичными «семантическими предками» 'руки' оказываются (а) глагольные основы, такие, как 'брать', 'хватать'; (б) именные основы-«жаргонизмы» — 'лапа', 'крыло' и т. п. Прочие известные случаи переносов значения (напр., из 'ладони' или 'кости руки') по сравнению с этими двумя группами оказываются скорее исключениями. Тип. сем. потомки: как правило, различные метафорические переносы — 'ручка' (двери и т. п.). Если значения *hand* и *arm* лексически не различаются (или совпали), метафоризация возможна и по отношению к длинным предметам ('хобот', 'шланг' и т. п.). Встречаются (нечасто) и развития, обратные по отношению к (б), т. е. 'рука' → 'крыло' и т. п.

19.1. **head** / **голова**. 19.2. Базисное понятие, чаще и естественнее всего применимое по отношению к человеческой голове как к конкретному объекту (т. е. не абстрактная 'голова' как 'верх', 'глава', 'начало' и т. п.). 19.3. *Ударить кого-л. в драке по голове; болит голова; камень размером с человекью голову*. 19.4. В разных языках мира нередко встречается «транзитная синонимия» — старое слово для 'головы', в основном служащее для выражения абстрактных значений, и новое, применимое исключительно к «физической» голове, хотя старое слово в отдельных контекстах также может применяться к анатомическому объекту. В этих

случаях ориентироваться при составлении списков следует на новое слово. 19.5. Тип. сем. предки: (а) жаргонизмы — чаще всего такие, как 'чаша', 'сосуд'; (б) конкретизация таких значений, как 'верхушка', 'край', 'конец' (чего-л.) — это развитие, впрочем, может идти в обоих направлениях; (в) из других основ с анатомическим значением возможным типовым предком может быть 'череп', с меньшей вероятностью — 'мозг'. Любопытно отметить также рекуррентность корреляции значений 'голова' и 'лысый' (помимо возможной, хотя и не доказанной, связи славянск. **golvā* и **golbь*, аналогичные связи отмечены также в алтайских языках), т. е. речь идет о возможности развития типа 'лысина', 'лысая голова' → 'голова' вообще. Тип. сем. потомки: наиболее частотным и естественным развитием является абстрактизация → 'верх, верхушка', 'глава', 'начальник, вождь' или → 'край', 'начало'. Из более «конкретных» развитий ср. упоминавшийся выше переход → 'волосы'.

20.1. **hear** / **слышать**. 20.2. Глагол восприятия ('слышать'), а не направленного действия ('слушать'). 20.3. *Я тебя не слышу; внезапно услышать выстрел*. 20.4. Важно, чтобы слово действительно применялось по отношению к процессу *звукового* восприятия, т. е. чтобы узус его не был ограничен такими контекстами, как *я слышал, что...* (где 'слышать' скорее является эквивалентом 'знать', 'обладать какой-л. (услышанной) информацией'). 20.5. Тип. сем. предки: для языков мира типична ситуация полисемии, в которой один и тот же глагол может использоваться для выражения разных типов ощущений: 'слышать' = 'чувять' = 'чувствовать' (осозательно) и т. п., иногда также = 'воспринимать', 'понимать' (интеллектуальное восприятие). Все эти значения довольно легко переходят друг в друга, расширяются и сужаются в зависимости от конкретной ситуации. Других типовых предков для значения 'слышать' (например, каких-либо именных основ) вроде бы не обнаруживается. Тип. сем. потомки: широко распространена морфологическая деривация 'слышать' → 'ухо' (орган слуха), см. №11; внутри глагольной системы частотна деривация 'слышать' → 'спрашивать' (например, через механизм образования каузатива), а также развитие → 'знать'.

21.1. **heart** / **сердце**. 21.2. 'Сердце' в первую очередь как анатомический орган, а не в переносном значении (*чувствовать*

сердцем и т. п.). 21.3. *Вырезать из туши сердце; человеческое сердце больше собачьего.* 21.5. Тип. сем. предки: любопытным образом, достаточно широко оказывается распространен переход от абстрактного к конкретному значению: 'сердце' (как часть тела) ← 'дух', 'душа', 'жизнь' и т. п. Наиболее надежные примеры известны из языков Восточной Азии (сино-тибетские, дравидийские и др.), но в отдельных случаях подобные ситуации можно предполагать и для других ареалов, так что развитие можно считать до некоторой степени типовым. Еще одно частотное развитие — из значения 'середина', 'центр' (хотя в данном случае можно, по-видимому, говорить о возможности развития в обоих направлениях). Наконец, в области более конкретных метонимических переходов хорошо известно развитие 'грудь' → 'сердце' (при этом в языках, лексически противопоставляющих мужскую и женскую грудь, обязательно развитие только из *мужской*). Тип. сем. потомки: развитие от конкретного значения 'сердце' к абстрактным 'храбрость', 'мужество', 'жизненная сила' и т. п. (хотя переход непосредственно в 'душу', 'дух', обратный вышеописанному, либо редок, либо вообще невозможен).

22.1. **horn / рог.** 22.2. Применительно к наиболее естественным и широко распространенным для данного географического ареала животным, при наличии скотоводства — домашним (напр., 'рога' крупного рогатого скота), при отсутствии — диким (буйволы, антилопы, носороги и т. п.). 22.3. *Звери бывают с рогами, а бывают без; схватить зверя за рог.* 22.4. 'Рог' — один из немногочисленных элементов 100-словника, для которого эксплицитно известны случаи отсутствия лексического эквивалента (например, в ряде языков Океании, где островные условия обитания до знакомства с европейской цивилизацией могли исключать присутствие соответствующей реалии); в подобного рода случаях слово просто исключается из подсчета, однако для языков Африки эта проблема нерелевантна. Как обычно, при составлении списков следует избегать узкоспецифических терминов — 'особые виды рогов' (= 'ветвистые рога', 'рожки молодого животного' и т. п.), а также 'рог (как сосуд для питья)' и 'рог, рожок (как музыкальный инструмент)'. 22.5. Тип. сем. предки: 'рог' может являться конкретизацией абстрактных значений 'край', 'конец', 'кончик', 'острие' (чего-л.); известны также метонимические переходы от значения

'лоб' и (реже) 'голова'. Тип. сем. потомки: шире всего представлена специализация в указанных выше значениях 'рог для питья', 'рог как инструмент'. Следует особо подчеркнуть, что для скотоводческих культур *или* культур, находящихся в тесном контакте со скотоводами, с высокой вероятностью можно ожидать заимствованного характера этого элемента (ср. заимствование 'рога' в большинство финно-угорских языков из индо-иранского источника).

23.1. **И / я.** 23.2. Стилистически нейтральная форма местоимения 1-го л. ед. ч. В отличие от неместоименных элементов списка, для этого и других местоимений *допускается* включение в список до двух супплетивных вариантов основы, чередующихся в рамках единой парадигмы (прямую и косвенную), при условии, что оба варианта могут функционировать как актанты (субъект или объект) — т. е., например, специальные посессивные аффиксы, присоединяемые к имени или глаголу, в лексикостатистических подсчетах остаются незадействованными¹. 23.3. *Кто там? Это я; Я знаю, а он не знает; он меня видит, а я его нет.* 23.4. Категорически не допускаются к подсчетам специально маркированные формы вежливости (что в значительной степени снимает вопрос о возможности заимствования местоименных форм 1-го л., актуальный, например, для языков Юго-Восточной Азии). 23.5. Поскольку местоимения 1-го и 2-го л. относятся к наиболее стабильному разряду элементов списка Сводеша, вопрос об их типовых семантических предках крайне сложен. На данный

¹ Отдельный вопрос — как поступать с языками, в которых самостоятельные формы личных местоимений употребляются скорее в эмфатической функции (аналогично латинскому *ego*), а нейтральное значение 1-го л. выражается связанными приглагольными показателями. С точки зрения унифицирующего подхода к материалу использование приглагольных показателей нежелательно. Учитывая, что даже в ситуациях, аналогичной латинской, в эмфатической функции все равно обычно выступает исконное местоимение, без замены на какую-то специально сконструированную форму, а также общее требование по возможности использовать в рамках 100-словника *несвязанные* формы, мы будем в любом случае ориентироваться на местоименную парадигму, а не на глагольные показатели — хотя обращать внимание на глагольные показатели в целях этимологической поддержки лексикостатистики все равно необходимо.

момент можно утверждать, что ни для одного из реконструированных на сколь-либо глубоком уровне местоимений 1-го л. не удается убедительно доказать его происхождение от какой-л. именной (и тем более глагольной) лексемы — за исключением, разумеется, специальных вежливых форм. В лучшем случае удастся зафиксировать «скачки» между функциями разных морфем, общим семантическим инвариантом которых является значение 1-го л. — например, выравнивание прямой основы по аналогии с (а) косвенной, (б) посессивной, (в) приглагольными показателями, (г) основой местоимения 1-го л. мн. ч., если они были изначально различными. В отдельных случаях *допустимо* — как и в ситуациях с другими местоимениями, напр., вопросительными, см. ниже — предполагать возможность развития из служебных морфем (например, эмфатических частиц), если в рамках сложной формы «А: морфема 1-го л. ед. ч. + Б: эмфатическая частица» основная функция выражения местоименного значения переносится на морфему Б, а морфема А, например, в ходе элементарной фонетической редукции, выпадает. Однако каждая такая гипотеза нуждается в тщательном историческом обосновании. Типовые семантические *потомки* для местоимения 1-го л. ед. ч. отсутствуют — за исключением разве что местоимения 1-го л. мн. ч., которое часто образуется от формы ед. ч.

24.1. **kill / убивать.** 24.2. Стилистически нейтральный глагол, применимый к максимально разнообразным ситуациям насильственного причинения смерти, в которых как агенсом, так и пациенсом может являться человек, а не только, например, животное, ср. 'задрать' (животное-агенса) или 'зарезать', 'забить' (животное-пациенса). 24.3. *Нехорошо человеку убивать другого человека; он такой сильный, что может убить кого угодно.* 24.4. Для глагола 'убивать' в языках мира, особенно литературных, свойственна обширная (квази-)синонимия, часто зависящая от различения очень тонких семантических оттенков и даже своеобразного разбиения возможных агенсов и пациенсов процесса умерщвления на классы, каждому из которых соответствует свой предикат. Тем не менее, и здесь, как правило (см. выше аналогичную ситуацию с глаголом 'умирать'), статистически выделим наиболее частотный и нейтральный вариант, хотя в отдельных случаях допустимо прибегать к постулированию синонимов

(например, в рамках «транзитной» синонимии). Как и в случае с 'умирать', при обнаружении супплетивизма 'убивать (одного)' : 'убивать (многих)', в лексикостатистических целях следует использовать сингулятивную основу. 24.5. Тип. сем. предки: (а) 'убивать' как морфологическое (каузативное) производное от 'умирать'; (б) из собственно семантических переходов наиболее известно развитие из значения 'бить', 'ударять' (ср. русск. *убить*), реже — из семантически близких глаголов, обозначающих сильное физическое воздействие на объект ('рубить', 'крушить' и т. п.).

25.1. *leaf* / *лист*. 25.2. 'Лист' как родовой термин, применимый к максимально большому количеству растений. 25.3. *Сорвать с дерева лист; на живом дереве растут листья, на мертвом дереве листьев нет*. 25.4. Если в языке различаются сингулятивная лексема 'лист' и коллективная 'листва', выбирать следует 'лист'. Отдельная «географическая» проблема — ситуации, в которых листовые растения, с которыми знакомы носители языка, отличаются отсутствием разнообразия (ситуаций, в которых носители вообще не знакомы ни с какими видами листьев, по-видимому, все же нет). Здесь «родовой» термин для понятия 'лист' может либо отсутствовать вообще, либо представлять собой какую-нибудь сложную идиому, «базисной» же лексемой будет скорее являться 'лист' конкретного «дерева *par excellence*» для соответствующего ареала (например, *Boscia albitrunca* для района Калахари и т. п.). По-видимому, здесь стоит обращать внимание на частотность узуса — если родовой термин отличается высокой степенью сложности или является редко используемым заимствованием, допустимо использовать вместо него слово, обозначающее лист конкретного дерева. Однако каждый такой случай должен быть максимально тщательно эксплицирован. 25.5. Тип. сем. предки: для 'листва' характерны (а) метонимические переносы: ← 'какая-л. часть или какой-л. тип растения' = 'стебель', 'цветок', 'трава' и т. п.; в максимально конкретном плане такие соответствия чаще всего описываются как сужение общего значения 'зелень', 'растительный покров'; (б) метафорические образования — 'лист' часто воспринимается как та или иная «часть тела» дерева. Чаще всего (особенно для африканского ареала) встречается 'лист' как 'ухо' (дерева), но попадаются также и 'лист' как 'глаз дерева', и, совсем редко, как 'рука дерева', 'крыло дерева'. Тип. сем.

потомки: хорошо известны случаи образования адъективной основы 'зеленый' ← 'лист'.

26.1. **louse** / **вошь**. 26.2. Имеется в виду в первую очередь 'головная' вошь (*Pediculus capitis*) — наиболее «естественная» разновидность, отличная от 'платяной вши' (*Pediculus corporis*), 'лобковой вши' (*Phthirus pubis*) и т. п. (впрочем, во многих языках эти виды принципиально не различаются), не говоря уже о 'гниде' (= 'яйцо вши') и о других видах насекомых-паразитов ('блоха' и т. п.). 26.3. *У него по голове ползают вши; вычесать из волос большую вошь*. 26.5. Вопрос о семантических связях слова 'вошь' с другими лексемами, как это ни странно, чрезвычайно сложен. За исключением окказиональных метонимических переносов, способных, как правило, иметь оба направления ('вошь' ↔ 'гнида', 'вошь' ↔ 'блоха'), проследить за образованием именной основы 'вошь' от какого-либо процесса ('кусать?'; 'ползать?') или признака ('маленький?') практически невозможно — при том, что само слово иногда обнаруживает нетривиальные и нерегулярные фонетические развития (например, в индоевропейских языках), иногда интерпретируемые как следы возможного табуирования первоначальной формы слова. При этом нельзя параллельно не отметить и удивительно высокий «рейтинг» 'вши' в различных индексах стабильности (ср. упоминавшееся выше *первое* место в индексе стабильности элементов 40-словного списка в работе [Holman et al. 2008] — очевидный методологический курьез, но и в индексе стабильности С. А. Старостина 'вошь' стоит довольно высоко, на 17-й позиции). Все эти наблюдения заслуживают, на наш взгляд, дальнейшего развития и уточнения в рамках отдельной монографии.

27.1. **meat** / **мясо**. 27.2. 'Мясо' как в первую очередь пищевой продукт; максимально общий термин, применимый как к разным подвидам мяса (снятого с разных частей туши животного), так и к разным его состояниям (сырое, приготовленное разными способами и т. п.). 27.3. *Я не ем мяса; это мясо какого животного?; он съел и мясо, и жир*. 27.4. Отдельные трудности могут возникать с отделением 'мяса' (*meat*), т. е. 'съедобной мышечной массы' от 'плоти' (*flesh*), т. е. 'сплошной твердой массы, из которой состоит внутренняя часть организма', противопоставленной 'крови' и / или 'коже' (толкование не претендует на строгую точность). В тех языках, где удастся обнаружить такое противопоставление, тре-

буется выбирать именно 'мясо', а не 'плоть', но в отдельных случаях, когда информации не хватает, «технической синонимии» избежать все же не удастся. 27.5. Тип. сем. предки: (а) 'мясо' часто бывает именным производным от глагольной основы 'есть'; (б) типичны и частотны разнонаправленные переходы типа 'мясо' ↔ 'плоть', причем последняя, в свою очередь, может быть тесно связана с 'телом' как таковым. Тип. сем. потомки: для языков Африки и ряда других ареалов хорошо известна полисемия 'мясо' / 'животное', где второе значение, как правило, вторично ('животное' как предмет охоты). Помимо этого, для некоторых ареалов характерно понятие «базисного пищевого продукта», который, в зависимости от конкретных обстоятельств существования носителей языка, может варьировать между 'мясом' и другими объектами (ср. семитский корень **lhm*, в др.-еврейском обозначающий 'хлеб', в арабском — 'мясо', или сино-тибетский корень **ɲa*, в большинстве ветвей обозначающий 'рыбу', но в ряде подгрупп опять-таки изменяющий значение на 'мясо' и т. п.).

28.1. **moon / луна**. 28.2. 'Луна' как небесное светило, противопоставленное 'солнцу' (и / или 'звездам'). 28.3. *Днем светит солнце, ночью — луна; на небе круглая луна*. 28.4. Следует избегать специальной лексики, обозначающей различные фазы луны ('месяц', 'лунный серп'); если в языке обнаруживается отдельный неидиоматический термин для 'полной луны', отличный от 'луны' вообще, его также надлежит отвергнуть, хотя типична все же ситуация, когда 'луна' и 'полная луна' выражаются одним корнем. 28.5. Тип. сем. предки: наиболее часто встречается образование слова 'луна' от функциональных признаков этого объекта: 'сиять, светить; (быть) светлым', а также от цветообозначения 'белый'. Тип. сем. потомки: широко распространенная полисемия 'луна' / 'месяц (года)' часто приводит к тому, что старое слово перестает употребляться в исходном значении и сохраняет только семантику 'месяц'.

29.1. **mouth / рот**. 29.2. 'Рот' как часть лица, включая губы и ротовую полость. 29.3. *Положить в рот кусок пищи; прополоскать рот водой*. 29.4. Узкоспецифические термины, такие, как, например, 'полость / внутренность рта', противопоставленный 'рту' вообще (ср. оппозицию *owurt* и *ayiz* в тюркских языках [Севортян 1974: 81-83, 407-409]), при составлении списка учитываться не

должны. 29.5. Тип. сем. предки: в отличие от других органов и частей тела, для значения 'рот' нетипично развитие из глагольных основ (хотя интуитивно естественной могла бы показаться семантическая связь с глаголами 'есть', 'кусать', 'говорить' и т. п.). Гораздо чаще встречаются переходы: (а) метонимические: 'рот' ↔ 'челюсть', а также 'рот' ↔ 'лицо'; (б) метафорические: 'рот' ↔ 'дыра, отверстие', 'рот' ↔ 'край, окончание'; сюда же можно подключить и жаргонизмы: 'рот' ← 'пасть', 'морда' (животного), 'клюв' (птицы). За исключением жаргонизмов, все остальные развития могут происходить в обоих направлениях.

30.1. **name** / **имя**. 30.3. Скажи мне имя твоего отца; у новорожденного еще нет имени. 30.5. Тип. сем. предки: 'имя' — одно из самых стабильных существительных в рамках 50-словного списка, что чрезвычайно затрудняет изучение семантических связей этого слова. В редких случаях отмечено развитие из более общего значения 'слово' (ср. литовск. *vardas*), причем, по-видимому, развитие это может идти в обоих направлениях (ср. лепча *miŋ* 'слово' ← прасино-тибетск. 'имя'). Тип. сем. потомки также не определяются (за исключением регулярного образования производных глагольных основ с значением 'именовать', 'называть').

31.1. **new** / **новый**. 31.2. 'Новый' в значении 'только что сделанный', 'только что появившийся / приобретенный' (в некоторых языках лексически отличается от 'нового' как 'не успевшего испортиться', 'не выработавшего свой ресурс'). 31.3. *Построить новый дом; изготовить новый инструмент*. Не подходит контекст типа **эта одежда пока что новая*. 31.4. Следует также отличать от значения 'молодой' ('young'), хотя во многих языках здесь наблюдается полисемия. 31.5. Тип. сем. предки: как правило, значение 'новый' оказывается побочным развитием из обозначений таких (в целом более конкретизированных) признаков, как 'свежий' (о продуктах или растительности); 'молодой' (о людях, животных); и даже 'зеленый', т. е. = 'молодой' (о растительности). В редких случаях удается предположить и обратное развитие ('новый' → 'свежий', 'новый' → 'молодой', хотя 'новый' → 'зеленый', по-видимому, все же невозможно).

32.1. **night** / **ночь**. 32.2. 'Ночь' как антоним 'дня' в значении 'темное время суток от заката до рассвета'. 32.3. *Ночь сменяет день; ночь светла, когда светит луна*. 32.4. Следует строго отличать от

более кратких периодов времени, в первую очередь 'вечера', но также 'сумерек', 'ранней ночи' и т. п. 32.5. Тип. сем. предки: прежде всего следует отметить тесную переплетенность перечисленных значений — во многих языковых семьях 'ночь', 'вечер', 'сумерки' и т. п. легко переходят друг в друга и даже меняются местами (ср. ситуацию в семитских языках, где аккадск. *lilāt* 'вечер' = запад.-семитск. **layl*- 'ночь', а аккадск. *mūš*- 'ночь' = эфиосемитск. **mās*- 'вечер'); это неизменно затрудняет семантическую реконструкцию и требует детальнейшего дистрибуционного анализа. Помимо этого, наиболее естественным семантическим предком для понятия 'ночь' могут быть адъективные основы 'темный' и 'черный' (обратные развития, т. е. 'темный' ← 'ночной', вроде бы встречаются редко).

33.1. **nose** / **нос**. 33.2. 'Нос' как «монолитная» часть лица, т. е. именно ед. ч. 'нос', а не мн. или дв. ч. 'ноздри' (за исключением случаев, когда эти значения выражаются одной и той же основой или даже одной и той же морфологической формой). 33.3. *Мы дышим и нюхаем носом; схватить кого-л. за нос; отрезать кому-л. нос*. 33.4. Как и в случае с 'глазом' и 'ухом', следует отличать от абстрактного 'носа' как 'обоняния', 'чутья'. 33.5. Тип. сем. предки: (а) как и другие органы, 'нос' иногда оказывается производным от глагольных основ, обозначающих его функции: 'дышать', 'нюхать', а также 'храпеть', 'сопеть' и т. п. (впрочем, по сравнению с такими развитиями, как 'глаз' ← 'видеть', 'ухо' ← 'слышать', образование 'носа' от глагольных основ встречается заметно реже); (б) жаргонизмы — 'нос' как 'клюв', 'рыло', 'морда'. Тип. сем. потомки: 'нос' легко развивает такие переносные значения, как 'край', 'кончик', 'выступ', которые затем конкретизируются применительно к тем или иным объектам (ср. *нос лодки* и т. п.). Возможен ли обратный переход, т. е. развитие от 'края' к 'носу', не вполне ясно, но исключить такую возможность нельзя.

34.1. **not** / **не**. 34.2. Под 'не' имеется в виду основной маркер отрицания, используемый при индикативных формах глагола; таким образом, это единственный элемент как 50-, так и 100-словника, который может быть представлен не только отдельным словом (частицей), но и связанным грамматическим формантом. Это не должно вызывать серьезных проблем, т. к. категория отрицания универсальна и имеет конкретные, сегментные (в

«худшем» случае — супрасегментные, если для какого-либо языка обнаружится, что отрицательная форма отличается от утвердительной только тональными характеристиками) средства выражения. 34.3. *Я не понимаю; ты его не видишь; он это не любит* и т. п. 34.4. Во многих языках система негативных маркеров развита очень детально, что приводит к существенным затруднениям при процессе отбора оптимального «синонима», на которых следует остановиться подробнее.

В первую очередь из рассмотрения *категорически* исключаются такие формы, как независимый «местопредикатив» *нет* (англ. *no*) — опыт показывает, что это слово отличается значительно меньшей стабильностью — и запретительное «не...!» (в таких контекстах, как *не делай этого* и т. п.), очень часто выражаемое этимологически особой основой.

Далее, базисное «не» бывает не только частицей или аффиксом, но и автономным «отрицательным глаголом», имеющим собственную парадигму спряжения; при этом, однако, необходимо следить за тем, чтобы в список включался отрицательный глагол *не быть* (*я не лезу/ твой друг, он не лезет/ делающий это = он этого не делает* и т. п.), а не *не иметь(ся)* (в таких контекстах, как *у меня нет чего-л., на столе нет тарелок* и т. п.).

Далее, в целом ряде лингвистических ареалов широко распространена супплетивная система негативных морфем, когда разные отрицательные основы употребляются при разных видо-временных формах глагола или при разных частеречных характеристиках сопровождаемого слова. Последняя проблема разрешается постулированием приглагольного отрицания как наиболее базисного, в отличие от, например, приаждективного или приадвербиального (*нехороший, нехорошо*). Что касается первой, то ее не удастся решить столь же просто. Рекомендуется все же ограничиваться исключительно морфемами, сопровождающими изъявительное наклонение глагола. *Внутри* этой системы допустима синонимия — например, одновременное включение в список отрицательных морфем, сочетающихся с перфективными и имперфективными формами (явление, достаточно широко распространенное в языках Африки).

Наконец, еще один «технический» вопрос — как поступать с таким явлением, как *циркумфикс* отрицания, также довольно

частотным как в Африке, так и в других лингвоареалах (ср. хотя бы французск. *ne... pas*). В диахроническом плане такие циркумфиксы обычно возникают как результат сочетания исходной отрицательной морфемы с дополнительной морфемой, носящей усилительный характер, причем, если употребление такого эфематического «модификатора» становится обязательным, в какой-то момент он может перенять на себя функцию главного члена отрицания: ср. в разговорном французском языке невозможность конструкции **je ne sais* 'я не знаю' при свободной допустимости *je sais pas*. Тем не менее, вряд ли корректно было бы утверждать, что в современном французском языке имела место лексико-статистическая замена *ne* → *pas*, поскольку опущение *ne* до сих пор ощущается как сознательный эллипсис, а сама эта частица легко восстанавливается в вежливой или даже стилистически нейтральной речи.

Правильным решением, по-видимому, было бы следующее: при анализе циркумфиксальных отрицаний нужно стараться (на основании сравнительных данных) выделить в них «старую» и «новую» части. Если «старая» часть, согласно описанию, до сих пор функционирует как «живой» элемент языка, т. е. не представляет собой статистически редкого архаизма, неохотно выдаваемого информантами, в такого рода ситуации нет серьезных оснований постулировать лексическую замену, и обе части циркумфикса можно рассматривать как подвид «транзитной синонимии» (для лексикостатистических подсчетов значимость будет, очевидным образом, иметь только «старая» часть циркумфикса). Аналогичное решение, кстати, будет пригодным и для ситуаций, в которых отрицание представляет собой монолитный комплекс, но состоит из двух скрепленных друг с другом морфем.

Тип. сем. предки/потомки: генезис отрицательных морфем пока что исследован недостаточно тщательно, но в целом можно сказать, что они обычно представляют собой замкнутую систему, в рамках которой возможны расширения или сужения функциональности тех или иных элементов. Так, нередко ситуации, в которых старый прохибитив начинает употребляться как индикативное отрицание, или же функции обычного отрицания переходят на один из отрицательных глаголов ('не быть', 'не иметь' и т. п.). Один из возможных выходов за рамки этой

системы — вышеупомянутые композитные основы, когда старое отрицание образует «тандем» с эмфатической частицей и впоследствии передает ей свою исходную функцию¹.

35.1. **one / один**. 35.2. Количественное числительное, способное выступать в функции определения к имени, обозначающему неодушевленный объект. 35.3. *На земле один лист, а на дереве много; один камень, два камня, три камня (много камней)*. 35.4. Если в языке существует специальная счетная форма, употребляемая при простом перечислении (*один = раз, два, три*), в список она не включается. 35.5. Тип. сем. предки: для числительного 'один' бывает характерна связь с системами указательных местоимений (← 'этот', 'тот'), а также развитие из значений 'единственный', 'другой', 'какой-нибудь / какой-то'. Тип. сем. потомки: хорошо известна грамматикализация слова 'один' в функции неопределенного артикля (если считать, что неопределенный артикль по своему значению — разновидность неопределенного местоимения, то соотношение 'один' ↔ 'некий, какой-то' можно тем самым расценивать как развитие, способное принимать оба направления).

36.1. **rain / дождь**. 36.2. Именная основа (следует строго отличать от глагола 'идти (о дожде)', часто представленного этимологически отличным корнем). 36.3. *Дождь — это вода, которая падает с неба; какой за дверью дождь — сильный или слабый?* 36.4. 'Дождь' как «неквантифицируемый» термин следует отличать от таких узкоспецифических слов, как 'ливень', 'шторм' («сильный» дождь), 'изморось' («слабый» дождь) и т. п. Для многих языков известна полисемия 'дождь' / 'вода' (в этом случае одно и то же слово занимает две позиции в списке) или же 'дождь' как композитная основа или идиоматическое сочетание 'вода неба' — в последнем случае ситуация должна разрешаться в соответствии с алгоритмом, описанным в разделе 1.6.3.4 (полиморфные основы). 36.5.

¹ Общей типологии систем выражения отрицательного значения в языках мира на самом деле посвящен колоссальный объем литературы, хотя, как и следовало бы ожидать, в основном соответствующие работы исследуют скорее синхронное, нежели диахроническое, поведение негативных морфем и конструкций. Из новейших публикаций можно порекомендовать коллективную монографию [Норт 2010], в которой излагаются наиболее современные взгляды на проблематику, а также приводится подробная библиография.

Тип. сем. предки: в первую очередь 'вода', 'влага'; из глагольных основ в качестве производящей для 'дождя' хорошо известна основа 'течь, струиться, лить(ся)', а также собственно 'идти (о дожде)' (последняя, в свою очередь, может восходить к тем же самым значениям — 'течь', 'литься' — или к 'падать', 'спускаться')¹. При сокращении композитной основы 'вода неба' до однокорневой возникает особое семантическое развитие 'небо' → 'дождь'. Тип. сем. потомки: за исключением того, что именная основа 'дождь' также может служить производящей для глагольной основы 'дождить', типовых семантических потомков для 'дождя' вроде бы не обнаруживается.

37.1. **smoke** / **дым**. 37.2. 'Дым' как продукт сгорания дерева и других «обычных» для сжигания материалов (т. е. для включения в список не годятся такие специальные термины, как 'дым от кремации', 'дым сигнального костра' и т. п.). 37.3. *Когда нет ветра, дым от костра поднимается прямо вверх; дым разъедает мне глаза; где огонь, там и дым.* 37.5. Тип. сем. предки: широко распространено образование именной основы 'дым' от глагольной основы 'дуть' (что касается глагола 'дымить', 'подниматься (о дыме)', то он может быть как производящим для 'дыма', так и производным от него). Помимо этого, известна связь между 'дымом' и такими значениями, как 'туман' (ср. *дымка*); 'пар'; (значительно реже) 'пыль' (метонимия — как дым, так и пыль могут разноситься ветром); и 'вонь', 'дурной запах'. Однако для всех этих связей трудно определить наиболее частотные направления развития значений;

¹ Любопытно в этой связи коснуться ситуации в тюркских языках, где **уау-тир* 'дождь' является композитной основой: **уау-* 'идти (о дожде или других осадках)' + **тир* 'вода' [ЭСТЯ 1989: 57]. Поиск внешних этимологических параллелей к основе **уау-* в других языках алтайской семьи показывает, что она может регулярно соответствовать таким коррелятам, как среднекорейск. *pi* 'дождь' и тунг.-маньчж. **pigi-n* 'буря' [S. Starostin et al. 2003: 1146], т. е. в конечном итоге все же восходить к именной или глагольно-именной основе. Если так, то в исторической перспективе для тюркского следует предполагать развитие от (а) *уау* 'дождь', *уау-* 'дождить' к (б) *уау-* 'дождить', *уау-тир* 'дождящая-вода' или 'дождь-вода', т. е. отсутствие лексикостатистической замены, т. к. основным «носителем» значения 'дождь' продолжает оставаться морфема **уау-*.

по-видимому, все они могут быть как «предками», так и «потомками» 'дыма'.

38.1. **star / звезда.** 38.2. Родовой термин, одинаково применимый к любой произвольно взятой звезде (или планете). 38.3. *Ночью на небе видны луна и звезды, а днем солнце; посмотри вон на ту яркую звезду.* 38.4. Из 38.2 следует, что для 100-словного списка неприемлемы такие термины, как собирательное 'созвездие' и тем более названия конкретных звезд и созвездий ('Полярная звезда', 'Большая Медведица', 'Плеяды' и т. п.). 38.5. Тип. сем. предки: две основные семантические коннотации 'звезды' — (а) 'свет', 'сияние' ('звезда' часто оказывается производной именной основой от глаголов 'сиять', 'светить', 'мерцать') и (б) ('яркая) точка', 'пятно', с метафорическими переносами с таких значений, как 'искра', 'метка', реже 'светлячок' и т. п. По-видимому, возможны и сужения значений от названия конкретных звезд к 'звезде' как таковой, но надежно доказанные случаи нам пока не известны. Тип. сем. потомки не идентифицированы.

39.1. **stone / камень.** 39.2. Родовой термин, применимый к различным камням средней величины (таким, которые можно взять в руку, бросить в человека, использовать как инструмент и т. п.). 39.3. *Он бросил в птицу камень и убил ее; забить колышек в землю камнем; на дороге валяется много камней, больших и маленьких.* 39.4. Семантическое поле 'камень' в зависимости от географических условий проживания носителей может оказаться исключительно богатым, с тщательной дифференциацией по форме, размеру и функциональности камня (ср. даже в русском — 'галька', 'щебень', 'булыжник', 'валун', 'глыба' и мн. др.; в «горных» языках эта терминология еще на порядок богаче). Тем не менее, пока что ни для одного языка из числа тех, по которым в рамках проекта «Вавилонская башня» составлялись 100-словные списки, не было доказано отсутствие общеродового термина, удовлетворяющего описанию в 39.2. Любопытное явление — широкое распространение полисемии 'камень' / 'гора', в диахроническом плане, по-видимому, двунаправленной (т. е. и 'камень' может оказаться 'маленькой горой', и 'гора' — 'большим камнем'). Если такая полисемия обнаружена в конкретном языке, не следует (как это делают отдельные составители списков) стараться во что бы то ни стало лексически «развести» 'камень' и 'гору' (например, выбирая

вместо 'камня' 'валун' или 'маленький камушек'); вполне естественно в такой ситуации заполнять обе позиции в 100-словном списке одним и тем же словом (для 50-словника эта проблема не актуальна, т. к. 'гора' в него не входит). 39.5. Тип. сем. предки / потомки: за исключением вышеупомянутой 'горы', широко распространенных семантических предков для 'камня', которые бы при этом относились к другому семантическому полю, установить не удается — как правило, все сводится к расширению значений тех или иных подвидов камней, или, наоборот, к сужению первоначального общего значения до отдельного подвида.

40.1. **sun** / **солнце**. 40.2. 'Солнце' как небесное тело, противопоставленное 'луне', 'звездам' и т. п. 40.3. *Не смотри на солнце, заболят глаза; ночью по небу ходит луна, а днем солнце*. 40.4. Во многих языках лексически противопоставлены 'солнце' как физический объект и 'солнечный свет', 'лучи солнца' — слова, в определенных контекстах взаимозаменяемые. В целях унификации подхода вторую лексему для лексикостатистики следует считать неприемлемой. 40.5. Тип. сем. предки: если для 'луны' и тем более 'звезды' наиболее типовой коннотацией является значение 'сиять, сверкать', то 'солнце' чаще ассоциируется с такими понятиями, как 'жара', 'тепло', 'огонь' (при этом обратное развитие, от 'солнца' к 'жаре', наоборот, нетипично, хотя и не может быть исключено). Помимо этого, в языках мира широчайшим образом распространена полисемия 'солнце' / 'день' — причем любопытен тот факт, что в ситуациях, когда этим двум значениям соответствуют морфологически производящая и производная основы, 'солнце' обычно оказывается производным от 'дня', а не наоборот (т. е. более естественным оказывается понимание 'солнца' как 'дневного светила', а не 'дня' как 'солнечного времени суток') — ср. хотя бы праенисейск. **xiʔg* 'день' : **xig-a* 'солнце', или описанную выше ситуацию в австронезийских языках, где 'солнце' = 'глаз дня'.

41.1. **tail** / **хвост**. 41.2. Слово должно быть применимо к максимальному разнообразию хвостатых животных, в первую очередь — к млекопитающим: такие специфические термины, как 'хвост птицы' (как любой, так и конкретной, напр. 'павлиний хвост'), 'рыбий хвост' и т. п., в список включать не следует. 41.3. *У зверей есть хвост, а у человека нет; собака машет хвостом*. 41.5. Тип. сем. предки: (а) метонимические — хорошо известны примеры

развития из значений 'задняя сторона', 'зад', реже 'спина'; в отдельных случаях развитие может идти от значения 'волосы', хотя обычно не от 'волос' вообще, а от более узких терминов ('густые волосы', 'сплетенные волосы' и т. п.); (б) метафорические — 'хвост' как 'палка', 'дубинка', 'шип, острие'. Тип. сем. потомки: одной из основных конкретных причин нестабильности слова 'хвост' может являться его тенденция к развитию побочного значения 'penis', первоначально в качестве эвфемизма, но в дальнейшем новый 'хвост' / 'penis' может легко вытеснить «старый» 'penis' на периферию и, в свою очередь, сам подвергается риску табуирования. В языках Африки такие процессы встречаются достаточно часто.

42.1. **thou / ты**¹. 42.2-42.5. Все сказанное выше относительно формы местоимения 1-го л. ед. ч. (№23) применимо и к местоимению 2-го л. ед. ч. (включая возможность рассмотрения супплетивных форм как синонимов, исключение специальных вежливых форм, проблему семантических предков и т. п.).

43.1. **tongue / язык**. 43.2. 'Язык' как часть тела (не в смысле *language*), в первую очередь человеческого. 43.3. *Показать кому-л. язык; он упал и прикусил язык; ему отрезали язык, он не может говорить*. 43.5. Тип. сем. предки: единственная более или менее стабильная корреляция прослеживается между 'языком' и глаголом 'лизать', морфологическим производным от которого 'язык' оказывается в различных ареалах планеты. Как это ни странно, но 'язык' как производное, например, от глагола 'говорить' встречается гораздо реже, несмотря на то, что вторичное значение 'язык (общения)' он развивает охотно и повсеместно².

¹ Корректным эквивалентом этого элемента списка в современном английском языке является, разумеется, *you*; нотация *thou* принята исключительно в целях снятия неоднозначности, чтобы *you* не воспринималось как форма 2-го л. мн. ч., каковой это слово является исторически.

² В некоторых языках обнаруживается формальное совпадение именной основы 'язык' и глагольной 'говорить', ср. в монгольск. *kele* и *kele-*, в инуитском эскимосском *uqaq* и *uqaq-* id. и т. п. В этимологически верифицируемых ситуациях (так, монгольское слово имеет высоковероятные параллели в алтайских и, далее, в уральских языках) первичным оказывается скорее 'язык', т. е. здесь налицо такая же корреляция между органом и тем сознательным направленным действием, инстру-

Другие семантические предки для этого слова устанавливаются с большим трудом. В пределах «очевидных» языковых таксонов 1-го и даже 2-го уровня оно, как правило, очень стабильно и не поддается внутренней этимологизации, а внешняя этимологизация часто затрудняется фонетическими нерегулярностями, которые «преследуют» этимон 'язык' в самых разных семьях, не связанных друг с другом ни близким родством, ни ареальными контактами (проблемный характер реконструкции 'языка' хорошо известен и в индоевропейской, и в дравидийской, и в сино-тибетской, и в многих других компаративистских традициях). Тип. сем. потомки: за исключением повсеместной метонимизации 'язык (орган)' → 'язык (средство общения)', широко распространена лишь метафора 'язык' как 'плоский (и) продолговатый объект', конкретизируемая многочисленными способами ('лезвие', 'лопасть' и т. п.), ни один из которых сам по себе не может быть признан типовым.

44.1. **tooth** / **зуб**. 44.2. 'Зуб' как общий термин для обозначения (по возможности) *любого* отдельно взятого зуба во рту человека. 44.3. *У взрослого человека много зубов, у новорожденных нет зубов; сломать зуб об камень (кость)*. 44.4. Термин следует строго отличать от названий конкретных видов зубов ('клык', 'резец', 'зуб мудрости' и т. п.). Для некоторых языков встречаются утверждения об отсутствии в них единого термина, покрывающего все 32 зуба (например, для баскского, где противопоставлены **hagin** 'коренной зуб' и **hartz** 'передний зуб', а собирательное понятие 'зубы' выражено композитной основой **hartz-haginak** [Trask 2008: 316]). Если такую ситуацию действительно можно подтвердить или хотя бы заподозрить на основании конкретных аргументов, то, по-видимому, можно допустить здесь синонимию, хотя в конечном итоге все должно зависеть от статистической частотности и степени контекстной связанности употребления обеих лексем. 44.5. Тип. сем. предки: как и другие «функциональные» части тела, 'зуб' чаще всего оказывается либо (а) именным производным от глагольной основы 'кусать' (но, любопытным образом, совершенно отсутствует наглядная деривация от 'есть', что, в частности,

ментом которого является этот орган, как и между лексемами 'глаз' : 'смотреть' (не 'видеть!'), 'ухо' : 'слушать' (не 'слышать!') и т. п.

ставит под сомнение популярную внутреннюю этимологию индоевропейск. **(e)dont-* 'зуб' от глагольного корня **ed-* 'есть'; либо (б) продуктом метафоризации таких понятий, как 'острие', 'острая палочка', 'колышек'. Следует, однако, отметить, что слово отличается исключительной стабильностью во всех ареалах мира; это существенно затрудняет нахождение для него надежных семантических источников. Тип. сем. потомки неизвестны (за исключением возможности обратной метафоризации, от 'зуба' к 'зубцу', 'зубчику' и т. п.).

45.1. **tree** / **дерево**. 45.2. 'Дерево' как родовой термин для обозначения любого вида древесных растений, растущего в дикой или домашней среде. 45.3. *У всякого дерева есть корень; у дерева толстый ствол, а у куста тонкий; залезть на высокое дерево*. 45.4. Помимо того, что это слово следует строго отличать от названий конкретных деревьев, нужно следить еще и за тем, чтобы в данной позиции не оказалось 'дерево' как строительный материал или как вид топлива (т. е. 'древесина', 'дрова' и т. п.). В отдельных ситуациях (как правило, там, где носители языка проживают в местности, лишенной разнообразия древесной флоры) 'дерево' как таковое может лексически отождествляться с конкретным видом дерева = единственным или одним из немногих представителей древовидных растений (см. выше 'лист'); при обнаружении таковых не следует стремиться к тому, чтобы во что бы то ни стало обнаружить отдельный родовой термин (который может, например, оказаться редко употребляемым заимствованием из какого-нибудь крупного соседнего языка). 45.5. Тип. сем. предки: 'дерево *растущее*', противопоставленное 'дереву как материалу', часто оказывается именным производным от таких глагольных основ, как 'расти', 'подниматься', 'стоять (вертикально)', при этом более старое 'дерево *растущее*' может сохраниться в языке как раз в значении 'дерево (материал)'. Другие возможности: (а) обобщение в значении 'дерево' конкретных названий деревьев ('дуб', 'пальма' и т. п.), см. выше; (б) сингуляризация коллективного существительного 'лес'; (в) редко, но все же встречаются метонимические переносы с таких значений, как 'ветка', 'ствол', 'стебель'. Тип. сем. потомки: вышеупомянутые значения, сводимые к общему инварианту «материалы / объекты, получаемые из (растущих) деревьев» — 'древесина', 'дрова', 'хворост' и т. п.

46.1. **two** / **два**. 46.2. Количественное числительное, выступающее в функции определения к имени (см. 'один'). 46.3. *На земле два листа, на дереве много листьев; один камень, два камня, три камня (много камней)*. 46.5. Тип. сем. предки: практически не известны, т. к. числительное 'два' — один из наиболее среднестатистически устойчивых элементов списка (индекс стабильности 2 согласно подсчетам С. А. Старостина), и успешная внутренняя этимологизация его в таксонах 1-го и даже 2-го уровня — явление чрезвычайно редкое. Иногда инновация случается в языках, лексически противопоставляющих значения 'два' (из многих) и 'пара' (= 'два из двух объектов', 'две одинаковые части одного объекта' и т. п.) — такова, например, ситуация в китайском. Другие возможные источники 'двойки' — такие слова, как 'половина'; 'другой' ('второй из двух'), однако надежных примеров на такие переходы известно очень мало. Тип. сем. потомки: иногда старая основа числительного 'два', исчезая в основном значении, сохраняется в таком производном образовании, как 'близнецы' (ср. *двойняшки* или англ. *twins*).

47.1. **water** / **вода**. 47.2. 'Вода' как субстанция, противопоставленная, например, 'огню', или же как одна из основных жидкостей для питья (языки, в которых эти две семантические функции 'воды' противопоставляются лексически, нам неизвестны). 47.3. *Человек не может жить без воды; огонь можно потушить водой; набрать в сосуд воды из реки*. 47.5. Тип. сем. предки: во многих языках мира наблюдается тесная связь между значениями 'вода' (питьевая = неподвижная) и 'река' (= 'проточная вода'), причем семантическое развитие 'вода' ↔ 'река' может, по-видимому, идти в обоих направлениях. С этим же, очевидно, связано и образование в ряде случаев 'воды' как именного производного от глаголов 'течь', 'струиться' и т. п. Наконец, 'вода' может оказаться производным от такой адъективно-вербальной основы, как '(быть) мокрым, влажным' (или сужением значения 'влага' = 'жидкость /любая/'); это развитие может протекать и в обратном направлении.

48.1. **we** / **мы**. 48.2-48.5. Для личного местоимения 1-го л. мн. ч. действительны все те же уточнения и ограничения, которые были описаны выше для местоимений 'я' и 'ты'. Следует, однако, добавить, что для слова 'мы' часто актуальны дополнительные парадигматические различия, невозможные для форм ед. ч. Это,

во-первых, противопоставление множественной и двойственной основ, во-вторых, различие эксклюзивных и инклюзивных форм. При этом, если относительно основы двойственного числа можно условиться исключать ее из рассмотрения (реальное противопоставление этимологически различных основ мн. и дв. ч. за пределами классических индоевропейских языков встречается редко), то противопоставление эксклюзива и инклюзива для многих языков носит фундаментальный характер, и определить, какая из них является более, а какая — менее «базисной», затруднительно.

Попытки обратиться к типологии генезиса соответствующей оппозиции также нельзя назвать успешными. Так, в рамках проекта ASJP при выборе корректного эквивалента предлагается брать только форму инклюзива — очевидно, для повышения корневого разнообразия списка, поскольку в типологическом плане эксклюзивная форма 'мы (без тебя)' часто оказывается образованной, с помощью аффикса мн. ч., от формы соответствующего местоимения ед. ч. 'я'. Однако и для форм инклюзива можно указать на типологические ситуации, когда последние образуются вторично — например, из сочетания местоимений 1-го и 2-го л. ('я и ты', 'мы и ты'). Для такой семьи, как дравидийская, более стабильной и неотъемлемой частью системы являются именно эксклюзивные формы, в то время как специальные формы инклюзива в дравидийских языках могут как присутствовать, так и отсутствовать, исчезать и образовываться вновь на не вполне ясной основе и т. п.

В силу этих обстоятельств разумным представляется допустить для данного случая супплетивную синонимию 'мы (экскл.)': 'мы (инкл.)', и считать, что формы местоимения 1-го л. мн. ч. в двух языках лексикостатистически тождественны в том случае, если налицо этимологическая корреляция хотя бы между одной из двух пар. (Формально к сравнению может быть допущено до *четырех* синонимов, учитывая возможные супплетивные прямые / косвенные основы, но реально такая ситуация пока что вроде бы неизвестна ни для одного языка мира).

49.1. **what / что**. 49.2. Вопросительное местоимение, отсылающее к неодушевленному референту, и способное занимать в предложении позиции актантов (субъекта и объекта). 49.3. *Кто*

там пришел и что он принес?; что ты делаешь из этого куска дерева?

49.4. При выявлении и внесении в список вопросительных местоимений 'кто' и 'что' необходимо по возможности следить за тем, чтобы вместо них (или параллельно с ними) в него не попадали *адъективные* основы ('какой', 'чей' и т. п.). Отдельная проблема — присвоение индексов когнации местоимениям, состоящим из нескольких морфем. Этимологический анализ показывает, что чаще всего такие композитные основы состоят из одной вопросительной морфемы и одной или нескольких «вспомогательных» — либо указательной ('это(т)', 'то(т)'), либо бывшего существительного (напр., 'вещь' для неодушевленного местоимения и 'человек' для одушевленного). Когнацию, таким образом, естественно считать по вопросительной морфеме, игнорируя «вспомогательные» морфемы точно так же, как игнорируются, например, различия в словообразовательных суффиксах при совпадающих корнях в родственных лексемах. 49.5. Диахроническая типология вопросительных местоимений исследована не намного подробнее, чем типология местоимений личных, однако в целом эта группа отличается меньшей устойчивостью. Один из возможных источников появления «новых» местоимений 'кто' и 'что' — перестановки в общей системе вопросительных и относительных местоимений; если совокупное число этимологически различных основ в этой системе достаточно велико, это повышает вероятность окказиональной смены или расширения их функций (например, 'что' ← 'какой' или 'который'). (Отдельный вопрос — сама по себе этимологическая причина такого разнообразия). Другой источник — вышеупомянутые композитные основы, частотность употребления которых может приводить к упрощению за счет элиминации «старой» вопросительной основы (ср. итальянск. *che cosa* 'что', букв. 'какая вещь', постепенно сдвигающееся в сторону сокращенного варианта *cosa*); аналогичные сценарии развития отмечались выше как возможные и для личных местоимений (когда старая основа, соединяясь с вспомогательной, рано или поздно «затирается» более новой морфемой).

50.1. **who** / **кто**. 50.2. То же, что и 49.2, только для одушевленных референтов. 50.3. *Кто пришел и что он принес?; кто убил эту собаку?* 50.4-50.5. См. комментарий к 49.4-49.5.

Часть II

Койсанские языки

Общий обзор.

Термин «койсанские» (англ. Khoisan, Khoesan) для обозначения общности языков, ранее известных под названием «бушменско-готтентотские», создан антропологом Й. Шапера [Scharera 1930], совместившим в одном слове этнонимы, которыми «готтентотские» племена кхой обозначали себя (*k^hoe*) и своих соседей «бушменов» (*san*). В языкознании этот термин твердо закрепился после выхода в свет итоговой работы Гринберга по генетической классификации языков Африки [Greenberg 1966]. Как старое, так и новое название изначально понимались в *ареальном* смысле: под ними объединялось несколько групп языков, распространенных на территории Южной и, в меньшей степени, Центральной Африки и характеризующихся рядом существенных общих типологических признаков, в первую очередь, наличием т. н. *щелчковых* звуков или *кликсов* (англ. *clicks*) — двухфокусных согласных, образующихся с помощью ингрессивного механизма артикуляции¹.

Вопрос о *генетическом* родстве всех этих языков до работ Дж. Гринберга активно не поднимался. Первая сознательная попытка как-то классифицировать весь массив известных на тот момент бушменских наречий была предпринята Доротеей Блик [Bleek 1927, 1929], которая разделила их на три основные группы: *северную*, *южную* и *центральную*, отмечая также многочисленные сходства между центральной группой и «готтентотскими» языками, но не настаивая на том, что они должны вместе образовывать одну таксономическую единицу. (До выхода в свет работ Д. Блик готтентотские языки, согласно гипотезе К. Майнгофа [Meinhof 1912], рассматривались как возможная подгруппа «хамитской» семьи; окончательно и бесповоротно эта гипотеза была отвергнута лишь в работах Дж. Гринберга, см. 2.4.1).

¹ Подробное описание основных артикуляторных механизмов, отвечающих за произнесение кликсов, их акустических признаков и типологического разнообразия см. в работе [Ladefoged & Traill 1994].

Массивный компендиум сравнительных данных Д. Блик по бушменским языкам [Bleek 1956] на сегодняшний день считается во многом устаревшим; новейшие исследования по тем из его языков, которые до сих пор остаются в живых (!кунг, !о!кунг, н|у, наро) показывают, что их фонологические инвентари были в то время установлены лишь приблизительно, а записи слов страдают от многочисленных (как системных, так и хаотических) ошибок. Тем не менее, на момент выхода в свет он, безусловно, представлял собой колоссальный прорыв в установлении койсанологии как отдельной субдисциплины в сравнительном языкознании, и вряд ли классификация Дж. Гринберга, во многом опиравшаяся на данные этого словаря, была бы возможна без тщательного учета содержащихся в нем данных.

Гринбергу удалось подтвердить деление койсанских языков на северную, центральную, и южную группу, а также убедительно показать, на основании многочисленных лексических и грамматических изоглосс, что языки готтентотов следует включать в центральную группу. Что касается более глубокого уровня родства, то, согласно его гипотезе, три вышеперечисленные группы должны объединяться в рамках «южноафриканско-койсанской» семьи (South African Khoisan), которой, в свою очередь, противопоставлены «восточноафриканско-койсанские» (East African Khoisan) языки-изоляты хадза и сандаве, объединяющиеся с южноафриканско-койсанскими на еще более глубоком, «общекойсанском» или «макро-койсанском» уровне (последний термин введен в обращение уже после выхода в свет работ Гринберга).

Койсанская классификация Гринберга как одна из составных частей его в целом позитивно принятой классификации языков Африки до сих пор регулярно воспроизводится в различных общих работах как по собственно африканистике, так и по сравнительно-историческому языкознанию в целом. Тем не менее, реакция на нее специалистов-койсанологов с самого начала была неоднозначной. В 1960-е — 1980-е гг. основным критиком идеи генетического единства всех «койсанских» языков был Э. Вестфаль, который в целом ряде публикаций [Westphal 1962, 1965, 1971, 1980] настаивал на том, что сходства, наблюдаемые даже между тремя подгруппами «южноафриканско-койсанских» языков (не говоря уже о хадза и сандаве), чересчур малочисленны и бессистемны,

чтобы можно было говорить о генетическом родстве, и, следовательно, любые бросающиеся в глаза сходства между этими языками следует объяснять как результат длительных межъязыковых контактов и ареальной диффузии.

Позиция Вестфалья в тот период немногочисленными койсанологами, которых хотя бы в какой-то степени интересовал вопрос исторических отношений между коренными языками Южной Африки, рассматривалась скорее как крайняя. Так, Э. Трэйлл, признавая необходимость четких стандартов для разграничения между контактной и генетически общей лексикой койсанских языков, утверждал, что «контактный сценарий» Вестфалья далеко не идеально согласуется с целым рядом разительных сходств в базисной лексике, список которых он опубликовал в работе [Traill 1986]. Широкую известность получила также его полемика с Вестфалем, развернувшаяся вокруг генетического статуса изолированного языка *ѓхоан*, открытого только в 1970-е гг. и поэтому не включенного Гринбергом в его общую классификацию. Трэйлл склонялся к тому, чтобы интерпретировать *ѓхоан* как своего рода «связующее генетическое звено» между северно- и южнокойсанскими семьями [Traill 1973]. Вестфаль, в свою очередь, был склонен согласиться с генетической привязкой *ѓхоан* к севернокойсанским, но никак не к южнокойсанским, и не готов был рассматривать данные *ѓхоан* как аргумент в пользу генетического родства этих двух семей [Westphal 1974], однако переубедить Трэйлла ему не удалось [Traill 1974].

Параллельно с развитием критического подхода к «койсанской» проблеме в этот же период предпринимались и отдельные попытки продемонстрировать общекойсанское родство на более серьезной и детальной основе, чем это было сделано у Гринберга. Здесь следует прежде всего упомянуть работы Г. Хонкена по койсанским системам личных местоимений [Honken 1977], примерным фонетическим соответствиям в сериях аффрикат и фрикативных [Honken 1988], а также общей типологии фонетических соответствий, наблюдаемых между различными «койсанскими» семьями [Honken 1998]. В дальнейшем, однако, Г. Хонкен отошел от идеи доказуемости общекойсанского родства, предпочитая, в рамках приобретающей все большую популярность «ареально-ориентированной» парадигмы, объяснять большинство лексико-

фонетических схождений между отдельными «койсанскими» семьями многотысячелетним периодом конвергентного развития [Honken 2006].

Многочисленные дополнения к лексическим параллелям Гринберга опубликованы в исследовании известного африканиста К. Эрета [Ehret 1986]. Ему же принадлежит авторство важнейшей гипотезы, согласно которой по крайней мере некоторая часть односложных слов, содержащих кликсы в «южноафриканско-койсанских» языках, могла быть образована из старых двусложных основ через редукцию корневого гласного и стяжение консонантного кластера в двухфокусный кликс — на это намекает ряд интересных сопоставлений между «южно-» и «восточноафриканской койсанской» лексикой (пожалуй, самый «классический» пример — сходство между сандаве *!ana* 'рог' и центр.-кой. *//a id., где предполагается трансформация латеральной аффрикаты в латеральный кликс, а срединного носового — в носовой исход кликса); идея развивается далее в работе [Starostin 2008a], хотя далеко не все потенциальные примеры на это явление, перечисленные в статье, одинаково убедительны.

К сожалению, дальнейшие изыскания К. Эрета в области койсанской реконструкции следует признать скорее неудачными; так, его попытка установить систему регулярных фонетических соответствий между тремя наиболее хорошо изученными представителями основных «койсанских» семей (жу|хоан для сев.-кой., !хонг для юж.-кой., нама для центр.-кой.) практически игнорирует факт наличия между этими семьями нетривиальных и неоднозначных соответствий, и зачастую отражает скорее ареальные, нежели генетические связи [Ehret 2003a, b]. Маловероятна эта система и с типологическо-исторической точки зрения: при опоре на фонетические соответствия Эрета оказывается, что в 100-словном списке анализируемые три языка имеют в среднем от 3 до 5% совпадений, т. е. их родство лежит в диапазоне статистической погрешности — и это при том, что фонетические соответствия между ними, по большому счету, тривиальны. Сколь бы ни была велика фонетическая уникальность «койсанских» языков, трудно поверить в то, что жу|хоан, !хонг и нама, будучи связаны генетическим родством на уровне не менее чем 12-15 тыс. лет тому назад, за столь длительный период почти

ничего не изменили в своих сверхсложных фонетических системах. Учитывая также и другие проблемы, легко обнаруживаемые в сопоставительных материалах Эрета (чрезвычайная натянутость семантических сопоставлений; отсутствие промежуточных реконструкций), вряд ли можно считать этот опыт серьезным вкладом в обоснование идеи южноафриканско-койсанского родства.

Список из 645 «койсанских этимологий» обнаруживается также в одной из обзорных работ М. Рулена [Ruhlen 1994: 45-69]; к сожалению, он составлен в соответствии с требованиями (точнее, отсутствием требований) метода «массового сравнения» и представляет собой не набор этимологий в сколь-либо строгом смысле этого слова, а лишь перечень фонетически сходных форм, что можно рассматривать только как предварительный шаг (набор первичного сопоставительного материала) на пути к установлению генетического родства.

К концу XX в., в свете многочисленных новых данных, собранных на профессиональном уровне, значительно превышающем уровень записей Д. Блик и прочих специалистов довоенного периода, а также в связи с общей тенденцией переноса акцента с исторической лингвистики на ареальную, в койсанистике явно наметился сильный крен в сторону «гиперкритической» позиции Э. Вестфала. Так, в сравнительном исследовании Б. Сэндс [Sands 1998a], анализирующем данные «койсанских» языков на предмет выявления неслучайных и системных совпадений, автор представляет свои выводы следующим образом:

«Despite the present inability to validate a significant number of cognate sets through the illustration of regular sound correspondences, the similarities seen between all major Khoisan groups with the exception of Hadza... indicate that there are similarities which go beyond what can be explained due to chance or borrowing. On balance, from the evidence that we now have on hand, *it seems a little more likely than not* that the Northern, Southern, Central Khoisan groups along with Sandawe are related but that additional data and further research is needed to elucidate the relationship» [курсив мой — Г. С.].

Однако уже в работе [Sands 2001] вопрос о генетических связях между «койсанскими» языками практически затирается более «насуцной» идеей списывания наблюдаемых между ними

сходств на эффект ареальной диффузии, ср.: «The widespread occurrence of a root in a given language family should not be taken as conclusive evidence that the root was inherited rather than borrowed or diffused from another language family, as we can expect multiple points of contact between Khoesan families» (стр. 213).

Разумеется, однозначно списывать изменение отношения к «койсанской проблеме» на общие усиления позиции ареальной лингвистики было бы неправомерно. В определенной степени оно оказалось обусловлено необходимостью привести классификацию «щелчковых» языков, в основе которой до сих пор лежит анализ Гринберга, в соответствии с современными представлениями об их фонетике, грамматике и лексике. (Подробная критика грамматической аргументации Гринберга содержится, в частности, в работе [Güldemann 2008]). Поскольку за последние полстолетия койсанистика успела обогатиться целым рядом детальных словарных, грамматических и сравнительных описаний тех койсанских языков, которые пока еще остаются в живых, наличие этих данных закладывает достаточно прочный фундамент для того, чтобы, наконец, перейти от этапа «массового сравнения» к настоящему сравнительно-историческому анализу. Однако непосредственное сопоставление в рамках такого анализа сразу всех койсанских языков вряд ли перспективно — как сторонники, так и противники «макро-койсанской» гипотезы сходятся во мнении, что, если все эти языки родственны между собой, то родство это должно быть очень глубоким¹; следовательно, корректное применение сравнительного метода к койсанскому материалу должно заключаться в том, чтобы сперва определить и реконструировать семьи более мелкого уровня, и только затем переходить к тестированию «макро-койсанской» теории.

Такой подход представлен, в частности, в работах Т. Гюльдеманна и Д. Элдеркина по обоснованию генетического родства центральнокойсанских языков с языком-изолятом квади, а также

¹ Так, К. Эрет приводит аргументы в пользу привязки «праюжноафриканского койсанского» к археологической культуре Уилтон (≈ VI-V тыс. до н. э.), отмечая, что это в целом сходится с результатами его глоттохронологических подсчетов по конкретным койсанским языкам [Ehret 2000: 391-2]. Очевидно, что хадза и сандаве («восточноафриканский койсанский») должны подключаться к этой ветви на еще более раннем уровне.

(на несколько более шатких основаниях) с языком-изолятом сандаве [Güldemann & Elderkin 2003; Güldemann 2004], что в корне противоречит старому делению на «южно-» и «восточноафриканско-койсанские» языки. Параллельно с этими исследованиями в работе [Starostin 2008a] приводятся — в форме лексического корпуса сопоставлений, между которыми в общих чертах намечены предварительные фонетические соответствия — аргументы в пользу тесного генетического родства северной койсанских и южной койсанских языков (для их объединения введен специальный термин «периферийно-койсанские», противопоставленный «центрально-койсанским»). Тем не менее, ни про «кхой-квади-сандаве», ни про «периферийно-койсанскую» гипотезы нельзя на данный момент считать, что они удовлетворяют всем основным требованиям сравнительно-исторического языкознания.

Вряд ли будет сильным преувеличением, если мы скажем, что популярность «койсанской» гипотезы, особенно среди лингвистов, не являющихся специалистами в области койсанистики, а также среди представителей смежных дисциплин (генетики, антропологии и т. д.), до сих пор основывается в первую очередь на типологической близости этих языков — в частности, наличия в них всех особых подсистем «щелчковых» согласных. Сам факт того, что хадза, сандаве и «южноафриканско-койсанские» языки обладают столь сложными и, вместе с тем, схожими друг с другом щелчковыми инвентарями, неминуемо наводит на мысль об их происхождении из общего источника. Сопоставляя эти типологические данные с результатами генетического анализа митохондриальной ДНК племен хадза и жу|хоан (северной койсанской ветви), которые на общеафриканском генетическом дереве отстоят друг от друга дальше, чем любые два других произвольно взятых этноса, некоторые исследователи автоматически заключают, что щелчковые согласные, оказавшиеся настолько стабильными в языках этих двух групп, должны были присутствовать и в праязыке человечества [Knight et al. 2003].

Резонную критику такой (во многом умозрительной) позиции можно найти в работах [Güldemann 2007; Güldemann & Stoneking 2008], где, в частности, подчеркивается, что ареальная диффузия «кликсов» от одной «койсанской» группы к другой является ничуть не менее реалистичным историческим сцена-

рием, чем представление о кликсах как реликтах древнейшего языкового наследия человечества. Действительно, учитывая хотя бы тот факт, что за последние две-три тысячи лет небольшие, но типологически схожие с «койсанскими» подсистемы щелчковых согласных образовались в нескольких языках банту, активно контактировавшими с бушменско-готтентотским населением Южной Африки (зулу, коса, сото, йейи)¹, нетрудно представить, как за временной отрезок в 5-6 тысяч лет языковая группа, первоначально вообще лишенная щелчковых согласных, могла бы, на базе сначала заимствований, а затем уже и внутренних фонетических процессов, развить целую сложную систему кликсов. Более того: активный процесс вторичного «кликсогенеза» предполагается даже теми исследователями, которые с симпатией относятся к «макро-койсанской» гипотезе (см. выше о работах К. Эрета).

На самом деле вопрос о степени архаичности систем щелчковых согласных в «койсанских» семьях неразрывно зависит от тщательной сравнительно-исторической обработки языкового материала этих семей и не может быть решен на основании чисто умозрительного подхода. На текущем этапе исследований следует воздержаться от использования «щелчкового аргумента» как серьезного довода в пользу макро-койсанского единства. Такое единство может быть продемонстрировано лишь на основе конкретных систематических сходжений между «койсанскими» языками в области базисной лексики. Подчеркнем, что по крайней мере в формальном плане «щелчковый аргумент» как таковой изначально отвергал и сам Дж. Гринберг, исходя из фундаментального положения о неприемлемости типологических изоморфизмов как свидетельств генетического языкового родства.

Забегая вперед, отметим также, что даже фонетико-типологическое «единство» всех потенциально койсанских языков не столь прочно, как может показаться на первый взгляд: подсистемы щелчковых согласных, наблюдаемые в «восточноафриканских койсанских» языках на самом деле существенно отличаются по своему устройству и функциональной нагрузке от аналогичных подсистем в «южноафриканских» языках — настолько, что в

¹ Подробнее о конкретных механизмах проникновения кликсов в языки банту см. [Argyle 1986; Vossen 1997b].

рамках обоснования их генетического единства в любом случае потребовалось бы вскрыть очень глубокий механизм «перестройки» этих подсистем (включая их отношение к подсистемам «обычных» согласных, а также к общей структуре корня), ничуть не менее глубокий, чем потенциально возможный механизм образования в языке инновативной системы кликсов «из ничего» (как в банту). Подробнее эти различия будут описаны ниже, в разделе 2.6 (язык сандаве).

В работе [Traill 1994b], опираясь на выявленное статистическое распределение щелчковых согласных в языках !хонг (южнокойсанская группа) и жу|хоан (севернокойсанская группа), автор делает вывод, что наблюдаемые корреляции удобнее интерпретировать как аргумент в пользу генетического родства этих языков, но родство это, скорее всего, настолько глубоко, что не может быть должным образом продемонстрировано с помощью сравнительно-исторического метода. Необходимость поиска других, альтернативных методик обоснования генетического родства Трэйлл напрямую объясняет тем, что (в случае койсанских языков) «...the standard comparative method is unable to establish the subgroupings» (стр. 440).

Неясно, однако, на каких конкретных основаниях зиждется данный аргумент. По сути, «стандартный сравнительный метод», о котором пишет Трэйлл, при сопоставлении разных *групп* койсанских языков до сих пор никем не применялся. Во-первых, большинство работ, посвященных этому вопросу, оперируют не столько группами, сколько отдельными — как правило, наиболее хорошо описанными — языками (жу|хоан, !хонг, нама, наро и т. п.), нарушая тем самым важнейший принцип ступенчатой реконструкции. Во-вторых, ни в одной из этих работ не видно целенаправленных усилий по разграничению между контактными и (потенциально) исконно родственными слоями лексики, что само по себе лишает исследователя возможности установить регулярные фонетические соответствия между исконно родственными слоями лексики в сравниваемых языках. А поскольку наиболее очевидные, бросающиеся в глаза лексические сходения относятся именно к разряду контактных, их обилие препятствует обнаружению нетривиальных соответствий, доказательность которых в вопросе установления родства намного выше.

Все это показывает, насколько важным для правильного разрешения вопроса о койсанской макросемье является упор на базисную лексику. В койсанистической литературе иногда эксплицитно высказывались сомнения в применимости стандартизированной лексикостатистики к койсанским языкам; в частности, Э. Трэйлл отмечает, что, как правило, лексические сходства между этими языками (нередко даже между географически удаленными друг от друга) наблюдаются в сфере более специфически культурной лексики, чем универсальных базисных понятий [Traill 1978: 126]; этот феномен, обычно списываемый на широко-масштабные заимствования, Трэйлл, тем не менее, не был готов автоматически считать признанием отсутствия койсанского родства. Но как раз для того, чтобы подтвердить или опровергнуть это интуитивное наблюдение, и требуется публикация подробных данных по предварительному лексикостатистическому обследованию обсуждаемых языков.

В отличие от остальных гипотетических африканских макросемей, анализ базисной лексики которых в рамках данного исследования проводится автором впервые, соответствующая обработка койсанского материала (по полным стословным спискам) уже была опубликована в работе [Starostin 2003]; на тот момент, однако, нам были недоступны материалы по ряду важных языков (н|у, экака !кунг и др.), а анализ материалов по доступным языкам был недостаточно подробным и содержал ряд серьезных ошибок (неполный учет заимствований, отдельные неверно установленные соответствия и т. п.), которые могли привести к искажению результатов. В связи с этим представляется необходимым провести ту же самую процедуру заново, на сей раз на материале 50-словных списков, анализ которых будет отличаться большей подробностью и, по возможности, методологической строгостью.

2.1. Севернокойсанская группа.

2.1.1. *Общие сведения и источники.* К севернокойсанской (North Khoisan) или, согласно новейшей терминологии, жу (Ju, Zhu) группе относится сравнительно небольшой пучок диалектов бушменских племен, распространенных на территории Намибии (северо-восточные районы) и, в меньшей степени, Анголы и Бот-

сваны (в областях, прилегающих к границе с Намибией); общее число носителей этих наречий колеблется в диапазоне 50,000 человек. Общепринятой точки зрения на то, сколько разных языков входит в эту группу, не существует; по-видимому, ее удобнее всего представлять в виде непрерывного диалектного континуума, в рамках которого последовательно сменяют друг друга *северный*, *центральный* и *южный* пучки диалектов ([Snyman 1997: 24-29]; [Treis 1998: 465-466]; к новейшим гипотезам относится попытка разделения северного пучка на собственно северный и отдельный «северно-центральный» [Sands 2010], хотя это утверждение еще нуждается в дополнительном историческом обосновании).

Несмотря на то, что многие диалекты до сих пор функционируют как активные средства общения, удовлетворительные лексикографические описания на настоящий момент существуют только для двух:

1) *жу|хоан* (жцх, южный пучок): общее число носителей в Намибии и Ботсване не превышает 35,000 человек, но диалект стабилен и даже оказывает ареальное влияние на ряд других «койсанских» языков в сопредельных регионах. Жу|хоан — один из немногих «бушменских» языков, для которых предпринимались попытки разработать официальную орфографию и даже программу школьного обучения (во многом благодаря стараниям выдающегося койсанолога П. Дикенса). В нашем исследовании все данные приводятся по словарю [Dickens 1994], с учетом более раннего словаря [Snyman 1975] и грамматики [Snyman 1970]);

2) *экока-!кунг* (эко, северный пучок): один из многочисленных диалектов, носители которых объединяются под автонимом !кунг или !хунг (= эко !хйй 'человек'); Экока — населенный пункт в Намибии, в окрестностях которого проживает один из наиболее многочисленных и, на данный момент, хорошо изученных анклавов !кунг. Основной источник данных — словарь [König & Heine 2008]¹.

¹ Этот словарь во многом обесценивает до недавнего времени самый подробный источник по лексике северных диалектов — работу [Heikkinen 1986], по крайней мере, в той ее части, которая релевантна для нашего исследования (базисная лексика); в связи с очевидной избыточностью, данные из нее не будут цитироваться при анализе 50-словных списков. Для исследований по синхронной и исторической грамматике

Для полноты картины, тем не менее, необходимо привлекать и менее надежные данные из более старых источников. Большую ценность имеют, в частности, записи Д. Блик по диалектам *!aυ||en* (*aye*; южный пучок; NI согласно «индексу Блик»)¹ и *!o!kунг* (*окн*; северный пучок; NIII согласно «индексу Блик»)², впервые опубликованные в [Bleek 1929] и затем перепечатанные и уточненные в [Bleek 1956]. Хуже всего представлены в источниках диалекты центрального пучка; наилучшее сочетание полноты и надежности записи, по-видимому, принадлежит К. Доуку [Doke 1925], опубликовавшему небольшое описание фонетики диалекта, локализованного рядом с городом Хрутфонтейн, который он называет просто *!кунг*. Лексическую близость к центральному пучку демонстрирует и «классический» диалект *!кунг* (NII согласно «индексу Блик»), по которому в 1879-1884 гг. большое количество текстов было собрано Л. Ллойд, хотя по своей фонетике и грамматическим особенностям этот диалект сегодня иногда оценивается и как тяготеющий к «северной», или, в терминологии Б. Сэндс, «северно-центральной» ветви [Lionnet 2009].

Не игнорируя материалы Л. Ллойд (в отдельных случаях ссылки на них могут быть полезны для прояснения спорных этимологических вопросов), мы все же в качестве «репрезентативного» центрально-северного диалекта севернокойсанской группы выберем более надежный в транскрипционном отношении говор Хрутфонтейн, записанный Доуком, для простоты нотации обозначая его как *!кунг* (*кнг*).

севернокойсанских языков вообще, и эока-!кунг в частности, большое значение имеют также подробная монография [Heikkinen 1987] и краткий очерк [König & Heine 2001].

¹ Более точное название — *!kxaυ||эйн* (*!kxaυ||ēi* или *!kxáó||āè*); диалект до сих пор жив (ок. 7,000 носителей в Намибии и Ботсване), но полевая работа с носителями возобновилась лишь совсем недавно, так что единственными существенными опубликованными записями до сих пор остаются только данные Д. Блик.

² Как и все прочие диалектные формы северного пучка, чрезвычайно близок к эока-!кунг, но все же имеет и небольшие лексические и, возможно, фонетические отличия (последние устанавливаются лишь с большим трудом, т. к. неясно, в какой степени зафиксированные отличия можно относить на счет элементарной неточности транскрипции Д. Блик).

Также в этимологических целях имеет смысл использовать обзорные сопоставительные данные, собранные по 15 севернокойсанским диалектам Я. Снейманом [Snyman 1997], хотя, в силу их «компаративной» организации, извлечь из них точные 100- или даже 50-словные списки затруднительно. Больше всего данных Снейману удалось собрать и опубликовать по т. н. «ангольскому !кунг» (в отдельной работе [Snyman 1980]) — наиболее «северной» из всех известных форм языков/диалектов жу; но в лексикостатистическом плане этот диалект почти неотличим от !о!кунг у Д. Блик, а слов из него при этом известно сильно меньше, так что мы будем привлекать данные Снеймана только в вспомогательных целях.

2.1.2. *Историческая характеристика.* Все рассматриваемые диалекты в лексическом, фонетическом и грамматическом отношении очень близки друг к другу; фонетические соответствия в целом тривиальны, а когнаты опознаются без особого труда, даже несмотря на то, что сами по себе фонологические системы отличаются исключительной сложностью.

Сопоставление данных по наиболее хорошо описанным диалектам — в первую очередь жцх и эко — позволяет реконструировать для прасев.-койсанского симметричный набор из 60 кликсов, сгруппированных в 5 серий по типу основы, т. е. первичного места артикуляции (дентальные, палатальные, альвеолярные, ретрофлексные, латеральные):

Исход кликса	Основа кликса				
	Дт.	Пл.	Ал.	Рт.	Лт.
Нулевой	*	* ̥	*	*	*
Озвончение	*	* ̥	* ̣	* ̣	* ̣
Назализация	* ̃	* ̥̃	* ̣̃	* ̣̃	* ̣̃
Гортанный взрыв	* ʔ	* ̥ʔ	* ̣ʔ	* ̣ʔ	* ̣ʔ
Аспирация	* ʰ	* ̥ʰ	* ̣ʰ	* ̣ʰ	* ̣ʰ
Гортанный взрыв + аспирация	* ʔʰ	* ̥ʔʰ	* ̣ʔʰ	* ̣ʔʰ	* ̣ʔʰ
Назализация+аспирация ¹	* ̃ʰ	* ̥̃ʰ	* ̣̃ʰ	* ̣̃ʰ	* ̣̃ʰ

¹ Очень редкая комбинация признаков, что, по-видимому, обуславливает и чрезвычайную нестабильность ее рефлексов в дочерних диа-

Исход кликса	Основа кликса				
	Дт.	Пл.	Ал.	Рт.	Лт.
Преглоттализация+назализация ¹	*ɣ n	*ɣ ɲ	*ɣ!n	*ɣ!!n	*ɣ n
Фрикативный велярный	* x	* ɣ	*!x	*!!x	* x
Фрикат. велярный+озвончение	* x	* ɣ	*!x	*!!x	* x
Велярная аффриката	* kx	* ɣkx	*!kx	*!!kx	* kx
Велярная аффриката+озвончение	* kx	* ɣkx	*!kx	*!!kx	* kx

В большинстве диалектов никаких существенных изменений этой системы по сравнению с праязыковой не произошло (отдельные «мелкие» развития, такие, как спорадическая утрата звонкой артикуляции кликсов *|kx, *|ɣkx и т. п. в диалектах северного пучка, не имеют существенной значимости для результатов нашего исследования).

Особо отметить необходимо только одну специфическую черту центрального пучка — наличие т. н. «ретрофлексных» кликсов (!! и производные), не встречающихся за пределами северной койсанской общности, но при этом безусловно выводимых на прасев.-кой. уровень, т. к. в южном пучке им регулярно соответствуют «обычные» альвеолярные (!), а в северном — «обычные» латеральные (||) кликсы. В рассматриваемых ниже диалектах ретрофлексные кликсы обнаруживаются в записях К. Доука в этимонах № 7, 10, 18, 28, 36, 47 (также восстанавливаются по соответствиям для № 41). Подробнее про реконструкцию ретрофлексных кликсов см. [Старостин 2005].

Относительно меньшей сложностью отличается система «обычных» (т. е. эктивных) согласных, восстанавливаемая для прасев.-койсанского. Ниже приводится подсистема эктивного консонантизма в *начальной* позиции:

лектах (как правило, утрачивающих назализацию; наиболее характерно такое упрощение для диалектов «северного» пучка).

¹ Этот ряд восстанавливается несколько условно, т. к. единственный диалект, в котором его существование отмечается эксплицитно — это эока !кунг (как в записях Т. Хейккинен, так и К. Кёниг). Тем не менее, из-за того, что между этим рядом и аналогичным ему в дальнородственном языке ɣхоан удается установить однозначные соответствия [Starostin 2008: 350], проекция его на прасев.-кой. уровень представляется оправданной.

(*p)	*t	*c	*č	*k
	*t ^h	*c ^h	*č ^h	*k ^h
(*b)	*d	*ʒ (*z)	*ʒ̥ (*ž)	*g
	*d ^h	*ʒ ^h	*ʒ̥ ^h	
		*cʔ	*čʔ	
		*ʒʔ	*ʒ̥ʔ	
		*s	*š	*x
*m	*r, (*n)			

Сразу отметим, что здесь и ниже нотация ʒʔ, ʒ̥ʔ и т. п. обозначает не «звонкий глоттализированный» тип артикуляции, фонетически невозможный по объективным причинам, а скорее «глоттализированный с предшествующим озвончением» (glottalized prevoiced), когда произнесение обычного глоттализированного согласного предваряется кратким интервалом вибрации голосовых связок, т. е. фонетически ʒʔ, ʒ̥ʔ как бы = ^dcʔ, ^dčʔ. Артикуляция обычных «звонких» согласных в некоторых койсанских языках также иногда описывается как «пред-озвонченная».

В начальной позиции зафиксированы и проецируются на прауровень также сложные сочетания согласных с веллярными фрикативными и аффрикатами: *tx, *cx, *čx, *tkx, *ckx, *čkx, *dx, *ʒx, *šx, *ʒkx, *škx.

Бросается в глаза практически полное отсутствие губных согласных: начальный сонант *m в исконно сев.-койсанской лексике встречается не более чем в пяти-шести случаях, начальный звонкий смычный *b — только в «детском» слове *ba 'отец', начальный глухой *p практически отсутствует как таковой. «Активного» запрета на губную артикуляцию, впрочем, нет, т. к. все эти согласные регулярно встречаются в заимствованиях из языков банту и европейских языков (английский, африкаанс, немецкий). Забегая вперед, отметим, что такая картина характерна для всех «южно-африканских койсанских» языков (небольшой подскок в частотности *m- наблюдается лишь в центр.-койсанской семье).

В аффрикатных и фрикативных рядах характерной сев.-койсанской чертой является последовательное противопоставление свистящих и шипящих (хотя в конкретных диалектах оно

довольно часто нейтрализуется в пользу того или другого ряда), а также наличие глоттализированных аффрикат, что в системном плане сближает их с щелчковыми рядами (для неаффрикат глоттализация несвойственна).

В *нена начальной* позиции консонантный инвентарь сев.-кой. языков намного беднее, т. к. на структуру морфемы накладываются строгие фонотактические ограничения. В срединной (интервокальной) позиции в словах исконного происхождения допустимы только сонорные *-m-*, *-n-*, *-r-* и смычный *-b-* (впрочем, в зависимости от конкретного диалекта последний может фонетически реализовываться и как глайд *-w-*, так что все эти возможные варианты допустимо рассматривать как проявления «правила сонорности нена начального согласного»). В конечной позиции встречаются только носовые сонорные *-m*, *-n*, а также *-ŋ*, хотя последний явно «дефективен»: не встречаясь ни в начальной, ни в срединной позициях, веларный носовой и в конечной позиции испытывает постоянную тенденцию к переходу в назализацию гласного (ср. эока-!кунг $\llbracket \tilde{a}ŋ \rrbracket$ 'подбородок' = жу|хоан $\llbracket \tilde{a}ŋ \rrbracket$ id.).

Вокалическая система сев.-кой. языков образует стандартную трапецию (**a*, **e*, **i*, **o*, **u*) с явным «перекосом» в сторону передних гласных: *e* и *i* встречаются довольно редко. В типологическом плане такой баланс сближает сев.-кой. систему с аналогичными системами и в других «южноафриканско-койсанских» семьях. Некоторые исследователи (в первую очередь Л. Ллойд и Д. Блик) отмечают противопоставление в этой семье открытых/закрытых (т. е., в современной терминологии, +ATR/-ATR) *e* : *ɛ*, *o* : *ɔ*, но оно, по-видимому, нефонологично. Широко распространены дифтонгические сочетания *ai*, *ae*, *ao*, *ai*, *oa*, *oe*, *ia*, *ui* (сочетания *ei*, *eu* в записях жу|хоан у Снеймана — фонетические варианты *ai* и *au*).

Гласные в сев.-кой. диалектах характеризуются также целым рядом дополнительных фонетических признаков. Это прежде всего назализация и фарингализация, которые не только обладают самостоятельной фонологической значимостью, но даже могут оказаться одновременно определенными на одном и том же гласном. В двухморных последовательностях конечные гласные и сонорные могут также отделяться от первой моры гортанной смычкой, наличие которой строго фонологично (ср. минимальную пару в жу|хоан: *tâm* 'не знать' : *tâʔm* 'чувствовать').

Сложнее обстоит дело с «шепотной» (breathy) артикуляцией отдельных гласных, которую П. Дикенс систематически отмечает для жу|хоан, а К. Кёниг и Б. Хайне — для экока-!кунг. Несмотря на то, что статистическая корреляция между этими данными, безусловно, существует (ср. жу|хоан t_i^h 'тяжелый' = экока t_i^h id. и мн. др.), т. е. признак «шепотности» реконструируем на уровне, нельзя также игнорировать явную связь между шепотной артикуляцией и низкими тонами: «шепот» возникает только в слогах низкого (реже) и сверхнизкого (чаще; см. ниже) регистров. Вопрос о том, насколько «взаимообязующей» является эта связь, равно как и о том, какая из характеристик в историческом плане первична («вокалическое придыхание» или низкий тональный регистр), нуждается в отдельном исследовании.

Для большинства сев.-кой. языков отмечена система из пяти регистровых тонов, в которой высокочастотны высокий, средний и низкий; реже встречается сверхнизкий и совсем редко — сверхвысокий. Контурные тоны возникают на стыке двух мор (см. ниже). Тональная реконструкция, тем не менее, существенно затрудняется запутанностью соответствий в целом ряде случаев, а также неточностью транскрипционной записи в большинстве работ (новейшие исследования, проводимые с использованием современной аппаратуры, показывают, что многочисленные ошибки в тональной нотации допускали даже П. Дикенс и Ян Снейман, не говоря уже о менее профессиональных исследователях). В приводимых ниже реконструкциях прасев.-кой. морфем тоны проставляются только там, где соответствия по конкретным диалектам более или менее однозначны. Следует отметить, что для жцх П. Дикенс не отмечает существование среднего тона, почти всегда записывая на его месте высокий; остается неясным, действительно ли в этом языке произошло совпадение двух тонов в один или же речь идет всего лишь о неточности записи.

Структура слога и морфемы в сев.-кой. языках подчиняется довольно строгим закономерностям. Поскольку, за незначительными исключениями, закономерности эти актуальны и для всех остальных языковых групп «южноафриканско-койсанского» ареала, на их описании имеет смысл остановиться подробнее.

В наиболее общем случае элементарная сев.-кой. словоформа, как именная, так и глагольная, складывается из двух мор, причем

выбор фонем внутри первой из них оказывается достаточно свободным, на структуру же второй накладываются очень жесткие ограничения. Поскольку такая комбинаторика чрезвычайно схожа с устройством сложных морфологических структур в языках мира («фонетически свободный» корень + «фонетически лимитированные» аффиксы), первую мору в таких сочетаниях можно обозначить как «условно-корневую», вторую — как «условно-суффиксальную»; «условно» потому, что и на синхронном, и, как правило, даже на «мелко-диахроническом» уровне вторая мора неразрывно связана с корнем и в лучшем случае может представлять собой «окаменевший» суффикс. Кое-где, впрочем, даже в пределах одного диалекта сохраняются реликтовые варианты одной основы, подтверждающие историческую оправданность такой терминологии — ср., например, в жу|хоан формы $\text{f}^h\text{íí}$ и $\text{f}^h\text{óá}$ 'собака', восходящие к $*\text{f}^h\text{o-í}$ и $*\text{f}^h\text{o-a}$ соответственно (где «чистый» корень — $*\text{f}^h\text{o-}$).

«Условно-корневая» мора имеет структуру CV-, где C может быть любым из перечисленных выше кликсов, «обычных» согласных или консонантных сочетаний с велярными аффрикатами и фрикативными, а V — любым простым гласным (не дифтонгом).

«Условно-суффиксальная» мора имеет структуру (R)V, где V — опять-таки любой простой гласный (не дифтонг) или «вокализированный» сонорный согласный *m, n, ŋ* («вокализация», в частности, означает, что на таких сонорных может быть определен самостоятельный тон); в структурах вида -RV первый элемент — сонорный *r, m, n*, или *w ~ b* (см. выше), второй — простой гласный (*a, i, u, e, o*).

Комбинирование обеих мор происходит следующим образом:

(а) Если гласный «условно-корневой» и «условно-суффиксальной» моры совпадают по качеству, они сливаются в единый гласный, количество которого в сев.-кой. языках не вполне ясно. В таких «южноафриканско-койсанских» языках, как !хонг и нама, гласный в словоформах со структурой CV почти всегда произносится долго, что, на наш взгляд, отражает их происхождение из двухморных $*\text{CV}_1\text{V}_1$; для сев.-кой. языков, однако, долгота в известных описаниях обычно не указывается даже как чисто фонетическая характеристика. Таким образом, например, жу|хоан $\text{f}^h\text{ò}$ 'быть жадным' в принципе допустимо описывать и как

одноморную последовательность, и как стяжение из двухморного $[\underline{d}-\delta \rightarrow \underline{d}:\rightarrow \underline{d}:\delta]$. Отметим при этом, что такого рода «одноморных» словоформ в сев.-кой. языках в принципе очень мало, и в целях унификации описания их может быть удобнее описывать как стяжения из двухморных.

(б) Если первый и второй гласный различаются по качеству, они образуют дифтонгические сочетания: жу|хоан ʃàì 'жираф' ← $\text{ʃà-} + \text{-ì}$. Реально зафиксированными оказываются при этом не все комбинации, скорее всего потому, что при образовании таких сочетаний в отдельных случаях имела место ассимиляция: так, недопустимы дифтонги $*ie$, $*io$ (на их месте обнаруживаем ассимилированные oe , $o \leftarrow oo$), а также любые сочетания с первым элементом e или i (т. е. $*ea$, $*ia$, $*eo$, $*io$ и т. п.); впрочем, следует помнить, что передние гласные e , i вообще имеют в сев.-кой. языках очень ограниченную дистрибуцию.

(в) В биконсонантных сочетаниях структуры CVRV первая мора может слегка «растягиваться» за счет ассимилятивных процессов. Так, типичной для большинства сев.-кой. диалектов является трансформация структуры $*Copa \rightarrow Coana$ (ср. жу|хоан ʃóáiná 'нести на плече'); при этом второй элемент новообразованного дифтонга — скорее «призвук», чем самостоятельный морообразующий гласный: в экака-!кунг на этом месте обнаруживается либо эксплицитно маркированная редукция (ʃòá'nà в [Heikkinen 1986]), либо, вместо дифтонгизации гласного, удвоение сонанта (ʃòp'nà в [König & Heine 2008]).

(г) На каждой из мор определена самостоятельная тональная характеристика. Если тоны обеих мор совпадают, вся форма произносится в одном регистре: ср. жу|хоан !kxáí 'нога', ʃʰòm 'встать на колени'¹, !hári 'тень'. Если они различаются, возникают контурные тоны: ʃkxáá (= ʃkxá) 'глина, грязь', ʃʰìì 'мышь', !àbèʰ 'выкапывать (зверя из норы)'. При этом допустимыми являются далеко не все тональные комбинации; по-видимому, уже на прасев.-кой.-уровне в каком-то объеме имела место аккомодация тонов, реконструкция которой, впрочем, выходит за рамки нашего исследования.

¹ В закрытых слогах, оканчивающихся на носовой сонорный, как в сев.-кой., так и в юж.-кой. языках тон сонанта, как правило, совпадает с тоном гласного и поэтому не обозначается специальной диакритикой.

(д) Дополнительные вокалические признаки назализации и фарингализации, по-видимому, изначально определены на каждой из мор в отдельности, но при образовании долгих гласных или дифтонгических сочетаний распространяются на весь слог в целом: ср. жу|хоан *táá* 'сердиться', *kòé* 'быть в безопасности'. Поскольку фонологических оппозиций типа $C\tilde{V}_1V_2 : CV_1\tilde{V}_2$, $C\tilde{V}_iV_2 : CV_1V_2$ быть не может, в транскрипционном отношении принципиально отмечать назализацию и фарингализацию только на первом гласном дифтонга.

Любопытно, что в структурах $CVRV$ первый гласный может быть фарингализованным (жу|хоан *fòàrà* ← **fòrà* 'ворона'), но не назализованным. Это позволяет поставить вопрос о возможном развитии назальной артикуляции из бывшего сегментного компонента, скорее всего — **n* (т. е. $C\tilde{V}$ ← * CVn , $C\tilde{V}_i\tilde{V}_2$ ← CV_1nV_2): такие ослабления носовой артикуляции иногда удается зафиксировать даже на синхронном уровне в пределах одного диалекта, например, в редупликациях: ср. жу|хоан *lã̀-|ani* 'трясти (дерево)' ← **|ani-|ani*, *šã̀šã̀ni* 'потирать руки' ← **šani-šani* и др. Тем не менее, назализация как фонологически дистинктивный признак никоим образом не снимается на прасев.-кой. уровне; вопрос о ее возможных источниках не может серьезно обсуждаться без выхода на уровень внешнего сравнения.

Двухморные структуры, состоящие из «условно-корневого» и «условно-суффиксального» элементов, можно обозначать как «первичные». Подавляющее большинство сев.-кой. словоформ, состоящих более чем из двух мор, морфологически производны даже на синхронном уровне отдельных диалектов. Это либо словоизменение (например, формы мн. ч.: жу|хоан *là̀* 'буйвол', мн. ч. *là̀-si*), либо словообразование, обычно представленное словосложением: ср. жу|хоан *kà̀-sí* 'крышка (горшка)', из *kà̀* 'горшок' + *sí* 'рот' и мн. др.

Морфологические аспекты сев.-кой. группы в целом незначимы для корневой реконструкции: суффиксальная деривация и парадигматика развиты слабо, глагольное словоизменение почти отсутствует, а именовое ограничивается лишь несколькими продуктивными способами образования множественного числа, не сопровождаемого (за редкими исключениями) нетривиальными морфонологическими процессами. В будущем задачи сев.-кой.

реконструкции, безусловно, входит решение вопроса об именных классах, которые в этих диалектах в основном выражаются на уровне анафорической референции (т. е. разным типам существительных соответствуют разные виды указательных местоимений); не исключено, что какая-то информация о классах первоначально кодировалась и в «условно-суффиксальных» морфемах первичных основ (как это до сих пор имеет место в ряде южнокойсанских языков, см. раздел 2.3.2).

Базисная лексика сев.-кой. языков в целом устойчива к заимствованиям, в отличие от периферийной, где активными донорами служат центральнокойсанские (особенно нама), бантусские и европейские языки (английский, африкаанс, немецкий). По крайней мере в жцх (возможно, также и в других диалектах, но проверить это трудно из-за нехватки данных) ощущается сильное влияние нама и на лексическое «ядро»; так, например, для большого количества названий частей тела в словаре [Dickens 1994] зафиксированы дублетные формы — одна «исконная», одна явно заимствованная из центр.-койсанского источника — без указаний на статистику их употребления (напр., 'сердце': !kxá / ǀáá, 'рот': cǀí / kǀáǀí и т. п.). Анализ контекстов в [Snyman 1970] и других источниках по жцх показывает, что в повседневном употреблении все же употребляются в основном исконные слова, так что словарные единицы типа ǀáá и kǀáǀí, скорее всего, представляют собой либо стилистически маркированные синонимы, либо особого рода «жаргонизмы» (что, впрочем, не означает, что любое заимствование из нама в жцх или в другие сев.-кой. диалекты должно по умолчанию проходить по разряду жаргонизмов).

2.1.3. *Лексикостатистика.* Подсчеты по полному 100-словному списку (см. таблицу 1) показывают, что минимальный процент совпадений наблюдается между ||ay||ен и эока !кунг (78% совпадений, что примерно соответствует распаду ок. III в. н. э.). Это означает, что в лексическом плане «северный» пучок диалектов противопоставлен более близким друг к другу центральному и южному пучкам, что в целом согласуется и с попытками классифицировать сев.-кой. диалекты по фонетическим инновациям: северные наречия в таких классификациях оказываются более изменчивыми (например, утрачивают фонологическую оппозицию свистящих и шипящих аффрикат и фрикативных).

Что касается способности лексикостатистики корректно разграничить центральные и южные диалекты, то ее приходится оценивать как ограниченную — как в силу технических причин (для таких старых источников, как [Doke 1925] и др., нельзя поручиться за семантическую точность всех приводимых английских эквивалентов), так и «сущностных» (тесная близость обоих пучков приводит к тому, что большинство процентных расхождений оказываются в рамках статистической погрешности). Так, диалект $\|ay\|en$ в записи Д. Блик обнаруживает меньший процент сходжений с жцх (86%), чем с хрутфонтейн-!кунг в записи К. Доука (92%), несмотря на то, что во всех «фонетически ориентированных» классификациях $\|ay\|en$ относится к южному, а не к центральному пучку. С другой стороны, результаты подсчетов легко могли оказаться искажены еще и за счет того, что по материалам Доука удалось набрать всего чуть более 70 потенциальных эквивалентов сводешевских значений. Не исключено, что статистические подсчеты по всему объему материала (по отношению к работе Доука — задача вполне реальная) в будущем позволят значительно уточнить общую картину.

Таблица 1.

Лексикостатистическая матрица севернойкойсанских языков
(100-словные списки).

	$\ Ay\ en$!Кунг	!О!кунг	Экока
Жу хоан	0.86	0.87	0.78	0.79
$\ Ay\ en$		0.92	0.82	0.78
!Кунг			0.83	0.87
!О!кунг				0.80

50-словный список для севернойкойсанских языков.

1) «пепел»: (а) жцх $t\dot{u}$, aye $t\dot{u}$; (б) окн $\dot{t}\dot{u}a$; (в) (?) эко $\|d^h\dot{a}$.

Начальные согласные в окн и эко не соответствуют друг другу, так что слова следует формально считать относящимися к разным корням; с другой стороны, речь может идти и об ошибочной записи, тем более что в эко отмечена также синонимичная форма

dàʔà *!!ðʰà*, где *dàʔà* = 'огонь' (см. ниже), а *!!ðʰà* = окн *ʔdà* (эко *!!* — регулярное отражение прасев.-кой. *ʔ).

Общность форм (б) и (в) тем более вероятна, что оба они похожи на относительно недавние заимствования из центрально-койсанского источника (см. центр.-кой. **ʔdà* 'пепел'); косвенным аргументом в пользу этого является присутствие этого же слова в жцх (*ʔdàʰ*) в значении 'мыло' (т. е. «культурный» продукт, для изготовления которого используется зола).

В этом случае единственным серьезным претендентом на прасев.-кой. статус оказывается праформа **tɔ*; следы ее присутствия в северном диалектном пучке обнаруживаются в редуцированной форме *t̪d̪o-t̪d̪o* в т. н. «ангольском!кунг» [Snyman 1980: 33]), хотя вокалические соответствия с жцх не вполне понятны¹.

2) «птица»: жцх *sàt̪à*, ауе *sata*, кнг *sāvà*, окн *sata*, эко *čát̪à*. || Прасев.-кой. **sata*.

Второй слог на синхронном уровне можно вычленить как отдельную морфему только в эко, ср. мн. ч. *čá-t̪hè*, указывающую на старый суффикс (общесев.-кой. уменьшительная морфема, в эко имеющая вид ед. ч. *-t̪à*, мн. ч. *-t̪hè*). Старый корень, таким образом, мог иметь вид **sá-* или **sát̪* (со стяжением **sát̪-ma* → **sata*). В связи с этим ср. в эко также отдельно форму *čát̪* 'птица', хотя это может быть и вторичной семантической контаминацией с прасев.-кой. **čát̪* 'вид птицы' (жцх *čát̪* 'сизоворонка (*Coracias garrulus*)' и др.).

3) «черный»: жцх *žó*, ауе *žɔ*, окн *žo ~ žu*, эко *žō*. || Прасев.-кой. **žō*.

4) «кровь»: (а) жцх *ʔán*, ауе *ʔi*; (б) окн *yalo ~ yalu*; (в) эко *ʔlú ~ ʔlú*.

Окн *yalo* — необычная форма (начальный *y-* для этого диалекта, равно как и для койсанских языков вообще, почти уникален), хотя явных источников заимствования для нее не обнаружено. Эко *ʔlú* также не имеет никакой этимологии за пределами северного диалектного пучка, хотя внутри этого пучка основа распространена широко (для ряда диалектов Снейман приводит (а) и (б)

¹ Имеет смысл упомянуть также особую «дублетную» форму: жцх *ʔáú* 'пепел (сигаретный)' — несмотря на фонетическое и семантическое сходство с **tɔ*, это слово можно рассматривать только как специализированное заимствование из центр.-кой. источника (см. ниже центр.-кой. **ʔai* 'пепел').

как синонимы — напр., оконго /*ɬaŋ* и //*ɔrù* — без объяснения семантических различий).

«Оптимальной» прасосновой, исходя из общей дистрибуции рефлексов, следует считать сев.-кой. */*ɬ(a)ŋ* как форму, наблюдаемую в подавляющем большинстве диалектов и представленную во всех трех пучках.

Отдельный вопрос — возможное отражение этой основы в ряде диалектов северного пучка с лабиальным ауслаутом, обнаруживаемое в обзорных списках Снеймана: /*ãt̪m̪* ~ /*ãt̪* (мпунгуфлей, кубанго и др.). Учитывая, однако, что, помимо ауслаута, эти основы различаются еще и исходом кликса (гортанная смычка в */*ɬ(a)ŋ*, нулевой исход в /*ãt̪m̪*), уместнее все же считать, что речь идет о двух разных основах, хотя отдельной этимологии для */*at̪* пока не обнаружено.

5) «кость»: жцх эко /*hú*, ауе /*hu*, кнг /*hú*; окн /*h* ~ /*ʔ*. || Прасев.-кой. */*hú*.

6) «ноготь»: жцх /*húʔurú*, ауе //*uru*, кнг //*hú*, окн //*ulu* ~ //*ɲu*, эко /*hú*. || Прасев.-кой. */*huʔru*.

Исходный ретрофлексный кликс сохраняется в кнг, и архаичность именно этого типа артикуляции подтверждается также развитием его в латеральный // в ряде северных диалектов (и даже в «южном» ауе, хотя здесь может на самом деле отражаться какое-то диалектное смешение, ср. ниже «разнобой» в рефлексах слова 'мясо'). Неясной остается рефлексация в эко, где регулярным отражением прасев.-кой. */*h* также должен быть латеральный //. Можно предложить два альтернативных объяснения нерегулярности: (а) фонетико-семантическая контаминация с квазиомонимом /*ɔlú* 'колчан' ('ноготь' как 'футляр'); (б) более вероятно — диссимилиация последовательности из двух ретрофлексных кликсов в составе композита //*ãɔ̄-ú* ← прасев.-кой. */*h* ~ /*ʔ*, букв. 'руки-ноготь'.

7) «умирать»: жцх /*hí*, ауе /*h* ~ /*ʔ*, кнг /*hí*, окн /*h* ~ /*ʔ*, эко //*h* ~ //*ʔ*. || Прасев.-кой. */*h*.

Соответствия начальных согласных, как и в № 6, 10, 17, указывают на прасев.-кой. «ретрофлексный» кликс, хотя ауе /*h* (*i*) имеет нерегулярный рефлекс (впрочем, нерегулярность может отражать ошибку в транскрипции Д. Блик или «субдиалектный» вариант, см. ниже 'мясо'). Стоит отметить также супплетивные

формы с тем же значением при субъекте мн. ч.: жцх !àð, ауе ʃau, окн !au 'умирать (о многих), вымирать' ← прасев.-кой. *!làð (этимологическая связь с *!áí возможна, но не очевидна).

8) «собака»: жцх ʃ^húí, ауе !ɔ, кнг ʃ^húí:, окн ʃúé, эко ʃ^hòē. || Прасев.-кой. *ʃ^ho-e ~ *ʃ^ho-í.

Колебания в артикуляции ауслатного гласного, по-видимому, отражают различия в морфологическом оформлении этой основы на «до-прасев.-койсанском» уровне. В жцх, помимо прочего, отмечен еще альтернативный вариант ʃ^hóá, с еще одним вариантом суффикса.

Форма !ɔ в ауе допустима к включению в эту же этимологию, т. к. в записях Д. Блик палатальный клик очень часто транскрибируется как альвеолярный, а назализация спорадически остается неотмеченной.

9) «пить»: жцх ʃ^hí, ауе ʃí, кнг ʃí:, окн ʃí, эко ʃí. || Прасев.-кой. *ʃ^hiN (?).

Все формы явно родственны, но реконструкция конечной моры затрудняется из-за неожиданного исчезновения назальной артикуляции в диалектах южного пучка (ср. очень похожий случай, где назализация не пропадает: жцх ʃ^hí, кнг ʃí:, эко ʃí 'печень'). Без детальной реконструкции можно предложить два варианта: (а) реконструировать минимальную пару 'пить' : 'печень' в виде *ʃ^hiN : *ʃ^hiŋ (именно в таком виде эта оппозиция выглядит в кнг Доука); (б) объяснять различия в рефлексации влиянием супrasegmentных признаков (для 'пить' характерен низкий или сверхнизкий тональный регистр, для 'печени' — высокий). В любом случае, носовой элемент должен быть каким-то образом отражен даже в предварительной реконструкции.

10) «сухой»: жцх !kǎí, кнг !ʔau, эко !kǎō. || Прасев.-кой. *!kǎu.

11) «ухо»: жцх, эко ʔ^húí, ауе !ui, кнг ʔ^húí, окн !úí. || Прасев.-кой. *ʔ^húí.

12) «есть»: жцх ʔ^hí, ауе t: ~ t, кнг ʔ^hí:, окн t ~ t, эко t. || Прасев.-кой. *ʔ^hí. Уникальный случай надежной реконструкции слогового губного сонанта на прасев.-кой. уровне.

13) «яйцо»: жцх !í, ауе !í:, окн !í ~ !í, эко !íy ~ !íy. || Прасев.-кой. *!í.

14) «глаз»: жцх !àʔá, ауе !a, кнг !əʔā ~ !āʔā ~ !ā, окн !a, эко !àʔā. || Прасев.-кой. *!aʔa (двухморная структура, прерываемая гортанной смычкой).

15) «огонь»: жцх *dàʔá*, ауе *dà*, кнг *dəʔā ~ dāʔā*, окн *dà ~ dà:*, эко *dàʔà*. || Прасев.-кой. **daʔa* (морная структура такая же, как в 'глазе', см. выше).

16) «нога»: жцх *|kái*, ауе *|e ~ |xe:*, кнг *|áí ~ |xái*, окн *|kxe ~ |kxe*, эко *|kái*. || Прасев.-кой. **|kái*.

Развитие велярной аффрикаты *-kx- в гортанную смычку в кнг не вполне регулярно, но отмечено и в других случаях (см. выше 'сухой') и может считаться свободным вариантом; напротив, отсутствие ее в ауе скорее следует списывать на небрежность транскрипции.

17) «волосы»: жцх *!kúí*, ауе *!kxi ~ !kxie ~ !ui*, кнг *!kúí*, окн *!ui ~ kxi*, эко *!kúí*. || Прасев.-кой. **!kúí*.

18) «рука»: жцх *!áú*, ауе *!au ~ !ou*, кнг *!áú*, окн *!au*, эко *!àð ~ !āð*. || Прасев.-кой. **!au*. Расщепление ретрофлексного кликса по диалектам здесь регулярно: альвеолярный рефлекс в южном пучке, латеральный в северном.

19) «голова»: жцх *!ái*, ауе *!e ~ !é ~ !i ~ !í*, кнг *!é:*, окн *!é*, эко *ʔé*. || Прасев.-кой. **ʔé* (с «преглоттализированным» носовым кликсом).

20) «слышать»: жцх *sàʔá*, ауе *sa ~ ča*, кнг *səʔá*, окн *sa: ~ sá:*, эко *čà ~ čàʔá*. || Прасев.-кой. **saʔa*.

Колебания между аффрикатами и спирантами, характерные для этой основы, для сев.-кой. языков в целом нетипичны. Здесь проще предположить вторичность аффрикатного рефлекса, вызванного, скорее всего, регрессивной ассимиляцией с гортанным смычным.

21) «сердце»: жцх *!ká*, ауе *!a*, кнг *!ʔa*, окн *kxa*, эко *!kā*. || Прасев.-кой. **!kxa*. В окн у Д. Блик зафиксирован такой же диалектный вариант с отпадением основы кликса, как и в 'волосах' (см. выше).

22) «рог»: жцх *!ʰú*, ауе *!u: ~ !ú*, кнг *!ú*, окн *!ʰú*, эко *!ʰú*. || Прасев.-кой. **!ʰu*.

Ретрофлексный кликс в записи К. Доука, не подтверждаемый сопоставительными данными Снеймана, по-видимому, ошибочен; на праязыковом уровне слово следует восстанавливать с «обычным» альвеолярным кликсом, о чем свидетельствуют данные всех остальных диалектов.

23) «я»: жцх *tí*, ауе *t ~ te ~ tí ~ ти*, кнг *t̄ ~ tí*, окн *t ~ te ~ tí*, эко *tí ~ t̄ā*. || Исходен, по-видимому, наиболее частотный вариант **tí*; остальные варианты вызваны редукцией, внутрифразо-

выми ассимиляциями или слиянием с какими-то эмфатическими частицами.

Стоит отметить также «альтернативную» основу местоимения 1-го л., чаще всего функционирующую в качестве притяжательной или косвенно объектной. В кнг она встречается в первую очередь в записях Л. Ллойд, где она в зависимости от контекста имеет вид *ŋ* (ср.: *ŋ a !ota* 'я низкого роста', *ŋ a !úí gu a* 'я не подниму тебя') или *n* (*n !wonní*: 'мой локоть'). Скорее всего, она же представлена и в таких формах из словарного списка К. Доука, как *ŋ=fo* 'брат' (наряду с *fo id.*, т. е. 'мой брат'), *ŋ=ʔhíí*: 'собака' (наряду с *ʔhíí id.*, т. е. 'моя собака') и ряд других.

В жцх прямых аналогов этой «велярно-дентальной» носовой основе не обнаружено; однако и в жцх, и в ряде диалектов северного пучка (эко, а также т. н. «ангольский !кунг» в [Snyman 1980]) существует основа *ná*, которая выполняет двоякую функцию: (а) императивно-предикативную, в значении 'дай мне (что-л.)'; (б) дативную, в значении 'мне', т. е. является косвенным объектом при бенефактивных глаголах (ср. жцх *!pama na šoro* 'купи мне табака').

В сегментном (но не функциональном!) плане эта основа тождественна морфеме *na* в кнг, которую К. Доук определяет как 'приглагольный формант 1-го л. ед. ч.' (при том, что глагольное спряжение, и тем более изменение по лицам, в сев.-кой. языках, строго говоря, отсутствует), приводя такие примеры, как *nā=č'í*: 'я мокрый', *nā=tám* 'я не знаю', *nā=!híí*: *!híí*: 'я убиваю бушмена' и др.). Функционально, однако, формант *na=* в хрутфонтейн-!кунг Доука скорее аналогичен цепочке *ŋ-a-* в !кунг Ллойд (см. выше), которую Д. Блик анализирует как сочетание из *ŋ* 'я' + вспомогательный глагол-связка *a* 'быть, являться'. Параллельно с этим, впрочем, Доук отмечает для кнг и существование глагола *na*: 'давать, приносить', скорее всего, выполняющего те же функции, что и в эко (т. е. не просто 'давать', а конкретное 'дай мне').

Связать все эти факты воедино довольно затруднительно. По-видимому, на прасев.-кой. уровень следует все же выводить по крайней мере основу **ŋ*, возможно, в качестве «реликтовой»: неустойчивость этого согласного в начальной позиции (отметим, что сама по себе фонема *ŋ* в сев.-кой. диалектах присутствует, но исключительно на конце слова, а в жцх и ряде других наречий пропадает и там) приводит к развитию **ŋ-a-* → **n-a* в сочетании

‘дай мне’ уже на прасев.-кой. уровне, но в некоторых диалектах велярная артикуляция еще «задерживается» в других синтаксических контекстах — например, в архаичных высокочастотных притяжательных конструкциях, записанных Доуком для кнг.

Что касается функционального соотношения основ $*m(i)$ и $*ŋ$, то ответ на этот вопрос следует искать уже на более глубоком уровне сравнения; на данном этапе можно лишь констатировать, что лексикостатистической значимостью в прасев.-кой. списке из этих двух морфем будет обладать только $*m(i)$, как наиболее надежно реконструируемая и функционально неограниченная основа.

24) «убивать»: жцх $!^h\acute{u}$, ауе $!ũ$, кнг $!^h\acute{u}$, окн $!ũ \sim !x\acute{u}$, эко $!^h\acute{u}ŋ \sim !^h\acute{u}$. || Прасев.-кой. $*!^h\acute{u}$ (возможно, $*!^h\acute{u}ŋ$ в связи с вариацией в эко).

25) «лист»: (а) жцх $d\grave{o}r\grave{a}$, ауе $d\grave{o}r\grave{a}$; (б) кнг $\acute{h}\acute{u}b\acute{u}$, окн $\acute{h}\acute{o}b\acute{a} \sim \acute{h}\acute{o}v\acute{a}$; (в) окн $g\grave{o}a$; эко $\acute{h}\acute{o}$.

К корням (б) и (в) ср. в жцх соответственно (б) $\acute{h}\acute{u}^?b\acute{u}$ ‘распускаться (о листьях)’ и (в) $\acute{h}\acute{o}$ ‘влажный лист’. Основа (а) не имеет по диалектам никаких других значений и поэтому могла бы считаться первичной, однако см. ниже обсуждение этого же этимона в центр.-кой. языке наро (№25 в разделе 2.4).

По-видимому, наибольшие шансы на архаичность в данной ситуации имеет все же основа (в) (значение ‘влажный лист’ в жцх можно рассматривать как вторичное сужение). Помимо эко и жцх, она встречается в записях кнг у Л. Ллойд, где в ряде контекстов переводится просто как ‘лист’, ср.: $\acute{h}\acute{a} \text{ ti } m\acute{a} \text{ !}\acute{a}\acute{u} \text{ } \acute{h}\acute{w}\acute{a}\text{-}s\acute{u}ŋ$ ‘ветер колышет листья на дереве’. В окн вариант $g\grave{o}a$: относится к говору, вторично утратившему альвеолярные кликсы (см. выше ‘волосы’).

26) «вошь»: жцх $s\acute{h}$, эко $\acute{h}\acute{o}ŋ$ (в прочих рассматриваемых диалектах слово не зафиксировано). || Прасев.-кой. $*s(i)ŋ$.

В жцх зафиксирована также основа $\acute{h}\acute{a}$ с тем же значением, но ближайшие параллели к ней также обнаруживаются только в центральнокойсанском языке наро (№26 в разделе 2.4), что говорит об ареальном («бродячем») характере основы.

27) «мясо»: жцх $!^h\acute{a}$, ауе $!^h\acute{a} \sim !^h\acute{a} \sim \acute{h}\acute{a}$; кнг $!^h\acute{a} \sim \acute{h}\acute{a}$; окн $!^h\acute{a}$, эко $!^h\acute{a}$. || Прасев.-кой. $*!^h\acute{a}$ (с ретрофлексным кликсом; в кнг Доука архаичный вариант не зафиксирован, но колебание альвеолярных и латеральных рефлексов по диалектам и говорам однозначно определяет реконструкцию).

28) «луна»: жцх *lúí*, ауе *lwi*, кнг *ll̥ī*, окн *ll̥wi ~ ll̥we*, эко *ll̥íí*. || Прасев.-кой. **ll̥íí* (ретрофлексный рефлекс сохраняется в кнг).

29) «рот»: жцх *cī*, ауе *ci ~ cī*, кнг *cī*, окн *ci ~ cī*, эко *čí*. || Прасев.-кой. **cī* (или **cī*, если сверхвысокий тон в жцх действительно архаичен).

30) «имя»: жцх *!ú*, ауе *!ú ~ !^hú*, кнг *!ú*, окн *!ú*, эко *!ú*. || Прасев.-кой. **!ú*; дентальный кликс в записи Доука, скорее всего, ошибочен (диалектные варианты с озвончением альвеолярного кликса в старых записях иногда встречаются, но дентальная артикуляция не подтверждается ни в одном из имеющихся источников).

31) «новый»: жцх *zé*, ауе *ze*, кнг *zē*, окн *ze*, эко *žə^hè*. || Прасев.-кой. **ze* (необычно появление в эко придыхательной артикуляции гласного, но вряд ли это может являться поводом для разделения корней).

32) «ночь»: жцх *!ú*, ауе *!u ~ !u:* ~ *!ú*, кнг *!ú*, окн *!ú*, эко *!ú*. || Прасев.-кой. **!ú*.

33) «нос»: жцх *cī*, ауе *čí*, кнг *cī*, окн *ciŋ ~ cáŋ ~ čŋ*, эко *čkxáŋ ~ čŋ*.

Здесь особенно показательна форма в эко — единственным удовлетворительно описанным из сев.-кой. диалектов, в которых обычные глоттализированные аффрикаты (прасев.-кой. **c̥* и **č̥* → эко *č*) последовательно отличаются от начальных сочетаний «переднеязычная аффриката + веллярная глоттализованная аффриката» (прасев.-кой. **ckx* и **čkx* → эко *čkx*). Таким образом, прасев.-кой. форма должна восстанавливаться в виде **ckx(u)ŋ* (реконструкция вокализма пока остается неясной — не исключен, в числе прочих, и вариант слогового **ŋ*).

34) «не»: жцх *!óá*, ауе *!wa ~ !ua:* ~ *!á*, окн *!wa ~ !wí ~ !we ~ kwé ~ kwí*, эко *!ōā*. || Прасев.-кой. **!oa*.

Я. Снейман отмечает в жцх не менее шести различных вариантов отрицательной частицы, стоящей в препозиции к глаголу: *!wa* (= *!oa*), *!wí* (= *!uí*), *!eu* (= *!au*), *!a*, *!ao*, *!u* [Snyman 1970: 154]. Это до некоторой степени согласуется с аналогичным разбросом вариантов в окн Д. Блик, но во всех остальных источниках, как правило, приводится только один вариант (включая описание в подробном грамматическом очерке эко [König & Heine 2001: 43]). Скорее всего, речь идет о каких-то индивидуальных произносительных особенностях или элементах сандхи, чем о составном ха-

рактуре таких вариантов (слияние с видо-временными или эмфатическими частицами и т. п.).

35) «один»: жцх *ǰéʼé*, ауе *ǰe ~ ǰéé*, кнг *ǰǰé*, окн *ǰé*, эко *ǰéʔè ~ ǰé*. || Прасев.-кой. *ǰéʔe (удвоенное написание гласного в ауе варианте *ǰéé* также отражает его «разрывную» артикуляцию).

36) «дождь»: жцх *ǰà*, ауе *ǰà*, кнг *ǰǰà*, окн *ǰǰa ~ ǰà ~ ga:*, эко *ǰà*. || Прасев.-кой. *ǰà (с звонким ретрофлексным кликсом; назализация в кнг Доука не подтверждается другими свидетельствами).

37) «дым»: (а) жцх *šòrà*, ауе *šore ~ šori*, кнг *šò:ɽi*, эко *šúle*; (б) окн *ʃoni*. || Прасев.-кой. *šore ~ *šora (гласный второго слога однозначно не реконструируется; возможно морфологическое варьирование).

Форма (б) в окн могла быть заимствована из кхойкхой (ср. нама *ʃanni-s* и др.), хотя при этом ожидался бы гортанный, а не нулевой исход кликса. Старая основа при этом все равно сохраняется: Д. Блик записывает ее как *šule ~ čuli* 'табак' (полисемия 'дым' / 'табак' характерна и для большинства прочих форм из группы (а)).

38) «звезда»: жцх *ǰǰǰ^h*, ауе *ǰǰǰ*, кнг *ǰǰǰ*, окн *ǰǰ ~ ǰǰ ~ !ǰǰ*, эко *ǰǰǰ*. || Прасев.-кой. *ǰǰ^(h). «Шепотная» артикуляция гласного отмечена только в жцх и несомненно связана с надежно реконструируемым для этого слова сверхнизким тоном (ср. рефлексy в жцх и эко).

39) «камень»: жцх *ǰǰm*, ауе *ǰǰm*, кнг *ǰǰ:m*, окн *ǰǰm*, эко *ǰǰm ~ ǰǰm ~ ǰǰm*. || Прасев.-кой. *ǰǰm (в жцх *-ит и *-от регулярно нейтрализуются в -от).

40) «солнце»: (а) жцх *ǰǰm*, ауе *ǰǰm*, кнг *ǰǰ:m*, окн *ǰǰm*; (б) эко *gàð ~ gàʔð*. || Прасев.-кой. *ǰǰm.

Из рассмотрения нельзя исключать гипотезу, согласно которой формы группы (а) могли быть заимствованы (как по отдельности, так и на прасев.-кой. уровне) из центральнокойсанского источника (*ǰǰm 'солнце, день', см. ниже). Однако, кроме точного фонетического совпадения форм, дополнительных аргументов в пользу этой гипотезы не существует. При этом маловероятно, что эко *gàʔð* — архаизм; данная форма сопоставима с жцх глаголом *gàʔáró*, который П. Дикенс переводит как 'to drink too little to quench one's thirst', т. е., по сути, 'продолжать испытывать жажду', что явно указывает на первичность значения 'жажда' для соответствующей простой именной основы. (Полисемия 'солнце' / 'жажда' для койсанских, как и для целого ряда других африканских языков, является нормой; в этих условиях допустимо не только

развитие 'солнце' → 'жажда', но и обратное). Таким образом, серьезных «конкурентов» на данную позицию в 50-словнике у прасев.-кой. **!ám* не оказывается.

41) «хвост»: жцх *!xúí*, aye *!^hwí ~ ɸwí*, окн *!wé*, эко *!xōē*. || Прасев.-кой. **!xōē* (дифтонг *-*oe* реконструируется на основании соответствия «жцх *uí* : эко *oe*»).

42) «ты»: жцх *à*, aye *a-hí*, кнг *à*, окн *a ~ a-hí*, эко *à*. || Прасев.-кой. **a* (с эмфатическим вариантом **a-hí ~ *a-híj* в ряде диалектов, ср. также у Доука эмфатич. форму *à-híj*).

В грамматическом очерке эко [König & Heine 2001] отмечено также существование особой формы местоимения 2-го л. *bà*, употребляемой только в функции субъекта в неначальной позиции во фразе. Никаких аналогов этой формы в жцх не обнаружено, но, скорее всего, эко *bà* как-то связано с редкой формой *ta ~ tʔa* 'ты', зафиксированной в записях Л. Ллойд по кнг и также выступающей исключительно в функции субъекта, хотя и не обязательно в неначальной позиции (ср. такие примеры, как *tá síŋ a* 'ты посмотри на себя (= тебя)'; *e júí gu a, ta tʔa júí e daba !^{ho}* 'мы не возьмем тебя, ибо ты не маленький ребенок' [Bleek 1956: 132] — в обоих примерах субъект *ta ~ tʔa* 'ты' контрастирует с объектом *a* 'тебя').

Судя в первую очередь по форме *tʔa*, фонетическое устройство которой предполагает морфологическую разложимость, речь все же идет не об особой супплетивной основе, а о каком-то старом стяжении обычного местоимения 2-го л. **a* с предшествующей ему («субъектной») частицей. Вопрос, безусловно, нуждается в дальнейшем изучении, но пока что к внешнему сравнению сев.-кой. **t-* или **b-* в функции местоимения 2-го л. привлекать недопустимо.

43) «язык»: жцх *d^hàrì*, aye *tari*, кнг *nt^háɾí*, окн *tali*, эко *d^hàlì*.

Варианты в жцх и эко указывают на прасев.-кой. реконструкцию **d^hari*, однако следует учитывать и такие нерегулярности, как систематическое оглушение в aye, окн и т. д., а также неожиданную преназализацию в кнг (у Доука; впрочем, это может быть опять-таки архаичным следом посессивного префикса 1-го л., см. 'я'). Учитывая типологическую частотность «необычных» фонем или кластеров в слове 'язык', не исключена реконструкция **nt^hari*, с уникальным для сев.-кой. сочетанием *nt^{h-}*.

44) «зуб»: жцх *sàì*, ауе *саи ~ сои ~ сѳои*, кнг *sáú*; окн *саи*, эко *č̣àā*. || Прасев.-кой. **сai*.

Для жцх П. Дикенс приводит также синонимичный эквивалент $\tilde{[a]}^h$, но это слово не имеет никакой этимологии.

45) «дерево»: жцх *lā̀t̃*, ауе *lā̀t̃*, кнг *lā̀t̃*, окн *lā̀t̃* ~ *lā̀* ~ *gā̀t̃*, эко *lā̀t̃* ~ *lā̀t̃*.

Слово имеет очень редкую структуру, что мешает однозначно установить первичность переднего дифтонга в жцх или заднего в кнг и окн; необычны также варианты с озвончением кликса в окн (подтверждаемые в более современном источнике [Snyman 1980], где это слово записано как *lā̀t̃* ~ *lā̀t̃*). Условная реконструкция — **lā̀t̃*: в этой форме «зашифрованы» все основные особенности ее рефлексии в языках-потомках, но ручаться за ее надежность не приходится.

46) «два»: жцх *cā̀* ~ *cā̀*, ауе *са ~ čа*, кнг *sā̀*; окн *cá* ~ *са ~ čа*, эко *čā̀*. || Прасев.-кой. **са* ~ **cā̀*.

Колебания между носовой и «чистой» артикуляцией гласного нерегулярны и могут отражать какие-то старые морфологические варианты. Фарингализация в жцх не подтверждается ни в эко, ни в менее надежных источниках, за исключением отдельных вариантов в записях Л. Ллойд (где зафиксирован значительный разброс вариантов: *sā̀* ~ *sā̀* ~ *s:a* ~ *са* ~ *cā̀* ~ *cā̀* ~ *čā̀*); природа ее непонятна.

47) «вода»: жцх *l̃ú*, ауе *l̃u* ~ *l̃ú* ~ *l̃ú*, кнг *l̃ú*, окн *l̃ú* ~ *l̃ó*, эко *l̃ú*. || Прасев.-кой. **l̃u*. Назализация в кнг, очевидно, вторична.

48а) «мы» (экскл.): жцх *è*, ауе *e*, окн *e*, эко *è*.

48б) «мы» (инкл.): жцх *m̃*, эко *m̃-hm̃*.

Инклюзивная форма местоимения 1-го лица мн. ч. в сев.-кой. диалектах, по-видимому, используется редко и поэтому не отмечена в большинстве старых описаний (Д. Блик цитирует форму *hm̃* из записей Л. Ллойд по языку !кунг, отмечая ее редкость в текстах). Однако данные жцх и эко однозначно говорят в пользу реконструкции оппозиции **e* (экскл.) : **m̃* (инкл.).

49) «что»: (а) жцх *hà-čé*; (б) ауе *če=ba*; (в) окн *m-pai*, эко *m-čá* ~ *m-čē* ~ *m-čí*.

50) «кто»: (а) жцх *hà-žòè*; (б) ауе *žu=ba*; (в) окн *m-žu*, эко *m-žē* ~ *m-žòè*.

Оба вопросительных местоимения образуются во всех диалектах по общей схеме: префиксальная (жцх, окн, эко) или суффиксальная (ауе) вопросительная основа в сочетании с существитель-

ными **çi* 'вещь, предмет' ('что') и **ži* 'человек' ('кто'). В качестве факультативного компонента обнаруживается также указательная основа *-*e*, вызывающая «мутации» конечного гласного основы (жцх *hà-çé* ← **ha-çi-e*, *hà-žòè* ← **ha-ži-e* и т. п.).

Сами по себе вопросительные основы, однако, оказываются этимологически различными по диалектам (теоретически не исключено, что ауе = *ba* и окн, эко *m*- как-то связаны, но только через нерегулярное фонетическое развитие). В плане дистрибуции, как показывают сопоставительные данные Я. Снеймана [Snyman 1997: 53], шире всего распространена морфема *(*h*)*a*-, что не снимает вопрос о происхождении двух других.

Ауе = *ba* фонетически тождественно жцх вопросительной частице *bà:h*, ср.: *hàçe bā:h ke* 'что это?' в [Dickens 1994: 189]. Именно эта конструкция, скорее всего, подвергается стяжению в ауе: **ha-çe bā:h* → **hàçeba* → *çe-ba*.

Происхождение **m*- в северном пучке диалектов остается неизвестным; никаких аналогов для этого корня в жцх не обнаруживается, и в этой связи его можно привлекать к дальнейшему внешнему сравнению наряду с **ha*-, хотя и с несколько большей опаской (вопросительное *ha*- в северных диалектах, судя по Снейману, все же присутствует, в то время как следов *m*- за пределами этого пучка нет в принципе).

2.2. Язык восточный ꜥхоан.

2.2.1. *Общие сведения и источники.* На вымирающем языке восточный ꜥхоан (сокращенно — просто ꜥхоан; уточнение «восточный» необходимо для того, чтобы отличать его от «западного ꜥхоан», в генетическом плане являющегося одним из диалектов южнокойсанского языка !хонг, см. ниже) сегодня говорит не более 200 носителей в юго-восточных районах Ботсваны (пустыня Калахари). Из-за столь низкой численности носителей, «затерянных» на фоне гораздо более многочисленных племен бушменов !хонг, язык был открыт сравнительно недавно, и в общекойсанской классификации Гринберга остался неучтенным.

Хронологически первым (но крайне несовершенным с точки зрения качества записи) источником по восточному ꜥхоан является небольшой глоссарий в [Traill 1973]. С тех пор исследованием

отдельных аспектов ꜥхоан занимались Дж. Грубер [Gruber 1973, 1975] и К. Коллинс [Collins 1998, 2001a, 2001b, 2001c; Bell & Collins 2001]. Однако, хотя полевая работа по ꜥхоан продолжается до сих пор, основательного словаря по этому языку пока что так и не опубликовано, и большую часть лексических данных приходится извлекать либо из перечисленных выше статей, либо из (довольно скудного) иллюстративного материала к описанию фонетики ꜥхоан, который до недавнего времени был доступен на веб-сайте Корнеллского университета (<http://ling.cornell.edu/khoisan>).

В генетическом отношении язык восточный ꜥхоан некоторое время формально считался изолятом. Однако уже в работах «первооткрывателя» ꜥхоан, Э. Трэйлла, высказывалась гипотеза, что ꜥхоан — своеобразное «промежуточное звено» между северно- и южнокойсанскими языками ([Traill 1974]; критику этого мнения см. в [Westphal 1974]).

Один из немногих исследователей, систематически исследовавший проблему генетических взаимоотношений койсанских языков, Г. Хонкен, уже в первой своей статье по этой тематике [Honken 1977], поместил восточный ꜥхоан внутрь севернокойсанской группы в качестве наиболее рано отделившейся ветви. Эта гипотеза была подтверждена в работах Г. Старостина в ходе лексикостатистического [G. Starostin 2003] и сравнительно-фонетического [G. Starostin 2008a: 356-363] анализа доступного материала. Параллельно с нашими исследованиями детальным обоснованием «жу-ꜥхоан» родства продолжал заниматься Г. Хонкен, уже в сотрудничестве с Б. Хайне; результатом их совместной работы стала публикация [Heine & Honken 2010], в которой подробно разбираются регулярные фонетические соответствия, представлен вариант праязыковой реконструкции и даже предложено название для «новой» языковой семьи — *кха*, от слова, обозначающего 'землю' в обоих таксонах (ꜥхоан, жу|хоан *kà*). (Впрочем, мы согласны с Б. Сэндс относительно неудачности такого названия, т. к. это слово не является эксклюзивной «жу-ꜥхоан» изоглоссой: так, оно встречается в том же значении и в языке квади, который, безусловно, не может быть включен в эту же семью, см. раздел 2.5).

Несмотря на то, что генетическое родство между сев.-кой. языками и ꜥхоан на сегодняшний день можно считать надежно

установленным и серьезным возражений со стороны специалистов не вызывает, семья «жу-џхоан», или «кха», лишь с большой натяжкой может быть признана «таксоном 1-го уровня». Общий процент совпадений в базисной лексике между прасев.-кой. и џхоан языками колеблется в районе 40% [G. Starostin 2003: 123], т. е. минимально допустимого числа для объединения языков в «группу», а не «семью»; «интуитивно» это родство не является самоочевидным (в противном случае вряд ли, например, Э. Трэйлл мог бы колебаться между вариантами отнесения џхоан к северно- или южнокойсанским языкам); и, наконец, между сев.-кой. языками и џхоан имеются также очень существенные типологические различия — в частности, фонологическая система џхоан в целом ближе к южно-, чем к северно-койсанским системам (см. ниже, п. 2.2.2).

Исходя из этих соображений, на начальном этапе исследования мы все же предпочтем рассмотреть базисно-лексический материал џхоан отдельно от сев.-кой. лексики. В заключительном разделе первого тома будет показано, что даже предварительный (как чисто автоматический, так и «с ручной коррекцией») анализ сходств не оставляет сомнений в генетическом родстве языков «жу-џхоан» («кха»); но реконструкция общего 50-словного списка для этих таксонов сразу на начальном этапе сравнения — задача чересчур сложная.

2.2.2. *Историческая характеристика.* В плане фонетики и фонотактики восточный џхоан устроен примерно так же, как южнокойсанский язык !хонг (см. раздел 2.3.2). Его отличительные черты — наличие подсерии лабиальных кликов (θ , θ , $\tilde{\theta}$ и др.), а также чрезвычайно богатой системы исходов кликов, в том числе увулярных (подробное описание содержится в работе [Bell & Collins 2001]).

В типологическом плане особенно сближает џхоан с юж.-кой. языками использование, хотя и в сильно ограниченном объеме, лабиальных кликов. Однако обращает на себя внимание тот факт (легко заметный даже при сравнении соответствующих 50-словных списков), что слова с лабиальными кликами в џхоан и слова с лабиальными кликами, реконструируемые для пра-юж.-кой. состояния, не совпадают (этому явлению посвящена специальная работа [Старостин 2006]). Следовательно, речь идет об

«ареальной» черте — например, вторичном развитии лабиально-щелчковой артикуляции в ɬ хоан под влиянием юж.-кой. окружения (обратное развитие менее вероятно, исходя из общей дистрибуции языков).

Специфическим образом устроена подсистема аффрикат и фрикативных. Во-первых, именно в этой области проявляется характерное типологическое сходство ɬ хоан и сев.-кой. языков: как и для последних, для ɬ хоан смыслоразличительной является оппозиция между альвеолярными (свистящими) и альвео-палатальными (шипящими) — ср. *sā* 'слышать', но *šā* 'приходить'. (Как показано в [G. Starostin 2008a] и в [Heine & Honken 2010], внешнее сравнение с сев.-кой. данными показывает, что это, скорее всего, архаизм, а не инновация).

Во-вторых, ɬ хоан — едва ли не единственный из всех «койсанских» языков, в котором на синхронном уровне противопоставлены *три* ряда аффрикат: помимо свистящих и шипящих, в его инвентарь входят также альвео-палатальные аффрикаты ɕ и ɟ . Здесь, однако, уже очевиден инновативный характер этого феномена, т. к. палатальный ряд на самом деле занимает место дентального: ɕ и ɟ — результат относительно недавней палатализации обычных **t* и **d*, на что, помимо типологических, недвусмысленно указывают и диалектные данные (см. ниже), и данные внешнего сравнения с сев.-кой. материалами.

Записи Дж. Грубера [Gr.] и К. Коллинса [Col.] отличаются большей точностью и детализированностью по сравнению с ранними транскрипциями Э. Трэйлла [Traill 1973]; в частности, только Грубер систематически маркирует для ɬ хоанских лексем тональные характеристики. Тем не менее, некоторые элементы базисной лексики до сих пор остаются зафиксированными только в материалах Трэйлла, так что к анализу необходимо привлекать и их (сопровождаемые ниже пометой [Т.]). Все данные Трэйлла цитируются по списку [Traill 1973]; материалы Грубера — по статьям [Gruber 1973, 1975], а также по компаративным работам Г. Хонкена и Б. Хайне, использовавшим их в качестве основного источника [Honken 1977, 1988, 1998; Heine & Honken 2010]; материалы Коллинса — по статьям [Collins 1998, 2001a, 2001b, 2001c; Bell & Collins 2001], а также транскрипциям, сопровождаемым звуковыми файлами, взятыми с бывшего веб-сайта по койсанской

лингвистике Корнеллского университета (по состоянию на сегодняшний день сайт официально закрыт).

Известны как минимум два диалекта восточного ɬхоан — собственно ɬхоан и цхаси, причем цхаси, по-видимому, с точки зрения исторической фонетики несколько более архаичен: так, старые дентальные смычные, которые в стандартном ɬхоан реализуются как палатальные аффрикаты (напр., *ɕe:ma* 'собака', *ɕao* 'идти'), в цхаси остаются дентальными (*te:ma*, *tao*), см. [Traill 1980: 175]. К сожалению, по диалекту цхаси нам известно всего чуть менее десятка лексических единиц (в передаче Трэйлла), и лексико-статистические подсчеты по диалектам ɬхоан невозможны.

«Обще-ɬхоан»/«пра-ɬхоан» формы для элементов 50-словного списка по очевидным причинам не приводятся, но на первом месте там, где это возможно, располагаются формы в записи Дж. Грубера и К. Коллинса, как заведомо более точно передающие звуковой и фонологический облик слова.

50-словный список для языка восточный ɬхоан.

- 1) «пепел»: хоа *ʔʰoe* [Col.], *ʔoe* ~ *ʔue* ~ *ʔueʰ* [T.].
- 2) «птица»: хоа *ʔi-si*: [Gr.], *ʔi:se* ~ *ʔiʔisi* ~ *ʔi:si* [T.]. Морфема *-si* — продуктивный уменьшительный суффикс. Г. Хонкен [Honken 1988: 60] приводит также форму *cʰa:ma* 'bird', сопоставимую с соответствующим сев.-кой. корнем; однако во всех остальных работах в качестве нейтральной «птицы» упоминается исключительно *ʔi-si*:.
- 3) «черный»: хоа *ʔkxau* [Col.], *ʔxau* ~ *!kxiu* [T.].
- 4) «кровь»: хоа *ʔqi* [Col.], *ʔi:* ~ *ʔkɕei* ~ *ʔi:* [T.].
- 5) «кость»: хоа *ɕa:* [Col.], *ɕeɕ* ~ *ɕeɕa:* ~ *ɕiɕ* [T.] (+ цхаси *tɕʔa*).
- 6) «ноготь»: хоа *!ɕ* [Gr.], *!ɕʔo* [Col.].
- 7) «умирать»: хоа *ʂi* [Gr.], *ʂi* [Col.], *ʂe:* ~ *sẽ:* ~ *sĩ:* [T.].
- 8) «собака»: хоа *ɕeata* [Col.], *ɕe:ta* ~ *ɕi:ta* [T.] (+ цхаси *te:ma*).
- 9) «пить»: хоа *ɕu* [Col.], *cʰu:* ~ *cu:* [T.].
- 10) «сухой»: хоа *ʔqʰau* [Col.], *ʔao* ~ *ʔkɕu* ~ *ʔau* [T.].
- 11) «ухо»: хоа *ʔqʰõẽ* [Gr.], *ʔqʰoe* [Col.], *ʔõẽ* ~ *ʔʰõẽ* ~ *ʔʰũĩ* [T.].
- 12) «есть»: хоа *ʔám* [Gr.], *ʔam* [Col.], *ʔam* ~ *ʔa:m* [T.].
- 13) «яйцо»: хоа *kʰõʔẽ* ~ *ɕxui* ~ *ɕxui* [T.] (?). Здесь неясно, идет ли речь о двух разных корнях или каких-то диалектных вариантах, т. к. подобного рода «палатализация» для других основ неиз-

вестна. К сожалению, в более достоверных источниках слово пока не зафиксировано.

14) «глаз»: хоа *θōā* [Gr.], *θoa* [Col.], *θoa* ~ *θūī* ~ *θua* [T.]. Необычная форма *θūī* у Трэйлла может на самом деле быть формой мн. ч.: ср. у Грубера — *θó:a*, мн. ч. *θqǒē-qà* (с назализацией в корне и неожиданным чередованием исходов кликсов).

15) «огонь»: хоа *θoa* [Col., T.].

16) «нога»: хоа *!aʔu* [Col.], *!aʔui* [T.]. Ср. *!a*: ~ *!ao* 'нога (выше стопы, т. е. leg)' в записях Грубера (неясно, идет ли речь о той же самой основе).

17) «волосы»: хоа *ʃu* [Col.], *ʃu*: ~ *ʃu*: [T.].

18) «рука»: хоа *ʃiu* [Gr.], *siu*: [T.].

19) «голова»: хоа *ʔθnū* [Col.], *ʔθnū* ~ *ʔθnu*: [T.].

20) «слышать»: хоа *cā* [Gr.], *ca*: [T.].

21) «сердце»: хоа *!qǒ* [Col.], *!ǒ*: ~ *!ū*: ~ *!ū*: [T.].

22) «рог»: хоа *!ʰǒ* [Gr.], *!ʰo-!ʰo* [Col.].

23) «я»: хоа *ta* [Col.].

24) «убивать»: хоа *!ǒʰ* [Gr.], *!ǒ:-ta* ~ *!ʰǒ:-wa*. [T.]. Корень, по видимому, один и тот же, хотя странный «глагольный суффикс» -*ta* ~ *-wa* у Трэйлла более нигде не встречается. Ср. также супплетивную основу множественного способа действия: *θó:a* [Gr.], *θoa* [Col.] 'убивать (многих)', 'вырезать'.

25) «лист»: хоа *kí=ʒǒba* [Gr.], *ʒǒba* [Col.].

26) «вошь»: хоа *cí* [Gr.], *cí*: [T.].

27) «мясо»: хоа *||ǵe* [Col.], *||ǵae* 'мясо', *||ǵa*: 'животное' [T.]. (последние две формы в морфологическом плане могут представлять различные лексемы, но корень, несомненно, один и тот же, учитывая естественность полисемии 'мясо': 'животное' для данного ареала).

28) «луна»: хоа *ʃibi* [Col.].

29) «рот»: хоа *šī* [Col.], *še*: ~ *sē*: ~ *sī*: [T.].

30) «имя»: хоа *!ó*: [Gr.].

31) «новый»: хоа *zà*: [Gr.] (цит. по [Honken 1988: 58]).

32) «ночь»: хоа *cʰao* [T.]. Фонологически, возможно, *cʰao*, т. к. транскрипция Трэйлла систематически не различает шипящие и свистящие аффрикаты; в более надежных источниках слово не зафиксировано.

33) «нос»: хоа *!qǒ* [Col.], *!ǒʔǒ* ~ *!ū*: [T.].

- 34) «не»: хоа /^hθ̄^hθ̄ [Col.] (частица, стоит в препозиции к глаголу).
- 35) «один»: хоа θ̄ĩ [Col.], θ̄ō:θ̄ō: ~ θ̄ĩ: ~ θ̄ikiθ̄i [T.] (ср. аналогичную редупликацию и даже «ретрипликацию» ниже для числительного '2').
- 36) «дождь»: хоа č̄ō^hā: ~ cō^hā: [T.]. Исходная артикуляция аффрикаты остается неясной из-за отсутствия более надежных данных.
- 37) «дым»: хоа z̄ōe [Gr.], z̄u(:)e [T.].
- 38) «звезда»: хоа ʃ̄ō [Gr.], ʃ̄ō [Col.], ʃ̄ō: ~ ʃ̄o: ~ ʃ̄u [T.].
- 39) «камень»: /^hθ̄:a [Gr.]. У Коллинса ср. также /^hqa 'rock'.
- 40) «солнце»: хоа č̄^ha: ~ z̄a ~ z̄a: [T.]. Исходная артикуляция аффрикаты остается неясной, однако слово как таковое, по-видимому, заимствовано либо из центр.-кой. источника (ср. нама *se:* 'день'), либо непосредственно из банту (ср. ПБ **ci* 'день, дневное время'). Скорее вероятен первый вариант, т. к. переход **e* > *a* в ʃ̄хоан наблюдается и в ряде других слов, имеющих внешние параллели, переход же **i* > *a* не отмечен совсем.
- 41) «хвост»: хоа θ̄xui [Col., T.].
- 42) «ты»: хоа ú: [Gr.].
- 43) «язык»: хоа cela [Col.], cela ~ cala [T.].
- 44) «зуб»: хоа c̄iú [Gr.], ciu [Col.]. В [Honken 1988: 56] парадигма (по данным Дж. Грубера) дается в виде *c̄iui*, мн. ч. č̄eo-qa (с неожиданным чередованием аффрикат).
- 45) «дерево»: хоа /^hō [Gr.], /^hō [Col.], /^hū: [T.].
- 46) «два»: хоа θ̄ó:a [Gr.], θ̄oa [Col.], θ̄oa ~ θ̄oθ̄oθ̄oa [T.].
- 47) «вода»: хоа ʒo [Col.], ʒo: ~ zo: [T.].
- 48) «мы»: хоа n-!a [Gr.]. Основная корневая морфема — слоговой *n*; -!a, по-видимому — показатель множественности.
- 49) «что»: хоа c̄ini [T.].
- 50) «кто»: в доступных источниках не зафиксировано.

2.3. Южнокойсанская группа.

2.3.1. *Общие сведения и источники.* В отличие от северной койсанской общности, которое с некоторой натяжкой можно охарактеризовать как диалектный континуум, термин «южнокойсанские» (South Khoisan; не следует путать с условным термином Дж. Гринберга «South African Khoisan», объединяющим все «койсанские»

языки, кроме хадза и сандаве) применяется по отношению к группе языков, значительно отличающихся друг от друга по грамматике, лексике и (насколько можно судить по обычно сомнительному качеству транскрипции) фонетике. К сожалению, подавляющее большинство этих языков, на которых, до прихода сначала племен банту, а затем — европейских колонизаторов, говорила значительная часть коренного населения современной ЮАР, на настоящий момент являются вымершими и известны лишь в записях исследователей конца XIX-го — середины XX-го столетия (большая часть этих записей собрана и опубликована в [Bleek 1956]). Это существенно затрудняет реконструкцию юж.-кой. состояния, которой должна предшествовать своеобразная «реконструкция» истинного фонемного инвентаря описанных мертвых языков, в значительной степени искаженного при передаче в «сырой» транскрипции.

Определенные сомнения в реальности юж.-кой. единства в свое время высказывались Э. Вестфалем в рамках «гиперкритического» подхода, которого последний придерживался в вопросе о генетических связях между «койсанскими» языками (см. особенно [Westphal 1965], где известные автору юж.-кой. языки разнесены по трем группам, возможность родства между которыми предполагается, но не считается достаточно обоснованной); не решаются однозначно положительно ответить на этот вопрос и такие современные исследователи, как Т. Гюльдеманн и Г. Хонкен.

Тем не менее, внимательное изучение материала показывает: количество системных лексических сходжений между этими языками столь велико (ниже будет показано, что примерно 70% вхождений в 50-словном списке праюж.-кой. языка имеют наглядные рефлексy во всех ветвях этой группы), что практически полностью исключает как возможность их случайного совпадения, так и «контактного» объяснения. Сомнения в юж.-кой. единстве скорее вызваны «техническими» факторами — например, трудностями в установлении корреляций между грамматическими системами таких языков, как !хонг и |хам, хотя на самом деле основным препятствием здесь является в первую очередь отсутствие адекватных грамматических описаний для вымерших языков.

Далеко не все юж.-кой. языки (диалекты), упоминаемые в доступной лингвистической литературе, оказываются пригод-

ными для лексикостатистического исследования; многие из них известны лишь по кратким описаниям, не содержащим статистически значимого числа элементов даже из 50-словного списка, не говоря уже о 100-словном. Ниже мы перечислим основные языки, лексикографические описания которых достаточны для включения их в исследование.

1) /Хам [хам] — бушменский язык, некогда самый многочисленный по числу носителей на территории совр. ЮАР, но находившийся на грани вымирания уже к 1910 г. (согласно информации Н. Кроухолла [Crawhall 2004: 172], последний носитель умер в 1990-е гг.). Хам известен почти исключительно по богатому архиву текстов, собранных В. Бликком и Л. Ллойд [Bleek & Lloyd 1911], а также по кратким грамматическим описаниям [Meriggi 1928/9] и [Bleek 1928-30], составленным по результатам анализа этих текстов. Лексические формы, извлеченные из перечисленных источников, цитируются по сравнительным словарям Д. Блик [Bleek 1929, 1956] («индекс Блик» SI); в качестве основного варианта по возможности цитируются формы в записи Л. Ллойд (как правило, одно и то же слово приводится в ее материалах в нескольких произносительных вариантах, что может отражать как ошибку восприятия, так и реальное диалектное варьирование).

2) *N/y* (*н/уки*, *ѓхомани*) — пучок близкородственных диалектов, в географическом плане расположенных к северу от ареала |хам (на границе современных ЮАР и Ботсваны). В первой половине XX в. активных носителей еще было достаточно много, так что язык удалось записать целому ряду лингвистов. С точки зрения качества транскрипционной записи более надежны данные краткого грамматического очерка Л. Майнгарда [Maingard 1937] и фонетического описания К. Доука [Doke 1936]; в количественном отношении, однако, их во много раз превосходят записи Д. Блик, частично опубликованные в недавно вышедшем из архивов очерке [Bleek 2000], частично — в ее сравнительных словарях [Bleek 1929, 1956] («индекс Блик» SII; данным Л. Майнгарда и К. Доука по т. н. «ѓхомани» присвоен индекс SIIIa, чтобы показать их особую близость к диалекту, записанному Д. Блик). Для лексикостатистических подсчетов мы используем в первую очередь данными Блик, но также приводим там, где это необходимо для прояснения отдельных историко-фонетических аспектов, более

точные в плане транскрипции варианты, записанные Л. Майнгардом и К. Доуком.

До относительно недавнего времени язык н|у считался полностью вымершим. Однако в середине 1990-х годов было обнаружено несколько пожилых носителей, уже давно перешедших на африкаанс, но не утративших при этом способность изъясняться на родном языке; в результате оказалось возможным получить некоторое количество языковых материалов, записанных с помощью новейших технических средств. К сожалению, большая часть этих материалов остается пока неопубликованной; для лексикостатистики нами были использованы данные из социолингвистической диссертации Н. Кроухолла [Crawhall 2004]; полевые записи Б. Сэндс и других исследователей, до недавнего времени доступные на сайте Корнеллского университета, а также частично любезно предоставленные в наше распоряжение самой Б. Сэндс [Sands et al. 2006]; и ряд недавних публикаций, посвященных отдельным аспектам фонетики и лексикологии н|у [Sands et al. 2007; Miller et al. 2009]¹. Вопрос о степени сохранности исконно юж.-кой. лексики в н|у специально обсуждается в первой из упомянутых статей, где показано, что, хотя н|у и находится в состоянии «быстрого вымирания» («radical death») в связи с поголовным переходом носителей на африкаанс, основной лексический инвентарь в ситуациях, когда носители все же переходят на родную речь, не вытесняется лексикой африкаанса, за исключением отдельных, обычно предсказуемых заимствований.

Четкую диалектную дифференциацию имеющихся источников провести оказывается довольно сложно, в первую очередь из-за неточностей в имеющихся описаниях. Лексикостатистическое сравнение данных Д. Блик и Б. Сэндс дает не менее 9-10 лексических отличий в пределах 100-словного списка, далеко не все из которых легко списать на неточности, допущенные при переводе н|у лексики на английский; и это при том, что не все

¹ Относительно недавно вышел в свет первый детальный грамматический очерк языка н|у [Collins & Namaseb 2011]; полностью включить опубликованные в нем данные в наше исследование мы уже не успели, однако в плане новой лексики очерк почти ничего не добавляет к вышеперечисленным публикациям.

позиции в списке удается заполнить. По этой причине, а также в целях повышения репрезентативности, в приводимых ниже списках отдельно будут цитироваться «старые» данные Д. Блик (в ее записях соответствующий диалект называется ||ng!ke, сокр. нгк) и «новые» данные Н. Кроухолла, Б. Сэндс и др. (сокр. нху; небольшие диалектные различия отмечены даже между отдельными информантами Н. Кроухолла, однако они не столь значительны, как отличия между его данными и записями Д. Блик, Л. Майнгарда и др.; как правило, речь идет о том, что некоторые из носителей нху подставляют на место «забытых» исконных слов эквиваленты из языков нама или африкаанс).

3) ||Хегви [хег]. Был локализован в районе озера Крисси (провинция Мпумаланга, ЮАР). Последний носитель, согласно данным Ethnologue, скончался в 1988 г.; язык остался частично зафиксированным в сравнительных словарях Д. Блик (где он фигурирует под местным бантусским названием *батва*; «индекс Блик» SIII); в двух кратких описательных статьях [Lanham & Hallowes 1956a, 1956b]; и в небольшом очерке [Ziervogel 1955], включающем несколько образцов текстов. Существенных диалектных различий между данными Блик [В.], Лэнхама / Хэллоуза [LH] и Цирфогеля [Zr.] не наблюдается.

4) |Ауни [аун]. Если район обитания носителей ||хегви отмечает восточную границу распространения юж.-кой. языков, то |ауни, наоборот, представлял собой западную границу; носители этого языка проживали примерно в районе слияния рек †Носсоб и Ауоб (граница ЮАР и Ботсваны). Язык известен исключительно по записям Д. Блик, опубликовавшей краткое грамматическое описание [Bleek 1937]; лексические материалы были включены в сравнительные словари [Bleek 1929, 1956] («индекс Блик» SIV).

5) |Хаси, или ку|хаси [хас]. Про этот язык, судя по разрозненным сведениям этнографического характера о его носителях, высказывалось предположение, что он еще в начале XX в. из всех юж.-кой. наречий мог быть вторым по числу и географическому разбросу носителей после |хам [Crawhall 2004: 175]. Тем не менее, единственный источник конкретных сведений о |хаси, пригодный для сравнительно-исторических целей — краткое грамматическое описание и словарь, составленные в 1937 г. Робертом Стори по итогам работы с одним-единственным информантом. Лексические

данные Стори были включены Д. Блик в ее словарь [Bleek 1956], но полное описание удалось опубликовать лишь благодаря усилиям Э. Трэйлла, разыскавшего оригинальную рукопись, в 1999 г. [Story 1999].

Д. Блик присвоила |хаси индекс SIVb, сочтя его, таким образом, одним из диалектов |ауни, наряду с диалектом SIVa, названным ей *кхатиа* или *хатиа* (в записных книжках Блик фигурирует также название *ǃэй-куси*), очень скудные данные по которому были записаны ей самой. Дать точную классификационную оценку кхатиа почти невозможно из-за нехватки данных; что касается отношений между |ауни и |хаси, то это, безусловно, разные языки, хотя, согласно мнению Т. Гюльдеманна, и близко родственные [Güldemann 2011].

б) !Хонг. Это — единственный из юж.-кой. языков, до сих пор не вымерший и не находящийся под непосредственной угрозой вымирания (по данным Ethnologue, на 2002 г. насчитывалось более 4,000 носителей). По разным оценкам, существует от двух до пяти-шести разных диалектов !хонг, из которых для нас наибольшее значение имеет диалект, представленный подробным словарем Э. Трэйлла [Traill 1994a], в котором язык называется просто !хонг (*!xóǃ*, [хнґ]; в литературе иногда встречается более специализированный термин *Lone Tree !Xoon*, указывающий на локализацию этого диалекта в районе населенного пункта Кацґгаэ в Ботсване).

Помимо этого, достаточные данные для лексикостатистического обследования имеются для вымершего диалекта *масарва*, или *какиа* (мас: собраны и опубликованы Д. Блик в [Bleek 1929, 1956]) и, по-видимому, еще живого диалекта н|у|ен (нүе: данные по тем же источникам). Известна также небольшая работа Л. Майнгарда [Maingard 1958], где он приводит сопоставительные данные по т. н. «южному» и «северному» диалектам !хонг, но для сколь-либо серьезных подсчетов их категорически недостаточно. Э. Вестфаль и Э. Трэйлл отмечают в своих работах существование и других диалектов (западный ǃхоан; н|амани; цхаси и др., полный список см. в [Treis 1998: 488-489]), но языковые данные по ним либо отсутствуют полностью, либо ограничены крошечными «ознакомительными» списками (например, в [Westphal 1965]).

Судя по лексике (заслуживающие доверия грамматические описания ограничены лишь краткими заметками Э. Трэйлла по собственно !хонг), все вышеперечисленные диалекты достаточно близки друг к другу, но при этом, например, между !хонг Трэйлла и масарва Д. Блик насчитывается порядка 15 расхождений в 100-словном списке; даже если некоторая часть из них фиктивна (т. е. обусловлена ошибочными семантическими эквивалентами в записях Блик), можно все равно подозревать, что «!хонг» на самом деле является небольшой языковой семьей или, по крайней мере, «макроязыком».

Из прочих юж.-кой. языков краткими словарными списками представлены также четыре отдельных наречия, которые, по мнению Д. Блик, тесно связаны с н|у; //ку//е («индекс Блик» SIIc; данные собраны самой Д. Блик и впервые опубликованы в [Bleek 1956]); *seroa* (индекс SIIId; [Wuras 1920a]); *Ɂlan!e* (индекс SIIe; [Anders 1934]); //кхау, или *Ɂунг-кве* (индекс SIIb; [Meinhof 1929]); все данные воспроизведены в [Bleek 1956]. Судя даже по этим скудным материалам, ни один из перечисленных языков нельзя отнести непосредственно к диалектам н|у, хотя сомневаться в принадлежности их к юж.-кой. группе также не приходится. Составить 100-словные списки по ним невозможно, но сам материал не следует игнорировать, т. к. он иногда позволяет подтвердить архаичность тех или иных реконструкций, а также внести существенные коррективы в наши представления об исторической типологии звуковых переходов в юж.-кой. языках.

До недавнего времени большинство специалистов по «койсанским» языкам поддерживали концепцию бинарного деления юж.-кой. семьи на две группы — *таа* (от !хонг *tâ* 'человек') и *!кви* (от |хам *!ui* 'человек'). При этом считалось, что в первую входит !хонг со всеми его диалектами, во вторую — все остальные известные языки (так, например, в [Güldemann & Vossen 2000]; в работе [Hastings 2001] эта классификация подтверждается на основании анализа грамматических и лексических данных по !хонг, |хам, н|у и |ауни). Существенные лексические и грамматические расхождения между *таа* и *!кви* даже подталкивали отдельных исследователей к скепсису в отношении их генетического родства; в частности, «юж.-кой. родство» никогда не принималось «гиперскептически» настроенным Э. Вестфалем (см. [Westphal 1962] и др.). В

современной койсанистике этот скепсис преодолен, в основном благодаря появлению новых расширенных данных (в первую очередь подробного словаря !хонг, составленного Э. Трэйллом), но вопрос о внутренней классификации семьи остается открытым и, опять-таки, в связи с расширением корпуса за счет как новооткрытых полевых, так и свежее опубликованных архивных данных, более актуальным, чем когда-либо.

Так, к числу новейших гипотез относится работа [Güldemann 2011], в которой автор предлагает пересмотреть бинарное членение юж.-кой. группы в пользу *тернарного*, отделяя от подгруппы !кви языки |ауни и |хаси и помещая их в особую подгруппу, называемую автором «нижненоссобской» (Lower Nossob, от названия реки, вдоль которой было зафиксировано присутствие носителей этих языков), причем ближайшими родственниками «нижненоссобских» наречий в классификации Гюльдемманна оказываются языки таа, а не !кви. В поддержку своей гипотезы Гюльдемманн приводит набор морфосинтаксических и лексических изоглосс, которые, действительно, объединяют «нижненоссобские» языки скорее с !хонг, чем с |хам и н|у. При этом, однако, автор не ставит целью показать *инновативный* характер этих изоглосс, что во многом обесценивает его выводы: для того, чтобы уверенно согласиться с исторической реальностью генетической общности «носсоб-таа», необходимо, чтобы значительный процент этих эксклюзивных изоглосс наглядно представлял собой общие инновации, а не архаизмы.

Ниже, в разделе 2.3.3, мы еще вернемся к этому вопросу по ходу презентации результатов юж.-кой. лексикостатистики; пока отметим лишь, что сама по себе спорность вопроса о внутреннем членении юж.-кой. группы в обязательном порядке требует анализа ее лексических данных в общей совокупности, а не отдельно по !кви и по таа. Впрочем, учитывая, что даже между наиболее удаленными друг от друга юж.-кой. языками процент базисно-лексических совпадений никогда не опускается ниже сорока пяти, такое решение вполне согласуется с нашей методикой и без оглядки на «общественное мнение» в отношении характера связей между таа, !кви и «нижненоссобским».

2.3.2. *Историческая характеристика.* Имеющиеся в нашем распоряжении на сегодняшний день полноценные фонетико-фоно-

логические описания !хонг [Traill 1985, 1994a] и н|у [Miller et al. 2007, 2009] позволяют утверждать, что юж.-кой. языки — абсолютные рекордсмены (как среди «койсанских», так и на общемировом уровне) по сложности устройства своих фонемных инвентарей. Помимо того, что во всех этих языках к «стандартным» четырем типам кликсов добавлен пятый (лабиальный), как !хонг, так и н|у отличаются очень большим числом исходов кликсов (для !хонг Трэйлл отмечает до 17 смыслоразличительных оппозиций; для н|у, правда, в работах А. Миллер и др. отмечено «всего» 10, т. е. на две оппозиции меньше, чем в жу|хоан; однако, во-первых, эти данные могут оказаться неокончательными, во-вторых, следует учитывать, что они записывались от носителей, которые уже много лет не использовали н|у в качестве активного языка общения и могли в связи с этим нейтрализовать отдельные «сверхредкие» оппозиции).

Как в !хонг, так и в н|у, сверх того на это накладываются: в подсистеме «обычных» согласных — оппозиции по придыхательности и абруптивности, а в подсистеме вокализма — дополнительные признаки назализованной и фарингализованной артикуляции (в !хонг также, по-видимому, еще и «шепотной»), не говоря уже о стандартных для всех койсанских языков тональных противопоставлениях (впрочем, тональная система н|у пока что остается неопищенной).

Все это превращает юж.-кой. языки в важнейшее поле исследования для компаративиста, т. к. даже в том случае, если фонетическая сложность языков типа !хонг окажется инновативной, сама по себе реконструкция ее исторических корней могло бы пролить свет на самые разные проблемы предистории южноафриканского языкового пространства. К сожалению, работа в этой области чрезвычайно затруднена из-за того, что большинство сведений о прочих языках этой семьи дошло до нас в весьма несовершенной транскрипционной записи. Тем самым проблема реконструкции фонологического состава праюж.-кой. языка осложняется дополнительной проблемой: необходимостью «реконструкции» по существующим записям конкретного фонологического состава исторически зафиксированных юж.-кой. языков.

Наглядным примером может служить подсистема увулярных согласных и исходов кликсов, уверенно зафиксированная для

языка !хонг Э. Трэйллоу и для языка н|у — А. Миллер, Б. Сэндс и др. исследователями, причем во многих случаях эти подсистемы согласуются друг с другом (ср., например, !хонг *ʔq^hʔie* 'ветер' = н|у *ʔq^hʔe* id. и т. п.), т. е., по всей видимости, отражают особенность праюж.-кой. фонетики. Тем не менее, из всего массива данных, накопленных за более чем сто лет изучения юж.-кой. языков, наличие увулярных согласных в других языках этой группы эксплицитно признается *только* для ||хегви, и то лишь в статьях Лэнхама и Хэллоуза (см. выше). В объемном словаре [Bleek 1956], где собраны материалы самых разных исследователей, нет даже намек на то, что эти фонемы существуют: в транскрипционных записях они записываются как обычные велярные согласные, что в каких-то случаях может отражать реальное историческое развитие, но в большинстве, судя по всему, должно трактоваться как дефект транскрипции.

Целый ряд типовых проблем, связанных с передачей юж.-кой. фонетики в старых источниках, обсуждается в [Starostin 2008a]; вопросу о принципах транскрибирования текстов на |хам в записях В. Блика и Л. Ллойд даже посвящено специальное мини-исследование [Traill 1995]. Облегчить ситуацию удастся лишь благодаря наличию «диагностических» языков — !хонг и н|у. С другой стороны, упрощение фонетики исследуемых языков, сколь бы грубым оно ни являлось, при уделении должного внимания материалу вряд ли может значительно исказить результаты лексикостатистических подсчетов, т. к. потенциальные когнаты между плохо и хорошо записанными языками, как правило, все же опознаются без особого труда. Удастся, хотя и не всегда надежно, даже установить некоторые нетривиальные фонетические (точнее, «графо-фонетические») соответствия между сравниваемыми языками; о некоторых из них см. ниже, при разборе конкретных этимологий.

Задача систематической реконструкции пра-юж.-кой. фонологической системы пока что остается нерешенной. В основе этой реконструкции, по-видимому, должно лежать бинарное сопоставление между !хонг и н|у; данные остальных языков пока что не дают оснований надежно реконструировать какие-либо дополнительные оппозиции, не прослеживаемые на «пересечении» соответствий между этими двумя языками, и играют скорее вспомо-

гательную роль. Однако сами эти соответствия далеко не столь «тривиальны», как, например, соответствия между сев.-кой. диалектами. В работе [Starostin 2008a: 365-373] перечислен целый ряд существенных расхождений между фонолого-фонетической репрезентацией родственных форм в юж.-кой. языках, включая даже отдельные случаи систематического несовпадения основ кликсов, не всегда объясняемые погрешностями транскрипции.

Поскольку наибольшее количество фонологических оппозиций из всех юж.-кой. языков на сегодняшний день зафиксировано в !хонг, и ни данные внутреннего анализа, ни данные внешнего сравнения пока что не позволяют убедительно показать их вторичное происхождение, предварительные реконструкции, представленные ниже в разделе, посвященном этимологическому анализу 50-словных списков, будут опираться в первую очередь на фонетико-фонологический облик слов в !хонг (разумеется, при наличии в !хонг соответствующего эквивалента). В связи с этим имеет смысл воспроизвести здесь общую систему фонологии !хонг в том виде, в котором ее постулирует Э. Трэйлл [Traill 1994: 11-12] (с небольшими изменениями транслитерационного характера):

Исход кликса	Основа кликса				
	Лаб.	Дт.	Пл.	Аль.	Лат.
Нулевой	∅		‡	!	
Озвончение	∅̥		‡	!̥	
Назализация	∅̃	̃	‡̃	!̃	̃
«Глухая» назализация	∅̥̃	̃	‡̃	!̃	̃
Гортанный взрыв	∅ʔ	ʔ	‡ʔ	!ʔ	ʔ
Аспирация	∅ ^h	^h	‡ ^h	! ^h	^h
Преглоттализация+назализация	ʔ∅n	ʔ n	‡n	!n	n
Фрикативный велярный	∅x	x	‡x	!x	x
Фрикат. велярный+озвончение	∅̥x	x	‡x	!̥x	x
Велярная аффриката	∅kx	kx	‡kx	!kx	kx
Велярная аффриката+озвончение	∅̥kx	kx	‡kx	!̥kx	kx
Увулярный смычный	∅q	q	‡q	!q	q
Увулярный смычный+озвончение	∅̥q	q	‡q	!̥q	q
Увулярный смычный+аспирация	∅q ^h	q ^h	‡q ^h	!q ^h	q ^h
Увулярный смычный+озв.+асп.	∅̥q ^h	q ^h	‡q ^h	!̥q ^h	q ^h

глухости/звонкости (аналогичные оглушения в словах со «сложными» исходами кликсов типологически свойственны, например, северному пучку сев.-кой. диалектов). Гораздо труднее понять, что в н|у произошло с исходом на глоттализированный увулярный, а также со «сложными» носовыми исходами: материала здесь очень мало, соответствия между !хонг и н|у неоднозначны, и к тому же (согласно личному сообщению Б. Сэндс) существуют определенные подозрения, что часть данных, несмотря на столь тщательную детализацию транскрипции, все же могла быть записана Трэйллом с ошибками.

Подсистема «обычных» согласных в !хонг имеет следующий вид (в скобках отмечены согласные, вряд ли выводимые на уровень, т. к. они встречаются в основном в заимствованиях):

(p)	t	c	k	kx	q	ʔ
(p ^h)	t ^h	c ^h	k ^h		q ^h	
(b)	d	ʒ	g	gx	ɠ	
	d ^h	ʒ ^h	g ^h		ɠ ^h	
	tʔ	cʔ	kʔ		qʔ	
		s	x			h
m	n, (l)					
ʔm	ʔn					

Основные отличия н|у от !хонг сводятся к следующим: (1) для н|у, как и для ряда других !кви-диалектов (а также, в типологическом плане, для западного !хоан) характерна палатализация дентальных согласных, т. е. в м. серии *t*, *t^h*, *d* в этом языке представлены *ɕ*, *ɕ^h*, *ʒ*; (2) звонкие придыхательные отсутствуют; (3) из всех альвеолярных аффрикат представлена только непридыхательная глухая *c* и глоттализированная *cʔ*; (4) увулярный ряд представлен только глухим смычным *q* и глоттализированной аффрикатой *qʔ* (= !хонг *qʔ*); (5) преглоттализированные сонанты *ʔm*, *ʔn* отсутствуют. Таким образом, и здесь мы наблюдаем существенное упрощение инвентаря по сравнению с !хонг.

В системе вокализма юж.-кой. языков наблюдается примерно такое же распределение, как и в сев.-кой. языках: высокая частотность средне- и заднерядных гласных *a*, *o*, *u* при относительной редкости (по крайней мере, в «корневой» море) передних *e*, *i*.

Назализованная и фарингализованная артикуляция регулярно отмечается как в !хонг, так и в н|у, и даже в «старых» записях по вымершим юж.-кой. языкам, но соответствия далеко не всегда бывают регулярными. Что касается шепотной артикуляции гласных в !хонг, последовательно отмечаемой Трэйллом, то ей пока что не найдено никаких эквивалентов ни в н|у, ни где-либо еще, и фонологичность этого признака еще предстоит верифицировать.

Структура слова в юж.-кой. языках в целом следует тем же правилам, которые в разделе 2.2.2 были описаны для сев.-кой. языков: в частности, !хонг практически идеально укладывается в «двухморную» схему, причем для этого языка она является еще более наглядной, чем, например, для жу|хоан, из-за того, что в !хонг сильнее развита парадигматическая морфология, и суффиксальный статус второй моры во многих случаях очевиден: ср., например, ^hā-be 'лук' (ед. ч.) : ^hā-n 'луки' (мн. ч.), ^lǃ-i 'самка' (ед. ч.) : ^lǃ-ba-tê 'самки' (мн. ч.). Вообще, !хонг отличает достаточно сложно устроенная система именных классов, маркируемая как в глаголе, так и с помощью суффиксальных показателей внутри самих имен, и связанное с этим богатое разнообразие моделей образования множественного числа существительных (подробное исчисление этих моделей с попыткой внутренней реконструкции исходной системы см. в [Старостин 2008б]; семантика именных классов в западном и восточном диалектах !хонг также обсуждается в работе [Kießling 2008]).

Наличие систем именных классов в прошлом можно предполагать и для языков !кви, хотя ситуация здесь очень сложная. Немногочисленные грамматические очерки по этим языкам, как правило, вообще не упоминают о соответствующей типологической характеристике, что объясняется двумя причинами. Во-первых, в таких языках, как |хам, н|у и др., именное словоизменение в целом характеризуется более простыми и продуктивными моделями, чем в !хонг, с гораздо меньшей степенью «подвижности» «суффиксальной моры» (так, в |хам формы мн. ч. чаще всего образуются либо просто с помощью суффиксов -kən, -tən и др., либо через механизм редупликации «корневой моры»). Во-вторых, представление об именных классах в языках Южной Африки на момент первых исследований по юж.-кой. языкам было настолько тесно связано с идеей классных префиксов (по

«модели банту»), что именные классы в языках !кви и таа, морфологическое выражение которых носит исключительно суффиксальный характер и к тому же проявляется в первую очередь в глагольных формах, согласующихся с именем по его классу, просто остались незамеченными. Так, Л. Майнгард в своем кратком описании !хонг [Maingard 1958] упоминает о загадочных «колебаниях» гласных, в том числе в глагольных основах; сопоставление таких его примеров, как $\tilde{e} \sim \tilde{ei} \sim \tilde{i}$ 'видеть' и др. с более современными данными Э. Трэйлла, показывают, что большинство таких «колебаний» — это, скорее всего, различные морфологические варианты одной основы, образовавшиеся из стяжения корня $*\tilde{V}$ - с теми или иными классными показателями. Не исключено, что тщательный анализ записей текстов из архива Блика и Ллойд по |хам, даже несмотря на фонетическую неточность транскрипции, сможет эксплицитно выявить для этого языка именные классы; однако это тема для отдельного и очень сложного исследования.

В любом случае, «суффиксальная мора» в юж.-кой. языках оказывается несколько более подвижной, чем в сев.-кой. пучке, что станет очевидным ниже, на примере разбора конкретных этимологий; поэтому в любом лексическом сопоставлении, охватывающем материал !кви и таа, в первую очередь требуется удостовериться в фонетической совместимости первой (корневой) моры, допуская возможность несовпадения второй (как следствие перехода слова из одного именного класса в другой, старого аналогического выравнивания парадигмы по ед. или мн. ч. и т. п.). Пренебрежимым оказывается также несоответствие вокализма в глагольных основах — оно легко может отражать праязыковую или синхронную алломорфию. В именных основах вокализм обычно имеет большую значимость, и здесь отдельные расхождения между языками таа и !кви будут обсуждаться подробно (в рамках конкретных этимологий).

2.3.3. *Лексикостатистика*. Лексикостатистическая обработка 100-словных списков по шести перечисленным выше языкам, проведенная нами, также свидетельствует в пользу бинарной схемы классификации: средний процент лексических совпадений между !хонг и другими юж.-кой. языками — около 45-50%, в то время как |хам и н|у обнаруживают не менее 70% совпадений, |хам

и ||хегви — около 60% и т. д. (приводимые цифры слегка занижены по сравнению с показателями в табл. 2 за счет возможных ошибок в этимологизации сравниваемого материала). Таким образом, гипотеза Т. Гюльдемманна относительно особого статуса т. н. «нижненоссобских» языков не подтверждается. Однако внутри группы !кви |ауни и |хаси, несомненно, образуют отдельную ветвь, о чем недвусмысленно свидетельствуют их многочисленные эксклюзивные изоглоссы, которая при этом обособилась от «обще-!кви» ствола на достаточно раннем этапе развития. По всей видимости, по крайней мере некоторые из грамматических и лексических изоглосс между таа и «нижненоссобскими» языками, которые приводит Т. Гюльдемманн, можно интерпретировать как общие архаизмы, утраченные в «узко-!кви» подгруппе (состоящей из ||хегви, н|у и |хам), но не требующие интерпретации в терминах особой генетической близости между подгруппами таа и «носсоб»¹.

При этом как минимум одно из заключений, сделанных Т. Гюльдемманном, все же находит подтверждение в лексикостатистике: обращает на себя внимание существенный аномальный «подскок» процентных совпадений между н|у и |ауни, что также согласуется с наблюдениями Д. Блик, которая еще в публикациях 1930-х гг. обращала внимание на тесные контакты между носителями этих языков и сама выявила несколько явных заимствований из н|у в |ауни; отдельные случаи будут обсуждены ниже (см. особенно 'голова'). Между |хаси и н|у, однако, контакты такого рода отсутствовали (ср. 59% совпадений между н|у и |ауни, но только 54% между н|у и |хаси), и поэтому вывод Т. Гюльдемманна (об «иллюзии» тесных связей между «нижненоссобскими» язы-

¹ Так, в частности, одним из ключевых аргументов Т. Гюльдемманна служит наличие в |хаси сложной системы согласовательных классов, во многом сходной или даже совпадающей с соответствующей системой в !хонг; этот изоморфизм противостоит ситуации в |хам и других «узко-!кви» языках, где такая система не просматривается. Однако классификационной значимостью этот аргумент обладает лишь в том случае, если система маркирования классовых отношений в этих языках оказывается относительно недавней инновацией; между тем, на самом деле она производит впечатление достаточно архаичной, и скорее следует предполагать разрушение исходной системы в |хам и других языках, чем «наращение» ее из неясных источников в фиктивном таксоне «таа + носсоб».

ками и !кви из-за лексических контактов) можно отчасти поддерживать только в отношении |ауни, но никак не в отношении |хаси. Впрочем, учитывая, что для |хаси удается заполнить лишь около 70% стословного списка, представленный сценарий, при всей своей оптимальности, не столь достоверен, как мог бы быть при наличии в нашем распоряжении более полных описаний.

Внутри 50-слового списка, как будет видно ниже, число совпадений между пра-!кви и пра-таа превышает 30 единиц, что позволяет нам безоговорочно рассматривать этот материал внутри одной и той же подсекции. К сожалению, там, где пра-!кви корень этимологически отличается от пра-таа корня, недостаточная разработанность юж.-кой. этимологии обычно мешает определить наиболее архаичный вариант.

Таблица 2. Лексикостатистическая матрица южнокойсанских языков (100-словные списки).

	Нг!ке	Н у	Хегви	Ауни	Хаси	!Хонг	Масарва	Н у ен
Хам	0.73	0.74	0.60	0.51	0.48	0.42	0.42	0.47
Нг!ке		0.88	0.62	0.59	0.48	0.45	0.45	0.48
Н у			0.69	0.59	0.54	0.49	0.49	0.52
Хегви				0.54	0.46	0.42	0.42	0.47
Ауни					0.69	0.46	0.49	0.51
Хаси						0.47	0.42	0.46
!Хонг							0.82	0.77
Масарва								0.78

50-словный список для южнокойсанских языков.

1) «пепел»: (а) хам *!ui*, нху *!qui*, нхе *!wi*; (б) аун *!ʰana*; (в) хнг *ʔda*, мас *ʔwa*. || Праюж.-кой. **!qui*.

Наиболее представительным является этимон (а), т. к. его присутствие отмечено как в ветви !кви, так и в ветви таа (нхе); судя по данным нху, первичным должен быть вариант **!qui* с увулярным исходом. Обращает на себя внимание фонетическое тождество между хам *!ui* 'пепел' и глаголом *!ui* 'гореть' (также в переносном значении: 'болеть (о частях тела)'); однако, за отсутствием глагольных параллелей в других юж.-кой. языках, делать выводы

о производности именной основы было бы преждевременно, тем более что в записях В. Блика, помимо значения 'пепел', для слова *!wi* зафиксировано также значение 'мука' — дополнительный аргумент в пользу первичности именного характера корня.

Формы *ʃda* в хнѣ и *ʃwa*: в мас более сомнительны, т. к. в фонетическом плане первая из них тождественна працентр.-кой. **ʃda* 'пепел' (см. ниже) и легко могла быть заимствована из центр.-кой. источника. Латеральный кликс в мас — скорее всего, результат искаженной передачи палатального кликса (таких примеров в словаре Д. Блик довольно много).

2) «птица»: (а) хам *kxarri* ~ *kxānni*; (б) нгк *!wi*, нху *!q^hui-si*, хег *!wi-θari* [LH], *!^hwi* [Zr.], аун *si=|u*, хас *si=|ɔ*, хнѣ *!ūh^ʔu*, мас *ʃi=|u*, нхе *si=|ou*. || Праож.-кой. **!(q)^hu(-i)*; в других языках !кви ср. также *!kxaу !^hwī*, *!an!e !wi-ŋ*.

Форма *kxarri* в хам не имеет никаких параллелей в других юж.-кой. языках; скорее всего, она заимствована из кхойкхой **kxani* 'птица' (см. ниже). Ее основное значение обычно передается в источниках как 'небольшая птица', что противопоставляет ее лексеме *!erri-tan-ti* 'большая птица'; лексема *kxarri* определена нами как «первичная» лишь на основании того, что она представлена в гораздо большем количестве примеров из текстов. Форма *!erri-tan-ti*, по-видимому, более уверенно претендует на исконный статус, но при этом является очевидным морфологическим производным от хам *!erri(ya)* 'перо'.

На роль общеюж.-кой. этимона 'птица' гораздо лучше подходит корень (б), возводимый к прототипу **!(q)^hu(-i)* (реконструкция исхода кликса затруднена из-за несовпадения артикуляции в нху и хоо, но обращает на себя внимание наличие придыхательного компонента в обоих языках, что повышает надежность сопоставления). В «нижненособских» языках и в группе таа отмечен также особый вариант основы, сочетающийся с «префиксальным» компонентом **ʃi*, значение которого не устанавливается (однако он обнаруживается и в ряде других случаев, например, в слове 'змея': мас *ʃi=|woi*, нхе *si=|wi* ~ *si=|wi*, в аун другой корень с тем же префиксом: *si=|kxaу*). Латеральный кликс в мас — скорее всего, результат опечатки (к сожалению, в целом ряде случаев анализ данных по источникам [Bleek 1929] и [Bleek 1956] дает основания подозревать чисто типографическое смешение знаков | и ||).

3) «черный»: (а) хам !we:n ~ !weŋ, нгк !we ~ !oe, нху !^hoe; (б) хег žwa: ~ žwā: [LH], čwa ~ nčwa [Zr.]; (в) хас //e; (г) хнг †á?ra, нуге †ana; (д) мас //ka.

Основной этимон с значением 'черный' для пра-!кви следует восстанавливать в виде *!^hoe (с суффиксацией в хам); к дистрибуции в !кви ср. также ||кхау ?u=!^hwe: 'черный (о лошади)' (компонент ?u= остается неясным). Хег žwa: и т. д. — скорее всего, заимствование из центр.-кой. источника (ср. вост.-кхой *nzi id.; заимствование позволяет объяснить и необычный для юж.-кой. языков анлаут на носовой кластер в диалектном варианте nčwa [Zr.]). Форма //e в хас не имеет этимологии.

В таа этому корню противопоставлена этимологически не сводимая к нему основа (г) *!fa (второй слог в таа — суффиксальный; ср. в хнг также †ā-be 'чернокожий' и др.). В мас наряду с основой неясного происхождения //ka, зафиксировано также слово *dani*, в [Bleek 1929] определяемое как просто 'черный', а в [Bleek 1956] — как 'черный, темный (например, перепачканный углем)'. Ни одну из этих форм пока не удается спроецировать даже на пратаа уровень, не говоря уже о пра-юж.-кой.

4) «кровь»: (а) хам //xáui-ka ~ //xáui-kəp ~ //xai-ki, нгк //xai, нху //xai-ke, аун //xai^ʔu, хас //xai; (б) хег λ^ʔēi [LH]; (в) хнг !ā:, мас !ā:a, нуге !ā:a.

Здесь налицо четкое разграничение между !кви *//xai и таа *!ā. Форма λ^ʔēi в хег вряд ли связана с *//xai; латеральная аффриката в этом языке обычно отражает старый палатальный, а не латеральный кликс (см. ниже 'собака', 'яйцо', 'луна'), который, напротив, должен сохраняться (см. 'рог', 'ночь' и др.). Впрочем, альтернативной этимологии у этого слова пока что также не обнаружено.

Д. Блик [Bleek 1956: 634] отмечает в мас также форму //xāui 'кровь', но в англо-койсанском словаре [Bleek 1929] она не приводится, поэтому серьезных оснований считать, что в пра-таа основным корнем со значением 'кровь' было то же слово, что и в пра-!кви, у нас нет. Не исключено, что в [Bleek 1956] наблюдается неверная языковая атрибуция корня (например, ошибочно представлен индекс SV вм. SIV).

5) «кость»: (а) хам !wá, хнг †ā, мас //a:, нуге †ā, (?) хег !a: [B.], !a [LH]; (б) нгк //abba, нху //oβa. || Праюж.-кой. *†(u)a (?).

Основной пра-таа корень со значением 'кость' — *!fa или *!ā (в мас налицо такая же ситуация передачи палатального кликса

через латеральный, как и в слове 'пепел', см. выше). При этом высока вероятность того, что хам *!wá* = пра-таа **!a*, поскольку:

— в хам обще-юж.-кой. палатальный кликс в транскрипциях В. Блика и Л. Ллойд регулярно передается как альвеолярный *!* (пока не вполне ясно, стоит ли за этим реальное фонетическое развитие или неадекватность транскрипции);

— соответствие «лабиализованный дифтонг в хам» (транскрибируется обычно как *-wa-*, *-we-*, *-wi-*): «монофтонг в остальных юж.-кой. языках» встречается очень часто и может считаться регулярным (см. ниже 'собака', 'пить' и др.). На данный момент трудно сказать, сохраняет ли здесь хам какой-то важный архаизм, но, с другой стороны, удовлетворительно объяснить вторичность лабиализованной артикуляции пока что также не удается. В любом случае, она никоим образом не может препятствовать совместной этимологизации рассматриваемых форм.

На первый взгляд, сюда же должна относиться и форма *!a*: в хег, однако с этой параллелью связана дополнительная сложность. Дело в том, что в хег альвеолярный кликс обычно отпадает (ср. хег *kwi* 'человек' = хам *!ui* id.; хег *kan* 'дорога' = нху *!an* id.), а палатальный — переходит в латеральную аффрикату (см. ниже 'собака', 'луна' и др.). Возможный выход из положения — считать все формы с *!-* в хег сравнительно недавними заимствованиями (например, из того же хам). Однако общая совокупность лексических данных по хег не дает оснований подозревать массивного «горизонтального» влияния на этот язык со стороны его юж.-кой. родственников, с заходом в сферу базисной лексики. Альтернативное решение — считать, что по крайней мере некоторые слова, имеющие в хег *!-*, все же отражают юж.-кой. основы со старым кликсом; в частности, можно заподозрить, что *!-* восходит к т. н. «второму палатальному» кликсу, о котором см. ниже ('нога', 'один'). В этом случае реконструкция **!a-* должна быть формально «конвертирована» в **!₁a-*.

К группе форм (в) ср. в хам слово *!àbba*, значение которого передается как 'кусочек кости антилопы канна'; по-видимому, инновация в нху.

б) «ноготь»: (а) хам *!uru*, нгк *!uri-si*, нху *!qoro-si*, хег *!ɔla* [В.], аун *!ora-sa*, хнг *!qûle*, мн. ч. *!qûn-sâ*; (б) хнг *!û?m*, мас *!kxa !am-te* (*!kxa* = 'рука', см. ниже); (в) нве *!nu*. || Праюж.-кой. **!qU(-)ro*.

Наиболее широкую дистрибуцию здесь демонстрирует корень (а), хотя в хнг, судя по данным словаря Э. Трэйлла, он сегодня «конкурирует» с синонимом $\ll\hat{a}^?m$, причем можно подозревать, что основным обозначением 'ногтя' (человеческого) в этом языке сегодня уже является именно $\ll\hat{a}^?m$: ср., во-первых, полный спектр значений, приводимых в словаре ($\ll q\hat{u}le =$ 'nail, hoof; $\ll\hat{a}^?m =$ 'fingernail, toenail'), во-вторых, то, что в мас также обнаруживается параллель только для $\ll\hat{a}^?m$: форма мн. ч. $|lm-te =$ хнг $\ll\hat{a}^?ma-t\hat{e}$ (дентальный кликс вместо латерального — скорее всего, типографическая ошибка, обратная наблюдавшейся выше для слова 'птица', где, наоборот, знак \ll ошибочно замещал требуемый).

Парадигма рассматриваемого этимона в хнг показывает, что слово, скорее всего, морфологически разложимо (ед. ч. $\ll q\hat{u}-le$: мн. ч. $\ll q\hat{u}-n-s\hat{a}$, с уже вторичным присоединением нового продуктивного аффикса мн. ч. $s\hat{a}$), однако в группе !кви двусложный вариант стабильно сохраняется во всех формах, а морфологически маркированным оказывается, наоборот, ед. ч. (с показателем сингулятива $-si$ или, в аун, $-sa$). Впрочем, для хам зафиксирована и особая форма мн. ч.: $\ll u\ll\hat{u}ddi \sim \ll u\ll\hat{u}ti$ (В. Блик), $\ll u-\ll\hat{u}tt\hat{a}n$ (Д. Ллойд), с удвоением первого слога и заменой второго на суффикс $-ti \sim -t\hat{a}n$ (точное значение которого неясно, поскольку, согласно глоссированию Д. Блик, он может появляться то в формах мн. ч., то в «эмфатических» формах ед. ч.). Эта парадигма также может служить аргументом в пользу исходной морфологической разложимости формы $*\ll qU(-)ro$.

Этимология формы (в), зафиксированной только в нуге, неясна. Учитывая, что значение 'ноготь' представлено для нее только в раннем словаре [Bleek 1929], в то время как в более позднем словаре [Bleek 1956] она глоссирована как 'палец', возможно, ее следует интерпретировать как заимствование из центр.-кой. источника (ср. нам $\ll^h u\hat{u}ni-$ 'палец').

7) «умирать»: (а) хам $\ll a$, нгк $\ll a$; нху $\ll a$; хег $\ll a$: [Zr.], аун $\ll \hat{a}$, хнг $\ll \hat{a}$; мас $\ll a$, нуге $\ll a$; (б) хас $\ll^h o$. \ll Праюж.-кой. $*\ll a$. В языках !кви ср. также $\ll ku\ll e$ $\ll a$, $\ll k\hat{x}a$ $\ll a$, $g\hat{a}n\ll e$ $\ll a$: id.

Исключительно устойчивый корень; единственная очевидная замена наблюдается в хас ($\ll^h o$; ср. также $\ll^h wa$: 'мертвый'), но очевидной юж.-кой. этимологии она не имеет. Любопытно, что в хег общеюж.-кой. этимон зафиксирован с нулевым исходом

вместо ожидаемой гортанной смычки, причем не только в материалах Цирфогеля, но и в других источниках (*/a:* 'мертвый' [LH]). Не исключено, что здесь имела место контаминация с глаголом 'убивать' (см. ниже).

8) «собака»: хам *!wɪŋ* ~ *!ʰwɪŋ*, нгк *!wɪŋ*, нху *ʰɪɪp* ~ *ʰɪɪp*, хег *||wi* [B.], *λwa* ~ *λweŋ* ~ *ɛwe* [Zr.], *λʰwɪŋ* [LH], аун *ʰɔ:*, хас *ʰʰɪŋ*, хнг *ʰqʰàì*, мас *ʰxai* ~ *!xàì* ~ *!àì*, нүе *ʰi* ~ *ʰi* ~ *ʰxi:*. || Прауж.-кой. **ʰ(q)ʰɔ-*. В языках !кви ср. также *||ку||e !wɪŋ* ~ *!ɪn-yi*, *||кхау ʰɪɪni*, сероа *kuen-ia* id.

Фонологическая реконструкция проблематична на всех уровнях, но, тем не менее, все формы можно считать фонетически совместимыми. На пра-!кви уровне начальный согласный восстанавливается как **ʰ-*, с регулярным отражением (то ли реальным, то ли чисто графическим) *!-* ~ *!ʰ-* в старых записях В. Блика, Л. Ллойд и Д. Блик по хам и нгк и столь же регулярным переходом в латеральную аффрикату в хег; исходная артикуляция лучше всего сохраняется в нху и, по-видимому, в хас. Второй элемент корня — губной гласный или полугласный, сохраняющийся везде, кроме хас, за которым начинается «суффиксальная мора», по-видимому, различная в зависимости от конкретного языка. Наиболее частый и, скорее всего, исходный вариант — **ʰɪ-i-ŋ*, нагляднее всего сохраняющийся в хам и в отдельных идиолектах хег (*λweŋ*, *λʰwɪŋ*); в современных диалектах нху сочетание *-iŋ* стянулось в *-ɪp* ~ *-ɪp*, а в аун практически «растворилось», оставив за собой след в виде назализации гласного.

В языках таа исходной корневой морой можно считать **ʰqʰa-*, судя по форме мн. ч. *ʰqʰà-ba-tê* в хнг (*-i* — один из классифицирующих суффиксов 1-го класса именных основ); транскрипционный разницей в мас и нүе связан либо с идиолектным варьированием, либо с обычными для «трудноразличимого» палатального кликса ошибками в записи.

Отождествление форм в таа и !кви на первый взгляд неочевидно, т. к. полное фонетическое совпадение налицо только для палатальной основы кликса. В частности, для !кви хотелось бы ожидать появления увулярного исхода, аналогичного отмеченному Трэйллом для хнг *ʰqʰàì*; но в том единственном языке !кви, для которого увулярные исходы кликсов отмечаются регулярно (нху), в данном случае наблюдается простое придыхание (*ʰɪɪp*). С другой стороны, соответствие «хнг *-qʰ-* : нху *-ʰ-*» является рекур-

рентным (см. ниже 'волосы'), и трудности, связанные с его интерпретацией, не должны препятствовать этимологизации (не исключена в том числе и неточность транскрипции в одном из этих двух языков).

Что касается различий в вокализме, то здесь возможно несколько сценариев, в частности: (а) реконструкция особого гласного *ɔ, регулярно делабиализованного в таа (а также в хас), но сохраняющего губную артикуляцию в !кви; (б) реконструкция простого *и в рамках трифтонгического сочетания *-иа-і, упрощенного на пра-таа уровне в *-а-і. Оба сценария имеют свои слабые и сильные стороны, но, в любом случае, соответствие «+лабиализация в !кви : -лабиализация в таа» также рекуррентно (ср. хотя бы такой пример, как нху *su-n ~ su-n* 'жир' = хн҃г *sã*: id.). На уровень внешнего сравнения допустимо выводить предварительную реконструкцию *ʃ(q)^hɔ-, в которой адекватно отражены как несомненные, так и спорные в плане интерпретации соответствия.

9) «пить»: хам *kxwā ~ kwī*: ~ *kxwī*.; н҃гк *kxa*: ~ *kxā ~ kxē*, нху *kxāi*, хег *kxi ~ kxa*: [Zr.], кxēi [LH], аун *kxā*: ~ *kxē*, хас *kxa*, хн҃г *kxā^h*, мас *kxā ~ kxe*, н҃уе *kxa*: ~ *kxa-u*. || Праюж.-кой. **kxa*-. В языках !кви ср. также ||ку||е *kxā ~ ||kxwāi*, ||кхау *kxa-ti*, г!ан!е *kxa*.; сероа *kxā* id.¹

Все формы очевидно родственны. Разнообразие вариантов огласовки отражает слияние исходного гласного корня с различными грамматическими морфемами, в том числе показателями согласовательных классов и видо-временными частицами (для хег, например, Цирфогель помечает вариант *kxi* как основу настоящего, а вариант *kxa*: — как основу прошедшего времени). Лабиализация в хам, по-видимому, вторична, поскольку она не подтверждается данными ни по одному из других юж.-кой. языков (в отличие от, например, 'собаки').

Бросается в глаза полное сегментное совпадение этой основы с працентр.-кой. **kxa* 'пить' (см. ниже); однако, учитывая надежность реконструкции на праюж.-кой. уровне, однозначно интер-

¹ В словаре [Bleek 1956: 153] эта форма напечатана как *daw*, что может создать иллюзию регулярного выпадения *kx*- в сероа (аналогично ситуации в нама, см. ниже); на самом деле это просто типографическая ошибка — в персональной орфографии К. Вурса, разработанной им для записи койсанских данных, он использовал для обозначения заднеязычной аффрикаты специальный знак *o*.

претировать это совпадение на контактной основе недопустимо, в отличие от многочисленных индивидуальных заимствований в юж.-кой. языки из нама, наро и других центр.-кой. языков.

10) «сухой»: (а) хам //o: ~ //ɔ:, нгк kxo: (?), нху //o:, хнг //úa; (б) аун //xom; (в) хнг /ɔ̀: . || Праюж.-кой. *//o- ~ *//u-.

Для большинства вымерших юж.-кой. языков слово вообще не засвидетельствовано. Тем не менее, фонетическая совместимость хам-нху //o: с хнг //úa (различия в вокализме здесь пренебрежимы при условии выделения в хнг «корневой моры» //ú-) позволяет вывести по крайней мере один этимон на прауровень. Сюда же можно с уверенностью подключить хег //o: [B.] 'жаждущий', а также, по-видимому, нуге //o 'теплый' (хотя в последнем случае уже не исключена семантическая контаминация или даже прямое заимствование из центр.-кой. *//o 'горячий, теплый').

Аун //xom — сомнительная форма, едва ли совместимая с остальными из-за невозможности удовлетворительно объяснить несовпадения как в исходе кликса, так и в «суффиксальной море». В хнг, помимо этого, наблюдается полисемия: по словарю Трэйлла невозможно точно установить разницу в семантике между //úa и /ɔ̀:. Последнее, однако, легко может быть интерпретировано как недавнее заимствование из центр.-кой. источника (см. ниже працентр.-кой. *//o 'сухой').

11) «ухо»: хам /ú-ntu, нгк /úwe: ~ /úwe-ntu ~ /ú:-ntu, нху /fui-si, хег /ʒwe [B.], /ɓwe [Zr.], /ɓwí: [LH], аун /fui, хас η=k'u=ʃa:m мн. ч. 'уши', хнг /fúáʰ, мас /wa:, нуге /fu-ša, мн. ч. /fui-te. || Праюж.-кой. *//u-. В языках !кви ср. также //kxaу /úwe-ntu и, вероятно, //ku||e de 'уши' (хотя данные по этому языку чрезвычайно скудны, не вызывает сомнения факт утраты кликсовой артикуляции с развитием в d- как минимум у нескольких щелчковых фонем: ср. dɔa-xu 'небо' = хам /wa-xu id., dɔwe 'три' = хам /wanna id., dɔ 'антилопа гну' = хам /ú).

Корень сам по себе устойчив, хотя праюж.-кой. парадигму восстановить сложно из-за морфологического разнообразия в языках-потомках. «Суффиксальная мора» оказывается представлена здесь и гласным -a- (в хнг и мас), и гласным -i- (в нуге, нху, аун), и нулем, на место которого встают более сложные именные суффиксы (сингулятив -ša в нуге; -ntu в хам — по-видимому, позиционный вариант суффикса -tu, часто встречающегося в языках !кви в сочетании с частями тела, см. ниже 'нос' и др.).

Тем не менее, практически все соответствия внутри «корневой моры» регулярны: налицо и закономерная передача палатального кликса как альвеолярного в хам и нгк транскрипциях Блик-Ллойд, и регулярный переход его в латеральную аффрикату в хег (см. 'собака'), хотя неясным при этом остается вариант с \check{z} - в записи Д. Блик (обычно латеральным аффрикатам, которые в этом языке последовательно отмечают исследователи 1950-х гг., в транскрипции Блик соответствуют латеральные кликсы). В «нижненособских» языках наблюдается также колебание между $\check{f}u$ - (аун $\check{f}u-i$) и $\check{f}a$ - (в сложной форме мн. ч. в хас: $\eta=k\check{u}=\check{f}a:m$, где $k\check{u}$ — уникальный для хас префикс мн. ч., а $\eta=$, скорее всего, посесивный префикс 1-го л.), ср. также аун $\check{f}ae$ 'ухо' в [Bleek 1929: 35] (менее надежная форма). Скорее всего, это результат стяжения из $*\check{f}u-a$ - (т. е. того же морфемного комплекса, который представлен в языках таа), хотя показать регулярность такого стяжения не удастся из-за отсутствия данных.

12) «есть»: хам \tilde{a} : ~ $h\tilde{a}$:, нгк \tilde{a} ~ \tilde{e} ~ $\tilde{e}\tilde{i}$, нху \tilde{a} , хег \tilde{a} ~ \tilde{e} [B.], $\tilde{a}\tilde{i}$ [Zr.], $\tilde{a}\tilde{i}$ ~ $\tilde{a}\tilde{i}\eta$ [LH], аун \tilde{a} ~ $h\tilde{a}$ ~ $h\tilde{a}a$, хас a :, хнг \tilde{a} :, мас \tilde{a} ~ a : ~ e : ~ ε :, нхе \tilde{a} ~ \tilde{e} . || В языках !кви ср. также ||ку||е \tilde{e} , ||кхау \tilde{a} , г!ан!е \tilde{a} - \tilde{a} id.

Праформа восстанавливается в виде $*\tilde{a}$ или, возможно, $*\tilde{e}$ (такая реконструкция позволила бы учесть вокалические колебания a ~ e , но вполне вероятно, что они на самом деле обусловлены морфологическими факторами, как и во многих других аналогичных ситуациях, в которых задействованы простые глагольные основы). Спорадическое исчезновение в отдельных рефлексах назализации также нелегко объяснить, но, учитывая дистрибуцию назализованных рефлексов, носовую артикуляцию в этом корне явно следует считать архаичной.

13) «яйцо»: (а) хам $!aui$ ~ $!auiwi$ ~ $!k\check{a}i:wi$, мн. ч. $!ui-t\check{a}n$, нгк $!^h\tilde{a}\tilde{i}$, нху $\check{f}ui$, хег $\check{a}wi\eta$ [Zr.], $\check{a}w\tilde{i}$ [LH], аун $!i\tilde{i}$ 'страусиное яйцо' (?), хнг $\check{f}\tilde{i}\tilde{a}$, мас $!wa$:, нхе $!w\check{o}\tilde{i}$; (б) хас $k\check{u}i$.

Все формы, кроме хас, сводимы к инварианту $*\check{f}u$ -, хотя определить старый исход кликса трудно: здесь наблюдается не только нулевой рефлекс (как в хам и нху), но и глоттализация (в «локальных» вариантах хег и хам), и придыхание (в нгк), и даже озвончение (в языках таа). Тем не менее, все эти «соответствия» (по меньшей мере несколько из которых — результат неверной транскрипции) до определенной степени рекуррентны и не

должны препятствовать этимологическому сближению всех форм, представленных в группе (а). Некоторое сомнение вызывает лишь форма в аун, не столько из-за более узкой семантики (другого «кандидата» на значение 'яйцо' в этом языке мы все равно не знаем), сколько из-за альвеолярного кликса; однако и здесь иногда наблюдаются ошибки или колебания в записи (ср. аун *ʃxo* ~ *!xo* 'иголка' = хнг *ʃxí*: id.), так что форму *!ũĩ* можно с некоторыми натяжками привлекать к сравнению.

Стоит также обратить внимание на любопытное вокалическое чередование в хам: *!au(-i)* в ед. ч., но *!ui-* в мн. ч.; не исключено, что оно отражает более старый вариант основы **ʃai-*, упростившийся в **ʃu-* вследствие присоединения еще одного суффикса.

14) «глаз»: (а) хам *сʰaxáũ* ~ *сaxáũ*, нгк *сáxũ* ~ *сaxe:t*, нху *сʰaxat*, хег *сáxũ* ~ *саи* [B.], *sagũ* [Zr.], *сʰagũ*, мн. ч. *сʰaŋ* [LH], аун *соо* ~ *сʰaxũ*, хас *сxɔ*; (б) хнг *!ũĩ*, мас *ʃkʷwē* ~ *ʃkʷwĩ*, нүе *ʃũ*. || В языках !кви к группе форм (а) ср. также *ʃku*||е *са-ху*, *ʃkxau* *сʷ-хо?*, *g!an!e tya-xũ*.

В таа и !кви здесь представлены разные корни. !Кви **сʰax(a)ũ* — нетипичная для юж.-кой. (и «койсанских») языков вообще) двусложная структура, в историческом плане представляющая собой словосложение, причем «основной» морфемой нужно считать первую — в хег еще сохранялась, по-видимому, более архаичная ситуация, когда компонент *-ху* был отделен (ср. мн. ч. *сʰa-ŋ*). Сам этот компонент можно уверенно отождествлять с пра-!кви **ху* 'лицо' (см. ниже 'голова'); при этом любопытно, что первая морфема (**сʰa-*) фонетически совместима с хнг (т. е. таа) лексемой *sãʔã*, которая также имеет значение 'лицо', 'поверхность'.

Дальнейший анализ, впрочем, показывает, что морфема **ху* в языках !кви находилась по меньшей мере на раннем этапе грамматикализации, т. к. ее десемантизированный вариант обнаруживается и в целом ряде других слов: ср. хам *ʃwa-xũ* 'небо' = нгк *!a-xũ* id., а также варьирование между записями *ʃŋ* и *ʃŋ-xũ* 'дом' в записях К. Доука по *ʃхомани* и т. д. Таким образом, пра-!кви **сʰa-xũ* следует анализировать не как композит сочинительного типа, состоящий из двух автономных лексических корней, а скорее как производную основу, состоящую из корня и суффикса. Семантическая связь между 'лицом' и 'глазом' типологически естественна, хотя направление развития однозначно не устанавливается; не исключено, однако, что на «ранне-!кви» уровне сложная основа

**сга-хи* 'глаз' была образована как раз для того, чтобы отличить ее от простой основы **сга* 'лицо'.

В этом случае большие шансы на юж.-кой. архаичность этимона имеет группа форм (б), представленная в таа. Здесь совершенно не ясными остаются колебания в записи кликсов: либо мы имеем дело с разными корнями (что все же маловероятно, ввиду надежности соответствий в остальной части слова), либо с ошибочной транскрипцией в нескольких источниках. Учитывая, что альвеолярный кликс в передаче Трэйлла подтверждается более ранними данными Э. Вестфаля (н|амани !*ūi*), пра-таа форму можно условно восстанавливать в виде *!*ūi*-.

15) «огонь»: хам /*ri*, нгк /*ri*, нху /*ri*, хег /*e* ~ /*ri* [B.], /*i* [Zr.], /*e* [LH], аун /*ri*, хас /*i*, хнг /*ã*ː, мас /*ã*ː ~ /*a*, нуге /*ã*. || В других языках !кви ср. также ||ку||*e* /*e* ~ /*ri*, ||кхау||*ri*, сероа /*ei*, г!ан!е /*ri*.

Все формы родственны. Начальный кликс практически однозначно восстанавливается в виде */*ɾ*- (необъяснимой остается лишь утрата глоттализации в варианте хег основы согласно Цирфогелю, а также в хас; возможно, это связано с тем, что гортанная смычка в позиции между переднеязычным кликсом и передним гласным *i* произносилась менее заметно, чем в других позициях). Труднее обстоит дело с реконструкцией вокализма; скорее всего, первичен вариант */*ri* в !кви, т. к. пра-таа */*riã* можно объяснить как результат вокалического слияния: корень */*ri*- + **-ã* (показатель 2-го класса существительных) → */*riã* → */*riã* (дифтонгические сочетания вида *iV* в языках таа запрещены).

Сразу бросается в глаза явное сходство основы */*ri* с працентр.-кой. /*e* 'огонь' (см. ниже); однако, поскольку основа уверенно реконструируется на праюж.-кой. уровне, анализировать ее как заимствование из центр.-кой. источника на данном этапе исследования недопустимо (совпадение может отражать и генетическое родство).

16) «нога»: (а) хам /*oá* ~ /*wa*, нгк /*a* ~ /*a*:*-хи*, хнг /*ɸi*ː, мас /*o* ~ /*o* ~ /*ɸ*, нуге /*ɸi*; (б) нху /*kxi*:*-ke*; (в) хег /*kxe* [B.], /*hi*ː [Zr.]; (г) аун /*kxai*, хас *n*=/*hai*.

«Графическое» соответствие «хам /*-* : нгк /*-*» в лексических материалах Блик/Ллойд встречается неоднократно (см. ниже 'один', а также, например, хам /*wei* ~ /*wai* 'женская грудь' = нгк /*e*: ~ /*ē*: id.; хам /*ai* 'ребро' = нгк /*ai* id.) и явно не может быть списано на неадекватность транскрипции: латеральная артикуляция в нгк

подтверждается современными записями по нху (ср. $!^h\tilde{a}i$ 'грудь, молоко'), а альвеолярный $!$ в хам теоретически мог бы передавать и палатальную артикуляцию ($\ddot{}$), но вряд ли латеральную ($!l$ - в старых источниках путаются довольно редко). В работе [Starostin 2008a] для этого соответствия на пра!кви и праюж.-кой. уровне постулировалась особая фонема $*\ddot{l}$, без попытки фонетической интерпретации. Не исключено, однако, что здесь все же представлено особое позиционное развитие старого палатального кликса, т. к. в таа этим случаям обычно соответствуют также палатальные рефлексы (см. 'один', а также $хнг \ddot{q}h\tilde{e}$ 'грудь, молоко').

В данной ситуации хнг также показывает начальный палатальный, так что все формы группы (а) фонетически совместимы и могут быть возведены к условной праформе $*\ddot{t}o$ или $*\ddot{t}u$ (\rightarrow !кви $*\ddot{t}o$ -а с присоединением дополнительного именного суффикса). Это автоматически решает вопрос об «оптимальном кандидате» на значение 'нога' в праюж.-кой. Что касается форм (б), (в), (г), то, несмотря на отдельные элементы звукового сходства, все они оказываются фонетически несовместимыми и к тому же не этимологизируются на взаимной основе.

17) «волосы»: хам $!u \sim !^h\acute{u}$, нгк $!u \sim !^h\acute{u}$, нху $!^hu:(-ke)$, хег $!^ho$ [B.], $!^hu$ -zi [Zr.], $!^h\ddot{u}$ [LH], аун $!^h\acute{o}o$, хас $!o$, хнг $!q^h\ddot{u}\tilde{a} \sim !q^h\ddot{u}\tilde{a}$, мас $!wa:-ni$, нуге $!un-te$.

Все формы очевидно родственны; общий инвариант восстанавливается в виде $*(q)^hu$ - с разными суффиксами (в !кви в основном представлены различные продуктивные показатели сингулятива, ср. в хег [Zr.] ед. ч. $!^hu$ -zi, мн. ч. $!^hu$ -η; в таа корневая мора сливается с показателем 2-го класса \tilde{a}). Точная артикуляция исхода кликса остается неясной; большинство языков сохраняет, по видимому, архаичную придыхательность, но сложный характер исхода и его идиолектное варьирование в хнг указывают на то, что полный набор артикуляторных признаков был богаче (возможно, совпадая с соотв. набором в хнг — «придыхательность» + «увулярность» + «звонкость»?). В хнг ср. также однокоренное слово $!q^h\acute{u}$: 'пучок волос, волосы в хвосте животного'; по своему морфологическому устройству эта лексема стоит ближе к формам в !кви.

18) «рука»: (а) хам $!kxa$, нгк $!kxa$, нху $!kxa$, хег $!kxa$ [B.], мн. ч. $!kxa$ -η [Zr.], аун $!kxa \sim !kxan$, хас $n=|xa$ -η $\sim k\eta\eta=k\eta u=|xa$ -η, хнг $!kx\acute{a}$, мас, нуге $!kxa$; (б) хег k^hi [Zr.], q^hi : [LH]. || Праюж.-кой. $*!kxa$. В других языках !кви ср. также $!ky||e$ $!kxa$; $!an!e$ $!a$, возможно, также, сероа ka : 'рука (arm)'

(последнее, впрочем, сомнительно, т. к. дентальный кликс в сероа обычно сохраняется).

Формы в хас представляют собой слияния либо с посессивными префиксами (*n*= 'мой, мои'), либо с префиксом мн. ч. *kni*=; не вполне ясным остается дополнительный префиксальный компонент *kaiŋ*=.

Отдельного объяснения требует странный супплетивизм основ ед. и мн. ч. в хег, согласованно отмеченный в материалах 1950-х гг., но, по-видимому, неизвестный в том диалекте хег, который был описан Д. Блик (ср., в частности, словосочетание *a /kxa*, которое она переводит как 'твоя рука', а не 'твои руки'). Полностью исключить архаичность оппозиции **q^hi* (ед. ч.) : **kxa(-ŋ)* (мн. ч.) нельзя, но вероятность этого очень низкая — ни в одном другом языке следов такого супплетивизма не обнаружено. С другой стороны, никакой внятной юж.-кой. этимологии для хег *q^hi*: предложить также не удастся.

19) «голова»: (а) хам *ʃa*: ~ *ʃā*:, нгк *ʃa*(:), нху *ʃa*:, хег *ʃa*: [B., Zr.], аун *ʃa*:, хнг *ʃan*, мас *ʃa* ~ *ʃan*, нуге *ʃan*; (б) хас *ŋ*=*xó*. || Праюж.-кой. **ʃa*-. В других языках !кви ср. также ||*ku*||e, ||*kxa*||*ʃa*:, сероа *ʃa*.

Реконструкция в целом однозначна. Только для хас записи Р. Стори эксплицитно предполагают лексическую замену на пра-!кви **хи* 'лицо', хотя необходимо отметить, что и в аун слово *хи*: несколько раз отмечено в контекстах, где оно может означать только всю 'голову' в целом, ср., например, *!wara tane k^ha: ke xu*: 'девочка несет на голове воду' [Bleek 1956: 342]. На основании такого рода примеров Д. Блик предположила, что «родным» в значении 'голова' словом для аун нужно считать именно *хи*:, в то время как *ʃa*: на самом деле отражает влияние нху («*ʃ*хомани»).

20) «слышать»: хам *ttu* ~ *ttúi*, нгк *tu* ~ *tu:i*, нху *ɕu*:, хег *tu:i* [B.], *tu* [Zr.], *tu[?]-bi* [LH], аун *tu*: ~ *tu:i*, хнг *tā́*, мас *tāa* ~ *tāan*, нуге *tān*. || Праюж.-кой. **tu*-. В других языках !кви ср. также ||*kxa*||*tū*, сероа *tu*.

Несмотря на «моноконсонантность» сравнения, различия в вокализме между формами, наблюдаемыми в !кви и таа, не должны служить серьезным препятствием для совместной этимологизации. Уже внутри одной только ветви !кви представлено несколько вариантов суффиксальной моры, иногда ассимилирующей исходный вокализм корня: ср. в нгк — *tu*: 'слышать', *tu-ā*: 'слышал' (прош. вр.), *tu-i* 'слушать' [Bleek 2000: 24]; в хег — наст. вр.

tu-bi, прош. вр. *tuw-a*, буд. вр. *tu*, императив ед. ч. *to*, мн. ч. *to-u* [Zr.]; в аун наряду с формой *tu*: в более раннем источнике [Bleek 1929: 46] приводится также вариант *ta.ã ~ ta:a*, практически совпадающий с аналогичными формами в языках таа.

С другой стороны, в то время как в таа глагол *tã* полисемичен (хн: 'слышать / чувствовать'), в !кви значению 'чувствовать' соответствуют следующие формы: хам *ta ~ tã ~ ttã*, нгк *tian*, нху *ɕ^hĩ*, аун *tian*, ||кхау *t^hã*, сероа *t^ha*. Вокализм этих форм полностью совместим с вокализмом в таа (за исключением признака глоттализиции, который в материалах по !кви не отмечен; впрочем, это может быть и случайностью), что предполагает возможность исходного этимологического различия между **tu* 'слышать' и **tã* 'чувствовать', с последующей лексико-семантической нейтрализацией этой оппозиции в таа.

Тем не менее, нельзя исключать и исходное этимологическое единство этих основ, не говоря уже о том, что за нейтрализацией в таа могли стоять и чисто фонетические причины, например, стяжение **tu-ã* → **tã*. Фонетическая совместимость огласовки глагольного корня *-i-* в !кви и *-a-* в таа подтверждается и другими случаями, ср., например, хам ||^hи: 'мочиться; моча' = хн ||^q*h*áa 'мочиться'.

Исходя из этих соображений, мы пока что все же предлагаем для всех перечисленных форм совместную этимологизацию и, следовательно, лексикостатистическое тождество.

21) «сердце»: (а) хам /ĩ:, нгк /ai ~ /e, нху /e:, аун /ɛ: ~ /ɛ:, хас n=^ha- /e, хн /q^han, мас /i:, нуе /an; (б) хег *kele* [Zr.], *kelen* [LH]. || Прауж.-кой. * /q^ha- (?). В других языках !кви ср. также ||ку||e /ɛ:, ||кхау /ae ~ /ai-si.

Форма (б) в хег — прозрачное заимствование из банту (ср. прабанту **kòdò* ~ **kòlò* 'сердце'). Все остальные формы фонетически совместимы, хотя реконструкция как исхода кликса, так и гласного корневой моры проблематична. Хаотичное варьирование исходов в языках !кви и таа, по-видимому, намекает на комплексную артикуляцию в прауж.-койсанском, близкую к глоттализированно-увулярной в хн. Что касается вокализма, то здесь самую широкую дистрибуцию обнаруживает гласный *a* (в нгк, хас, хн, нуе): формы с огласовкой *e*, *i* можно трактовать как результат стяжения корневой моры * /q^ha- с суффиксальной **-i* ~ **-e* (ср. особенно варьирование в нгк по данным Д. Блик). В языках

таа исходный вокализм корня сохраняется лучше из-за того, что основа содержит другую суффиксальную мору (*-n). Неясной остается форма в хас (редупликация?), за исключением префикса *n=*, скорее всего, отражающего старую possessивную форму.

22) «рог»: хам //^hē: ~ //^hēī: ~ //ē:, нк //āī, нху //q^hoe-si, хег //ē [B.], //i: [LH], аун //ēī, хнг //āē, мас //m-ša, нуе //ā. || Праюж.-кой. *(q)^hō-.

В словаре [Bleek 1956] варианты этого корня в разных юж.-кой. языках чрезвычайно тяжело отделить от фонетически сходного слова 'зуб' (см. ниже); однако рефлексy в качественно записанных диалектах, таких, как нху и хнг, однозначно свидетельствуют о том, что речь идет о двух совершенно разных словах (впрочем, отдельные случаи контаминации исключить также нельзя). Тем не менее, соответствие «нху -q^h- : хнг -Ø-» уникально, и от совместной этимологизации данных форм мы не отказываемся лишь потому, что и в этих двух случаях нельзя уверенно исключить какую-то транскрипционную погрешность.

Исходный вокализм корня хотелось бы восстанавливать в виде *a, но этому эксплицитно препятствует вариант //q^hoe- в нху (подтверждаемый, помимо прочего, записанной ранее формой //oi-si [Westphal 1965: 144]). На пра-!кви уровне более вероятен вариант *//q^ho-i или *//q^hō-ī (с деназализацией в нху); на пра-таа уровне уже был, несомненно, представлен вариант *//ā-, что опять-таки возвращает к вопросу о возможности реконструкции для праюж.-кой. языка отдельной фонемы *э. Не исключено, тем не менее, и какое-то особое позиционное развитие (например, юж.-кой. *-ōī → пра-таа *-ōē → *-āē); вопрос требует дополнительного изучения.

23) «я»: хам η ~ n, нк η ~ n, нху η, хег ?η ~ ?n ~ ?in ~ ?iη ~ ?ī [Zr.], ?iη-?e [LH], n ~ η ~ aη [B.], аун n ~ η ~ na, хас η ~ n, хнг ñ, мас n ~ na ~ η ~ ηa, нуе η ~ n ~ na. || Праюж.-кой. *η. В других языках !кви ср. также //ку||e η ~ nie, //кхау n ~ η, сероа ?n, !ан!e n ~ η ~ ηe.

Исходная форма местоимения 1-го л. ед. ч. восстанавливается как *η, поскольку заднеязычную артикуляцию носового, имеющую столь широкую дистрибуцию, невозможно объяснить вторичным развитием. Переднеязычный вариант n (na), который, согласно словарю Трэйлла, является единственной формой анализируемого местоимения в хнг (ср., однако, варианты na ~ η в старой работе [Maingard 1958: 106]) — такой же результат аккомодации «необычного» фонетического слова в различных позициях,

как и отмеченный в словаре [Bleek 1956] для всех без исключения юж.-кой. языков вариант *t ~ at* в позиции перед начальным губным согласным (ср. пример из нуге: *t θwa kei e* 'это мой ребенок', а также из хеґ: *t |ue:ŋ* 'я сплю' ← *ŋ θue:ŋ с переходом огубленности кликса на предшествующее местоимение).

Варианты с начальной гортанной смычкой (?ŋ и т. д.) следует рассматривать как свободные: поскольку сонанты в юж.-кой. языках функционально приравнены к гласным (практически не встречаясь в начальной «консонантной» позиции), праюж.-кой. *ŋ 'я', скорее всего, реализовывалось с автоматическим «твердым приступом», как любое слово, начинающееся с гласного сегмента; следы этой преглоттализиции наиболее наглядно отражены в хеґ транскрипции Цирфогеля.

24) «убивать»: (а) хам |á ~ |i: ~ |^há, нгк |a ~ |i: ~ |^hi, нху |^ha; (б) хеґ |iŋ ~ |eŋ [Zr.]; (в) хас !au; (г) хнг qâi, мас |^ha, нуге |^hwan ~ |^hwan-naко (?).

Формы группы (а) обнаруживают очевидное фонетическое сходство с юж.-кой. основой *|a- 'умирать', см. выше, но детальный анализ показывает, что оно, скорее всего, случайно. Наиболее показательна оппозиция в нху, где новейшие данные однозначно подтверждают глоттализированный исход для глагола 'умирать' и придыхательный — для 'убивать', но и в более старых источниках это различие проводится довольно последовательно, хотя отдельные случаи контаминации (как среди носителей языков, так и в результате исследовательских ошибок) исключить нельзя.

Относительно группы форм (г) следует отметить, что латеральный кликс в мас и нуге на самом деле может ошибочно передавать в транскрипции Д. Блик некликсовую увулярную артикуляцию. Надежно обосновать это трудно, т. к. начальные увулярные смычные в хнг встречаются редко, а достоверные параллели к ним в мас и нуге — еще реже; ср., однако, такие примеры, как хнг qâ?na 'соль', 'соленый' = мас !xane id.; хнг qâla 'копать' = мас |ale ~ |a:la id. (|a:la — скорее всего, опечатка вместо *|a:la), где один и тот же увулярный q- в хнг передается различными кликсами. Таким образом, по меньшей мере хнг qâi и мас |^ha: можно свести к единой праформе *qa-; нуге |^hwan остается под вопросом из-за необъяснимости дифтонгического элемента wa (= oa).

В целом этимон 'убивать' оказывается довольно неустойчивым — все три подгруппы !кви имеют разные эквиваленты,

несводимые друг к другу и (пока что) не этимологизируемые на взаимной основе. Для некоторых языков отмечена полисемия: так, хег *λiη* = 'ударять', 'убивать' (в [LH] этот же глагол транскрибируется как *λeиη* и приводится только в значении 'ударять'); мас *||a:*, скорее всего, является диалектным вариантом той же основы, что и *||kai* 'убивать ударом по голове'; в хнц обращает на себя внимание полное сегментное тождество *qâi* 'убивать' и директивного наречия *qâi* 'с резким движением вниз' (ср. *qâi ||úm* 'топнуть (ногой)', *qâi uLV ||ân* 'резко наклонить голову' и т. п.).

Исходя из этих наблюдений, при ранжировании «кандидатов» на значение 'убивать' в пракож.-кой. языке предпочтение имеет смысл отдать !кви основе **||^ha-*, не только потому, что из всех «кандидатов» она выводится на максимальную хронологическую глубину (будучи общей для хам и всего кластера нгк-нху), но и потому, что в подавляющем большинстве примеров из текстов она засвидетельствована именно в значении 'убивать' (разными способами) и никак другом. Тем не менее, к внешнему сравнению вполне допустимо с осторожностью привлекать и все остальные формы.

25) «лист»: (а) нгк *xerro:*; (б) нху *bla:r-si*; (в) хег *k^ha-si* [LH]; (г) хнц *||āna*, мас *||a:na*; (д) нуге *||abu*.

Серьезный разброс вариантов говорит не столько об общей неустойчивости этого этимона в юж.-кой. языках, сколько о возможном отсутствии его как такового на пракож.-кой. уровне. Форма, зафиксированная в нху — недавнее заимствование из африкаанс; к более старому варианту (нгк *xerro:*) ср., вероятно, хам *kerru* 'зеленый', 'зелень, трава, листва'. Форма в хег — заимствование из свази (*li=k^hasi* 'лист'). Со словом *||abu* связана путаница: в [Bleek 1929] эта форма дается как S₆ (т. е. нуге), в [Bleek 1956] — как SIV (т. е. аун), и есть основания предполагать в одном из этих источников опечатку в нумерации (скорее всего, SIV по ошибке записано вм. SVI, т. к. в главной работе Д. Блик по аун [Bleek 1937] формы *||abu* нет; показательна также форма мн. ч. *||abu-te*, т. к. суффикс мн. ч. *-te* более типичен для языков таа, чем для «носсобской» ветви !кви).

Таким образом, единственная форма, уверенно отмеченная хотя бы в двух разных языках (диалектах) — это таа **||ana*. К сожалению, и здесь, скорее всего, мы имеем дело с относительно не-

давним заимствованием: внешнего подтверждения в языках !кви этот корень не находит, зато полностью совпадает с працентр.-кой. **ʔáná* 'трава', 'лист' (см. ниже), рефлексy которого представлены и в западнохойских языках, до сих пор находящихся в активном контакте с диалектами таа. Учитывая общую неустойчивость понятия 'лист' в юж.-кой. языках, заимствование здесь намного более вероятно, чем сохранение старого корня.

26) «вошь»: хам *θwɪŋ* ~ *θoen*, нгк *θoin-ya*, нху *θu-si*, хег *θe-zi* [LH], хнг *θú:*. || Праюж.-кой. **θu-*.

За наиболее архаичный вид корня условно принимается вариант в таа (с фарингализацией гласного); разнообразие рефлексов по языкам объясняется сочетаниями с различными именными суффиксами. Для хнг Трэйлл приводит также альтернативную лексему *ʔkóni* 'вид вши', однако последняя не имеет параллелей в других юж.-кой. языках, а ее фонетическое сходство с общекалахари-кхой **kuni* 'вошь' (см. ниже) заставляет предполагать возможность заимствования (что, правда, не объясняет латеральный кликс в хнг, если только эта запись не ошибочна).

27) «мясо»: (а) нгк *θwai:*, нху *θoe*, хег *θwa:* [B.], *θa:* [Zr., LH], аун *θwe* ~ *θwi*, хас *θwi:*, хнг *θàye*, мас *θwe*, нхе *θwe:* ~ *θwi:*; (б) хам *ā:* ~ *āŋ* ~ *eŋ* ~ *eŋ-ya*.

Корень (б), встречающийся и в других языках !кви (||кхау *ʔāŋ*; возможно, также ||ку||е *ða-si*, хотя в последнем трудно объяснить огубленность гласного) — довольно прозрачное именное производное от **ʔā* 'есть' (см. выше) и, таким образом, скорее инновативен в значении 'мясо' по сравнению с более архаичной основой **θa-* (представленной в сочетаниях с различными суффиксальными морфемами: **θa-i* ~ **θa-e*). Форма *θwai:* зафиксирована и в хам, но, судя по тому, что в [Bleek 1929] в качестве единственного эквивалента слова 'meat' приводится именно хам *ā:*, можно думать, что *θwai:* в этом языке на момент фиксации уже употреблялось скорее в переносном значении 'дичь', 'животное', которое свойственно и другим юж.-кой. языкам и также было представлено у этого корня уже на праюж.-кой. уровне (ср. хам *θwai:-tən* *ʔe* и *ti te:ŋte:ŋ* 'животные быстро ложатся на землю' [Bleek 1956: 685] и др. примеры там же).

В хнг ср. также однокоренное *θā* ← **θa-ā* 'стадо антилоп', 'плоть, мясо'.

28) «луна»: (а) хам !au!áuru ~ !a!áuru ~ !au!árru, нгк !orре ~ !urru ~ turro, нху ʃoro, хег klolo [B.], ʃolo [LH]; (б) аун !oi ~ !ōĩ, хас !^hwi; (в) хнг !q^han, мас !xan, нуге !xan. || В других языках !кви ср. также ||кхау ʃóró; ||ку||е t^olo (к группе форм (а)).

Три перечисленных группы форм несводимы друг к другу. Основа (а), представленная в «узко-!кви» языках (не включая «нособские»), должна восстанавливаться как *ʃoro, с регулярным развитием палатального кликса в латеральную аффрикату в хег (а также в глоттализированный t- в ||ку||е, ср. аналогичный пример: ||ку||е t^o 'мужчина' = нху ʃo: id.). Вариант turro в нгк помечен как «необычный» уже в записях самой Д. Блик (т. е. должен рассматриваться как редкое идиолектное развитие). Странная «редупликация» в хам скорее должна рассматриваться как словосложение: удвоение корневого морфемы в этом языке регулярно служит для образования мн. ч., но в формах ед. ч. обычно не встречается. Один из возможных вариантов — развитие из исходного *!aui-ʃoro 'полная луна' с контекстно обусловленным выпадением носового согласного; возможны и другие гипотезы, менее убедительные с точки зрения семантической типологии.

В «нособских» языках вместо этой основы обнаруживается *!oi ~ *!ōĩ, без очевидных этимологических параллелей (ср., может быть, хам !ui-ta, ||кхау !ui 'белый?'); в языках таа — основа *!q^han, несовместимая ни с *ʃoro, ни с *!oi. Все три формы можно, таким образом, почти на равных основаниях привлекать к внешнему сравнению.

29) «рот»: (а) хам ttú, нгк tu, нху ʃu; хег tu [B.], tu ~ tui [LH], аун tu ~ t^hu; (б) хас n=ʃa; (в) аун ʃu; хнг ʃûe, мас !we; нуге ʃûê. || В других языках !кви ср. также ||ку||е tu, ||кхау tu- в tuʔs !^hui 'борода' (букв. 'волосы рта'), ɠan!e tu ~ t^hu (к группе форм (а)); сероа gi (возможно, к группе форм (в)).

В «узко-!кви» подгруппе этимон 'рот' однозначно восстанавливается в виде *tu; в подгруппе таа представлен другой корень, имеющий вид *ʃu- (ср. в хнг мн. ч. ʃû-m-sâ). Любопытное «промежуточное» положение занимает здесь аун, где отмечено существование обоих вариантов. Комментируя этот факт, Д. Блик предполагает, что форма tu в этом языке может рассматриваться как заимствование из ʃкхмани (= нуки) [Bleek 1937: 195, 207]. Косвенно это подтверждается тем, что в ранних данных по аун,

собранных Д. Блик в 1911 г. и опубликованных в [Bleek 1929], форма *tu* вообще отсутствует; она появляется лишь в материалах экспедиции 1936 г., в ходе которой Д. Блик действительно работала с информантами, находившимися под сильным ареальным влиянием $\text{ᠰ}k\text{х}omani$.

Если $\text{a}u\text{ᠨ} tu$ — заимствование, шансы $\text{a}u\text{ᠨ} \text{ᠰ}u$: на отражение исконного пра-!кви этимона 'рот' значительно возрастают, особенно ввиду внешних данных по языкам таа. К сожалению, поддержки со стороны второго известного нам «носсобского» языка, $\text{ᠰ}as$, не обнаруживается, т. к. в последнем значении 'рот' выражается корнем $\text{ᠰ}a$, категорически не сопоставимым с $\text{ᠰ}u$: из-за различий в вокализме. С другой стороны, следует отметить, что для корня *tu* в $\text{ᠰ}am$ надежно зафиксированы и абстрактные значения ('дыра', 'отверстие' и т. п.); это позволяет отождествлять его с суффиксальным элементом *-tu* в таких словах, как $\text{ᠰ}^n a\text{-}tu$ 'подмышка', $\text{ᠰ}^n u\text{-}tu$ 'нос' и др. и может служить дополнительным косвенным аргументом в пользу инновативности этой основы в анатомическом значении 'рот'.

Очень любопытна также форма *gi* в языке сероа, зафиксированная в записях, опубликованных в 1846 г. [Arbousset & Damas 1846: 250]. В этом источнике буква *g*- обычно передает щелчковую артикуляцию (ср. *gni* 'антилопа гну' = $\text{ᠰ}am \text{ᠰ}i$, $\text{ᠰ}ka\text{ᠰ}u \text{ᠰ}i$), так что можно предполагать, что в данном случае *gi* = юж.-кой. $\text{ᠰ}^n \text{ᠰ}u$. Поскольку других существенных лексических аргументов в пользу принадлежности наречия «сероа», записанного Арбуссе, к «носсобским» языкам вроде бы не обнаружено, это еще один аргумент в пользу того, что «узко-!кви» $\text{ᠰ}^n tu$ 'рот' — это инновация.

30) «имя»: $\text{ᠰ}am \text{ᠰ}e \sim \text{ᠰ}e$, $\text{ᠰ}ng \text{ᠰ}e$, $\text{ᠰ}ny \text{ka}=\text{ᠰ}i \sim \text{ka}=\text{ᠰ}e$, $\text{ᠰ}eg \text{ᠰ}e$: [LH], $\text{a}u\text{ᠨ} \text{ᠰ}e \sim \text{ᠰ}ep$, $\text{ᠰ}as a=\text{ᠰ}aṅa$, $\text{ᠰ}ng \text{ᠰ}i$, мн. ч. $\text{ᠰ}à$, $\text{ᠰ}as \text{ᠰ}k\text{ᠰ}i$, $\text{ᠰ}ue \text{ᠰ}i$. || В других языках !кви ср. $\text{ᠰ}ka\text{ᠰ}u \text{ᠰ}e$. Праюж.-кой. $\text{ᠰ}^n \text{ᠰ}e$.

Почти все языки согласованно имеют в данном этимоне ден-тальный кликс с нулевым исходом и назализованную артикуляцию гласного, что может намекать и на более ранний вариант $\text{ᠰ}^n a\text{-}$; ср. в этой связи транскрипцию в «носсобских» языках, где *-n* как согласный отмечен в $\text{a}u\text{ᠨ}$. (В $\text{ᠰ}as$ носовой почему-то имеет веллярную окраску, однако вполне вероятно, что форму $a=\text{ᠰ}aṅa$ нужно анализировать как *a* 'твой' /притяж. префикс/ + $\text{ᠰ}aṅ$ 'имя' + *a* 'это' /указательно-связочная морфема/; в этом случае носовой

согласный на самом деле оказывается в позиции ауслаута, где *-n* в хас не встречается, ср. хас $f^h\acute{a}ŋ$ 'собака' = нху $f^h\acute{u}n$ id. и др.).

Вокализм обнаруживает, как и в ряде других случаев, колебания между огласовками *e*, *a* и (очень редко) *i*, что можно интерпретировать и как развитие из особой фонемы (например, $*\epsilon$), и как следы различных морфонологических процессов ($*|\bar{a} \sim *|\bar{a}-u$ в таа, но $*|\bar{a}-e \rightarrow *|\bar{e}$ в !кви). Второй вариант выглядит более экономным, однако на данном этапе наших знаний о юж.-кой. синхронной и исторической морфонологии при внешнем сравнении следует учитывать различные варианты реконструкции вокализма (пока что праформу можно условно записать как $*|\bar{E}-$).

31) «новый»: (а) хам $\|a:\eta \sim \|^{h}a:\eta \sim \|a:\eta$; (б) нгк $!xe: \sim !xe:k^{h}a \sim !xe:t^{h}a$; (в) хег $\|i$ [LH]; (г) хнг $\|q\mu V$, мас $\|xwe$.

Данный этимон очевидно неустойчив и к тому же довольно плохо зафиксирован в вымерших языках (так, полностью отсутствуют данные для «носсобской» подветви). Тем не менее, более или менее надежно вырисовывается праформа $*\|q\mu-$ для ветви таа, обнаруживающая надежную параллель и в !кви: хам $\|we \sim \|w\acute{e}: \sim \|w\acute{e}:$ 'новый' (судя по конкретным примерам из текстов — только применительно к 'месяцу', что не дает возможности рассматривать это слово как «базисное»). Это автоматически превращает этимон (г) в наиболее перспективную «кандидатуру» на значение 'новый' и на праюж.-кой. уровне.

Для хам $\|a:\eta$ отмечена типологически естественная многозначность 'новый / свежий / сырой'; это дает возможность сравнить соотв. этимон с хнг $\|q\acute{a}n$ 'мягкий', 'нежный' (о вкусе) и предположить направление развития 'свежий' → 'новый', с вытеснением старой глагольно-адъективной основы 'новый' на периферию. Формы в нгк и в хег пока что остаются без убедительной этимологии.

32) «ночь»: (а) хам, нгк $\|a \sim \|a:$, нху $\|a:$, хег $\|a$ [B.], $\|a$: [LH], аун $\|a\acute{u} \sim \|\delta$, хас $\|a\|a$; (б) хнг $\|\acute{u}e$, мас $\|\acute{o}e$, нуе $\|\acute{d}e \sim \|\acute{w}e$. || В других языках !кви ср. $\|ku\|e \|\acute{a}$, $\|k\acute{x}a\acute{u} \|a:$, $!an\acute{e} \|a$ 'вечер'.

Реконструкции $*\|a$ на уровне пра-!кви и $*|\bar{u}-$ на уровне пра-таа более или менее однозначны (в «носсобских» языках зафиксированы производные морфологические варианты — либо с именным суффиксом $*-u$, как в аун, либо с редупликацией, как в хас). Взаимной этимологизации они, тем не менее, не обнаруживают,

и к внешнему сравнению могут привлекаться с равными основаниями.

33) «нос»: хам $\tilde{h}i\tilde{u}$ ~ $\tilde{h}i\tilde{u}ni$, нгк $\tilde{h}utu$, нху $\tilde{h}u\tilde{ci}$, хег $\tilde{h}u$ [B., Zr., LH], аун $\tilde{h}u$: ~ $\tilde{h}o$, хас $\tilde{h}u$, хнг $\tilde{h}i^h\tilde{u}a$, мас $\tilde{h}u$ ~ $\tilde{h}u\text{-}\tilde{c}a$, нхе $\tilde{h}u\text{-}\tilde{c}a$. || Праюж.-кой. * $\tilde{h}u$ -. В других языках !кви ср. также $\|\text{ку}\|e$ $\tilde{h}utu$, $\|\text{кхау}\|u$, $!an!e$ $\tilde{h}u\text{-}\tilde{h}u\tilde{w}i\tilde{u}n\text{-}ti\tilde{h}$ (последняя — скорее всего, редуцированная форма мн. ч.).

Устойчивый этимон; реконструкция корневой моры однозначна и тривиальна. В мас и нхе представлен тот же суффикс (частей тела?), что и в нхе $\tilde{h}u\text{-}\tilde{c}a$ 'ухо' и др.; в языках !кви эту позицию заполняет элемент $-tu$ = пра-!кви * tu 'рот', 'отверстие' (см. выше 'рот', 'ухо'), иногда ассимилируемый под влиянием носовой артикуляции кликса.

34) «не»: (а) хам $k\tilde{h}ai$ ~ $k\tilde{h}ai\text{-}ki$; (б) нгк $\|\tilde{h}u$ ~ $\|\tilde{c}$ ~ $\|\tilde{e}$ ~ $\|\tilde{h}ai$, нху $\|\tilde{h}u$, хнг $\|\tilde{q}^h\tilde{u}a$, мас $\|\tilde{k}a$ ~ $\|\tilde{a}$: ~ $\|\tilde{a}i$ ~ $\|\tilde{t}wa$, нхе $\|\tilde{h}u$; (в) нгк $\tilde{h}a$ ~ $\tilde{h}e$ ~ $\tilde{h}i$ ~ $\tilde{h}o$; (г) хег $^?a$ [Zr., LH]; (д) аун $k^v\tilde{a}$ ~ $t^v\tilde{a}$ ~ $t\tilde{a}$, хас $t^v\tilde{u}$ ~ $t^v\tilde{a}$.

Основная (немаркированная) общеюж.-кой. отрицательная частица, исходя из дистрибуции по ветвям и языкам, однозначно восстанавливается в виде * $\|\tilde{h}u$ -; колебания в вокализме там, где они наблюдаются, скорее всего, обусловлены требованиями глагольного согласования.

Не вполне ясно, можно ли отнести сюда же бескликсовую основу $k\tilde{h}ai$ в хам. Сценарий, объединяющий формы (а) и (б), должен предполагать нерегулярное отпадение латерального кликса, развития исхода кликса в веларную аффрикату и дифтонгизацию старого монофтонга, т. е. чрезмерное количество индивидуальных изменений, которые можно пытаться объяснить в лучшем случае «особой частотностью» служебного слова. Повидимому, на данном этапе стоит все же отделять $k\tilde{h}ai$ от форм с латеральным кликсом, хотя никаких особых параллелей для этого слова в других юж.-кой. языках не обнаружено.

Скудность данных также не позволяет понять, чем отрицательная основа $\tilde{h}V$ в нгк отличается от основы $\|\tilde{h}V$ в этом же языке. В кратком грамматическом описании [Bleek 2000: 21] упоминается только первое из них (ср. такие контексты, как u $\tilde{h}i$ $kiwa$ 'ты не спишь' и др.), но в [Bleek 1956] приводятся надежные примеры также и на второе, ср. $!q$: $\|\tilde{h}u$ $\tilde{h}o$: 'дождь не идет'. Никаких внешних

параллелей у основы \tilde{V} нет; отсутствует она и в доступных нам данных по нху.

Аналогичная проблема связана и с материалами по хег: в [Bleek 1929] отмечено существование отрицания $\|a$, явно связанного с общеож.-кой. негативной морфемой, но в [Bleek 1956] такого вхождения для хег нет, а в [Lanham & Hallows 1956b: 115] эксплицитно отмечено, что главным отрицанием в этом языке служит морфема $\text{?}a$, ср. *ha-?a-siŋ-we* 'он не курит' и др. примеры. Этимологическая связь между $\text{?}a$ и $*\|V-$ невозможна (даже при нерегулярном отпадении кликса ожидалось бы сохранение смычного исхода, т. е. скорее $*ka$, чем $\text{?}a$).

Несколько яснее ситуация с отрицанием в аун и в хас: формы $k^y a \sim t^y a \sim t^y u$ отражают глагольную отрицательную основу $*tV-$, зафиксированную и в других языках !кви, напр. хам *tã* 'не делать'; в аун как самостоятельный глагол основа *ta* отмечена в значении 'не уметь', 'быть неспособным'. Есть основания полагать, что в «ниженоссобских» языках произошло серьезное расширение ее функциональности, хотя различия в огласовке между предикативными и адвербиальными вариантами пока не имеют убедительного объяснения.

35) «один»: хам *!wa:i* ~ *!w?a:i*, нгк $\|we:$ ~ $\|rwe:$, нху $\|?oe$, хег $\|a$: [B.], *!oa* [Zr.], *!wa*: [LH], аун *ʔú* ~ *ʔú-u*, хас *ʔj-ka*, хнг *ʔú-ã*, мас *!kwe*, нхе *!oe*. || В других языках !кви ср. также $\|ku\|e$ $\|k\text{?}oa$, $\|k\text{?}au$ *?oe*, *seroa* $\|oi$.

Все перечисленные формы представляются фонетически совместимыми, но конкретную реконструкцию предложить трудно из-за нетривиальности соответствий. Основа кликса здесь явно та же самая, что и в этимоне 'нога': соответствие «хам ! : нгк || : хнг ʔ» встречается нечасто, но стабильно, и выше уже предлагалось интерпретировать его как «вариант» $*ʔ_1$ (то ли позиционный, то ли фонологичный) обычного палатального кликса $*ʔ$, дающий в диалектном пучке «нуки-ʔхомани» латеральный рефлекс вместо палатального. В хег рефлекс !- уже был выше предложен как регулярное отражение $*ʔ_1$ (см. 'кость'); то, что в записи Д. Блик кликс затранскрибирован как латеральный, можно интерпретировать как ошибку, но не исключена и возможность различного отражения $*ʔ_1$ по диалектам этого языка.

Исход кликса удивительным образом колеблется между нулевым и глоттализированным, иногда в пределах одного и того же языка или диалекта. Чаще всего такие колебания в старых транскрипциях указывают на структуры вида $*\zeta V^?V$, где ζ — основа кликса (с нулевым исходом); при редуцированной артикуляции первой гласной гортанная смычка может начать восприниматься как исход кликса (а в каких-то ситуациях, возможно, становится таковой по-настоящему). С другой стороны, конкретные транскрипции вида $*\|o^?ai$ и т. п. отсутствуют (в том числе и в современных записях по хнѳ и нху), так что не исключена и реконструкция какого-то сложного исхода (например, $*\dagger_{1}q^?$, с утратой увулярной артикуляции в нху и хнѳ).

В плане вокализма большинство диалектов согласованно указывает на старое дифтонгическое сочетание $*-oe$ (возможно, $*-o^?e$, с трифтонгизацией $*-oe \rightarrow -oai$ в хам), но в «носсобской» ветви, а также в хнѳ вместо этого наблюдается вокализм $*-ũ$ (хас $\dagger_{1j}-ka \leftarrow * \dagger_{1}ũ$ + глагольная связка ka), что можно интерпретировать как стяжение из $*-u-$ + $*-ã$. Общая предварительная реконструкция, учитывающая все разнообразие рефлексов, может иметь вид $*\dagger_{1}u^?-$ (корневая мора) с вариантами основы $*\dagger_{1}u^?-e$ ($\rightarrow * \dagger_{1}o^?e$ с ассимиляцией) и $*\dagger_{1}u^?-ã$.

36) «дождь»: (а) хам $!wa: \sim !^hwa$, нѳк $!^ha \sim !a$, нху $\dagger_{1}qau$, (?) хег $\phi^h eij$ [LH], аун $!^h \grave{a}:a$; (б) хнѳ $!kx\delta e$; (в) мас $!we-ga$ $!a$, нѳе $!xwe$.

В !кви, скорее всего, первичен вариант с палатальным кликсом, отмеченный в нху (\leftarrow пра-!кви $*\dagger_{1}q^{(h)}au$). Неясно, относится ли сюда же форма в хег. С одной стороны, окказиональная палатализация $*\dagger$ в этом языке вроде бы встречается, хотя условия ее неочевидны: ср. такой убедительный пример, как нху $\dagger_{1}q^h oe$ 'ветер' = хег $\acute{s}uwi$ [B.], $\acute{s}we$: [LH] id. С другой стороны, появление в этой позиции рефлекса ϕ^h - уникально (ожидалось бы скорее $*\acute{s}eij$). В [Bleek 1929: 68], помимо этого, зафиксирован явно другой корень — хег $ga:a$ 'дождь', неясного происхождения.

Ситуация осложняется еще и тем, что некоторые из приводимых форм — напр., хам $!^hwa$ — омонимичны (или, по крайней мере, омографичны) соотв. рефлексам праюж.-кой. корня 'вода' (см. ниже), и, если бы не однозначная их дифференциация в надежно записанном нху ($\dagger_{1}qau$ 'дождь' : $!^ha$: 'вода'), они, скорее всего, оказались бы записанными в один и тот же исторический

корень. Но даже зная ситуацию в нху, нельзя до конца исключить вероятность контаминации в хам и других языках !кви.

В хнг представлен другой корень — **!kxo* — который, вследствие явного несовпадения артикуляции основ кликса, нельзя отождествлять с !кви **!q(h)ai*; по крайней мере, надежных случаев соответствия «нху † : хнг !» на данный момент не обнаружено. В других диалектах таа значение 'дождь' выражается тем же словом или, по крайней мере, корнем, что и 'облако': ср. мас *!wé* 'облако', нуе *!xwe*: *!arri* 'облако' = 'небо дождя'. Неясно, следует ли объединять эти формы с «плавающим» исходом кликса (*!we* ~ *!we* ~ *!xwe*) с хнг *!kxɔ*: велярная аффриката в хнг должна соответствовать такой же артикуляции в мас и нуе. На всякий случай при лексикостатистических подсчетах их пока что рекомендуется различать.

37) «дым»: (а) хам *!ʃ*; нху *!q:-ke*, хас *!ai*; (б) нгк *!wi*; (в) хег *k^ha?a-zi* [LH]; (г) хнг *skâye*; (д) мас *!a:li*.

Этимон очевидно неустойчив. Для пра-!кви удается восстановить вариант **!q-* или, возможно, **!qi* со стяжением в хам и нху, который на более глубоком уровне можно сравнивать с хнг глаголом *!i[?]-i* 'выкуривать пчел'. Впрочем, и эта реконструкция напрямую зависит от достоверности формы *!ai*, зафиксированной в хас, для которой в словарики Р. Стори приводятся значения 'дым' (именное) и 'курить' (глагольное), но в текстах из них подтверждается только глагольное. Из остальных форм (б, в, г, д) ни одна без выхода на внешний уровень не этимологизируется.

38) «звезда»: (а) хам *!wq-tən*, хег *!ou-ni* (мн. ч.) [Zr.]; (б) нгк *!kxwe-sa* ~ *!wqi-sa*, нху *!kxɔ-si*, хнг *!bna*, мас *!wana-te* ~ *!wana-te* (мн. ч.), нуе *!pana-te* (мн. ч.); (в) (?) хег *kale*: (мн. ч.) [B.]; (г) аун *!^ha*. || В других языках !кви ср. также кку *!an-te* (мн. ч.), ккау *!roa:n-si*, сероа *koan-koan* (мн. ч.).

В языках !кви наиболее широкую дистрибуцию имеет этимон **!q-a* ~ **!q-i*; формы в хам и в хег фонетически совместимы, хотя определенное сомнение вызывают различия в суффиксации. С другой стороны, нху *!kxɔ-*, хотя и никак не совместимо с **!q-*, вполне сопоставимо с таа **!ona*, при условии признания элемента *-na* в таа старым «окаменевшим» суффиксом, что вполне естественно в рамках концепции разделения любой основы на корневую и суффиксальную моры. Особую значимость здесь имеют формы в кку *!e* и в ккау, где также виден носовой суффикс.

Основная проблема с группой форм (б) — корреляции исходов кликсов. Велярная аффриката в нху в принципе не должна соответствовать нулевому исходу в таа; однако есть веские основания усомниться в том, что на пра-!кви уровне здесь действительно была велярная аффриката, учитывая весь разброс вариантов в старых записях нгк ||ку||е и ||кхау. Даже в нху Н. Кроухолл (данные которого в целом менее надежны, чем записи Б. Сэндс, по которым цитируется основной вариант) фиксирует не велярную аффрикату, а увулярный исход, причем еще и со странной морфологической дифференциацией (если это не опечатка в записи): ||qʔe-si для ед. ч., но ||qʔe-ke для мн. ч.¹ Скорее всего, исход кликса был все же менее «тривиален», чем, например, в слове 'рука' (см. выше), и вплоть до более тщательного разрешения этого вопроса этимоны в !кви и таа мы будем считать совместимыми.

В итоге между группами форм (а) и (б) наблюдается дистрибуционное пересечение, разрешаемое в пользу (б): в хам и хег наблюдаются, скорее всего, независимые инновации, хотя и неясного происхождения. Совершенно особым корнем является аун ^{!h}a:, допускающее сравнение с хнг глагольной конструкцией ^{!h}ai tí ~ ^{!h}à: tí 'заходить', 'падать (о звезде)' (изначально = 'звездопад'?).

39) «камень»: (а) хам !au ~ !ou, нгк !au, нху !ao, хег žeu [Zr.], ʔeə [LH] (?), хас !dè; (б) аун ||kxɔ; (в) хнг !i-le, мас !i-le; (г) нве !um; (д) нве ʔoiye. // В других языках !кви ср. также ||кхау !ao, ||ку||е dʔ 'скала' (см. выше комментарии к слову 'ухо' отн. утраты щелчковой артикуляции в ||ку||е).

Пра-!кви 'камень' устанавливается однозначно в том случае, если хам-ʔхомани *!ao (с полисемией 'камень / гора') отражает тот же этимон, что и хас !dè: проблематичны вокалические соответствия, но, поскольку хороших примеров на отражение в хас пра-!кви дифтонга *ao нет, на данном этапе формы можно считать фонетически совместимыми (полное совпадение консонантизма + фонетически сходный вокализм).

¹ Нельзя не отметить также вариацию, отмеченную в записи нху слова 'еж': ||qʔe-si ~ ||qʔe-si (с глоттализиацией) ~ ||qʔae-si (без глоттализиации). Метафорическое определение 'ежа' как 'звезды' согласуется с традиционным бушменским представлением о еже как одушевленной форме падающей звезды [Traill 1994a: 59].

Сложной остается ситуация в хег. Здесь также неизвестна регулярная рефлексия пра-!кви **!ao*, и допустимо, что этот дифтонг подвергнулся упереднению, как в ѳоан. В этом случае **!ao* → **!eo* → **keo* (с регулярной утратой щелчковой артикуляции) → **ǰeo* (потенциально регулярная палатализация, ср. ниже этимон 'два'). Сомнения вызывает глоттализованность начального согласного в транскрипции [LH]: *ǰeo* должно было бы в таком случае отражать праформу **!ʔao*, а не **!ao*. Однако ситуация осложняется тем, что в транскрипции Цирфогеля здесь вообще зафиксирован звонкий спирант *ǰ* (!). Не исключены контаминации: в [LH] отмечено существование параллельного синонима *ǰwe* 'камень', скорее всего, заимствованного из банту, ср. сото *li:ǰwe* id. При этом *ǰeu* [Zr.] и *ǰwe* [LH], очевидным образом, тоже не обнаруживают идеальной корреляции друг с другом. Наконец, в [Bleek 1956] находим два похожих, но также отличающихся по вокализму варианта: *ǰe* и *ǰi:* 'камень'. Все это наводит на мысль о том, что в разных диалектах хег исконная основа 'камень', случайно приобретя фонетическое сходство с синонимичным заимствованием, разными способами модифицировалась под его влиянием; к сожалению, разобраться в этих процессах детально вряд ли удастся из-за невозможности отследить ситуацию в полевых условиях. Пока что мы все же будем считать хег *ǰeo* потенциальным рефлексом **!ao*, списывая глоттализацию на возможную ошибку в записи, а варианты с *ǰ*-объяснять результатами контаминаций.

Аун //кх: — сомнительная форма, отмеченная только в раннем источнике [Bleek 1929]; в дальнейшем из записей исчезает.

В языках таа наиболее широкую дистрибуцию имеет основа **!i-* (также с полисемией 'камень / гора'; *-le* в хнг и мас — показатель класса, ср. в хнг мн. ч. *ǰi-n*, а также мас *ǰi:n* 'гора', на самом деле = мн. ч. 'камни'). Нуе *!ɨit* 'камень' — скорее всего, ошибочный или идиолектный вариант нуе *!ɨit* 'гора' = хнг *!iʰit* 'холм (невысокий)'. Этимология синонимичной формы *ǰoiye* неизвестна.

Надежной взаимной этимологизации !кви **!ao* и таа **!i-* не имеют; однако основа в !кви потенциально совместима с хнг *!qǰi* 'каliche (т. е. налет карбонат кальция)'. Если этимология верна, то сложное фонетическое устройство этого слова в хнг (увулярный исход + фарингализованный гласный) может каким-то образом быть связано с нетривиальным поведением этимона в хег; вопрос

требует комплексного исследования. В любом случае, на данном этапе к внешнему сравнению можно привлекать оба этимона (тем более что даже при условии корректности этимологии неясным остается направление семантического развития).

40) «солнце»: (а) хам $\| \acute{o}i\eta \sim \| \acute{r}\acute{o}i: \sim \| \acute{r}\acute{o}i:\eta \sim \| \acute{r}\acute{u}i\eta \sim \| \kappa\acute{o}i\eta$, нгк $\| \acute{r}\acute{o}\acute{e} \sim \| \acute{r}\acute{o}i\eta \sim \| \acute{r}\acute{o}\acute{e} \sim \| \acute{o}i\eta$, нху $\| \acute{r}\acute{u}i$, хег $\| \acute{o}i \sim \| \acute{r}i:n$ [B.], $\| \acute{u}mi$ [LH], хнг $\| \acute{r}\acute{a}n$, мас $\| \acute{r}li$, нхе $\| \acute{r}li \sim \| \acute{r}\acute{e}$; (б) аун $| \acute{e} \sim | \epsilon n$, хас $| \acute{r}i$. || В других языках !кви ср. также $\| \acute{r}\acute{o}\acute{e}$.

подавляющее большинство вариантов основы (а) в языках !кви сводится к вариантам $*\| \acute{u}i$ или $*\| \acute{r}\acute{u}i$ (фонетическая реализация дифтонга в конкретных диалектах могла быть ближе к $\acute{o}i$, но пока что нет серьезных оснований считать, что в пра-!кви oi и ui противопоставлялись фонологически). Очень интересен вариант $\| \acute{u}mi$ в хег: если он представляет ту же самую основу (а другой этимологии у него нет), его можно трактовать исключительно как архаизм, сохраняющий исходную двусложную структуру, в то время как в диалектном пучке $| \acute{x}am$ - $| \acute{x}omani$ (равно как и в части диалектов хег, описанных Д. Блик) срединный носовой ослаб и выпал, трансформировавшись в носовую окраску гласного. Тогда реконструкция на пра-!кви уровне должна иметь вид $*\| (\acute{r}) \acute{u}mi$ или $*\| (\acute{r}) \acute{u}ni$ (если $-m$ в хег — результат ассимиляции после $*-u-$).

Невозможно также твердо определить исход кликса в пра-!кви: «надежная» форма в нху свидетельствует в пользу глоттализиции, но наличие колебаний между глоттализированным и нулевым исходом и в хам, и в нгк, и в хег (а в хам представлен еще и вариант с веллярной аффрикатой) говорит о том, что на прауровне исход был менее тривиальным, возможно, чем-то вроде глоттализированного увулярного (т. е. $*\| q\acute{r}i\acute{u}ni \sim * \| q\acute{r}i\acute{u}ti$). Еще один вариант — глоттализация самого гласного, т. е. реконструкция $*\| \acute{u}mi \sim * \| \acute{u}mi$, с «переползанием» глоттализиции на исход кликса (реальной или фиктивно-ошибочной).

В любом случае пра-!кви этимон можно считать совместимым с пра-таа $*\| \acute{r}an$: вокалические соответствия здесь примерно такие же, как в этимоне 'собака' (см. выше) и ряде других (т. е. возможна, например, реконструкция особой фонемы $*\acute{r}$), а консонантизм полностью совместим при условии, что в пра-!кви исходный носовой согласный корня был $*n$, а не $*m$. В этом случае пра-!кви $*\| q\acute{r}i\acute{u}n-i$ ($\sim * \| \acute{u}m-i$) = пра-таа $*\| \acute{r}an$ ← праюж.-кой. $*\| \acute{r}an$ или $*\| q\acute{r}an$

(впрочем, финальную реконструкцию можно будет предложить только после подробного разбора всего материала).

Отдельная группа форм представлена только в «носсобских» языках: хотя аун /ɛn и хас /i не соответствуют друг другу идеально (особенно в части исхода кликса), это можно списать либо на нетривиальность соответствия, либо, скорее, на неточность в записи. Что касается этимологизации этой явной (с точки зрения дистрибуции) инновации, то здесь хороших гипотез пока нет. Этимологическая связь с 'огнем' маловероятна из-за серьезных фонетических расхождений в обоих языках; не исключено заимствование из хкойкхой *се: 'день' (или непосредственно из тех языков банту, откуда это слово попало в хкойкхой), но в этом случае требуется подтвердить возможность перехода при заимствовании аффрикаты в кликс другими примерами, что пока не сделано.

41) «хвост»: (а) хам !^hwi, нгк !ei, нху !^hai, хег k^hi [Zr.]; (б) аун ɬwi; (в) хас i=^ha:-a, хнг |ãÿ, мн. ч. |ã:, мас |ãÿ ~ |ãÿ, нуе |ãÿ.

На пра-!кви уровне без учета «носсобских» данных надежно восстанавливается форма *!^hai ~ *!^hi (с вторичной лабиализацией в хам, см. выше 'пить' и другие аналогичные примеры, и с регулярным отпадением простого альвеолярного кликса в хег). В ветви таа ей противостоит основа *|ã- (или *|ãÿ); на взаимной основе они не этимологизируются.

Трудная ситуация складывается с «носсобской» ветвью. Аун ɬwi, безусловно, хотелось бы сравнить с группой форм (а), однако этому препятствует как явное несоответствие основ кликса, так и наличие в аун лабиализации, которая в этом языке, в отличие от хам, откровенно вторичной не бывает; если палатальный ɬ еще можно как-то списать за счет ошибочной транскрипции (хотя ɬ вместо ! в аун отмечен максимум в одном-двух случаях), то дифтонгическое сочетание -wi — вряд ли. С другой стороны, сложная форма в хас (где начальный i=^h — скорее всего, посессивный префикс 3-го л.) гораздо лучше согласуется с пра-таа *|ã- (совпадает все, кроме носовой артикуляции гласного), так что более вероятен следующий сценарий: прауж.-кой. */a- ~ */ã- 'хвост' с замещением в пра-!кви и отдельным замещением в аун.

42) «ты»: хам a ~ a:, нгк a, нху a, хег a [B., LH], ?a ~ ?a-ŋ [Zr.], аун a, хас a, хнг ã^h, мас a, нуе a. || Прауж.-кой. *a. В других языках !кви ср. также ||ку||e a ~ an, ||кхау ?a, сероа, г!ан!e a.

Реконструкция корневой морфемы однозначна (возможно, в дальнейшем ее придется изменить на $*a^h$ с придыхательной артикуляцией, если можно будет доказать архаичность варианта в хн); сопровождающие ее в отдельных языках суффиксальные элементы — продуктивные или окаменевшие эмфатические частицы (ср. хам *a-keŋ*, хн *āh-ʔā* и др.).

43) «язык»: хам *ʔéřři* ~ *léřři*: ~ *ʔénni*, нгк *ʔē*, мн. ч. *ʔe:n-yaŋ*, ху *ʔāŋ* ~ *ʔāī*, хег *ʔē* [B.], аун *ʔāri*, хн *ʔnāŋ*, мас *ʔa:n*, нуе *ʔani*. || В других языках !кви ср. также ||*хау* *ʔanan-si*.

Формы в !кви лучше всего объясняются через исходную основу $*ʔan$ ~ $*ʔani$; наличие во втором слоге ауслузного *-i* объясняет переогласовку корневого вокализма в хам, нгк и хег, но остается неясным, насколько тесно это *-i* было связано с корнем (продуктивный именной суффикс или неотделимая часть корня, отпадающая в ряде языков из-за редукции?).

На пра-таа уровне, опираясь на данные по хн, можно восстановить $*ʔnan$ ~ $*ʔnan-i$, с точно таким же варьированием между одно- и двусложными вариантами, как и в !кви. Конкретное соответствие «!кви *ʔ*: таа *ʔn*» другими примерами не подтверждается, но есть основания подозревать, что на самом деле таа $*ʔnan$ — результат ассимилятивного распространения носовой артикуляции на исход кликса, т. к. в материалах Э. Трэйлла по хн наблюдается довольно высокая корреляция между наличием в корне срединного (или конечного) *-n*- и назализованного преглоттализованного кликса, на основании которой можно сформулировать правило: $*ʔVn(V) \rightarrow *ʔnVn(V)$ (хотя из него бывают отдельные исключения).

В любом случае, учитывая также частые случаи «нестандартного» поведения корня 'язык' в языках Африки (см. ниже нерегулярную рефлексацию этого этимона в центр.-койсанской группе и др.), вряд ли уместно отделять друг от друга формы в !кви и таа из-за нестандартности соответствий в нетривиальном исходе кликса. На прауж.-кой. уровне, скорее всего, нужно восстанавливать форму $*ʔan(-i)$.

44) «зуб»: хам $\|^{h}ēi$: ~ $\|ēi$, нгк $\|āi$: ~ $\|ē$: ~ $\|ēi$, ху $\|^{h}āi-si$, хег $\|ēi$ [B.], $\|^{h}i$ [Zr., LH], аун $\|ēi$, хас *kʔi=|ε*, хн $\|q^{h}ā$, мас $\|xū$, нуе *ʔan-te* (мн. ч.). || В других языках !кви ср. также ||*ку* $\|ε$ *kε*, $\|an!ε$ $\|i-$ $\|iŋ$ 'зубы' (с дубликацией).

На пра-!кви уровне надежно вырисовывается праформа $*\|^{h}āī$ (фонетически = $\|^{h}ēī$ с ассимиляцией); от этого прототипа несколько отклоняется только основа $\|ε$ в хас, которая, возможно, отражает «чистую» корневую мору ($kʷi=$ — префикс мн. ч.). Помимо этого, выше было уже отмечено, что в записях В. Блика и Д. Блик !кви формы слов 'зуб' и 'рог' тяжело отличать друг от друга, но в самом крайнем случае речь может идти лишь об отдельных случаях вторичной омонимии или контаминации, поскольку в подавляющем большинстве языков эти два корня так или иначе различаются.

Судя по данным таа, назализация корневого гласного может в конечном итоге восходить к полноценному носовому согласному; ср. особенно такие формы мн. ч., как нүе $\|^{h}in-te$ и мас $\|xani \sim \|xani$. В этом случае хнġ $\|q^{h}ā \leftarrow * \|q^{h}an-ā$, где $-ā$ — продуктивный показатель 2-го грамматического класса, а пра-!кви $*\|^{h}āī$, в свою очередь, может быть стяжением из старой формы мн. ч. $*\|^{h}an-i$. Возможны и другие сценарии, но все они должны учитывать архаичность носового артикуляционного компонента в ауслауте корневой морфемы; на уровень внешнего сравнения перспективнее всего выносить реконструкцию $*\|^{h}an$ (увулярный исход $-q^{h}$ в хнġ, не подтверждаемый новыми данными по нху, сомнителен).

45) «дерево»: (а) хам $\theta^{h}o$, нġк $\theta o \sim \theta o$: $\sim \theta^{h}o$, нху θo ; хег $\theta^{h}o$: [B.], $\theta o \sim \theta^{h}o \sim \theta^{h}oŋ$ [Zr.], $\theta\delta\text{:}-z\grave{i}$ [LH], аун $\theta wa:a \sim \theta wa\text{:}sa$, хнġ $^{\theta}On\grave{a}ye$, мас $\theta oe\text{:} \sim \theta oi$, нүе $\theta a\text{:}$; (б) хас $\#^{h}ai$. || В других языках !кви ср. также $\|k\text{хау}$ $\theta\bar{o}$; $\#an'e$ $\theta o\text{-}s$ 'дерево, древесина (материал)'.

Учитывая дистрибуционную редкость этимонов с губным кликсом, очевидно, что практически все юж.-кой. языки отражают один и тот же корень, условно записываемый нами на пра-уровне как $*^{\theta}on\text{-}$, хотя конкретный исход кликса неоднозначен: назализованная артикуляция в хнġ согласуется с формой, которую Лэнхэм и Хэллоуз записывают для хег, но не с прочими вариантами !кви, в которых назализация не отмечена совсем. Возможно, что такая ситуация представляет собой нормальное развитие редкого «преглотализованно-назализованного» кликса, зафиксированного в хнġ, но не в других юж.-кой. языках.

В большинстве языков для данного слова отмечена полисемия 'дерево (растущее) / дерево (как материал)', что в типологическом плане может привести к утрате первого из этих двух значений.

Такая утрата действительно имеет место в хас, где Р. Стори фиксирует для слова *θoei* только значение 'дерево (как материал)'; значение 'дерево (растущее)' выражается инновацией *ʰai*. Аналогичную ситуацию отмечает для современного нху и Б. Сэндс (личн. сообщ.): старая форма *θo*: в значении 'дерево (растущее)' замещается на *ʰʰi*: 'босция белокорая (*Boscia albitrunca*)' = хас *ʰai* (впрочем, в [Maingard 1937: 256] для «*фхомани*» еще приводится форма *θo* 'дерево', так что не вполне ясно, насколько данное замещение, отмеченное в речи группы информантов, давно уже не использующих нху в бытовом общении, можно считать «естественным» результатом стандартного диахронического процесса).

46) «два»: (а) хам *!ú*: ~ *!ʰi*:, нкк *!u* ~ *!ʰi*:, нху *!ʰi*:, хег *!u* ~ *!ʰi* [B.], *kʰuʷi*: [Zr.], *kʰi*: ~ *ʕʰi*: [LH], аун *!ʰi* (?); (б) аун *!at*; (в) хас *s=||a:-ma*; (г) хнг *ʰít*, мас *ʰít* ~ *ʰit*, нуге *ʰit*. || В других языках !кви ср. также *||kʷe* *!ʰi*, *||kʰau* *!ʰi*.

Для пра-гаа можно уверенно предполагать реконструкцию **ʰit*, поскольку палатальный кликс в мас и нуге в транскрипциях Д. Блик регулярно заменяется на альвеолярный (реже — на латеральный).

Что касается !кви, то тут наибольшее удивление вызывает варьирование между *!-* и *!ʰ-* и в хам, и в нкк, и в хег (по записям В. Блика, Л. Ллойд и Д. Блик) — чрезвычайно редкое совпадение даже для записей, отличающихся своей неточностью. Возможно, что, как и в ряде других случаев, здесь следует предполагать исходный вариант **!ʰiʷi*, стягивающийся в **!ʰi* либо «реально», либо «фиктивно» (т. е. в результате ошибки транскрибирующего исследователя). Необычен также вариант с латеральным кликсом в хег по версии Д. Блик, при том, что транскрипция в [LH] скорее отражает регулярный рефлекс старой формы **!ʰi*: → *kʰi*: (с ожидаемым выпадением основы кликса) и далее → *ʕʰi*: с ожидаемой палатализацией велярного (см. выше 'камень'); корреляция уникальна и, скорее всего, отражает какое-то редкое идиолектное развитие, возможно, еще и неправильно затранскрибированное (например, **!ʰi* → *qʰi* или даже *ʎʰi* вместо *kʰi*, с неверным «распознаванием» увулярной или латеральной аффрикаты как латерального кликса).

В аун Д. Блик чаще всего дает для значения 'два' эквивалент *!at* — явное заимствование из центр.-койсанского источника (см.

ниже); однако в [Bleek 1956: 492] дается и форма *!ʉi*, сопровождаемая конкретным примером (*i a si-n !ʉi* 'нас двое'). В грамматическом очерке [Bleek 1937] *!ʉi*, тем не менее, вообще не упоминается в перечне числительных, так что есть все основания подозревать, что в приведенном примере архаизм *!ʉi* просто сохраняется в составе идиоматической «двойственной» формы личного местоимения. Тем не менее, и этот факт очень важен, т. к. он подтверждает возможность реконструкции **!(u)ʉi* на самом «верхнем» уровне пра-!кви, включая «носсобские» языки, и, соответственно, инновативный статус странной хас формы *s=||a:-ma:*. Последняя состоит из трех морфем, в первой из которых, очевидно, выпал корневой гласный, но предложить приемлемую внутреннюю этимологию для всей конструкции пока что не удается.

Пра-!кви **!(u)ʉi* и пра-таа **ʃit*, даже при условии морфологической делимости **ʃit* (недоказуемой на уровне таа), несовместимы друг с другом и не этимологизируются на взаимной основе (в фонетическом плане **!(u)ʉi* сравнимо с хнц глагольной основой *!ūʔa kV* 'отделять', но возможность такого рода семантической деривации требует серьезного типологического подтверждения). На уровень внешнего сравнения допустимо выносить обе формы.

47) «вода»: хам *!wá ~ !wa: ~ !wā ~ !ʰoa: ~ !ʰwà;* нцк *!ʰa: ~ !a: ~ !ʰa ~ !ʔa: ~ ||ʰa;* нху *!qʰa;* хег *kʰa: ~ ||ʰa [B.], kʰa: [Zr.], qʰa: [LH];* аун *kʰá: ~ kʰáá ~ kʰái ~ ||ʰa;* хас *kà;* хнц *!qʰà;* мас *!ʰá ~ !ʰa: ~ !xa;* нуге *!ʰa.* || В других языках !кви ср. также *||ky||e kʰo, !ʔan!e !oa ~ !owa.*

Большой разброс вариантов, зафиксированных в «старых» записях, может быть связан с увулярным придыхательным исходом кликса *-qʰ*, непосредственно зафиксированном в профессионально записанных формах нху (*!qʰa:*), хнц (*!qʰà:*) и даже хег (*qʰa:* [LH], с регулярным отпадением альвеолярной основы кликса); неумение определить этот тип артикуляции приводило к тому, что формы иногда записывались, в том числе, и как якобы содержащие латеральный кликс — последнее крайне маловероятно. Относительно возможных контаминаций с этимологически отдельной основой 'дождь' см. выше. На прауж.-кой. уровне основа восстанавливается в виде **!qʰa* (лабиализация в хам, как и в других случаях, по-видимому, вторична).

В хег, помимо данного слова, все исследователи, кроме Цирфогеля, отмечают также наличие синонима *ša* [B.], *ša:* [LH]. Хег -

может отражать не только старый сибилант (ср. *šo*: [B.] 'сидеть' = нху *so*: id. и др.), но и старый палатальный кликс с исходом на $-q^h$ - (возможно, также и с другими исходами), ср. *šwe*: [LH] 'ветер' = нху *ʃq^hoe* id., *ša-gu* [Zr.] 'грудь' = нху *ʃq^ha*: id.; поэтому хег *ša* 'вода' может отражать и старую основу $*ʃq^h_a$, и в этом случае можно было бы допустить диалектное варьирование в пра-!кви ($*!q^h_a \sim *ʃq^h_a$), которому, однако, сложно дать разумную интерпретацию. Более вероятно все же их раздельное происхождение, отчасти подтвержденное синтаксическими контекстами: ср. *n* |*o:wa*, *k^ha*: *n* |*eo* 'я хочу пить, а воды нет' [B.], *sa ne k^ha*: [Zr.] 'дай мне воды', но *kan* |*owa ke e ša* [B.] 'он стоит в воде', *ija ša*: *gi=t^hama* [LH] 'я красивый', букв. 'я — [как] вода озера Крисси', т. е. *k^ha*: скорее обозначает 'воду для питья' (как субстанцию), а *ša* — 'воду в водоеме' (текущую или стоячую). Если эта гипотеза верна, форма *ša* вообще оказывается нерелевантной для лексикостатистики.

48а) «мы» (экскл.): хам *si ~ ssi*, нгк *si*, нху *si*, аун *si ~ se ~ ci*, хас *ci*, хнг $\bar{i}=s\hat{i}$ (?), мас *si ~ ši ~ šia ~ ša*, нхе *si ~ si-sa*. || В других языках !кви ср. также ||кхау, ||ку||е, г!ан!е *si*.

48б) «мы» (инкл.): хам *i ~ i*, нгк *i*, нху *i*, хег *i* [B.], *ʔi* [Zr.], *i-ʔe ~ ʔi-ʔe* [LH], аун *i ~ e*, хас *i ~ i-tyo:a*, хнг $\bar{i}^h \sim \bar{i}^h-\bar{i} \sim \bar{i}-s\hat{i}$, мас *i*, нхе *i ~ e*. || В других языках !кви ср. также ||кхау *ʔi*, ||ку||е *i*, сероа *i*.

При анализе форм личного местоимения 1 л. мн. ч. наблюдается частичная дифференциация между «старыми» записями и «новыми». В хег и в хнг зафиксирован единственный вариант этого местоимения — пражж.-кой. $*i$, употребляемый для выражения как эксклюзива, так и инклюзива (в хнг есть еще специальная форма для двойственного числа $\bar{i}\bar{i}$, но это, скорее всего, локальная инновация).

Напротив, в «старых» записях В. Блика, Д. Блик и других исследователей пражж.-кой. $*i$ — это в первую очередь инклюзивное местоимение, в то время как на роль эксклюзивного «мы» претендует морфема $*s\hat{i}$. Не исключено, что в хнг именно эта морфема представлена в эмфатической форме $\bar{i}-s\hat{i}$, восходящей к коллокации обоих местоимений, хотя и не имеющей специфически инклюзивного значения. Подчеркнем, что в описании [Maingard 1958: 103, 106], которое в целом не выдерживает никакого сравнения с последующими работами Э. Трэйлла, для хнг, тем не менее, также эксплицитно постулируется оппозиция между инклюзивом

i и эксклюзивом *si* ~ *ši*, так что современная ситуация в хнг по любому должна определяться как инновативная.

Что касается нху, то в [Maingard 1937: 244] отмечено существование варианта *i* для инклюзива и вариантов *sa*, *si* для эксклюзива; в [Crawhall 2004] также приводятся все три варианта, но без указания на какие-либо смысловые различия; в новейших полевых данных Б. Сэндс и др. подтверждается противопоставление между инкл. *i* и эксл. *si*.

Чисто умозрительно можно допустить, что в праюж.-кой. существовала только форма **i*, а вариант **si* сформировался под влиянием контактов с языками кхойкхой, где эта морфема регулярно маркирует эксклюзив (см. ниже). Однако дистрибуция **si* в юж.-кой. языках — как !кви, так и таа — чересчур широка, чтобы можно было настаивать на отсутствии этого местоимения в праюж.-кой.; к тому же в языках кхойкхой маркер эксклюзивности *si* практически никогда не встречается в отрыве от других местоименных морфем, ни в функции субъектного, ни объектного местоимения, и возможный механизм заимствования получается чересчур запутанным, чтобы можно было рассматривать его в качестве разумной альтернативы исходному присутствию на праюж.-кой. (и, возможно, еще более глубоком) уровне.

49) «что»: (а) хам *ca=de* ~ *ca=de* ~ *ca-de* ~ *=de*, нгк *gi-si* ~ *ki-si* ~ *d^{hi}-si*, нху *çiĩ*, хег *t^{hĩ}*: [LH]; (б) хас *ʃ^ha* ~ *ʃi*, хнг /V ... *è^h*. || В других языках !кви ср. также ||кхау *de*: ~ *dən* ~ *den*.

50) «кто»: (а) хам *!u=de* ~ *!u=de-xa*; (б) нгк *tú-e* (?), нху *çu*, хег *to*: [B.], *towa* [LH]; (в) хас *ci*; (г) хнг /V ... *èh*. || В других языках !кви ср. также ||кхау *tú*.

Как и в сев.-кой. языках, вопросительные местоимения в юж.-кой. группе могут образовываться по модели «общая вопросительная частица» + 'вещь' (= 'что') или + 'человек' (= 'кто'). Однако конкретная ситуация более сложная. На пра-!кви уровне удастся восстановить вопросительную частицу *TI = хам *dε*, нгк *d^{hi}* ~ *gi* ~ *ki*, нху *zi* ← **di* (с ожидаемой палатализацией дентального перед передним гласным), хег *t^{hĩ}*, без явных параллелей в языках таа¹;

¹ В словаре [Bleek 1929] для значения 'что?' в мас и в нхе дается эквивалент *-ta* (дефис, по-видимому, обозначает предположительно энклитический статус), который при сильном желании можно было бы

при этом в образовании обоих вопросительных местоимений она участвует только в хам, ср. *сʷa=dε* 'что' (*сʷa* = 'вещь') и *!u=dε(-xa)* 'кто' (*!u* = 'человек'), в прочих же языках встречается только в составе местоимения 'что' (или, возможно, даже выполняет его функции сама по себе, как в хег).

Другая вопросительная частица, **!V*, надежно зафиксирована в хнг, где она образует разрывное вопросительное местоимение, сочетаясь с указательной основой *ε^h* (причем любопытно отметить полное отсутствие, согласно Э. Трэйллу, лексического различия между одушевленной и неодушевленной формами). В хас эта же частица, по-видимому, сама по себе берет на себя функции местоимения 'что?' (ср. у Р. Стори: *!i a* 'что это?'). Этимологически связать ее с **!TI* невозможно.

Наконец, в ряде языков местоимение 'кто?' может просто совпадать со словом 'человек', ср. нху *ɕi*, \llbracket хау *tu* 'кто?' = хам *tu-kən* 'мужчины', нгк *tu* 'человек', хнг *tû*: 'люди' и т. п. Механизм такого сдвига не вполне ясен; скорее всего, изначально это слово сочеталось с какой-то вопросительной частицей, впоследствии отпавшей. В этом плане интересен вариант, зафиксированный Д. Блик для нгк: *tu e ki* [Bleek 2000: 23], букв. 'человек — есть — вопр. частица' (ср. *ki-si* 'что?'). К сожалению, единственный пример, который она приводит на употребление этого местоимения, неоднозначен: *tu e se kia* 'кто идет?' (англ. 'who comes here?'), где *kia* может означать просто 'здесь'.

Все это означает, что для праюж.-кой., как и для прасев.-кой., реконструировать нужно не столько особые формы вопросительных местоимений, сколько общие вопросительные частицы. При этом более широкую дистрибуцию по ветвям имеет **!V* (таа + возможное сохранение в «носсобской» подветви *!кви*), но исключать из внешнего сравнения пра-*!кви* **!TI*, ввиду невозможности внутренней этимологизации этой частицы, также недопустимо.

сопоставить с пра-*!кви* **!TI*. Однако в более полном и подробном словаре [Bleek 1956] это вхождение отсутствует, и никаких следов соотв. вопросительной морфемы в близкородственном хнг также не обнаружено. Скорее всего, речь на самом деле идет об одной из указательных местоименных морфем, ошибочно принятой за вопросительную.

2.4. Центральнокойсанская группа.

2.4.1. *Общие сведения и источники.* К центральнокойсанской группе (альтернативное название — *кхой*, от центр.-кой. **k^hoe* ‘человек’) относится около 20 отдельных языков, до сих пор распространенных на территории Намибии и Ботсваны, в меньшей степени — Анголы, Замбии и Зимбабве. Уже один этот факт подчеркивает особый статус центр.-кой. среди прочих «койсанских» таксонов — это единственная группа, для которой качественно записанные материалы удается собрать по большому количеству языков, серьезно отличающихся друг от друга в плане фонетики, лексики и грамматики. Тем самым значительно облегчается работа компаративиста, ставящего своей целью системную реконструкцию працентр.-кой. фонологической системы и этимологического корпуса.

Генетическое единство языков, составляющих центр.-кой. группу, долгое время оставалось под сомнением из-за значительных этнических и антропологических различий между их носителями. Д. Блик [Bleek 1929] впервые ввела в оборот термин «центральнобушменские» (Central Bushman) для обозначения языков бушменских племен наро и тати (дети), очевидно связанных близким родством, но плохо сопоставимых как с северно-, так и с южнобушменскими языками; позже в состав этой группы были включены и другие языки бушменов Бечуаналенд (Ботсваны). При этом она не могла не отметить многочисленные сходства между этими наречиями и готтентотским языком нама; однако, ввиду значительного культурно-антропологического контраста между бушменами и готтентотами, ни она сама, ни другие исследователи первой половины XX в. не решались эксплицитно поместить готтентотские и «центральнобушменские» языки в одну группу.

Ситуация осложнялась еще и окказиональными попытками связать готтентотские языки с некойсанскими языками Африки, в первую очередь «хамитскими» (т. е. чадскими и кушитскими), причем появление в них системы кликсов в рамках этой концепции объяснялось воздействием бушменского субстрата [Meinhof 1912]. Отдельные лексические и грамматические сходства между афразийскими и готтентотскими языками, действительно,

наблюдаются, но даже поверхностный анализ показывает, что количество их ничтожно мало по сравнению с изоглоссами, тесно связывающими готтентотские языки с «центральнобушменскими». Точку в этом вопросе поставило исследование Дж. Гринберга [Greenberg 1966: 66-72], который убедительно показал, что готтентотские языки следует объединять вместе с наро и дети в единую «центральнокойсанскую» группу.

Первая попытка обобщить накопленный материал по центр.-кой. языкам и предложить на его основе працентр.-кой. реконструкцию была предпринята уже в 1972 г. К. Баукомом [Baucoum 1974]; к сожалению, ее приходится признать во многом неудачной, т. к. автор был вынужден опираться в основном на старые, фонетически ненадежные записи «бушменских» языков. Однако уже в 1980-е гг. происходит значительный прорыв в полевой работе над языками Ботсваны и прилегающих территорий, в первую очередь благодаря экспедициям Р. Фоссена (продолжившего традицию, заложенную выдающимся койсанологом О. Кёллером), труды которого увенчались выходом в 1997 г. основополагающей монографии «Die Khoe-Sprachen» [Vossen 1997a], где не только подробно описывались фонологические и грамматические структуры не менее десятка ранее не известных центральнокойсанских языков, но и была предложена — впервые в истории койсанистики — строгая системная реконструкция фонетики, лексики и грамматики працентр.-кой. языка, последовательно опирающаяся на сравнительно-исторический метод.

Ввиду всех этих обстоятельств можно утверждать, что из всех «койсанских» групп центральнокойсанская семья на сегодняшний день изучена наиболее подробно, и, таким образом, составление 50-словного списка для працентр.-кой. языка будет сопряжено со значительно меньшими трудностями, чем для праож.-кой. языка.

Перечислим основных представителей центр.-кой. семьи, данные которых можно использовать для лексикостатистических подсчетов и этимологического анализа базисной лексики.

I. «Готтентотские» («кхойкхой») языки:

1) *Нама* [нам]. Общее название для большого пучка диалектов, на котором в совокупности говорят более 160,000 носителей в Намибии (где нама является одним из официальных языков) и еще несколько тысяч в Ботсване. Помимо собственно племен нама,

на близкородственных диалектах разговаривают также племена дамара, хай||ом, ʃакхой и другие; в настоящее время постепенно начинает закрепляться стандарт, согласно которому термин «нама» обозначает только конкретный диалект племен нама, а в роли общего термина для обозначения всего кластера используется слово *кхойкхойговаб* (*k^hoek^hoe gowa-b*), букв. «язык людей».

Вопрос о диалектном членении кхойкхойговаб подробно обсуждается в работе [Haacke et al. 1997], где, в частности, приводится список из 96 лексических элементов (отличный от списка Сводеша) с перечислением их отражений в десяти диалектах этого языка. Судя по приводимым данным, уровень лексической дифференциации в нама незначителен, и с точки зрения целей нашего исследования этой информацией можно пожертвовать без ощутимого ущерба для результатов. (В процессе составления подробного 100-словного списка по данным словаря [Haacke & Eiseb 2002], учитывающего диалектные особенности, оказалось, что процент явных расхождений между диалектами никогда не превышает 5-6 лексических единиц из 100).

В силу своей распространенности и высокого социального статуса язык нама исследован намного подробнее, чем любой другой «койсанский» язык. Статус классического давно приобрело фонологическое описание Д. Бича [Beach 1938], в котором, в частности, проводится попытка внутренней реконструкции предыстории нама, основанная на глубоком анализе его фонотактики и морфологии. Наиболее детальное грамматическое описание, отвечающее основным требованиям современной лингвистики, можно обнаружить в работе Р. Хэгмана [Hagman 1977].

Что касается лексикографических описаний, то первый словарь нама был составлен уже Й. Крёнлейном в конце XIX в. [Krönlein 1889]; до недавнего времени основным источником сведений по лексике нама было переиздание этого словаря, существенно дополненное и исправленное Ф. Рустом [Rust 1969]. Лишь относительно недавно вышел в свет совершенно новый вариант словаря нама, составленный южноафриканским лингвистом В. Хааке вместе с носителем этого языка Э. Эйсебом, сначала в виде предварительного «школьного» варианта [Haacke & Eiseb 1999], затем — в полном виде, включая тональную нотацию и большое количество языковых примеров [Haacke & Eiseb 2002].

Этот последний вариант по умолчанию является основным источником данных по нама, приводимых в нашей работе.

2) *!Кора* или *!корана* [кор]. Единственный из готтентотских языков, помимо нама, о котором можно составить довольно подробное представление, несмотря на то, что уже к середине XX в. он, по-видимому, являлся вымершим; информация сохранилась в основном благодаря грамматическим описаниям и словарям начала К. Вураса [Wuras 1920b], а затем К. Майнгофа [Meinhof 1930]. По своим объемам эти словари значительно уступают словарям нама, но для лексикостатистических целей их более чем достаточно. Источником по умолчанию ниже будет считаться словарь Майнгофа, более адекватно воспроизводящий фонологическую систему *!кора*, но в отдельных случаях имеет смысл приводить данные также и по словарю Вураса [Wu.], поскольку между этими двумя источниками имеется ряд небольших, но существенных расхождений (в том числе лексического характера).

II. «Бушменские» («калахари-кхойские») языки:

3) *||Ани* [ани]. Около 1,000 носителей в северо-западной части провинции Нгамиленд (Ботсвана), проживающих вдоль берегов р. Окаванго. Известны лишь данные, опубликованные в статье [Vossen et al. 1988] и в монографии [Vossen 1997a], а также в кратком фонетическом описании [Vossen 1986] и в грамматическом очерке [Vossen 2000].

4) *Г|анда* [анд]. Точное количество носителей неизвестно; основной информант Р. Фоссена, родившийся и всю жизнь проживший в Кхвай, относил себя к суб-этносу буга-кхой, хотя язык его явно отличался от языка буга. Данные приводятся только в работах [Vossen et al. 1988] и [Vossen 1997a].

4) *Буга* [буг]. От 1,000 до 2,000 носителей, проживающих в районе населенного пункта Кхвай непосредственно к югу от *||ани*. Все данные известны только по [Vossen et al. 1988] и [Vossen 1997a]. Расхождения между буга и г|анда минимальны; возможно, их следует рассматривать как два близких говора.

6) *Кхве*, или *кхой* (также *ху-кве*) [кхо]¹. Довольно многочисленное племя (не менее 4,000 носителей), раскиданное по южным

¹ *Кхве* и *кхой* (*кхое*, *кхое*, *кхой*, *кхве* 'человек, люди' в разных системах транслитерации) — разные произносительные варианты одного и того

областям Анголы и Замбии. Язык кхве был подробно исследован уже в 1960-е годы О. Кёлером, и большое количество лексической информации обнаруживается как в его грамматическом описании кхве [Köhler 1981], так и в отдельных статьях, исследующих те или иные аспекты лексикологии кхой или внешних связей этого языка [Köhler 1966, 1973/74, 1986 и др.]. Данные из этих публикаций были инкорпорированы в [Vossen 1997a].

Относительно недавно вышел в свет и подробный словарь кхве [Kilian-Hatz 2003], данные которого имеют большое значение для уточнения центр.-кой. реконструкции и пополнения этимологического корпуса, а в некоторых местах позволяют и уточнить отдельные вхождения в 100-словный список, составленный по материалам Кёлера. В нашем исследовании формы кхве по умолчанию приводятся в записи Кёлера; уточнения по словарю Килиан-Хатц помечаются как [КН].

7) *Наро* [нар]. Крупнейшее «не-готтентотское» племенное объединение в рамках центр.-кой. общности (от 6,000 до 9,000 носителей по разным подсчетам). Ареал распространения — пограничная область между западной Ботсваной и Намибией.

В силу многочисленности своих носителей язык *наро* был впервые описан уже Д. Блик, которая даже посвятила ему отдельную монографию [Bleek 1928]; лексические данные включены как в [Bleek 1929], так и в [Bleek 1956] (где он маркируется как СII). Следующий крупный этап — выход в свет словаря А. Барнарда [Barnard 1985], значительно более подробного, чем записи Д. Блик, но, к сожалению, разделяющего их основной недостаток — неточность транскрипции.

же слова, причем оно же служит и для обозначения центр.-койсанской этнолингвистической общности как таковой. В монографии Р. Фоссена автор, вслед за своим учителем О. Кёлером, вводит условное транскрипционное разграничение между *Khoe* (группа в целом) и *Kxoe* (язык кхой или кхве), пользуясь тем, что в некоторых диалектах кхве придыхательный велярный смычный k^h произносится как велярная аффриката *kx*. Чтобы избежать путаницы в кириллической транслитерации, мы будем записывать название *языка* как «кхве», а всей группы центральных наречий — как «кхой» (восходящая или нисходящая артикуляция дифтонга в этом слове также зависят от диалектных или идиолектных особенностей носителей, т. е. все эти разграничения носят чисто условный характер).

Ситуация изменяется к лучшему в результате полевых исследований Р. Фоссена, который приводит большой и, впервые, понастоящему качественно записанный материал по наро сначала в [Vossen et al. 1988], затем в [Vossen 1997a]. Наконец, из новейших достижений нельзя не отметить словарь Г. Фиссера [Visser 2001], много лет проработавшего с носителями наро и составившего, по-видимому, наиболее детальное на сегодняшний день лексикографическое описание наро, к тому же полностью адекватно передающее фонологию этого языка. Для наших целей в общем достаточно материалов Р. Фоссена; там, где между этими данными и словарем Фиссера обнаруживаются существенные фонетические или лексические расхождения, вариант по Фиссеру будет сопровождаться пометой [Vi.].

8-9) *Г/ви* [гви], *Г/ана* [ган]. Два близкородственных языка, носители которых проживают в центральных и южных областях Ботсваны (центральная Калахари). Общее число носителей достоверно не определено, по-видимому, около десятка тысяч. Оба языка были впервые описаны японским лингвистом Дзиро Танака, составившего по ним подробный словарь [Tanaka 1978]; к сожалению, качество записи оставляет желать лучшего, и там, где это возможно, предпочтительно доверять более поздним полевым записям Р. Фоссена [Vossen et al. 1988; [Vossen 1997a]. В приводимых ниже списках данные по [Tanaka 1978] помечены как [Та.].

Из наиболее современных исследований по этим языкам стоит отметить ряд статей Хироси Накагава [Nakagawa 1996, 1998], который, анализируя фонетику этих языков с помощью профессиональной звуковой аппаратуры, обнаруживает в ней ряд особенностей, ранее не фиксировавшихся койсанологами в принципе — например, наличие особого фонологически значимого абруптивного велярного исхода кликса (т. е. *kʔ*, *k!ʔ* и т. д.), противопоставленного как простому исходу на гортанную смычку (*p*, *!p*), так и простому велярному исходу (*kl*, *k!*), и даже абруптивному увулярному исходу (*qʔ*, *q!ʔ*). Хотя точность этого открытия пока не подтверждена, оно наглядно показывает, что, по-видимому, наши знания о фонетической и фонологической типологии щелчковых согласных еще далеки от идеального состояния.

10) *ᶑХаби* [хаб]. По-видимому, вымирающий язык, представленный лишь небольшой группой носителей, проживающих в

горах Кгвебе к югу от основного ареала распространения языка кхве. Единственные данные — полевые записи Р. Фоссена [Vossen 1997a].

11) *Тсixa* [тсх]. По данным Р. Фоссена, около 300-400 носителей в районе населенного пункта Мабабе (север Ботсваны). Лексика и грамматика известны только в передаче Р. Фоссена [Vossen et al. 1988; Vossen 1997a].

12) *Даниси* [дан]. Ареал распространения примерно такой же, как и у тсixa; число носителей неизвестно. Языковые данные, опять-таки, известны только по записям Р. Фоссена [Vossen et al. 1988; Vossen 1997a].

13) *Дети* [дет]. Около 1,000 носителей, проживающих в районе г. Ракопс (центральная Ботсвана). Язык известен только по [Vossen et al. 1988] и [Vossen 1997a]. Эндоним «дети» (также «тети», «тати») происходит от названия района Тати, в котором проживает это племя. В этой связи может возникнуть некоторая путаница, т. к. название «тати» употребляет, в частности, С. Дорнан [Dornan 1917], называющий так группу бушменов, язык которых на самом деле представляет собой не диалект дети, а наречие, близкое к диалектам куа, чуа и тсуа (см. ниже).

14-15) *Чара* [чар], *Хайсе* [хай]. Два близкородственных языка, занимающих соседние территории на юго-востоке Ботсваны; число носителей неизвестно. Все данные известны только по [Vossen et al. 1988] и [Vossen 1997a].

(Стоит отметить, что по классификации [Ethnologue 2009] языки 11-15 отмечены как диалекты единого языка *шуа*, в то время как Р. Фоссен склонен скорее считать их отдельными языками, объединяющимися в *подгруппу* шуа. Примирить эти позиции тяжело из-за нехватки данных; в целом, однако, классификация койсанских языков, приводимая в системе «Этнолог», не вызывает особого доверия, т. к. содержит большое количество явных таксономических ошибок).

16-18) *Чуа* [чуа], *куа* [куа], *цуа* [цуа]. Чрезвычайно близкородственные наречия — по-видимому, диалекты одного языка (именно так они классифицируются в [Ethnologue 2009]; все три названия представляют собой диалектные варианты одного и того же пратимона **kia*, точное значение которого, однако, остается неизвестным). Общее число носителей — ок. 800 человек. Область

распространения — непосредственно к западу от языков чара и |хайсе. Почти все данные по этим диалектам известны только из сравнительного описания Р. Фоссена [Vossen 1997a], однако попадают и отдельные публикации, посвященные фонетике цуа [Snyman 2000] и чуа [Chebanne 2000], которые, впрочем, для наших текущих целей не имеют особой значимости.

К этой же группе наречий, несомненно, относится диалект, описанный С. Дорнаном [Dornan 1917]; последний называет исследуемое им племя либо «бушменами тати», либо «масарва» (т. е. тем же словом, которое Д. Блик использует для обозначения одного из диалектов таа, см. выше), но в [Bleek 1956], куда данные Дорнана были инкорпорированы с ярлыком С1, этому диалекту присвоено название «хие» или «хие-чваре». Анализ данных по хие-чваре показывает, что этот диалект нельзя целиком и полностью отождествить ни с чуа, ни с тсуа Р. Фоссена (хотя по признаку палатализации начальных велярных он явно ближе к этим двум, чем к куа), так что данные эти будут цитироваться отдельно под ярлыком [хие]. Разумеется, качество транскрипции в работе С. Дорнана крайне низкое по сравнению с записями Фоссена; с другой стороны, Дорнан дает довольно подробное словарное описание, в отличие от Фоссена, данных которого по чуа, например, так мало, что по этому диалекту невозможно даже составить нормальный 50-словный список.

Выше были перечислены все центр.-кой. языки и диалекты, по которым мы либо располагаем подробными словарными описаниями, либо краткими списками слов, записанными на профессиональном уровне и имеющих важное значение для классификации. За пределами исследования остаются лишь отдельные мелкие списки слов, в основном по вымершим языкам или диалектам, записанные в старых работах и вряд ли способные сколько-либо серьезно повлиять на наше представление о працентр.-кой. языке или общей таксономической характеристике этой семьи — такие, как, например, краткий словарь языка гриква (гири-ква), близкородственного нама и !кора [Meinhof 1930: 147-152]; немногочисленные сравнительные данные по ряду центр.-кой. диалектов, приводимые в [Maingard 1961] и некоторые другие.

Общая классификация, предложенная Р. Фоссеном для центр.-кой. группы, в целом несильно отличается от предыдущих

моделей (О. Кёлер, Э. Вестфаль и др.), но является более детальной и учитывает большее количество языков. Согласно Фоссену, центральнокойсанские языки, или языки кхой, группируются следующим образом [Vossen 1997a: 386]:

(А) Языки *кхойкхой* (= готтентотские): нама, !кора, гири-ква и др.

(Б) Языки *калахари-кхой*¹:

(Б.1) Западнокхойские языки:

(Б.1.1) Подгруппа *кхве*: кхве, буга, ||ани, г|анда;

(Б.1.2) Подгруппа *наро-г||ана*:

(Б.1.2.1) наро;

(Б.1.2.2) г||ана, г|ви, †хаба;

(Б.2) Восточнокхойские языки:

(Б.2.1) Подгруппа *шуа*: чара, дети, |хайсе, тсиха, даниси;

(Б.2.2) Подгруппа *чуа*: куа, чуа, цуа.

Эта классификация была получена Фоссеном на основании тщательного анализа в первую очередь фонетических и грамматических изоглосс между центр.-кой. языками, а также примерных лексикостатистических подсчетов (результаты опубликованы в [Vossen 1984] и в [Vossen et al. 1988]; вообще Р. Фоссен является одним из немногих африканистов, последовательно применяющих лексикостатистическую методику — ср., в частности, его тщательную работу над стословными списками по языкам восточнонилотской группы [Vossen 1982: 105-156]).

2.4.2. *Историческая характеристика.* Благодаря тому, что для центр.-кой. языков сегодня есть как минимум три подробных словаря, отвечающих современным требованиям (нама, наро, кхве), а также большое количество самых разнообразных и, как правило, достаточно хорошо записанных материалов по другим

¹ Р. Фоссен в своей монографии использует для обозначения этой ветви термин «не-кхойкхой» (Nicht-Khoekhoe), но в силу своего необычного устройства он сегодня используется реже, чем *чу-кхве* (Tshu-Khwe, по названиям двух пучков диалектов, маркирующих восточную и западную границу распространения этих языков) или *калахари-кхой* (т. к. подавляющее большинство носителей этих языков действительно локализовано в границах пустыни Калахари).

языкам, сравнительно-исторические исследования в рамках этой группы проводить намного проще, чем в рамках юж.-кой. группы. На наш взгляд, Р. Фоссену удалось вполне надежно восстановить основные фонетические и грамматические характеристики працентр.-кой. языка, потомки которого к тому же отличаются высокой консервативностью.

Ближе всего к праязыку из его современных наследников оказываются, по-видимому, такие языки, как наро, г|ви и г||ана. Здесь почти в неизменном виде сохраняется как система основ кликсов (стандартные четыре типа: дентальный /, альвеолярный /, палатальный /, латеральный //), так и система исходов, где Фоссен восстанавливает семь возможных типов, среди которых: глухой веларный / «нулевой» *k*, звонкий веларный *g*, фрикативный веларный *x*, носовой *n*, гортанный смычный *ʔ*, гортанный придыхательный *h* и веларная абруптивная аффриката *kx*.

Под вопросом остается разграничение между двумя типами носовых исходов — «простой» преназализацией кликса, которую Фоссен, вслед за О. Кёлером, обозначает надстрочной тильдой ($\tilde{}$ и т. п.), и преназализацией «коартикуляционного» характера, автоматически обусловленной эксплицитным произнесением заднеязычного носового исхода, которую он обозначает надстрочной тильдой в сочетании с последующим *n*, противопоставляя, таким образом, фонемы $\tilde{}$: $\tilde{}/n$, $\tilde{}$: $\tilde{}/n$ и т. д. В описании кхве по версии К. Килиан-Хатц [Kilian-Hatz 2003: 8] эта оппозиция трактуется иначе: «простые преназализованные» кликсы определяются как преназализованные кликсы с исходом на звонкий веларный согласный (т. е. $\tilde{}$ Кёлера и Фоссена = *n/g* у Килиан-Хатц), а «преназализованные + исход на носовой» — как обычные назализованные (т. е. $\tilde{}/n$ Кёлера и Фоссена = *n/* у Килиан-Хатц).

Фонологическая оппозиция между этими двумя типами исходов постулируется Фоссеном практически для всех языков калахари-кхой, с которыми он имел дело в своих полевых исследованиях, но независимое подтверждение получила пока что только в работах К. Килиан-Хатц по кхве, и для того, чтобы уверенно восстановить ее на працентр.-кой. уровне, необходимы дальнейшие подробные исследования, тем более, что и в полевых записях Фоссена нередко наблюдаются бессистемные колебания в рамках даже близкородственных языков, несводимые к простой и

однозначной системе соответствий. При этимологическом анализе элементов 50-словного списка мы будем, вслед за Фоссеном, транскрипционно различать два типа исходов как \tilde{r} : $\tilde{r}n$ и т. п., но в реконструкциях для простоты будем обобщать оба типа как «простую назализацию» (\tilde{r}). В любом случае, выносить это противопоставление на внешний уровень сравнения *до* прояснения ситуации на материале самих центр.-кой. языков было бы явно преждевременно.

Важная типологическая особенность, отличающая центр.-кой. языки от их сев.-кой. и юж.-кой. соседей — явно более высокий процент среди исконной лексики слов, начинающихся с обычных нещелчковых согласных. Только для працентр.-кой., в частности, вполне обычными (хотя и не высокочастотными) оказываются корни, начинающиеся с губного смычного **p-* (**pá* 'кусать', **pê* 'прыгать', **pî* 'грудь, молоко' и др.) и губного носового **m-* (**mî* 'видеть', **mâ* 'голова', **mâ* 'давать' и др.)¹. При этом сама система подчеркнута «экономна» по сравнению с гораздо более детализированными системами в сев.-кой. и юж.-кой. языках. Р. Фоссен восстанавливает ее в следующем виде:

<i>*p</i>	<i>*t</i>	<i>*c</i>	<i>*k</i>	<i>*?</i>
	<i>*t^h</i>		<i>*k^h</i>	
<i>*b</i>	<i>*d</i>		<i>*g</i>	
		<i>*cʔ</i>	<i>*kx</i>	
		<i>*s</i>	<i>*x</i>	<i>*h</i>
<i>*m</i>	<i>*n</i>			

Впрочем, можно отметить, что и система кликсов в этих языках также воспринимается как «упрощенная»: в ней отсутствуют не только нетривиальные типы основ кликсов (такие, как ретрофлексная в сев.-кой. и лабиальная в юж.-кой. языках), но и некоторые типы исходов, довольно стандартные для прочих «бушменских» групп.

Впрочем, детальный анализ материала показывает, что система Фоссена может оказаться несколько упрощенной. Так,

¹ В сев.-кой. и юж.-кой. языках исконные корни на **p-* не встречаются вообще, а надежные корни на **m-* исчисляются буквально единицами.

например, во многих центр.-кой. языках надежно зафиксированы подсистемы увулярных смычных (глухой *q*) и кликсов с исходом на увулярные звуки (обычно на тот же глухой смычный *q*, ср. бунд, анд *lqíè* 'усталый', *llqé* 'свежевать' и др.). Посвящая этому вопросу отдельную статью [Vossen 1992], Р. Фоссен приходит к выводу, что эти фонемы в тех языках, где они встречаются, скорее должны рассматриваться как инновации. Эта позиция, однако, оспаривается в [Starostin 2008a: 435-436] (основной аргумент сводится к тому, что невозможно установить регулярные условия «увуляризации» начальных согласных или исходов кликсов в тех корнях, которые надежно прослеживаются на працентр.-кой. уровне). С другой стороны, нельзя исключить и то, что увулярные согласные и исходы кликсов в центр.-кой. языках на самом деле являются результатом переноса на согласный фарингализованной артикуляции соседнего гласного (см. ниже).

В той же работе предлагаются и некоторые коррективы в области реконструкции вокализма, где Фоссен восстанавливает стандартную пятичленную схему (**a*, **e*, **i*, **o*, **u*) с возможным наложением дополнительного признака назализации, а также с возможностью образования дифтонгических сочетаний (**ae*, **ai*, **ao*, **au*, **oe*, **oa*, **ui*) в рамках «двухморной» базовой структуры словоформы, которая в центр.-кой. языках устроена практически так же, как и в сев.- и юж.-кой. группах. В частности, на основании ряда нетривиальных соответствий, где «калахари-кхойские» гласные среднего подъема *e*, *o* в кхойкхой немотивированно соответствуют гласному нижнего подъема *a*, предполагается в качестве отдельных фонем восстанавливать «открытые» гласные **ɛ*, **ɔ* (ср. выше аналогичное предложение для юж.-кой. реконструкции).

Из дополнительных артикуляторных признаков Р. Фоссен признает безусловную фонологическую значимость для працентр.-кой. языка назализации гласных (хотя в отдельных случаях сравнительный материал убедительно показывает, что назализованные дифтонги могут возникать из стяжения двусложных словоформ с срединным носовым, ср. кор *kxàní-* 'птица' = нар *kxáí* *id.* и др.).

В подветви наро-г|ви ветви калахари-кхой фонологическую значимость имеет также признак фарингализации; вследствие столь узкой дистрибуции Р. Фоссен склонен считать его ареальным и

не проецирует фарингализацию на працентр.-кой. уровень, но и определить источник возникновения этой оппозиции ему также не удастся. На данном этапе имеет, по-видимому, смысл с ним согласиться по «техническим» причинам: фарингализация систематически отмечается только в словаре наро Г. Фиссера, а полевые записи Фоссена по нар, гви, ган и хаб (тем четырем языкам, в которых фарингализованные гласные — отдельные фонемы) внутренне противоречивы и не позволяют установить однозначные соответствия или четкие контексты, в которых простые и фарингализованные гласные нейтрализуются в пользу простых. Ввиду этого обстоятельства в наших предварительных реконструкциях фарингализация формально учитываться не будет, хотя при разборе конкретных этимологий мы все же будем акцентировать внимание на случаях ее появления в нар и др.

В целом большинство предложенных корректив к системе Фоссена носит сугубо уточняющий характер; необходимости ее радикального пересмотра нет, т. к., насколько удастся установить в процессе дальнейшей этимологизации центр.-кой. базисной лексики, она оказывается вполне приемлемой даже за пределами того сравнительного лексического материала, которым Р. Фоссен подкрепляет свою реконструкцию в приложении к монографии.

Из основных тенденций эволюции працентр.-кой. фонологического инвентаря следует упомянуть упрощение (иногда довольно кардинальное) системы кликсов в языках-потомках. Типичной рефлексацией в языках калахари-кхой является, например, отпадение альвеолярной основы кликса (ср. **!xáð* 'бегемот' → нар *!xáð*, но ани, бут, анд, хай, дет, цуа *xáð*); в восточном ареале калахари-кхой столь же типична палатализация и утрата ингрессивной артикуляции у палатальной основы кликса (ср. **!kǣ* 'плевать' → ани *!kǣ*, гви *!kǣi*, но хай *ǣʼ*, дет *ǣʼ*). В некоторых случаях (обычно также в восточной подгруппе, см. ниже) отмечается даже спорадическое отпадение латеральной основы кликса (ср. **!ráù* 'рыба' → нар *!ráù*, но хай, дет *ʼáù*, дан *ʼáù*).

Таким же образом упрощается в ряде языков и система исходов. Здесь своего рода «лидером» в подгруппе кхойкхой является нама, где исход на велярную абруптивную аффрикату сливается с простым исходом на гортанную смычку (ср. **!kǣ* 'плевать' → нам *ʼáǎ-b* 'слюна'), а «нулевой» исход на глухой велярный — с исходом

на звонкий велярный (ср. */ú '(быть) близким' → нар /ú, но нам /ú); в результате число фонологически дистинктивных исходов в нам сводится к пяти (по сравнению с, например, семнадцатью в !хонг).

Разумеется, типологически сходные тенденции прослеживаются и в других койсанских языках (ср. описанную выше ситуацию в ||хегви), но только в центр.-кой. группе они носят столь массовый характер. Одно из возможных объяснений — более тесные языковые контакты носителей центр.-кой. языков с носителями местных языков банту, что приводит к своеобразному «фонетическому нивелированию» обеих семей: с одной стороны, упрощенные подсистемы кликсов активно проникают в фонологические системы банту, с другой — их общее количество, наоборот, снижается в системах самих центр.-кой. языков.

Основные изменения в системах вокализма и «нешелчкового» консонантизма подробно описаны в главе, посвященной реконструкции працентр.-кой. состояния в [Vossen 1997a]; отдельные моменты будут откомментированы ниже в ходе анализа 50-словных списков. Что касается тональных систем центр.-кой. языков, то вопрос этот еще недостаточно хорошо изучен: Р. Фоссен последовательно реконструирует тоны только на промежуточных уровнях, и то лишь в тех случаях, где тональные соответствия оказываются «тривиальными». Следуя его примеру, мы пока что предлагаем тональную реконструкцию только для тех статистически надежных случаев, где подавляющее большинство языков демонстрирует примерно одинаковые тональные характеристики¹.

Грамматическая система в центр.-кой. языках имеет довольно значительные отличия от соотв. систем в сев.- и юж.-кой. языках. В частности, сложно устроенным системам именных классов (как

¹ Из последних работ по сопоставительной тонологии центр.-кой. языков следует особо отметить [Elderkin 2004, 2008], в которых выделены основные ряды соответствий и предложена двухтональная реконструкция для працентр.-кой. состояния, и [Honken 2008], где исследуется связь между тоногенезом в отдельных центр.-кой. подгруппах и ларингальными признаками начальных согласных. Тем не менее, ни реконструкцию Элдеркина, ни систематизацию Хонкена нельзя считать окончательными из-за выборочности материала, на котором они демонстрируются; достоверность их можно будет показать только в процессе составления подробного этимологического словаря центр.-кой. языков.

продуктивным, в таких языках, как !хонг, так и реликтовым, как в сев.-кой. языках) здесь противопоставлена система из трех родов (мужского, женского, среднего), надежно выводимая на пра-центр.-кой. уровень — именно эта типологическая черта в свое время стимулировала К. Майнгофа отнести «готтентотские» языки в разряд «хамитских». Глагольная морфология отличается повышенной (по сравнению с другими «койсанскими» языками) сложностью: словоформа может быть выстроена по агглютинативной модели, с цепочками из трех-четырех суффиксальных показателей времени, наклонения, залога и т. п., однако личное спряжение при этом отсутствует. Подробное описание и историческая интерпретация всех этих деталей приводятся в монографии Р. Фоссена; для нашего исследования они в целом несущественны, т. к. правила стыковки лексических основ и продуктивных словоизменяемых морфов в центр.-кой. языках, как правило, тривиальны.

Значительно более сложен вопрос о возможной исторической членимости самих лексических основ, устроенных по «двухморному» принципу. «Условно-суффиксальные» моры в языках этой группы имеют примерно такие же фонотактические ограничения, как и в сев.- и юж.-кой. группах, что стимулирует желание трактовать их как следы старых «окаменевших» суффиксов. Однако конкретных аргументов в пользу такой трактовки собрано пока немного: бросаются в глаза лишь отдельные случаи, такие, как ани *lábè*, кхо *ngábè*, нар *lábé* и т. д. 'жираф', но нам, кор *lai- id.* — здесь основа **la-be* в языках калахари-кхой противопоставлена основе **la-i* в языках кхойкхой, что может быть истолковано лишь как результат относительно раннего распределения по диалектам пра-кхой двух морфологических вариантов. Аналогичное варьирование наблюдается и в глагольных основах, ср. ани *l'óbè*, буг *l'oébé*, нар *l'óbè* и др. 'целовать', но нам *l'oa id.*

До тех пор, пока не будет проведена полная инвентаризация такого рода расхождений (включая, может быть, даже попытку их семантической интерпретации), полностью игнорировать «суффиксальную мору» — например, отбрасывая ее при внешнем сравнении — непозволительно; тем не менее, основной акцент при таком сравнении все равно должен ставиться на «корневую»

мору как безусловную составляющую основного лексического значения.

В плане *лексики* современные центр.-кой. языки — особенно крупные, такие, как нам или нар — по отношению к своим «койсанским» соседям чаще оказываются донорами, нежели акцепторами. В работе [Sands 2001: 201] перечислено не менее пяти контактных зон, где центр.-кой. языки (в основном нам и нар) однозначно определены как источники заимствований в жу|хоан, !хонг, |хам и т. д. Обратные заимствования встречаются реже и, как правило, легко отсеиваются через дистрибуционный анализ, так что в ходе этимологического анализа 50-словного списка проблема отсеивания «псевдо-кандидатов» на статус выражения того или иного сводешевского значения как заимствований из внешних источников встает крайне редко (исключение — несколько любопытных случаев ареальных изоглосс между нар и жу|хоан, возможно, заимствованных из какого-то третьего источника, см. ниже 'лист', 'вошь').

2.4.3. *Лексикостатистика*. Проверка классификационной модели Р. Фоссена с помощью лексикостатистического метода серьезно осложняется тем, что по большинству «мелких» языков, основные данные по которым берутся из опубликованных полевых записей самого Фоссена, невозможно составить качественные 100-словные списки. В монографии [Vossen 1997a] автор приводит в первую очередь сравнительно-этимологический материал, не цитируя при этом случаи лексических инноваций в отдельных языках-потомках. Намного большую ценность имеют сравнительные лексикостатистические списки, опубликованные в [Vossen et al. 1988], однако и здесь налицо две серьезные проблемы: во-первых, Р. Фоссен оперирует собственным списком, который совпадает со стандартным инвентарем Сводеша примерно наполовину, во-вторых, есть основания сомневаться в семантической аккуратности этих данных, особенно в случае неустойчивых этимонов (см. ниже, например, замечания к этимону 'лист'), имеющих ключевое значение для внутренней классификации группы.

По-настоящему надежные данные имеются только по языкам, представленным большими словарями (нама, !кора, кхве, наро, г|ви, г|ана) или хотя бы относительно детальными лексическими и/или грамматическими описаниями, позволяющими заполнить

большую часть списка (||ани, хие-чваре). В таблице 3 приводятся данные по процентам лексических совпадений между всеми этими языками, а также *примерные* цифры совпадений между ними и языками ʃхаба, тсиха, дети и куа (последние три — «диагностические» представители подгрупп шуа и чуа, выделяемых Р. Фоссеном).

Таблица 3. Лексикостатистическая матрица
центральнокойсанских языков (100-словные списки).

Язык	!Кора	Ани	Кхве	Наро	Г ви	Г ана	ʃХаба	Тсиха	Дети	Куа	Хие.
Нама	0.78	0.44	0.41	0.46	0.48	0.47	0.54	0.47	0.49	0.48	0.45
!Кора		0.47	0.43	0.48	0.49	0.49	0.56	0.49	0.52	0.52	0.48
Ани			0.89	0.74	0.77	0.77	0.82	0.89	0.81	0.85	0.73
Кхве				0.69	0.75	0.74	0.78	0.87	0.76	0.82	0.71
Наро					0.82	0.84	0.94	0.73	0.73	0.80	0.72
Г ви						0.98	0.91	0.76	0.75	0.85	0.78
Г ана							0.91	0.78	0.77	0.84	0.78
ʃХаба								0.80	0.76	0.86	0.80
Тсиха									0.81	0.85	0.75
Дети										0.83	0.77
Куа											0.89

Данная матрица, равно как и соответствующее ей генеалогическое древо центр.-кой. языков (схема 3), показывает, что лексикостатистическая классификация во многом согласуется со схемой Р. Фоссена, но в ряде случаев содержит существенные противоречия, требующие внимательного рассмотрения; к сожалению, по-настоящему снять их окажется возможным лишь после выхода в свет подробных словарей по всем перечисленным языкам (или, по крайней мере, после сбора и публикации более аккуратно составленных 100-словных списков).

Бросается в глаза значительная лексическая — и соответствующая ей хронологическая — дифференциация между кхойкхой и калахари-кхойскими языками (средний процент совпадений между нама и калахари-кхой примерно равен 50%, а между

языковыми подгруппами калахари-кхой никогда не опускается ниже 70%). Такая дифференциация хронологически соотносима со временем распада праюж.-кхой. языка на !кви и таа (\approx рубеж I тыс. до н. э.) и может создавать серьезные проблемы с точным определением працентр.-кхой. этимонов, соответствующих отдельным элементам 50-словника. Тем не менее, вряд ли имеет смысл разбирать кхойкхой и калахари-кхой лексику в отдельных разделах, учитывая, что общие элементы 50-словника, реконструируемые для соотв. праязыков, все же значительно превышают число различий.

Чрезвычайно сомнительным по итогам подсчетов оказывается выделение Р. Фоссенем генетически единой *западнокхойской* группы, противостоящей восточнокхойской группе, т. к. средний процент совпадений между ||ани-кхой и наро-г||ана и соответствующий процент между любой из этих двух групп и восточнокхойскими языками примерно одинаков. Аналогичным образом не выявляется и специфическая лексическая близость друг к другу ветвей шуа и чуа, объединяемых Р. Фоссенем в *восточнокхойскую* группу.

По всей видимости, распад подгруппы калахари-кхой на четыре ветви (кхой, наро-г||ана, шуа и чуа) произошел примерно на рубеже I тыс. н. э. и носил достаточно стремительный характер — настолько, что, даже если «западно-» и «восточнокхойские» узлы и соответствуют какой-либо исторической реальности, их кратковременное существование на лексикостатистической основе либо не выявляется вовсе, либо требует гораздо более тщательно составленных списков (возможно, даже 200-словных).

Отметим отдельно, что сопоставление реконструированных стословных прасписков по четырем основным ветвям калахари-кхой, где за праформу принимается этимон, имеющий наиболее широкую дистрибуцию внутри ветви, дает следующие результаты:

	Наро-г ана	Шуа	Чуа
Кхой	0.84	0.91	0.84
Наро-г ана		0.91	0.89
Шуа			0.96

Эти цифры в некотором смысле заслуживают большего доверия, чем результаты по отдельным языкам, т. к. сравнительный материал в работах Фоссена и других исследователей часто позволяет с гораздо большей уверенностью определять корень, выражавший 100-словное значение на одном из промежуточных уровней, чем корень, выражающий его в том или ином языке-потомке. В частности, они подсказывают, что лексикостатистических оснований для постулирования западнохой единства действительно нет, в то время как восточнохойские языки все же связаны друг с другом более тесно и, вероятно, какое-то время были представлены единым праязыком.

Одна из причин такого расхождения между подсчетами по живым языкам и по реконструкциям — очевидно аномальный подскок процентных совпадений между западнохойской подгруппой ||ани-хой и восточнохойским языком тсixa. С точки зрения общей лексикостатистической методологии такая ситуация может отражать неопознанные эффекты дополнительных языковых контактов между этими языками; и действительно, Р. Фоссен [1997а: 394-396] отмечает, что в социолингвистическом плане язык тсixa находится под сильным «конвергентным» влиянием со стороны языка кхве, которое даже привело к тому, что в тсixa, наравне с прочими восточнохойскими языками затронутым процессом «декликсизации» палатальных кликсов (*ʃ*), эти кликсы впоследствии оказались «восстановлены» в результате интенсивных контактов с кхве (ср. ниже такие примеры, как 'черный', 'есть', 'ухо' и др.).

Таким образом, в рамках «конвергентно-дивергентной» модели языкового развития язык тсixa, наверное, пришлось бы считать «смешанным» (западно-восточным). В рамках более традиционной, чисто дивергентной модели главным вопросом является корректная идентификация относительно недавних заимствований в тсixa из кхве и устранение их из лексикостатистических подсчетов (задача очень сложная, учитывая очевидную фонетическую близость обоих языков, а также возможную адаптацию-«пересчет» хойских заимствований под фонетические правила тсixa). Окончательное решение этого вопроса вряд ли возможно до публикации полного грамматического и лексикографического описания тсixa.

Еще одно расхождение между итоговой классификацией Фоссена и нашими подсчетами касается языка ꠘхаба, который, согласно Фоссену (а также более ранней классификационной модели О. Кёлера), относится к подгруппе г||ана-г|ви, при том, что максимальное число совпадений он имеет не с г||ана и не с г|ви, а с наро. Данный лексикостатистический результат хорошо соотносится и с недавними наблюдениями Х. Накагава [Nakagawa 2011], который приходит к выводу о необходимости перемещения ꠘхаба на генеалогическом древе ближе к наро на основании ряда важных фонетических изоглосс (таких, как, например, соответствие увулярных исходов кликсов в г||ана-г|ви фарингализованным гласным в наро и в ꠘхаба). По-видимому, в более ранних классификационных моделях на позицию ꠘхаба оказывали влияние соображения этнолингвистического характера, т. к. племена, говорящие на нем, образуют единую конгломерацию с г||ана и г|ви, противопоставленную ареально обособленному этносу наро.

50-словный список для центральнокойсанских языков.

1) «пепел»: (а) нам *ca'ó-b*, кор *t^hao-b*, нар, ган, хаб *t^hái*, дан *t^hái*; (б) ани, ган *ʃòà*, кхо *ʃòá*, бут *ʃòá*, анд *ʃòà*, гви *ʃúà*, хай, куа *zòá*, чар *zòà*, тсх, дан *du'íà*, цуа *z'íá*, хие *ʒoa*; (в) дет *dì*.

Для прахойкхой однозначно восстанавливается этимон (а) **t^hao* ← працентр.-кой. **t^hái* (с регулярным развитием **t^h* → *c*- в нама); он также сохраняется и в подгруппе наро-г||ана, в других же ветвях либо исчезает, либо изменяет семантику (= кхо *t^héú* 'труг', ани *t^hái* 'пламя').

Працентр.-кой. корень (б) **ʃòà* в значении 'пепел' представлен только в языках калахари-кхой (подгруппа кхве западнокхойской ветви и восточнокхойская ветвь); в кхойкхой он имеет значение 'грязь', 'глина' (нам *ʃgòà-b*, кор *ʃòà-s*). Однозначно предпочесть (а) или (б) для працентр.-кой. этимона затруднительно, но **t^hái* отличается лучшей репрезентативностью, а семантическое сужение значения 'грязь' → 'пепел' представляется более естественным, чем обратное расширение 'пепел' → 'грязь'.

Дет *dì* этимологически связано с кхо *dí* 'угли, используемые для лечения', далее, вероятно, с нам *dì-nú-b* 'горячий уголек'; очевидная инновация ('горячие угли' → 'пепел').

2) «птица»: (а) нам àní-b, кор kxàní-s; (б) ани, бут, хаб, дан zàrá¹, кхо žárá, анд, гви zárá, нар qará [Vi.], ган zára, хай, цуа zèrà, дет, чар zàrà, тсх zìrà, куа zérá, хие zera.

Корень (а) отражает працентр.-кой. *kxani = калахари-кхой *kxāi (с моносиллабизацией слога) 'стервятник' = ани kxāi, бут, анд, нар kxāi, хай, дет, чар kxāi и др. Корень (б), отражающий калахари-кхой *zara или *zera, в языках кхойкхой не зафиксирован.

Исходя из семантики корней, разумно предположить, что в кхойкхой имело место семантическое развитие 'стервятник' → 'птица (par excellence)' (к типологии такого рода развитий ср., например, греч. ὄρνις 'птица' ← индоевроп. *er- / *or- 'орел' [Pokorny 1958: 325-326]), в результате которого из употребления оказался вытесненным старый працентр.-кой. корень *zara / *zera (последний, кстати, должен был бы закономерно развиваться в пракхойкхой *cara, что привело бы к омонимии с основой *cara 'пыль, облако пыли' → нам cārá-b и др., так что процесс мог иметь и обратную хронологическую последовательность).

3) «черный»: нам fūi, кор fū, ани, гви, ган, хаб fñi, кхо fú, бут fū, анд, тсх fú, нар fñi, хай nzi, дет uí, чар uí, дан ndú, куа zú, цуа dú, хие ži-nye.

Все формы закономерно восходят к працентр.-кой. *fū; в правост.-кхой. налицо регулярная «декликсизация» *fñi → *nzi, с последующим «восстановлением» щелчковой артикуляции в тсх. Хие ži-nye расширено за счет дополнительного суффикса, но зафиксирована и чистая форма: ži 'темнота'.

4) «кровь»: (а) нам /áá-b, кор /rai-b, ани, кхо, бут, анд, нар, гви, ган, хаб, хай, дет, чар, тсх, дан, чуа /áá; (б) куа, цуа tákà, хие t^haka.

¹ В записи Р. Фоссена в этой и аналогичных формах вместо интервокального -r- записывается -d-. Реально, насколько удается судить по имеющимся описаниям, ни в одном центр.-кой. языке фонологической оппозиции между -d- и -r- не существует, а конкретная фонетическая реализация варьирует от языка к языку, но чаще всего носит все же сонорный характер. В целях экономии и унификации мы будем все случаи интервокального фонологического -d- в записях Фоссена транскрибировать как -r-. В реконструированных формах, следуя общему правилу, принятому в данной работе, варьирующая артикуляция -r/-d- будет отмечаться с помощью символа *r.

Почти все языки стабильно сохраняют працентр.-кой. основу **ʔád*; частичная замена имеет место только в подгруппе куа, где для диалектов куа и цуа приводится этимологически неясная форма *tâkà* (относительная архаичность которой подтверждается записью начала XX в. *tʰaka* у Дорнана в хие). Фонотактическое устройство этого слова выдает явно не «исконно-койсанское» происхождение, но конкретный источник заимствования установить пока что не удается.

5) «кость»: (а) нам *ʔxòǝ-*; (б) ани, кхо, буг, нар, гви, ган, хаб, хай, чар, тсх, дан *ʔǝǎ*, анд *ʔǝǎ*, дет, куа *ʔǝǎ*, чуа, цуа *ʔǝǎ*, хие *n|gwa*.

Пракхойкхой корень однозначно не устанавливается, так как слово 'кость' не зафиксировано для кор ни в записях Майнгофа, ни у Вураса; что касается нам *ʔxòǝ-*, то ему в кор соответствует *ʔxǝ-b* 'косточка (плода)'. Это заставляет предположить, что кхойкхой **ʔxo* по крайней мере в нам вытесняет старый этимон 'кость', сохраняющийся в калахари-кхой **ʔǝǎ* (← працентр.-кхой. **ʔǝǎ* или **ʔona* с вокализацией носового согласного).

6) «ноготь»: (а) нам *||óró-s*, кор *||ǝrǝ-b*, нар *||órǝ* [Vi.], гви, ган *!ore* [Ta.], дет *||óró*, чар, куа *||óró*, дан *||áró*, хие *!ara* (?); (б) ани, тсх, дан *||à*, кхо *||à*, буг, чар *||à*, хай *||à*.

Налицо сложно устроенная в дистрибуционном плане «конкуренция» двух корней — **||oro* и **||a*, первый из которых формально является общецентр.-кой., второй — обще-калахари-кхой. Оба корня одновременно встречаются только в ряде языков подгруппы шуа, но разница в значении остается неясной; судя по уточнениям Р. Фоссена, только **||oro* может встречаться в значении 'копыто', т. е. первичное распределение могло иметь вид **||oro* = 'ноготь (на пальце ноги)', 'копыто', **||a* = 'ноготь (на пальце руки)', но этот вывод крайне ненадежен без эксплицитного подтверждения такой семантической оппозиции хотя бы в одном словарном описании. На уровень внешнего сравнения допустимо выводить оба этимона, отдавая, тем не менее, предпочтение **||oro* как имеющему более широкую дистрибуцию.

В рефлексии отмечаются отдельные нерегулярности, как реальные (хай *||à* с неожиданным придыханием), так и фиктивные: скорее всего, ошибочно записана дентальная артикуляция кликса в хие (см. аналогичную орфографию в слове 'луна') и альвеолярная в гви и ган в записях Танака (также частая ошибка).

7) «умирать»: нам //ʔó, кор //ʔo, ани, буг, анд, кхо, нар, ган, гви, хаб, хай, дет, чар, дан, тсх //ʔó, цya ʔó, хие o.

В кya ср. также //ʔó 'смерть' (глагольное значение не зафиксировано). Все формы стабильно отражают працентр.-кой. *//ʔó. Отпадение латеральной основы кликса в диалектах подгруппы чуа наблюдается и в других случаях, причем исключительно в случае исхода на гортанный взрыв (ср. *//ʔan 'созревать' → кya ʔàí, цya ʔàí и др.), однако нехватка материала мешает разобраться в ситуации подробно.

8) «собака»: (а) нам àrí-, кор ʔàrí-b, ани ʔèrí-kù, нар hq-gù [Vi.], гви hárí-gù, хаб hárí-gù; (б) кхо ʔápā, буг, анд ʔápà, гви aba [Ta.], ган, хай, чар, кya, цya ʔábà, дет, тсх, дан ʔábá, хие aba.

Для пракхойской однозначно реконструируется форма *ʔaci-; за пределами кхойкхойских языков ситуация более сложная.

Если для (гипотетического) правост.-кхой. языка реконструкция *ʔábà однозначна, то в зап.-кхой. языках фонетически сходные формы конкурируют с этимоном *[h]ari-gu (форма hq-gù в нар — стяжение, ср. старый вариант *aru-gu* из публикации 1891 г., упоминаемый в [Bleek 1956: 11], а также промежуточный вариант *hàí-gù* в записи Фоссена). При этом в «нормальной» форме ʔábà корень (б) в зап.-кхой. языках налицо только в подгруппе гви-гана; в подгруппе кхве представлен инвариант *ʔara с глухим интервокальным -p-, что нарушает стандартные законы центр.-кой. фонотактики.

Все это наводит на мысль, что такие формы, как *ʔaba, *ʔara не являются исконно центр.-койсанскими, а появились независимо друг от друга в правост.-кхой. и других мелких подгруппах калахари-кхой в результате активных контактов с языками банту, где наиболее распространенным корнем со значением 'собака' является прабанту *-bíá. Более архаичная ситуация сохраняется в кхойкхой и отчасти в зап.-кхой. языках, где старая основа *ʔari была расширена за счет дополнительной суффиксации.

9) «пить»: нам â, кор kxâ, ани, буг, анд, кхо, нар, гви, ган, хаб, дан kxâ, хай, дет, чар, тсх, кya, цya kʔâ, хие ʔʰa.

Все формы закономерно отражают працентр.-кой. *kxâ (с регулярным отпадением велярной аффрикаты в нам). Запись палатального кликса в хие, скорее всего, ошибочна (отмечено еще по меньшей мере несколько случаев, когда Дорнан почему-то

передает правост.-кхой. *kʷ- через ʃ(h)-; ср. куа, цуа kʷai̯ 'смеяться' = хие ʃʰae-yo 'смех' и т. п.).

10) «сухой»: (а) нам ʃai̯-sá; (б) кор /ʔō, нар. куа, цуа /ʔō, хие /o; (в) ани, кхо, гви, ган, хаб, тсх, дан //xó, буг, дет //xǒ.

Нам ʃai̯-sá, адъективное производное от глагольной основы ʃai̯ 'высыхать', 'пересыхать', 'увядать' — явная инновация (исторически, скорее всего, связано с основами ʃai̯ 'течь, протекать (о сосуде)' и ʃai̯ 'лить, орошать' посредством дополнительной «скрытой» суффиксации, вызывающей тональные чередования).

Что касается корней (б) и (в), то они отличаются смешанной дистрибуцией, и без подробных данных об их семантике и сочетаемости в языках калахари-кхой однозначно определить их статус в працентр.-кхой. довольно затруднительно. Оба имеют параллели в нам: к корню (б) ср. нам /ʔō 'быть бесплодной (о корове)', прилаг. /ʔō-sá (значение 'бесплодная', наряду с исходным значением 'сухой', отмечено и в ряде языков калахари-кхой); к корню (в) ср. нам //xóó 'суровый, опасный; интенсивный (в т. ч. о засухе)', кор //xǒ 'злой, неприятный'.

Наиболее репрезентативна дистрибуция для працентр.-кхой. *ʔo (в основном значении 'сухой' этот корень присутствует и в кхойкхой, и в зап.-кхой, и в вост.-кхой. языках), причем развитие значения 'сухой' → 'бесплодная' вряд ли может быть обратным. Працентр.-кхой. *//xó, напротив, имеет значение 'сухой' только в калахари-кхой, и здесь как раз можно предполагать метафорическое развитие 'суровый, тяжелый' → 'сухой'. Не исключено, что семантика 'сухости' в какой-то степени присутствовала у слова *//xó уже на працентр.-кхой. уровне (как одно из переносных значений), но на уровень внешнего сравнения следует в первую очередь выносить все же этимон *ʔo.

11) «ухо»: (а) нам ʃàé-s, ани, кхо, анд, нар. хаб, тсх ʃé, буг, гви, ган ʃé, хай, дет, чар, дан, чуа ʃé, куа, цуа kʷé, хие ʃe; (б) кор //ai̯-b.

Подавляющее большинство языков сохраняет працентр.-кхой. корень *ʃé (с закономерным развитием *ʃ- → *ʃ- в правост.-кхой.). Только в кор он оказывается заменен на именное производное от глагольного корня *//ai̯ 'слышать' (см. ниже).

12) «есть»: нам ʃi̯i̯, кор ʃi̯i̯, ани, кхо, анд, нар. ган, хаб ʃi̯i̯, буг, тсх ʃi̯i̯, гви !ópi [Ta.], хай, дет, чар, дан, куа, цуа ʔui̯, хие jo.

В отличие от сев.- и юж.-кой. языков, в центр.-кой. семье последовательно проводится семантическое различие между понятиями 'есть (мягкую пищу)' и 'есть (жесткую пищу = мясо)'; второе из них в працентр.-кхой. обозначалось тем же корнем, что и имя 'мясо' (т. е. **kó*, см. ниже). В соответствии с принципом семантической детализации для лексикостатистики как более базисное отбирается понятие 'есть мягкую пищу', которое во всех центр.-кой. языках восходит к общему этимону **ʃí*. Начальный мягкий носовой в хие *no*, по-видимому, отражает в передаче Дорнана наализацию гласного.

13) «яйцо»: (а) нам *!n̄iwí-s*, кор *!n̄ibú-b*; (б) ани, кхо, бут, нар, гви, ган, хаб *ʃúbí*, дет *ʔyùbí*, чар *ʔyíbí*, куа, цуа *ʔíbí*, хие *íbí*.

Пракхойкхой **!n̄ibu* и пракалахари-кхой **ʃubi*, несмотря на очевидное фонетическое сходство, не могут быть сведены друг к другу, т. к. соответствие «кхойкхой ! : калахари-кхой ʃ» не является регулярным. В работе [Starostin 2008a: 435] приводится еще несколько возможных примеров на такое соответствие (напр., нам *!n̄ini-b* 'локоть' = калахари-кхой **ʃh̄uni* id. и т. п.), но исчисляются они буквально единицами, и на текущий момент предпочтительнее считать эти формы отдельными корнями — несмотря на то, что они дополнительно распределены между подгруппами.

14) «глаз»: (а) нам *t̄íí-s*, кор *t̄í-b*; (б) ани, бут, анд, ган, хаб *ʃxái*, кхо, нар, гви *ʃxéí*, хай *ʃáí*, дет, куа, цуа *ʃxàí*, чар, дан, чуа *ʃxái*, хие *ʃáí*.

Пракалахари-кхой корень **ʃxái* противопоставлен здесь пракхойкхой корню **t̄í*. Инновацию следует постулировать для кхойкхой уровня, т. к. **t̄í* даже синхронно можно рассматривать как именное производное от общецентр.-кой. глагольного корня **t̄í* 'видеть', одинаково хорошо сохраняющегося и в кхойкхой, и в калахари-кхойских языках.

Працентр.-кхой. **ʃxái* при этом, возможно, отражается в кхойкхой как нам *ʃxái*, кор *ʃxai* 'просыпаться' (← 'открывать глаза'), что опять-таки подчеркивает большую семантическую архаичность значения 'глаз' в калахари-кхой. (Аналогичные примеры конверсии по принципу «часть тела» → «пребывать в состоянии, имеющем отношение к части тела» есть и в других центр.-кой. языках: ср. кхо *ʃáó* 'сердце' → *ʃáó* 'быть счастливым, веселиться').

15) «огонь»: (а) нам *ʔáé-s*, кор *ʔae-b*, ани, бут, чар, тсх, дан *ʔé*, кхо, анд, нар, гви, хаб, куа, цуа *ʔé*, ган *ʔé*, хие *ʔe*; (б) дет *ʔí*.

Все формы группы (а) регулярно отражают працентр.-кой. **ʔe* (с закономерной дифтонгизацией монофтонга в кхойкхой). В *дет* имела место локальная инновация (конверсия из глагольной основы **ʔu* 'разводить огонь' → *нар ʔú*, *кхо ʔúú* id.).

16) «нога»: (а) *нам ʔáí-b/s*, *кор ʔai-b*; (б) *ани ʔnàré*, *кхо káré*, *буд, цуа kàré*, *анд, куа kári*, *гви ngare* [Та.], *ган ng!àre* [Та.], *чуа kàré*; (в) *буд, хай, дет, тсх, дан ʒí*, *чар ʒí*.

Пракхойкхой **ʔai* хорошо сопоставимо с такими формами, как *кхо, нар ʔí* 'пинать', 'лягать'; это повышает вероятность инновативной семантики в кхойкхой, т. к. деривация 'лягать' → 'нога' (букв. 'лягалка') с общетипологической точки зрения более естественна, чем обратная.

Ситуация с двумя другими корнями более сложная. Первый имеет широкую дистрибуцию и в зап.-кхой., и в вост.-кхой., хотя соответствия не являются полностью регулярными: назализация кликса, обнаруживаемая в *ани* и, возможно, также в подгруппе *ʔви-ʔ*ана, неожиданно исчезает в других языках. В кхойкхой родственными формами, скорее всего, следует считать *нам ʔàrì* 'ехать (на транспорте); вести (повозку, машину и т. п.)', *кор ʔári-s* 'колесо', при условии сужения из первоначально более широкого значения 'передвигаться' (вообще). В этом случае калахари-кхой **ʔare* также можно рассматривать как инновацию, тем более, что глагольные значения в этой подгруппе также представлены: ср. в *кхо* по данным Килиан-Хац — *kʷááré* 'нога', 'колесо', но *kʷáré* 'возвращаться'. Глагольная семантика «передвижения», таким образом, имеет более широкую дистрибуцию, чем именная семантика (семантическое развитие 'нога' → 'колесо' в языках Африки, правда, встречается нередко).

Что касается второго корня, то, помимо перечисленных выше случаев, он встречается также в *нам (síi-b* 'коготь страуса', 'большой палец (на ноге)'); в *ани (zê* 'палец (ноги)', с не ясной утратой назализации), и в *кхо (ʒí* 'лапа (птицы)', 'коготь'). Учитывая скромную дистрибуцию центр.-кой. **ʒí* (в значении 'нога' — только в языках шуа и, возможно, в результате ареального развития, в буга), а также естественность семантического развития 'коготь, палец ноги' → 'нога' (ср. русск. *нога* и *ноготь*), здесь также можно было бы предполагать инновацию. С другой стороны, обращает на себя внимание тот факт, что из трех разобранных основ только

*zĩ обладает исключительно именной семантикой, а сужение значения 'нога (вообще)' → 'нога (птичья)' также вполне допустимо.

В конечном итоге решающих аргументов, которые позволили бы однозначно исключить из рассмотрения хотя бы одного из трех «кандидатов», представить нельзя. Ранжировать основы при привлечении их к внешнему сравнению можно исключительно по масштабу их дистрибуции: (а) *l̥are, (б) *zĩ, (в) *ɸ(a)i.

17) «волосы»: нам /ɸiː-b, кор /ɸiː-b, ани, буг, анд, нар, гви, ган, хаб, хай, дет, дан, чуа, куа, цуа /ɸiː, кхо, чар /ɸiː, тсх /ɸiː, хие /ʰo.

Все формы закономерно отражают працентр.-кой. *ɸi (в хие, скорее всего, некорректно затранскрибирован исход кликса).

18) «рука»: (а) нам /ʔóm-mi¹; (б) кор /ùm-ta; (в) ани, буг, чар, дан cʰai, кхо ʂèi, анд, нар, хай, дет, тсх, чуа, куа, цуа sai, гви, ган sai, хаб sai.

Обще-калахари-кхой корень однозначно восстанавливается в виде *cʰai; явных параллелей в языках кхойкхой не обнаружено. Что касается форм (а) и (б) в нам и кор, то, несмотря на фонетическое сходство, они несводимы друг к другу, поскольку различаются исходом кликса. То, что они восходят к разным кхойкхой и общецентр.-кой. корням, доказывается наличием альтернативных этимологий: к нам /ʔóm-i ср. скорее кор /ɸit- 'кулак', а кор /ùm-ta, хотя и не имеет соответствий в нам, сопоставимо с пракалахари-кхой *l̥ota ~ *l̥uta 'рука (от локтя до запястья)' → ани /bòtà, буг gòtà, дет, чуа gítà и др.

Однозначно выбрать вариант (а), (б) или (в) на роль основной працентр.-кой. 'руки' не удастся, но, учитывая «разнобой» в языках кхойкхой, наиболее вероятно утрата исходного працентр.-кой. корня *cʰai в кхойкхой с заменой его на близкие по семантике, но разного происхождения корни в нам и кор.

19) «голова»: (а) нам dàná-s/b; (б) кор /ʔā-b; (в) ани, кхо, буг, анд, нар ɸi, тсх ɸi ~ ɸi; (г) гви mâ [Ta.], m-ʔm [V.], ган, дет, чар má, хаб m-ʔá, хай, чуа, куа mâ, цуа m-ʔà, дан tà, хие hta.

Разнообразие перечисленных корней трудно свести к общему сценарию. Очевидной инновацией можно считать лишь кор /ʔā-b:

¹ Второй -m- в этой и других подобных формах — не самостоятельная часть суффикса, а результат морфонологической ассимиляции на стыке с показателем среднего рода ʔi: /om-ʔi → /om-mi.

это именное производное от пракхойкхой глагольной основы **!ʔā* 'чувствовать', 'слышать' (т. е. = '(то, что) чувствует, воспринимает').

Оставшиеся три корня оказываются дополнительно распределены по подгруппам: **dana-* в кхойкхой, **fú* в подгруппах кхве и наро, **ta* в подгруппе г|ви-г|ана и вост.-кхой. языках. (Исключением является дублет *fú* ~ *ǰú* в вост.-кхой. языке тсх, но его появление здесь, скорее всего, связано с сильным ареальным влиянием зап.-кхой. наречий, см. выше). В нескольких языках представлена также сложная основа, где первый компонент — скорее всего, редуцированный вариант корня **ta*, а второй допустимо сравнивать все с той же основой **!ʔā* ~ **!ʔa*, которая стала главным носителем значения 'голова' в кор.

Кхой / наро **fú* правомерно отождествлять с первым слогом в нам *fúú-rǝ* 'начальный', 'первый'; если эта этимология верна, она предполагает несомненное существование по меньшей мере в «ранне-практикхой» именного корня **fu-* 'голова' и, соответственно, семантически инновативный статус как у нам *dānǎ-s/b*, так и у вост.-кхой **ta*. Однако этот аргумент нельзя считать решающим: во-первых, не исключены и другие семантические сценарии (например, редкий, но допустимый переход **fu-* 'начало' → 'голова'), во-вторых, пока что отсутствуют убедительные внешние этимологии как для **dana-*, так и для **ta*. К внешнему сравнению, таким образом, допустимо привлекать все три основы, ранжируя их в следующем порядке: (в), (г), (а).

20) «слышать»: (а) нам *ǰǰú*, кор *ǰǰū*; (б) ани, кхо, буг, анд, дет, чар *kóŋ*, нар, ган, хаб, тсх, дан, куа *kúŋ*, гви *kiáŋ*, цуа *sóŋ*, хие *ǰot*.

Налицо распределение между пракхойкхой **ǰǰū* и пракалахари-кхой **kúŋ* ~ **kóŋ* (качество гласного однозначно не определяется). Ключевым моментом здесь оказывается то, что в словаре Вураса для кор, помимо ожидаемого *ǰǰū* 'слышать', приводятся также следующие слова: *kat-sin* 'hear (attentively)', *kum-sin* 'listen (attentively)'. Гласный *-a-* в *kat-sin* неясен (возможно, речь идет о каком-то ошибочном разграничении), однако *kum-sin* — несомненный архаизм, не сохранившийся в нам, но явно указывающий на архаичность семантики пракалахари-кхой формы **kúŋ*.

21) «сердце»: нам *fǎó-b*, кор *fǎó-b*, ани, кхо, анд, нар, гви, ган, хаб, тсх *fǎó*, буг *fǎó*, хай *só*; дет *sǎó*, чар, дан *sǎó*, чуа *só*, куа *kʷó*, цуа *kʷó*, хие *ǰo*.

Все формы закономерно отражают працентр.-кхой. **fáb* без существенных отклонений (за исключением отдельных неясностей с тонами).

22) «рог»: нам *ʃáá-b*, ани, кхо, буг, анд, нар, гви, ган, хаб, хай, дет, чар, тсх, дан, кya, чуа *ʃná*, хие *ɲga*.

Формы, по-видимому, отражают працентр.-кхой. **ʃá*; Р. Фоссен восстанавливает **ʃna* с носовым гласным, но в этой ситуации убедительнее выглядит сценарий диссимиляции в языках калахари-кхой (**ʃá* → **ʃa*), нежели ассимиляции в кхойкхой, т. к. фонотактическое сочетание «кликс с носовым исходом + носовой гласный» в калахари-кхой встречается крайне редко, в отличие от нам (в кор слово 'рог' не зафиксировано).

23) «я»: нам *tà* (эмф. *tíí-tà*, *tíí-r*), кор *ti-re* (м.), *ti-ta* (ж.), ани, кхо, буг, анд, хаб, тсх *tí*, нар *tí*, *tí-rá*, гви *tí-rè*, ган *tê*, хай, дет, чар, дан *tá*, чуа *tjí-re* (диал. *tyé*), кya *kʷé ~ kʷé-dì*, чуа *kʷé ~ kʷé-dè*, хие *čí*.

Фонетическое сходство всех перечисленных форм очевидно, однако разнообразие вариантов требует комментария. Прозрачнее всего ситуация устроена в кор, где в роли «чистых» местоименных морфем выступают **-re* для мужского рода и **-ta* для женского. Именно такой вид они имеют как объектные показатели при глаголе (*-r* и *-ta* соответственно); полные формы независимых местоимений расширены за счет префиксации, по-видимому, эмфатической морфемы *tí-*. Почти такая же ситуация наблюдается и в нам, где, однако, оппозиция по роду для местоимения 1-го л. ед. ч. не отмечена (полные формы *tíí-tà* и *tíí-r* отмечены как свободные варианты). Сразу заметим, что, в отличие от местоимения 2-го л. (см. ниже), где оппозиция по роду зафиксирована во всех центр.-кой. языках без исключения, для местоимения 'я' она наблюдается только в кор, что похоже на инновацию; с другой стороны, никакого другого объяснения «вариации» **de* ~ **ta* предложить также невозможно.

Если предполагать архаичность ситуации в кор (исходные формы: **de* 'я (м. р.)', **ta* 'я (ж. р.)', **ti-re*, **ti-ta* 'я (эмф.)'), то разнообразие развитий в калахари-кхойских языках объясняется довольно легко. В одних языках собственно местоименное значение переносится на эмфатический префикс, откуда возникают «усеченные» формы *tí*, *tê* (в подгруппе чуа они подвергаются дальнейшей регулярной палатализации: *tʷé*, *kʷé*, *čí*). В других, на-

оборот, из употребления выходит эмфатический вариант — так, в языках шуа обобщается простой вариант **ta* (выбор именно его вместо варианта м. р. **de*, возможно, мотивирован фонетической аналогией с местоимениями 2-го л. **ca*, **sa*).

В нар мы наблюдаем, по сути, инверсию исходной ситуации: здесь вариант *tí* (← старый эмфатический префикс) становится немаркированным 'я', в то время как «полный» вариант *tí-rá* (где *-rá* ← **de* 'я' /м. р./), наоборот, используется в усилительных контекстах.

Такой анализ в целом согласуется и с выводами Р. Фоссена [Vossen 1997a: 367-370], который восстанавливает для працентр.-кой. эмфатический местоименный префикс **ti* и местоимение 1-го л. ед. ч. **ta* (правда, удовлетворительное объяснение элемента **-de* в его реконструкции отсутствует). Отметим, что «эмфатический местоименный префикс» можно интерпретировать на працентр.-кой. уровне и просто как слегка модифицированную редупликацию исходного местоимения (см. ниже аналогичную ситуацию с местоимением 2-го л.).

24) «убивать»: (а) нам *!āń*, кор *!ām*; (б) ани, дан */kũ*, кхо, буг, нар, гви, ган, хаб */kũ*, анд */kũ*, хай, чар */ũ*, тсх */ũ*, куа, цуа */ũ*, хие */o*; (в) дет *kʰāé*.

Для пра-кхойкхой восстанавливается форма **!am*, для пра-калахари-кхой — **/kũ*. В дет локальная инновация: *kʰāé* ← пр.-вост.-кхой **kʰae* 'втыкать, пронзать' ← працентр.-кой. **!ʰae* id. (ср. нар *!ʰāé* id. и др.).

Вопрос о первичности того или иного корня в исходном значении 'убивать' остается открытым. Отметим, что в нам регулярным соответствием для калахари-кхой **/kũ* является этимон */ũ* 'прекращать, останавливать [что-л.]' (с закономерным развитием велярной аффрикаты в гортанную смычку). В случае весьма вероятного этимологического тождества этих корней семантика 'убивать' в калахари-кхой, скорее всего, вторична (эвфемистическое развитие 'прекращать' → 'убивать' типологически более частотно, чем обратное).

25) «лист»: (а) нам *!āẽ-s/b*; (б) кор *!āũ-b* [Wu.]; (в) ани *!ā*, кхо, дет *!ā*, анд *!ā*, ган *!ānā*, хай *!ānā*, чар, тсх, дан *!ānā*, хие *!ānā*; (г) нар *tɔ̀r̀r̀* [Vi.]; (д) нар, хаб *dānā* [V.], гви, ган *dana* [Ta.]; (е) ган */xéũ*, буг */kxéũ*.

Языки кхойкхой демонстрируют здесь типично «африканское» метафорическое развитие 'ухо' → 'лист' (относительно этимологии форм см. выше 'ухо'). На этом фоне явно более архаичным выглядит обще калахари-кхойский корень **lana*, стягивающийся в моносиллабическое *lã* в ани, дет и анд; он же, по-видимому, представлен и в кхойкхой, но в значении 'трава': нам *lãã-b*, кор *lã-b* (для калахари-кхой семантическая оппозиция 'лист' : 'трава' коррелирует с лексической оппозицией **lana* : **doa*).

Не вполне ясными остаются формы (г) и (д) в нар, гви и ган. Нар *tòràrà* (Барнард приводит варианты *tò:àrà* ~ *dò:àrà*, наряду с *lana*, не отмечая каких-либо различий в значении) явно тождественно жу|'хоан *dòràrà* (см. 'лист' в сев.-кой. языках); поскольку дистрибуция этой формы как в центр.-кхой., так и в сев.-кой. группах крайне ограничена, совпадение следует считать ареальной изоглоссой — но направление заимствования при этом установить затруднительно. Что касается гви и ган *dana*, то их сходство с альтернативной формой *dãná* в нар бросается в глаза (особенно если учесть более аккуратную транскрипцию в [Nakagawa 1998: 249] — гви *dãna*, *zãna*, ган *dãna*, также с фарингализацией первого гласного), но остается неясным, как *dãná* Фоссена соотносится с *tòràrà* Фиссера — списывать это соотношение на диалектное разнообразие трудно, т. к. свободные колебания *r* / *n*, а также начальной глухости / звонкости для диалектов нар вроде бы не свойственны. Все это наводит на мысль о каком-то ареально-субстратном лексическом слое, не поддающемся четкой идентификации, из которого в разные периоды могли быть заимствованы все соответствующие формы. В любом случае к дальнейшему внешнему сравнению привлекать их рискованно.

Исключен вывод на внешний уровень сравнения форм группы (е), известных лишь по спискам, опубликованным в [Vossen 1986]: судя по всему, это ошибочно записанные в значении 'лист' варианты обще-|ани-кхой прилагательного 'зеленый' (ср. кхо *kxáó* 'зеленый' в надежном источнике [Killian-Hatz 2003: 173]).

26) «вошь»: (а) нам *úri-b/s*, кор *kúri-b*, ани *kúni*, кхо, буг, гви *kúni*, ган *kúni*, хай, дет, чар, куа, цуа *kúni*, тсх *kúni*; (б) нар *ḥã* [Vi.]; (в) хие *lat*.

Подавляющее большинство языков отражает единый центр.-кхой. корень **kuri* ~ **kuni*. Причины вариации *-r-* (кхойкхой) : *-n-*

(калахари-кхой) остаются неясными, но она наблюдается и в ряде других случаев (ср., например, пра-кхойкхой **!xaru* 'храпеть' = пракалахари-кхой **!xuni*).

Изолированная форма $\hat{f}\hat{a}$ в нар не имеет каких-либо соответствий в других центр.-кой. языках, однако почти идеально совпадает с формой $\hat{f}\hat{a}$, которая дается в качестве синонима обычного слова *с'и* в жу|'хоан в словаре Дикенса. Налицо такая же ситуация, как и в случае со словом «лист»: ареальная «наро-жу|'хоан» изоглосса, не поддерживаемая другими койсанскими языками — ни северными, ни центральными. Вероятно, речь опять идет о каком-то субстратном элементе.

Хие $\parallel at$ не имеет никаких параллелей за пределами списка Дорнана; все «ближайшие родственники» этого диалекта отражают регулярную працентр.-кой. основу **kuni*.

27) «мясо»: (а) нам $\parallel \hat{a}i-i$; (б) кор *kxo-b*, ани *kxò-xú*, дан *kxò-xú*, кхо *kxó-xò*, бу, анд, хаб *kxó-xò*, нар *kxò-xó*, хай *kò-hú*, дет, тсх *kò-xú*, чар *kò-hú*, куа, цуа *kó-xì*, хие *ko-ho*; (в) гви $\parallel xá$, ган $\parallel xâ$.

Общесред.-кой. корень (б) **kxo* уже на прауровне употреблялся и как глагольный ('есть мясо', 'есть жесткую пищу'), и как именной ('мясо'), хотя видно, что в большинстве языков в именном значении представлена сложная основа **kxo-xi*, где **-xi* — номинализирующий суффикс, восходящий к общесред.-кой. имени **xi* 'вещь'. Это позволяет предположить, что первичным для корня **kxo* было именно глагольное значение, и, следовательно, если на раннем этапе працентр.-кой. языка и существовал отдельный именной корень 'мясо', то он был утрачен (или утратил это значение) еще до начала распада группы.

Не исключено, впрочем, что архаизм здесь сохраняют языки ветви $\parallel vi-g \parallel ana$, где в значении 'мясо' выступает общесред.-кой. именной корень **xa* 'тело; плоть' [Vossen 1997: 468]. Семантическое различие между **xa* и **kxo* в некоторых языках может быть очень тонким: ср., например, в нар в словаре Г. Фиссера — *kxò* 'мясо с костями (о животных)', но $\parallel xââ$ 'мясо без костей (о животных), плоть (человеческая)' (отметим, что форма *kxò* у Фиссера может быть фонетическим стяжением из **kxò-xò*). Если такое семантическое разграничение в дальнейшем подтвердится как архаичное на подробном материале и других языков, возможно, оба слова нужно будет учитывать как синонимы.

В нам корень **kxo* до сих пор сохраняется в глагольном употреблении: *ðǝ* (← **kxo*) 'есть (о животных); жрать (*зруб.*)), но в именной функции он заместился на форму *ʃǝi-i*, возможно, изначально 'мышца', 'мускул', 'кусочек мяса' (такие значения зафиксированы в словаре нам наряду с 'мясом'), хотя явных параллелей к нему в других центр.-кой. языках не обнаружено. То, что эта ситуация инновативна, а не архаична, видно хотя бы по тому, что в близкородственном кхойкхой языке кор старый корень сохраняется также и в именном значении.

28) «луна»: (а) нам *ʃxã-b*, кор *ʃxã-s*; (б) ани *ʃnóé*, кхо *ʃóé*, буг, тсх *ʃóé*, анд, нар, дет, чар, куа *ʃóé*, гви, ган *ʃòè*, хаб, хай *ʃnòè*, цуа *ʃúé*, хие *ʃwe*.

Пра-кхойкхой **ʃxã* и пракалахари-кхой **ʃóe* находятся в дополнительной дистрибуции по ветвям; определить, какой из этих корней более архаичен, не прибегая к внешнему сравнению, невозможно. Дентальный кликс в хие — след ошибочной транскрипции или опечатки (см. выше аналогичную ситуацию с 'ногтем').

29) «рот»: нам *ǎt̪-s*, кор *amt̪-a* [Wu.] (?), ани, кхо, буг, анд, нар, гви, ган, хаб *kǎt̪*, хай, чар, тсх, чуа *kǎt̪*, дет, куа *kǎt̪*, цуа *kǎt̪*, хие *ʃat̪*.

Все формы отражают общецентр.-кой. этимон **kam* (палатальный кликс в хие, скорее всего, отражает неправильное восприятие глоттализованной велярной аффрикаты, ср. выше 'пить').

Неясной остается ситуация в кор. В словаре Майнгофа слово 'рот' почему-то не приводится вообще, а регулярно соответствующей перечисленным форме *kam* приписывается значение 'ворота': очевидно, метафорический перенос, но при этом неясно, могла ли она иметь в кор также и значение 'рот'. В словаре Вураса в значении 'рот' приводится форма *amma*, которая, скорее всего, связана с **kam* — но при этом никак не может соответствовать працентр.-кой. корню регулярно, т. к. велярная аффриката **kx* в кор должна сохраняться (см. выше 'пить' и другие корни). Естественно было бы трактовать вариант *amma* как заимствование из нам, где выпадение **kx* регулярно, но это также странно, т. к. других случаев явных заимствований в кор из нам не отмечено.

30) «имя»: нам *ʃòt̪-s*, кор *ʃkòt̪-á*, ани, ган *ʃkòt̪*, кхо *ʃkòt̪*, буг, анд, нар *ʃkúit̪*, гви *ʃkòit̪*, хаб, дан *ʃkúit̪*, хай, чар *ʃòt̪*, дет *ʃit̪*, тсх *ʃòt̪*, куа, цуа *ʃit̪*, хие *ʃun ~ ʃun*.

Все формы явно отражают один и тот же працентр.-кой. корень, хотя реконструкция вокализма неоднозначна: */kxon ~ */kxiŋ. С точки зрения дистрибуции более перспективна реконструкция с исходным вокализмом */o, отмеченным в обеих основных ветвях группы, но регулярность соответствий еще предстоит уточнить. В нам, помимо этого, отмечена также любопытная «переогласованная» диалектная форма /ɾèŋ-s; возможно, ее стоит сравнить с вариантом /kxiŋ в буг и нар ← */kxon-i (где -i — старый показатель рода, слившийся с корнем), но рекуррентность такой корреляции пока не продемонстрирована.

31) «новый»: (а) нам /ɾá-sà, кор /ɾà-sà; (б) ани kxóà, кхо kxóa, дет, чар kʔóá, тсх kʔóá; (в) нар kaba, гви, ган qábà; (г) цуа /áà; (д) хие /ao.

Крайне неустойчивый этимон; однозначная реконструкция общецентр.-кой. корня не представляется возможной.

В языках кхойкхой здесь налицо продуктивное адъективное образование от корня */ɾa, который соблазнительно отождествить с нам /ɾáà 'острый', '(свеже)заточенный'; последнее, однако, отражает общецентр.-кхой. */kxa (ср. кор /kxá 'острый'), так что либо речь идет все же о двух разных корнях, либо кор /ɾà-sà следует считать относительно недавним заимствованием из нам (см. дискуссию вокруг корня 'рот').

Корень (б) отражает общецентр.-кой. основу *kxoa. В языках калахари-кхой она бывает полисемична: ср. в кхо значения 'сырой', 'неспелый', 'зеленый'; в нар она обнаружена с суффиксальным расширением как kxóà-rà 'сырой' (о мясе) [Vi.]. В плане фонетики в языках кхойкхой ей идеально соответствует глагольная основа 'возвращаться': нам óà, кор kxóá. Если это не случайное совпадение, такая семантика — сильный аргумент в пользу исходности значения 'новый' уже на працентр.-кой. уровне: в противном случае для кхо и нар нужно реконструировать двойной семантический переход 'возвращаться', 'делать снова' → 'новый' → 'сырой', что, учитывая относительную «молодость» группы, не очень вероятно.

Корень (в), специфичный для подгруппы наро-г|ана, обнаружен и в нам (kàwá 'новый'), но только в диалектах Сесфонтейн и Хай||ом, где она, в принципе, могла оказаться заимствованной из нар. Обращает на себя внимание идеальное фонетическое соответствие форме qává 'треснуть, расколоться' (о дереве) в кхо [KH],

но семантическая корреляция спекулятивна ('трещина' как 'новообразование?').

Еще более странной оказывается ситуация в чуа: для формы ʃáà Р. Фоссен действительно отмечает значение 'новый', но сама форма при этом полностью соответствует общецентр.-кой. $*\text{ʃáà}$ 'старый' (!). Наконец, хие $!ao$ — совсем необычная форма, т. к. старый альвеолярный кликс в языках чуа выпадает уже на прауровне и обнаруживается буквально в единичных случаях, которые обычно интерпретируются как заимствования. В материалах Фоссена можно найти такие формы, как куа, чуа $!áà$, но только в значении 'длинный' (семантически оно никак не совместимо с 'новым') и с подозрением как минимум на «обратную реструктуризацию» из $*kád$ под влиянием языков других подгрупп. Скорее можно было бы думать о том, что хие $!ao = \text{чуа } \text{ʃáà}$, т. к. в отдельных случаях Дорнан ошибочно записывает альвеолярный кликс вместо латерального (ср. хие $!hoe$ 'спать' = куа ʃóé 'лежать', хие $!hao$ 'дротик' = дан ʃxáá id.), но в этом случае неясно также, куда «пропал» в записи Дорнана носовой исход. На данном этапе, по-видимому, стоит все же разделять эти формы.

Таким образом, «фундаментальная» оппозиция выстраивается в первую очередь между кхойкхой $*/a-$ и калахари-кхой $*khoa$, причем $*khoa$ чуть более вероятен как основной носитель значения 'новый' в працентр.-кой. на основании своих семантических связей, но в целом «вероятностный зазор», скорее всего, пренебрежим ввиду общей неустойчивости 'нового' по подгруппам и отдельным языкам.

32) «ночь»: (а) нам $síy-xí-b$, кор $t^hí-xu-b$, ани, кхо, буг, анд, чар, тсх, дан $t^hí$; (б) нар $fñí$; (в) гви, ган $hxae-shika$ [Та.], хие $haie$.

Корень (а) $*t^hi$, представленный как в языках кхойкхой (где он расширен за счет именного суффикса $*xi \leftarrow$ общецентр.-кой. $*xi$ 'вещь'), так и калахари-кхой, имеет во всех языках, где он встречается, исключительно значение 'ночь' и, таким образом, должен считаться первичным. Развитие $*t^h- \rightarrow \text{нам } c-$ регулярно (см. выше 'пепел' и др.).

В нар имела место лексическая замена на 'черный' (см. выше). Формы группы (в) отражают общецентр.-кой. основу $*!q^hae$ 'темный, темнота' = нам $!xàè$, ган $qáè$ id. (нотация $hx-$ в записи Танака, по-видимому, отражает увулярную артикуляцию) и, скорее

всего, вторичны в значении 'ночь' (ср. аналогичное развитие и в диалектной форме кор !xai-b 'ночь', которую Майнгоф отмечает для идиолекта одного из своих информантов наряду с более привычным t^hũ-xi-b).

33) «нос»: нам fũi-s, кор fũi-b, ани, кхо, бут, анд, гви, тсх fũi, нар, хаб fũi, ган fũi, хай, дан, чуа, куа, цуа fũi, дет, чар fũi, хие çui.

Все формы согласованно отражают общецентр.-кой. *fũi. Причины озвончения кликса в ган по версии Фоссена непонятны; в словаре Танака для ган и гви записана одна и та же форма !tui (= fũi в нашей транслитерации, с глухим исходом).

34) «не»: (а) нам tãmã, кор tama, нар -tã, -tãmã, ган tãmã, хай, дет, чар, дан, куа, цуа -tã, тсх -tã; (б) ани, кхо, бут, анд -bé.

Р. Фоссен [Vossen 1997a: 366] восстанавливает праформу для (а) в виде *tama, предполагая нерегулярное усечение → ta в нар и вост.-кхой языках. Поскольку морфема эта чаще всего выступает в качестве приглагольной энклитики, такое развитие представляется вполне вероятным, тем более, что промежуточная форма tã с назализацией гласного, вызванной падением носового m, эксплицитно зафиксирована по меньшей мере в одном языке (тсх). Обратное решение — выбор в качестве первичного варианта *ta — не объясняет ни назализации в тсх, ни функции и происхождения элемента -ta в остальных языках.

Дистрибуционный анализ показывает, что в подгруппе кхве имела место инновация, в ходе которой старая морфема *tama оказалась полностью вытеснена энклитикой *-bé. Происхождение последней неясно; помимо данной подгруппы, она встречается только в дан, где, в отличие от морфемы -tã, сочетающейся с формами настоящего времени, сочетается с формами будущего и прошедшего времен [Vossen 1997a: 227].

Поскольку для большинства центр.-кой. языков пока что отсутствуют подробные грамматические описания, не исключено, что ситуации, аналогичные зафиксированной Фоссеном для дан, будут обнаружены и в каких-то других языках. В общетипологическом плане стоит отметить, что супплетивные основы отрицательных морфем, распределенные в зависимости от видо-временных характеристик глагольных форм — явление, абсолютно нормальное для самых разных языковых групп Африки, но при этом нехарактерное ни для северно-, ни для южнокойсанских

языков; вне зависимости от того, удастся ли оппозицию **tama* / **be* убедительно вынести на працентр.-кой. уровень, ее присутствие даже в одном-двух языках этой группы свидетельствует о «некой-санских» связях или влияниях (либо ареального плана, либо генетических).

К внешнему сравнению, разумеется, необходимо привлекать в первую очередь морфему **tama*, но, ввиду отсутствия внутренней этимологии **be* и возможности постулировать для обеих хотя бы слабо аргументированную ситуацию праязыкового супплетивизма, исключить вторую морфему при выходе на внешний уровень также не следует.

35) «один»: нам /*íí*, кор /*í*, ани, кхо, буг, анд, нар, ган, гви, хаб, хай, дет, чар, тсх, дан, куа, цуа /*íí*.

Все формы регулярно отражают працентр.-кой. **íí*. Стоит обратить внимание на фонетико-семантическое сходство с працентр.-кой. **í* 'другой' → кхо, буг, хаб, чар, тсх, дан /*í*; в ани и в гви оба значения выражаются одной и той же формой /*íí*, что можно расценивать либо как контаминацию, либо как след этимологического тождества (**íí* 'один' ← **í* 'другой; один из двух' + суффикс **-i*).

36) «дождь»: (а) нам *tíí-s*, кор *tí-s*, ани, кхо, анд, нар, гви, хаб, хай, тсх, чар, дан, куа, цуа *tí*, буг, дет *tí*, ган *kʰi*: [Ta.], хие *tu*; (б) нам /*àrí-b*.

Общецентр.-кой. форма **tí*; в большинстве языков корень выступает одновременно как именной ('дождь') и как глагольный ('идти (о дожде)').

В современных диалектах нам, судя по словарной информации, идет процесс постепенного вытеснения старой основы: форма *tíí-s* еще употребляется в значении 'дождь' в диалектах дамара, топнаар и др., в то время как в собственно нама главным эквивалентом этого значения является форма /*àrí-b* — именное образование от глагола /*àwí* ~ /*àrí* 'идти (о дожде)', в свою очередь, производного от простой основы /*àíí* 'течь, протекать (о ручье, источнике)'. В ряде диалектов встречаются и другие лексические эквиваленты, например, /*àní-b* (по-видимому, семантический сдвиг из старого значения 'ливень, буря'); подробная картина приводится в [Naascke et al. 1997: 184].

37) «дым»: нам /ʔáŋ-ni/s, кор /kxan-ni ~ /kxān-na, ани, кхо, буг, анд *sáni*, нар, ган, хаб, тсх, дан, куа, цуа *síni*, гви *séŋè*, хай, дет, чар *sáni*, хие *seŋe*.

Все формы восходят к общецентр.-кой. **cani* (так в реконструкции Р. Фоссена; фонетически начальный сегмент корня, возможно, мог артикулироваться как сочетание **ckx-*). Развитие **c-* (**ckx-*) → пракхойкхой */*kx-* абсолютно регулярно; это один из немногих доказанных случаев откровенно вторичного появления щелчковой артикуляции из нещелчковой в сравнительно «молодых» койсанских подгруппах. Вокалические колебания в языках калахари-кхой связаны с ассимилятивными тенденциями, протекающими скорее на уровне отдельных говоров и идиолектов, чем языков (ср. в аналогичном контексте: */*lani* 'подбородок' → ани *gani*, нар /*lani*, но гви *gini*, ган *gini* id.); тем не менее, не исключена возможность реконструкции и иного вокалического ауслauta, напр. **cane* (ср. форму в гви, необъяснимую при реконструкции гласного верхнего подъема).

В любом случае именная основа **cani* — несомненное производное от глагольной основы **cān* 'дымиться', успешно сохраняющейся в нам (/ʔáŋ) и в кор (/kxān); в языках калахари-кхой глагольная и именное значения нейтрализовались в общей форме **cani*.

38) «звезда»: (а) нам /*lami-ró-s* ~ /*lami-ró-s*, кор /*lami-ró-s/b*; (б) ани, кхо, хай /*xáni*, буг, анд, дет, тсх, дан, куа /*xáni*, чар /*xini*, цуа /*xáni*, хие /*haine*; (в) нар /*ɬonò* [Vi.], гви /*ɬonù* [V.], /*onò* [Ta.], ган /*ɬonù* [V.], /*onò* [Ta.], хаб /*ɬonù*.

Наиболее архаичное впечатление производит группа форм (б), отражающая пракалахари-кхой */*xani*. В нам этот же корень сохраняется в производном метафорическом образовании /*xini-ni-s*, диал. /*xini-ŋ-s*, /*xini-ã-s* 'вид якорцевых (Tribulus), «утренняя звезда»'.

Вряд ли случайным является и совпадение перечисленных форм со словом 'цесарка' в зап.-кхой языках: кхо /*xáni*, нар /*xané*, ср. также нам /*xéni-s*, диал. /*xini-s* ~ /*xini-s* (имеется в виду сходство между звездным небом и мелкими белыми пятнышками, которые покрывают оперение цесарки). Однозначно установить направление метафоризации вряд ли возможно, но на первичность значения 'звезда' может, в качестве ареальной изоглоссы, намекать также любопытная форма в !хонг (юж.-кой.): /*xãna*: 'еж' (в комментарии Трэйлла указывается, что еж, согласно верованиям

бушменов !хонг, считается одушевленной формой упавшей звезды [Traill 1994a: 59]). В центр.-кой. языках значение 'еж' для этого корня не зафиксировано, но в принципе такая метафоризация могла появиться уже на юж.-кой. почве после заимствования слова из центр.-кой. источника. Если так, то и значение 'цесарка' в центр.-кой. языках, по аналогии, может быть вторичным.

В основном значении 'звезда' в языках кхойкхой слово оказалось вытеснено вполне прозрачным именным производным от глагола 'мерцать, мигать' (нам [ãtĩ]: *|amĩ-ro-, с ассимиляцией вокализма второго слога в кор (и, в отдельных диалектах нама, с переходом звонкого исхода кликса в назализованный под влиянием срединного носового сонанта).

Неоднозначна ситуация в подгруппе наро-г||ана. Вряд ли случайно сходство между нар [ɲnɔ] и прасев.-кой. *ɲĩ (см. 2.1, № 38), что снова ставит вопрос об ареальных связях между зап.-кхой и сев.-кой. языками (см. выше 'лист', 'вошь'); в данном случае, однако, направление заимствования будет, по всей вероятности, из сев.-кой. источника в пранаро-г||ана. В других центр.-кой. языках параллелей к этим формам не отмечено, так что на общецентр.-кой. статус они в любом случае не претендуют.

39) «камень»: (а) нам [ɲĩi-s/b, кор [ɲi-, дет [ɲĩi, чар [ɲiĩ; (б) ани [ɲnɔ́á, кхо [ɲnɔ́á [КН], буг, анд, ган, тсх, дан [ɲnɔ́á, нар, хаб [ɲnɔ́á, гви [ɲnɔ́á, куа [ɲnɔ́á, хие [ɲwa.

Формы (а) отражают общецентр.-кой. основу *|ɲi, (б) — общекалахари-кхой основу *||oa (фарингализация дифтонга в нар и хаб необъяснима и, возможно, архаична, но решение этого вопроса приходится пока отложить из-за неустановленности соответствий).

Р. Фоссен, исходя из дистрибуции обоих корней, считает исходным працентр.-кой. этимолом 'камень' основу *|ɲi, а основу *||oa объявляет инновацией в калахари-кхой [Vossen 1997a: 495]. Однако такое решение оставляет открытым вопрос о семантических различиях между основами на уровне пракалахари-кхой.

На самом деле ситуация несколько более сложная. С одной стороны, для основы *|ɲi в языках кхойкхой отчетливо зафиксирована полисемия 'камень' / 'гора' (типичная, но не обязательная для языков Африки). С другой стороны, для основы *||oa значение 'гора' не отмечено ни в одном языке из тех, для которых

существуют подробные лексикографические описания (кхо, нар, хие). При этом в кхойкхой потенциальным соответствием для этой основы (идеальным в плане фонетики) может быть нам ʃbá-s 'мышеловка (устроенная из плоского камня и палочки)', что опять же намекает на первичность значения 'камень'.

Таким образом, общая совокупность данных требует следующего разведения семантики этих двух основ на працентр.-кой. уровне: $*\text{ʃui}$ 'гора, холм' vs. $*\text{ʃoa}$ 'камень'. Независимое (типологически вполне естественное) развитие значения 'гора' → 'камень' имело место в пра-кхойкхой и в ряде диалектов подгруппы чуа; в остальных языках сохранилась старая ситуация.

40) «солнце»: (а) нам sórè-s , кор sōrē-b ; (б) ани, кхо, буг, анд, нар, ган, гви, хаб, хай, чар, тсх, дан, чуа, куа, цуа ʃám , хие ʃam .

Формы группы (а) отражают пракхойкхой $*\text{sore}$, группы (б) — пракалахари-кхой $*\text{ʃám}$. Степень архаичности семантики каждой из этих основ установить затруднительно, но отметим, что в нам регулярным соответствием пракалахари-кхой $*\text{ʃám}$ является глагольная основа ʃám 'нагреваться, становиться горячим' (с дальнейшим именным производным ʃám-mi 'тепло, жара'; ср. также диал. форму ʃám-b/s 'восток'). Типологически более частотным и естественным является развитие 'тепло, жар' → 'солнце', нежели обратное, так что в данном случае семантическая инновация более вероятна в языках калахари-кхой, чем в нам и кор.

(Косвенным аргументом в поддержку именно такого направления развития можно считать и форму $k^h\text{óbò}$, которую Р. Фоссен приводит в значении 'солнце' — по-видимому, как дополнительный синоним — для языков дет и чар [Vossen 1988: 99]; ср. эту же основу в других языках — ани $k^h\text{óbò}$ 'быть теплым, жарким', нар $k^h\text{óbò}$ 'потеть' и др.).

41) «хвост»: (а) нам ʃáré-b ; (б) кор sáó-b , ани, буг, анд, нар, гви, ган, хай, дет, чар, тсх, дан, чуа, куа, цуа sáó , кхо ǰáó , хаб sàó , хие saui .

Группа форм (б) отражает устойчивый общецентр.-кой. корень $*\text{cao}$, который в нам сохраняется только в переносном значении: sáó 'следовать (друг за другом), идти гуськом'.

Что касается инновации ʃáré-b (форма м. р.), то ее исходная семантика, скорее всего, видна в соответствующей форме ж. р.: ʃáré-s 'зад (человека), курдюк (овцы)'; это связано с тем, что в нам формы женского рода устойчиво ассоциируются с признаками

‘большой’, ‘круглый’, а мужского — с признаками ‘длинный’, ‘узкий’ (ср. м. р. *hâi-b* ‘высокое и стройное дерево’, ж. р. *hâi-s* ‘низкое, раскидистое дерево с густой кроной’ и т. п.), так что связь между значениями ‘зад’ и ‘хвост’ здесь полностью вписывается в общую картину морфологической семантики нам.

42) «ты»: нам *sá=c* (м.), *sá=s* (ж.), кор *sa=c* (м.), *sa=s* (ж.), ани, бут, анд *cá* (м.), *há ~ hǎ́* (ж.), кхо *çá* (м.), *hǎ́* (ж.), нар *cá ~ cá-ci* (м.), *sá ~ sá-sí* (ж.), гви *cí* (м.), *sí* (ж.), ган, хаб *câ* (м.), *sâ* (ж.), хай *çá* (м.), *sá* (ж.), дет, чар, тсх, дан, куа, цуа *cá* (м.), *sá* (ж.), чуа *çá* (м.), *šá* (ж.), хие *çá*.

Все перечисленные формы связаны друг с другом этимологически, но парадигмы в кхойкхой и калахари-кхой устроены разным образом. Для пракалахари-кхой общая ситуация восстанавливается в виде **ca* (м. р.), **sa* (ж. р.); развитие **sa* → **ha* в подгруппе кхве нерегулярно и, скорее всего, объясняется какими-то контекстными особенностями употребления этого местоимения как высокочастотной формы).

Для пракхойкхой восстанавливается общая парадигма **sa=c* (м. р.), **sa=s* (ж. р.). При этом элемент **sa=* имеет очевидно префиксальный статус, т. к. в соответствующих клитических формах, выражающих значение прямого объекта, он не наблюдается: ср. нам -*c* (м. р.), -*s* (ж. р.). Соответственно, клитические морфемы -*c*, -*s* уместно считать редуцированными вариантами тех же самых морфем **ca*, **sa*, которые реконструируются для пракалахари-кхой. Что касается «префикса» **sa-*, то он исторически может восходить к редупликации: **sa-sa* ‘ты (ж. р.)’ → **sa-s*, **ca-ca* ‘ты (м. р.)’ → **sa-c* (диссимилиация из **ca-c*).

Применив к полученной для общецентр.-кой. оппозиции **ca* (м. р.) : **sa* (ж. р.) принцип внутренней реконструкции, эти формы можно разложить на корневую морфему **a* и «родовые» префиксы **c-* и **s-*. В таком случае местоименный префикс женского рода **s-* будет, по всей видимости, тождествен общецентр.-кой. суффиксу женского рода **-sa* [Vossen 1997a: 342]. Для префикса мужского рода подобной аналогии провести нельзя (общецентр.-кой. суффикс мужского рода восстанавливается в виде **-ba*), но отметим, что точно такая же оппозиция наблюдается и в других местоимениях (ср. в нар: *sì-c-ám* ‘мы (м. р.)’, *sì-s-ám* ‘мы (ж. р.)’; *sà-c-áo* ‘вы (м. р.)’, *sà-s-áo* ‘вы (ж. р.)’; *xà-c-árá* ‘они (м. р.)’, *xà-s-árá* ‘они (ж. р.)’), так что даже с синхронной точки зрения морфологическое

разложение местоимений *ca, sa* на *c-/s-* + *-a* в определенной степени оправдано. К внешнему сравнению, таким образом, в качестве базовой морфемы 2-го л. ед. ч. привлекать следует моновокалический сегмент **a*.

43) «язык»: нам *nám-mi* ~ *là'm-mi* ~ *tám-mi*, кор *tám-mà*, ани, кхо, бут, нар, хай, дет, чар, тсх, дан, чуа, куа, цуа *dám*, анд *dám*, гви *gyám*, ган, хаб *dám*, хие *d'am*.

Р. Фоссен [Vossen 1997a: 510], справедливо восстанавливая для пракалахари-кхой форму **dam*, из соображений осторожности цитирует параллель в кор как этимологически сомнительную, а форму в нам не приводит вообще. Однако наличие в нам диалектного варианта *tám-mi* показывает, что речь, скорее всего, идет о том же самом корне.

Чередование *n- : l- : t-* в диалектах нам уникально для данной основы, но само наличие таких нерегулярностей типично для этимона «язык», особенно в койсанских языках (см. выше аналогичные нерегулярности для соотв. сев.- и юж.-кой. корней). Возможно, речь идет об уникальном случае працентр.-кой. фонемы **l-* или **n-* (первая для центр.-кой. не реконструируется вообще, вторая постулируется только в срединной позиции), или о каком-то нетривиальном и неустойчивом сочетании согласных. В любом случае отрицать этимологическую связь между формами в нам, кор и калахари-кхой языках, одновременно признавая ее между, например, германским **tungōn-*, др.-индийским *jihvá-* и славянским **językъ* вряд ли разумно (тем более, что между нам и большинством калахари-кхойских языков налицо даже прямая просодическая корреляция).

44) «зуб»: (а) нам *||úú-b*, кор *||ú-b*, кхо, гви, ган, хай, чар, тсх, дан, чуа, цуа *||ú*, бут, нар, дет *||ú*, хаб *||ú*, куа *||ú*, хие *||o*; (б) ани *ságári*.

Все формы (кроме неясной замены в ани, судя по фонетике, явно иноязычного происхождения) согласованно отражают общецентр.-кой. корень **||ú*. Фарингализация гласного в полевых записях Р. Фоссена отмечена только для хаб, но ср. также нар *||ǫ̃*: в словаре Г. Фиссера; не исключено, что признак фарингализованности также следует выводить на прауровень.

45) «дерево»: нам *hái-i*, кор *hei-b* [Wu.], ани, кхо, бут, нар, хаб, цуа *ú*, анд, гви, ган, дет, чар, тсх, дан *ú*, хай *zi*, куа *ú*, хие *hi*.

Пракалахари-кхойская форма восстанавливается в виде **y*, т. е. с чрезвычайно редкой начальной фонемой **y*; вследствие этой редкости точные соответствия ее в кхойкхой неизвестны. Единственный более или менее надежный пример — как раз данный корень, где пра-кхойкхой **h*- (**hai*-) теоретически может оказаться регулярным рефлексом працентр.-кхой. **y*-. Обратное (т. е. первичность кхойкхой **h*-) менее вероятно вследствие наличия такого яркого контрпримера, как нам *hii* ~ *hi* 'делать' = ани, кхо, буг, хай, дет и др. *hi* id. ← працентр.-кой. **hi* (здесь, как видно, ларингальный согласный сохраняется во всех языках, не подвергаясь ассимилятивному воздействию со стороны переднего гласного).

46) «два»: нам *lám*, кор *lam*, ани, кхо, буг, анд, нар, гви, ган, хаб, хай, дет, чар, тсх, дан, куа, цуа *lám*, хие *lame*.

Все формы закономерно отражают працентр.-кой. **lám*.

47) «вода»: (а) нам *lám-mi*, кор *lam-mi* ~ *lam-ta*; (б) ани, буг, хай, дет, чар, тсх, дан, куа, цуа *c^há*, кхо *šá*, анд, нар *c^há*, гви, ган *c^há*, хие *sa*.

Формы группы (а) отражают пракхойкхой **lam*, группы (б) — пракалахари-кхой **c^ha*. Определить относительную степень семантической архаичности этих двух этимонов довольно сложно.

На первый взгляд, с формами группы (б) вполне сопоставима диалектная форма *sàà-b* 'вода' в нам (диалект хай||ом). Однако в остальных диалектах этого языка главное и единственное значение именной формы *sàà-b* — 'слиюна', а соответствующая ей глагольная основа *sàà* имеет значения 'лизать', 'лакать' (= кор *t^há* id.). Помимо этого, на данный момент отсутствует твердая уверенность в том, что именно пракхойкхой **c*- (→ нам *c*-, кор *t^h-*) регулярно соответствует редкой пракалахари-кхой фонеме **c^h-*. Не исключено, что ситуация в диалекте хай||ом вторична. Стоит учесть, что в диалектологическом исследовании [Naascke et al. 1997: 139] отмечается довольно высокий подскок лексических изоглосс между хай||ом и калахари-кхой языком нар — серьезный аргумент в пользу ареальных контактов (хай||ом — диалект, распространенный на северо-восточной периферии общего ареала нам), и, в частности, на необычное развитие значения 'слиюна' → 'вода' в хай||ом могла оказать влияние семантическая контаминация со значением нар *c^há*.

Внешней этимологии у кхойкхой **lam* не обнаружено; стоит отметить необычный вариант *lám-mi* в диалекте Сесфонтейн,

который не удается пока объяснить ни диалектным фонетическим развитием («декликсизация» *// → t- в нам беспрецедентна), ни контаминацией (возможный источник таковой не обнаружен).

48а) «мы» (дв.): нам [инкл.] $sá=kx=ñ$ (м.), $sá=ñ$ (ж., общ.), [экскл.] $sí=kx=ñ$ (м.), $sí=ñ$ (ж., общ.); кор [инкл.] $sa=k^h a=m$ (м.), $sa=sa=m$ (ж.), $sa=m$ (общ.), [экскл.] $si=k^h a=m$ (м.), $si=sa=m$ (ж.), $si=m$ (общ.); ани $có=ñ$ (м.), $só=ñ$ (ж.), $k^h á=ñ$ (общ.); кхо $çá=ñ$ (м.), $šá=ñ$ (ж.), $k^h á=ñ$ (общ.); буг, анд $cá=ñ$ (м.), $yá=ñ$ (ж.), $k^h á=ñ$ (общ.); нар $sì=cá=ñ$ (м.), $sì=sá=ñ$ (ж.), $sì=k^h á=ñ$ (общ.); гви, ган $?í=cè=bè$ (м.), $?í=sè=bè$ (ж.), $?í=k^h è=bè$ (общ.); хаб $cè=bè$ (м.), $sè=bè$ (ж.), $k^h è=bè$ (общ.); хай $cá=ñ$ (м.), $sá=ñ$ (ж.), $k^h á=ñ$ (общ.); дет $?à=cá=ñ$ (м.), $?à=sá=ñ$ (ж.), $?à=k^h á=ñ$ (общ.); чар $zá=ñ$ (м.), $sá=ñ$ (ж.), $k^h á=ñ$ (общ.); тсх $cú=ñ$ (м.), $sú=ñ$ (ж.), $k^h ú=ñ$ (общ.); дан $cú=ñ$ (м.), $sú=ñ$ (ж.), $k^h á=ñ$ (общ.); чуа, куа, цуа $cá=bè$ (м.), $sá=bè$ (ж.), $k^h á=bè$ (общ.);

48б) «мы» (мн.): нам [инкл.] $sá=k=è$ (м.), $sá=s=è$ (ж.), $sá=t=à$ (общ.), [экскл.] $sí=k=è$ (м.), $sí=s=è$ (ж.), $sí=t=à$ (общ.); кор [инкл.] $sa=t^y=e$ (м.), $sa=s=e$ (ж.), $sa=d=a$ (общ.), [экскл.] $si=t^y=e$ (м.), $si=s=e$ (ж.), $si=d=a$ (общ.); ани $\parallel=é$ (м.), $s=é$ (ж.), $t=é$ (общ.); кхо $\parallel=é$ (м.), $š=é$ (ж.), $t=é$ (общ.); буг, анд $\parallel=é$ (м.), $y=é$ (ж.), $t=é$ (общ.); нар $sì=\parallel=áé$ (м.), $sì=s=é$ (ж.), $sì=t=á$ (общ.); гви, ган $?í=\parallel=àè$ (м.), $?í=s=è$ (ж.), $?í=t=àè$ (общ.); хаб $\parallel=àè$ (м.), $s=è$ (ж.), $t^y=àè$ (общ.); хай, дан $k=é$ (м.), $s=é$ (ж.), $c=é$ (общ.); дет $?à=k^y=é$ (м.), $?à=s=é$ (ж.), $?à=c=é$ (общ.); чар $z=é$ (м.), $s=é$ (ж.), $z=é$ (общ.); тсх $\parallel=é$ (м.), $s=é$ (ж.), $c=é$ (общ.); чуа, куа $k=á$ (м.), $sí=è$ (ж.), $cí=è$ (общ.); куа $k=â$ (м.), $s=í$ (ж.), $c=í$ (общ.).

Подробный исторический разбор этой достаточно запутанной системы форм, где комбинации значений трех параметров (клюдивность, число, род) могут давать до 12 разных форм местоимения 'мы', можно найти в [Vossen 1997a: 370-376]. Нас интересует в первую очередь проблема выделения и максимально корректной реконструкции корневых морфем. Основное противопоставление, в отличие от сев.- и юж.-кой. языков, обнаруживается здесь не между формами инклюзива и эксклюзива (эта оппозиция налицо только в кхойкхой), а между формами двойственного и множественного числа.

В дв. ч. за вычетом префиксов клюдивности и рода корневая морфема чаще всего имеет вид =m, за исключением подгрупп гви-гана и чуа, где зафиксирован вариант =be. Р. Фоссен склоняется к мысли о первичности именно второго варианта, сравнивая

развитие $*=be \rightarrow *=m$ с аналогичной «мутацией» $*=ba \rightarrow *=ma$ в показателе мужского рода в целом ряде языков калахари-кхой.

На самом деле аналогия здесь весьма сомнительна. Во-первых, развитие $*=ba \rightarrow *=ma$ наблюдается гораздо реже, чем $*=be \rightarrow *=m$; взрывной характер губного согласного в показателе м. р. сохраняется как в нар ($=ba$), так и в языках кхойкхой (нам, кор $=b$), при том, что местоименная основа во всех этих языках содержит носовой $-m$. Во-вторых, любопытно то, что даже в « be -языках» соответствующая форма *объектного* местоимения часто все равно имеет носовой (ср. в ган: субъект $?i=cè=bè$, объект $?i=cè=mà$).

Оба эти наблюдения скорее свидетельствуют в пользу того, что именно $*=m$ является здесь наиболее архаичным вариантом; форму $*=be$, скорее всего, стоит трактовать как какое-то позиционное искажение или слияние с конечной частицей, в результате которой имело место контекстное изменение $*=m-e \rightarrow *=be$.

В языках калахари-кхой морфема $*=m$, сочетаясь с показателями рода, образовывала сложные основы $*ca=m$, $*sa=m$, $*k^h a=m$. В кхойкхой, помимо этого, система дополнительно усложнялась за счет добавления префикса эксклюзивности $*si=$ (по-видимому, старой эмфатической частицы, ср. ее наличие в формах нар) или префикса инклюзивности $*sa=$ (скорее всего, = местоимение 2-го л. 'ты'); при этом старая форма общего рода $*k^h am$ перешла в мужской род, а в нам, к тому же, старая форма женского рода оказалась вытесненной инновативной формой общего рода.

В множественном числе основная корневая морфема однозначно реконструируется как $*=e$. С ней сочетается другая система родовых показателей: наиболее архаичный вариант, по-видимому, без изменений сохраняется в ани — $*//e$ (м. р.), $*s=e$ (ж. р.), $*t=e$ (общ. р.). Высокочастотная основа $*//e$ из-за наличия в ней латерального кликса часто подвергается в языках-потомках различным упрощениям, обычно $\rightarrow *ke$.

49) «что»: (а) нам $tá(r)é$, кор $da-$; (б) ани, тсх $né$, куа $ná-ù$, хие $na-o$, $na-te$; (в) кхо $máa-xù$ [KH]; (г) буг, чар, дан $ndú$, нар $dǎ$, гви, ган ni : [Та.], дет $dú$;

50) «кто»: (а) нам $tà(r)ǎ$, кор $tǎé-$; (б) ани ta , кхо $má$, буг ta , ган $mâ$, гви $mâ$, чар, дан, тсх, дет, куа $má$, хие $na-ma$; (в) нар $dí-$.

Системы вопросительных местоимений в кхойкхой и калахари-кхой устроены по-разному. В пра-кхойкхой оба местоимения

образовывались от одной и той же основы **ta-* (причины озвончения начального согласного в одушевленном варианте в кор непонятны). В пракалахари-кхой, если обобщить самую распространенную из наблюдаемых оппозиций, основным этимоном в значении 'что?' следует считать основу **ndu-*, в значении 'кто?' — основу **ma-*. Последняя, кстати, сохраняется и в кхойкхой, но только в адъективной функции (ср. в нам: *mā gomasa* 'какая корова?', *mā-sa xi* 'по какой причине?' и т. п.).

Сочетание согласных *nd-*, тем более в начальной позиции, уникально для центр.-кой. лексик и в историческом плане, скорее всего, представляет собой стяжение из **nV-d-*, где **nV-* может быть той же морфемой, что и *ne*, *na* в группе форм (496). В этом случае **-d-* можно отождествить с начальным дентальным согласным в кхойкхой основе **ta-*. Впрочем, оба эти предположения достаточно спекулятивны. Бесспорны лишь (а) выводимость на общецентр.-койсанский уровень вопросительной морфемы **ma*; (б) широкое распространение вопросительных форм с начальным дентальным согласным, по-видимому, также отражающее архаичное состояние. Менее убедительна реконструкция пракалахари-кхой **nV*, но и эта морфема также представлена довольно широко, особенно в адъективном значении 'какой?' [Vossen 1997a: 265-267]. По-видимому, на уровень внешнего лексикостатистического сравнения можно уверенно выводить морфемы **ma* и **ta-*, менее уверенно — **nV* (скорее как адъективную, чем именную основу).

2.5. Язык квади.

2.5.1. *Общие сведения и источники.* Изолированный язык квади [квд] был примерно до середины XX в. распространен на юго-западе Анголы, однако уже в 1950-е гг. число его носителей, по примерным оценкам, не превышало десятка [Güldemann 2004: 252]. Единственными специалистами, успевшими собрать полевые записи по квади, оказались португальский антрополог А. де Алмейда, работавший с информантами в 1950-х гг., и лингвист Э. Вестфаль, причем подавляющее большинство этих данных до недавнего времени оставалось неопубликованным, за исключением небольшого списка из примерно 50 слов в сопоставлении с другими койсанскими материалами в [Westphal 1965] и отдельных

лексических и грамматических примеров в [Westphal 1971]. После того, как в начале 2000-х гг. Кейптаунский университет запустил проект по оцифровке и онлайн-публикации архивов Э. Вестфала, информация по квади стала более доступной: сперва в виде цитаций в работах койсанологов по генетическим связям квади (таких, как [Güldemann & Elderkin 2003]), затем, с 2011 г. – в оригинальном виде (сканированные рукописи) на официальном сайте Кейптаунского университета.

В совокупности данных Вестфала [W.] и Алмейды [Alm.] хватает для заполнения 35 позиций в 50-словном списке: этого достаточно для того, чтобы хотя бы прикидочным образом учесть квади в нашем исследовании и попытаться объективно определить место этого языка в выстраиваемой классификации. Разумеется, поручиться за фонетическую и семантическую точность данных в подобного рода ситуации невозможно. Исходя из общих соображений, транскрипции Э. Вестфала, по-видимому, заслуживают большего доверия, чем формы в фиксации де Алмейда, но при этом существенная часть базисной лексики квади известна только по записям последнего.

2.5.2. *Историческая характеристика.* Из-за скудности данных сказать что-либо определенное о предыстории квади не представляется возможным. С точки зрения синхронного анализа бросается в глаза определенная «типологическая промежуточность» квади между южноафриканским койсанским и восточноафриканским койсанским типом. С одной стороны, структура корня в квади явно тяготеет к моносиллабичности, как в центр.-кой. языках; с другой — обращает на себя внимание относительная бедность системы кликсов (надежно выделяются лишь дентальный и палатально-альвеолярный, с подсистемами из четырех или пяти исходов) при наличии полной системы латеральных аффрикат и фрикативных, что типологически сближает квади с сандаве (см. ниже, п. 2.6.2).

В плане грамматики наиболее значительной структурно-типологической изоглоссой между квади и центр.-кой. языками является классифицирующая категория рода, имеющая одни и те же значения (мужской, женский, «средний» или «общий») и выражающаяся в обеих группах суффиксальными показателями.

Генетический статус квади на сегодня остается неясным. В исходную койсанскую классификацию Гринберга этот язык не попал, т. к. данные по квади Гринбергу в 1950-е/1960-е гг. оказались недоступными. Сам Э. Вестфаль, который, как известно, придерживался «радикально-отрицательных» взглядов на генетический статус термина «койсанские языки», считал квади лингвистическим изолятом, не имеющим не только близких, но и относительно дальних родственников; отмечая отдельные лексические сходжения между квади и центр.-кой. языками, он, тем не менее, предпочитал объяснять их контактными связями квади с !кора и, возможно, другими языками кхойкхой.

Несогласие с позицией Вестфалья высказал современный койсанолог Т. Гюльдемманн, который, проанализировав грамматические и лексические данные квади, пришел к выводу, что существуют веские основания помещать его вместе с центр.-кой. языками в единую семью кхой-квади [Güldemann, Elderkin 2003; Güldemann 2004]. Критика грамматической аргументации Т. Гюльдемманна, которая представляется нам убедительной лишь отчасти, не входит в цели данного исследования; наша задача сводится к тому, чтобы показать, насколько эта аргументация подкрепляется параллелями в области базисной лексики. В любом случае на сегодняшний день факт генетического родства между квади и центр.-кой. языками не является ни общепринятым, ни интуитивно очевидным; даже если оно существует, расстояние между працентр.-кой. языком и квади чересчур значительно, чтобы можно было на первом же этапе успешно и наглядно реконструировать общий для этой семьи 50-словный список. Все это вынуждает нас рассматривать данные квади в отдельном разделе.

50-словный список для квади (заполнены 35 позиций).

- 3) «черный»: квд *ʒu*.
- 4) «кровь»: квд *|ó-ʔè* [Alm.].
- 5) «кость»: квд *ʔã-*.
- 7) «умирать»: квд *kádé ʔò* [Alm.].
- 8) «собака»: квд *ʔau-de* [W.], *ʔau-di* (м. р.), *ʔau-yi* (ж. р.), *ʔayi* (общ. р.) [Alm.].
- 9) «пить»: квд *kca* [W.], *ká-lá-* [Alm.].

- 11) «ухо»: квд *go-de; gó-dì* [Alm.].
 12) «есть»: квд *ʔnũ-*.
 13) «яйцо»: квд *ʔi-dì*, мн. ч. *ʔi-wà* [Alm.].
 14) «глаз»: квд *ší-di* [W.], *ší-dì* [Alm.].
 15) «огонь»: квд *ʔē:* [W.], *ʔē-dè* [Alm.].
 16) «нога»: квд *ze-*.
 17) «волосы»: квд *lòʔm̄* [Alm.].
 19) «голова»: квд *c^hě*, мн. ч. *c^hũʔe* [W.]; (?) *fúʔm̄* [Alm.].
 20) «слышать»: квд *kum* [W.], *kú-lá-* [Alm.].
 21) «сердце»: квд *cʔo-de*.
 23) «я»: квд *ta* (посс. *čí*).
 27) «мясо»: квд *kxóle* (Э. Вестфаль приводит также отдельную форму *kxóʔ-e* 'плоть', в которой исходный корень **kxo-* виден более отчетливо).
 28) «луна»: квд *k^hā-de* [W.], *xá-bè*, мн. ч. *xá-wà* [Alm.].
 29) «рот»: квд *kxami-* [W.], *kámè* [Alm.].
 32) «ночь»: квд *t^hwi:* [W.], *t^hwî*, мн. ч. *t^hũ-wà* [Alm.].
 33) «нос»: квд *čwi-de* [W.], *čí-dì* [Alm.].
 35) «один»: квд *lwi* [Alm.].
 36) «дождь»: квд *fétè* [Alm.].
 38) «звезда»: квд *lxo-*, мн. ч. *lxwé-ya* [W.]; *lxwê:*, мн. ч. *lxó-wà* [Alm.].
 39) «камень»: квд *ʔo-de* [W.], *gú-dè* [Alm.].
 40) «солнце»: квд *ʔui-de* [W.], *ʔúdú-dì*, мн. ч. *ʔúdú-wà* [Alm.].
 41) «хвост»: квд *θo-de*.
 42) «ты»: квд *sa*.
 43) «язык»: квд *tamen-de* [W.], *támē-dì* [Alm.].
 44) «зуб»: квд *ʔwe*, мн. ч. *ʔōʔ-e* [W.], *čǒ-ʔi* [Alm.].
 45) «дерево»: квд *č^hi-dì¹*.
 46) «два»: квд *lā ~ lá* [Alm.].
 47) «вода»: квд *kxoʔ-e*.

¹ Форма приводится по [Westphal 1971: 396]. В [Westphal 1965: 144] это же слово дается как *t hí-dì*, где на месте пробела, скорее всего, должен был находиться вставной рукописный знак *f*. Тот факт, что корень следует читать как *č^hi-*, а не *thi-*, несколько снижает, но не отменяет вероятность того, что он является заимствованием из какого-то бантуского источника, ср. прабанту **tí* 'дерево' (в языках групп L и R нередко представлен в палатализованном виде как *=çi*, *=ši*, см. [Guthrie 1967: IV, 105]).

48) «мы»: (а) *дв. ч.: квд а-ти, ha-ти*, объектная форма *-ти*; (б) *мн. ч.: квд ina, hina, u-hina*.

2.6. Язык сандаве.

2.6.1. *Общие сведения и источники.* В отличие от квади, изолированный язык сандаве известен исследователям с гораздо более раннего времени; к тому же на нем до сих пор активно говорят ≈ 40,000 человек в центральных областях Танзании, а лингвистические исследования его фонетики, грамматики и лексики продолжают до сих пор.

Первое подробное описание сандаве, включавшее словарь, было составлено О. Демпвольфом уже в начале XX в. [Dempwolff 1916]. Значимость его на сегодня является скорее исторической, но в лексикографическом плане оно полезно возможностью проверки современных записей на (хотя бы относительную) «древность». Подробный, но также отчасти устаревший словарь был спустя несколько десятилетий составлен М. ван де Кимменаде [Kimmenade 1954]. Наконец, последним по времени и наиболее профессиональным общим лексикографическим описанием сандаве является словарь Р. Кагайя [Kagaya 1993], который мы принимаем в качестве основного источника по умолчанию [Ka.]¹.

Впрочем, сопоставление с другими, менее подробными, но более точными исследованиями по лексике и фонетике сандаве показывают, что записи Кагайя (не говоря уже о его предшественниках) содержат многочисленные неточности и ошибки, в том числе системные (например, отсутствие маркировки тонов). Поэтому необходимым представляется там, где это возможно, дополнять транскрипции Кагайя альтернативными записями из других источников. Это: (а) словарь Демпвольфа [Dm.]; (б) работы Д. Элдеркина по отдельным аспектам фонетики, грамматики и

¹ Уже в процессе подготовки рукописи к печати стало известно о выходе в свет еще одного словаря сандаве под редакцией К. Эрета и П. Эрета, обработавших материалы, собранные недавно скончавшимся антропологом Э. тен Раа [Raa 2012]. К сожалению, полноценно учесть данные этого источника нам уже не удалось, но (при наличии не менее трех других словарей этого языка) маловероятно, что они могли бы существенно повлиять на результаты исследования.

возможной предыстории сандаве, основанные на полевой работе автора [Elderkin 1983, 1998] (помечены как [Eld.]); (в) лексикофонетические примеры из работы [Tucker, Bryan, & Woodburn 1977], помечены как [TBW]; (г) записи Х. Итон и ее коллег по Летнему институту лингвистики [Eaton 2002; Hunziker & Eaton 2007, 2008], помечены как [Еа.].

Вопрос о диалектном членении сандаве и, соответственно, возможности какой-либо внутренней реконструкции остается открытым. Судя по результатам недавнего исследования [Hunziker & Eaton 2007], диалектные различия между центральной, западной и восточной группами сандавеязычных деревень существуют, но примеров таких различий приводится явно недостаточно, чтобы оценить их значимость. По крайней мере в плане лексики их так мало, что проведение внутреннего лексико-статистического исследования по диалектам сандаве вряд ли имеет смысл (из 246 слов, диалектные данные по которым собраны в упоминаемом исследовании, собственно лексические расхождения были обнаружены только в шести случаях).

2.6.2. *Историческая характеристика.* Дж. Гринберг относил сандаве вместе с хадза к «восточноафриканским» койсанским языкам (не настаивая, однако, на генетической интерпретации этого термина), противопоставляя их «южноафриканским» койсанским в первую очередь по своим типологическим характеристикам. В фонологическом плане основными отличиями сандаве от центр.-, юж.- и сев.-кой. языков можно считать следующие:

а) преимущественно двусложная структура корня (CVCV); односложные корни структуры CV также частотны, но не преобладают над двусложными. В «южноафриканских» койсанских в структурах C_1VC_2V накладываются жесткие ограничения на позицию C_2 (второй согласный может быть только сонорным или глайдом); эти ограничения, равно как и данные сравнительного анализа, наводят на мысль, что второй слог в таких структурах представляет собой застывший (а в некоторых случаях — все еще продуктивный, как, например, в юж.-кой. языке !хонг) суффикс класса. В сандаве, напротив, на позицию C_2 никаких сколь-либо значимых ограничений нет (ср.: *ala* 'стрела', *cade* 'шкура', *scha* 'роса'), вплоть до того, что в ней могут иногда встречаться даже кликсы (ср.: *u//ri* 'кашлять', *ha!a* 'звать');

б) наличие полного ряда латеральных аффрикат (глухая λ , звонкая L , абруптивная $\lambda^ʔ$) и сибилантов (глухой t); в «южноафриканских» койсанских λ и $\lambda^ʔ$ появляются лишь окказионально (как в юж.-кой. языке ||хегви), а L и t вообще неизвестны;

в) отсутствие жестких ограничений на употребление губных согласных p , b , m , дистрибуция которых, в том числе и в начальной позиции, вполне стандартна;

г) относительная скудность подсистемы кликсов — различаются всего три основы (дентальный $!$, палатально-альвеолярный $!$, латеральный $||$) и четыре исхода (нулевой/велярный, носовой, придыхательный $-h$ -, гортанная смычка).

Все это, равно как и очевидное отсутствие близкого родства с другими койсанскими языками, очень затрудняет сравнение лексического материала сандаве с потенциальными когнатами в «южноафриканских» койсанских. В плане грамматики основные изоглоссы обычно отмечаются между сандаве и центр.-кой. группой; Д. Элдеркин и Т. Гюльдемманн, в частности, выдвигают идею языковой семьи «сандаве-квади-кхой», оговариваясь, однако, что позиция сандаве в этой семье наиболее шаткая и нуждается в дополнительном обосновании [Elderkin 1986; Güldemann, Elderkin 2003]. Дж. Гринберг, отмечая такую типологическую общность между сандаве и центр.-кой. языками, как систему родов, тем не менее, считал, что особой лексической близости именно между этими двумя группами не наблюдается. Отсутствуют подобного рода таксономические выводы и в сравнительном исследовании Б. Сэндс [Sands 1998a].

Мы будем, таким образом, исходить из статуса сандаве как языка-изолята, генетические связи которого с другими языковыми семьями Африки либо отсутствуют, либо являются удаленными и требуют первичного обоснования с помощью предварительно-лексикостатистической методики.

В связи с тем, что все диалекты сандаве очень близки друг к другу, возможности внутренней реконструкции применительно к этому языку чрезвычайно ограничены. Первичный этимологический анализ лексики сандаве, по сути, сводится лишь к фильтрации очевидных заимствований: носители сандаве находятся в тесных контактах с носителями языков банту, а также с южнокушитскими языками (иракв, алаве, бурунге). При этом, однако,

сандаве очевидным образом обладает более высоким «лексическим иммунитетом» по сравнению, например, с хадза — в частности, в составе 50-словного списка ни одного прозрачного заимствования из соседних языков не обнаружено (отдельные подозрительные случаи будут обсуждены ниже, но все они по большому счету бездоказательны).

50-словный список для сандаве.

- 1) «пепел»: сан *ʃur^ha* [Ka.], *!ɲura* [Dm.], *!ɲip^há* [Eld., Ea.]¹.
- 2) «птица»: сан *t^hui* [Ka.], *t^hwi* [Dm.], *twi:* [TBW], *t^hwi:* [Ea.].
- 3) «черный»: сан *kʷaŋkʷara* [Ka., Dm.], *kʷá(ŋ)kʷàrà* [TBW], *kʷaŋkʷàrà* [Ea.].

В [Hunziker & Eaton 2007] приводится также диалектный этимон *sɲiʔá*, отмеченный в говоре одного информанта; исторически это слово, скорее всего, связано с *sɲisɲi-n* [Ka.], *sɲísɲí:* [Eld.] 'уголь', но архаичность его в значении 'черный' маловероятна в связи со столь скромной дистрибуцией и возможностью вторичной этимологизации.

- 4) «кровь»: сан *ʃekʷa* [Ka., Dm.], *ʃékʷà* [Eld.], *ʃékʷà* [TBW], *ʃékʷá* [Ea.].
- 5) «кость»: сан *ʃi* [Ka.], *!i* [Dm.], *!i-ŋ* [TBW], *!i* [Ea.].
- 6) «ноготь»: сан *sɲwaʔa* [Ka.], *soa* [Dm.], *sɲwáʔá* [TBW, Ea.].
- 7) «умирать»: сан *ʎa:-si* [Ka.], *ʎa-si* [Dm.], *ʎá:s ~ ʎá:sʔ* [TBW], *ʎá:-si* [Ea.].

- 8) «собака»: сан *ka:ka* [Ka.], *ká:kà* [Eld., TBW], *ká:ká* [Ea.].
- 9) «пить»: сан *sɛ:* [Ka.], *sɛ* [Dm.], *sé:* [Eld., TBW, Ea.].
- 10) «сухой»: сан *ʃiŋ-* [Ka.], *ʃi(ŋ)-* [Dm.], *ʃi-* [Eld.].

В списке [Hunziker & Eaton 2007] этого глагольного корня ('сохнуть, высухать') нет; вместо него приводится форма *sɲimé* 'сухой; быть сухим', подтверждения которой в других источниках найти не удалось.

- 11) «ухо»: сан *keke* [Ka., Dm.], *kéké* [TBW, Ea.].
- 12) «есть»: сан *manɕ^ha* [Ka.], *manca* [Dm.], *mânca* [TBW], *mánɕ^há* [Ea.].
- 13) «яйцо»: сан *diʔa* [Ka.], *tiʔa* [Dm.], *diʔá* [TBW], *diʔá* [Ea.].

¹ В сандаве отсутствует фонологическое различие между альвеолярным (!) и палатальным (ʃ) кликсами; реальная артикуляция в зависимости от говора колеблется между обоими типами, что и объясняет различие в нотации у Кагайя (ʃ) и остальных исследователей (!).

В [TBW] приводится также синонимичная форма *cʷè*, однако последняя не находит подтверждения в других источниках.

14) «глаз»: сан /*oe*(:), /*we*(:) [Ka.], /*wé* [Dm.], /*wé*: [TBW], /*wè*: [Ea.].

15) «огонь»: сан /*piŋ* [Ka., Dm.], /*piŋ* [TBW], /*ṗi*: [Ea.].

16) «нога»: сан /*ʰata* [Ka.], /*ada* [Dm.], /*ʰatá* [Eld., Ea.], /*ʰatá* [TBW].

17) «волосы»: сан *cʷe* [Dm.], *cʷé* [Eld., Ea.], *cʷe* [TBW].

Именно это слово употребляется по умолчанию в значении 'волосы на голове'. В словаре Кагайя оно отсутствует, а соответствующее значение передается композитом *se: ɸu*, букв. 'головы-волосы', где первая часть = 'голова' (см. ниже), а вторая = 'волосы, волосяной покров' (вообще), ср. также [Dm.] /*ʔù* 'волосы', 'перья', [TBW] /*ʔùŋ* ~ /*ʔũ* 'перо'.

18) «рука»: сан /*ʔuŋ* [Ka., Dm.], /*ʔũ* ~ /*ʔùŋ* [TBW], /*ʔũ* [Ea.].

19) «голова»: сан *ɸe*: [Ka.], *ce* [Dm.], *cʰé*: ~ *cé*: [TBW], *čé*: [Ea.].

20) «слышать»: сан /*kʰeʔe* [Ka.], /*keʔe* [Dm.], /*kʰéʔé* [Eld., Ea.].

21) «сердце»: сан /*ʒigida* [Ka., Dm.], /*ʒigidá* [TBW], /*ʒigidá* [Ea.].

22) «рог»: сан /*ʔana* [Ka., Dm.], /*ʔána* [TBW, Ea.].

Это слово в [Kießling, Mous 2003: 193] сопоставляется с бурунге /*ʔana* 'верхняя часть ноги, бедро' (!), но, во-первых, авторами не оговорено возможное направление заимствования, во-вторых, семантическая связь между этими формами носит довольно фантастический характер.

23) «я»: сан /*ci* [Ka.], /*ci* ~ /*ciŋ* [Dm.], /*ci* [Eld.], /*ciŋ* [TBW], /*či* [Ea.].

24) «убивать»: сан /*kʷoe*(:), /*kʷe*(:) [Ka.], /*kʷa* [Dm.].

В [Eaton 2007] в качестве эквивалента значения 'kill' приводится другая основа: /*wákʷã*, которая обнаруживается и в остальных источниках (Ka.: /*wakʷa*, Eld.: /*wàkʷà*), но лишь применительно к множественным объектам, т. е. в более маркированном значении, не подходящем для наших целей.

25) «лист»: сан /*a*: [Ka.], /*a* [Dm.], /*á*: [TBW], /*ǎ*: [Ea.].

26) «вошь»: сан /*ma:ŋʔa* [Ka.], /*maŋʔa* [Dm.], /*má:cʷà* ~ /*mǎʔã* [TBW], /*má:ʔã* [Ea.].

Редкий, но не исключительный, пример двусложной основы с кликсом посередине (вариант с аффрикатой в [TBW], судя по согласованным показаниям всех остальных источников, вторичен).

27) «мясо»: сан /*ʔin* [Ka.], /*ʔi* [Dm.], /*ʔiŋ* ~ /*ʔi* [TBW], /*ʔi*: [Ea.].

28) «луна»: сан /*ʔa:p(u)so* [Ka.], /*!a-biso* ~ /*!a-oso* [Dm.], /*!a-bisò* [TBW], /*!â* ~ /*!ǎ:sò* [Ea.].

Чистая основа **!a-* в значении 'луна' налицо только в списке [Hunziker & Eaton 2007]; во всех остальных источниках этот корень сам по себе имеет значение 'лунный свет' (Ka.: *ʃa*, Dm.: *!a*, TBW: *!â* и др.). Смысл морфемы **-biso* неизвестен (возможно, этот же суффикс представлен и в слове *ʃa-biso* [Ka.], *ʃa-bísó* [Ea.] 'живот', но в данном случае неясным остается даже отдельное значение первой морфемы).

29) «рот»: сан *ʃum* [Ka.], *!ɲuɲ* [Dm.], *ʃû ~ ʃûŋ* [TBW], *ʃû:* [Ea.].

30) «имя»: сан *ʃwa* [Ka.], *ʃoa* [Dm.], *ʃwâ* [TBW, Ea.].

31) «новый»: сан *ʃae(:)* [Ka.], *ʃae* [Dm.], *ʃâ:* [Ea.].

32) «ночь»: сан *tue(:)* ~ *twe(:)* [Ka.], *tue ~ tuwe ~ twe* [Dm.], *twé:* [TBW], *twë:* [Ea.].

33) «нос»: сан *ʃat^{hi}* [Ka.], *ʃati* [Dm.], *ʃát^{hî}* [Eld.], *ʃátin* [TBW], *ʃáti* [Ea.].

34) «не»: сан *-ce* [Ka.], *ce(:)* [Dm.], *č^hē* [Ea.].

Абруптивная артикуляция аффрикаты в этой энклитической частице, зафиксированная в словаре Кагайя, не подтверждается другими источниками.

35) «один»: сан *cece* [Ka.], *ceɬ* [Dm.], *ceʒé* [TBW], *ceʒê* [Ea.].

Форма, возможно, является морфологически составной, ср. дальнейший числовой ряд [Ea.]: '2' *kísô-xi*, '3' *swámkí-xi*, '4' *hâká-xi*. Против разложения ее на старый корень **ce* и «нумеративный» суффикс **-xe*, однако, говорят (а) разница в вокализме (в остальных числительных суффикс имеет вид **-xi*), (б) незафиксированность простого варианта **ce*, в то время как все прочие числительные имеют и нерасширенные формы (см. ниже «два»).

36) «дождь»: сан *ʃwawɲ* [Ka.], *ʃwoa* [Dm.], *ʃwá ~ ʃwâ* [TBW], *ʃwâ:* [Ea.].

37) «дым»: сан *ɬukɬa* [Ka., TBW], *ɬuka* [Dm.], *ɬúkâ* [Ea.].

38) «звезда»: сан *ʃowan* [Ka., Dm.], *hí=ʃáwâ:* [Ea.].

Префикс *hi-* обычно наблюдается в глагольных основах (ср. *hi=ki* 'идти', *hi=ʃa* 'привязывать', *hi=!a* 'распределять' и др.); это дает возможность предположить, что мы имеем здесь дело с каким-то отглагольным производным от незафиксированной отдельно основы **ʃawa*. Подозрительно, однако, то, что в сочетании с этим префиксом лексема встречается только в одном источнике ([Ea.]), в то время как все остальные, включая старый словарь Демпвольфа, дают только «корневой» вариант.

39) «камень»: сан *dí(:)n* [Ka.], *din* [Dm.], *din ~ dî* [TBW], *dî:* [Ea.].

40) «солнце»: сан *ʔakas(u)* [Ka.], *ʔagasu* [Dm.], *ʔakásù* [TBW], *ʔàkásu* [Ea.].

41) «хвост»: сан *сʷa:* [Ka.], *сʷá:* [TBW], *сʷǎ:* [Ea.].

42) «ты»: сан *hapi* [Ka.], *hapa ~ hapi* [Dm.], *hàpú* [Eld., Ea.], *hapú* [TBW].

Зависимая форма этого местоимения (объектный и посессивный суффикс) имеет вид *-po* [Dempwolff 1916: 25], что в принципе разрешает членить ее на морфемы **ha-* и **-pu* соответственно; не исключено, однако, что энклитика *-po* — это просто стяжение из более старого **-hapo*.

43) «язык»: сан *ʔheŋ* [Ka.], *!eŋ* [Dm.], *!ʰêŋ ~ !ê* [TBW], *!ʰê:* [Ea.].

44) «зуб»: сан *ʔakʰaŋ* [Ka.], *ʔakaŋ* [Dm.], *!aká* [TBW], *!àkʰǎ:* [Ea.].

45) «дерево»: сан *tʰe:* [Ka.], *tʰe* [Dm.], *tʰé:* [TBW], *tʰě:* [Ea.].

Как и для соответствующего этимона в квади (см. выше), не исключена возможность заимствования из банту (прабанту **=tí* 'дерево').

46) «два»: сан *kiso-x(o)* [Ka.], *ki, ki-so-x* [Dm.], *kísòx* [TBW], *kísòxi* [Ea.].

Односложная форма *ki* встречается только в словаре Демпвольфа, в значениях 'два' и 'тоже, также'; все прочие источники дают только двусложный вариант *kiso-*. К сожалению, данных явно недостаточно, чтобы уверенно охарактеризировать форму *ki* как архаизм, редукцию двусложного *kiso-*, или отдельный корень. Отметим, что в текстах, записанных Демпвольфом, в значении 'два' встречается только *kiso-*.

47) «вода»: сан *сʷa(:)* [Ka.], *сʷa* [Dm.], *сʷâ* [Eld., Ea., TBW].

48) «мы»: сан *s=uiŋ* [Ka., Dm.], *s=ú:* [Eld., Ea.], *s=üŋ* [TBW].

Сравнение с формой 2-го л. мн. ч. *siŋ* [Ka., Dm.], *sí:* [Eld.] позволяет предположить, что начальный элемент *s-* исторически представляет собой отделимый префикс. Выделение корневых морфем **=u(iŋ)*, **=i(iŋ)* хорошо согласуется и с наличием соответствующих глагольных объектных суффиксов: *-o* для 1-го л. мн. ч., *-e* для 2-го л. мн. ч. [Dempwolff 1916: 25].

49) «что»: сан *ho-ʔo:* [Ka.], *ho-s, ho-c* [Dm.], *hó-cò* [TBW], *hó-čō:* ~ *hó-bê* [Ea.].

50) «кто»: сан *ho(:)* [Ka.], *ho*, ж. р. *ho-su* [Dm.], *hò* [Ea.].

Оба вопросительных местоимения образуются от одной и той же интеррогативной основы *ho-*, либо представленной в чистом виде ('кто'), либо расширенной за счет неэтимологизируемых суффиксальных элементов *-co*, *-be* ('что').

2.7. Язык хадза.

2.7.1. *Общие сведения и источники.* Первые краткие описания изолированного языка хадза, на котором сегодня говорит не более 800 носителей в северо-западных районах Танзании, появляются уже в начале XX в. [Obst 1912; Dempwolff 1917; Bleek 1931], но сегодня носят скорее исторический характер. Большой лексический материал по этому языку был накоплен Д. Блик, которая на основании обнаруженных лексических сходств включила его в центральную группу бушменских языков (= центральнокойсанские) наравне с известными ей двумя другими языками этой группы — наро и хие-чваре; в [Bleek 1956] материалы по хадза в связи с этим маркируются как СШ (в анализируемом ниже 50-словном списке формы в записи Д. Блик сопровождаются пометой [В.]).

После выхода в свет работ Д. Блик в изучении хадза наступила затяжная пауза; за пятьдесят лет не было опубликовано никаких новых существенных материалов, за исключением краткого фонетического описания в [Tucker, Bryan, & Woodburn 1977] (в списке формы из этой статьи, а также из доступных нам неопубликованных записей А. Такера помечены как [TBW]). Основной упор делался на попытках более точного определения генетического статуса хадза: Дж. Гринберг [Greenberg 1966] подтверждает «койсанский» статус этого языка, хотя и считает его родство с «южноафриканскими койсанскими» отдаленным; Г. Флеминг [Fleming 1986] предпринимает малоубедительную попытку придать реальный исторический статус термину «восточноафриканские койсанские», сопоставляя лексику хадза и сандаве (а также привлекая, в качестве косвенного аргумента, данные генетических исследований по этим племенам). Д. Элдеркин [Elderkin 1982] занимает более осторожную позицию, указывая на то, что лексика хадза обнаруживает значительное количество сходжений не только с койсанскими, но и с афразийскими (особенно кушитскими и омотскими) языками, а также с кулякскими языками, традиционно относящимися к нило-сахарской макросемье.

Новые перспективы для внутреннего и сопоставительного изучения хадза открываются в 1990-е гг., в первую очередь в связи

с полевыми и аналитическими исследованиями Б. Сэндс, которой удалось собрать много нового лексического материала и перепроверить большую часть старого на предмет аккуратности транскрипционной записи. Подробный словарь хадза до сих пор так и не был формально издан в виде монографии, однако большая часть полевых записей Б. Сэндс была любезно предоставлена ей в наше распоряжение; частично, впрочем, они все же опубликованы — так, слегка модифицированный вариант 100-словного списка по хадза обнаруживается в [Sands 1998a], а более подробный список из 232 слов можно найти в фонетическом описании [Sands, Maddieson, & Ladefoged 1996]. В приводимом ниже списке записи Б. Сэндс [Sd.] будут использованы как формы по умолчанию, т. е. качество их транскрипции отличается высоким уровнем профессионализма.

2.7.2. *Историческая характеристика.* Фонетическая структура словоформы в хадза устроена примерно так же, как и в сандаве: основы, как правило, двусложны, серьезные фонотактические ограничения на консонантизм почти отсутствуют, позицию второго согласного корня могут иногда занимать кликсы. При этом сама подсистема кликсов в хадза еще беднее, чем в сандаве: отмечены те же самые три типа основ (дентальный /, альвеолярный /, латеральный //), но всего три типа исходов (нулевой / велярный; носовой; гортанно-смычный, причем последний всегда сопровождается избыточной преназализацией и в записях Б. Сэндс имеет обычно вид $\eta/ʔ$, $\eta!ʔ$, $\eta//ʔ$; во избежание путаницы между этими кликсами и их аналогами с обычным носовым исходом мы будем маркировать их просто как /ʔ, !ʔ, //ʔ, исходя из нефонологичности преназализации).

Как и сандаве, хадза характеризуется отсутствием ограничений на дистрибуцию губных согласных и наличием ряда латеральных аффрикат и сонантов; помимо этого, твердо противопоставлены свистящие (c, ʒ, s) и шипящие (č, ʒ̣, š) фонемы, а также удается выделить отдельный ряд лабиовелярных смычных и носовых (k^{w} , k^{hw} , $k^{ʷ}$, g^{w} , η^{w} ; в остальных «койсанских» языках транскрипционные нотации типа kw - обычно отражают сочетание обычного k - с первым элементом восходящего дифтонга).

В плане внутренней реконструкции наши данные чересчур ограничены, чтобы можно было добиться каких-либо важных

результатов. Как Б. Сэндс, так и другие исследователи регулярно упоминают о незначительных диалектных различиях в фонетике и лексике хадза, но фонетические различия, по-видимому, носят скорее свободно-вариативный характер, а каких-либо расхождений в базисной лексике хадза между основными доступными нам источниками обнаружить не удалось, так что для целей исследования вполне достаточным будет представление о хадза как о едином языке без диалектов.

Сопоставительные исследования между хадза и прочими «койсанскими» языками, проведенные Б. Сэндс, приводят ее к заключению, что оснований для включения хадза в койсанскую макросемью значительно меньше, чем оснований для выделения койсанской макросемьи как таковой (см. [Sands 1998a] и особенно [Sands 1998b]). То, что некоторые источники (например, система «Этнолог») до сих пор характеризуют хадза как «койсанский» язык, объясняется лишь данью традиции (точнее, уважением к классической таксономии Дж. Гринберга), а также соображениями типологического характера, т. к. хадза, как и сандаве, представляет собой язык «щелчкового типа», при этом не имеющий явно близких родственников в некойсанских макросемьях.

Совсем недавно, в развитие идей Д. Элдеркина, нами было показано [Старостин 2008a], что предварительное лексикостатистическое обследование хадза на предмет возможных связей с некойсанскими языками действительно обнаруживает намного большее количество сходств между хадза и афразийскими языками — в первую очередь омотскими — чем между хадза и койсанскими; на хадза-афразийскую связь намекает и целый ряд серьезных грамматических изоглосс. При этом, однако, вряд ли подлежит сомнению наличие в хадза и «койсанского» слоя. Вопрос сводится скорее к тому, какой из этих слоев в хадза является более древним и детерминирующим генетический статус этого языка. Очень важно также четко различать между «древними» лексическими слоями в хадза и большим количеством явно «новых» заимствований из окружающих языков, иногда проникающих даже в базисную лексику — таких, как банту (кисуахили, кисукума, киньямвези и др.), южнонилотские (датога) и южнокушитские (иракв). По сравнению с сандаве хадза, безусловно, производит впечатление языка с «пониженным лексическим

иммунитетом», что чрезвычайно затрудняет реконструкцию его генетического статуса и ареальной эволюции.

Вплоть до получения четкого ответа на этот вопрос мы будем придерживаться традиционного статуса хадза как «потенциально койсанского» языка, связанного с остальными «койсанскими» семьями отдаленным родством, помещая основные сведения о хадза и разбор 50-словного списка по этому языку в раздел «Койсанские языки».

50-словный список для хадза.

1) «пепел»: хад *hócʔó-* [Sd.], *hócʔò-ko* [TBW].

Д. Блик приводит в значении 'пепел' форму *cʔko*, т. е. 'огонь' (см. ниже); нельзя не отметить явную сегментную близость этих основ — возможно, 'пепел' следует анализировать как *hó=cʔó-*, с застывшим префиксом *hó-*. О других возможных случаях выделения такого рода префиксов см. ниже.

2) «птица»: (а) хад *tʰitʰi-* [Sd.]; (б) хад *cʔélàlò-* [TBW, Sd.], *cealo*, *celealo* [B.].

Различия в значении или в сфере употребления этих двух корней остаются неясными. В такой ситуации нередки случаи, когда один из терминов — исконный, второй же представляет собой относительно недавнее заимствование; в этой связи любопытно отметить такие южнокушитские формы, как иракв, горова *círʔi*, алагва *círʔi*, бурунге *ciraʔa* ← праиракв-бурунге **ciraʔa* [Kießling, Mous 2003: 299]. Тесные контакты между хадза и южнокушитскими языками не подлежат сомнению, а фонетическая близость этих форм не может быть случайной; ср. абруптивную артикуляцию аффрикаты, а также колебание между вариантами с одной и двумя фонемами *-l-* (фонологической разницы между *l* и *r* в хадза нет) в записях Блик, что, возможно, отражает альтернативную передачу кушитского фарингала ʕ как *r* (*l*) или нуля.

Таким образом, при выборе из двух корней более архаичное впечатление производит форма *tʰitʰi-*, хотя происхождение ее может, конечно, быть и звукоподражательным.

3) «черный»: хад *tičʔi-* [Sd.], *ticʔi* [TBW], *diči* [B.].

4) «кровь»: хад *ʔátʰaʔmá-* [Sd.], *átàmâ* [TBW], *átama* [B.].

Д. Блик приводит в своем словаре также форму *áta-be* 'кровеносные сосуды'; поскольку морфема *-be* является регулярным показателем мн. ч., можно предположить, что исторически корнем в слове 'кровь' была морфема **at^ha-*, а *-ma* представляет собой старый застывший суффикс. Против этого, однако, свидетельствует, во-первых, то, что ни для одного другого слова такой суффикс даже гипотетически выделить не удастся; во-вторых — то, что форма *áta-be* не подтверждается другими источниками. Так, в записях Б. Сэндс обнаружены формы *atama-pi* 'кровеносные сосуды (на шее)', а также *atama-pe* 'свернувшаяся кровь' — ни в той, ни в другой слог *ma* не выпадает. Скорее всего, *áta-be* — какое-то диалектное стяжение из **at^hama-be*, и корнем в этом слове является **at^hama* (или **t^hama*, если считать элемент *a-* застывшим префиксом — последнее намного более вероятно, чем застывший суффикс **-ma*).

5) «кость»: хад *miλʷa* [Sd.], *miλʷâ* [TBW].

Бросаются в глаза афразийские параллели — ср. дахало *miλʷo*, онгота *mič'a* и др. — однако ни один из языков, в которых встречаются эти слова, не контактирует с хадза, так что по крайней мере к «новым» заимствованиям этот этимон отнести нельзя.

6) «ноготь»: хад *baʷu-* [Sd.], *baxlu*, мн. ч. *buxlu-pe* [B.].

7) «умирать»: (а) хад *talʷi-* [Sd.], *talʷia* [B.]; (б) хад *misi* [Sd.], *missi*, *misi* [B.].

Различие между этими двумя синонимами неизвестно, так что к внешнему сравнению допустимо привлекать оба.

8) «собака»: (а) хад *ʷáʷano-* [Sd.], *ʷáʷanò ~ ʷáʷano-wa* [TBW], *ʷka:po* [B.].

Исход на велярную аффрикату в записи Д. Блик, скорее всего, фиктивен (современные записи не дают оснований для его выделения).

9) «пить»: хад *fá-* [Sd.], *fuá*, *fuá-ta* [B.].

10) «сухой»: хад *ʷape* [Sd., B.].

11) «ухо»: хад *hačʷapič'i* [Sd.], *háčʷapičì-kò* [TBW], *čʷapičì* [B.].

Только в записях Д. Блик эта основа встречается без первого слога *ha-*, что заставляет подозревать в нем застывший префикс. К дальнейшей этимологизации см. ниже 'лист'.

12) «есть»: хад *séme-* [Sd.], *sama ~ sameta ~ seme ~ seme* [B.].

13) «яйцо»: хад *uʷe-* [Sd.], *uxle ~ uxliaku* [B.].

14) «глаз»: хад *ʷak^hwa-* [Sd.], *ák^hwa-kò* [TBW], *akwa-ko* [B.].

15) «огонь»: *cɔkɔ*- [Sd.], *cɔko* [TBW], *cɔko* [B.].

Судя по словарным данным Б. Сэндс, корень в этом слове двухсложный — ср. *cɔkɔ-ko* 'большой костер, пожар'. Любопытно, однако, что в записях А. Такера второй согласный не имеет абруптивной артикуляции. Если эта ситуация первична (а глоттализация появилась в результате ассимиляции), второй слог может представлять собой исторический суффикс (показатель единичности *-ko*), первый же можно и нужно сопоставлять с *ho=cɔ*-(*ko*) 'пепел' (см. выше) в качестве односложного корня **cɔ*, от которого в «прото-хадза» образовывались обе эти основы.

16) «нога»: хад *ʔapukwa*- [Sd.], *áp^húkwa-kò* ~ *ʔúp^húk^hwa* [TBW], *fukwa* ~ *upukwa* [B.].

Записи Д. Блик снова показывают вероятный старый вариант (*fukwa*) без префикса (*ʔa*=; *ʔu*= по ассимиляции с гласным корня).

17) «волосы»: хад *háλɛ̀*- [Sd.], *háλɛ̀-ko* [TBW].

Как и во многих других случаях, начальный слог *ha-* может быть старым префиксом.

18) «рука»: хад *ʔuk^hwa* [Sd.], *úk^hwá-kò* [TBW], *ukwa*, *ukwa-ko* [B.].

19) «голова»: хад *λóma*- [Sd.], *λóma-kò* [TBW], *klɔma* ~ *ʃkɔma* [B.].

В работе [Kießling, Mous, Nurse 2007: 193] это слово сопоставляется с празападнорифтом **λo.ma* 'гора; висок', причем семантические соображения (метафорическое развитие 'гора' → 'голова' якобы оказывается типологически более вероятным для африканского ареала, чем обратное) заставляют авторов предполагать заимствование из кушитского источника в хадза, а не наоборот.

С таким сценарием трудно согласиться. Во-первых, хадза *λóma*- не имеет явно различимых семантических коннотаций вида 'гора' или 'верхушка', в то время как рифтское слово нигде не значит собственно 'голова'; следовательно, речь идет в лучшем случае об очень старом заимствовании. Но этому противоречит тот факт, что рифтское слово не имеет надежной южнокушитской этимологии (в [Ehret 1980: 216] оно сопоставляется с *ma'ta=łóma* 'жир' ← **λoma* 'нарост, вздутие', но вряд ли эту этимологию можно воспринимать всерьез). Более вероятно, поэтому, что мы имеем здесь дело с субстратным заимствованием в рифт, либо из «старого хадза», либо из каких-то вымерших языков, родственных хадза, в которых этот корень был нагружен побочными коннотациями. Отметим, что никаких других потенциальных

заимствований анатомических терминов в пределах 50-словного списка из западноорифтских языков в хадза не наблюдается.

20) «слышать»: хад //áʔe- [Sd.], //áʔè- [TBW], //ae [B.].

21) «сердце»: (а) хад ʔasači-ko, ʔasaši-ko [Sd.], ášàši-kò [TBW]; (б) хад ηkólo- [Sd.], ηk^hólò-ko [TBW].

Из этих двух корней, различие в употреблении которых неизвестно, более архаичен, скорее всего, первый (ʔasači- или *=sači-, если ʔa- восходит к префиксу); второй явно заимствован из банту (прабанту *=kódò 'сердце', в сочетании с классным префиксом обычно *η=kódò [Guthrie 1967: III, 292]).

22) «рог»: хад loʔo-p^he [Sd.], lɔa-ko, мн. ч. lɔ:-be ~ lɔ:-pe [B.].

23) «я»: хад ono (м. р.), ono-ko (ж. р.) [Sd.], na ~ on ~ ona ~ ona ~ ona ~ ona ~ onə [B.].

В зависимой (суффиксальной) форме это местоимение имеет варианты -na, -naʔa (при глаголе), -ne (при имени) [Sd.], что позволяет выделить в нем консонантную основу *n. Местоименный суффикс 1-го л. -ta (который Дж. Гринберг сопоставляет с t-образными местоименными основами в центр.-кой. языках [Greenberg 1966: 74]) употребляется только в формах настоящего времени; степень его архаичности остается неясной.

24) «убивать»: хад //ó- [Sd.], //ʔo ~ //kɔ ~ //kɔna [B.].

25) «лист»: хад hacápe [Sd.], hácɔp^hɪp^hɪ (мн. ч.) [TBW], hacápe, мн. ч. hišepe [B.].

Форма исторически представляет собой plurale tantum, т. к. заканчивается на показатель мн. ч. -pe; «новое» множественное число образуется от старого либо удвоением показателя (как в записи А. Такера), либо с помощью странного вокалического чередования (как у Д. Блик, хотя записанные у нее варианты не подтверждаются другими исследователями, равно как и возможность постулирования в хадза внутренней флексии).

В свете типичной для африканского региона семантической коннотации «лист : ухо» ('лист' как 'ухо дерева') обращает на себя внимание явное сходство основы *hacá- с основой *[ha]cɔpɪči- ~ *[ha]čɔpɪči- 'ухо' (см. выше). Если первый согласный в корне слова 'ухо' исходно представлял собой шипящий č, сходство это случайно; если, однако, первична свистящая артикуляция (в том виде, в котором слово фигурирует в записи А. Такера: hácɔpɪči-kò), а шипящая является результатом ассимиляции (*hacɔpɪči →

hačʰaričʰi), то 'ухо' следует анализировать как **ha=čʰa-pičʰi*, где первая морфема — застывший префикс, вторая — старый корень с полисемией 'ухо / лист', а третья — какой-то уточняющий корень, позволяющий разграничить между этими двумя значениями; к сожалению, в самостоятельном употреблении форма **pičʰi* нигде не засвидетельствована, что не позволяет надежно верифицировать такой анализ.

26) «вошь»: хад *ʔátacʰi-ya-kʰo* [Sd.], *ʔátacʰi-ko* [TBW].

27) «мясо»: хад *mana-ko* [Sd.], *mànà-kô* [TBW], *mana* ~ *man* [B.].

28) «луна»: хад *sétʰa* [Sd.], *sêta*, *sé:ta* [TBW], *seta* [B.].

Это слово, по всей видимости, является относительно недавним заимствованием из южнонилотского источника, ср. датаго *šê:da* (диалекты баджута, гисамджанга), *šé:ra* (диалекты ротигенга, бурадига), а также омотик *té:ta* [Rottland 1982: 459, 505]; согласно реконструкции Ф. Роттланда (**ʔE-*), первичной для южнонилотского следует считать латеральную артикуляцию, что говорит о направлении заимствования из nilотского в хадза, а не наоборот.

29) «рот»: хад *ʔawanika-* [Sd.], *awanika-ku* [B.].

30) «имя»: хад *akana-be* [B.].

Слово встречается также в полевых записях Д. Элдеркина в виде *ʔakana*. В записях Б. Сэндс слово 'имя' не приводится, а форма *akana-be* переводится как глагол 'звать'.

31) «новый»: хад *zana* [Sd.].

32) «ночь»: хад *çifi-* [Sd.], *çifi* [TBW], *cifu* ~ *cifua* [B.].

33) «нос»: хад *ʔintʰawe-* [Sd.], *intʰà-wà* [TBW], *ndaawe* ~ *ntawe* ~ *ntawa* [B.].

Как и в ряде других случаев, в записях Д. Блик обнаруживается вариант без начального гласного — неясно, идет ли речь о редукции или о диалектной форме без «застывшего префикса».

34) «не»: хад *ʔakw-* [Sd.], *akwá* ~ *akwé* ~ *kwene* [B.].

Слово, судя по описанию Б. Сэндс, представляет собой отрицательную глагольную основу, способную принимать личные окончания (отсюда разнообразие вариантов в словаре Д. Блик).

35) «один»: хад *ʔičáme-* [Sd.], *ičame* ~ *čame* [B.].

Первый слог (*ʔi-*), возможно, имеет префиксальный характер.

36) «дождь»: = «вода», см. ниже.

37) «дым»: хад *çikʷo-wa* [Sd.], *çikʷò* [TBW].

38) «звезда»: хад *çʰa* ~ *ncʰa* [Sd.].

Д. Блик приводит для этого значения форму *maŋwanda*, которая в записях Б. Сэндс имеет вид *magwanda* и переводится как 'название звезды в Западном полушарии' (по-видимому, заимствование из банту). Преназализация отмечается и в записях Д. Элдеркина (*пса*).

39) «камень»: хад *ha!^há-* [Sd.], *há!^hà ~ hán!^hà* [TBW], *han-!^ha* [B.].

Двусложная основа со срединным кликсом; остается неясным, следует ли считать носовую артикуляцию, отмеченную у А. Такера и Д. Блик, сегментной частью основы (т. е. **han!^ha*), или же она просто отражает обязательную преназализацию кликса с исходом на гортанную смычку (ср. детальную фонетическую транскрипцию *haŋ!^há-* у Б. Сэндс).

40) «солнце»: хад *?išo-* [Sd.], *išo ~ išo-ko ~ išo-wa* [B.].

41) «хвост»: хад *cahó-* [Sd.], *càhò-kò* [TBW], *caho ~ cau* [B.].

42) «ты»: хад *te* (м. р.), *te-ko* (ж. р.) [Sd.], *te* [B.].

В зависимой форме это местоимение имеет вид *te*, *ti* (при имени), *ta*, *ta^ha* (при глаголе), что позволяет видеть в качестве общего «костяка» консонантную основу **t-*.

43) «язык»: хад *ǀát^hâ-* [Sd.], *ǀátà* [TBW], *ǀata-ku* [B.].

44) «зуб»: хад *?áhà-* [Sd.], *áhà-bê?* (мн. ч.) [TBW], *aha-pi* (мн. ч.) [B.].

45) «дерево»: хад *cítì-* [Sd.], *cítì-ko* [TBW], *ceti* [B.].

46) «два»: хад *píye-* [Sd.], *pie ~ pie-be ~ pie-pi ~ pie-pe* [B.].

Для этого слова высока вероятность заимствования из банту (ср. прабанту **-bàdí* 'два'), хотя однозначно определить точный источник заимствования сложно (в ряде языков банту действительно имело место развитие **-d-* → *-y-*, но к ближайшим соседям хадза оно не относится).

47) «вода»: хад *?át^hi-* [Sd.], *átì* [TBW], *ati ~ ate ~ ati:ye* [B.].

Это же слово употребляется и в значении «дождь».

48) «мы»: хад *?ubi?i* (м. р.), *?obe?e* (ж. р.) [Sd.], *obe ~ ube ~ obi ~ ubi* [B.].

Д. Блик отмечает существование вариантов с *o-* и *u-*, но не соотносит их с категорией рода. Для правильного выделения корня необходимо сопоставление с другими формами этой же местоименной парадигмы, ср. у Б. Сэндс: *?itibi* 'вы (м. р.)', *?etebe* 'вы (ж. р.)', *bi?i* 'они (м. р.)', *be?e* 'они (ж. р.)'. Видно, что во всех этих лексемах выделяется элемент *-b-*, отвечающий за значение мн. ч. (и, очевидно, совпадающий с показателем мн. ч. у обычных имен), а также суффиксальные родовые показатели *-i* (м. р.) и *-e* (ж. р.).

Следовательно, основную корневую нагрузку в формах $\text{?}u\text{-}b\text{-}i\text{?}i$, $\text{?}o\text{-}b\text{-}e\text{?}e$ несет начальный губной гласный, а артикуляторное различие по подъему легко объясняется ассимиляцией с гласным показателя рода. Учитывая данные зависимых посессивных форм (м. р. $\text{?}u\text{-}bi\text{-}\check{c}e\text{-}ya$, ж. р. $\text{?}u\text{-}bi\text{-}te\text{-}ya$; ср. от местоимения 2-го л. мн. ч. — м. р. $\text{?}iti\text{-}bi\text{-}\check{c}e\text{-}ya$, ж. р. $\text{?}iti\text{-}bi\text{-}te\text{-}ya$), логично предположить, что исходной основой было $*u\text{-}$.

Зафиксированы также инклюзивные формы, расширенные за счет включения в состав лексемы дополнительной морфемы $-ne\text{-}$ / $-ni\text{-}$: $\text{?}u\text{-}ni\text{-}b\text{-}i\text{?}i$ м. р., $\text{?}o\text{-}ne\text{-}b\text{-}e\text{?}e$ ж. р. Данную морфему соблазнительно объяснять как зависимый вариант местоимения ед. ч. 'я' (см. выше), но это было бы типологически странным (скорее ожидалась бы зависимая форма местоимения 2-го л.).

49) «что»: (а) хад $\text{?}aku \sim \text{?}akwa\check{z}a$ [Sd.]; (б) хад $ta\text{-}$ [Sd.].

50) «кто»: (а) хад $\text{?}aku \sim \text{?}akwa\check{z}a$ [Sd.]; (б) хад $ta\text{-}$ [Sd.].

Две вопросительные основы — велярная $*(\text{?}a)ku\text{-}$ и дентальная $*ta\text{-}$ — которые приводит Б. Сэндс, распределены не по принципу одушевленности/неодушевленности, а более сложным образом, требующим дополнительного изучения. В значении 'что', по видимому, вариантом по умолчанию следует считать скорее велярную основу ($ta\text{-}$ чаще всего появляется в составе сложной глагольной основы $ta\text{-}\check{s}i\text{-}$ 'делать что?'), а в значении 'кто' велярная и дентальная основа практически взаимозаменяемы, ср. примеры Сэндс: $ta\text{-}o\text{-}ne\text{?}e$ 'кто я?', $\text{?}akwa\check{z}a\text{-}neko$ 'кто я?' (ж. р.).

2.8. Предварительные выводы.

2.8.1. *Критерии оценки.* В заключение данного раздела исследования уместно предложить предварительную оценку степени генетической близости «койсанских» языков, т. е. провести *предварительный анализ прасписков 1-го уровня*, о котором шла речь в разделе 7.5 вводной части, пока что без привлечения данных по другим таксонам 1-го уровня, локализованным на африканском континенте.

Здесь, на этапе как ручной, так и автоматической обработки реконструкций, полученных для элементов 50-словного списка, мы вплотную сталкиваемся с проблемой, уникальной для «койсанских» языков — вопросом о том, каким образом можно

встроить систему щелчковых согласных в множество выделенных А. Б. Долгопольским и С. А. Старостиным консонантных классов. *A priori* можно думать о двух подходах:

а) как внутри конкретных фонологических систем «койсанских» языков, так и в рамках абстрактно-универсальной системы фонологических оппозиций «кликсы» на самом деле скоррелированы с обычными эгрессивными согласными. Грубо говоря, например, дентальный кликс / можно рассматривать как «щелчковый вариант» обычного дентального согласного *t*, отличающийся от него не «фундаментально», а всего лишь способом образования, точно так же, как отличаются от *t* звонкий *d*, глоттализированный *tʰ* и т. п. Такой подход был бы идеологически близок к позиции тех современных койсанологов, которые придерживаются мнения о «неэкссклюзивности» койсанской фонетики;

б) обратное: кликсы представляют собой совершенно автономную систему, члены которой не имеют никаких систематических связей с элементами нещелчковой (эгрессивной) подсистемы фонологического инвентаря (хотя отдельные исторические переходы кликсов в «не-кликсы» и наоборот допустимы, аналогично тому, как допустимы отдельные переходы «обычных» согласных из одного консонантного класса в другой). Такой подход, наоборот, подчеркивает уникальность койсанской фонетики: если щелчковые согласные, по большому счету, не коррелируют ни с какими «обычными» элементами консонантного инвентаря, это может считаться существенным теоретическим аргументом в пользу свержаичности кликсов и, соответственно, глубочайшей древности отделения «общекойсанского» языкового континуума от прочих языков Африки (а, возможно, и всего мира).

Впрочем, на том этапе сопоставления, когда к данным «койсанских» языков еще не подключены данные по другим языковым семьям, различие между этими двумя подходами не столь принципиально: полномасштабно оно проявляется лишь в ситуации, когда язык или языковую группу, насыщенную «кликсованной» лексикой, требуется сравнить с языками, в которых кликсы отсутствуют как класс. *Внутри* же «койсанского» континуума очевидно, что лексемы, содержащие кликсы, будут в первую очередь сопоставляться с «кликсованными» же лексемами, и наоборот — за исключением разве что сопоставлений между языками с

богатой системой щелчковых (три основные семьи + †хоан) и относительно бедной (квади, сандаве, хадза), где действительно имеет смысл озаботиться вопросом о том, какие «обычные» согласные могли бы соответствовать тем или иным типам щелчковой артикуляции.

Поскольку вопрос о генезисе как сложных, так и простых систем щелчковых согласных пока что далек от решения, на этапе автоматической обработки имеет смысл предложить по меньшей мере *три* различных способа анализа материала, после чего сопоставить полученные результаты на предмет выявления сходств и расхождений. Подчеркнем, что речь пока что идет исключительно об *основах* кликсов; методика анализа *исходов* будет описана ниже, чуть более кратко.

(1) Для всех известных основ кликсов (четыре универсальных и двух редких) предлагаются конкретные корреляции среди «обычных» консонантных классов, исходя в первую очередь из тех немногочисленных сведений о типологии эволюции кликсов, которыми мы обладаем (непременное условие — забыть о традиционно принятых *названиях* основ кликсов, которые изначально носили сугубо условный характер, а в отдельных ситуациях могут ввести в серьезное заблуждение относительно истинных артикуляторных характеристик той или иной фонемы):

(1.1) Т. н. *дентальные* кликсы (/ и производные) могут быть включены в обширный класс *переднеязычных аффрикат и сибилантов* (С). Важнейший аргумент — эксплицитное наличие такой корреляции в языковой группе кхойкхой, где некоторые працентральнокойсанские аффрикаты регулярно развиваются в дентальные кликсы (см. выше: **сани* 'дым' → кхойкхой **/xani* и др.). Некоторые случаи возможного развития дентальных кликсов из слов, ранее содержавших аффрикаты, приводятся в [Starostin 2008a] на материале семантически точных параллелей между сандаве и другими койсанскими языками.

(1.2) Т. н. *альвеолярные* кликсы (! и производные) скорее должны коррелировать с *велярными* согласными («класс К»), чем с альвеолярными, т. к. наиболее типологически частотный механизм их развития — выпадение основы кликса с последующим переходом сопровождающего его исхода в разряд велярных

согласных (см. выше для центральнокойсанских языков такие развития, как $*!- \rightarrow k-$, $*!- \rightarrow g-$, $*\tilde{!}- \rightarrow \eta-$ и т. д.).

(1.3) Т. н. *палатальные* кликсы ($\#$ и производные), на первый взгляд, должны коррелировать с палатальными (среднеязычными) аффрикатами — хотя бы потому, что в случае утраты щелчкового типа артикуляции эти согласные, как правило, дают палатальные рефлексы (см. опять-таки в первую очередь рефлексацию в центральнокойсанских языках, где $*\#- \rightarrow \text{ç-}$ и т. п.). Однако в собственно койсанских языках отдельный палатальный ряд нещелчковых согласных либо отсутствует совсем, либо оказывается результатом относительно недавней палатализации дентальных ($t-$, $d-$, как в западном !xoaŋ , отдельных центральнокойсанских языках и т. п.) или велярных ($k-$, $g-$, как в ряде языков группы !kvi) согласных. Поскольку с велярным рядом в нашей системе уже скоррелированы альвеолярные кликсы, палатальный ряд логичнее включить в тот же класс, в который уже включены переднеязычные взрывные (Т). На это же косвенно намекают и отдельные любопытные сопоставления между центральнокойсанскими словами, содержащими палатальные кликсы, и словами в сандаве, начинающимися с дентальных согласных, которые приводятся в работе [Starostin 2008a].

(1.4) Т. н. *латеральные* кликсы ($\|$ и производные) на уровне отдельных койсанских групп не обнаруживают заметных корреляций ни с одним типом «обычных» согласных, что может быть связано с отсутствием (а в исторической перспективе — возможным исчезновением в ходе слияния с другими артикуляторными типами) в этих группах особого класса латеральных аффрикат и спирантов. Однако такой класс надежно засвидетельствован в хада и сандаве, и, более того, обнаруживаются отдельные случаи возможных соответствий между конкретными лексемами в сандаве, содержащими латеральные аффрикаты/фрикативные, и центральнокойсанскими основами, начинающимися на латеральные кликсы («классический» пример — сандаве !ana 'рог': центр.-кой. $*\|ā$ id). Исходя из этого, уместнее всего относить латеральные кликсы к латеральному же консонантному классу (L), включающему аффрикаты и фрикативные (но не плавные спиранты).

(1.5) Редкий подтип *лабиальных* кликсов (θ и производные), очевидно, следует объединять с классом губных смычных (P) —

хотя, забегая вперед, можно сразу сказать, что продуктивных результатов от такого объединения ожидать не придется: сам по себе класс губных эгрессивных смычных в койсанских языках представлен чрезвычайно скудно, за исключением опять-таки сандаве и хадза. В одной из наших статей [Старостин 2006], на основании негативных результатов сопоставления лексики, содержащей губные кликсы, в южнокойсанских языках и в западном ꠄхоан, была даже высказана гипотеза о возможности относительно недавнего вторичного образования этих фонем и в том, и в другом таксоне (например, под влиянием губного вокализма корня). Поскольку, однако, на данном этапе у нас нет твердых представлений о наиболее типичных предках губных кликсов, остается лишь, пусть и сугубо условно, отнести их к «обычному» классу губных согласных.

(1.6) Наконец, свехредкий класс ретрофлексных кликсов, засвидетельствованный только для северно-койсанской группы, можно объединить с латеральным классом (L), т. к., судя по сопоставительному материалу в [Starostin 2008a], и в южно-, и в центральнокойсанских языках слова с ретрофлексными кликсами чаще всего обнаруживают фонетически сходные параллели с латеральными кликсами (оставляя пока что за рамками обсуждения вопрос о том, являются ли эти сходства генетическими или ареальными).

(2) Альтернативный подход ориентирован на полную автономность систем щелчковых согласных: корреляции кликсов с «не-кликсами» при нем считаются, если и не полностью исключенными, то, по крайней мере, нетипичными и недопустимыми в рамках определения фонетической созвучности по методу консонантных классов. Здесь имеет смысл проводить анализ двумя различными способами:

(2.A) Все основы кликсов помещаются в единый класс «щелчковых» (обозначим его, например, латинской буквой Q), внутри которого возможен «свободный» переход от одного типа щелчковой артикуляции к другой. Это довольно радикальный и маловероятный метод, который, очевидным образом, приведет к обнаружению большого количества псевдо-когнатов между отдельными «койсанскими» семьями; тем не менее, учитывая, что (а) переходы кликсов от одного типа артикуляции к другой действительно

известны, но при этом (б) общая типология диахронических изменений в рамках подсистемы кликсов находится пока в зачаточном состоянии — его вполне допустимо испробовать наряду с другими, более детализированными подходами;

(2.Б) Каждая из основ кликсов помещается в самостоятельный класс, с запретом на переходы от одного типа артикуляции к другому. Этот подход также в некотором смысле радикален. Все попытки предварительного этимологического анализа «койсанского» материала на сколь-либо глубоких хронологических уровнях (Г. Хонкен, Г. Старостин и др.) показывают, что такие переходы более чем реальны, и что конкретные успехи праязыковых реконструкций на таких уровнях будут во многом зависеть от того, насколько тщательно удастся в дальнейшем определить типологию этих переходов. Напротив, полагаясь исключительно на одно-однозначные соответствия (как делает, например, К. Эрет в своих койсанских сопоставлениях), мы рискуем во многих случаях принять за генетические связи результаты недавних контактов между отдельными семьями; для машинного же анализа вероятность допущения таких ошибок близка к ста процентам. Следовательно, и у такого подхода нет явных и очевидных преимуществ над перечисленными выше, т. е. результатам такого автоматического анализа не следует отдавать явного преимущества над результатами предыдущих двух.

Не менее сложной является ситуация с *исходами* кликсов. Она напрямую связана с вопросами фонологического анализа щелчковых подсистем в «койсанских» языках, в первую очередь — с вопросом допустимости анализа отдельных или даже всех типов сложных кликсов не как двухфокусных согласных, а как консонантных кластеров. В современной койсанистике более популярным является первый подход; на наш взгляд, однако, это связано с тем, что *фонетические* исследования на базе койсанских языков сегодня гораздо более развиты, чем *фонологические*, не говоря уже об *историко-фонологических*.

При этом, однако, и в чисто фонетических исследованиях последнего времени прослеживается определенная тенденция к разграничению между несколькими, принципиально отличными друг от друга типами образования кликсов. Так, в работе [Miller et al. 2009] предлагается вообще отказаться от «устаревшего» термина

«исход кликса» (англ. *click accompaniment*) и различать вместо этого т. н. «язычные» (*lingual*), «легочно-язычные» (*linguo-pulmonic*) и «гортанно-язычные» (*linguo-glottalic*) кликсы. В рамках данной классификации, иллюстрируемой авторами на материале южно-койсанского языка нју:

— подсистема «язычных» кликсов оказывается более или менее параллельной подсистеме «обычных» согласных, включая кликсы с нулевым исходом (= «обычные» глухие согласные), с озвончением (= звонкие согласные), с исходом на придыхание (= придыхательные согласные), гортанную смычку (= глоттализированные согласные) и с назализацией (= носовые / преназализованные согласные);

— подсистема «легочно-язычных» кликсов включает фонемы, в которых традиционно выделялись исходы на увулярную смычку (*/q* и т. д.) и на велярный фрикативный (*/x* и т. д.), с или без дополнительных фонетических признаков (озвончение, придыхание, назализация). В образовании этих кликсов последовательно задействованы два механизма — язычный (лингвальный), т. е. собственно «щелчковый», и легочный (пульмонический), типично приущий «обычным» (нещелчковым) согласным;

— подсистема «гортанно-язычных» кликсов включает фонемы, в которых традиционно выделялся исход на велярную (глоттализированную) аффрикату (*/kx* и т. д.) с или без дополнительного фонетического признака (озвончения).

В фонологических терминах эту классификацию можно интерпретировать как разделение щелчковых согласных на два класса. В первый попадают «простые» согласные, различающиеся по ряду дистинктивных признаков (глухость / звонкость / глоттализированность, аспирация, назализация). Во второй — «сложные» согласные, которые можно анализировать как сочетания «простых» кликсов с «простыми» заднеязычными взрывными, фрикативными и аффрикатами, аналогичные по устройству таким «нещелчковым» кластерам, хорошо зафиксированным в койсанских языках, как *tx*, *tkx*, *cx*, *ckx* и др. В историческом плане для слов, начинающихся с таких кластеров, уместно предположить (пусть даже скорее умозрительно, чем основываясь на конкретных этимологиях) происхождение из структур типа *CVC(V)* с редукцией начального гласного и последующим образованием консонантного сочетания.

Исходя из этих соображений, для процедуры автоматического анализа будет принят следующий подход:

(а) «простые» («язычные») кликсы будут анализироваться как один согласный; все язычные кликсы, имеющие одну и ту же основу, будут включаться в один и тот же класс (т. е. будет считаться, что l, l', l^h, l' находятся друг с другом в таких же отношениях, как, например, t, d, t^h, t');

(б) «сложные» («легочно-» и «гортанно-язычные») кликсы будут анализироваться как состоящие из двух согласных: «простого» кликса, относимого к соответствующему типу основы, и заднеязычного согласного-исхода, который, очевидным образом, может быть отнесен только к заднеязычному классу (K);

(в) *исключением* из (а) будет являться трактовка назализованных и преназализованных кликсов как представляющих «сложные» образования — сочетания «простого» кликса с носовым «квази-согласным» (N). Это связано с тем, что и для простых согласных мы принимаем разграничение между носовыми и неносовыми звуками одного места образования (т. е. t и n , p и m , k и η соответственно попадают в разные классы). В отличие от «звонкости», «придыхательности» и «глоттализованности», «назализация» применительно к консонантному инвентарю все же обычно не является дополнительным фонетическим признаком такого же характера, особенно в историческом плане. И, хотя некоторые этимологические сопоставления показывают, что назальная артикуляция кликсов иногда может рассматриваться как возникающая вторично, под воздействием носовой артикуляции соседнего гласного [Starostin 2008a: 399], таких случаев относительно немного; гораздо более частотны ситуации, в которых носовой исход кликса «неустраим» даже на историко-фонологическом уровне, а, если сюда добавить еще такие случаи соответствий, как упомянутый в (1.4) 'рог' в центральнокойсанских языках и сандаве, автоматический анализ «консонантного костяка» последовательности $\int a$ как [CN] оказывается резонным.

2.8.2. *Результаты автоматических подсчетов по методу 1* (= основы кликсов коррелируют с классами «обычных» согласных). Как для этих, так и для всех последующих подсчетов использовались 50-словные прасписки для прасевернокойсанского (\approx 6 в. н. э.), пра-!кви (\approx 1 в. до н. э.), пра-таа (\approx 5 в. н. э.), пра-кхойкхой (\approx 2 в. н. э.),

пракалахари-кхой (\approx 2 в. н. э.), а также живых или вымерших языков-изолятов: восточный ꜥхоан, квади, сандаве и хадза. Автоматическая расстановка «псевдо-когнаций» по первому из предложенных методов позволяет получить следующую лексикостатистическую матрицу:

	ꜥХоан	!Кви	Таа	Кхойкхой	Калахари	Квади	Сандаве	Хадза
Сев.-кой.	33%	12%	10%	10%	12%	24%	12%	2%
ꜥХоан		2%	2%	2%	6%	3%	6%	4%
!Кви			49%	8%	12%	15%	10%	6%
Таа				10%	14%	9%	10%	12%
Кхойкхой					52%	29%	12%	10%
Калахари						38%	12%	4%
Квади							12%	0%
Сандаве								8%

Прежде чем переходить к другим методам, соотнесем эти показатели, в порядке убывания, с текущими представлениями о возможных генетических и ареальных отношениях «койсанских» семей.

а) На первом месте по числу лексических совпадений (52% = 26 «когнатов», из них этимологически корректных — 25) стоит центральнокойсанская группа/семья, родство между двумя основными ветвями которой (кхойкхой и калахари-кхой) убедительно продемонстрировано в рамках сравнительно-исторического метода.

б) 49% имеют между собой южнокойсанские подгруппы !кви и таа, родство между которыми также не вызывает сомнений у большинства койсанологов (алгоритм обнаруживает 24 «когната», из них 23 соответствуют этимологическим ожиданиям).

в) На следующем этапе, с 38% совпадений, объединяются языки калахари-кхой и квади (меньший процент между кхойкхой и квади — случайность, вызванная несколько большей фонетической инновативностью языков кхойкхой). Из 13 автоматически определенных «когнатов» 12 не вызывают сомнений и согласуются с лексико-этимологическим анализом в работе [Güldemann & Elderkin 2003], подтверждая реальность генеалогического узла «кхой-квади», постулированного авторами.

г) Наконец, последний «крупный» процент соответствий показывают между собой северно-койсанские языки и восточный

ɬхоан — 33% (16 совпадений, из них 14 не вызывают существенных возражений). Этот результат хорошо ложится на гипотезу о специфически близком родстве между этими двумя таксонами (сравнительно-исторические данные в поддержку которой уже публиковались в работах Г. Старостина, Г. Хонкена и др.).

д) Обращает на себя внимание и неожиданно высокий результат, полученный между северно-койсанскими языками и квади (24%). Здесь, правда, из восьми потенциальных «когнатов» четыре вызывают определенные сомнения, т. к. представляют собой отождествления лексем, содержащих кликсы в северно-койсанском и «обычные» согласные в квади или наоборот: сев.-кой. **laʔa* 'глаз' = квади *ši*, **daʔa* 'огонь' = квади *ɬē* и т. д. С другой стороны, наличие таких корреляций, как сев.-кой. **ʒo* 'черный' = квади *ʒu*, а также наличие непровержимых параллелей в сфере местоимений наводит на мысль о том, что данный «подскок» может являться результатом переплетения сразу трех факторов — случайных совпадений, явлений ареально-контактного характера и следов глубокого (а не специфически «жу-квади») родства.

е) Во многих случаях число схождений колеблется от 8% до 14-15%, что в свете статистических данных аналогичных подсчетов по языковым семьям Евразии (см. раздел 1.7.5 вводной части), в рамках которых такие цифры обычно соответствуют надежно установленным ситуациям языкового родства, могло бы рассматриваться как «позитивный» результат. Однако не следует забывать, что в данной ситуации мы все же используем особо «дозволенную» разновидность метода, позволяющую отождествлять «кликсы» с обычными согласными и, тем самым, способствующую приросту случайных схождений. Помимо этого, довольно большое число опознанных «когнатов» моносиллабично, что также снижает надежность и без того ненадежного метода (скажем, сев.-кой. вопросительные местоимения с основой *ha-* опознаются как родственные сандаве вопросительным местоимениям с основой *ho-*; этимологическая «подпитка» такого сопоставления в обязательном порядке требовала бы объяснения различия в вокализме, т. к. иначе этимология держалась бы исключительно на общей структуре корня и начальном ларингале).

Помимо этого, сомнительным выглядит тот факт, что результаты сопоставлений выявляют дистрибуционные нарушения: ср.,

в частности, данные по хадза, где метод дает 12% схождений с одной южнокойсанской подгруппой (таа), но всего 6% с другой (!кви), или 10% с языками кхойкхой, но всего 4% с языками калахари-кхой (2 схождения, из которых перспективным выглядит только фонетическое сходство слов со значением 'хвост').

Таким образом, на данном этапе резонно было бы признать неслучайность результатов (а)-(г), на порядок отличающихся от всех остальных и к тому же подтверждающих результаты уже проведенных этимологических исследований, «принять к сведению» результат (д) и отказаться от использования в нашей рабочей модели всех прочих результатов до тех пор, пока они не будут сопоставлены с результатами анализа по двум другим методам разбиения на консонантные классы.

2.8.3. *Результаты автоматических подсчетов по методам 2.А (= основы кликсов не коррелируют с «обычными» согласными, все типы основ кликсов помещаются в один класс) и 2.Б (= каждый тип основ кликсов составляет самостоятельный класс).*

Автоматическая расстановка когнатов по методу 2.А дает результаты, существенно расходящиеся с представленными выше:

	‡Хоан	!Кви	Таа	Кхойкхой	Калахари	Квади	Сандаве	Хадза
Сев.-кой.	37%	29%	26%	12%	20%	15%	18%	0%
‡Хоан		17%	14%	2%	8%	3%	6%	0%
!Кви			49%	20%	24%	24%	12%	4%
Таа				16%	24%	15%	14%	6%
Кхойкхой					52%	29%	14%	6%
Калахари						47%	16%	2%
Квади							18%	0%
Сандаве								6%

В целом эти изменения предсказуемы: налицо резкое увеличение числа «когнатов» между языками со сложными системами кликсов, встречающихся в начальной позиции как минимум в половине всех элементов списка, и, наоборот, некоторое снижение числа «когнатов» между этими языками и языком хадза, в котором щелчковые согласные играют несколько меньшую роль. Так, если в первой версии подсчетов нуль процентов совпадений наблюдался только между хадза и квади, то теперь нулевые

И здесь также «непоколебимыми» остаются только выводы (а), (б), (в), (г): очевидно, что связи между сев.-кой. языками и ɬхоан, отдельными ветвями южно- и центральнокойсанских групп, а также центральнокойсанскими языками и квади ни одна из предложенных методик «порвать» не в состоянии, чего нельзя сказать о других подскоках процентных соотношений, существенно зависящих от того, какому из трех методов мы отдаем предпочтение.

Общую картину можно суммировать следующим образом:

(а) автоматический анализ уверенно сводит *девять* исходных узлов нижнего уровня, предложенных к объединению в рамках единого древа, к *пяти* узлам «среднего уровня»: (1) сев.-кой. + ɬхоан; 2) юж.-кой.; 3) центр.-кой. + квади; 4) сандаве; 5) хадза;

(б) во *всех* трех предложенных матрицах обращает на себя внимание среднестатистический процент, который демонстрирует с остальными языками хадза — существенно более низкий, чем проценты схождения между любыми другими «узлами среднего уровня». Это заставляет предположить, что даже «слабые» (на уровне 8-15% сходств) связи, которые представленные три разновидности алгоритма обнаруживают между 4 из 5 узлов «среднего уровня», все же не целиком случайны и либо отражают какие-то достаточно древние отношения родства, либо старые языковые контакты, либо смесь того и другого.

2.8.4. *Результаты, полученные методом «ручной коррекции».* Теперь, получив и сравнив между собой несколько результатов автоматических подсчетов, попробуем получить предварительное «ручное» дерево, несвободное от мелких элементов субъективности, но в целом опирающееся на логически и эмпирически обоснованную коррекцию результатов автоматической идентификации когнатов.

Первый опыт «ручного» построения общекойсанского древа на основании фонетических соответствий, установленных «на глазок», был в свое время опубликован нами в работе [Starostin 2003]; некоторые из заподозренных нетривиальных соответствий впоследствии были подтверждены в более подробном исследовании [Starostin 2008a] (хотя многие из них до сих пор нельзя считать «строго доказанными» с точки зрения стандартного применения сравнительно-исторического метода). Однако на данном этапе, в ходе «ручной коррекции» матрицы, полученной автоматически,

мы будем ориентироваться *исключительно* на параметры, описанные в разделе 1.7.5, т. е. скорее на исправление бросающихся в глаза ошибок автоматического анализа, чем на попытку тщательного определения фонетических соответствий между узлами, родство которых остается под существенным сомнением; последнее, по сути, относится уже к дальнейшему этапу работы.

«Ручная коррекция» будет заключаться в следующем.

1) Из трех матриц, полученных в ходе автоматического анализа, мы выбираем *третью*, т. е. опирающуюся на наименьшее количество непроверенных презумпций. Поскольку у нас нет твердых гарантий ни того, что «конкретные типы кликсов регулярно соответствуют конкретным типам нещелчковых согласных», ни того, что «разные типы кликсов свободно переходят друг в друга», здравомыслящему лингвисту-компаративисту на *первом* этапе анализа не следует включать в свою модель много-многозначные соответствия, если только они не подтверждаются «здесь и сейчас» самым наглядным образом — так, чтобы для этого хватало, например, одного только материала стословного списка.

2) «Нетривиальные» соответствия, т. е. такие, в которых представлен переход фонемы из одного класса в другой, учитываются между теми узлами, родство которых (а) наглядно продемонстрировано в компаративистской литературе, (б) столь же наглядно определено в рамках каждой из трех автоматических матриц. Так, например, соответствие «кхойкхой */kxan : калахари-кхой */cʰan 'дым'» при построении третьей матрицы не было опознано из-за запрета на корреляции между кликсами и не-кликсами; «ручная коррекция» разрешает его восстановить в силу того, что это — регулярное фонетическое соответствие, надежно установленное для центральнокойсанской семьи.

3) Как «позитивные» отмечены несколько типов потенциальных «нетривиальных» соответствий, в которых задействованы свойственные одному или нескольким узлам *редкие* типы фонем или фонемных позиций, не имеющие прямых аналогов в остальных узлах. В частности:

— язык западный Ꞥхоан, по всем возможным методам подсчета тяготеющий к сев.-кой. языкам, отличается от последних наличием губных кликсов, причем даже в пределах 100-словного списка рекуррентной оказывается корреляция между этими

кликсами и дентальными кликсами в сев.-кой. языках (ср. в 50-словных таблицах 'глаз', 'голова', 'один'). Для таких случаев «ручная коррекция» допускает возможность когнации;

— бросается в глаза возможная связь между сев.-кой. *ʔm 'есть', *mi ~ *ma 'я' и юж.-кой. *ʔā (если ← *ʔη) 'есть', *η 'я', исторически вполне реальная вследствие позиционных ограничений на заднеязычный носовой в сев.-кой. Эти два сходства также можно пометить как совпадения;

— ретрофлексные кликсы в сев.-кой. языках, не имеющие прямых фонетико-фонологических аналогов в других семьях, допускаются к сопоставлению с теми классами кликсов, в которые они сами могут переходить в сев.-кой. диалектах, т. е. с альвеолярными и латеральными (таким образом, например, пра-сев.-кой. *ʔli 'луна' и калахари-кхой *ʔbó id. получают один и тот же индекс);

— отмеченное выше сходство между сандаве *ʔana* и центр.-кой. *ʔa 'рог' также отмечено как потенциальный когнат в силу своей «идеальности»: в отличие от других гипотетических случаев перехода не-кликсов, засвидетельствованных в лексике сандаве и хадза, в кликсы в центр.-кой. и других языках, где соответствия обычно являются неполными (скажем, остается непонятным генезис исхода кликса, вокализм и т. п.), для данного случая весь сценарий развития оказывается наглядным и тривиальным: переход латеральной аффрикаты в латеральную же основу кликса и развитие носового согласного в носовой же исход. (Другой возможный случай вторичного развития щелчковой артикуляции, представленный корреляцией сандаве *keke* : центр.-кой. *ʔe, мы в число потенциальных когнатов не включаем, т. к. механизм возможного превращения велярного согласного в палатальный кликс остается неясным — разве что генезис щелчковой артикуляции попутно сопровождался еще и палатализацией перед передним гласным, но это усложняет презумпцию настолько, что без дополнительных этимологических подтверждений постулировать здесь когнацию неправомерно даже на сугубо предварительном уровне).

Все эти коррекции проделаны либо на основании тщательно выполненного и задокументированного сравнительно-исторического исследования (например, работ Р. Фоссена по центрально-койсанской реконструкции), либо на основании непротиворечивых

корреляций в рамках 50-словного списка, для обнаружения и исторического обоснования которых не требуется подключение дополнительного этимологического материала. Напротив, «предварительная» нетривиальная система соответствий между, например, сев.- и юж.-койсанскими языками, описанная в работе [Starostin 2008a], в рамках которой одному сев.-кой. кликсу могут соответствовать несколько юж.-кой. и наоборот, в ходе «ручной коррекции» сознательно не была отражена, т. к. обнаружение и проверка таких соответствий — задача уже *следующего* этапа работы.

Результирующая лексикостатистическая матрица, отталкивающаяся от автоматического метода 2.Б и прошедшая ручную проверку, имеет следующий вид:

	‡Хоан	!Кви	Таа	Кхойкхой	Калахари	Квади	Сандаве	Хадза
Сев.-кой.	41%	14%	12%	8%	10%	9%	4%	0%
‡Хоан		8%	6%	2%	6%	0%	4%	0%
!Кви			69%	10%	12%	14%	6%	2%
Таа				10%	12%	12%	8%	2%
Кхойкхой					62%	38%	8%	6%
Калахари						41%	12%	2%
Квади							15%	3%
Сандаве								6%

Существенные отличия от матрицы 2.Б и других автоматизированных матриц можно суммировать следующим образом:

а) за счет дополнительных заподозренных корреляций, не установленных автоматическими методами подсчета, несколько повысилось число параллелей между сев.-кой.-‡хоан и юж.-кой. узлами, при том, что число параллелей между этими узлами и центр.-кой. осталось неизменным или даже слегка занизились;

б) существенно снизилось число схождений между сев.-кой. языками и сандаве, в то время как схождения между центр.-кой. узлом и сандаве, напротив, возросли;

в) катастрофически снизилось число схождений между хадза и большинством остальных узлов (до нуля процентов в двух из восьми возможных парных сопоставлений).

При *условном допущении* реальности «общекойсанского» или «макрокойсанского» таксона представленная матрица, обрабо-

танная методом «ближайших соседей», должна трансформироваться в следующую «древовидную» классификацию:

1. Хадза.
2. Макрокойсанские «пост-хадза»:
 - 2.1. Сандаве-кхой-квади:
 - 2.1.1. Сандаве.
 - 2.1.2. Кхой-квади:
 - 2.1.2.1. Квади.
 - 2.1.2.2. Центральнокойсанские (= кхой):
 - 2.1.2.2.1. Кхойкхой.
 - 2.1.2.2.2. Калахари-кхой.
 - 2.2. Периферийно-койсанские:
 - 2.2.1. Южнокойсанские (= !кви-таа ~ туу):
 - 2.2.1.1. !Кви.
 - 2.2.1.2. Таа.
 - 2.2.2. Жу-ѳхоан:
 - 2.2.2.1. Западный ѳхоан.
 - 2.2.2.2. Севернокойсанские (= жу).

Учитывая, однако, что процентные совпадения для многих узлов на глубоких хронологических уровнях ничтожно малы и едва ли отличимы от случайных, результаты классификации делятся на три группы:

(а) *достоверные*: схождения исчисляются двумя и более десятками процентов, что фактически исключает как случайность (даже если определить ее возможный порог десятью процентами — числом, на практике ни разу не отмеченным для 50-словников по основным языковым группам Евразии), так и «ареальность». Это узлы 2.1.2 («кхой-квади»), 2.2.1 (!кви-таа) и 2.2.2 (жу-ѳхоан). Для этих таксонов уже в пределах 50-слового списка можно наметить отдельные фонетические соответствия; на следующем (этимологическом) этапе исследования число опознанных когнатов должно будет увеличиться за счет опознания дополнительных, менее тривиальных соответствий;

(б) *тестируемые*: схождения ограничиваются небольшой группой в 10-12%, что не исключает случайность или ареальность, но должно, тем не менее, являться стимулом к попыткам

дальнейшего этимологического сравнения, т. е. к «ручному» определению не опознанных автоматически когнатов и нетривиальных соответствий на основании привлечения к сравнению дополнительного материала. «Тестируемый» статус присваивается таким узлам, как периферийно-койсанский и сандаве-кхой-квади (напомним еще раз, что обе гипотезы уже были предложены, на основании подмеченных лексических и грамматических регулярностей, соответственно Г. Старостиным и Д. Элдеркином);

(в) *нейнтерпретируемые*: узлы, определенные «древостроительным» алгоритмом чисто формально, на основании единичных процентов совпадений, неотличимых от случайностей. В нашем случае это узлы 1 и 2, т. е. как раз те, признание реальности которых — обязательное условие подтверждения «макрокойсанской» гипотезы.

В добавление к основным подсчетам «прочность» выделенных узлов можно также оценить, применив к ним критерий «динамической градации»; для этого нужно разбить тестовый список на две половины, в первую из которых попадает более устойчивая (по «индексу С. А. Старостина») базисная лексика, и проанализировать соотношение процентных совпадений по обеим частям списка. Здесь картина оказывается следующей (данные приводятся только для тех бинарных пар, общее число совпадений в которых равно или превышает 10%, т. е. 5 общих слов; «динамическая градация» на меньшем числе сходств, по-видимому, вообще утрачивает какой-либо смысл; в ячейках таблицы первое число указывает кол-во совпадений в первых 25 «сверхстабильных» элементах списка, второе — общее число совпадений по всему 50-словнику):

	‡Хоан	!Кви	Таа	Кхойкхой	Калахари	Квади	Сандаве
Сев.-кой.	12/20	6/7	5/6		4/5		
!Кви			17/34	5/5	5/6	4/5	
Таа				4/5	4/6	3/4	
Кхойкхой					17/31	8/13	
Калахари						8/14	1/6
Квади							2/5

В подавляющем большинстве случаев «динамическая градация» дает позитивные результаты, причем ожидаемым образом: соотношение между «сильноустойчивой» и «слабоустойчивой» половинами списка достаточно сбалансировано на «достоверных» узлах (т. е. между близкородственными группами) и явно крениется в сторону «сильноустойчивой» половины на «тестируемых» узлах, что скорее свидетельствует в пользу генетического родства между, например, сев.-кой. и юж.-кой языками, чем случайностью или «ареальным характером» совпадений. Единственное серьезное исключение из этой тенденции — сандаве, бóльшая часть параллелей которого с квади и центр.-кой. языками принадлежит «слабоустойчивой» половине 50-словного списка.

Перейдем к общей оценке результатов. В ситуации с хадза все достаточно наглядно: проценты совпадений настолько ничтожны, что говорить даже о сугубо гипотетической *возможности* генетического родства хадза с другими «койсанскими» языками не имеет смысла до тех пор, пока аналогичные сопоставления не будут проведены между хадза и хотя бы *некоторыми* (желательно — всеми) из остальных языковых групп африканского континента.

Сандаве обнаруживает вроде бы значимое число сходств с центр.-кой. языками, но при этом, странным образом, не в первой, а во второй половине 50-словника. Согласно принципу динамической градации, это скорее должно указывать на тесные контактные связи между этими таксономическими единицами, чем на их генетическую общность. Учитывая, однако, глубокие типологические расхождения между фонологическими и фонотактическими характеристиками сравниваемых таксонов, а также статус сандаве как современного языка-изолята, результаты этого теста на данном этапе исследования, безусловно, надлежит принять во внимание, но не настолько, чтобы на их основании перестраивать предварительную классификацию. Более тщательно вопрос о характере связей между кхой-квади и сандаве должен обсуждаться на уровне этимологического анализа материала.

Ситуация со связями между «сандаве-квади-кхой» и «периферийно-койсанскими» языками еще сложнее, т. к. число соответствий между отдельными узлами в составе этих таксонов также иногда оказывается равно или превышает 10% — например, в случае схождения между центрально- и южнокойсанскими

языками. Данный подскок, однако, выглядит скорее как аномалия, т. к. столь высокого числа схождения не наблюдается между северно- и южнокойсанскими языками; если же при этом учитывать еще и повышенно «престижный» характер центральнокойсанской семьи (особенно языковой группы кхойкхой), многочисленные заимствования из отдельных представителей которой хорошо известны и в северно-, и в южнокойсанских языках, завышенность процентов схождения между !кви-таа и кхой технически объяснима как отражение возможных контактных связей. С другой стороны, «динамическая градация» может намекать и на то, что за отдельными случаями контактных наслоений скрывается генетическое родство «сверхглубокого» уровня.

В любом случае ситуация такова, что говорить о «макрокойсанском» родстве (даже без хадза) можно будет по-настоящему только после того, как (а) этимологический анализ подтвердит — или опровергнет? — гипотезы более низкого уровня («периферийно-койсанскую» и «сандаве-кхой-квади»); (б) будет проведено предварительное автоматическое и откорректированное вручную сопоставление данных «койсанских» групп с материалами по другим языковым группам Африки.

Таким образом, предварительный анализ 50-словников, составленных для всех «койсанских» праязыков 1-го уровня и языков-изолятов, не дает позитивных свидетельств в пользу «макрокойсанской» гипотезы, но позволяет свести девять койсанских «мини-узлов» к трем:

- (а) хадза;
- (б) сандаве-кхой-квади;
- (в) периферийно-койсанские,

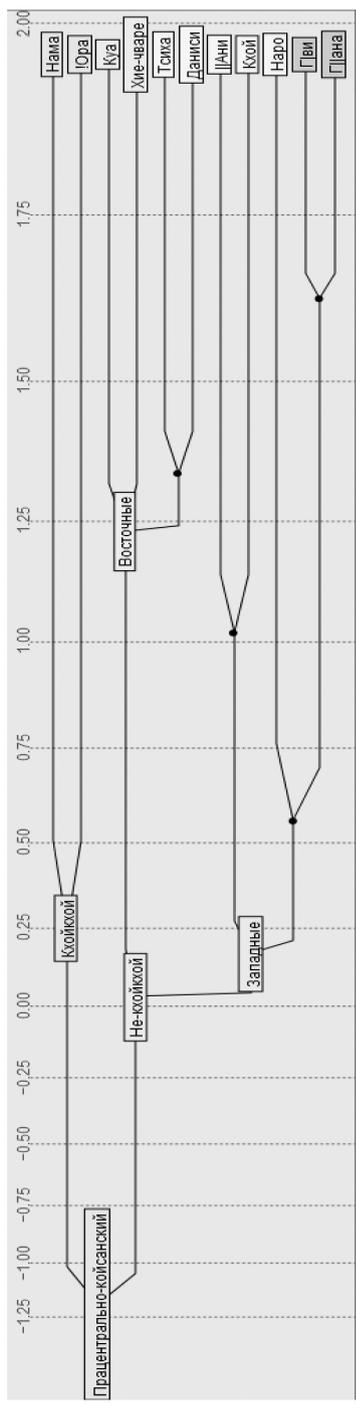
два из которых — (б) и (в) — для того, чтобы получить статус «наглядно продемонстрированных», нуждаются в дальнейшем этимологическом обосновании, но могут при этом являться объектами автономного этимологического исследования. Что касается потенциальных родственных связей между (а), (б) и (в), на «макросемейном» уровне, то о них на данном этапе говорить преждевременно; к этому вопросу можно будет вернуться лишь на самой последней стадии исследования.

Приложение 1. Генеалогические деревья.

Приводимые ниже генеалогические классификационные схемы порождены автоматически в рамках компьютерной лингвистической среды StarLing (авторство программных алгоритмов принадлежит С. А. Старостину, А. С. Старостину и Ф. С. Крылову) на материале полных 100-словных списков Сводеша, составленных для перечисленных языков Г. С. Старостиным. Подробный этимологический разбор первой («сильноустойчивой») половины 100-словных списков содержится в настоящем издании; с полными списками, включая подробные библиографические и лингвистические аннотации к каждому вхождению, можно ознакомиться на официальном сайте проекта «Глобальная лексикостатистическая база данных» (<http://starling.rinet.ru/new100>).

Цифры, проставленные поверх деревьев, соответствуют хронологическим датам (за единицу принимается временной отрезок в 1000 лет); приблизительная глоттохронологическая датировка распада исходного и промежуточных праязыковых узлов определяется исходя из варианта глоттохронологической формулы С. А. Старостина с «подвижным» коэффициентом скорости распада (высчитывается на основании усреднения значений индивидуальных коэффициентов каждого из элементов 100-словного списка).

Схема 3. Генеалогическое дерево центральнокойсанских языков



Приложение 2. Сопоставительные таблицы.

В приводимых ниже таблицах максимально сжато и схематично представлены основные результаты этимологического анализа 50-словных списков по отдельным группам низкого уровня. В случаях, когда на выражение одного и того же базисного значения в праязыке с более или менее равной вероятностью претендуют две основы, в таблице приводятся обе. «Слабые» кандидаты на роль праэтимона в таблицу не включаются.

В квадратные скобки в таблице заключены реконструкции или формы, зафиксированные в языках-изолятах, базисный статус которых находится под существенным сомнением. Тильдой (~) отмечены праязыковые варианты (диалектные или морфо/но/логические).

Подстрочными индексами (1, 2 и т. п.) обозначаются «псевдокогнации» между отдельными формами, полученные в результате ручной коррекции результатов автоматической обработки 100-словных списков алгоритмом определения фонетического сходства (см. раздел 2.8). Необходимо учитывать, что *эти цифры не отражают результаты строгого этимологического анализа на основании сравнительно-исторического метода*, а представляют собой сугубо предварительный результат, на базе которого впоследствии будет производиться более тщательная обработка данных для определения хронологически более глубоких таксонов и реконструкции прасписков базисной лексики «2-го уровня».

Слово	Сев.-кой.		Хоан	Кви	Таа	
ashes	*t _Q 1	*//oa	ʃ ^h oe	*!qui 2	*!qui 2	[*!oa]
bird	*c ^a (-)ma 1		ʃi [c ^h q:ma] 1	*!q ^h u- 2	*!(^h)u- 2	
black	*žō 1		ʃkxau	*! ^h oe	*!ʃa-	
blood	*!ʃ(a)ŋ 1		qʀi 1	*//xau	*!q̃	
bone	*!ʔú 1		q̃a:	*!ʃa 2	*!ʃa 2	
claw(nail)	*!!uʔuru 1		!q̃oʔ 1	*!q̃Uro 2	*!aʔm	
die	*!!é		šĩ	*!ʃa 1	*!ʃa 1	
dog	*!ʃ ^h oe 1		q̃eama	*!ʃUŋ 1	*!q ^h a- 1	
drink	*č ^h iN 1		čú 1	*kxa 2	*kxa 2	
dry	*!!kxau		qʀau	*!o- 1	*!//ú- 1	
ear	*!ʃ ^h úí 1		q ^h oe 1	*!ʃu- 2	*!ʃu- 2	
eat	*ʔm̃ 1		ʔam 1	*ʔā 1	*ʔā 1	
egg	*!šú		[k ^h ōʔē]	*!ʃ(a)u- 1	*!ʃu- 1	
eye	*!aʔa 1		θoa 1	*c ^a x(a)u 2	*!ʔū-	
fire	*daʔa		θoa	*!ʃi 1	*!ʔā 1	
foot	*!kxái		!aʔu	*!ʃioa 1	*!ʃú 1	
hair	*!kxúí		šú	*!ʃ ^h u- 1	*!q ^h u- 1	
hand	*!šau		šiu 2	*!kxa 3	*!kxa 3	
head	*ʔšē 1		ʔonū 1	*!ʃa- 1	*!ʃa- 1	
hear	*saʔa 1		ca: 1	*tu 2	*!q̃ 2	
heart	*!kxa 1		!qʀō 1	*!(q)ʔa-i 2	*!qʀa- ~ *!qʀi- 2	
horn	*!ʃ ^h u 1		! ^h o 1	*! (q) ^h ō- 2	*!//ā- 2	
I	*m(i) 1		ma 1	*ŋ 1	*ŋ 1	
kill	*!ʃ ^h ū 1		! ^h ō 1	*!ʃ ^h a- 2	*qa-	
leaf	*!oa		ʒoba	—	*!jana 1	
louse	*c(i)ŋ 1		cʀi 1	*!ōy- 2	*!ōy- 2	
meat	*!ʃ ^h a		q̃e	*!θa- 1	*!θa- 1	
moon	*!šúí		ʃibi	*!ʃoro	*!oi	*!q ^h an

Слово	Сев.-кой.		ЃХоан	!Кви		Таа	
mouth	*c <i>í</i> ₁		š <i>ĩ</i> ₁	*tu		*ʃue	
name	*!ú ₁		!o ₁	* ē- ₂		* ā- ₂	
new	*ze ₁		za: ₁	* ue ₂		* qe ₂	
night	* ú		c ^h ao	* a-		* y-	
nose	*ckx(u)ŋ		!qɔ̃	* u- ₁		* u- ₁	
not	* oa ₁		^h ɔ̃ɔ̃ ₁	* V- ₂		* q ^h V- ₂	
one	* eʔe ₁		Ōū ₁	* ɪʔu- ₂		* ɪʔu- ₂	
rain	* à		čɔ̃ʔā:	* q(ʰ)au		*!kxoe	
smoke	*šorV		ʒue	* ɔ-		*ckxa- ₁	* alu
star	* ʔú ₁		ʔō ₁	* kxɔ- ₂		* o(-)na ₂	
stone	* ūm		^h qa	*!ao		* u(-le)	
sun	* án, *gao		č ^h a:	* qʔuni ₁		* ʔan ₁	
tail	* xōē		θxui	* ^h (a)i		* ā-	
thou	*a ₁		u	*a ₁		*a ₁	
tongue	*(n)d ^h ari		cela	* ʔan(i) ₁		* nɔn(i) ₁	
tooth	*cʔau ₁		ciu ₁	* ^h āi ₂		* q ^h an ₂	
tree	* ^h āŋi		ɔ̃	* ōɔ- ₁		* ʔonɔ- ₁	
two	*ca ~ *cā		θoa	* (u)ʔu		* ʔum	
water	* u		žo ₁	* q ^h a ₂		* q ^h a ₂	
we	E.*e ₁	I.*m ₂	n-!a	E.*si ₃	I.*i ₁	E.*si ₃	I.*i ₁
what	*ha	*m	çini	*TI ₁	* V ₂	* V ₂	
who	*ha	*m	—	*TI		* V	

Слово	Кхойкхой	Калахари-кхой	Квади	Сандаве	Хадза
ashes	*t ^h ao ₁	*f ₁ óà	—	!úip ^h á	hoc ^o
bird	*kxani	*zara	—	t ^h ui ₃	t ^h it ^h i ₃
black	*f ₂ u ₂	*f ₂ ú ₂	zu ₁	kʌŋkʌra	tič <i>i</i>
blood	* ʌo ₂	* ʌò ₂	ʌo ₂	ɛkʌ	át ^h a ^o má
bone	*f ₃ xo	* ʌóǎ	fǎ ₂	!i ₁	miλʌ
claw (nail)	* oro ₂	* oro ₂	—	cʌwáʌá	baʌu
die	* ʌo ₂	* ʌó ₂	kade-ʔo	λá-	misi ta i
dog	*ari ₂	*ari- ₂ *aba	ʌau-	kákà	áʔano
drink	*kxa ₂	*kxá ₂	kxa ₂	cʌé	fá
dry	* ʌo ₂	* ʌo ₂	—	iŋ-	ɔpe
ear	*f ₃ ae ₃	*f ₃ é ₃	gɔ-	kéké	hačʌpič <i>i</i>
eat	*f ₃ ũ ₂	*f ₃ ũ ₂	ʔjũ-	manɸ ^h a	séme
egg	*!ʌubu	*f ₃ úbí	ʔi-	dĩʔa	uʌe
eye	*mũ	*f ₃ xéí	ši- ₂	óé ₁	ʔak ^h wa
fire	* ʌae ₁	* ʌe ₁	fě	ĩ	cʌokʌo
foot	*f ₃ ai	*f ^h are	ze-	^h átá	ʔapukwa
hair	* ʌũ ₂	* ʌũ ₂	oʔm ₂	cʌê	haλʌe
hand	*!ʌom	*c ^h áũ ₂	—	λũ	ʔuk ^h wa
head	*dana	*f ₃ ú *mâ	c ^h ē ₂	ɸé ₂	λóma
hear	* áũ ₃	*kúm ₄	kum ₄	k ^h éʔé	aʔe ₃
heart	*f ₃ ao ₃	*f ₃ áo ₃	cʌo-	zigida	ʔasač <i>i</i>
horn	* ǎ ₃	* á ₃	—	λáná ₃	loʔo
I	*ta ~ *de ₂	*ta ~ *de ₂	ta ₂	ɸi ₂	o=no
kill	*!am	* kxũ ₂	—	kʌoe	ó
leaf	*f ₃ ae	* áná ₁	—	á ₁	hačʌape
louse	*kxuri ₃	*kxuni ₃	—	mǎ ǎ	ámáç <i>i</i>
meat	*kxo ₂	*kxo ₂	kxóle ₂	ĩ	mana
moon	* xǎ	* óé	k ^h ǎ-	!a-biso	sét ^h a
mouth	*kxam ₂	*kxám ₂	kxami- ₂	ũ	ʔawanika

Слово	Кхойкхой	Калахари-кхой	Квади	Сандаве	Хадза			
name	* kxon ₃	* kxòh̃ ₃	—	oa	?akana			
new	* pa-sa	*kxòà	—	ae ₁	zana			
night	*t ^h u ₁	*t ^h ũ ₁	t ^h wi: ₁	tue ₁	cɪfi			
nose	* fui ₂	* fui ₂	čwi-	at ^h i	?int ^h awe			
not	*tama ₃	*tama ₃	*be	—	-cɛ	?akw-		
one	* ui ₃	* úí ₃	ui ₃	cɛxe	?ičáme			
rain	*tu ₁	*tú ₁	fete	λoa	?at ^h i			
smoke	* kxan-i ₁	*cʔan-i ₁	—	cʔukʔa ₂	cʔikʔo ₂			
star	* ami-ro	* xáni	xo-	owã	nc ^h a			
stone	* pui	* lóa	ʒo-	di(ŋ)	ha(n)!ʔa			
sun	*sore	* ám̃	?ui-	ʔakasu ₁	?išo			
tail	*sao ₁	*cáo ₁	θo- ₁	cʔoa ₁	cʔahó ₁			
thou	*=c / *=s ₁	*c=a / *s=a ₁	s=a ₁	ha=pu	te			
tongue	*lam ₂	*dàh̃ ₂	tamen- ₂	! ^h e(ŋ)	át ^h â ₁			
tooth	* ũ ₃	* ú ₃	ʔo-	!ʔak ^h a(ŋ)	?áhà			
tree	*hai ₂	*yi ₂	č ^h i-	t ^h é	citi			
two	* am ₁	* ám̃ ₁	a ~ ã ₁	ki(-)so	piye			
water	* lam	*c ^h ǎ ₁	kxoʔ-	cʔa ₁	?at ^h i			
we	*=e ₁	*=m ₂	*=e ₁	*=m ₂	(h)ina ₄	=mu ₂	s ^u uŋ ₄	u-
what	*ta- ₁	*ndu	—	ho-ɕo	?akwaza			
who	*ta- ₁	*ma	—	ho	ta- ₁	?aku		

Литература

- Андронов 1971: М. С. Андронов. *Язык брауи*. М.: «Наука».
- Апресян 1995: Ю. Д. Апресян. *Лексическая семантика: Синонимические средства языка*. 2-е издание. М.: «Восточная литература».
- Беликов 1998: В. И. Беликов. *Пиджины и креольские языки Океании. Социолингвистический очерк*. М.: «Восточная литература».
- Беликов 2006: В. И. Беликов. *Конвергентные процессы в лингвогенезе*. Диссертация в виде научного доклада на соискание ученой степени д. ф. н. М.
- Бурлак 2007: С. А. Бурлак. Креольские языки и глоттохронология. // *Аспекты компаративистики III*. М.: РГГУ, стр. 499-508.
- Бурлак & С. Старостин 2005: С. А. Бурлак, С. А. Старостин. *Сравнительно-историческое языкознание*. 2-е изд. М.: «Академия».
- Долгопольский 1964: А. Б. Долгопольский. Гипотеза древнейшего родства языков Северной Евразии с вероятностной точки зрения. // *Вопросы языкознания*, № 2, стр. 53-63.
- Долгопольский 1965: А. Б. Долгопольский. Сохраняемость лексики, универсалии и ареальная типология. // *Лингвистическая типология и восточные языки*. М.: «Наука», стр. 189-195.
- Климов 1964: Г. А. Климов. *Этимологический словарь картвельских языков*. М.: Издательство Академии Наук СССР.
- Курилов 2001: Г. Н. Курилов. *Юкагирско-русский словарь*. Новосибирск: «Наука».
- Мудрак 2009: О. А. Мудрак. *Классификация тюркских языков и диалектов с помощью методов глоттохронологии на основе вопросов по морфологии и исторической фонетике*. М.: РГГУ.
- Расторгуева, Эдельман 2007: В. С. Расторгуева, Д. И. Эдельман. *Этимологический словарь иранских языков. Том 3: f-h*. М.: «Восточная литература».
- Сводеш 1960: М. Сводеш. Лексикостатистическое датирование доисторических этнических контактов. // *Новое в лингвистике, в. I*. М.: Издательство иностранной литературы (перевод [Swadesh 1952]).
- Севортян 1974: Э. В. Севортян. *Этимологический словарь тюркских языков (общетюркские и межтюркские основы на гласные)*. М.: «Наука».

Старостин 2005: Г. С. Старостин. Некоторые аспекты развития кликсов в койсанских языках. // *Аспекты компаративистики 1*. М.: РГГУ, стр. 281-299.

Старостин 2006: Г. С. Старостин. Лабиальные кликсы в койсанских языках. // *Аспекты компаративистики 2*. М.: РГГУ, стр. 353-374.

Старостин 2008а: Г. С. Старостин. К вопросу о генетической принадлежности языка хадза. // *Африканский сборник 2007*. Санкт-Петербург: «Наука», стр. 262-278.

Старостин 2008б: Г. С. Старостин. Согласовательные классы и способы выражения множественного числа в языке !хонг. // *Вопросы языкознания*, 3 (2008), стр. 51-75.

С. Старостин 1989: С. А. Старостин. Сравнительно-историческое языкознание и лексикостатистика. // *Лингвистическая реконструкция и древнейшая история Востока*. М.: «Наука», стр. 3-39.

С. Старостин 1991: С. А. Старостин. Алтайская проблема и происхождение японского языка. М.: «Наука».

С. Старостин 1995: С. А. Старостин. Сравнительный словарь енисейских языков. // *Кетский сборник. Лингвистика*. М.: «Восточная литература», стр. 176-315.

С. Старостин 1999: С. А. Старостин. О доказательстве языкового родства. // *Типология и теория языка (к 60-летию А. Е. Кибрика)*. М.: 1999, стр. 57-69. Перепечатано в: С. А. Старостин. *Труды по языкознанию*. М.: Языки славянских культур, 2007, стр. 779-793.

С. Старостин 2004: С. А. Старостин. Определение устойчивости базисной лексики. // В: С. А. Старостин. *Труды по языкознанию*. М.: Языки славянских культур, 2007, стр. 827-839.

С. Старостин 2005: С. А. Старостин. Computer-based simulation of the glottochronological process (Letter to M. Gell-Mann). // В: С. А. Старостин. *Труды по языкознанию*. М.: Языки славянских культур, 2007, стр. 854-861.

Хелимский 1989: Е. А. Хелимский. К оценке надежности индоевропейско-семитских лексических сопоставлений. // *Палеобалканистика и античность*. М., стр. 13-20.

ЭСТЯ 1989: *Этимологический словарь тюркских языков. Общетюркские и межтюркские основы на буквы "Ж", "Ж", "Й"*. Под редакцией Л. С. Левитской. М.: «Наука».

Anders 1934: H. Anders. A note on a southeastern Bushman dialect. // *Zeitschrift für Eingeborenen-Sprachen*, 25, pp. 81-89.

Arbousset & Damas 1846: T. Arbousset, F. Damas. *Narrative of an Exploratory Tour to the North-East of the Colony of the Cape of Good Hope*. Cape Town: A. S. Robertson, Heerengracht, Saul Solomon & Co.

Argyle 1986: W. J. Argyle. The Extent and Nature of Khoisan Influence on Zulu. // *Sprache und Geschichte in Afrika* 7.2, pp. 43-71.

Atkinson et al. 2005: Quentin Atkinson, Geoff Nicholls, David Welch, Russell Gray. From Words to Dates: Water into Wine, Mathe-magic Or Phylogenetic Inference? // *Transactions of the Philological Society*, 103: 2, pp. 193-219.

Barnard 1985: Alan Barnard. *A Nharo Wordlist with Notes on Grammar*. Durban: University of Natal.

Bartholomae 1961: Christian Bartholomae. *Altiranisches Wörterbuch*. 2. *Unveränderte Auflage*. Berlin: Walter de Gruyter & Co.

Baucom 1974: Kenneth Baucom. Proto-Central Khoisan. // *Third Annual Conference on African Linguistics, 7-8 April 1972*. Ed. by E. Voeltz. Bloomington, Indiana University, pp. 3-37.

Baxter 1995: W. Baxter. 'A stronger affinity... than could have been produced by accident': A probabilistic comparison of Old Chinese and Tibeto-Burman. // *The Ancestry of the Chinese Language*. Journal of Chinese Linguistics, Monograph Series No. 8. Ed. by W. Wang, pp. 1-39.

Baxter & Manaster-Ramer 2000: W. Baxter, A. Manaster-Ramer. Beyond lumping and splitting: probabilistic issues in historical linguistics. // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 167-188.

BCCW: *Benue-Congo Comparative Wordlist*. Ed. by Kay Williamson & Kiyosi Shimizu. Ibadan: West African Linguistic Society, 1968.

Beach 1938: *The Phonetics of the Hottentot Language*. Cambridge: W. Heffer and Sons.

Bell & Collins 2001: Arthur Bell, Chris Collins. †Hoan and the typology of click accompaniments in Khoisan. // *Cornell working papers in linguistics*, 18, pp. 126-253.

Bender 1971: Lionel M. Bender. The Languages of Ethiopia: A New Lexicostatistic Classification and Some Problems of Diffusion. // *Anthropological Linguistics*, 13, 5, pp. 165-288.

Bender 1983: Lionel M. Bender. Proto-Koman Phonology and Lexicon. // *Afrika und Übersee*, 66, pp. 259-297.

Bergsland & Vogt 1962: K. Bergsland, H. Vogt. On the Validity of Glottochronology. // *Current Anthropology*, 3, pp. 115-153.

Biggs 1990: Bruce Biggs. *English—Maori / Maori—English Dictionary*. Auckland University Press.

Blažek 2007: Václav Blažek. Nilo-Saharan Stratum of Ongota. // // *Advances in Nilo-Saharan Linguistics. Proceedings of the 8th Nilo-Saharan Linguistics Colloquium. Hamburg, August 22-25, 2001. Ed. by Doris Payne & Mechthild Reh*. Köln: Rüdiger Köppe, pp. 1-10.

Bleek 1927: Dorothea F. Bleek. The distribution of Bushman languages in South Africa. // *Festschrift Meinhof*. Hamburg: L. Friedrichsen, pp. 55-64.

Bleek 1928: Dorothea F. Bleek. *The Naron: A Bushman Tribe of the Central Kalahari*. Cambridge University Press.

Bleek 1928-30: Dorothea F. Bleek. Bushman grammar: A grammatical sketch of the language of the |xam-ka-!kʷe. *Zeitschrift für Eingeborenen-Sprachen* 19, pp. 81-98; 20, pp. 161-174.

Bleek 1929: Dorothea F. Bleek. *Comparative vocabularies of Bushman languages*. Publications of the School of African Life and Language, University of Cape Town. University Press. Cambridge.

Bleek 1931: Dorothea F. Bleek. The Hadzapi or Watindega of Tanganyika territory. // *Africa*, 4, pp. 273-286.

Bleek 1937: Dorothea F. Bleek. Grammatical Notes and Texts in the |Auni Language. // *Bushmen of the Southern Kalahari*. Ed. by J. D. R. Jones, University of Witwatersrand Press, Johannesburg, pp. 195-220.

Bleek 1956: Dorothea F. Bleek. *A Bushman Dictionary*. American Oriental Series, vol. 41. New Haven, Connecticut.

Bleek 2000: Dorothea F. Bleek. *The ||N!ke or Bushmen of Griqualand West*. Ed. by Tom Güldemann. Khoisan forum, working paper No. 15. Köln: University of Cologne.

Bleek & Lloyd 1911: Wilhelm H. I. Bleek, Lucy L. Lloyd. *Specimens of Bushman Folklore*. London: George Allen & Co.

Blust 2000: Robert Blust. Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages. // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 311-331.

Bosworth & Toller 1898: Joseph Bosworth. *An Anglo-Saxon Dictionary. Edited and Enlarged by T. Northcote Toller*. Oxford University Press.

Burrow & Emeneau 1984: Thomas Burrow, M. B. Emeneau. *A Dravidian Etymological Dictionary. Second Edition.* // Oxford: Clarendon Press.

Campbell 1988: Lyle Campbell. Review article on Language in the Americas by Joseph H. Greenberg. // *Language*, v. 64, pp. 591–615.

Campbell 1998: Lyle Campbell. Nostratic: a personal assessment. // *Nostratic: Sifting the Evidence.* Ed. by Joseph C. Salmons and Brian D. Joseph. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 107-152.

Campbell 2004: Lyle Campbell. *Historical Linguistics: an Introduction. 2nd edition.* Edinburgh University Press, MIT Press.

Campbell & Poser 2008: Lyle Campbell, William J. Poser. *Language Classification: History and Method.* Cambridge University Press.

Chebanne 2000: Andy M. Chebanne. The Phonological System of the Cuaa Language. // *The State of Khoesan Languages in Botswana.* Ed. by H. M. Batibo & J. Tsonope. Tasalls Publishing, University of Botswana, pp. 18-32.

Childs 2003: G. Tucker Childs. *An Introduction to African Languages.* Amsterdam: John Benjamins.

Clements & Rialland 2008: G. N. Clements, Annie Rialland. Africa as a phonological area. // *A Linguistic Geography of Africa.* Ed. by Bernd Heine & Derek Nurse. Cambridge University Press, pp. 36-85.

Collins 1998: Chris Collins. *Plurality in †Hoan.* Khoisan Forum, working paper No. 9. Köln, University of Cologne.

Collins 2001a: Chris Collins. Aspects of plurality in †Hoan. // *Language*, 77 (3), pp. 456-476.

Collins 2001b: Chris Collins. Multiple verb movement in †Hoan. // *Cornell working papers in linguistics*, 18, pp. 75-104.

Collins 2001c: Chris Collins. VP internal structure in Jul'hoan and †Hoan. // *Cornell working papers in linguistics*, 18, pp. 1-27.

Collins & Namaseb 2011: Chris Collins, Levi Namaseb. *A Grammatical Sketch of Njuuki with Stories.* Köln: Rüdiger Köppe Verlag.

Crabb 1965: David W. Crabb. *Ekoid Bantu Languages of Ogoja, Eastern Nigeria. Part I: Introduction, Phonology and Comparative Vocabulary.* Cambridge University Press.

Crawhall 2004: Nigel Crawhall. *!Ui-Taa language shift in Gordonia and Postmasburg Districts, South Africa.* PhD thesis submitted to the Faculty of Humanities, University of Cape Town.

Dempwolff 1916: Otto Dempwolff. *Die Sandawe. Linguistisches und ethnographisches Material aus Deutsch-Ostafrika*. Hamburg: L. Friedrichsen.

Dempwolff 1917: Otto Dempwolff. Beiträge zur Kenntnis der Sprachen in Deutsch-Ostafrika. // *Zeitschrift für Kolonialsprachen*, 7.4, pp. 309-325.

Dickens 1994: Patrick J. Dickens. *English-Ju|'hoan, Ju|'hoan-English dictionary*. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 8. Köln: Rüdiger Köppe Verlag.

Dimmendaal 2011: Gerrit Dimmendaal. *Historical linguistics and the comparative study of African languages*. Amsterdam: John Benjamins Publishing Company.

Doke 1925: Clement M. Doke. An outline of the phonetics of the language of the Chû: Bushmen of north-west Kalahari. *Bantu studies and general South African anthropology*, 2 (3), pp. 129-165.

Doke 1936: Clement M. Doke. An outline of #Khomani Bushman phonetics. *Bantu studies* (Johannesburg), 10, pp. 433-461.

Dolgopolsky 2000: Aharon Dolgopolsky. Sources of linguistic chronology. // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 401-409.

Dolgopolsky 2008: Aharon Dolgopolsky. *Nostratic Dictionary*. McDonald Institute for Archaeological Research, Oxford Publishing Press.

Dornan 1917: S. S. Dornan. The Tati Bushmen (Masarwas) and their language. // *Journal of the Royal Anthropological Society of Great Britain and Ireland*, 47, pp. 37-112.

Dybo 1989: Vladimir Dybo. V. M. Illich-Svitych and the Development of Uralic and Dravidian linguistics (preliminary report). // In: *Explorations in Language Macrofamilies*. Ed. by V. Shevoroshkin. Bochum, Ann Arbor, pp. 20-29.

Dybo & Starostin 2007: A. V. Dybo, G. S. Starostin. In Defense of the Comparative Method, or the End of the Vovin Controversy. // *Акнекты компаративистики III*. Moscow: RSUH Publishers, pp. 119-258.

Eaton 2002: Helen Eaton. *The grammar of focus in Sandawe*. Ph. D. dissertation. University of Reading.

Ehret 1980: Christopher Ehret. *The historical reconstruction of Southern Cushitic: Phonology and Vocabulary*. Berlin: Dietrich Reimer Verlag.

Ehret 1986: Christopher Ehret. Proposals on Khoisan reconstruction. // *Sprache und Geschichte in Afrika*, 7.2, pp. 105-130.

Ehret 2000: Christopher Ehret. Testing the expectations of glotto-chronology against the correlations of language and archaeology in Africa. // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 373-399.

Ehret 2001: Christopher Ehret. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln: Rüdiger Köppe Verlag.

Ehret 2003a: Christopher Ehret. Toward Reconstructing Proto-South Khoisan (PSAK). // *Mother Tongue VIII*, 2003, pp. 63-80.

Ehret 2003b: Christopher Ehret. *Proto-South Khoisan Etymological Dictionary (Draft, 2/2/2003)*. Ms.

Elderkin 1982: Derek Elderkin. On the classification of Hadza. // *Sugia*, 4, pp. 67-82.

Elderkin 1983: Derek Elderkin. Tanzanian and Ugandan isolates. // *Nilotic studies: proceedings of the international symposium on languages and history of the Nilotic peoples, Cologne, January 4-6, 1982*. Ed. by Rainer Vossen & Marianne Bechhaus-Gerst. Vol. 2, pp. 499-521.

Elderkin 1986: Derek Elderkin. Diachronic inferences from basic sentence and noun phrase structure in Central Khoisan and Sandawe. // *Sprache und Geschichte in Afrika* 7, 2, pp. 131-156.

Elderkin 1998: Derek Elderkin. Sandawe Stems. // *Language, Identity, and Conceptualization among the Khoisan*. Ed. by Mathias Schladt. Köln: Rüdiger Köppe Verlag, pp. 13-33.

Elderkin 2004: Edward D. Elderkin. The starred tones of Central Khoisan. // *Afrika und Übersee*, 87, pp. 3-77.

Elderkin 2008: Edward D. Elderkin. Proto-Khoe tones in Western Kalahari. // *Khoisan Languages and Linguistics*. Ed. by Sonja Ermisch. Köln: Rüdiger Köppe Verlag, pp. 87-136.

Embleton 1986: Sheila M. Embleton. *Statistics in Historical Linguistics*. Bochum: Studienverlag Dr. N. Brockmeyer.

Ethnologue 2009: *Ethnologue: Languages of the World. 16th edition*. Dallas: SIL International. On-line version: <http://www.ethnologue.com>.

Fleming 1986: Harold Fleming. Hadza and Sandawe genetic relations. // *Sprache und Geschichte in Afrika*, 7.2, pp. 157-187.

Fodor 1966: István Fodor. *The Problems in the Classification of the African Languages*. Budapest: Hungarian Academy of Sciences.

Fraenkel 1962: Ernst Fraenkel. *Litauisches Etymologisches Wörterbuch*. Heidelberg: Carl Winters Universitätsverlag.

Gell-Mann, Peiros, & Starostin 2009: Murray Gell-Mann, Ilia Peiros, George Starostin. Distant Language Relationship: The Current Perspective. // *Вопросы языкового родства (Journal of Language Relationship)*, v. 1, pp. 13-30.

Gleason 1959: H. A. Gleason Jr. Counting and calculating for historical reconstruction. // *Anthropological linguistics*, 1, 2, pp. 22-32.

Goodman 1971: M. Goodman. The strange case of Mbugu. // *Pidginization and Creolization of Languages*. Ed. by D. Hymes. Cambridge University Press, pp. 243-254.

Greenberg 1955: Joseph H. Greenberg. *Studies in African linguistic classification*. New Haven, CT: Compass.

Greenberg 1966: Joseph H. Greenberg. *The Languages of Africa*. Bloomington, Indiana University; Mouton & Co., The Hague.

Greenberg 1981: Joseph H. Greenberg. Nilo-Saharan Moveable-k as a Stage III article (with a Penutian Typological Parallel). // *Journal of African Languages and Linguistics*, 3, pp. 105-112.

Greenberg 1987: Joseph H. Greenberg. *Language in the Americas*. Stanford University Press.

Greenberg 1989: Joseph H. Greenberg. Classification of American Indian languages: a reply to Campbell. // *Language*, 65, pp. 107-114.

Greenberg 1995: Joseph H. Greenberg. The Concept of Proof in Genetic Linguistics. // *Mother Tongue I*, pp. 207-216.

Greenberg 2001: Joseph H. Greenberg. The Methods and Purposes of Linguistic Genetic Classification. // *Language and Linguistics* 2.2, pp. 111-135.

Gruber 1973: Jeffrey S. Gruber. †Hòã kinship terms. // *Linguistic Inquiry*, vol. IV, 4, pp. 427-449.

Gruber 1975: Jeffrey S. Gruber. Plural predicates in †Hòã. // *Bushman and Hottentot linguistic studies*. Ed. by Anthony Traill. Communications from the African Studies Institute, no 2. Johannesburg: University of the Witwatersrand, pp. 1-50.

Güldemann 2004: Tom Güldemann. Reconstruction through 'deconstruction': the marking of person, gender, and number in the Khoe family and Kwadi. // *Diachronica* 21:2, pp. 251-306.

Güldemann 2007: Tom Güldemann. Clicks, genetics, and "proto-world" from a linguistic perspective. // *University of Leipzig Papers on Africa, Languages and Literatures*, 29. Leipzig: Institut für Afrikanistik, Universität Leipzig.

Güldemann 2008: Tom Güldemann. Greenberg's «case» for Khoisan: the morphological evidence. // *Problems of linguistic-historical reconstruction in Africa*. Ed. by Dymitr Ibriszimow and Rainer Voßen. Köln: Rüdiger Köppe, pp. 123-155.

Güldemann 2010: Tom Güldemann. Proto-Bantu and Proto-Niger-Congo: Macro-areal typology and linguistic reconstruction. // To appear in: *Christa König & Osamu Hieda (eds). International Symposium of the Center of Corpus-Based Linguistics and Language Education (CbLLE)*. Amsterdam & Philadelphia: John Benjamins.

Güldemann 2011: Tom Güldemann. *The Lower Nossob varieties of Tuu: !Ui, Taa, or neither?* // Paper presented at ICHL XX, National Museum of Ethnology, Osaka, July 25-30 2011.

Güldemann & Elderkin 2003: Tom Güldemann, Derek Elderkin. *On external genealogical relationships of the Khoe family*. Ms. available at: http://email.eva.mpg.de/~gueldema/pdf/Gueldemann_Elderkin.pdf.

Güldemann & Stoneking 2008: Tom Güldemann, Mark Stoneking. A historical appraisal of clicks: a linguistic and genetic population perspective. // *Annual Review of Anthropology*, 37, pp. 93-109.

Güldemann & Vossen 2000: Tom Güldemann, Rainer Vossen. Khoisan. // *African languages: an introduction*. Ed. by Bernd Heine & Derek Nurse. Cambridge University Press, pp. 99-122.

Guthrie 1964: Malcolm Guthrie. Review of *The Languages of Africa* by J. Greenberg. // *Journal of African History*, 5, pp. 135-136.

Guthrie 1967: Malcolm Guthrie. *Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages, vols. I-IV*. London: Gregg Press Ltd.

Haacke et al. 1997: Wilfrid Haacke, Eliphaz Eiseb, Levi Namaseb. Internal and external relations of Khoekhoe dialects: a preliminary survey. // *Namibian languages: reports and papers*. Ed. by Wilfrid Haacke & Edward Elderkin. Cologne: Rüdiger Köppe Verlag for the University of Namibia, pp. 125-203.

Haacke & Eiseb 1999: Wilfrid Haacke, Eliphaz Eiseb. *Khoekhoegowab-English / English-Khoekhoegowab Glossary*. Gamsberg Macmillan.

Haacke & Eiseb 2002: Wilfrid Haacke, Eliphaz Eiseb. *A Khoekhoegowab Dictionary with an English-Khoekhoegowab Index*. Gamsberg Macmillan.

Haarmann 1990a: Harald Haarmann. *Language in its Cultural Embedding*. Berlin: Mouton de Gruyter.

Haarmann 1990b: Harald Haarmann. 'Basic vocabulary' and Language Contacts: the Disillusion of Glottochronology. // *Indogermanische Forschungen*, 95, pp.1-37.

Hagman 1977: R. S. Hagman. *Nama Hottentot Grammar*. Bloomington: Indiana University.

Hastings 2001: Rachel Hastings. Evidence for the genetic unity of Southern Khoesan. // *Cornell working papers in linguistics*, 18, pp. 225-246.

Heggarty 2000: Paul Heggarty. Quantifying change over time in phonetics. // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 531-562.

Heikkinen 1986: Terttu Heikkinen. Phonology of the !Xū dialect spoken in Ovamboland and western Kavango. // *South African Journal of African Languages* 6, 1, pp. 18-28.

Heikkinen 1987: Terttu Heikkinen. *An Outline of the Grammar of the !Xū Language Spoken in Ovamboland and Western Kavango*. Pretoria.

Heine & Honken 2010: Bernd Heine, Henry Honken. The Kx'a Family: A New Khoisan Genealogy. // *Journal of Asian and African Studies*, 79, pp. 5-36.

Heine & Kuteva 2001: Bernd Heine, Tania Kuteva. Convergence and Divergence in the Development of African Languages. // *Areal Diffusion and Genetic Inheritance*. Ed. by Alexandra Y. Aikhenvald and R. M. W. Dixon. Oxford University Press, pp. 393-411.

Heine & Leyew 2008: Bernd Heine, Zelealem Leyew. Is Africa a linguistic area? // *A Linguistic Geography of Africa*. Ed. by Bernd Heine & Derek Nurse. Cambridge University Press, pp. 15-35.

Heine & Nurse 2000: Bernd Heine, Derek Nurse (eds.). *African Languages: An Introduction*. Cambridge University Press.

Heine & Nurse 2008: Bernd Heine, Derek Nurse. Introduction. // *A Linguistic Geography of Africa*. Ed. by Bernd Heine & Derek Nurse. Cambridge University Press, pp. 1-14.

Hetzron 1972: Robert Hetzron. *Ethiopian Semitic: Studies in Classification*. Manchester.

Hock & Joseph 1996: Hans H. Hock, Brian D. Joseph. *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin & New York: Mouton de Gruyter.

Hojjer 1956: Harry Hojjer. Lexicostatistics: a critique. // *Language*, 32, pp. 49-60.

Holm 2007: Hans J. Holm. The new Arboretum of Indo-European "Trees": can new algorithms reveal the phylogeny and even prehistory of IE? // *Journal of Quantitative Linguistics*, 14: 2, pp. 167-214.

Holm 2009: Hans J. Holm. Albanische Basiswortlisten und die Stellung des Albanischen in den indogermanischen Sprachen. // *Zeitschrift für Balkanologie*, Nr. 45-2, pp. 171-205.

Holman et al. 2008: Eric W. Holman, Søren Wichmann, Cecil Brown, Viveka Velupillai, André Müller, Dik Bakker. Explorations in automated lexicostatistics. // *Folia Linguistica* 42.2, pp. 331-354.

Honken 1977: Henry Honken. Submerged features and proto-Khoisan. // *Khoisan linguistic studies* 3. Ed. by Anthony Traill. Communications from the African Studies Institute, no 6. Johannesburg: University of the Witwatersrand, pp. 145-169.

Honken 1988: Henry Honken. Phonetic correspondences among Khoisan affricates. // *New perspectives on the study of Khoisan*. Ed. by Rainer Vossen. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 7. Köln: Rüdiger Köppe Verlag, pp. 47-65.

Honken 1998: Henry Honken. Types of Sound Correspondence Patterns in Khoisan languages. // *Language, Identity, and Conceptualization among the Khoisan*. Ed. by Mathias Schladt. Köln: Rüdiger Köppe Verlag, pp. 171-193.

Honken 2006: Henry Honken. Fused loans in Khoisan. // *Pula*, 20, 1, pp. 75-85.

Honken 2008: Henry Honken. The split tones in Central Khoisan. // *Khoisan Languages and Linguistics*. Ed. by Sonja Ermisch. Köln: Rüdiger Köppe Verlag, pp. 185-224.

Horn 2010: *The Expression of Negation*. Ed. by Laurence R. Horn. Berlin / New York: Mouton De Gruyter.

Hunziker & Eaton 2007: Daniel A. Hunziker, Elisabeth Hunziker, Helen Eaton. *A Sandawe Dialect Survey*. SIL International: SIL Electronic Survey Report 2007-014, August 2007.

Hunziker & Eaton 2008: Daniel A. Hunziker, Elisabeth Hunziker, Helen Eaton. *A Description of the Phonology of the Sandawe Language*. SIL International: SIL Electronic Working Papers.

Jakobi 1990: Angelika Jakobi. *A Fur Grammar. Phonology, Morphophonology and Morphology*. Helmut Buske Verlag, Hamburg.

Jakobi & Crass 2004: Angelika Jakobi, Joachim Crass. *Grammaire du beria (langue saharienne)*. Köln: Rüdiger Köppe Verlag.

Jungrauthmayr & Ibriszimow 1994: Hermann Jungrauthmayr, Dymitr Ibriszimow. *Chadic Lexical Roots. Volume I: Tentative Reconstruction, Grading, Distribution, and Comments. Volume II: Documentation.* Berlin: Dietrich Reimer Verlag.

Kagaya 1993: Ryohei Kagaya. *A classified vocabulary of the Sandawe language.* Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

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.

Kessler 2001: Brett Kessler. *The Significance of Word Lists.* Stanford, California: CSLI Publications.

Kießling 2008: Roland Kießling. Noun Classification in !Xoon. // *Khoisan Languages and Linguistics.* Ed. by Sonja Ermisch. Köln: Rüdiger Köppe Verlag, pp. 87-136.

Kießling, Mous 2003: Roland Kießling, Maarten Mous. *The Lexical Reconstruction of West-Rift Southern Cushitic.* Köln: Rüdiger Köppe Verlag.

Kießling, Mous, Nurse 2008: Roland Kießling, Maarten Mous, Derek Nurse. The Tanzanian Rift Valley Area. // *A Linguistic Geography of Africa.* Ed. by Bernd Heine & Derek Nurse. Cambridge University Press, pp. 186-227.

Kilian-Hatz 2003: Christa Kilian-Hatz. *Khwe-English Dictionary.* Köln: Rüdiger Köppe Verlag.

Kimmenade 1954: Martin van de Kimmenade. *Essai de grammaire et vocabulaire de la langue Sandawe.* Micro bibliotheca anthropos, no 9, pp. 120. Posieux: Anthropos-Institut.

Knight et al. 2003: Alec Knight et al. African Y chromosome and mtDNA divergence provides insight into the history of click languages. // *Current Biology* 13, pp. 464-473.

Koelle 1854: Sigismund Wilhelm Koelle. *Polyglotta Africana, or a comparative vocabulary of nearly three hundred words and phrases, in more than one hundred distinct African languages.* London: Church Missionary House.

Kogan 2006: Leonid Kogan. Lexical Evidence and the Genealogical Position of Ugaritic (I). // *Babel und Bibel* 3. *Orientalia et Classica* vol. XIV. Winona Lake, Indiana: Eisenbrauns, pp. 429-488.

Köhler 1966: Oswin Köhler. Die Wortbeziehungen zwischen der Sprache der Kxoe-Buschmänner und dem Hottentottischen als ge-

schichtliches Problem. // *Neue Afrikanische Studien*. Ed. by J. Lukas. Hamburg: Deutsches Institut für Afrika-Forschung, pp. 144-165.

Köhler 1973/74: Oswin Köhler. Neuere Ergebnisse und Hypothesen der Sprachforschung in ihrer Bedeutung für die Geschichte Afrikas. // *Paideuma*, 19-20, pp. 162-199.

Köhler 1981: Oswin Köhler. Les langues khoisan. // *Les langues dans le monde ancien et moderne. I: Les langues de l'Afrique subsaharienne*. Ed. by G. Manessy. Paris: Editions du CNRS, pp. 455-615.

Köhler 1986: Oswin Köhler. Allgemeine und sprachliche Bemerkungen zum Feldbau nach Oraltexen der Kxoe-Buschleute. // *Sprache und Geschichte in Afrika*, 7.1, pp. 205-272.

König 2006: Christa König. Marked nominative in Africa. // *Studies in Language* 30, 4, pp. 705-82.

König & Heine 2001: Christa König, Bernd Heine. *The !Xun of Ekoka. A demographic and linguistic report*. Khoisan Forum, working paper No. 17. Köln: University of Cologne.

König & Heine 2008: Christa König, Bernd Heine. *A Concise Dictionary of Northwestern !Xun*. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 21. Köln: Rüdiger Köppe Verlag.

Kraft 1981: Charles H. Kraft. *Chadic wordlists. Vols. 1-2*. Berlin: Verlag von Dietrich Reimer.

Krönlein 1889: J. G. Krönlein. *Wortschatz der Khoi-khoïn (Namaqua-Hottentotten)*. Berlin: Deutsche Kolonialgesellschaft.

Ladefoged & Traill 1994: Peter Ladefoged, Anthony Traill. Clicks and their accompaniments. // *Journal of Phonetics*, 22, pp. 33-64.

Lanham & Hallows 1956a: Leonard Lanham, D. P. Hallows. Linguistic relationships and contacts expressed in the vocabulary of Eastern Bushman. // *African studies* (Johannesburg), 15 (1), pp. 45-48.

Lanham & Hallows 1956b: Leonard Lanham, D. P. Hallows. An outline of the structure of eastern Bushman. // *African studies* (Johannesburg), 15 (3), pp. 97-118.

Lionnet 2009: Florian Lionnet. *Lucy Lloyd's !Xun and the Ju Dialects*. // Presentation: Linguistisches Kolloquium, Humboldt Universität zu Berlin 10 Nov. 2009.

Maingard 1937: Louis F. Maingard. The †Khomani dialect of Bushman: its morphology and other characteristics. // *Bushmen of the southern Kalahari*. Ed. by J. D. Rheinallt Jones & C. M. Doke. Witwatersrand University Press, Johannesburg, pp. 237-275.

Maingard 1958: Louis F. Maingard. Three Bushman Languages. Part II: The Third Bushman Language. // *African studies* (Johannesburg), 17, pp. 100-115.

Maingard 1961: Louis F. Maingard. The central group of click languages of the Kalahari. // *African Studies* (Johannesburg), 20, pp. 114-122.

Matisoff 2000: James A. Matisoff. On the uselessness of glottochronology for the subgrouping of Tibeto-Burman. // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 333-371.

McAlpin 1981: David W. McAlpin. *Proto-Elamo-Dravidian: the Evidence and its Implications*. Philadelphia: The American Philosophical Society.

McMahon & McMahon 2005: April McMahon, Robert McMahon. *Language classification by numbers*. New York: Oxford University Press.

Meeussen 1963: A. E. Meeussen. Review of *The Languages of Africa* by J. Greenberg. // *Journal of African Languages*, 2, pp. 170-171.

Meinhof 1912: Carl Meinhof. *Die Sprachen der Hamiten*. Hamburg: L. Friedrichsen.

Meinhof 1929: Carl Meinhof. Versuch einer grammatischen Skizze einer Buschmannsprache. // *Zeitschrift für Eingeborenen-Sprachen*, 19, pp. 161-188.

Meinhof 1930: Carl Meinhof. *Der Koranadialekt des Hottentotischen*. Berlin: Dietrich Reimer Verlag.

Meriggi 1928/9: Piero Meriggi. Versuch einer Grammatik des [Xam-Buschmännischen]. *Zeitschrift für Eingeborenen-Sprachen* 19, pp. 117-153, 188-205.

Meyer-Lübke 1972: W. Meyer-Lübke. *Romanisches Etymologisches Wörterbuch*. Heidelberg: Carl Winter Universitätsverlag.

Militarev 2000: Alexander Militarev. Towards the Chronology of Afrasian (Afroasiatic) and its Daughter Families. // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 267-307.

Militarev 2005: Alexander Militarev. Once more about glottochronology and the comparative method: the Omotic-Afrasian case. // *Аспекты компаративистики I*. Moscow: RSUH Publishers, pp. 339-408.

Militarev & Kogan 2000: Alexander Militarev, Leonid Kogan. *Semitic Etymological Dictionary. Vol. I: Anatomy of Man and Animals*. Münster: Ugarit-Verlag.

Miller et al. 2007: Amanda Miller, Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter, Chris Collins. The Sounds of N|uu: Place and Airstream Contrasts. // *Working Papers of the Cornell Phonetics Laboratory*, 16, pp. 101-160.

Miller et al. 2009: Amanda Miller, Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter, Chris Collins. Differences in airstream and posterior place of articulation among N|uu clicks. // *Journal of the International Phonetic Association*, 39, 2, pp. 129-161.

Mous 2003: Maarten Mous. *The making of a mixed language: The case of Ma'a/Mbugu*. Amsterdam: John Benjamins' Publishing Company.

Nakagawa 1996: Hiroshi Nakagawa. A first report on the click accompaniments of |Gui. // *Journal of the International Phonetic Association*, 26 (1), pp. 41-54.

Nakagawa 1998: Hiroshi Nakagawa. Unnatural Palatalization in |Gui and ||Gana? // *Language, Identity, and Conceptualization among the Khoisan*. Ed. by Mathias Schladt. Köln: Rüdiger Köppe Verlag, pp. 245-263.

Nichols 1992: Johanna Nichols. *Linguistic Diversity in Space and Time*. University of Chicago Press.

Nichols 2010: Johanna Nichols. Macrofamilies, Macroareas, and Contact. // *The Handbook of Language Contact*. Ed. by Raymond Hickey. Wiley-Blackwell, pp. 361-379.

Nikolayev & Starostin 1994: Sergei Nikolayev, Sergei Starostin. *A North Caucasian Etymological Dictionary*. Moscow: Asterisk Publishers.

Nikolayeva 2006: Irina Nikolayeva. *A Historical Dictionary of Yukaghir*. Berlin — New York: Mouton de Gruyter.

Obst 1912: E. Obst. Von Mkalama ins Land der Wakindiga. II. Die Sprache der Wakindiga. // *Mitteilungen der Geographischen Gesellschaft in Hamburg*, 26, 1, pp. 29-45.

Orel 2003: Vladimir Orel. *A Handbook of Germanic Etymology*. Leiden — Boston: Brill.

Orel & Stolbova 1995: Vladimir E. Orel, Olga V. Stolbova. *Hamito-Semitic Etymological Dictionary: Materials for a reconstruction*. Leiden — New York — Köln: Brill.

Peiros 2000: Ilia Peiros. Family diversity and time depth. // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 75-108.

Pokorny 1958: Julius Pokorny. *Indogermanisches etymologisches Wörterbuch*. Francke Verlag, Bern & München.

Prost 1956: André Prost. *La langue soñay et ses dialectes*. Dakar: Mémoire IFAN 47.

Raa 2012: Eric ten Raa. *A Dictionary of Sandawe: The Lexicon and Culture of a Khoesan People of Tanzania*. Ed. by Christopher Ehret & Patricia Ehret. Köln: Rüdiger Köppe Verlag.

Reader 1997: John Reader. *Africa: A Biography of the Continent*. London: Hamish Hamilton Ltd.

Reh 1985: Mechthild Reh. *Die Krongo-Sprache (Ninò Mó-di). Beschreibung, Texte, Wörterverzeichnis*. Berlin: Dietrich Reimer Verlag.

Renfrew et al. 2000: Colin Renfrew, April McMahon, Larry Trask (eds.). *Time Depth in Historical Linguistics*. Cambridge: McDonald Institute for Archaeological Research.

Renfrew et al. 2006: Colin Renfrew, Peter Forster, James Clackson (eds.). *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute for Archaeological Research.

Ringe 1992: D. A. Ringe, Jr. *On Calculating the Factor of Chance in Language Comparison*. Philadelphia: The American Philosophical Society.

Ringe 1998: D. A. Ringe, Jr. Probabilistic Evidence for Indo-Uralic. // *Nostratic: Sifting the Evidence*. Ed. by J. C. Salmons & B. D. Joseph. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 153-197.

Rottland 1982: Franz Rottland. *Die Südnilotischen Sprachen: Beschreibung, Vergleichung und Rekonstruktion*. Berlin: Dietrich Reimer Verlag.

Ruhlen 1994: M. Ruhlen. *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford, California: Stanford University Press.

Rust 1969: F. Rust. *Nama Wörterbuch (Krönlein Redivivus)*. Pietermaritzburg: University of Natal Press.

Sands 1998a: Bonny Sands. *Eastern and Southern African Khoisan. Evaluating Claims in Distant Linguistic Relationships*. Ed. by Rainer Vossen. Köln: Rüdiger Köppe Verlag.

Sands 1998b: Bonny Sands. The Linguistic Relationship between Hadza and Khoisan. // *Language, Identity, and Conceptualization among*

the Khoisan. Ed. by Mathias Schladt. Köln: Rüdiger Köppe Verlag, pp. 265-283.

Sands 2001: Bonny Sands. Borrowing & Diffusion as a source of lexical similarities in Khoesan. // *Cornell working papers in linguistics*, 18, pp. 200-224.

Sands 2010: Bonny Sands. Juu subgroups based on phonological patterns. // *Proceedings of the 1st International Symposium, January 4-8, 2003, Riezlern / Kleinwalsertal*. Ed. by M. Brenzinger & Ch. König. Köln: Rüdiger Köppe Verlag, pp. 40-73.

Sands, Maddieson, & Ladefoged 1996: Bonny Sands, Ian Maddieson, Peter Ladefoged. The Phonetic Structures of Hadza. // *Studies in African Linguistics*, 25:2, pp. 171-204.

Sands et al. 2006: Bonny Sands, Amanda Miller, Johanna Brugman, Levi Namaseb, Chris Collins, Mats Exter. *1400 item N|uu Dictionary*. Ms.

Sands et al. 2007 = Bonny Sands, Amanda Miller, Johanna Brugman. The Lexicon in Language Attrition: The Case of N|uu. // *Selected Proceedings of the 37th Annual Conference on African Linguistics*. Ed. by Doris L. Payne & Jaime Peña. Somerville, MA, USA, pp. 55-65.

Schapera 1930: J. Schapera. The Khoisan Languages. // *The Khoisan peoples of South Africa*. London, Routledge, pp. 419-438.

Sidwell 1999: Paul Sidwell. 2nd Preface. // *Historical Linguistics & Lexicostatistics*. Ed. by Vitaly Shevoroshkin & Paul J. Sidwell. Melbourne: Association for the History of Language, pp. ix-xiii.

Snyman 1970: Jan W. Snyman. *An Introduction to the !Xū (!Kung) Language*. Cape Town: A. A. Balkema.

Snyman 1975: Jan W. Snyman. *ǀZu^hoāsi fonologie en woordeboek*. Cape Town: A. A. Balkema.

Snyman 1980: Jan W. Snyman. The relationship between Angolan !Xu and ǀZu^hoāsi. // *Bushman and Hottentot linguistic studies 1979*. Ed. by Jan W. Snyman. Pretoria: University of South Africa, pp. 1-58.

Snyman 1997: Jan W. Snyman. A preliminary classification of the !Xū and ǀZu^hoāsi dialects. // *Namibian languages: reports and papers*. Ed. by Wilfrid Haacke & Edward Elderkin. Cologne: Rüdiger Köppe Verlag for the University of Namibia, pp. 21-106.

Snyman 2000: Jan W. Snyman. Palatalisation in the Tsowaa and G||ana Languages of Central Botswana. // *The State of Khoesan*

Languages in Botswana. Ed. by H. M. Batibo & J. Tsonope. Tasalls Publishing, University of Botswana, pp. 33-43.

S. Starostin 1995a: Sergei Starostin. Response To L. Sagart's "Some Remarks on the Ancestry of Chinese". // *The Ancestry of the Chinese Language*. Journal of Chinese Linguistics Monograph Series, 8, pp. 393-404.

S. Starostin 1995b: Sergei Starostin. The historical position of Bai. // *The Moscow Linguistic Journal*, v. 1, pp. 174-190.

S. Starostin et al. 2003: Sergei Starostin, Anna Dybo, & Oleg Mudrak. *An Etymological Dictionary of Altaic Languages*. Leiden: Brill.

Starostin 2002: George Starostin. On the Genetic Affiliation of the Elamite Language. // *Mother Tongue* VII, pp. 147-170.

Starostin 2003: George Starostin. A Lexicostatistical Approach Towards Reconstructing Proto-Khoisan. // *Mother Tongue* VIII, pp. 81-126.

Starostin 2008a: George Starostin. From Modern Khoisan Languages to Proto-Khoisan: the Value of Intermediate Reconstructions. // *Аспекты компаративистики III. Orientalia et Classica (Труды Института восточных культур и античности)*, в. XIX. М., RSUH Publishers, pp. 337-470.

Starostin 2008b: George Starostin. *Making a comparative linguist out of your computer: problems and achievements*. Talk given at the Santa Fe Institute, August 12, 2008. On-line version: <http://starling.rinet.ru/Texts/computer.pdf>.

Starostin 2010: George Starostin. Preliminary lexicostatistics as a basis for language classification: a new approach. // *Вопросы языкового родства (Journal of Language Relationship)*, v. 3, pp. 79-117.

Steel et al. 1988: M. Steel, M. Hendy, & D. Penny. Loss of information in genetic distances. // *Nature*, 333, pp. 494-495.

Steiner et al. 2011: Lydia Steiner, Peter F. Stadler, Michael Cysouw. A Pipeline for Computational Historical Linguistics. // *Language Dynamics and Change*, 1 (1), pp. 89-127.

Story 1999: Robert Story. *K'u^ha:si manuscript (MS collections of the Ki^hazi dialect of Bushman, 1937)*. Ed. by Anthony Traill. Khoisan forum, working paper No. 13. Köln, University of Cologne.

Struck 1911: Bernhard Struck. Über die Sprachen der Tatoga und Irakwleute. // *Mittheilungen aus den Deutschen Schutzgebieten*, 4, pp. 107-132.

Swadesh 1952: Morris Swadesh. Lexico-statistical dating of pre-historic ethnic contacts. With special reference to North American

Indians and Eskimos. // *Proceedings of the American Philosophical Society*, 96, pp. 452-463.

Swadesh 1955: Morris Swadesh. Toward greater accuracy in lexicostatistical dating. // *International Journal of American Linguistics*, 21, pp. 121-137.

Tanaka 1978: Jiro Tanaka. *A San Vocabulary of the Central Kalahari: Gllana and G!wi Dialects*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Teeter 1963: Karl V. Teeter. Lexicostatistics and Genetic Relationship. // *Language*, vol. 39, No. 4, pp. 638-648.

Theil 2006: *Is Omotic Afroasiatic? A Critical Discussion*. Paper presented at The David Dwyer Retirement Symposium, Michigan State University, October 2006.

Thomason 1983: Sarah G. Thomason. Genetic relationship and the case of Ma'a (Mbugu). // *Studies in African Linguistics*, 14, 2, pp. 195-231.

Traill 1973: Anthony Traill. "N4 or S7": another Bushman language. // *African studies*, 32 (1), pp. 25-32.

Traill 1974: Anthony Traill. Westphal on "N4 or S7": a reply. // *African studies*, 33 (4), pp. 249-255.

Traill 1978: Anthony Traill. Research on the non-Bantu African languages. // *Language and communication studies in South Africa*. Ed. by Leonard Walter Lanham and Karel P. Prinsloo. Cape Town: Oxford University Press, pp. 117-137.

Traill 1980: Anthony Traill. Phonetic diversity in the Khoisan languages. // *Bushman and Hottentot linguistic studies 1979*. Ed. by Jan W. Snyman. Pretoria: University of South Africa, pp. 167-189.

Traill 1985: Anthony Traill. Phonetic and phonological studies of !Xóõ Bushman. *Quellen zur Khoisan-Forschung/Research in Khoisan studies*, Bd 1. Köln: Rüdiger Köppe Verlag.

Traill 1986: Anthony Traill. Do The Khoi have a place in the San? New data on Khoisan linguistic relationships. // *Sprache und Geschichte in Afrika*, 7.1, pp. 407-430.

Traill 1994a: Anthony Traill. *!Xóõ dictionary*. *Quellen zur Khoisan-Forschung/Research in Khoisan studies*, Bd 9. Köln: Rüdiger Köppe Verlag.

Traill 1994b: Anthony Traill. Structural Typology and Remote Relationships between Zhu and !Xóõ. // *Sprache und Geschichte in Afrika*, 16/17, pp. 437-454.

Traill 1995: Anthony Traill. Interpreting |Xam phonology: the need for typological cleansing. // *The complete linguist: papers in memory of Patrick J. Dickens*. Ed. by A. Traill, R. Vossen & M. Biesele. (Afrikanische Sprachen und Kulturen.) Köln: Rüdiger Köppe Verlag, pp. 509-523.

Trask 1996: Larry Trask. Response to Starostin. // *Mother Tongue II*, pp. 111-118.

Trask 2008: Larry Trask. *Etymological Dictionary of Basque*. Ed. for Web publication by Max W. Wheeler. University of Sussex.

Treis 1998: Yvonne Treis. Names of Khoisan Languages and their Variants. // *Language, Identity, and Conceptualization among the Khoisan*. Ed. by Mathias Schladt. Köln: Rüdiger Köppe Verlag, pp. 463-495.

Tryon 1995: Darrell T. Tryon (ed.). *Comparative Austronesian Dictionary: An Introduction to Austronesian Studies*. Berlin-New York: Mouton de Gruyter.

Tucker & Bryan 1974: A. N. Tucker, M. Bryan. The 'Mbugu' anomaly. // *Bulletin of the School of Oriental and African Studies*, 37, pp. 188-207.

Tucker, Bryan, & Woodburn 1977: A. Tucker, M. Bryan, J. Woodburn. The East African click languages: a phonetic comparison. // *Zur Sprachgeschichte und Ethnohistorie in Afrika*. Ed. by Wilhelm J. G. Möhlig, Franz Rottland & Bernd Heine, pp. 301-323.

Turner 1966: R. L. Turner. *A Comparative Dictionary of the Indo-Aryan Languages*. Oxford University Press.

Ullmann 1951: Stephen Ullmann. *The Principles of Semantics*. Glasgow: Jackson, Son & Company.

Vallaëys 1986: A. Vallaëys. *Dictionnaire Logo-Français suivi d'un Index Français-Logo*. Musée Royal de L'Afrique Centrale. Tervuren, Belgique.

Visser 2001: Hessel Visser. *Naro Dictionary*. Botswana: Printing & Publishing Co.

Vorbichler 1965: *Die Phonologie und Morphologie des Balese (Ituri-Urwald, Kongo)*. Glückstadt: Verlag J. J. Augustins.

Vossen 1982: Rainer Vossen. *The Eastern Nilotes: Linguistic and Historical Reconstructions*. Berlin: Dietrich Reimer Verlag.

Vossen 1984: Rainer Vossen. Studying the linguistic and ethno-history of the Khoe-speaking (central Khoisan) peoples of Botswana, research in progress. // *Botswana Notes and Records*, 16, pp. 19-35.

Vossen 1986: Rainer Vossen. Zur Phonologie der ||Ani-Sprache. // *Contemporary studies on Khoisan (Festschrift Oswin R. A. Köhler)*, v. 2. Ed. by R. Vossen & Klaus Keuthmann. Helmut Buske Verlag, Hamburg, pp. 321-345.

Vossen 1992: Rainer Vossen. Q in Khoe: borrowing, substrate or innovation? // *African linguistic contributions (Festschrift Ernst Westphal)*. Ed. by Derek F. Gowlett. Hamburg: Helmut Buske Verlag, pp. 363-388.

Vossen 1997a: Rainer Vossen. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Köln, Rüdiger Köppe Verlag.

Vossen 1997b: Rainer Vossen. What click sounds got to do in Bantu. // *Human Contact through Language and Linguistics*. Ed. by Birgit Smeja & Meike Tasch. Peter Lang, Europäischer Verlag der Wissenschaften, pp. 353-366.

Vossen 2000: Rainer Vossen. Khoisan languages, with a grammatical sketch of ||Ani (Khoe). // *Areal and genetic factors in language classification and description: Africa south of the Sahara*. Ed. by Petr Zima. Lincom Europa, München, pp. 129-145.

Vossen et al. 1988: Rainer Vossen, Sabine Neumann, Christina Patriarchi, Margit Rottland, Rainer Spörl, Beate Vagt. Khoe Linguistic Relationships Reconsidered: The Data. // *New perspectives on the study of Khoisan*. Ed. by Rainer Vossen. Quellen zur Khoisan-Forschung / Research in Khoisan studies, Bd 7. Köln: Rüdiger Köppe Verlag, pp. 67-108.

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.

Vries 1962: Vries, Jan de. *Altnordisches Etymologisches Wörterbuch*. Leiden: Brill.

WALDS I: *West African Language Data Sheets I*. Ed. by Mary-Esther Kropp-Dakubu. Legon & Leiden: West African Linguistic Society, 1977.

WALDS II: *West African Language Data Sheets I*. Ed. by Mary-Esther Kropp-Dakubu. Legon & Leiden: West African Linguistic Society, 1980.

Wang 2006: Wang Feng. *Comparison of Languages in Contact: The Distillation Method and the Case of Bai*. Institute of Linguistics, Academia Sinica.

Westphal 1962: Ernst O. J. Westphal. On classifying Bushman and Hottentot languages. // *African Language Studies*, 3, pp. 30-48.

Westphal 1964: Ernst O. J. Westphal. Review of *The Languages of Africa* by J. Greenberg. // *American Anthropologist*, 66, pp. 1446-1449.

Westphal 1965: Ernst O. J. Westphal. Linguistic research in SWA and Angola. // *Die ethnischen Gruppen Südwestafrikas*. Wissenschaftliche Forschung in Südwestafrika, Bd 3. Windhoek: Südwestafrikanische Wissenschaftliche Gesellschaft, pp. 125-144.

Westphal 1971: Ernst O. J. Westphal. The Click Languages of Southern and Eastern Africa. // *Linguistics in Sub-Saharan Africa*. The Hague/Paris: Mouton, pp. 367–420.

Westphal 1974: Ernst O. J. Westphal. Notes on A. Traill: "N4 or S7?". // *African studies*, v. 33 (1973), pp. 243-247.

Westphal 1980: Ernst O. J. Westphal. The age of "Bushman" languages in Southern African pre-history. // *Bushman and Hottentot linguistic studies*. Ed. by Anthony Traill. Communications from the African Studies Institute, no 2. Johannesburg: University of the Witwatersrand, pp. 59-79.

Wuras 1920a: C. F. Wuras. An Outline of the Bushman Language. // *Zeitschrift für Eingeborenen-Sprachen*, 10, pp. 81-87.

Wuras 1920b: C. F. Wuras. *Vokabular der Korana-Sprache*. Herausgegeben und mit kritischen Anmerkungen versehen von Walther Bourquin. Berlin: Dietrich Reimer Verlag.

Yeon-Ju, Sagart 2008: Lee Yeon-Ju, Laurent Sagart. No limits to borrowing: the case of Bai and Chinese. // *Diachronica* 25, 3, pp. 357-385.

Yigezu 2002: Moses Yigezu. *A comparative study of the phonetics and phonology of Surmic languages*. Laboratoire de phonologie, Université libre de Bruxelles (Thèse de doctorat).

Yigezu & Dimmendaal 1998: Moses Yigezu, Gerrit Dimmendaal. Notes on Baale. // *Surmic Languages and Cultures*. Ed. by Gerrit J. Dimmendaal & Marco Last. Köln: Rüdiger Köppe Verlag, pp. 273-317.

Ziervogel 1955: Dirk Ziervogel. Notes on the language of the eastern Transvaal Bushmen. // *The disappearing Bushmen of Lake Chrissie*, ed. by E. F. Potgieter. (Hidding-Currie publications of the University of South Africa, no 1.). J L van Schaik. Pretoria, pp. 34-64.

SUMMARY

«*Languages of Africa: an attempt at a lexicostatistical classification*» has been planned as a multi-volume monograph that aims at a complete, step-by-step re-evaluation of current hypotheses on the genetic classification of most of the languages, currently or until recently spoken on the African continent. The relevance of this task goes far beyond the current needs and issues of historical linguistics. In recent decades, significant progress has been achieved in recreating the human pre-history of Africa through important discoveries and systematizations in the fields of anthropology, archaeology, and population genetics, allowing for a thorough reassessment of earlier conceptions and beliefs on the subject. At the same time, the general «standard model» for the overall classification of Africa's languages, introduced by Joseph Greenberg more than half a century ago, still continues to serve as the default scheme of reference for linguists and non-linguists alike — not so much due to any exceptional robustness, inherent in the principles and methods according to which it was originally constructed, but rather due to a complete lack of a well-grounded alternative. Despite a plethora of new high-quality linguistic material that has been accumulated over the past fifty years, and despite the fact that Greenberg's methodology of «multilateral comparison» has been harshly criticized over the same period, leading more and more specialists in the field to doubt or even completely reject most of his «macrofamily» groupings, it remains obvious that, as long as no *constructive* challenge is presented, Greenberg's «quadripartite» scheme, according to which the absolute majority of Africa's languages falls into one of the four macrofamilies (Khoisan, Nilo-Saharan, Niger-Kordofanian, or Afro-Asiatic), will remain in active usage — for technical and pragmatic reasons, if nothing else.

The current study is the first part of an ongoing attempt to offer just such an alternative — present a re-classification of a large part (hopefully, all) of Africa's linguistic groupings that would be in better agreement not only with the enlarged and improved pool of language data which we now have at our disposition, but also with state-of-the-

art methodology that could offer objective, formalized, and historically credible criteria for properly separating chance resemblances, areal contacts, and traces of genuine genetic relationship within the data pool offered for comparison.

Reflecting this latter purpose, the larger half of Vol. I of the study represents a detailed account of this methodology. It is explained why the most diagnostic traces of genetic relationship are to be sought primarily in the so-called «basic lexicon», particularly when significant time depths are involved, and that, consequently, the major technique behind classification schemes has to be of a lexicostatistical nature. At the same time, it is clearly emphasized that «crude» lexicostatistics alone, based on quantifying degrees of phonetic similarity between words with the same meanings in different languages, regardless of whether it is conducted on a fully automated basis or by means of manual «cognate scoring» (both procedures have their relative ups and downs), is utterly insufficient to come up with a reliable classification model — because each individual case of such «cognate scoring» lacks a proper historical interpretation.

The proposed methodology is a complex combination of automated and manual lexicostatistical techniques with a detailed evaluation of «basic lexicon» subsets as to the phonetic, semantic, and distributional properties and individual histories of each of their elements. The most important stage of this procedure is the step-by-step reconstruction of «proto-lists» — identification of cognate sets that may be considered «optimal candidates» for expressing certain meanings, first on the proto-level of small language groups, then on the level of deeper language families, and so on. This involves elements of both phonetic and semantic reconstruction, carried out in accordance with the traditional comparative method and also relying quite heavily on the constantly increasing bank of data on the diachronic typology of phonetic and semantic change. The result is a fruitful synthesis of classic comparative linguistics with the better sides of J. Greenberg's methodology (in particular, the emphasis on «multilateral» rather than «binary» comparison) and a set of simple, but effective quantitative methods, allowing for the (relatively!) quick creation of working models for language classification that are based on historically interpreted evidence and, consequently, may serve as a launchpad for more detailed activity in the sphere of linguistic reconstruction.

The second part of the study applies the initial stages of this methodology to one of the most problematic of Greenberg's «macrofamilies» — the so-called «Khoisan» stock of Southern and Southeast Africa. Although, in recent decades, the work of such Khoisanologists as Tom Güldemann, Henry Honken, Bonny Sands, and others, has made it clear that serious comparative evidence for a genetic «Khoisan» unity is actually quite scarce, the term itself has survived, not only because, in addition to a genetic interpretation, it also has an areal / typological one (where «Khoisan» languages are basically understood as «click-containing» languages), but also because, so far, it has not been demonstrated beyond reasonable doubt that the evidence in question, as scarce as it might be, is altogether more typical for a situation of areal contact than it is for one of genetic relationship. In any case, with all the serious progress achieved in descriptive «Khoisan» linguistics over the past half-century, re-evaluating the old evidence on par with the new becomes a matter of the utmost relevance.

The «Khoisan» section of the study surveys most of the important data sources and uses them to reconstruct realistic proto-lists of 50 basic lexical items for each of the surveyed small uncontroversial groupings: North Khoisan (Ju), South Khoisan (!Wi-Taa), Central Khoisan (Khoe), as well as the alleged linguistic isolates (ǀHōã, Sandawe, Hadza, and the extinct Kwadi). The procedure allows to strip away many lexical and semantic innovations, as well as areal borrowings that hinder genetically-oriented comparison. The reconstructed lists are then subjected to several «rounds» of lexicostatistical analysis, including fully automated procedures of «blind» cognate scoring based on phonetic similarity, and re-calculations based on the results of manual correction that takes into account individual structural and historic properties of the compared data.

The intermediate results of this research yield «relationship signals» of different degrees of strength that confirm some of the existing hypotheses: in particular, very high cognate percentages are scored between North Khoisan and ǀHoan, which agrees with the earlier studies by the author of the current study and H. Honken (carried out independently of each other) on a «Ju-ǀHoan» grouping, and also between Central Khoisan and Kwadi, confirming the recent proposal, by T. Güldemann and D. Elderkin, of a «Khoe-Kwadi» family. Somewhat weaker, but statistically and historically significant signals, are attes-

ted for a deeper genetic relationship between «Ju-†Hoan» and South Khoisan (!Kwi-Taa), indirectly confirming the «Peripheral Khoisan» hypothesis, suggested by the author in some of his earlier publications; still weaker, but worth investigating, are the signals of an even deeper relationship between «Khoe-Kwadi» and Sandawe (an idea also supported by the research of D. Elderkin).

On the other hand, any «relationship signal» between «Peripheral Khoisan» and «Sandawe-Khoe-Kwadi» that could be spotted in the results turns out to be virtually indistinguishable from noise (which could be easily interpreted as a few chance resemblances mixed in with some arealisms that still slipped through the filters). The issue of whether the two groupings could still be related within an overall «Khoisan» macrofamily at a very deep time level remains open, but it can hardly be solved without taking into consideration the data of other African linguistic groupings, to some of which one or both of these families might, in fact, be related closer than to each other (the «click» factor notwithstanding). Likewise, no evidence whatsoever, apart from a few superficial similarities that do not exceed chance expectations, has been accumulated for a «Khoisan» affiliation of Hadza (this conclusion is similar to the results of Bonny Sands' research, conducted more than a decade earlier on a different basis).

It is important to understand, however, that all of the described results must be viewed as intermediate, since the existence of larger groupings, such as «Peripheral Khoisan», has yet to be further confirmed by the rigorous reconstruction of a higher level proto-list, and all the lists reconstructed for the smaller groupings have to be tested against analogous reconstructions for «non-Khoisan» groupings. These issues will be addressed in later volumes, once the initial survey of all the small level groupings has been completed.

The second volume of the study, expected to appear in print in the nearest future, will be applying the same methodology of study to the largest «building block» of Greenberg's equally suspicious «Nilo-Saharan» macrofamily — the so-called «Eastern Sudanic» languages, in an attempt to verify whether this term defines a genuine «genetic» unit, and if yes, which of the currently competing classification models of «Eastern Sudanic» stands closer to the historic truth.

Г. С. Старостин
ЯЗЫКИ АФРИКИ. ОПЫТ ПОСТРОЕНИЯ
ЛЕКСИКОСТАТИСТИЧЕСКОЙ КЛАССИФИКАЦИИ
Том 1: Методология. Койсанские языки

Художественное оформление обложки
Сергея Жигалкина

Подписано в печать 28.01.2013. Формат 60×90 ¹/₁₆.
Бумага офсетная № 1. Печать офсетная. Гарнитура Старлинг.
Усл. п. л. 31,9. Тираж 200. Заказ № 1462

Издательство «Языки славянской культуры».
№ госрегистрации 1037739118449.
Phone: +7 495 959-52-60 E-mail: lrc.phouse@gmail.com
Site: <http://www.lrc-press.ru>, <http://www.lrc-lib.ru>

Отпечатано в цифровой типографии «Буки Веди» на оборудовании Konica Minolta
ООО «Ваш полиграфический партнер», ул. Ильменский пр-д, д. 1, корп. 6
Тел.: (495) 926-63-96, www.bukivedi.com, info@bukivedi.com

Оптовая и розничная реализация — магазин «Гнозис».
117342, Москва, ул. Бутлерова, 17Б, офис 313
Тел.: 8 (499) 793-57-01, e-mail: gnoxis@pochta.ru
Костюшин Павел Юрьевич (с 10 до 18 ч.)