

Г. С. СТАРОСТИН

ЯЗЫКИ АФРИКИ

**ОПЫТ ПОСТРОЕНИЯ
ЛЕКСИКОСТАТИСТИЧЕСКОЙ
КЛАССИФИКАЦИИ**



Том I

Методология. Койсанские языки

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 количественный 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

Электронная версия данного издания является собственностью издательства, и ее распространение без согласия издательства запрещается.

СОДЕРЖАНИЕ

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 ʃхоан	
2.2.1. Общие сведения и источники	333
2.2.2. Историческая характеристика	335
50-словный список для языка восточный ʃхоан	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

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, \dots, 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, \dots, 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. Бэксстером [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 огорчиваясь, 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 сходжений, ср., например¹:

'кора' — кор. *k̄ap-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) ауслаутное **-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 лишённые 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 «нанализуется» 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. Вовиним 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 Гюльдемманн; 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 сходжений. Неравномерное распределение сходжений по подмножествам, выделенным на основании формального критерия (например, «устойчивости»), будет однозначно свидетельствовать в пользу того, что случайность — не единственный фактор, отвечающий за наличие таких сходжений.

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 вес; 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 схождений между которыми (по подсчетам А. Ю. Милитарева) также не превышает 10 единиц, 9 из 10 схождений 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. *irit* '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 схождений 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.