

### О пороге родства и индексе стабильности в базисной лексике при массовом сравнении: атлантические языки

В статье рассматриваются некоторые ключевые проблемы установления дальнего родства языков в условиях, когда отсутствует классическая ступенчатая реконструкция, а именно: 1) проблема установления порога случайных сходств в стословном списке Сводеша (при различных алгоритмах их фиксации); 2) проблема ранжирования базисной лексики по ее устойчивости; 3) проблема многочисленных изолированных корней, не находящих параллелей в других языках семьи. Эти общие сюжеты рассматриваются на основе конкретной проблематики построения генетической классификации атлантических языков. Атлантические языки традиционно считаются одной из ветвей нигеро-конголезской макросемьи, но при этом лексикостатистика говорит о том, что дата их разделения составляет не менее 8000 лет. Публикуемые два критических отзыва на статью в определенном смысле являются ее продолжением, развивая сложную проблематику установления дальнего родства языков; авторский ответ на эти отзывы следует также рассматривать как важный фрагмент публикуемой статьи.

*Ключевые слова:* праязыковая реконструкция, базисная лексика, лексикостатистика, этимология, атлантические языки, макрокомпаративистика.

В статье рассматривается ряд ключевых проблем в русле давней дискуссии о критериях установления дальнего родства языков. Для семей старше индоевропейской (ВІЕ) такая дискуссия по-прежнему актуальна, несмотря на публикацию нескольких фундаментальных работ, посвященных этому сюжету<sup>1</sup>. Речь пойдет о трудных случаях классификации (например, проблемы классификации языков семьи нигер-конго), когда «нельзя» опереться на классические методы (в отсутствие ступенчатой реконструкции), но «очень хочется» получить предварительные результаты, т. е. нас будут интересовать проблемы построения классификации не на основе регулярных соответствий, а на основе сходств в базисной лексике. Мы сосредоточимся в первую очередь не на общих рассуждениях на эту тему (они в изобилии присутствуют в литературе), а на конкретных критериях установления дальнего родства и границ их реального применения — прежде всего для таких трудных случаев, какими являются, в частности, атлантические языки, которыми я

---

<sup>1</sup> Так, в 2013 г. вышла в свет исключительно интересная монография Г. С. Старостина [Г. Старостин 2013], с которой во многом перекликается данная статья. В этой монографии представлен прекрасный обзор литературы по установлению критериев дальнего родства, в том числе и статистических. Публикация [Г. Старостин 2013] избавляет автора от необходимости системного представления всех идей и экспериментов в этой обширной области. Здесь же будут рассмотрены только такие аспекты проблемы, которые, как представляется, не привлекали специального внимания компаративистов.

занимаюсь. Если принять терминологию, предлагаемую Г. С. Старостиным («группа» — таксоны, показывающие от 40% и выше совпадений, «семья» — от 15 до 40%, «макросемья» — менее 15%, [Г. Старостин 2013: 140]), то атлантические языки — это в лучшем случае макросемья (по подсчетам Д. Сэпира, языки волоф-темне дают 3% (!) сходств [Sapir 1971]).

В [Г. Старостин 2013] постулируются и аргументируются три основных принципа, которые можно использовать в качестве критериев отделения родственных языков от неродственных в случаях предполагаемого дальнего родства: 1) принцип конкурирующих гипотез; 2) принцип регулярности соответствий и, наконец, 3) принцип «динамической градации базисной лексики». Рассматривая эти принципы на заведомо нелепом примере доказательства «родства» между центральносуданским языком лого и корейским языком, которые дают 10% сходств в стословнике Сводеша, Старостин показывает, что минимум 2 критерия из трех (1 и 3) не позволяют нам постулировать родство между этими языками. Эти же принципы, как показал Старостин, вполне позволяют «разобраться» с примером [Campbell & Poser 2008: 382—384], в котором рассматриваются сближения английского и маори (также 10 созвучных слов а стословнике).

Не будем здесь останавливаться на критерии «конкурирующих гипотез». Он основан на вполне разумном рассуждении (действительно, с учетом накопленных знаний легче допустить родство английского и немецкого, чем английского и маори), но практически неприменим для установления дальнего родства в Африке. Дело в том, что в большинстве трудных случаев у нас нет «конкурирующих гипотез», а точнее, у нас есть определенный набор равновероятных конкурирующих гипотез: классифицируя язык лимба, традиционно относимый к атлантическим, мы, учитывая крайне низкий процент его сходств с другими языками, можем с равным успехом допустить, что он относится, например, к бенуэ-конголезским языкам, или представляет собой отдельную нигер-конголезскую ветвь, или же вообще не является нигер-конголезским языком.

Остановимся подробнее на других критериях, главным из которых, конечно же, является процент сходств в базисной лексике.

## 1. О пороге дальнего родства

Существует ли некое минимальное количество сходств в списке Сводеша по двум языкам, которое позволило бы нам подозревать, что эти языки родственны? По мнению Г. С. Старостина, «наличие 10% совпадений в пределах стословного списка в принципе достаточно для постулирования глубокого родства между сравниваемыми языками — разумеется, при условии того, что совпадения эти не являются случайными, а отражают систему регулярных соответствий» [Г. Старостин 2013: 33]. Действительно, разнообразные эмпирические данные по макросемьям дают нам многочисленные примеры родственных языков, показывающих не более 10% сходств по Сводешу. Вопрос в том, сколько сходств должны давать случайные совпадения в неродственных языках? Чтобы определить порог случайности, необходимо: а) сформулировать, что мы будем считать фонетическим сходством, б) знать инвентарь фонем и словарные частоты фонем в сравниваемых языках.

Что считать фонетическим сходством? Предположим, что мы имеем дело только со структурами CVC (это основная структура в атлантических языках) и допускаем любые сближения согласных в рамках одного локального ряда (например, любых губных или любых палатальных). Тот же критерий мы применяем к сравнению гласных — с тем ис-

ключением, что центральные гласные мы имеем право сближать как с передними, так и с задними гласными. В среднестатистическом языке это дает нам 4 ряда у согласных и 2 ряда у гласных. При равных частотах для каждого ряда в начальной и в конечной позиции мы должны получить 3 случайных сходства в стословнике Сводеша по неродственным языкам. Четверть слов (т. е. 25) совпадет по рядности начального согласного. Четверть из них (т. е. 6) одновременно совпадет по рядности конечного согласного. Половина из них (т. е. 3) покажет сопоставимый гласный. Англ. *meat* ~ русск. *m'as-o* сходны по принятому критерию, поскольку они родственны, а их сходство с манджак *maf* 'мясо' случайно, но удивительно, как и сходство англ. *mouth* и бедик *m'isy* или англ. *liver* и бапен *dembar*.

Мы посчитали эту вероятность для случая равных частот согласных в каждой из двух позиций и в каждом из четырех рядов. Допустим теперь, что в конечной позиции у нас в обоих сравниваемых языках встречаются только согласные двух рядов — 50% дентальных и 50% велярных. В этом случае число случайных сходств в стословном списке = 6, то есть число ожидаемых случайных сходств увеличилось в два раза по сравнению с предыдущим примером. В целом отметим, что чем больше разброс частот в сравниваемых языках, тем больше оказывается вероятность случайных совпадений. Это, в частности, означает, что упрощенные прикидки порога случайности, исходящие из равенства частот всех согласных, позволяют определить **минимум** случайных сходств — их число, полученное по реальным частотам, всегда будет больше.

Подставим в наш пример реальные частоты. Средние частоты согласных, сгруппированных по рядам, в 40 атлантических языках представлены в табл. 1.

Таблица 1.

ATL (%)	Начальные	Конечные
LAB	31,3	15,0
DENT	22,5	46,9
PAL	18,0	14,6
VEL	28,0	24,0

Мы видим, что в конечной позиции ряды согласных представлены неравномерно: половина согласных в словарях атлантических языков относится к дентальному ряду. Для двух языков, близких к среднестатистическому атлантическому языку, эти частоты предполагают в стословнике 10 случайных совпадений начального и одновременно конечного согласного по ряду. И это без такой «роскоши», как допуск банальной палатализации, предполагающей сближение велярных и палатальных или дентальных и палатальных<sup>2</sup>. Учитывая, что качество гласных при хорошем совпадении согласных практически не принимается во внимание при установлении сближений в африканских языках<sup>3</sup>, приходится констатировать, что грань между случайным сходством и сходством при дальнем родстве практически стерта. Это становится тем более ясным, если учесть, что при пороге 10 число сходств от 7 до 13 не может считаться существенным отклонением, оставаясь в пределах трех сигм.

Как было показано выше, на порог случайных сходств существенно влияют частоты согласных. А это означает, что лексикостатистика на основе сходств в атлантических язы-

<sup>2</sup> То есть такие сближения, как диола-гусилай *teeη* с англ. *tau*, не будут считаться сходством.

<sup>3</sup> Гласные, как известно, как правило менее стабильны, чем согласные. К тому же избранный принцип (сближение только внутри ряда) для гласных существенно более уязвим, чем для согласных.

ках, показывающих от 5 до 20% сходств для языков разных ветвей, практически неприменима для построения атлантического дерева, поскольку небольшие различия коэффициентов могут определяться особенностями частотного распределения в конкретных парах языков.

Приведем два фактора, существенно повышающих и без того высокий порог случайных сходств.

Первый фактор отчетливо проявляется в языках с включенными в корни архаичными префиксами именных классов, то есть в большинстве нигеро-конголезских языков. Проиллюстрируем его на примере атлантических языков. Когда мы вычисляем процент односложных слов (суммарно для структур V, VC, CV, CVC) отдельно для существительных и глаголов в больших словарях атлантических языков, мы обнаруживаем, что исключительно дивергентные атлантические языки едины в одном: процент моносиллабов в существительных никогда не превышает их процент в глаголах, а в подавляющем большинстве языков он существенно выше, чем в глаголах (в среднем по атлантическим языкам в 2 раза [23% моносиллабов в существительных и 47% — в глаголах]), о чем наглядно говорят данные табл. 2 по 24 атлантическим языкам.

Таблица 2.

%	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	СРЕДН
N	3	9	50	49	42	2	1	67	32	69	3	75	42	45	7	4	38	32	38	14	48	6	4	4	23
V	60	88	61	62	47	2	1	83	49	91	63	92	59	79	7	80	67	94	87	39	49	6	32	16	47

Атлантические существительные существенно длиннее, чем глаголы. Отмечу, что для многих языков подсчеты производились по словарям корней, а не слов. Это может означать только одно: в атлантических существительных системно присутствуют утратившие продуктивность префиксы именных классов, которые мы вынуждены сегодня воспринимать как корневые фонемы. Следовательно, в наших сближениях по сходству мы в реальности часто сближаем не корневые начальные согласные, а согласные архаичных префиксов, что неизбежно повышает процент иллюзорных сходств — уже потому, что согласных, входящих в префиксы классов, существенно меньше, чем корневых согласных.

Приведенная выше таблица хорошо иллюстрирует и еще один фактор, резко повышающий процент случайных сходств в таких языках, как атлантические. В многочисленных моносиллабичных корнях, многие из которых имеют структуру CV, для сходства достаточно совпадения лишь одного согласного. К тому же для пары родственных языков, в одном из которых наблюдается повышенное число корней CV, а в другом — CVC, логично предположить соответствие этих структур (возникшее, например, в результате падения некоторых конечных согласных в одном из языков). Но включение пар вроде *bow* ~ *bo* в потенциально возможные сближения также существенно повышает процент ожидаемых случайных сходств.

Конечно же, формализация критериев сходства неоднократно становилась предметом обсуждения в литературе, посвященной дальнему родству. В монографии Г. С. Старостина подробно рассматриваются различные критерии: от наиболее строгих («согласные и гласные одинаковы или различаются только по одному признаку» [Г. Старостин 2013: 61, 62]<sup>4</sup> до существенно более гибких — введение «фонетически изоморф-

<sup>4</sup> Этот критерий использует, в частности, Д. Сэпир в своей лексикостатистике по атлантическим языкам [Sapir 1971].

ных параллелей» как «некоторого компромиссного варианта между „классическими регулярными соответствиями“ и аморфным „фонетическим сходством“ в рамках теории „массового сравнения“» [Г. Старостин 2013: 55]. К компромиссным попыткам формализации консонантных сходств относится и выделение 10 непересекающихся консонантных классов А. Б. Долгопольского и 12 классов С. А. Старостина [Г. Старостин 2013: 185—186].

Для интересующего нас сюжета важно то, что каждый из этих подходов предполагает очень разное число ожидаемых случайных сходств. Максимально строгий подход сводит их число практически к нулю, однако и число наблюдаемых сходств сокращается настолько, что коэффициенты бессмысленно сравнивать. При строгом критерии консонантных сближений для русского и английского мы получаем 10 сходств, и это не учитывая расхождения гласных: *cold* — *xolodnyj*, *full* — *polnyj*, *meat* — *m'aso*, *new* — *novyj*, *no* — *ne*, *nose* — *nos*, *sit* — *sid'et'*, *sun* — *solntse*, *two* — *dva*, *water* — *voda*. Отмечаем, что критерий позволяет выявить приблизительно треть фактических соответствий. Для более древнего родства можно в лучшем случае рассчитывать на несколько сходств. В атлантических языках, например между языками волоф и лимба, мы получим только 2 подходящих сближения, а именно: 'птица' (волоф *picc* ~ лимба *beci*, где *be-*, возможно, является префиксом именного класса!) и 'сидеть' (волоф *toog* ~ лимба *dɔŋɔ*, если считать, что **-g** и **-ŋ** различаются одним признаком, а не двумя). Можно еще добавить сюда 'колени' (волоф *oot* ~ лимба *hibiŋ*), но и здесь, строго говоря, мы выходим за рамки различия по одному признаку между **-m** и **-b**, учитывая наличие **-b** в системе. Но и порог случайности, исчисленный для этой пары языков на основе частот их согласных, дает 2 случайных совпадения первого и второго согласных. Несмотря на то что «строгий» критерий, в силу указанных причин, непригоден для классификации, он может быть использован для предварительного определения статуса языков, которые мы считаем изолированными. С большой вероятностью можно утверждать, что языки, дающие по строгому критерию 5 сходств, являются родственными<sup>5</sup>.

Выделение десятка непересекающихся классов не улучшает дела. Проблемы, связанные с применением классов и построенных на их использовании формальных алгоритмов сближений, прекрасно рассмотрены в [Г. Старостин 2013: 185—190]. Абсолютно справедливо замечание Г. С. Старостина о принципиальной ложности исходного посыла: «... фонетические изменения происходят в пределах одного и того же консонантного класса» [Г. Старостин 2013: 188]. Совершенно прав Г. С. Старостин и в том, что реально мы имеем дело с «пересекающимися подмножествами общего консонантного инвентаря» [Г. Старостин 2013, 189]. Заметим, что критерий консонантных классов является столь же рестриктивным, что и критерий полного сходства (или отличия по одному признаку внутри локального ряда). Исчисленный по обоим критериям на основе реальных частот порог случайных совпадений для пары языков волоф и гола дает одинаковый результат: 2 случайных совпадения по согласным. К тому же возникает простой методологический вопрос: как быть, когда, например, класс R, включающий только вибранты, не представлен в одном из языков? В гола отсутствует фонема /r/. Это означает, что /r/ волоф вообще ничему не может соответствовать в гола.

<sup>5</sup> Этот вывод подкрепляется, в частности, серией тестов, в которых сопоставлялись списки атлантических языков со списками многочисленных языков за пределами Африки. Применение «строгого» критерия (полное совпадение либо различие по одному признаку в рамках локального ряда) ни в одной паре языков не дает больше двух сходств. 10 «строгих» сходств между русским и английским при глубине расхождения более 6000 лет представляется на этом фоне беспорным свидетельством родства этих языков.

С более «либеральными критериями» дело обстоит сложнее. В поисках сходств имеются три ресурса, позволяющие довести порог случайных совпадений до 100%, а именно: а) вольное обращение со сравниваемыми формами; б) допущение семантических изменений; в) подмена парного сравнения сравнением произвольно выбранных пар языков из двух семей для каждого отдельного значения. Чаще всего эти допущения ничем не оправданы и служат для «научного обоснования» националистических теорий: приблизительно по одному алгоритму строятся доказательства происхождения волоф от языка египетских пирамид или патриотические этимологии Михаила Задорнова. С Задорновым спорить бессмысленно — отметим лишь, что мы имеем дело практически со стопроцентной вероятностью обнаружения случайных внешних сходств для словаря любого языка. Более интересны, естественно, алгоритмы поиска сходств, которые используются профессиональными лингвистами.

Приходится признать, что в лексикостатистике мы допускаем фонетические изменения, но не признаем изменений семантических не потому, что таковых нет. Просто, как только мы «отпустим» тормоза с семантикой, мы мгновенно потеряем возможность опереться на список значений, а вероятность случайных сходств возрастет на порядок. Поэтому большинство лингвистов, прибегающих к лексикостатистике, семантику «не трогают»<sup>6</sup>.

Редко нарушается и принцип парного сравнения. Если попарное сравнение языков подменяется выбором созвучной формы в одном из многочисленных языков семьи или группы, так, как это сделано у Дж. Гринберга, то вероятность обнаружения случайных сходств приближается к 100%. Так, в первой десятке списка Сводеша, без введения семантических изменений, можно обнаружить следующие сходства между английским и атлантическими языками (табл. 3):

Таблица 3.

1	all	
2	ashes	limba <i>hɔtɔ</i> , temne <i>r-os</i>
3	bark	joola-kasa <i>ka-fongol</i> , temne <i>foki</i>
4	belly	wolof <i>biir</i> , joola <i>far</i> , limba <i>fɔle</i>
5	big	sereer <i>mag-in</i> , joola <i>mlomp bə-məki</i> , manjak / pepel <i>mag</i> , nalu <i>baŋ, gboŋ</i> , бага <i>maduri baki</i>
6	bird	
7	bite	wolof <i>màtt</i>
8	black	bedik <i>-bàlà</i> , biafada <i>baŋ</i> , peul <i>balee</i> , sereer <i>baal</i> , nalu <i>bal</i> , limba <i>boloi</i>
9	blood	jaad <i>poada</i> , buy <i>biler</i> ,
10	bone	joola <i>wul</i> , manjak <i>moɓ</i> , pepel <i>mɔ</i> , nalu <i>wol</i> , бага- <i>maduri</i> / temne <i>ke-bant</i>

<sup>6</sup> В анонимной рецензии на рукопись статьи содержалось следующее рассуждение: «Автор не прав, строгий ответ иной. Мы „не трогаем“ семантику, потому что семантическое тождество предполагает *по определению*. Без условия тождества значений лексикостатистика превращается в т. н. этимостатистику (фактически не разработанный метод сравнения)». Это возражение я принять не могу, и не только потому, что оно сформулировано не вполне внятно, но и по сути. Если имеется в виду, что семантическое тождество предполагается по определению, то законен вопрос — почему? Представляется, что эта конвенция, предложенная еще Сводешем, продиктована в первую очередь тем, что учет любого семантического изменения автоматически уводит нас от фиксированного списка значений, состоящего ровно из 100 единиц.

Удивительно не то, что для каждого английского слова в семье из 40 языков находятся созвучные слова. Удивительно скорее, что для двух из них не нашлось близких по звучанию слов.

Отдельно подчеркну следующее. Меньше всего автору хотелось бы, чтобы эти строки воспринимались читателем как очередная попытка ревизии уникальной работы, проделанной Дж. Гринбергом. Именно Гринберг сформировал облик современной африканской компаративистики. Г. С. Старостин пишет о «парадоксальной ситуации», когда большинство лингвистов отвергают метод «массового сравнения», но при этом опираются на гринберговскую классификацию, «в силу полного отсутствия альтернативных схем» [Г. Старостин 2013: 7]. Мне представляется, что этот «парадокс» решается просто: у самого Гринберга не было никакой формализованной методики сравнения. Его этимологические сближения следует воспринимать как «точечные» (в силу огромного материала) иллюстрации к результатам, основанным не на конкретной процедуре анализа, а на энциклопедических знаниях и исключительном исследовательском таланте. Два последних качества не всегда имеются у последователей Гринберга, которые используют его публикации для методологического обоснования массового непарного сравнения, которое может гарантировать «доказательство» любого родства.

Оставим общие рассуждения на эту тему и посмотрим, какие именно звуковые сходства допускал Дж. Гринберг. Это позволит нам оценить вероятность случайных сходств, если использовать его разрешения на совместимость в случае попарных языковых сравнений. Обратимся к одной из наиболее представительных его публикаций [Greenberg 1963]. В ней на примере 49 корней и их рефлексов обосновывается родство нигеро-конголезских языков. В своих иллюстрациях Гринберг старается подбирать слова с одинаковыми гласными, но, если это не получается, сближает любые гласные без ограничений. Минимум 22 корня из 49 имеют во всех рефлексах структуру CV, т. е. включают только один согласный. В 27 корнях для многих языков дается второй — неначальный — согласный. Понятно, что одинаковые согласные сближать разрешается. Если говорить о сближениях согласных разного качества или о соответствиях различных согласных нулю, то разных сближений этого типа в списке имеется 111 для начальных согласных и 157 для неначальных. Многие неоднозначные сближения представлены в различных корнях: так, например, сближение  $n = \eta$  в начальной позиции отмечено в 10 корнях (корни 1, 11, 16, 19, 23, 24, 32, 34, 36, 46 в нумерации Гринберга).

Выделим 20 наиболее часто встречающихся у Гринберга согласных и представим сближения согласных разного качества у Гринберга в таблице 4.

Отметим только основные особенности таблицы. Лабиальные и дентальные сближаются в основном внутри локального ряда. При этом для дентальных это справедливо как для начальных, так и для неначальных, а для лабиальных — исключительно для начальных: у Гринберга нет ни одного примера сближения в неначальной позиции двух разных лабиальных согласных. Палатальные, а также велярные согласные практически сближаются только с одинаковыми согласными в обеих позициях. Носовые согласные свободно сближаются как по ряду, так и по серии (поэтому в таблице они сгруппированы). В начальной позиции сближения согласных с нулем не допускаются. Напротив, в неначальной позиции они используются Гринбергом чрезвычайно широко — особенно для плавных и носовых.

Определим, сколько случайных совпадений при этом алгоритме мы должны получить, сравнивая список Сводеша в двух любых языках (при равных частотах всех согласных в обеих позициях). Разрешения для начальных согласных составляют 25% (в таблице, включая диагональ, заполнено 58 клеток из 231). Разрешения для неначальных согласных составляют 32% (включая диагональ, заполнено 75 клеток из 231). Вспомним, что

Таблица 4.

	p	f	b	v	w	t	r	d	l	c	s	j	y	k	h	g	m	n	ɲ	η	∅		
p	—	5	7	4													1						
f		—	5	3																			
b			—	7													1						
v				—													1						
w					—																		
t					1	—	1	2		1	6												
r					1	3	—	7	7									1	1				
d					1	1	4	—	8									1	1				
l					1	2	8	6	—									1	1				
c										—	1												
s						1	1				—	1		1		1						1	
j												—											
y										1			—										
k					1									—		2						1	
h															—								
g					1					1			1	1		—						1	
m																	—	2	1	1			
n							2	1	1					1			1	—	10	1			
ɲ					1	1	2	1	1									3	—	1			
η							2	2	2	1			1	1		1	2	6	1	—			
∅					3	3	12	6	10	1	1		2	2	1	2	3	9	4	13			—

Комментарий к таблице: в ячейках указаны количества корней, в которых отмечено то или иное сближение; выше диагонали в таблице фиксируются сближения для начальных согласных, ниже диагонали — для неначальных.

22 корня из 49 у Гринберга имеют структуру CV, то есть для 48% слов достаточно совпадения только начального согласного. Это дает нам в списке Сводеша уже 12 случайно совпавших слов ( $100 \cdot 0,25 \cdot 0,48$ ). Для 52% нам необходимо совпадение начального и неначального согласных, что дает еще 4 слова ( $52 \cdot 0,25 \cdot 0,32$ ). Итак, при алгоритме Гринберга в списке Сводеша случайное сходство должно обнаружиться у 16 слов! А это уже не просто родство — это уже семья, а не макросемья по Г. С. Старостину. К тому же напомним, что допущение о равенстве частот согласных (которого никогда не бывает) дает минимальное число случайных сходств — чем больше различия частот, тем больше окажется случайных сходств. Это значит, что реально по алгоритму Гринберга, не учитывающему различия гласных, число случайных сходств будет больше 16. В таком случае, имея два любых словаря по 1000 слов, мы можем продемонстрировать родство языков, представив убедительный список в 160 корней<sup>7</sup>.

<sup>7</sup> Привожу это очевидное рассуждение, учитывая явное статистическое недоразумение в [Г. Старостин 2013, 42], касающееся процента случайных совпадений в «слабой» лексике. В приводимом примере, как пишет автор, «ожидалось бы скорее число, близкое к нулю», хотя в действительности ожидание составляет 50%.



Таблица 5.

	p	t	c	k	f	r	s	h	b	d	j	g	w	l	y	∅	m	n	ñ	η
p	1				1			1	1				1			1				
t		1	1			1	1			1				1						
c		1	1	1		1	1	1			1				1					
k			1	1			1	1				1	1			1				
f	1				1	1	1	1	1				1			1				
r		1	1		1	1	1	1			1			1	1	1				
s		1	1	1	1	1	1	1						1	1					
h	1		1	1	1	1	1	1				1	1		1	1				
b	1				1				1			1	1			1	1			
d		1								1	1			1				1		
j			1			1				1	1	1			1				1	
g				1				1	1		1	1	1		1	1				1
w	1			1	1			1	1			1	1		1	1	1			1
l		1				1	1			1				1	1	1		1		
y			1			1	1	1			1	1	1	1	1	1			1	
∅	1			1	1	1		1	1			1	1	1	1	1		1		
m									1				1				1	1		1
n										1				1		1	1	1	1	1
ñ											1				1			1	1	1
η												1	1				1	1	1	1

На практике в африканской макрокомпаративистике, как показывает анализ конкретных публикаций, алгоритмы выделения когнатов достаточно «либеральны». Наиболее типичными являются следующие критерии фонетической совместимости:

- 1) признаются сходными любые два согласных одного места образования (например, любые два лабиальных);
- 2) применительно к велярным и денальным (особенно перед передними гласными) допускается «реконструкция палатализации», т. е. допускается сближение  $ti \sim ci$ ,  $ge \sim je$  и им подобные;
- 3) допускается сопоставление слов CV и CVC, особенно в случаях, когда конечный согласный «слабый» — например, носовые сонанты **w, l, y** (а также **n**) или фрикативные — например, **f, h**;
- 4) допускается в тех же случаях совпадение слов VC и CVC с предположительным падением начального согласного или его интерпретацией как префикса именного класса;
- 5) допускаются сближения CCV и CV (особенно если второй согласный кластера — плавный), а также CC и CVC;
- 6) различие гласных часто игнорируется, особенно в тех случаях, когда два согласных в структурах CVC показывают хорошую совместимость;
- 7) если три первые звука предполагаемых корней показывают сходство по указанным признакам, различие последующих звуков объясняется суффиксной морфологией или просто игнорируется.

Впрочем, иногда легче дать список, чем формализовать все правила его построения. Чтобы приблизительно оценить порог случайных сходств при допущении наиболее типичных соответствий в атлантических языках, допустим, что у нас есть два языка с одинаковым инвентарем, включающим 20 согласных в обеих позициях, причем частоты всех согласных равны. Отметим в таблице допускаемые сходства (табл. 5).

В нашей таблице 5 заполнено 158 ячеек из 400, то есть 40%. При равных частотах это дает  $100 * 0,4 * 0,4 = 16$  случайных сходств начального и конечного согласных в словнике по двум языкам, а если ввести 2 непересекающихся класса гласных — соответственно 8 сходств. Если у строгого читателя 10 сходств вызывают активный протест, он может исключить их, и в таком случае вероятность случайного сходства согласных уменьшится до 14.

И наконец, рассмотрим реальный пример с реальными частотами. Определим, сколько сходств мы должны случайно получить для языков волоф и гола. Пример интересен тем, что гола в определенном отношении резко отличается от других языков, традиционно относимых к атлантическим, — в конечной позиции 88% слов гола имеют гласный (в волоф — 33%). Приведем частоты конечных в гола (табл. 6).

Таблица 6.

j	g	m	ŋ	n	l	∅
0,1%	0,1%	1,0%	1,6%	4,1%	5,4%	87,8%

При таком распределении, казалось бы, уже практически не важны частоты конечных согласных, как, впрочем, и индивидуальные решения лингвиста, касающиеся допускаемых сближений конечных согласных волоф и гола: количество случайных совпадений будет определяться по существу частотами только начальных согласных. И тем не менее наши индивидуальные решения даже в данном нетипичном случае принципиально меняют порог случайных сходств. Для «чистоты эксперимента» я попросил заполнить матрицу разрешений волоф—гола моего коллегу, специалиста по атлантическим языкам Гийома Сежерера, который исповедует строгий подход к сближениям (на мой взгляд, слишком строгий). На основе его заполнений случайность определяется как 11 слов (без учета различия гласных). Я применял достаточно либеральный подход, и на основе моих заполнений порог случайности оказался в 2 раза выше и составил 22 слова. В свою защиту отмечу интересное обстоятельство. При «строгом» подходе Г. Сежерера слова волоф с конечными глухими смычными (в частности, высокочастотный конечный -t) и тем более конечные геминированные глухие смычные не могут считаться совместимыми ни с одним из конечных звуков языка гола. Но ведь если мы имеем дело с родственными языками, то чему-то они соответствуют, например, нулю?! Этот пример интересен следующим. Он наглядно показывает, что, в противоположность часто высказываемому мнению, опора на внешнее сходство часто дает меньше сближений, чем выявление регулярных соответствий. Никакое самое вольное понимание сходства не позволит выявить соответствие русск. *дым* и фр. *fumée*. В любом случае отметим, что при имеющихся частотах согласных вероятность их случайного сходства в волоф и гола превышает 10, а приблизительно столько реальных сходств атлантические языки разных ветвей и показывают.

Проверим наши подсчеты на реальных примерах. Сопоставим английский список Сводеша со списками атлантических языков. Смысл такого рода сопоставлений не в том, чтобы предоставить читателю еще несколько примеров курьезных сходств в заведомо неродственных языках, а в том, чтобы по мере накопления данных эмпирически нащупать реальный порог случайных сходств при реально используемых алгоритмах сближения слов.

Таблица 7.

Temne	
ashes	r-os
bark	foki
bone	ka-bant
fly	faliɾ <sup>w</sup>
go	kɔ
me	mi
little	lɔl
one	in
sand	a-sənt
seed	ma-sənt, ma-ther
woman	bom

Biafada	
fire	fuuru
fly	fiili
me	mma
mouth	mməsə
nose	nasin
say	cii, re
white	paas
you sg.	yi
black	baŋg?
breast	bəl?
ground	gal?
fish	yesä / bwa-?
little	tiiti?
rain	səna?

Bedik	
black	-bála
breast	ε-bār
burn	o-bód
hand	ε-kònd
hear	o-hér
me	-mé- / mé-
mouth	b̄-más
white	-fèšà / -pèšà
eye(s)	gí-nguùs?
knee	é-dà-kàlà?
liver	i-dám-bél?
meat	ɲás?

Nalu	
black	bal
drink	dig
go	kɔ
star	ma-saar
you sg.	yi
big	wək? -baŋ?
bone	wol?
breast	n-bin?
fire	mət?
good	kuun?
tree	n-ti?
we	biye?

Limba	
belly	fɔle
black	boloi
breast	hu-bili
fish	fee
tooth	hu-titi
you sg.	yi
ashes	hɔtɔ? ~ fɔtɔ?
drink	tie?
eye	hoya ~ foya?
kill	kɔra?

Mankany	
big	-mag
liver	u-deeb
meat	u-ye-maɕ
nail	ka-ñiiri
what	we
white	p-faaθ
breast	p-wil?
hair	ka-weel?
leaf	ka-toh?
nose	b-yiis?

В наш список были включены 41 атлантический язык (включая несколько языков мель), а также один язык другой нигеро-конголезской группы — манде, а именно язык мандинка. Число сходств с английским списком колеблется в них от 3 (языки биджого, суа, шербро) до 14 (язык биафада). В среднем атлантический язык дает 8,5% случайных сходств с английским<sup>8</sup>. 12 языков показывают 10 и более процентов сходств. Приведем созвучные слова в 6 из них<sup>9</sup> (табл. 7).

<sup>8</sup> Отметим, что сравнение атлантических слов с английским системно дает существенно больше сходств, чем с французским или с русским. Мне не представить, с чем это могло бы быть связано. Английские слова более «прототипичны», легче модифицируются в соответствии с указанным алгоритмом сравнения?

<sup>9</sup> Знаком вопроса отмечены менее «убедительные» сближения.

Следует сжиться с фактом, что десяток сходств при реально существующих алгоритмах их выделения — это всего лишь норма случайных созвучий между любыми языками. А это означает, что на глубине ВЕ расхождения в процентах сходств ни о чем не говорят. Иными словами, они не могут быть, например, использованы для установления конфигурации нигеро-конголезского и даже атлантического древа.

В таких случаях в теории макрокомпаративистики предлагается подключать другой критерий, подробно обсуждаемый в [Г. Старостин 2013], а именно регулярность звуковых «соответствий» в наблюдаемых сходствах. Но этот критерий применить очень непросто.

Во-первых, в десятке слов регулярность может не проявиться — слишком беден материал, а привлекая сходства за пределами списка Сводеша, нетрудно подтвердить все что угодно.

Во-вторых, и заведомо неродственные языки могут обнаруживать регулярность сходств. Рассмотрим приведенные в таблице сходства английского и темне. В 11 приведенных примерах мы обнаруживаем 4 «соответствия» звонких в английском — глухим в темне: *bark* — *foki*, *go* — *ko*, *sand* — *sant*, *seed* — *sant*. При этом ни одного обратного примера в списке нет! В одной паре отмечено «соответствие» **b-** ~ **b-**, но при желании компаративист легко может объяснить это отклонение, реконструировав в этом случае, например, **\*bh-**, и уточнить диахроническое правило: в английском звонкие простые оглушаются, а придыхательные утрачивают аспирацию.

Рассмотрим регулярность «соответствий» в английском и бедик. Констатируем, что во всех трех случаях, англ. **b-** соответствует преглоттализованному **ɸ-** в бедик и никогда простому **b-**. Отмечаем, что когда в английском имеется кластер со вторым плавным или **-n**, в бедик ему регулярно соответствует последовательность CVC (конечно же V в бедик — эпентетический гласный!): *black* — *bálà*, *breast* — *ɸár*, *knee* — *kàlà*. Конечное **-t** (вне кластеров) дает в бедик фрикативный: *meat* — *ɸás*, *white* — *fèšà*. Трудно ожидать большей регулярности в 12 словах!

Признаюсь, что в отдельных случаях сближения явно неправомерны. Так, *nose* нельзя сближать с биафада *ɸasin*, поскольку **ɸa-** является в данном случае префиксом именного класса, а корень здесь — *sin*. Но дело в том, что именно этот фактор представляет основную проблему и при сравнении атлантических языков между собой, когда мы очень часто не можем провести аргументированное морфемное членение слов и, соответственно, получаем сближения, аналогичные рассмотренному.

Если десяток созвучных слов не является признаком родства, а критерий регулярности соответствий, как и критерий «конкурирующих гипотез», также не вполне эффективен, остается только один критерий, который был впервые четко сформулирован С. А. Старостиным и подробно рассмотрен в [Г. Старостин 2013] — ранг стабильности единиц списка Сводеша. Рассмотрим перспективы его применения к атлантическим языкам.

## 2. О рангах стабильности

В своей фундаментальной работе «Определение устойчивости базисной лексики» С. А. Старостин заложил основу для сопоставления диахронической стабильности слов в языках мира [С. Старостин 2007]. На огромном материале (737 языков в 14 семьях!) автор эмпирически определил относительную устойчивость слов в списке Сводеша и предложил критерий ее абсолютной оценки. Создана, по существу, новая область диахронической семантики. К сожалению, полученные Старостиным результаты широко не обсуждаются в публикациях по компаративистике — практически единственным исключением является

ся монография Г. С. Старостина [Г. Старостин 2013], в которой предложенный С. А. Старостиним критерий системно рассматривается в связи с проблемой установления дальнего генеалогического родства. В своей небольшой и предельно насыщенной статье С. А. Старостин не имел возможности не только останавливаться на крайне любопытных деталях, но и дать разностороннюю оценку обработанных им данных. Прежде чем интегрировать атлантический материал в сравнительные данные Старостина, остановимся на некоторых следствиях, вытекающих из его анализа, которые в указанной статье не эксплицированы.

С. А. Старостин вводит следующий индекс стабильности: «отношение максимально-го числа языков, использующих для данного значения один и тот же корень, к общему числу языков в семье» [С. Старостин 2007: 827]. Так, по Старостину, индекс стабильности в славянских языках для значения «кора» = 1 (корень *\*(s)korā* представлен в 13 языках из 13), а для значения «живот» = 0,38 (максимально широко представленный корень *\*brjuxъ* / *\*brjuxo* отмечен в 5 языках из 13, менее распространенные корни *\*životъ*, *\*měxъ*, *\*trъbuxъ*, а также заимствования не учитываются).

Вычислим по данным Старостина средний индекс стабильности для одного значения списка в 14 семьях<sup>10</sup> (табл. 8).

Таблица 8.

	YEN (енисейск.)	TH (тайск.)	DR (дравид.)	KT (картв.)	UR (урал.)	IE (и.-е.)	NC (сев.-кавк.)
№ языка	5	47	21	7	17	97	33
сред. стаб.	<b>0,67</b>	<b>0,65</b>	<b>0,56</b>	<b>0,54</b>	<b>0,48</b>	<b>0,45</b>	<b>0,43</b>

	KS (койсанск.)	Alt (алтайск.)	ST (сино-тиб.)	AN (австронез.)	AA (австроаз.)	AUS (австрал.)	AF (афразий.)
№ языка	26	54	132	94	99	36	69
сред. стаб.	<b>0,43</b>	<b>0,41</b>	<b>0,41</b>	<b>0,40</b>	<b>0,37</b>	<b>0,29</b>	<b>0,20</b>

Действительно, как отмечал Старостин, во многом «индекс стабильности зависит от времени, прошедшего с момента распада праязыка данной семьи» [С. Старостин 2007, 828]. Этот фактор хорошо виден в таблице, упорядоченной по убыванию индексов стабильности. Наиболее низкий индекс у афразийских языков — 0,20, что интерпретируется следующим образом: в афразийской семье в среднем максимально широко представленный корень в значениях списка Сводеша отмечен в 20% языков, то есть приблизительно в 14 из 69 языков. При этом, возможно, афразийская семья является наиболее древней из перечисленных. На этом фоне тем более интересен соответствующий индекс в атлантических языках (включая мель), где он равен 0,28<sup>11</sup>.

Для упорядочивания списка по рангам устойчивости С. А. Старостин прибегает к остроумной, но достаточно сложной процедуре: «1) Ранжируем все значения (от 1 до 110) по индексу стабильности в каждой из рассмотренных семей... 2) Из значений, занимающих позицию  $n$  (начиная с  $n = 1$  и кончая  $n = 11(0)$ ), будем выбирать такие, которые занимают эту или более высокую ( $n-1$ ) позицию как минимум в двух семьях. При этом при одинаковом ранге (позиции  $n$ ) „удельный вес“ выбранного значения будет увеличи-

<sup>10</sup> Данные получены по стословнику Сводеша — 10 дополнительных значений при подсчетах были исключены.

<sup>11</sup> В атлантическом списке значения «green» и «yellow» заменены на «walk» и «you pl.».

ваться в соответствии с числом семей, в котором он эту позицию занимает» [С. Старостин 2007, 833, 838]. Используя эти процедуры, Старостин получает обобщенные ранги стабильности для каждого значения списка в языках мира (табл. 9).

Таблица 9.

	1	2	3	4	5	6	7	8	9	10
1—10	we	two	I	eye	you sg.	who	fire	tongue	stone	name
11—20	hand	what	die	heart	drink	dog	louse	moon	nail	blood
21—30	one	tooth	new	dry	liver	eat	tail	this	hair	water
31—40	nose	no	mouth	full	ear	that	bird	bone	sun	smoke
41—50	stand	tree	ashes	give	rain	star	fish	neck	breast	leaf
51—60	come	kill	foot	sit	root	horn	fly	hear	skin	long
61—70	road	know	say	egg	seed	knee	black	head	sleep	burn
71—80	ground	feather	swim	white	bite	meat	fat	man	person	all
81—90	night	see	go	hot	red	cold	woman	round	lie	cloud
91—100	big	bark	sand	good	many	yellow	mountain	green	belly	little

Этот список, естественно, очень интересен для выявления порядков, но использованный Старостиным сложный алгоритм его получения не вполне удобен: во-первых, он не прозрачен, во-вторых, он не дает возможности количественного сравнения индексов. С. А. Старостин отказался от сопоставления средних индексов, учитывая «зависимость коэффициента стабильности от глубины семьи и числа сравниваемых языков» [С. Старостин 2007: 833]. Однако и алгоритм, использованный Старостиным, не вполне преодолевает указанные факторы. Вместе с тем средние (числовые) коэффициенты конечно же удобнее для сопоставления. Получим по материалам Старостина иерархию значений по средним коэффициентам (табл. 10)<sup>12</sup>.

Таблица 10.

	1	2	3	4	5	6	7	8	9	10	
1—10	I	you sg.	two	eye	we	tongue	name	one ++	dog +	nail +	0,71
11—20	die	water +	tooth +	louse	fire -	blood	hand	moon	nose ++	liver -	0,59
21—30	ear -	eat	heart -	sun +	star ++	stone --	new	what -	who --	smoke +	0,54
31—40	bone	full	hair -	give +	rain +	tail -	tree +	fish +	stand -	drink --	0,5
41—50	no -	leaf	foot +	this --	dry --	horn +	head ++	hear +	black ++	come +	0,44
51—60	sleep +	ground ++	egg +	knee +	long	skin	road +	kill	meat ++	sit	0,41
61—70	night ++	root -	mouth ---	breast --	red ++	white +	ashes --	know	bird ---	feather +	0,38
71—80	person	hot +	see +	bite	seed -	fat	go +	that ----	swim	cloud +	0,35
81—90	fly ---	neck ----	good +	man -	say --	woman	mountain +	burn --	lie	yellow +	0,31
91—100	bark	green	big	cold -	little	belly	sand	round -	all --	Many	0,25

<sup>12</sup> Плюсами и минусами обозначены все изменения десятков: так, два плюса для head означает, что его ранг поднялся на 2 десятка: у С. А. Старостина это значение было в седьмом десятке, а в исчислении по средним коэффициентам — в пятом; четыре минуса у neck обозначает, что из пятого десятка у Старостина оно перешло в девятый десяток.

Какой список «лучше»? Московская школа компаративистики использует в лексикостатистике первую половину списка, ранжированного по коэффициентам стабильности, полученным С. А. Старостиным. Изменение критерия привело к тому, что в первой половине списка мы «потеряли» 6 значений и соответственно 6 «приобрели». Вышли за пределы первой половины: *mouth, breast, ashes, bird, that, neck*. Они были заменены следующими значениями: *foot, horn, head, hear, black, come*. Чтобы оценить эти изменения, обратимся к интуиции компаративистов. По мнению Г. С. Старостина, «в первую очередь из-за технических обстоятельств» 9 значений из первой половины списка (*liver, neck, breast, this, that, full, give, stand, fish*) следует заменить на 9 значений из второй половины, а именно: *kill, foot, horn, hear, meat, egg, black, head, night* [Г. Старостин 2013: 141–143]. Оказывается, с интуицией у автора все в порядке. Пять из девяти предложенных Старостиным значений для включения в стабильную часть вошли в первую половину списка *foot, horn, hear, black, head*, значение *egg* вплотную приблизилось к нему, оказавшись в шестом десятке вместо седьмого, а значение *night* повысило свой ранг на два десятка. Как и хотелось автору, значения *neck, breast, that* вышли из первой половины, а значения *liver, this, stand* существенно понизили свой ранг (*this* теперь занимает 44-е место вместо 28-го). Эти результаты однозначно свидетельствуют о том, что критерий средних коэффициентов не только более удобен, но и лучше работает. Так, например, с его помощью легко оценить различия в стабильности первой и второй половин списка в разных семьях.

Исчислим коэффициент стабильности для каждой из двух половин списка (включив в список данные по атлантическим языкам — ATL, 41 язык). Отсортируем семьи по различиям коэффициентов первой и второй половин (третья строка) — табл. 11.

Таблица 11.

	ST	DR	AN	NC	UR	IE	KT	AA	TH	YEN	Alt	KS	AF	AUS	ATL
1. 1–50	0,56	0,71	0,55	0,58	0,62	0,59	0,65	0,47	0,75	0,75	0,48	0,50	0,26	0,35	0,32
2. 51–100	0,25	0,41	0,25	0,28	0,34	0,32	0,42	0,27	0,56	0,60	0,34	0,36	0,14	0,24	0,25
1–2	0,31	0,31	0,30	0,29	0,28	0,27	0,23	0,21	0,19	0,15	0,14	0,14	0,12	0,11	0,07

Прежде всего, все без исключения семьи показывают существенно более высокую стабильность первой половины по сравнению со второй. Даже в атлантических языках, занимающих по этому критерию последнее место, эта разница большая: в среднем 30% языков для основных корней в первой половине («главные» корни заполняют матрицу на треть) и 25% — во второй половине («главные» корни заполняют матрицу на четверть). В большинстве же семей эта разница исключительно велика — в 8 семьях она превышает 20%.

Интересно, что в самом конце списка оказываются все три африканские семьи плюс австралийская. Индексы стабильности значений, полученные в основном на языках Евразии, существенно отличаются за ее пределами. Это означает только, что накопление материалов по стабильности базисной лексики должно быть продолжено.

Приведем атлантический список, отсортированный по убыванию индексов стабильности (табл. 12).

Очевидно, что атлантический список существенно отличается от среднестатистического. Так, например, значение *pame* перешло из первого десятка в последний, значение *fly* — из девятого десятка в первый. Значение *head*, занимающее второй ранг в атлантических языках, ни в одной из семей не входит не только в первую десятку, но и в первую два-

Таблица 12.

	1	2	3	4	5	6	7	8	9	10
1—10	eye 0,83	head 0,76	ear 0,71	tongue 0,61	tooth 0,56	star 0,54	hair 0,51	tree 0,51	white 0,49	fly 0,46
11—20	knee 0,46	liver 0,44	one 0,44	rain 0,44	sun 0,44	little 0,41	water 0,41	foot 0,39	I 0,39	you sg. 0,39
21—30	die 0,37	bark 0,34	bird 0,34	horn 0,34	mouth 0,34	neck 0,34	bite 0,32	black 0,32	breast 0,32	dry 0,32
31—40	feather 0,32	new 0,32	person 0,32	two 0,32	eat 0,29	egg 0,29	go 0,29	swim 0,29	cold 0,27	louse 0,27
41—50	ground 0,27	man 0,27	road 0,27	root 0,27	stone 0,27	walk 0,27	what 0,27	you pl. 0,27	blood 0,24	good 0,24
51—60	leaf 0,24	lie 0,24	nose 0,24	see 0,24	skin 0,24	come 0,22	drink 0,22	fire 0,22	sit 0,22	who 0,22
61—70	belly 0,20	bone 0,20	dog 0,20	give 0,20	hand 0,20	hear 0,20	kill 0,20	meat 0,20	nail 0,20	night 0,20
71—80	red 0,20	sand 0,20	seed 0,20	smoke 0,20	tail 0,20	this 0,20	woman 0,20	all 0,17	fat 0,17	fish 0,17
81—90	full 0,17	hot 0,17	know 0,17	many 0,17	sleep 0,17	stand 0,17	we 0,17	burn 0,15	cloud 0,15	moon 0,15
91—100	mountain 0,15	say 0,15	ashes 0,12	big 0,12	heart 0,12	long 0,12	name 0,12	that 0,12	no 0,10	round 0,07

дцатку стабильных значений. Вместе с тем очень существенными отличиями характеризуются по признаку стабильности и другие семьи. Так, значение *name* оказывается в последней десятке и в тайских языках. Значение *drink* в койсанских и австронезийских языках входит в первую десятку, а в алтайских языках занимает последнее 100-е место. Хотелось бы прояснить самые общие вопросы: какие языки наиболее близки к среднестатистическому, а какие наиболее удалены от него и насколько? Одинаковы ли отклонения от среднестатистических индексов в каждой десятке стословника? Если мы делим список на две половины, сколько значений в каждой семье языков оказываются «не на своем месте»?

Эти ответы (по крайней мере, приближения к ним) нетрудно получить. Для того чтобы изложить количественный алгоритм сравнения, воспользуемся иллюстративной таблицей (табл. 13) с рассмотренным значением *name*.

- 1) Отметим для каждого из 100 значений номер десятки, в которую оно входит в семье языков. Значение *name* (строка 1 таблицы) входит в большинстве семей в первую десятку, в DR — в десятку 2, в AUS — в десятку 3, в AN — в десятку 5, в AA и Alt — в десятку 6, в TH и Atl (атлантические языки) — в последнюю десятку<sup>13</sup>.
- 2) Вычтем из полученных индексов (1—10) среднестатистический индекс (в случае с *name* — единицу) — строка 2. Таким образом, мы определяем, на сколько («десятков») индекс в семье отличается от среднестатистического.

<sup>13</sup> Точно разбить индексы на группы по 10 невозможно. Одинаковый индекс могут иметь, например, значения 8—12, и в таких многочисленных случаях отнесение двух первых по алфавиту значений к десятке 1, а вторых — к десятке 2 безосновательно. Пренебрежем этим фактором, чтобы избежать существенного усложнения алгоритма сравнения. Исключим лишь енисейские языки из сравнения — в семье только 5 языков, а значит, только 5 различных индексов, а в таком случае указанная погрешность слишком велика.



Таблица 13.

'name'		СРЕДН	KT	UR	IE	NC	KS	ST	AF	DR	AUS	AN	AA	Alt	TH	ATL	$\Sigma 1$	$\Sigma 2$
1	10	1	1	1	1	1	1	1	1	2	3	5	6	6	10	10		
2	10		0	0	0	0	0	0	0	1	2	4	5	5	9	9		35
3	10		0	0	0	0	0	0	0	1	1	1	1	1	1	1	7	
4	20	1	1	1	1	1	1	1	1	1	2	3	3	3	5	5		
5	20		0	0	0	0	0	0	0	0	1	2	2	2	4	4		15
6	20		0	0	0	0	0	0	0	0	1	1	1	1	1	1	6	
7	50	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2		
8	50		0	0	0	0	0	0	0	0	0	0	1	1	1	1		4
9	50		0	0	0	0	0	0	0	0	0	0	1	1	1	1	4	

3) Возможен менее детальный подсчет (строка 3) — вне зависимости от того, насколько отличается индекс от среднестатистического, будем приписывать любому отклонению значение 1, а отсутствию отклонений — значение 0. Например, в картвельских — ранг значение пале приблизительно совпадает со средним (оно оказывается в той же десятке), а в австралийских — отличается от него (оно оказывается в другой десятке).

Прделаем те же процедуры, разбив список не на десятки, а на двадцатки (строки 4—6 в таблице). В тайских и в атлантических языках значение пале находится в пятой — последней — двадцатке (строка 4); его ранг на 4 отличается от среднего (строка 5); его ранг отличается от среднего (строка 6).

И наконец, мы можем разбить список на две половины (строки 7—9). В уральских языках значение пале находится в первой половине списка (строка 7). Это не отличает его от среднестатистического списка (в данном случае строки 8 и 9 принципиально одинаковы).

В последних столбцах вычислены суммы по строкам.  $\Sigma 1$  принципиально меньше, чем  $\Sigma 2$ . Суммы в строке 6 и тем более 9 принципиально меньше, чем в строке 3.

Применение указанных алгоритмов позволяет сопоставлять семьи, значения и группы значений (по 10, по 20, по 50) по рангам стабильности. Ограничимся здесь суммарными данными по десяткам (аналог строки 3 в табл. 13), см. — табл. 14.

Вдумаемся в цифры, приведенные в таблице. Начнем с сумм по столбцам. В афразийских языках, показывающих наименьшее отклонение от среднего, 75 значений из 100 входят не в те десятки, которые вычислены для среднего, а в алтайских языках таких отклонений — 88! Подчеркну, речь идет не о том, что значения имеют другой ранг — они имеют существенно другой ранг. Все 10 алтайских значений, составляющих вторую десятку, не входят в нее в среднестатистическом списке. В первой десятке (наиболее стабильных значений) 8 значений в AA, TH, Atl имеют более низкий ранг в усредненном списке. При этом различие между семьями в целом небольшое — от четверти совпадающих по десяткам рангов значений в афразийском до 12 совпадающих рангов из 100 в алтайском. Интересно сравнить суммы по строкам. Мы видим существенно большие совпадения со средним в первой и последней десятках — то есть ранги, отмеченные в усредненном списке, лучше всего работают в качестве эталона для самых стабильных и для самых нестабильных значений.

Таблица 14.

	AF	ST	IE	NC	KS	UR	DR	KT	AA	AUS	AN	TH	Atl	Alt	Σ
1—10	5	4	4	5	5	4	4	4	8	6	6	8	8	7	<b>78</b>
11—20	7	9	8	7	7	7	8	8	8	9	8	8	8	10	<b>112</b>
21—30	9	8	9	8	9	10	9	8	6	8	8	9	10	9	<b>120</b>
31—40	7	9	10	9	8	10	8	9	9	10	9	7	10	8	<b>123</b>
41—50	9	8	8	9	7	9	7	9	9	7	10	9	10	10	<b>121</b>
51—60	6	7	8	10	10	8	9	9	8	8	9	10	8	9	<b>119</b>
61—70	8	9	7	8	8	9	9	9	9	9	9	10	8	10	<b>122</b>
71—80	8	9	9	8	9	9	10	9	9	10	9	8	9	9	<b>125</b>
81—90	9	7	8	10	9	8	8	9	9	9	9	10	8	9	<b>122</b>
91—100	7	6	6	5	7	6	9	7	6	5	7	6	8	7	<b>92</b>
Σ	<b>75</b>	<b>76</b>	<b>77</b>	<b>79</b>	<b>79</b>	<b>80</b>	<b>81</b>	<b>81</b>	<b>81</b>	<b>81</b>	<b>84</b>	<b>85</b>	<b>87</b>	<b>88</b>	

Таблица 15.

	ST	AN	KT	NC	UR	IE	AF	DR	KS	AA	TH	AUS	Alt	Atl	Σ
1—10	9	13	7	6	10	9	12	9	17	24	41	22	20	34	<b>233</b>
11—20	14	11	19	11	18	31	24	18	17	11	13	26	35	27	<b>275</b>
21—30	14	14	18	20	22	19	26	20	31	22	24	24	18	25	<b>297</b>
31—40	14	18	16	22	17	24	17	23	20	20	16	24	22	33	<b>286</b>
41—50	16	28	19	19	22	19	18	15	14	30	25	15	27	23	<b>290</b>
51—60	17	19	17	24	17	16	12	24	24	18	18	22	15	19	<b>262</b>
61—70	17	22	23	17	20	17	22	19	15	28	23	21	15	29	<b>288</b>
71—80	17	17	22	17	24	18	16	27	24	18	16	29	22	24	<b>291</b>
81—90	15	13	16	21	14	15	22	14	29	25	25	34	31	29	<b>303</b>
91—100	15	15	15	15	16	12	21	25	17	14	13	7	27	31	<b>243</b>
Σ	<b>148</b>	<b>170</b>	<b>172</b>	<b>172</b>	<b>180</b>	<b>180</b>	<b>190</b>	<b>194</b>	<b>208</b>	<b>210</b>	<b>214</b>	<b>224</b>	<b>232</b>	<b>274</b>	

Если мы применим другой критерий и будем учитывать не только сам факт отклонения, но и его величину (аналог строки 2 в иллюстративной таблице 13), существенно поменяется относительная степень близости различных семей к среднему (табл. 15).

По-прежнему наиболее сильно отличаются атлантические и алтайские языки, остальные же языки перегруппировались. В атлантических языках значения первой десятки не просто соответствуют второй десятке в среднестатистическом списке — они в среднем отличаются по рангу на три десятка каждое! Неизменным остаются более высокие суммарные индексы стабильности для первой и последней десятков.

Применим более толерантный критерий — будем фиксировать отличия только в том случае, если значение входит в другую двадцатку, и при этом не будем учитывать радикальность этих отличий (см. пример в строке 6 табл. 13) — табл. 16.

Порядок языков опять поменялся (атлантические языки по-прежнему наиболее сильно отличаются). Первая и последняя десятки по-прежнему дают наименьшие отклонения в рангах стабильности, причем следующими идут вторая и предпоследняя де-

Таблица 16.

	ST	NC	KT	AF	TH	KS	AA	DR	AN	IE	AUS	Alt	UR	Atl	Σ
1—10	3	1	3	3	8	4	5	3	4	4	4	5	3	5	55
11—20	5	5	6	5	3	5	4	7	5	8	7	10	6	7	83
21—30	7	6	5	9	7	8	5	7	6	9	7	8	9	7	100
31—40	7	7	8	6	6	6	8	7	9	9	10	5	9	10	107
41—50	5	5	7	7	8	4	8	6	9	7	7	9	9	8	99
51—60	7	8	8	5	9	9	7	9	7	6	8	7	7	6	103
61—70	5	8	8	7	6	6	8	8	6	6	6	5	8	7	94
71—80	8	6	5	8	6	8	6	7	8	8	9	9	9	9	106
81—90	5	8	5	6	6	7	8	4	6	5	7	7	6	7	87
91—100	4	3	3	5	3	5	4	6	4	3	2	4	5	6	57
Σ	56	57	58	61	62	62	63	64	64	65	67	69	71	72	

Таблица 17.

	TH	DR	KT	UR	IE	NC	KS	Alt	ST	AN	AA	AUS	AF	Atl	Σ
1—10	5	0	0	1	0	0	2	1	0	0	2	2	0	4	17
11—20	1	2	1	1	4	0	1	5	1	0	0	3	3	3	25
21—30	4	3	2	2	2	3	5	1	2	1	4	4	4	3	40
31—40	2	4	3	3	5	4	3	3	1	4	2	5	4	7	50
41—50	5	2	6	4	3	7	5	5	7	7	7	2	3	5	68
51—60	6	1	2	4	7	6	6	4	4	5	5	5	2	4	61
61—70	4	3	5	3	3	3	2	1	3	5	5	3	6	6	52
71—80	3	4	3	3	3	2	4	3	1	2	2	3	1	4	38
81—90	3	1	1	1	1	2	3	4	2	0	2	5	3	4	32
91—100	1	2	1	0	0	1	1	3	1	0	1	0	2	4	17
Σ	34	22	24	22	28	28	32	30	22	24	30	32	28	44	

сятки. Варьирование относительно усредненного списка по-прежнему исключительно высоко. Так, ни одно из 10 значений, составляющих в алтайских языках вторую десятку, не входит в первую двадцатку среднестатистического списка. В таблице мы не видим ни одного ноля и видим только одну единицу: в северокавказских языках 9 слов из 10 самых стабильных слов входит в первую двадцатку усредненного списка.

И в заключение приведем наиболее толерантную суммарную таблицу. В ней в качестве отклонений фиксируются только случаи вхождения значения не в ту половину. Данная таблица, по-видимому, наиболее интересна для оценки возможности использования только первой половины списка в лексикостатистике (табл. 17).

Отметим, что суммы по столбцам принципиально четные: если значение пропало из первой половины, то оно по определению заменилось другим. Это значит, что, например, в тайских языках, которые в этом списке наиболее близки к среднему, 34 значения из 100 оказались не в той половине, а поскольку количество отклонений в первой и второй половинах списка принципиально равно, можно заключить, что из 50 наиболее стабильных

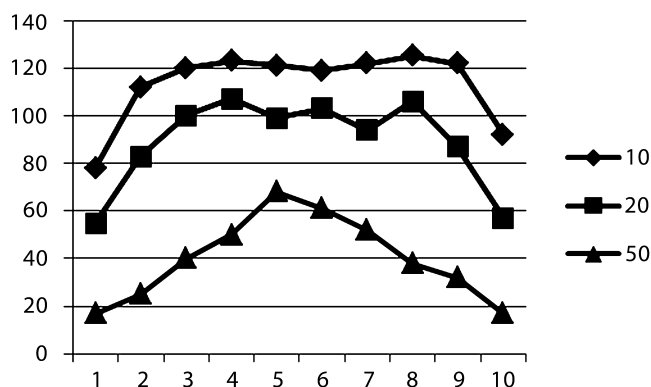


График 1.

слов 18 (приблизительно треть) не являются таковыми в усредненном списке. В атлантических же языках эти два списка не совпадают почти на половину. В таблице наконец появились нули. Например, все значения первой десятки в индоевропейских языках входят в первую половину усредненного списка, а последней десятки — во вторую.

Суммарные данные по строкам очень интересны. Представим на графике (график 1) три кривые: отклонения в пределах десятков, в пределах двадцаток и в пределах двух половин (без учета их радикальности).

График 1 показывает четкую закономерность. Наиболее универсальны начало и конец списка. Наиболее стабильные значения универсального списка являются наиболее стабильными в большинстве языковых семей. Наименее стабильные значения являются таковыми в большинстве семей. Средняя часть «универсального списка» наименее универсальна.

Подведем итог. При таком сильном варьировании индексов стабильности по семьям трудно рассчитывать на эффективное использование усредненных рангов значений для определения языкового родства. Утешает лишь тот факт, что даже в атлантических языках усредненный список показывает несколько большую степень стабильности в первой половине, чем во второй. Возможно, иерархию стабильности значений продуктивно использовать для определения места изолированных языков в семье или макросемье, например для сопоставления языка лимба с различными ветвями атлантических языков и языков мель, однако в этом случае полезнее, вероятно, использовать именно атлантический список, а не «универсальный».

Можно, впрочем, попробовать другие критерии стабильности. Один из альтернативных критериев я использовал для обработки атлантических стословников. Он элементарно прост. Можно, вслед за С. А. Старостиным, брать для каждого значения только максимально представленные в семье корни. А можно просто брать количество корней для каждого значения в семье языков. Умозрительно легко допустить, что эти два критерия дадут разные результаты. Представим 2 теоретические ситуации: (1) в 10 языках 5 корней по 2 языка в каждом — индекс устойчивости 2 по Старостину, индекс 5 — по другой методике; (2) один корень представлен в 6 языках, а в оставшихся 4 языках — 4 разных корня: индекс устойчивости 6 по Старостину (существенно выше) и 5 по сумме корней (тот же индекс). Поэтому неожиданным для меня оказалось то, что эти две разные техники определения рангов дают для атлантических языков абсолютно одинаковые суммарные результаты при исчислении по двадцаткам и по двум половинам, а при исчислении по десяткам оказывается, что вторая методика дает даже большее приближение к рассмотренной выше усредненной последовательности в языках мира, чем первая: если по методике Старостина мы вынуждены были констатировать, что 87 значений ока-

зываются не в тех десятках, что значения усредненного списка, то при ранжировании по количеству разных корней таких значений только 80, что является средним отклонением для языков 14 семей.

Давая те же результаты или даже улучшая их, сумма разных корней как критерий более прозрачна, поскольку учитывает все корни определенного значения, а не один из них (максимальный), и, в частности, учитывает многочисленные изолированные корни, представленные в атлантических языках, то есть корни, отмеченные только в одном из 40 языков. Исследование таких изолированных корней, как представляется, с полным основанием может быть выделено в отдельную область компаративистского исследования. Рассмотрим это положение подробнее.

### 3. Об изолированных корнях

Буквально в каждом из атлантических языков, в том числе и в тех, которые входят в достаточно «молодые» подгруппы (с совпадением по Сводешу более 50%), наблюдается исключительно высокий процент «уникальной» базисной лексики, которая не находит параллелей **ни в одном** другом атлантическом языке, в том числе и в языках своей подгруппы. Само по себе наличие изолированных слов тривиально для любого языка. Нетривиален высокий процент таких слов. В лимба уникальных корней в стословнике — 47, то есть практически половина списка! Казалось бы, если группа настолько гетерогенна, что само ее существование ставится под сомнение, то наличие исключительно больших пластов уникальной лексики вполне закономерно. Конечно, это так. Но эта ситуация вызывает вопросы, на которые нет вразумительных ответов. В частности, мы имеем все основания задуматься — откуда взялась эта масса уникальных базисных слов?

Первое естественное объяснение — заимствования. Но мы их предварительно исключили. В списках нет ни французских слов, ни арабских. Возможно, там есть еще не опознанные нами древние заимствования из манде, но их не может быть много.

Второе объяснение — лексические дериваты и семантические сдвиги (один из типичных примеров для атлантических языков: «калебаса» > «сердце»). Для большинства наших случаев эти объяснения не годятся — речь идет о непроемных базовых лексических корнях.

Может быть, в этих уникальных корнях отражена древняя нигеро-конголезская лексика, которая сохраняется только в одном из атлантических языков, а в других — исчезает? Отдельные такие случаи вполне вероятны, но предположение о том, что в отдельных атлантических языках, в стословниках Сводеша, отражается 600 нигеро-конголезских корней помимо корней, отраженных в нескольких языках, вряд ли можно рассматривать всерьез.

Тогда что это? Субстрат? Но шесть десятков субстратов — не много ли это для западноафриканского атлантического побережья? С субстратами тоже не получается. Что еще осталось у нас в запасе в качестве возможных объяснений? Может, это до-нигеро-конголезские корни, которые дошли до нас через многие тысячелетия, и в таком случае, выделяя такие изолированные корни, мы получаем ключ к самому древнему лингвистическому наследию человечества? Уже тот факт, что за неимением лучшего мы не можем отбросить эту провокационную космическую гипотезу, говорит о том, что изолированными корнями имеет смысл системно заниматься.

Стратегия компаративистики побуждает нас к поиску соответствий, сходств, сближений, параллелей. Если они не находятся, мы выносим уникальные корни за скобки, отправляя их «в отвал» — и, вероятно, напрасно. Такие корни имеет смысл выделять и

Таблица 18.

10–16	many	full	know	see	woman	sit	night	sand	stand					
9	eat	foot	horn	kill	little	moon	name	neck	root	round	say	sleep	star	water
8	come	feather	fire	nail	rain	red	road	white						
7	blood	dry	good	heart	knee	liver	long	one	seed	skin	stone	walk		
6	big	die	fat	fish	hair	hear	leaf	louse	mountain	swim	two			
5	all	belly	bird	cold	dog	fly	hand	hot	meat	mouth	tree	what	who	
4	ashes	black	breast	burn	cloud	drink	give	lie	nose	smoke	sun	tooth		
3	bite	bone	go	man	person	tail	that	you sg.						
2	egg	new												
1	ear	eye	ground	head	I	this	We	you pl.						
0	bark	no	tongue											

Таблица 19.

Limba	47	Joola-Kwaatay	22
Gola	34	Kisi	22
Bijogo	31	Buy	20
Nalu	29	Konyagi	19
Sua	29	Bayot	19
Wolof	28	Bedik	18
Balant	26	Sereer	17
Sherbro	25	Peul	14
Nyun	23	Biafada	12
Temne	23	Karon	12
Jaad	22	Baga-Maduri	11

публиковать — с тем, чтобы специалисты в смежных областях компаративистики могли провести их экспертизу.

К тому же количество изолированных корней в том или ином значении можно рассматривать как вполне независимый критерий определения ранга нестабильности внутри списка Сводеша. Ранжируем по этому критерию атлантический список — табл. 18 (цифры обозначают количество изолированных корней для каждого значения в 34 языках).

Приведем атлантические языки, в которых количество изолированных корней превышает 10 (табл. 19).

Отметим, что данные, приведенные в табл. 18, получены на представительном, но все же ограниченном списке, включающем 35 языков. Если же учесть все корни во всех атлантических языках и языках мель, количество изолированных корней еще более возрастает, причем возрастает существенно. Приведем один характерный пример — атлантические корни для слова со значением «женщина».

В нашей с Г. Сежерером базе собраны лексические данные по 64 языкам (атлантические и мель)<sup>14</sup>. В этих языках выделяется 28 различных корней со значением «женщина».

<sup>14</sup> Число не учитывает многочисленные диалекты и говоры, которые также представлены в нашей базе.

Естественно предположить, что сравнительно-этимологический анализ позволит существенно сократить это число, но два десятка корней в списке останется. Если мы вернемся к первому алгоритму установления сходства, который обсуждался в связи с порогом случайных совпадений (а именно: «разрешается сблизать любые согласные внутри локального ряда»), то теоретически при четырех локальных рядах у нас возможны 16 комбинаций структуры CVC. Добавим к ним 4 комбинации структуры CV. Итого получаем 20 теоретически возможных структур, к которым сводятся слова всех языков мира в рамках выделенных комбинаций. В атлантических корнях со значением «женщина» представлены 16 из 20 возможных структур, причем многие структуры представлены разными корнями<sup>15</sup>. Во-первых, это возвращает нас к утверждению, что в этом значении в атлантических языках мы легко можем найти звуковую параллель для английского, русского или практически любого другого языка. Во-вторых, и это для нас главное в данном параграфе, даже если мы не реконструируем корни для праатлантического уровня, это не избавляет нас от необходимости ответить на вопрос: откуда эти корни появились?

Приведем из базы слова со значением «женщина», сгруппировав их по корням в рамках указанных 20 звуковых комбинаций<sup>16</sup>.

- PP: бага koba *bum*, темне *bom*.  
 PT/ПК: балант *fag / far*.  
 PT: темне *-bèrà*.  
 PC: лимба *basa*.  
 PK: темне *boko / borko*.  
 PV: *sua mfo, nalu fay*.  
 TP: *palor 'be-leb / 'be-lep, ndut 'be-lep, laala leb / 'be-te, safen bi-tip, palor bi-tif* ('femme adulte'), *noon 'be-te' / bè-dop, peul debb-o / rew-be*.  
 TP: *konyagi -nəm, bapen -nam, bullom -noma, sherbro -nóma* (но *sherbro no-ma / à-ma* – возможно, корень \**ma*) .  
 TP / CP: *joola fogny naaba, bayot naw*.  
 TT: *kisi nord lààndò-, kisi sud láno, бага Fore laŋən, бага koba rani, бага sitemu ran, бага maduri / landuma raani*.  
 TT / (TP): *balant niin / ninn / nim, gola -ni(e)* — ср. бенуэ-конго \**nine / \*nene / \*nina*.  
 TT / CT: *joola* (разные языки) *naale / naare, joola banjal -njalena, joola fogny -nalen, biafada naali, bayot -rər*.  
 TK: *bainunk* (разные языки) *-dik-am, kasanga / kobiana lik-aam, bullom -lak-an, sei lag-an*.  
 TK / CK: *limba nani, joola ejamat nakan, joola kasa nakunta*.  
 TV: *bullom la, bedik lui-m / -nuí-m, limba nda, wolof ndaw* 'jeune fille' — ср. Jukunoid \**ndá*.  
 CP: *konyagi -səvælá, bedik sʷámبالàŋ / -sàmbar, jaad caafe, gola dzàve*.  
 CP: *bapen -suwon*.  
 CT: *gola dzân*.  
 CC: *manjak jije, joola ejamat ?änjä'ijä' / nejäu*.  
 CC: *nalu θay*.  
 CK: *wolof jigéen*.

<sup>15</sup> Так, комбинация «дентальный» — «V — лабиальный» отмечена в трех корнях, родство которых далеко не очевидно.

<sup>16</sup> Принятые условные обозначения: P — любой лабиальный, T — любой дентальный, C — палатальный, K — велярный. Например, обозначение PT будет обозначать присутствие начального лабиального и конечного дентального в структуре CVC.

- СК: basari *tóxár* / *sóxár*, bedik *sóyár*, konyagi *ryónkáx*, tanda *zogari*, bainunk-Kas *toxol*.  
 СК: joola fogny *seek*, joola ejamat *seh*.  
 CV: pepel *šo*.  
 КТ: bijogo *-kanto* — ср. Benue-Congo *\*-kwazi*.  
 КС: бага Mboteni *gìsì*.  
 КС: joola gusilay *koona*.  
 VT: joola (разные языки) *aal* / *aar*, karon *haal*, manjak / mankany *-aats* / *aaf*, manjak churo *aah*, pepel *arş* / *aş*.

Итак, в атлантических языках для значения «женщина» отсутствуют только корни структур ТС, КР и КК, но и в других языках эти структуры корней низкочастотны (подробнее об этом — см. [Pozdniakov & Segerer 2007]).

Вернемся в заключение к исходному вопросу: как быть, когда «нельзя», но «очень хочется»? Из всего вышесказанного вроде бы следует, что все равно «нельзя». На глубине ВЕ у нас нет, к сожалению, надежных инструментов, чтобы строить классификации на основе внешних сходств в базисной лексике, и единственный путь — классическая ступенчатая реконструкция. А пока ее нет, любые разумные критерии — морфологические изоглоссы, фонологическая статистика, данные внешнего сравнения, сопоставимость «слабой» лексики в неконтактирующих языках и многие другие — не являются ни «контролирующими», ни «сопутствующими», а имеют не меньший вес, чем сходства в стословниках.

Д. Р. Лецинер  
 Google Inc. (Маунтин-Вью, США)

### Заметки к статье К. И. Позднякова «О пороге родства...»

Мне хотелось бы обсудить некоторые выводы статьи, обобщения, сделанные на их основе, методы, приведшие к этим выводам, и границы применимости этих обобщений. Позволю себе кратко перечислить разделы статьи и те ее выводы, которые привлекли мое внимание.

Статья посвящена экспериментальному исследованию (на материале атлантических языков) следующих вопросов:

- существует ли некое минимальное количество сходств в списке Сводеша по двум языкам, которое позволило бы нам подозревать, что эти языки родственны?
- можно ли использовать индексы стабильности лексических значений для процедур определения языкового родства?
- каково происхождение аномально большого количества «уникальной» базисной лексики в отдельных атлантических языках?

Автор, в частности, приходит к следующим выводам:

- на глубине старше индоевропейской (ВЕ) у нас нет надежных инструментов, чтобы строить классификации на основе внешних сходств в базисной лексике, и единственный путь — классическая ступенчатая реконструкция;
- массовое непарное сравнение может гарантировать «доказательство» любого родства;
- при существующем сильном варьировании индексов стабильности (в версии С. А. Старостина) по семьям трудно рассчитывать на эффективное использование усредненных рангов значений для определения языкового родства;
- изолированными корнями имеет смысл системно заниматься. Такие корни имеет смысл выделять и публиковать — с тем, чтобы специалисты в смежных областях компаративистики могли провести их экспертизу.



Представлен подробный анализ варьирования индексов стабильности. Предложена и модификация методики Старостина для расчета индексов стабильности: ранжирование по количеству разных корней для каждого значения в семье языков. Эта модификация представляется весьма интересной. С выводом о нежелательности игнорирования изолированных корней трудно не согласиться. Тот факт, что на вопрос об их природе не находится очевидного ответа, делает эту проблематику еще более интригующей.

Значительно труднее, однако, согласиться с логикой автора в тех пунктах, где он делает выводы о применимости тех или иных эмпирических данных к вопросу определения языкового родства. Сразу должен сказать, впрочем, что основная часть раздела 1, посвященная оценкам вероятности случайного фонетического сходства слов, не вызывает никаких возражений. Напротив — проведен убедительный, аккуратный и впечатляющий анализ, получаемые оценки следует принять<sup>1</sup>. Однако, на мой взгляд, из них никоим образом не вытекают окончательные выводы, приведенные выше. Также и вывод о неприменимости индексов стабильности представляется несколько поспешным и недостаточно обоснованным. О индексах стабильности будет ниже сказано несколько слов. Но главным образом я хотел бы обсудить логику, связанную с применимостью процедур массового и попарного сравнения.

Напомню, что исходный вопрос, поставленный автором, следующий: «**Существует ли некое минимальное количество сходств в списке Сводеша по двум языкам, которое позволило бы нам подозревать, что эти языки родственны?**» Однако вопрос этот, в той форме, как он поставлен в статье, имеет значение в первую очередь для случая сравнения **пары изолированных** далеких языков (скажем, шумерского и сонгаи). Именно в этом случае количество сходств при попарном сравнении списков является единственным доступным критерием для принятия решения. И действительно, в этом случае — вряд ли кто против это-

<sup>1</sup> Одним исключением, впрочем, является трактовка локальных рядов согласных при наличии структур VC и CV — нулевой согласный не трактуется автором как самостоятельный ряд. В статье не предлагается никаких критериев сравнения локальных рядов с нулевым рядом. Если допустить его сравнение с любым локальным рядом (такую возможность автор далее упоминает), то уровень случайности действительно зашкалит. Если же не разрешать сравнение с нулевым согласным, тогда проводимые вычисления вероятностей нуждаются в корректировке в сторону понижения.

го возразит — метод поиска фонетических сходств в стословном списке не является надежным инструментом установления даже и гипотетического родства. Однако — именно постольку, поскольку на этот счет среди лингвистов имеется общее согласие — никто из них и не применяет метод в этом варианте (я не имею тут в виду случаи очевидно близкого родства — когда «минимальное количество» действительно «существует»). На практике, область его применения другая — а именно, когда сравниваются:

- две установленные языковые семьи друг с другом;
- несколько установленных языковых семей друг с другом;
- или в крайнем случае (достаточно редко) — один изолированный язык с установленной языковой семьей.

В случае же сравнения языка (или группы языков) с **языковой семьей** — критерий сходства должен быть другим. А именно, допустимыми сближениями должны считаться только сближения с **праязыковой** лексикой семьи (в случае же сравнения двух или более семей — только сближения праязыковой лексики с праязыковой лексикой) — «исконные» сближения. Те сближения, для которых может быть показано, что это не так, должны отбрасываться. Те, для которых их «исконный» (праязыковой) характер сомнителен, должны и учитываться, как сомнительные.

Приведем табл. 7 из статьи (для удобства сопоставления таблиц я переупорядочил их строки в порядке английского списка слов), см. стр. 212. Я выделил серым фоном те значения, которые повторяются хотя бы в паре списков из приведенных шести. Обращает на себя внимание следующее:

- в шести (из 41) «самых похожих» списках сближений 5 значений (black, breast, me, white, you sg.) встречаются чаще чем дважды. Хочу заметить, что я не дал себе труда погрузиться в сравнительное изучение атлантических языков (и никогда ранее ими также не занимался), поэтому любое мое впечатление о сравнении атлантической лексики свободно может оказаться ошибочным. Тем не менее, хотя сближения для black, breast, me, white, you sg. и производят впечатление исконных атлантических корней, я не уверен, относятся ли **banj** для black, **n-bin** для breast к тому же исконному корню, что и прочие формы (если же нет, то их следует исключить). Остается следующее число сближений: black — 3, breast — 4, me — 3, white — 3, you sg. — 3. Итого 16;

Таблица 7а.

Temne	
ashes	r-os
bark	foki
bone	ka-bant
fly	falir <sup>w</sup>
go	kq
little	lɔl
me	mi
one	in
sand	a-sənt
seed	ma-sənt, ma-ther
woman	bom

Nalu	
black	bal
big	wɔk? -baŋ?
bone	wol?
breast	n-bin?
drink	dig
fire	mɛt?
go	kɔ
good	kuun?
star	ma-saar
tree	n-ti?
we	biye?
you sg.	yi

Biafada	
black	baŋ?
breast	bəl?
fire	fuuru
fish	yesä / bwa-?
fly	fiili
ground	gal?
little	tiiti?
me	mma
mouth	mməsə
nose	nasin
rain	səna?
say	cii, re
white	paas
you sg.	yi

Limba	
ashes	hɔtɔ? ~ fɔtɔ?
belly	fɔle
black	boloi
breast	hu-bili
drink	tie?
eye	hoya ~ foya?
fish	fee
kill	kɔra?
tooth	hu-titi
you sg.	yi

Bedik	
black	-bála
breast	ɛ-bɔr
burn	o-bód
eye(s)	gi-nguùs?
hand	ɛ-kònd
hear	o-hér
knee	é-dɔl-kɔlɔ?
liver	i-dámbél?
me	-mé- / mé-
meat	ɲás?
mouth	bɔ-más
white	-fɛsà / -pɛsà

Mankany	
big	-mag
breast	p-wil?
hair	ka-weel?
leaf	ka-toh?
liver	u-deeb
meat	u-ye-maɕ
nail	ka-ñiiri
nose	b-yiis?
what	we
white	p-faaθ

- из остальных 13 «парных» сближений (а именно: ashes, big, bone, drink, eye, fire, fish, fly, little, liver, meat, mouth, nose) я бы (с той же степенью достоверности) усомнился в по крайней мере в 5 (ashes, bone, eye, fire, little). Подозреваю, что статистически моя оценка близка к правде. Хочу заметить, что отсутствие внутриатлантического родства в данном случае не говорит о необходимости отвергнуть данное сходство с английским целиком — оно только говорит о необходимости отвергнуть **по крайней мере одно из двух** этих сближений. Ито-

го, верхняя оценка числа возможных сближений:  $(13 - 5) * 2 + 5 = 21$ ;

- что же касается остальных 25 «одиночных» сближений — у меня нет материала, чтобы судить об исконности (праатлантической древности) этих корней (однако для большинства из них он, вероятно, есть у автора статьи). Из «общих» соображений я бы не думал, что процент исконных среди них выше, чем достоверно исконных парных сближений. Если это так, получаем верхнюю оценку числа исконных одиночных сближений:  $25 * (13 - 5) / 13 \approx 15$ ;

- итого имеем верхнюю оценку суммарного числа «возможных» исконных сближений на 6 списков:  $16 + 21 + 15 = 52$ . В среднем же — 8,5% (но это **на шести самых лучших списках** — а не в среднем по семье, как подсчитано в статье). По методике же статьи, в этих списках находится 69 сближений, в среднем — 11,5%. Если мы примем тот же коэффициент завышения и по остальным 35 спискам, то цифра «в среднем 8,5% случайных сходств с английским» превратится в  $8,5\% * (8,5\% / 11,5\%) = 6,3\%$ .

Казалось бы, 6,3% все еще вполне себе отличная от нуля цифра — хотя уже как-то легче поверить в ее случайность... И действительно, сам автор и приводит один из решающих аргументов против предположения о родстве атлантических языков с английским (в примечании 8): «Отметим, что сравнение атлантических слов с английским системно дает существенно больше сходств, чем с французским или с русским. Мне не представить, с чем это могло бы быть связано». Я тоже затрудняюсь сказать, с чем это связано (вероятнее всего, случайность; если бы мы имели дело с действительным родством, это теоретически могло бы объясняться большим фонетическим или лексическим консерватизмом английского по сравнению с русским и французским — впрочем, мы сейчас даже не говорим о том, что по отношению к конкретной паре английский—русский такое утверждение само по себе вряд ли справедливо), однако очевидно, тем не менее, что наличие подобной ситуации с несомненностью **доказывает случайность** сходства атлантических слов с английским (или, как максимум — наличие «атлантического субстрата» в английском или германских языках; впрочем, насколько я понимаю, о наличии такого субстрата науке на сегодня ничего не известно). Если бы сходжения объяснялись исконным родством, эта близость была бы **воспроизводима** на большинстве индоевропейских языков (а их, как известно, немало, и достаточно разнообразных — а тем самым есть материал для накопления статистики).

В действительности следует усреднить процент сходств не только по подгруппам атлантических языков (именно по подгруппам, а не по отдельным языкам: чтобы не давать неоправданно большого веса ни одной большой совокупности похожих — близкородственных — языков), но и по подгруппам индоевропейской семьи. И далее, если подсчитанный таким образом средний процент будет **статистически достоверно превышать** средний процент при сравнении с **другими** семьями языков (т. е. равновероятными конкурирующими ги-

потезами) — тогда можно говорить, что есть подозрение на наличие глубокого родства (или же древних контактов). Это — а не заранее фиксированный «единый» числовой порог — и будет правильным<sup>1</sup> тестом на родство. При отсутствии же достоверного превышения из подсчетов никакого вывода сделать нельзя. Тогда априорная «равновероятность» гипотез превращается в такую же апостериорную «равновероятность». Однако хочу подчеркнуть, что апостериорная «равновероятность» не является свойством множества гипотез, как такового: она является свойством множества гипотез **и процедуры**, примененной для его анализа.

Фактически аргументы автора статьи сводятся к тезису о большой зашумленности (и тем самым статистической недостоверности) процедуры **парного** сравнения языков на основании фонетического сходства 100-словных списков. С этими аргументами, применительно к обсуждаемой в статье процедуре, следует согласиться<sup>2</sup>. Увы, обсуждаемая процедура плоха не только этим, но и тем, что она отличается от лингвистически оправданной процедуры (см. выше). Но не позволит ли лингвистическая коррекция процедуры исправить и ее статистические недостатки (в том числе и в варианте т. н. массового сравнения)? Рассмотрим этот вопрос.

В этой связи я бы попробовал также уточнить утверждение: «у самого Гринберга не было никакой формализованной методики сравнения». Формально говоря, с этим нельзя не согласиться: формализованной методики сравнения у Гринберга не было. Однако я думаю, что некая **неформализованная** методика у него была — не сводящаяся «лишь» к его энциклопедическим знаниям и исключительному исследовательскому таланту (которые действительно формализовать отнюдь не просто), но в том числе и к более приземленным вещам, которые, вообще говоря, вполне могут поддаваться формализации. Более конкретно, я

<sup>1</sup> Не «единственно правильным» — могут, вообще говоря, существовать и другие разумные тесты (см., в частности, ниже) — а просто «статистически правильным».

<sup>2</sup> Я в данном случае соглашаюсь не со всеми конкретными обобщениями автора относительно применимости оценок зашумленности сравнения списков к решению вопроса о надежности количественных критериев родства (и в частности, с выводом о полной неработоспособности указанной процедуры «на глубине старше индоевропейской» — должен заметить, что, например, для ностратических языков подобная процедура, при усреднении, работает очень устойчиво), и не с конкретным выбором процедур фонетического сравнения, но лишь с общими оценками вероятности фонетических сходжений (с оговоркой, упомянутой в сн. 1 на стр. 211).

думаю, что именно соображения, подобные изложенным выше, находились в ее основе (и они, собственно говоря, и представляют собой «метод массового сравнения» *per se*). Рассмотрим их суть несколько подробнее.

Пусть нам дано некоторое количество «гипотетических» языковых семей. В отличие от предыдущего рассмотрения, эти семьи не являются «установленными». В данном случае важно не то, что мы, может быть, не полностью уверены в реальности этих семей — пока что можно ограничиться и гипотезой — а то, что у нас нет достоверной информации о том, какие корни в каждой из этих семей являются или не являются исконными. Поэтому, для того чтобы проводить осмысленную процедуру сравнения семей, нам придется заменять эту «внутрисемейную» информацию статистическими гипотезами, основанными на фонетическом сходстве — так же, как и при сравнении семей между собой.

В статье приведены достаточно простые и разумные способы оценки вероятности случайного фонетического сходства — на самом деле не только слов 100-словного списка в **среднем**, но и каждой **данной** пары «сходных» слов. Понятно, что эти числа (как и указано в статье) зависят, в частности, от частотности фонем в данных языках — а также от их позиционного распределения и сочетаемости. Например, вероятность случайного совпадения слов, в которых обнаруживается редкая фонема или сочетание фонем, будет крайне мала, а вероятность случайного совпадения слов типичной структуры — куда больше. (Заметим, что того же нельзя сказать про вероятность совпадения в результате независимого заимствования.) Для нас важно то, что эти числа достаточно несложно получить и воспользоваться ими. И что далее важно: если у нас имеются, скажем, **три** фонетически сходных слова с одинаковым значением в трех языках — что мы можем сказать о вероятности каждого из них быть «исконным» словом в своем языке? (Мы предполагаем, что все три языка родственны друг другу «на одном уровне» — т. е. являются продуктом одновременного распада праязыка на три ветви.) Здесь на самом деле два вопроса: 1) какова вероятность, что по крайней мере два слова из трех восходят к общему корню? 2) если (1) верно — какова вероятность того, что **данное** слово к нему восходит? Я бы затруднился дать точное определение этих величин (не так просто сформулировать полный набор необходимых предположений), но следующие простые оценки будут, видимо, достаточно разумны:

- вероятность (1) =  $1 - (1 - p_{12})(1 - p_{13})(1 - p_{23})$
- вероятность (2) =  $1 - (1 - p_{12})(1 - p_{13})$   
[это для слова в языке 1]

Здесь  $p_{12}$  и т. д. — **апостериорная** вероятность родства слов в языках 1 и 2. Она зависит от **априорной** вероятности родства и от вероятности случайного сходства (включая сюда и вероятность общего заимствования!) по следующей формуле:

- $p_{12} = p_{apr} / (p_{apr} + p_{rand}(1 - p_{apr}))$

Здесь мы приняли, что в случае родства вероятность сходства = 1, что, конечно же, неверно. Если же оценить ее как  $p_0$ , тогда будет:

- $p_{12} = p_{apr} * p_0 / (p_{apr} * p_0 + p_{rand}(1 - p_{apr}))$ .

Обратим внимание: если  $p_{apr} * p_0 \ll p_{rand}$ , то  $p_{12} \approx p_{apr} * p_0 / p_{rand}$  — и это именно та ситуация, которую мы имеем при достаточно дальнем сравнении. (Если же мы заранее отвергаем всякую возможность родства — т. е. если  $p_{apr} = 0$ , то и никакое количество сколь угодно убедительных данных нас все равно не убедит; тогда и  $p_{12} = 0$ .)

Отсюда видно, во-первых, что от оценки  $p_{apr}$  зависит очень многое. (И откуда взять ее?) Во-вторых, видно, что если у нас имеется не три, а **много** сходных корней, тогда даже и малые (но не нулевые, конечно!) значения  $p_{apr}$  нам не так уже страшны.

А именно, пусть у нас имеется  $k$  языков, а вероятность попарного родства для всех пар одинакова (в общем случае это неверно —  $p_{rand}$  зависит от конкретных форм слов) и равна  $\epsilon$ . Тогда вероятность (1) — обобщая формулу на  $k$  языков — равна  $1 - (1 - \epsilon)^{k(k-1)/2}$ . Если  $\epsilon$  очень мало, это будет приблизительно  $\epsilon k(k-1)/2$  — то есть, взяв приблизительно  $\epsilon^{-1/2}$  языков, мы уже в неплохой форме. Если, скажем,  $\epsilon = 1\%$ , десяти языков нам хватит. (Для вероятности (2), к сожалению, значения будут меньше — порядка  $\epsilon k \dots$ ) Это и есть ключ к правильному массовому сравнению — когда примеры по отдельным языкам используются не скопом и без разбора, а проверяя и контролируя друг друга — и только достоверные примеры используются для дальнейшего сравнения.

Но откуда все-таки взять  $p_{apr}$  (и тем самым  $\epsilon$ )? Очень просто. Не забудем, что мы говорили сейчас об установлении исконного родства **внутри** «готовой» гипотетической семьи. Внутри этой семьи мы уже имеем примерные оценки степени родства ее языков — т. е. среднего количества совпадений в 100-словном списке. И кстати, обычно эти оценки не так уж малы — скорее порядка 10% или выше, и уж никак не 1%. Эта оценка, в первом приближении, и есть  $p_{apr}$ . Эту оценку можно уточнить в зави-

симости от индекса стабильности значений слов — каким бы методом это не делать это, для высокостабильных корней различие будет очень заметным. Далее, в зависимости от значений  $p_{\text{rand}}$  получаем  $\varepsilon$  — которое, вообще говоря, будет заметно выше  $p_{\text{app}}$  при низких значениях  $p_{\text{rand}}$ .

В действительности все эти оценки — пока лишь подготовка к главному: к сравнению **между** семьями. Вот тут-то и должна заработать процедура массового сравнения. Необходимые ингредиенты — **высокие уровни уверенности в исконности определенной лексики** внутри отдельных семей (см. расчеты выше) и **наличие большого числа семей для сравнения**, чтобы на основании тех же рассуждений, что и выше, большого числа фонетически сходных форм в разных семьях и высокого уровня уверенности в их исконности прийти к вы-

воду о вероятном исконном родстве этих форм на макроуровне: уровне родства **между** рассматриваемыми семьями. При попарных сравнениях, по очевидным причинам, такой вывод будет заметно затруднен.

На самом деле, как нетрудно заметить, приведенные выше процедуры являются, в определенном смысле, процедурами «ступенчатой реконструкции». Речь не идет, разумеется, о реконструкции исторической фонетики и установлении праформ — объектом «реконструкции» выступает в данном случае гипотетический список форм-потомков в языках-потомках. Это существенно меньше, чем достоверно установленные праформы и исторические соотношения, но с точки зрения статистики почти столь же хорошо для лексикостатистического анализа с целью выявления дальнего родства.

Г. С. Старостин

Российский государственный гуманитарный университет / РАНХиГС (Москва)

### Ответ на статью К. И. Позднякова «О пороге родства...»

Статья Константина Позднякова (далее — КП) вновь поднимает ряд вопросов, чрезвычайно актуальных для современного этапа развития компаративистики, и, поскольку автор во многом отталкивается от недавно опубликованной мной монографии («Языки Африки. Опыт лексикостатистической классификации»), большая часть которой посвящена как раз этим вопросам, я с удовольствием использую эту возможность для того, чтобы уточнить отдельные моменты собственной концепции, а заодно и снять некоторые критические возражения автора или предложить возможные пути решения обозначенных им проблем.

В первую очередь, я бы охарактеризовал статью КП как набор любопытных и во многом полезных размышлений, формально связанных друг с другом, но не имеющих единой общей цели (по крайней мере, очевидной). С одной стороны, автор ставит в начале текста вполне конкретный вопрос — «можно ли классифицировать языки на уровне макросемей, не используя ступенчатую реконструкцию на базе сравнительно-исторического метода?» — и в конце текста получает на него столь же конкретный ответ (нельзя). С другой стороны, остается не вполне яс-

ным, к кому обращен этот вопрос. Активные «сплиттеры» (убежденные противники любых гипотез о дальнем языковом родстве) и без статистических выкладок КП уже давно пришли к тем же самым выводам. «Ламперы»-гринбергианцы, оперирующие псевдометодом массового сравнения, принципиально работают не со списком Сводеша (который, очевидным образом, не удовлетворяет критерию «массовости» сравнения), а с полным корпусом лексического и грамматического материала, и для полемики с ними требуется совсем другая статистика. С таким же полным корпусом предпочитают работать и этимологи-традиционалисты, отказывающиеся замечать какие-либо фундаментальные различия между «обычным» и «макросравнением» и относящиеся к лексикостатистике предельно скептически (такие, как, например, покойный А. Б. Долгопольский), так что их из потенциальных адресатов статьи тоже, наверное, можно выкинуть.

Естественно было бы предположить (и какое-то время мне именно так и казалось), что статья обращена в первую очередь на критику некоторых теоретических положений, вынесенных мной во вводную (методологическую) часть вышеупомянутой

монографии, где предлагается некоторая, сама по себе довольно сложная, процедура установления относительного языкового родства для макросемей на «сверхглубоких» уровнях, не предполагающая при этом полной реконструкции праязыков ветвей-потомков. Эта процедура на определенных этапах действительно предполагает сравнение поверхностной фонетической структуры языковых форм, которое может осуществляться в виде автоматического алгоритма, т. е. претендует на некоторую степень объективности, не попадая напрямую в зависимость от интуиции исследователя-компаративиста.

Однако между тем, как предлагается осуществлять сравнение в монографии «Языки Африки», и той процедурой сравнения, которую тестирует на материале атлантических языков в своей статье КП, есть и немало существенных отличий — таких, которые, скорее всего, фундаментально повлияли бы на результаты, полученные автором. Кратко суммирую здесь лишь некоторые из них.

(1) По утверждению КП, «следует сжиться с фактом, что десяток сходств при реально существующих алгоритмах их выделения — это всего лишь норма случайных созвучий между любыми языками». Но при каких алгоритмах? КП иллюстрирует свое утверждение на примере сопоставления слов английского языка и атлантических языков темне (11 сходств), биафада (12 сходств, из них 4 под вопросом), бедик (12 сходств, из них 4 под вопросом), налу (14 сходств, из них 9 под вопросом), лимба (10 сходств, из них 4 под вопросом), манкань (10 сходств, из них 4 под вопросом). При этом, однако:

- автор использует *чрезвычайно* либеральный алгоритм сопоставления; семь его основных принципов, описанных выше в тексте статьи, которые, помимо прочего, допускают объединение в рамках одного класса чересчур большое число согласных, а также игнорирование «слабых» согласных в середине слова. По сравнению с этим алгоритмом применяемый в нашей монографии подход не позволил бы считать «сходством» такие «англо-темне изоглоссы», как *bark : foki*, *little : lol*, *seed : ma-sənt*, и даже *woman : bom*;
- автор бывает непоследователен в вопросах отделения от корня как «застывших», так и вполне продуктивных префиксов. Например, темне *ka-bant* ‘кость’ сравнивается с английским *bone* с отделением префикса, в то время как биафада *nasin* ‘нос’ сравнивается с английским *nose* несмотря на то, что *na-* в этой форме, судя по источникам, — точно такой же

именной префикс. Такого рода непоследовательности, разумеется, ведут к дальнейшему увеличению случайностей;

- наблюдаются и отдельные, возможно случайные, семантические неточности: например, английское *hand* сравнивается с бедик *ε-kənd*, хотя в словаре М.-П. Ферри однозначно указывается, что *ε-kənd* — это ‘bras, de l'épaule au coude’, в то время как *hand (main)* на бедик будет *ε-ɲēt* и с *hand* несопоставимо.

Таким образом, утверждение о *норме* в десять случайных созвучий справедливо далеко не для всех «реально существующих алгоритмов», а лишь для таких, которые основаны на чрезвычайно дозволительных и размытых стандартах поверхностного сравнения. Если же применять алгоритм, аналогичный описанному в «Языках Африки», то, например, число поверхностных сходств между бедик и английским с двенадцати по версии КП падает до четырех, и т. д. Так что, вопреки утверждению КП о том, что «выделение десятка пересекающихся классов не улучшает дела», количество выделяемых классов на самом деле влияет на результат даже при принятии универсального стандарта — и, на мой взгляд, система, изначально предложенная А. Б. Долгопольским, вполне подходит на роль такого стандарта, соблюдая разумный и типологически обоснованный баланс между «чрезмерной дозволительностью» и «чрезмерной строгостью» сравнения.

(2) «Возникает простой методологический вопрос: как быть, когда, например, класс R, включающий только вибранты, не представлен в одном из языков? В гола отсутствует фонема /r/. Это означает, что /r/ волоф вообще ничему не может соответствовать в гола». Вообще говоря, в рамках нашей системы классов в класс R попадает не только вибрант /r/, но и плавный /l/ (что основано на типологически частотном в языках мира колебании между этими двумя фонемами), который, насколько мне известно, в гола наблюдается, так что данный пример не очень удачен.

Допустим, впрочем, что ситуация сложнее, и в сравниваемом языке /r/ изменился не в /l/, а в /d/, т. е. соотв. фонема действительно перешла в другой класс. В этом случае в рамках принятой в нашем исследовании методологии в дело вступает т. н. *метод ручной коррекции* — переоценка результатов подсчетов «поверхностных сходств», основанных на универсальной системе критериев, с учетом «локальной специфики», т. е. легко обнаруживаемых и описываемых особенностей фонологической системы языка, не «отлавливаемых»

универсальным алгоритмом. Грубо говоря, если фонему /r/ языка А «не с чем сравнивать» в языке Б, «ручная коррекция» находит для нее в языке Б ближайший эквивалент, основываясь на данных типологии фонетических изменений (где «ближайшим соседом» класса R оказывается фонема *d* в классе T). При желании «ручную коррекцию» можно формализовать и более жестким способом (например, с помощью статистических подсчетов частотностей фонем или через привязку к т. н. дистанциям Левенштейна), но в целом все эти проблемы носят технический характер (и уж во всяком случае, представляются тривиальными по сравнению, например, с гораздо более фундаментальной проблемой распределения по консонантным классам такой «экзотики», как щелчковые фонемы).

(3) Разумеется, и при более строгих алгоритмах поверхностного сравнения не только не исключены, но и засвидетельствованы отдельные (редкие) случаи, когда число случайных сходств между заведомо неродственными или чрезвычайно дальнородственными языками превышает ожидания и оказывается сопоставимым с числом неслучайных сходств для других пар языков (хороший пример: 10–12 случаев совпадений между английским и омотским, описанный в недавней работе А. Касьяном). Если на это наложить еще и актуальный для некоторых ситуаций ареальный фактор, то, конечно, само по себе поверхностное сходство двух языков, потенциально родственных на уровне очень старой семьи или макросемьи, серьезным аргументом признать нельзя.

Однако ни в работах С. А. Старостина, оперирующих методами «поверхностного» сопоставления языковых форм, ни в моей монографии, собственно говоря, и не предлагается выдвигать даже предварительные гипотезы о языковом родстве на основании прямолинейного сравнения современных языков, связанных потенциальным родством на «макросемейном» уровне. Вообще, очень важно, чтобы любая ситуация родства, так сказать, осознавалась максимально *динамично* — не на основании отдельно взятого анализа сходств между двумя статическими точками, а в рамках некоторого хронологического пространства, где это родство можно проанализировать в том числе и на предмет естественности его эволюции.

Так, в частности, как уже неоднократно в своих работах отмечал С. А. Старостин, если число совпадений между языками при перемещении сравнения с более современного на более древний уровень остается неизменным, совпадения эти, скорее всего, случайны; если же родство настоящее, число

это должно возрастать (т. е. между древнеанглийским и древнерусским должно быть больше сходжений, чем между их современными потомками, а между прагерманским и праславянским оно должно быть еще больше). Любопытно в этом смысле посмотреть на то, как будет вести себя «английско-темне» сопоставление при выходе, например, на уровень «прагерманско-мель» — подозреваю, что общее число сходжений не только не увеличится, но и, хуже того, на место одних сходжений хотя бы частично встанут другие (поскольку прагерманские эквиваленты современных английских слов будут существенно отличаться по своей консонантной структуре).

В монографии «Языки Африки» неоднократно подчеркивается неправомерность «поверхностного сопоставления» по методу консонантных классов для языков, связанных потенциальным родством, более глубоким, чем уровень группы, т. е. в случае атлантических языков такому сравнению должна предшествовать «очистка» материала — выявление наиболее перспективных претендентов на выражение того или иного сводешевского значения на уровне отдельной группы или ветви. Если такая «очистка» приведет к тому, что число сходжений между условным «пра-мель» и прагерманским уменьшится, а возрастет, причем так, что «английско-темне» параллели не будут *заменены* новыми параллелями, а, наоборот, будут ими *дополнены* — значит, языки мель и германские языки, скорее всего, родственны. Однако с большой вероятностью этого не произойдет.

Теперь несколько слов относительно проблематичности критерия динамической градации, который КП подробно рассматривает на примере материалов статьи С. А. Старостина, подключая к ним собственные данные по атлантическим языкам. Хотя сам КП оценивает полученные выводы пессимистично («при таком сильном варьировании индексов стабильности по семьям трудно рассчитывать на эффективное использование усредненных рангов значений для определения языкового родства»), мне, наоборот, кажется, что выводы только подтверждают значимость метода динамической градации.

Дело в том, что выведенный автором график показывает, что *начальная* и *конечная* секции ранжированного 100-словного списка действительно стремятся к универсальной значимости, а это как раз то, что нам и нужно. По существу, на «среднюю» часть списка в такой ситуации можно вообще не обращать внимания — достаточно посмотреть, сколько сходжений, отловленных на этапе

поверхностного сопоставления, относятся, скажем, к первым 10—15 элементам ранжированного списка, а сколько — к последним 10—15. Грубо говоря, если в ваших сравнительных материалах совпадают слова типа ‘маленький’, ‘круглый’ или ‘гора’, но отсутствуют при этом сходжения в личных местоимениях, числительных 1—2, или таких частях тела, как ‘глаз’, ‘ухо’, ‘зуб’ или ‘рука’, или, если они все же присутствуют, но при этом не обнаруживается явного превалирования сходств второго типа над сходствами первого типа, то ситуация должна расцениваться как подозрительная — своего рода «красная лампочка», которая указывает на то, что не все в порядке и следует обстоятельно протестировать идею возможных контактных связей между языками, или нулевую гипотезу (случайные созвучия).

Рассмотрение ранжированного атлантического списка, представленного КП, показывает, что первые 50 позиций в нем на 30 пунктов совпадают со среднестатистическим 50-словным списком, отобранным для «Языков Африки». С одной стороны, кажется, что это не сильно превышает порог случайности (при котором ожидалось бы порядка 25 совпадений). С другой стороны, нельзя не отметить, что «атлантический список» — понятие само по себе довольно смутное, т. к. исследования самого КП и Г. Сежерера вроде бы говорят о том, что «атлантической» семьи как таковой не существует, и не является ли в этом смысле попытка замерить индекс стабильности по «атлантическому списку» аналогом попытки замерить такой индекс, скажем, по «ностратическому списку», в пределах которого мы зачастую просто не можем с уверенностью говорить об этимологических (и, соответственно, лексикостатистических) связях между сравниваемыми формами?

Чрезвычайно подозрителен, например, тот факт, что в «атлантическом списке» столь низкие рейтинги (19-е и 20-е место соответственно) занимают личные местоимения ‘я’ и ‘ты’; в этом они сближаются разве что с тайскими языками, но в тайских нестабильность личных местоимений, как правило, связана с вытеснением старых форм «вежливыми» новообразованиями — явление, типичное для юговосточноазиатского ареала, но вроде бы не очень распространенное в Африке. Уже один этот факт заставляет подозревать что-то «неладное» — например, неверную расстановку когнаций, или же просто неправильность подхода к атлантическому списку как к единой сущности. Может быть, следовало бы вместо единого атлантического списка просчитать рейтинги стабильности по твердо установленным группам (мель, бак и т. п.).

Кстати говоря, замечание КП относительно того, что «в самом конце списка [по различиям коэффициентов первой и второй половин. — Г. С.] оказываются все три африканские семьи плюс австралийская», совершенно не обязательно связано с тем, что за пределами Евразии языковые семьи ведут себя как-то по особому в плане среднестатистической устойчивости своих индексов; на самом деле все эти семьи объединяет не столько их «неевразийность», сколько элементарная неразработанность этимологии. Афразийская семья — это макросемья, где определение когнаций на самых верхних уровнях (кушитские и омотские языки + все остальные) чрезвычайно сложно и противоречиво; койсанская семья, как и атлантическая, скорее всего, представляет собой несколько семей, а про австралийскую этимологию и говорить нечего. Так что, вне всякого сомнения, «накопление материалов по стабильности базовой лексики должно быть продолжено», но и с тем, что уже накоплено, как показывает текущий опыт работы с койсанскими и нило-сахарскими языками Африки, работать вполне можно.

Что касается последнего пункта статьи КП, о корнях-изолятах, то я полностью согласен с идеей автора о том, что такие корни следует как минимум инвентаризировать и публиковать. Более того, в рамках моего собственного исследования, объединяющего этимологический подход с лексико-статистическим, ровно это и делается — поскольку в основу анализа кладется не этимологический корпус, а полный список Сводеша, все такие корни оказываются как минимум замеченными, как максимум же иногда проводится попытка их этимологизации, например, как возможные заимствования (не обязательно из древних субстратов, иногда получается и объяснить их как относительно недавние заимствования из соседних семей). Вообще, на мой взгляд, одна из основных проблем современной макрокомпаративистики связана с чрезмерным фокусированием исследователя на «общем» между дальнородственными языками — качественная теория родства требует не только обнаружения между ними системных сходств, но и реалистичного учета и объяснения накопленных расхождений.

Откуда берутся «корни-изоляты»? При условии, что речь идет действительно о родственных языках (что для атлантического массива справедливо на «низких» уровнях дерева, но не столь очевидно на «верхних»), здесь могут быть только два объяснения, и, думаю, что оба должны быть в разной степени верны. Во-первых, до тех пор, пока у нас нет четкой системы фонетических соответствий между



атлантическими языками, не содержащей существенных лагун и наглядно иллюстрированной на материале опубликованного корпуса, вряд ли можно поручиться, что все эти «корни-изоляты» действительно несводимы к тем или иным корням праатлантического или, за неимением такового, нескольких «праатлантических» языков (пра-мель, пра-бак и т. п.).

Во-вторых, в отличие от КП, я был бы готов вполне серьезно отнестись к гипотезе о «шести десятках субстратов» на западноафриканском атлантическом побережье. Не очень понятно, откуда берется число 60 — т. е. понятно, что оно соответствует общему числу атлантических языков, в каждом из которых обнаруживаются «корни-изоляты», но почему мы обязаны считать, что каждый язык обязательно заимствовал из уникального субстрата? Размышляя логически, для этого нам потребовалось бы выполнение следующих условий: (а) некоторые значения (будь то «сводешевские» или любые другие) в *каждом* атлантическом языке выражаются формами, несводимыми друг к другу; (б) ни для одной из этих форм мы вообще не можем предложить никакой внутренней этимологии и вынуждены считать их заимствованиями. Однако речь, насколько я понимаю, идет о другом — в каждом атлантическом языке то и дело встречаются «загадочные» слова с неизвестной этимологией. Это совершенно не эксклюзивно-уникальная особенность атлантических языков: она характерна для подавляющего большинства языков Африки, да, пожалуй, и всего мира.

Единственный тип ситуации, при котором «субстратная» гипотеза была бы априорно маловероятна — это ситуация, при которой территория Западной Африки заселяется носителями атлантических языков «с нуля»; но вроде бы история атлантических народов и археологические данные твердо говорят об обратном, так что в факте смешения атлантических племен с «доатлантическим» населением региона вряд ли приходится сомневаться. Насколько лингвистически разнообразен был этот регион до того, как в нем обосновались разные группы «атлантов», можно только догадываться, в том числе, кстати, привлекая как раз «субстратные» данные, особенно таких богатых на «странные» слова языки, как лимба и т. п.

Можно ли использовать «корни-изоляты» для определения рангов стабильности элементов списка Сводеша? Думаю, что нет или, по крайней мере, такую процедуру следует проводить обособленно от процедуры определения рангов стабильности, описанной в статье С. А. Старостина. Дело в

том, что если в статье Старостина стабильность замеряется в первую очередь на основании *внутренних* замен, то «корни-изоляты» (если только это действительно корни-изоляты, а не нормальные унаследованные основы, для которых не удалось пока найти достоверную этимологию) — это в первую очередь замены *внешние*, т. е. обусловленные специфическими социо-культурными условиями. Ранжирование сводешевских элементов на этой основе, скорее всего, будет как минимум сильно отличаться от ранжирования по методу Старостина, а как максимум, составление для него некоторого общеуниверсального индекса (который учитывал бы и атлантические субстраты, и сонгайско-берберские контакты, и судьбу дравидийского языка брахуи, и китаизмы в тайских языках, и латинизмы в албанском, и т. д. и т. п.) вообще не будет иметь никакого смысла — впрочем, попробовать, конечно, можно.

Резюмируя вышесказанное, хочется в любом случае выразить благодарность автору статьи за поднятые им вопросы и высказанные предложения — хотя бы уже за то, что в статье кратко и наглядно суммируются все основные проблемы, которые ожидают исследователя при попытке применения метода «предварительного лексикостатистического обследования» к материалу столь трудного языкового массива, как атлантический (хотя, на мой взгляд, «трудность» его вполне сопоставима, а во многих отношениях существенно уступает «трудности» анализа, скажем, койсанского языкового массива, к которому этот метод был применен вполне успешно). Будущее покажет, насколько успешно эти проблемы получится преодолеть (в том числе и в рамках продолжения моего собственного проекта по детальному лексикостатистическому обследованию языков Африки).

В определенной степени согласен я и с основным выводом автора («получается, что все равно нельзя»), *при условии*, что мы точно понимаем, чего конкретно нельзя — а именно, формулировать гипотезы языкового родства на глубине «*before Indo-European*» через поверхностное сопоставление базисной лексики современных языков, следующее размытой и предельно обобщенной системе правил и критериев. Там же, где поверхностное сопоставление (а) осуществляется на основе четко сформулированной и типологически обоснованной системы правил, (б) сочетается с дополнительными проверочными фильтрами, такими как ясно оговоренная процедура «ручной коррекции» и тест на динамическую градацию результатов, (в) учитывает принцип ступенчатости, т. е. требует

хотя бы прикидочную реконструкцию *прасписков* для сравнения на глубоких хронологических уровнях, оно, на наш взгляд, представляет собой вполне законный и, может быть, даже оптимальный способ формулирования серьезных гипотез языкового родства на разных уровнях *без* сопутствующего требования полной реконструкции этимологического корпуса (которая, кстати говоря, для многих языковых семей по ряду технических причин вообще не может быть осуществлена).

Что касается альтернативных критериев, защищаемых в последнем абзаце текста КП (морфологические изоглоссы, фонологическая статистика и т. п.), то, разумеется, ни «предварительная», ни «финализирующая» лексикостатистика не отменяют их значимости. Дело здесь скорее в следующем: для любой науки важнейшая задача — правильная расстановка приоритетов, а для лингвиста, занимающегося проблемами языкового родства, важнейший приоритет — выявление в языке максимально устойчивого «ядра». На роль такого «ядра» в качестве *универсального* стандарта, как неоднократно аргументировал в своих работах С. А.

Старостин (и с этим я полностью согласен), в наибольшей степени претендует именно базисная лексика, а если это так, то, разумеется, все остальные критерии по определению могут быть обозначены как «сопутствующие». Другое дело, что бывают ситуации, когда (а) базисная лексика сравниваемых языков известна нам в недостаточной степени — например, из-за неаккуратности описания, или из-за того, что описание фокусируется исключительно на грамматике, или, как это часто бывает с древними письменными языками, известные нам тексты представлены лексикой в основном культурного характера; (б) сравнение базисной лексики осуществляется некорректным образом — это уже вина исследователя, а не искривление закономерного процесса языковой эволюции. И с той, и с другой ситуацией надлежит бороться — в первом случае стремиться к полноте и аккуратности описания, а также к полноценному учитыванию всех доступных источников, во втором — к повышению качества методологии, в чем мы все, разумеется, заинтересованы и чему, хотелось бы надеяться, данная дискуссия тоже будет способствовать.

К. И. Поздняков  
INALCO, IUF, LLACAN-CNRS (Paris)

*От дальнего родства позиций — к ближнему  
(ответ на отзывы Г. С. Старостина и Д. Р. Лещинера)*

Прежде всего хочу поблагодарить журнал за предложение опубликовать статью в рубрике «Дискуссии». Это идеальный формат для любого автора: он позволяет получить квалифицированные отзывы на публикацию и к тому же ответить на критику. В определенном смысле проще отвечать, когда в отзывах выражается прямое несогласие с авторской позицией. Сложнее реагировать, когда рецензенты обсуждают мотивы написания статьи, цели, которые преследовал автор, круг ее потенциальных читателей и другие вопросы, имеющие лишь косвенное отношение к развиваемым в статье положениям, и при этом выражают согласие с ее основными тезисами. Если оставить в стороне дипломатические обороты, суть отзыва Г. С. Старостина (далее — Г.С.) я понял так: в статье содержатся отдельные интересные наблюдения, но не вполне

ясно, зачем и для кого их нужно публиковать. Г.С. прямо характеризует статью «как набор любопытных и во многом полезных размышлений, формально связанных друг с другом, но не имеющих единой общей цели (по крайней мере, очевидной)». Прежде всего, я охотно принимаю первую часть этой формулировки. Что касается «целей», то они вполне точно сформулированы самим Г.С. в конце его отзыва: «в статье кратко и наглядно суммируются все основные проблемы, которые ожидают исследователя при попытке применения метода „предварительного лексикостатистического обследования“ к материалу столь трудного языкового массива, как атлантический». Надеюсь, что это именно так — для этого статья и написана.

Тогда в чем же основная проблема, которая перевела статью в русло полемики? В отсутствии по-

тенциального читателя, которому эти проблемы могли бы быть интересны? В отзыве Г.С. высказывается мнение, что вопрос, поставленный в первой части статьи<sup>1</sup>, не имеет адресата, поскольку он не интересен ни «сплиттерам» (а), «ни ламперам» (б) и при этом статья не «обращена на критику» общих теоретических положений, сформулированных в монографии Г. С. Старостина (с), ибо подменяет важные исходные посылки другими. Охотно принимаю и эти три пункта, хотя не скрою, что выстроенная триада меня несколько озадачила и заставила задуматься, поскольку мне действительно нечего сказать ни сплиттерам, ни ламперам, ни теоретикам глобальной макрокомпаративистики.

Последний пункт для меня по ряду причин особенно важен, и мне хотелось бы, чтобы именно в нем авторская позиция была сформулирована предельно четко. Мне будет искренне жаль, если у читателя сложится впечатление, что в статье делается попытка ревизии основных принципов использования лексикостатистики для установления дальнего родства, разработанных московской школой компаративистики. Эти принципы системно обсуждаются в монографии Г. С. Старостина «Языки Африки», которая на сегодняшний день является, по моему убеждению, лучшей обобщающей публикацией на эту тему. Еще раз подчеркну то, что прямо написано в статье: я не только высоко ценю эту монографию, но и полностью разделяю ее основные положения. В целом я, как и многие компаративисты, с интересом слежу за глобальным проектом «Evolution of Human Languages» (одним из руководителей которого является Г. С. Старостин), нацеленным на создание единой генеалогической классификации языков мира. Отсюда ясно, что предложенная мной статья не полемизирует с изложенными в монографии Г. С. Старостина принципами — она на них опирается. Что же касается сплиттеров и ламперов, то позволю себе напомнить, что параллельно с этими достойными людьми существуют конкретные специалисты в конкретных областях сравнительно-исторического языкознания. Занимаясь малоизученными семьями языков, они вправе задать себе вопрос, который, строго говоря, не имеет отношения к классическому сравнительно-историческому методу<sup>2</sup>:

можно ли построить генеалогическую классификацию этих языков, и если можно, то как это сделать?

К таким специалистам относит себя и автор, размышления которого адресованы специалистам в смежных областях. Почему бы в таком случае не взять «атлантический фрагмент» мировой лексикостатистической классификации и попросту не опубликовать его с минимальными комментариями? Ответ на этот вопрос и может, вероятно, объяснить суть «полемики».

Года четыре назад с моим коллегой Гийомом Сежерером мы решили пересмотреть генеалогическую классификацию атлантических языков, которыми давно занимаемся, и даже опрометчиво объявили о готовящейся публикации на эту тему. Казалось бы, в чем здесь проблема? Мы заполнили матрицу по стословнику Сводеша и, пользуясь личными связями, обратились к Г. С. Старостину с просьбой обработать наши данные. Через несколько дней (ноябрь 2010) мы получили красивое дерево<sup>3</sup>. Казалось бы, можно было опубликовать его и считать вопрос закрытым. Но дело в том, что, изучив это дерево, мы увидели, что публиковать его нам совсем не хочется.

Какие результаты мы получили? Праатлантический язык датировался приблизительно 10 000 до н. э. Для тех, кто следит за развитием атлантической компаративистики, это может показаться неудивительным. Накапливается все больше аргументов для того, чтобы считать, что традиционно объединяемые в одну группу языки мель и собственно атлантические языки представляют собой две самостоятельные ветви нигер-конго, а это значит, что указанная дата может относиться к времени распада нигеро-конголезского языка. Но далее оказывается, что ряд языков, традиционно относимых к североатлантическим языкам (волоф, нун, буй), отстоят дальше от остальных северных языков (фула, сереер, тенда и др.), чем, например, языки бак (диола и проч.)! Эта «революция» в атлантическом языкознании прослеживается в период 5000—6000 до н. э., причем очевидно, что по-другому интерпретированные три-четыре слова могут радикально изменить классификацию в ее наиболее важных фрагментах<sup>4</sup>. Понятно, что такое дерево публиковать не только бессмысленно, но и

<sup>1</sup> Г.С. формулирует его так: «Можно ли классифицировать языки на уровне макросемей, не используя ступенчатую реконструкцию на базе сравнительно-исторического метода?»

<sup>2</sup> О соотношении генеалогической классификации и собственно сравнительно-исторического метода интересно писал В. Я. Порхомовский (Порхомовский 1982).

<sup>3</sup> Еще раз хочу поблагодарить Г. С. Старостина и его коллег за оперативную помощь.

<sup>4</sup> Близкие языки легко объединяются в группы без использования лексикостатистики. Можно сказать иначе: они объединяются в результате применения любой лексикостатистической методики, кроме заведомо абсурдных.

вредно, учитывая, что классификации языков имеют свойство широко тиражироваться и в итоге возвращаться к авторам в виде незыблемых истин.

Снова с помощью Г. С. Старостина мы сделали вторую попытку. Отказавшись от формальных критериев установления сходств, мы заново заполнили матрицу, используя наши (пока еще очень фрагментарные) знания в области атлантической этимологии, и получили новое дерево (март 2011). Естественно, наш праязык значительно помолодел. Разделение собственно атлантических языков и языков мель датируется в новой «классификации» как 6500 до н. э., а разделение собственно атлантических языков (без изолированных языков лимба и гола) на две большие группы — как 5600 до н. э. Исчезли явно нелепые группировки языков, но процент сходств, определяющий отнесение некоторых языков (или подгрупп) к одной из двух групп остался настолько мал, что та или иная интерпретация по-прежнему не может считаться надежной. Так, язык налу объединился с северными языками (при этом индексы его сходств колеблются от 0,10 с волоф до 0,21 с джаад) и, соответственно, отделился от центральных языков, индексы сходств с которыми колеблются у налу от 0,08 с диола керак до 0,18 с балант и манкань. Индексы не только мало различаются в двух группах, они и очень малы абсолютно — настолько, что приближаются к интуитивному порогу случайности. Так, языки пепель и суа показывают 3% сходств, а внутри собственно атлантических биджого-буй дают 8%, как и рассмотренная выше пара налу — диола керак.

Отсюда, в ходе работы над конкретным атлантическим материалом, и возник вопрос — а собственно, сколько случайных сходств между двумя неродственными языками мы должны получить? Оказалось, что этот вопрос изучен явно недостаточно, причем немногочисленные публикации на эту тему ограничиваются несколькими конкретными критериями формального определения сходства и не затрагивают другие алгоритмы (формализованные или интуитивные), которые применяются компаративистами на практике. Так, автору пришлось самому анализировать критерии сходств у Гринберга — их никто, насколько мне известно, не публиковал, хотя ссылок на Гринберга имеются сотни, если не тысячи. Мне показалось интересным поделиться результатом моего анализа с компаративистами, работающими с другими языками. Так, например, Гринберг сближает только **одинаковые** веллярные и палатальные и практически никогда не сближает **разные** согласные

внутри этих двух рядов (например, [ç] и [j]). Это значительно более строгий алгоритм, чем допуск различия по одному признаку и тем более допуск сближения в рамках консонантного класса. И тем не менее случайные сходства по критериям Гринберга должны составить не менее 16%, если логика моих подсчетов верна (а ее вроде бы мои оппоненты не оспаривают). Предлагая для публикации результаты такого рода, я надеялся (и продолжаю эту надежду питать), что они могут быть интереснее коллегам, чем очередные общие рассуждения «за» или «против» массового сравнения.

При тестировании количества ожидаемых случайных сходств я использовал очень разные критерии — от самых строгих до наиболее свободных. В отзыве Г.С. выводы первой части статьи ставятся под сомнение в связи с тем, что автор «использует чрезвычайно либеральный алгоритм сопоставления». Но единственная «позитивная» гипотеза статьи, касающаяся минимального числа сходств, ориентирована на самый строгий из всех возможных критериев! В статье на основе исследования вероятностей совпадений высказывается предположение о том, что 5% **полных** сходств убедительно свидетельствует о родстве языков. Если мы будем считать только примеры полной идентичности согласных в двухконсонантных корнях списков французского и русского языков, мы получим 7 сходств, а именно: *tourir* — *умирать* 13<sup>5</sup>, *lune* — *луна* 18, *poiveau* — *новый* 23, *plein* — *полный* 34, *soleil* — *солнце* 39, *blanc* — *белый* 79, *écorce* — *кора* 102. Вероятность случайности этого события настолько низка, что языки можно считать родственными. Конкретно, при 20 согласных в каждом языке и при самых благоприятных условиях для их случайных совпадений (одинаковый инвентарь согласных в двух языках и равные частоты согласных во всех позициях<sup>6</sup>) вероятность полного совпадения одного слова стословного списка (по двум согласным) составляет всего четверть процента ( $100 * 0,05 * 0,05$ ). Подчеркну, что обратная ситуация (отсутствие совпадений) не говорит об отсутствии родства. Так, сопоставление английского и русского списков дает только один пример, удовлетворяющий заданным жестким условиям (если отвлечься от ситуации с гласными): *many* — *много* 106.

<sup>5</sup> Цифрами обозначены ранги стабильности приведенных значений по С. А. Старостину. Тестирование проводится без учета фактора заимствований.

<sup>6</sup> Важный в практическом отношении вывод, обоснованный в статье: чем больше варьируют частоты согласных, тем больше вероятность их совпадений / сходств в разных языках.

Вернемся к атлантическим данным. Сходства в стословнике для языка лимба, который традиционно объединяют с атлантическими, дают следующие проценты:

- лимба — североатлантические языки — от 7% до 13%.
- лимба — центральноатлантические языки — от 4% до 14%.
- лимба — мель языки — от 6% до 12%.

Убежден, что, как для меня, так и для моих оппонентов бесспорно следующее: за фрагментом дерева, на котором лимба оказывается на 250 лет ближе к северно- и центральноатлантическим языкам (6250 до н. э.), чем к мель (6500 до н. э.), ровно ничего не стоит. Такая конфигурация определяется одним-единственным сходством. Вместе с тем, очевидно, мы не можем утверждать, что язык лимба неродственен атлантическим языкам и языкам мель, хотя мой «слишком вольный алгоритм» предполагает, что чисто случайные сходства должны давать не меньший процент. То есть про связи лимба мы не можем сказать ничего. Посмотреть, сколько слов из семи имеет высокий ранг семантической стабильности, а сколько — низкий, конечно, можно, но и на этом классификацию не построишь.

Можно объявить язык лимба отдельной ветвью нигер-конго, т. е. приписать языку лимба тот же статус, что и, например, семье бенуэ-конго. Такое решение легко обосновать, и формально оно будет выглядеть безукоризненным. Но все ли возможности мы испробовали, чтобы объединить лимба с атлантическими языками, что соответствовало бы традиционному взгляду на положение лимба? На чем основан (кроме географии) этот интуитивный традиционный взгляд, который разделяется ведущими специалистами? Не стоит ли в этом случае опереться на данные исторической морфологии и, прежде всего, на системы именных классов, которые по ряду важных параметров оказываются более близкими у лимба и у атлантических языков и не похожи в этом плане на системы языков банту? Не имеет ли смысл, в ситуации, когда индексы сходств слишком малы, изучить в первую очередь те 47 слов лимба из стословника Сводеша, которые не находят видимых параллелей ни в одном из десятков атлантических языков? Это далеко не полный круг вопросов, которые должны, на мой взгляд, интересовать специалистов, занимающихся генеалогической классификацией языков Африки, и которые мне представлялось важным обсудить в статье. Если «максимально устойчивое „ядро“» (Г.С.) в языках уже **известно и является универсальным**, то почему же «для лингвиста, занимающего-

ся проблемами языкового родства, важнейший приоритет — выявление в языке максимально устойчивого „ядра“» (Г.С.)? Зачем так важно выявлять заново то, что известно заранее? Применительно к атлантическим языкам могу сказать ответственно: если выбирать из двух предложенных формулировок Г.С., следует выбрать вторую, а именно выявить наиболее устойчивый пласт форм, применяя весь комплекс возможных подходов, уже апробированных или совсем новых.

Отмечу, что, на наш взгляд, наиболее показательным для атлантической классификации является традиционный поиск инноваций в лексике и особенно в морфологии и максимальная тщательная верификация этих инноваций с привлечением данных внешнего сравнения. Неожиданно интересные результаты дает и анализ «изолированной» лексики — подход, который рассматривается в третьей части статьи. Я рад, что Г.С. «полностью согласен с идеей автора о том, что такие корни следует как минимум инвентаризировать и публиковать». Более того, резюме Г.С. как нельзя лучше суммирует то, что я пытаюсь донести до читателя — с большим удовольствием приведу формулировку из отзыва Г.С.: «Вообще, на мой взгляд, одна из основных проблем современной макрокомпаративистики связана с чрезмерным фокусированием исследователя на „общем“ между дальнородственными языками — качественная теория родства требует не только обнаружения между ними системных сходств, но и реалистичного учета и объяснения накопленных расхождений». По-моему, лучше не скажешь. Замечу, что далеко не все компаративисты готовы принять эти положения.

Допуская принципиальную ценность изучения изолированных корней, Г.С. выражает сомнение в перспективности их использования для определения рангов стабильности. В этой связи хочу воспользоваться этим пунктом, чтобы высказать некоторые общие соображения о применении статистики в сравнительно-историческом языкознании (помимо традиционной лексикостатистики). Давние поиски в этой области<sup>7</sup> и многочисленные обсуждения разнообразных методик с коллегами, а также со специалистами в области статистики сформировали у автора твердое убеждение, что «идеальных» формул в этой области не существует.

<sup>7</sup> Взгляды автора на этот вопрос впервые системно изложены в разделе «Статистика и компаративистика» книги [Поздняков 1993]. Препринт последней общей публикации на эту тему — «Statistique et comparatisme» — доступен на <http://pozdnjakov.free.fr>.

Прежде чем раскрыть это «крамольное» положение, позволю себе привести легенду, которую мне довелось слышать в аспирантские годы в Институте языкознания АН СССР (еще на Волхонке). На одной из конференций к трибуне вышел очередной докладчик. Он выдержал эффектную паузу, а затем резко выхватил длинную линейку и, потрясая ей, сообщил притихшей публике: «С этого инструмента должно начинаться любое лингвистическое исследование!» Занимался докладчик полисемией. Он ее мерил... в сантиметрах, сравнивая длину словарных статей. Логика была неувязима: в статье к слову *телефон* — мало сантиметров, а к слову *резать* — много; следовательно, слово *резать* — полисемичнее. Зададим себе вопрос: работает эта статистическая методика или нет? В принципе работает. Если у скептиков найдутся контрпримеры, методику можно будет скорректировать. Возможно, кто-то предложит новую методику, которую проще применить сегодня: можно мерить не в сантиметрах, а в количестве слов словарной статьи. Более того, если поставить себе глобальную задачу — выделить 100 наиболее полисемичных слов в языках мира — возможно, это будет единственная статистическая методика, у которой будет шанс дать любопытные результаты при минимуме временных затрат.

Подчеркну, что, приводя этот лингвистический анекдот, автор не предлагает читателю мерить ранг стабильности слов «в попугаях». Просто одна из ключевых проблем в подборе статистической формулы к сложнейшим языковым массивам, на формирование которых оказывает влияние множество разнообразных факторов, состоит в подборе такой формулы, которая, минимально искажая результат, была бы максимально проста и максимально прозрачна. Под прозрачностью здесь понимается, прежде всего, «физический смысл» формулы, который понятен лингвисту. Так, процентное отклонение от среднего физически понятнее, чем квадратичное отклонение, и при этом для многих массивов применение критерия  $\chi^2$  при всех его очевидных преимуществах существенно не меняет получаемые результаты.

В формулу С. А. Старостина, рассмотренную в статье, заложена попытка нейтрализовать очевидную «зависимость коэффициента стабильности от глубины семьи и числа сравниваемых языков». Вместе с тем применение предложенной процедуры подсчета индексов разрушает прозрачность формулы в указанном выше смысле и к тому же предельно затрудняет количественное сопоставление индексов (они задуманы как относительные,

а не абсолютные). Последнее для меня было особенно важно, потому меня интересовало именно сопоставление индексов по каждому отдельному значению. Достаточно неожиданным для меня оказалось то, что применение более простой и физически понятной формулы дает не менее точный результат, а это означает, что влиянием указанного фактора, на который вводился коэффициент, вполне можно пренебречь. Более того, выяснилось, что по ряду важных параметров «примитивная» формула работает лучше: полученные с ее помощью результаты позволяют «узаконить» именно те изменения в первой половине списка Сводеша, которые московская школа и предлагает ввести, обосновывая эти изменения «техническими (?) причинами» (Г.С.). И наконец, оказалось, что очень неплохие результаты дает использование другого простейшего критерия — отношение количества корней в рамках одного значения к количеству языков. Достоинство этого критерия в том, что он учитывает и корни-изоляты, и, таким образом, выявляется, например, особый статус языка лимба при его сопоставлении с атлантическими языками. В связи с этим мне непонятен комментарий Г.С.: «Ранжирование сводешевских элементов на этой основе, скорее всего, будет как минимум сильно отличаться от ранжирования по методу Старостина [...] впрочем, попробовать, конечно, можно». Вообще-то говоря, я попробовал. Результаты такого ранжирования — в тексте статьи, что позволяет рассчитывать на более предметный уровень критики.

В целом приходится признать, что ответственность за целый ряд указанных недоразумений несет автор — вероятно, следовало более четко обозначить задачи статьи в начале и суммировать выводы в заключении. Однако дело не только в этом: суть полемики, как мне показалось, не в столкновении двух разных точек зрения, а в существовании разных углов зрения на одну проблему, — в разных **ракурсах** рассмотрения проблем дальнего родства. Рискну предположить, что у меня будет еще меньше расхождений с Г.С. — автором генеалогической классификации койсанских языков, чем с Г.С., защищающим глобальный проект классификации языков мира.

В этом плане мне был чрезвычайно интересен и третий угол зрения на количественные критерии установления родства и рангов семантической стабильности, представленный в отзыве авторитетного специалиста в области статистики Д. Р. Лещинера. Прежде всего, для меня важно, что предлагаемые в статье количественные критерии в основ-

ном успешно прошли профессиональную экспертизу. Помимо этого, считаю нужным отметить, что отзыв Д. Р. Лещинера, изложенный в виде «Заметок», имеет, на мой взгляд, самостоятельное научное значение безотносительно к обсуждаемой статье. Я имею в виду последнюю часть «Заметок», посвященную возможностям исчисления «апостериорной вероятности родства слов» и подходам к

«ступенчатой» лексикостатистике. Эти новые подходы представляются мне перспективными, тем более, что они заставляют задуматься над рядом других, практически не исследованных вопросов, и, в частности, над вопросом о влиянии конфигурации генеалогического древа на возможности реконструкции. Впрочем, это уже предмет отдельной дискуссии.

### Литература

- Поздняков 1993 — К. И. ПОЗДНЯКОВ. *Сравнительная грамматика атлантических языков*. М.: Наука, 1993. [K. I. POZDNYAKOV. *Sravnitel'naya grammatika atlanticheskikh yazykov*. М.: Nauka, 1993.]
- Порхомовский 1982 — В. Я. ПОРХОМОВСКИЙ. Проблемы генетической классификации языков Африки // *Теоретические основы классификации языков мира. Проблемы родства*. М.: Наука, 1982. [V. Ya. PORKHOMOVSKIY. *Problemy geneticheskoy klassifikatsii yazykov Afriki // Teoreticheskie osnovy klassifikatsii yazykov mira. Problemy rodstva*. М.: Nauka, 1982.]
- Г. Старостин 2013 — Г. С. СТАРОСТИН. *Языки Африки. Опыт построения лексикостатистической классификации*. Т. 1: *Методология. Койсанские языки*. М.: ЯСК, 2013. [G. S. STAROSTIN. *Yazyki Afriki. Opyt postroeniya leksikostatisticheskoy klassifikatsii*. Т. 1: *Metodologiya. Kojzanskie yazyki*. М.: YaSK, 2013.]
- С. Старостин 2007 — С. А. СТАРОСТИН. Определение устойчивости базисной лексики // С. А. СТАРОСТИН. *Труды по языкознанию*. М.: ЯСК, 2007. С. 827—839. [S. A. STAROSTIN. *Opreделение ustojchivosti bazisnoj leksiki // S. A. STAROSTIN. Trudy po yazykoznaniiyu*. М.: YaSK, 2007. S. 827—839.]
- Campbell & Poser 2008 — L. CAMPBELL, W. J. POSER. *Language Classification: History and Method*. Cambridge University Press, 2008.
- Greenberg 1963 — J. GREENBERG. *The Languages of Africa*. Bloomington: Indiana University Press, 1963.
- Pozdniakov & Segerer 2007 — K. POZDNYAKOV, G. SEGERER. Similar Place Avoidance: A Statistical Universal // *Linguistic Typology* 11, 2. 2007. P. 307—348
- Sapir 1971 — J. D. SAPIR. West Atlantic: an inventory of the languages, their noun class systems and consonant alternation // T. A. SEBEOK (ed.). *Current trends in linguistics 7: Linguistics in Sub-Saharan Africa*. The Hague & Paris: Mouton & Co, 1971. P. 45—112.

Konstantin POZDNYAKOV. On the threshold of relationship and the “stability index” of basic lexicon in mass comparison: Atlantic languages

The paper deals with certain key aspects of the procedure of establishing long-range genetic relationship between languages in the absence of a classic step-by-step reconstruction, such as: 1) the issue of establishing the randomness threshold for lexical matches on the Swadesh 100-item wordlist, depending on the particularities of the employed algorithm; 2) the issue of stability ranking in the sphere of the basic lexicon; 3) the issue of multiple isolated roots that do not find any parallels in the other languages of the family. These general topics are discussed in the context of specific problems that arise during research on the genetic classification of Atlantic languages. The latter are traditionally considered to constitute a branch of the Niger-Congo macrofamily, but lexicostatistics suggests that their date of separation is no younger than 8000 BP. Two critical responses to the paper present further opinions on the difficult problematics of establishing long-range relationship; the author’s final response is also to be regarded as an important constituent of the published paper.

*Keywords:* protolanguage reconstruction, basic lexicon, lexicostatistics, etymology, Atlantic languages, macro-comparative studies.