

К.И.Поздняков
(INALCO-IUF-LLACAN)

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

В статье мы продолжим обсуждение критериев установления дальнего родства языков. Для семей старше индоевропейской (BIE) такая дискуссия по-прежнему актуальна, несмотря на публикацию нескольких фундаментальных работ, посвященных этому сюжету¹. Речь пойдет о трудных случаях классификации (например, проблемы классификации Нигер-Конго), когда «нельзя» опереться на классические методы (в отсутствие ступенчатой реконструкции), «но очень хочется» получить предварительные результаты. То есть 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, в таких трудных случаях, какие представляют, в частности, атлантические языки, которыми я занимаюсь. Если принять терминологию, предлагаемую Г.С.Старостиным («группа» – таксоны, показывающие от 40% и выше совпадений, «семья» - от 15 до 40%, «макросемья» - менее 15%, [GS, 140]), то атлантические языки – это в лучшем случае макросемья (по подсчетам Д.Сэпира, языки волоф-темне дают 3% сходств! [Sapir, 1971]).

В [GS] 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, 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

¹ В 2013 вышла в свет исключительно интересная монография Г.С.Старостина (в дальнейшем, [GS]), с которой во многом перекликается данная статья.

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 нигеро-конголезскую ветвь или же вообще не является нигеро-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.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» [GS, 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 мы 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 исключением, 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 (то есть 25) совпадет по рядности начального согласного. Четверть из них (то есть 6) одновременно совпадет по рядности конечного согласного. Половина из них (то есть 3) покажет сопоставимый гласный. Англ. *meat* ~ русск. *m'as-o* сходны по принятому критерию, поскольку они родственны, а их сходство с манджак *maʃ* 'мясо' случайно, но неудивительно, как и сходство англ. *mouth* и бедик *m'isy* или англ. *liver* и бапен *dɛmbar*.

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.

Табл. 1

ATL (%)	initial	final
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 относятся к дентальному ряду. Для двух языков, близких в 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.² Учитывая, что качество гласных при хорошем совпадении согласных практически не принимается во внимание при установлении сближений в африканских языках³, приходится констатировать, что грань между случайным сходством и сходством при дальнем родстве практически стерта. Это становится тем более ясным, если учесть, что при пороге 10 число сходств от 7 до 13 не может считаться существенным отклонением, оставаясь в пределах трех сигм.

Как было показано выше, на порог случайных сходств существенно влияют частоты согласных. А это означает, что лексикостатистика на основе сходств в атлантических языках, показывающих от 5% до 20% для языков разных ветвей, практически неприменима для построения атлантического дерева, поскольку небольшие различия коэффициентов могут определяться особенностями частотного распределения в конкретных парах языков.

² То есть такие сближения как диола-гусилай *mεεη* с англ. *manu* не будут считаться сходством.

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

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 монсиллабов 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 таблицы 2 по 24 атлантическим языкам.

Табл.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»

([GS,61,62]⁴) до существенно более гибких – введение «фонетически изоморфных параллелей» как «некоторого компромиссного варианта между «классическими регулярными соответствиями» и аморфным «фонетическим сходством» в рамках теории «массового сравнения» ([GS,55]). К компромиссным попыткам формализации консонантных сходств относятся и выделение 10-ти непересекающихся консонантных классов А.Б.Долгопольского и 12-ти классов С.А.Старостина ([GS, 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 – holodnyj, 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' *wolof picc ~ limba beci* (где *be-* возможно является префиксом именного класса) и 'сидеть' *wolof toog ~ limba dɔŋɔ* (если считать, что *-g* и *-ŋ* различаются одним признаком, а не двумя). Можно еще добавить сюда 'колени' *wolof oom ~ limba hubuŋ*, 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.

Выделение десятка непересекающихся классов не улучшает дела. Проблемы, связанные с использованием классов и построенных на их использовании формальных алгоритмов сближений, прекрасно рассмотрены в [GS,185-190]. Абсолютно справедливо замечание Г.С.Старостина о принципиальной ложности исходного посыла – «фонетические изменения происходят в пределах одного и того же консонантного класса» [GS,188]. Совершенно прав Г.С.Старостин и в том, что реально мы имеем дело с «пересекающимися подмножествами общего консонантного инвентаря» [GS,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

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

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.

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 неинтересно – 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».

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):

Табл. 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</i> , <i>gbɔŋ</i> , бага <i>maduri baki</i>
6	bird	
7	bite	wolof <i>màtt</i>
8	black	bedik <i>-bálà</i> , biafada <i>baŋg</i> , peul <i>balee</i> , sereer <i>baal</i> , nalu <i>bal</i> , limba <i>boloi</i>

9	blood	jaad poadə, buy bilɛr,
10	bone	joola wul, manjak moe, pepel mɔ, nalu wol, бага-мaдuri / temne ke-bant

Удивительно не то, что для каждого английского слова в семье из 40 языков находятся созвучные слова. Удивительно скорее, что для 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 отвергает метод «массового сравнения», но при этом опирается на гринберговскую классификацию, «в силу полного отсутствия альтернативных схем» [GS,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 последние 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, то есть 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 $n = \tilde{n}$ 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 (таблица 4).

Табл. 4

	p	f	b	v	w	t	r	d	L	c	s	j	Y	k	h	g	m	n	Ñ	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		
ñ					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.

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 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 ($100 * 0,25 * 0,48$). Для 52% нам необходимо совпадение начального и неначального согласных что дает еще 4 слова ($52 * 0,25 * 0,32$). 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 учитывая 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 корней⁵.

На практике в африканской макрокомпаративистике, как показывает анализ конкретных публикаций, алгоритмы выделения когнатов достаточно «либеральны». Наиболее типичным 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», то есть допускается сближение *ti ~ ci, ge ~ 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.

⁵ Привожу это очевидное рассуждение, учитывая явное статистическое недоразумение в [GS,42], касающееся процента случайных совпадений в «слабой» лексике. В приводимом там примере, как пишет автор, «ожидалось бы скорее число, близкое к нулю», хотя в действительности ожидание составляет 50%.

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).

Табл.5

	P	t	c	k	f	r	S	h	b	d	j	g	w	l	y	Ø	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	
Ø	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

В нашей таблице 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).

Табл. 6

j	ηη	m	η	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. *дум* 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 эмпирически нащупать реальный порог случайных сходств при реально используемых алгоритмах сближения слов.

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 английским⁶. 12 языков показывают 10 и более процентов сходств. Приведем созвучные слова в 6-ти из них⁷ (табл. 7).

Табл. 7

temne		biafada		bedik	
ashes	r-os	fire	fuuru	black	-bálà
bark	foki	fly	fiili	breast	ε-bār
bone	ka-bant	me	mma	burn	o-bóđ
fly	faliɾ^w	mouth	mməsə	hand	ε-kònd
go	kɔ	nose	nasin	hear	o-hér
me	mi	say	cii, re	me	-mé- / mé-
little	lɔl	white	paas	mouth	bɔ-más
one	in	you sg	yi	white	-fèsà / -pèsà
sand	a-sənt	black	baŋg?	eye(s)	gí-ngùs?
seed	ma-sənt, ma-ther	breast	bəl?	knee	é-đλ-klà?
woman	bom	ground	gal?	liver	i-dáméí?
		fish	yesä / bwa-~?	meat	nas?
		little	tiiti?		
nalu		rain	səna?		
black	bal				
drink	dig	limba		mankany	
go	kɔ	belly	fɔle	big	-mag
star	ma-saar	black	boloi	liver	u-deeb
you sg	yi	breast	hu-bili	meat	u-ye-maɕ
big	wɔk? -baŋ?	fish	fee	nail	ka-ñiiri
bone	wol?	tooth	hu-titi	what	we
breast	n-bin?	you sg	yi	white	p-faaθ
fire	mɛt?	ashes	hɔtɔ? ~ fɔtɔ?	breast	p-wil?
good	kuun?	drink	tie?	hair	ka-weel?
tree	n-ti?	eye	hoya ~ foya?	leaf	ka-toh?
we	biye?	kill	kɔra?	nose	b-yiiş?

Следует сжиться с фактом, что десяток сходств при реально существующих алгоритмах их выделения – это всего лишь норма случайных созвучий между любыми языками. А

⁶ 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?

⁷ Знаком вопроса отмечены менее «убедительные» сближения.

это означает, что на глубине BIE 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 [GS], 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-ти приведенных примерах мы обнаруживаем 4 «соответствия» звонких в английском – глухим в темне: *bark – foki, go – ko, sand – sant, seed – sɛnt*. 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 **ɓ-** 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 - b̃l̃r, knee - k̃l̃à*. Конечное **-t** (вне кластеров) дает в бедик фрикативный: *meat - jás, white - fèSà*. Трудно ожидать большей регулярности в 12-ти словах!

Признаюсь, что в отдельных случаях сближения явно неправомерны. Так, *nose* нельзя сближать с биафада *jasin*, поскольку **ja-** 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 эффективны, остается только один критерий, который был впервые четко сформулирован С.А.Старостиным и подробно рассмотрен в [GS] - ранг стабильности единиц списка Сводеша. Рассмотрим перспективы его применения к атлантическим языкам.

2. О рангах стабильности.

В своей фундамертальной работе «Определение устойчивости базисной лексики» [Starostin 2007] С.А.Старостин заложил основу для сопоставления диахронической стабильности слов в языках мира. На огромном материале (737 языков в 14-ти 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 монография Г.С.Старостина [GS], в которой предложенный С.А.Старостиным 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 максимального числа языков, использующих для данного значения один и тот же корень, к общему числу языков в семье» [Starostin 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, *trьbuxъ*, а также заимствования, не учитываются).

Вычислим по данным Старостина средний индекс стабильности для одного значения списка в 14-ти семьях⁸ (табл. 8).

Табл. 8

	YEN	TH	DR	KT	UR	IE	NC
	енисейск.	тайск.	дравид.	картвельские	урал.	и.-е.	сев.кавказск.
Н яз.	5	47	21	7	17	97	33
СР стаб.	0,67	0,65	0,56	0,54	0,48	0,45	0,43
	KS	Alt	ST	AN	AA	AUS	AF
	койсанск.	алтайск.	синотибет.	австронезийские	австроаз.	австрал.	афразийск.
Н яз.	26	54	132	94	99	36	69
СР стаб.	0,43	0,41	0,41	0,40	0,37	0,29	0,20

Действительно, как отмечал Старостин, во многом «индекс стабильности зависит от времени, прошедшего с момента распада праязыка данной семьи» [Starostin 2007, 828]. Этот фактор хорошо виден в таблице, упорядоченной по убыванию индексов

⁸ Данные получены по стословнику Сводеша – 10 дополнительных значений при подсчетах были исключены.

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 0,28⁹.

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

Табл. 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» [Starostin 2007, 833]. Однако и алгоритм, использованный Старостиным, не вполне преодолевает указанные факторы. Вместе с тем, средние (числовые) коэффициенты

⁹ В атлантическом списке значения « green » и « yellow » заменены на « walk » и « you pl. ».

конечно же удобнее для сопоставления. Получим по материалам Старостина иерархию значений по средним коэффициентам (табл. 10)¹⁰.

Табл. 10

	1	2	3	4	5	6	7	8	9	10	
01-10	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

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.А.Старостина, «в первую очередь из-зв 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 [GS, 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 место вместо 28ого). 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.

¹⁰ Плюсами и минусами обозначены все изменения десятков: так, два плюса для **head** обозначает, 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.

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.

Табл. 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 это разница исключительно велика – в 8ми 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 то, накопление материалов по стабильности базисной лексики должно быть продолжено.

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

Табл. 12

	1	2	3	4	5	6	7	8	9	10
01-10	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

Очевидно, что атлантический список существенно отличается от среднестатистического. Так, например, значение name 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 двадцатку стабильных значений. Вместе с тем, очень существенными отличиями характеризуются по признаку стабильности и другие семьи. Так, значение *pame* 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-ое 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 *pame*.

Табл. 13

	name	CP.	KT	UR	IE	NC	KS	ST	AF	DR	AUS	AN	AA	Alt	TH	ATL	Σ1	Σ2
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	

- 1) Отметим для каждого из 100 значений номер десятки, в которую оно входит в семье языков. Значение *pame* (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 десятку¹¹.
- 2) Вычтем из полученных индексов (1-10) среднестатистический индекс (в случае с *pame* – единицу) – строка 2. Таким образом, мы определяем, на сколько («десятков») индекс в семье отличается от среднестатистического.
- 3) Возможен менее детальный подсчет (строка 3) – вне зависимости от того, насколько отличается индекс от среднестатистического, будем приписывать

¹¹ 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.

люому отклонению значение 1, а отсутствию отклонений – значение 0. Например, в картвельских – ранг значение пате 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 пате 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 пате 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. Полную сопоставительную таблицу можно найти на <http://pozdniakov.free.fr>. Здесь же ограничимся суммарными данными по десяткам (аналог строки 3 в табл. 13) – табл. 14.

Табл. 14

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

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! Почеркну, 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-ти 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.

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

Табл. 15

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

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 будет учитывать радикальность этих отличий (см. пример в строке 6 табл. 13) – табл. 16.

Табл. 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	

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

Табл. 17

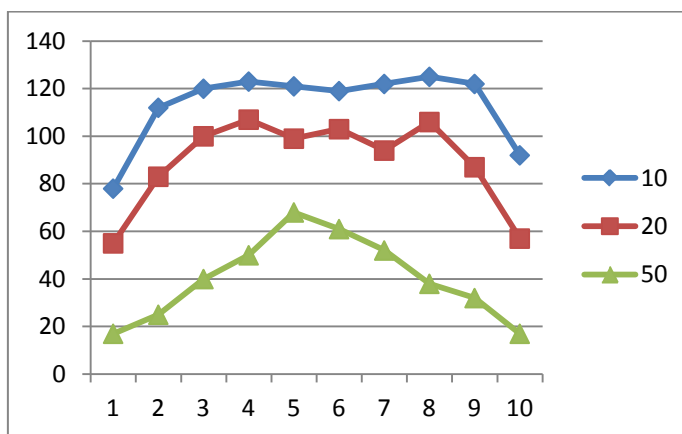
	TH	DR	KT	UR	IE	NC	KS	Alt	ST	AN	AA	AUS	AF	Atl	Σ
1-10	5	0	0	1	0	0	2	1	0	0	2	2	0	4	17
11-20	1	2	1	1	4	0	1	5	1	0	0	3	3	3	25
21-30	4	3	2	2	2	3	5	1	2	1	4	4	4	3	40
31-40	2	4	3	3	5	4	3	3	1	4	2	5	4	7	50
41-50	5	2	6	4	3	7	5	5	7	7	7	2	3	5	68
51-60	6	1	2	4	7	6	6	4	4	5	5	5	2	4	61
61-70	4	3	5	3	3	3	2	1	3	5	5	3	6	6	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	

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-ти наиболее стабильных 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



Безнадежному гуманитарии трудно понять, является ли характерная конфигурация нижней кривой (для двух половин) математически предсказуемой при подсчете отклонений от средней величины или же она рельефнее подчеркивает закономерность, наблюдаемую уже на первых двух кривых, по которой начало и конец списка более стабильны в разных семьях, чем его середина? Вроде бы, математики говорят, что случайная кривая при принятой технике подсчетов так и должна идти. Но тогда мы не можем сказать, что первая десятка стабильнее, чем остальные? В любом случае, если мы принимаем ранги стабильности усредненного списка за основу для установления родства языков, этот фактор просто необходимо учитывать.

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-ти 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 оказываются 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-ти 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. О 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 словниках 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 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)

Табл. 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											

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

Табл. 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

Отметим, что данные, приведенные в табл. 18, получены на представительном, но все 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 базе¹² собраны лексические данные по 64 языкам (атлантические и мель)¹³. В этих языках выделяется 28 различных корней со значением «женщина». 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-ти возможных структур, причем многие структуры

¹² База G.Segerer-K.Pozdniakov

¹³ Цифра не учитывает многочисленные диалекты и говоры, которые также представлены в нашей базе.

представлены разными корнями¹⁴. 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-ти звуковых комбинаций¹⁵.

PP : бага koba *bum*, темне *bom*

PT/ПК : балант *fag / far*

PT : темне *-bèrâ*

PC : лимба *ɓasa*

PK : темне *boko / borko*

PV : sua *mfo, nalu fay*

TP : palor *bɛ-leb / 'bɛ-lɛp*, ndut *'bɛ-lɛp*, laala *leb / 'bɛ-tɛ*, safen *bi-tip*, palor *bi-tif* ('femme adulte'), noon *'bɛ-teⁱ / bè-dop*, peul *debb-o / rew-bɛ*

TP : konyagi *-nəm*, bapen *-nɪm*, bullom *-nomɑ*, sherbro *-nóma* (но sherbro *no-mɑ / à-mɑ* – возможно, корень **mɑ*)

TP / CP : joola fogny *naaba*, bayot *ɲaw*

TT : kisi nord *làndò-*, kisi sud *láno*, бага Fore *laŋən*, бага koba *rani*, бага sitemu *ran*, бага maduri / landuma *raani*

TT / (TP) : балант *niin / ninn / nim*, gola *-ni(e)* – ср. *Бенуэ-Конго **nine / *nene / *nina*

TT / CT : joola (разные языки) *naalɛ / naarɛ*, joola banjal *-ɲalɛna*, joola fogny *-ɲalɛn*, biafada *naali*, bayot *-rɛr*

TK : bainunk (разные языки) *-dik-am*, kasanga / kobiana *lik-aam*, bullom *-lak-an*, sei *lag-an*

TK / CK : лимба *nañi*, joola ejabat *ɲakan*, joola kasa *ɲakunta*

TV : bullom *la*, bedik *lú-m / -nú-m*, лимба *nda*, wolof *ndaw* 'jeune fille' – ср. *Jukunoid **ndá*

CP : konyagi *-sàvɛ̀lɛ́*, bedik *sʷámɓálàŋ / -sɪmbár*, jaad *caafɛ*, gola *džàvɛ*

CP : bapen *-suwon*

CT : gola *džân*

CC : manjak *jije*, joola ejabat *ʔäŋä'ijäⁱ / ɲejäu*

CC : nalu *θay*

CK : wolof *jigéen*

CK : basari *tóxár / sóxár*, bedik *sóyár*, konyagi *ryónkæx*, tanda *zogari*, bainunk-Kas *toxol*

CK : joola fogny *seek*, joola ejabat *sɛɰ*

CV : pepel *ʃo*

¹⁴ Так, комбинация «дентальный» - V – лабиальный» отмечена в трех корнях, родство которых далеко не очевидно.

¹⁵ 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.

КТ : bijogo *-kanto* – ср. VenueCongo **-kwazi*

КС : бага Mboteni *gisi*

КС : joola gusilay *koona*.

И наконец еще одна структура – VT : joola (разные языки) *aal / aar*, karon *haal*, manjak / mankany *-aats / aat*, 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. Вроде бы, получается, 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» критериями, а имеют не меньший вес, чем сходства в стословниках.

ССЫЛКИ :

[Campbell & Poser 2008] Campbell, L., W.J.Poser. *Language Classification : History and Method*. Cambridge University Press, 2008

[Greenberg 1963] Greenberg, J. 1963. *The Languages of Africa*. Bloomington: Indiana University Press, 1963.

[GS] Г.С.Старостин. *Языки Африки. Опыт построения лексикостатистической классификации*. Т.1: Методология. Койсанские языки. Москва, ЯСК, 2013

[Pozdniakov-Segerer 2007] K.Pozdniakov, G.Segerer. *Similar Place Avoidance: A Statistical Universal // Linguistic Typology* 11, 2 2007, p. 307-348

[Sapir 1971] Sapir, J.D. West Atlantic: an inventory of the languages, their noun class systems and consonant alternation. In Sebeok, T. A. (Ed), *Current trends in linguistics, 7: linguistics in sub-Saharan Africa*, The Hague & Paris : Mouton & Co., 1971, pp. 45-112.

[Starostin 2007] С.А. Старостин. Определение устойчивости базисной лексики. // *Труды по языкознанию*. Москва: ЯСК, 2007, с. 827-839.