ВОПРОСЫ ЯЗЫКОЗНАНИЯ

№ 4 1988

ЖУРАВЛЕВ А. Ф.

ЛЕКСИКОСТАТИСТИЧЕСКАЯ ОЦЕНКА ГЕНЕТИЧЕСКОЙ БЛИЗОСТИ СЛАВЯНСКИХ ЯЗЫКОВ

Лексикостатистический подход к оценке генетической близости языков, заявленный в названии настоящей работы, не должен отождествляться с получившим в свое время большую известность методом глоттохронологии М. Сводеша [1—3]. Цели данной работы и, главным образом, объем и характер лингвистического (лексического) материала, используемого для достижения этих целей, равно как и способ его обработки, отличны от задач глоттохронологических исследований и материала, на котором они базируются, вследствие чего точек соприкосновения нашего анализа с методом Сводеша оказывается немного. Воспользуемся, однако, возможностью сравнения с глоттохронологией для того, чтобы отчетливее выявить особенности нашего подхода.

Цель настоящей работы сравнительно узка: испытание метода, основанного на лексикостатистических данных, с помощью которого «измеряется» родство языков (точнее, идиомов — лингвистических объектов в принципе любого уровня классификации, от говоров одного языка до родственных языковых семей, например, внутри ностратической гиперсемьи; в данной статье метод будет опробован применительно к нескольким группам языков, составляющих славянскую ветвь). Это означает, что, в отличие от глоттохронологических исследований в духе Сводеша, здесь не ставится задача абсолютного датирования этапов «распада» праязыка. Если у Сводеша и сторонников его метода установление картины языкового родства является хотя и необходимо важной, но лишь частью искомого результата (конечная цель все-таки хронологизация «распада»), то задачи нашего исследования этой «частью» и ограничиваются.

Отказ от датировки этапов дивергентного развития славянских языков вызван прежде всего сохраняющимися сомнениями в доказуемости постулата о постоянстве скорости, с которой меняется так называемый «базовый словарь». Однако это не единственная причина уклонения от абсолютной хронологизации процессов «распада». Метод Сводеша был подвергнут серьезной критике. Не повторяя многих замечаний в адрес глоттохронологии и не стремясь умножить ряды ее недоброжелателей (это было бы и несвоевременно и несправедливо по отношению к работам, которые породили целое лингвистическое направление, отчасти стимулировавшее и данное исследование), отметим здесь все же, что ее существенным недостатком является молчаливое допущение диалектной монолитности праязыка, т. е. фактическая опора на одностороннюю концепцию родословного древа при принятии во внимание только процессов дивергенции. Между тем представления о диалектной расчлененности в принципе любого языка, в том числе и реконструируемых праязыков, становятся уже почти банальностью, и никто не может поручиться за то, что и «базовый словарь», как его определил Сводеш, непременно с самого начала един для всех ком-

понентов (диалектов) праязыка, выделяющихся затем в самостоятельные лингвистические образования. Но даже если признать унитаристскую концепцию безальтернативной, уязвимость попыток датирования дивергенции этим не устраняется, поскольку в этом случае в упрек глоттохронологическим работам может быть поставлен неучет того обстоятельства, что «распад» праязыкового единства начинается, как это логично предположить, с низкочастотной и неустойчивой периферийной лексики и лишь какое-то время спустя затрагивает «базовый словарь». Следовательно, полученные оценки времени «распада» праязыка (или, что то же, возраста языков-потомков) неизбежно окажутся заниженными, причем неясно насколько.

К этому можно добавить, что, оперируя понятиями «праязык», «распад», «автономное развитие языков» и под., сторонники глоттохронологии мало интересуются реальным содержанием процессов дивергенции, отказываясь обсуждать сами критерии языковой самостоятельности. Отделяющиеся от языка-основы идиомы выступают в глоттохронологических работах некими бесплотными сущностями, языками «без свойств», если прибегнуть к парафразе известной формулы Р. Музиля. Это впечатление усиливается несколько гротескной формой представления результатов подсчета, принятой в глоттохронологии: начало дивергенции между чешским (!) и венгерским (!) языками (!) приходится на 6 677 год до н. э. (плюс-минус 998 лет), а между русским (!) и финским (!) — на 6 274 год до н. э. (плюс-минус 944 года) [4]. Применительно к подобным выкладкам можно сказать, заостряя ситуацию, что специалисты по глоттохронологии не знают, что случилось, но достаточно уверенно судят о том, когда это произошло.

Мало ясности вносят и такого рода теоретические поправки: «Ни один язык не существует без диалектных различий внутри его. Но суть не в материальных различиях между диалектами. Совокупность диалектов остается единым языком при всех различиях между диалектами до тех пор, пока они эволюционируют согласованно. Наоборот, самостоятельные языки характеризует независимость изменения» [5, с. 17]. Этот привлекательный на первый взгляд структурно-эволюционный критерий разграничения языка и диалекта в действительности требует разрешения множества вопросов: в чем должна быть выражена эта согласованность? можно ли выявить «порог» несогласованности? одинаков ли будет он для разных в генетическом отношении «распадающихся» идиомов?... Думается, что и в пределах славянского лингвоисторического пространства существует немало феноменов, не вписывающихся в эту жесткую и умозрительную схему, особенно если учесть, что проблема «диалект — язык» относится не только и даже не столько к сфере структурно-эволюционных контроверз, сколько, быть может, к компетенции социолингвистики и даже этнопсихологии (впрочем, по вполне понятным причинам интересы социолингвистики до праязыка не простираются [1]). Характерно, что авторы приведенного выше высказывания о согласованности эволюции диалектов и независимости изменения языков десятью страницами спустя, не усматривая в том никакого противоречия, допускают возможность «параллельного развития языков после распада праславянского единства» [5, с. 27].

[1] Ср.: «...мы исходим из признания исторического характера таких безусловно соотносительных по своей природе понятий, как „язык“ и „диалект“, а также из принципиальной невозможности использования по отношению к праязыковым состояниям социолингвистических критериев, единственно релевантных для языковой индентификации диалектов» [6].

Ввиду отказа от датировок дивергенции и ограничения лишь задачей воссоздания картины родства наша работа оказывается по характеру гораздо ближе к попыткам количественной таксономии языков, предпринимаемых, например, А. Я. Шайкевичем [7].

Другое отличие нашей работы от исследований в рамках глоттохронологического направления состоит в критическом взгляде на возможность адекватного воссоздания картины генетической близости языков путем обращения к диагностическому списку в 100—200 слов (или, точнее, «смыслов»). Увеличение числа языковых единиц («признаков»), привлекаемых к статистическим подсчетам, у А. Я. Шайкевича («627 понятий из первых 14 разделов» известного словаря избранных синонимов основных индоевропейских языков К. Бака [8]) также не кажется нам достаточным, хотя полученная им схема [7, с. 335] отличается определенной убедительностью и дает автору возможность сделать некоторые интересные нетривиальные заключения (например, «о необходимости повысить ранг таксона „кельтские языки“» и о том, что «следует считать их надгруппой внутри индоевропейских языков», включающей в себя гаэльскую и бриттскую группы» [7, с. 334]) [2].

Вряд ли кто будет спорить с тем, что для точности «измерения» языкового родства большие массивы лексики и особенно полный праязыковой словарь в его отражении современными языками лучше, чем любые выборочные диагностические перечни слов. Однако большинство исследуемых лингвистами родственных групп и семей языков такими массивами и словарями не обеспечено, так что глоттохронологический метод в предложенном М. Сводешем виде является вынужденным и в общем нескрываемым компромиссом.

С предельной строгостью картина языкового родства лексикостатистическими методами может быть воссоздана на основе квантитативного анализа в с е й праязыковой лексики, сохраняющейся в языках-потомках. Применительно к славянским языкам это означает, что подсчеты следует вести на базе приблизительно 20 тыс. лексем, входящих в реконструированный словарный состав праславянского языка, включая в него и праславянские диалектизмы (Ф. П. Филин [9] указывает ориентировочную цифру 22 000 лексем; оценки Т. Лер-Сплавинского, Ф. Копечного и др. в 1 700, 2 000, 9 000 единиц, ср. [10—13], могут считаться безнадежно устаревшими). Таким образом, объем материала, привлекаемого для статистического анализа, должен возрасти по сравнению со списком Сводеша на два порядка, что несравненно повысит надежность результатов. Для современных средств вычисления этот объем, разумеется, отнюдь не является чрезмерным. Вопрос состоит лишь в том, чтобы исследователь располагал подобным материалом.

---

[2] Нужно заметить, что в схеме на с. 335 по недосмотру автора (или редактора) оказались не отраженными связи древнегреческого с латинским, литовского с древнеисландским, старославянского с древнеанглийским, древневерхненемецким и нидерландским, по своей силе входящие в тот же интервал показателей генетической близости, что и отмеченные в схеме связи латинского с готским и древневерхненемецким, а также между языками бриттской группы с языками гаэльской группы и языков балтийской группы с языками славянской группы. Любопытно, что в парных связях этой мощности фигурируют прежде всего мертвые индоевропейские языки, включенные А. Я. Шайкевичем в схему,— древнегреческий, латинский, древнеирландский, готский, древневерхненемецкий, древнеанглийский, старославянский; группы современных индоевропейских языков связываются между собою относительно большей близостью друг к другу предшествующих языковых состояний: изящное проникновение диахронии в ахроническую схему родства.

Начавшаяся в середине 70-х годов публикация словарей, которые ставят своей задачей реконструкцию славянского праязыкового лексического фонда в его полном объеме («Этимологический словарь славянских языков. Праславянский лексический фонд» под ред. О. Н. Трубачева в Москве и «Słownik prasłowiański» под ред. Ф. Славского в Кракове) позволяет оценить родство славянских языков попарно не на крайне ограниченном материале «универсального» диагностического базового списка в 165 (см. [14]) — 215 — 200 — 100 понятий, а с опорой на реконструированный словарь праязыка in corpore в его проекции на унифицированно представленные континуанты — лексиконы современных славянских языков.

Использование этимологических словарей, ориентированных на предельно полный охват и статейное перечисление праязыковой лексики, сохраненной в современных языках, принципиально важно для работы подобного рода: оно позволит если не избежать полностью, то свести к минимуму опасность, от которой не гарантированы (более того — на которую почти обречены) подсчеты по методу Сводеша. Имеется в виду бессилие глоттохронологии в ее классическом варианте перед явлениями лексического взаимопроникновения в языках-потомках. Разумеется, и современный этимологический анализ не обеспечивает стопроцентной уверенности в том, что перед нами именно только унаследованная праязыковая лексика, а не следствие взаимовлияний эпохи «после распада». Однако, во-первых, нынешнее состояние славянской этимологии следует оценивать очень высоко, и в большинстве своем послепраязыковые лексические перемещения из языка в язык выявляются достаточно надежно. Во-вторых, при оперировании лексическими массивами в несколько тысяч единиц возможна (и статистически вполне корректна) элиминация сомнительных случаев, к каковой мы и будем прибегать при необходимости.

Как выявить численную меру языкового родства? Самый простой, на первый взгляд, способ определить относительную генетическую близость языков друг к другу на основании лексикостатистических данных — сопоставить число общих для данных двух языков лексем, восходящих к восстановленному праязыковому словарному фонду, с аналогичными цифрами, характеризующими другие пары родственных языков: можно предполагать, что идиомы A и B, имеющие в своих словарях 1 000 общих для них праязыковых лексем, генетически менее близки, чем идиомы A и C, у которых число общих лексем, восходящих к праязыковой эпохе, допустим, 2 000.

Однако этот наиболее простой способ прямого сличения цифр одновременно наименее надежен и не пригоден для каких бы то ни было количественных сравнений лингвогенетической направленности. Он может работать только в том случае, когда все сравниваемые идиомы, находящиеся между собою в родстве, сохраняют по совершенно одинаковому числу единиц праязыкового лексического фонда. Такая ситуация крайне маловероятна для любых лингвистических общностей и существует как возможная лишь теоретически. Реально представленность праязыковой лексики в разных языках-потомках, естественно, различна: в верхнелужицком языке праславянских слов, по-видимому, всего около 5 000, тогда как в чешском или сербохорватском их не меньше 10 000, а в русском эта цифра, возможно, значительно превышает 12 000 лексических единиц. Понятно, что количество общих праславянских слов, сохраненных, скажем, польским и полабским языками, заведомо д о л ж н о б ы т ь меньшим, чем количество общих праславянских слов для польского и русского языков, поскольку объем доступного реконструкции праславянского пласта в по-

лабской лексике примерно в десять раз меньше праславянского словника русского языка, чем бы это ни вызывалось — деградацией полабского языка или просто скудостью наших сведений о нем. Однако столь же понятно и то, что польский и полабский характеризуются более тесными генетическими связями, чем польский и русский.

Следовательно, прямое сопоставление абсолютных цифр, описывающих лексические связи праязыкового характера в языках-потомках попарно, дает искаженную картину генетической близости языков, причем это искажение тем значительнее, чем заметнее расхождения в цифрах, отражающих объемы унаследованной или сохраненной поздними языками праязыковой лексики.

Устранение упомянутых искажений не представляет большой сложности. Очевидно, что показательны в данном случае числа лексических совпадений праязыкового характера для разных пар языков не в абсолютном выражении, а в отношении к количеству праязыковых лексем, сохраненных каждым из сравниваемых языков вообще, т. е. к объемам праязыковых словников обоих языков: для славянских языков A и B количество общих праславянских слов должно быть отнесено к произведению числа всех праславянских слов, обнаруженных в языке A, и числа всех праславянских слов, выявленных в языке B (в знаменателе вместо абсолютного числа всех праславянских лексем в данном языке можно брать и его долевое выражение — отношение объема праславянского лексического пласта в данном языке к объему всего реконструируемого этимологическим анализом праязыкового словаря в целом; выбор того или другого регулируется только удобством масштаба численного выражения конечного результата).

Но и при таком подсчете полученная картина соотношений будет весьма далека от ожидаемой на основании интуитивных представлений. Причиной этому — неравноценность разных конкретных лексических корреспонденций между языками, во-первых, и различия в пропорциях между разнородными связями для разных пар языков, во-вторых.

Ценность лексических изоглосс в установлении степени генетической близости языков путем статистических подсчетов находится в очевидной обратной зависимости от числа охватываемых ими языков: наиболее весомы в этом отношении сепаратные лексические связи между двумя языками, наименее — лексические изоглоссы, охватывающие все исследуемые родственные языки. В соответствие последовательности изоглоссных классов, выделяемых в зависимости от числа связываемых идиомов, должна быть поставлена шкала коэффициентов, позволяющих устранить неравнозначимость показателей близости. Пробные подсчеты показали, что при сравнительно небольшом числе сопоставляемых языков удовлетворительно работает шкала коэффициентов, представляющая собою последовательность десятичных логарифмов чисел, определяющих класс изоглосс (т. е. количество охватываемых ими идиомов), но расположенных по отношению к числам, выражающим лексические связи, в обратном порядке: число эксклюзивных лексических связей между двумя данными идиомами умножается на коэффициент $\lg n$; число изоглосс, охватывающих три идиома, — на $\lg (n - 1)$; четыре — на $\lg (n - 2)$ и т. д., до коэффициента $\lg 2$, на который умножается число связей, охватывающих все n идиомов.

Взвешенные с помощью указанной шкалы коэффициентов величины суммируются, полученная сумма делится на произведение чисел, выражающих объемы унаследованного праязыкового словаря отдельно в языках A и B. Результат и служит показателем генетической близости этих языков. Таким образом, индекс генетической близости для пары языков

(идиомов) определяется по формуле:

$$G(\mathrm{A}, \mathrm{B}) = \frac{\sum_{2}^{n} \lg (n + 2 - i) \cdot V(\mathrm{A}, \mathrm{B})_i}{H(\mathrm{A}) \cdot H(\mathrm{B})},$$

где G (лат. *gentilitas* «родство») — показатель генетической близости, «степени родства»; A и B — два данных идиома из всей совокупности исследуемых идиомов; n — общее число сравниваемых идиомов; V (лат. *verbum* «слово») — число общих для двух данных идиомов слов, входящих в изоглоссный тип данного класса; H (лат. *hereditas* «наследство») — число всех восстанавливаемых (или привлекаемых к анализу) единиц праязыкового лексического фонда, отмеченных в данном идиоме; i — класс изоглоссы (число охватываемых ею идиомов).

Вычисление уровня родства каждой пары идиомов по предложенной формуле несложно, однако нуждается в весьма трудоемком «нулевом цикле» работ.

Если подсчеты ведутся вручную, то для их облегчения разумно использовать рабочее понятие типовой изоглоссной конфигурации. Под нею здесь понимается обобщенная характеристика изоглоссы, освобожденная от сведений о конкретном очертании ареала и представляющая собою лишь перечисление идиомов, в которых отмечены континуанты заголовочной праформы (составляющей элемент словника привлекаемого к анализу праязыкового лексикона). Необходимость этого рабочего понятия вызывается тем, что в зависимости от мощности типа, т. е. числа охватываемых им идиомов, в формуле, которую мы предлагаем, разным изоглоссным объединениям приписываются разные коэффициенты. Предварительная разметка используемого словаря — отнесение списков межъязыковых («межидиомных») соответствий справа от каждой заголовочной праформы к той или иной типовой изоглоссной конфигурации — и требует наибольших затрат труда.

Количество типовых изоглоссных конфигураций, которые обобщают конкретные изоглоссы с идентичным распределением по изучаемым идиомам, зависит от количества (n) привлекаемых к сравнению идиомов и выражается числом 2^n (вернее, $2^n - 1$, так как одна из типовых изоглоссных конфигураций представлена прочерками против всех включаемых в анализ идиомов и является «пустой»). Таким образом, если сравниваются четырнадцать славянских языков (болгарский, македонский, сербохорватский, словенский, чешский, словацкий, верхнелужицкий, нижнелужицкий, полабский, польский, кашубско-словинский, русский, белорусский и украинский), то число возможных типовых изоглоссных конфигураций будет равным $2^{14} - 1 = 16\ 383$. Разумеется, далеко не все из теоретически возможных типовых изоглоссных конфигураций окажутся реально воплотимыми в имеющихся реконструкциях; более того, реальных воплощений не будет иметь, вероятно, подавляющая часть исчисленных изоглоссных типов. Тем не менее их количество слишком велико, чтобы с вычислениями на уровне отдельных славянских языков, при сравнении их друг с другом, можно было справиться вручную. Подобную работу можно поручить только ЭВМ. Поэтому в исследовании предварительного характера, каковым является настоящая статья, целесообразно в качестве пробного варианта остановиться на выяснении генетической близости меньшего числа идиомов — не отдельных славянских языков, а их группировок. Снижение числа идиомов (за счет не сокращения, а «обобще-

ния», «укрупнения» их) до 6 или 7, что вполне приемлемо для экспериментальной проверки формулы, дает $2^6 - 1 = 63$ или $2^7 - 1 = 127$ изоглоссных типов.

Разметив словарь, т. е. проставив против каждой словарной статьи номер типовой конфигурации, которой соответствует список континуантов праформы в заголовке статьи, нетрудно подсчитать, каким количеством праязыковых реконструкций («изолекс») представлен каждый из исчисленных изоглоссных типов. Имея таблицу изоглоссных типов с данными о количестве конкретных изолекс, которым отражена каждая типовая конфигурация, далее путем суммирования несложно получить все необходимые для подстановки в предложенную выше формулу цифры.

В настоящей работе славянские языки были разбиты на шесть групп, отношения особо тесного родства внутри которых бесспорны и которые соответствуют наиболее вероятному, по теперешним представлениям, диалектному членению позднепраславянского языка: восточная южнославянская (болгарско-македонская), западная южнославянская (сербохорватско-словенская), чешско-словацкая, лужицкая, лехитская и восточнославянская (русская) языковые группы (ср. [15—18]). Учитывая несколько особое положение полабского языка в составе лехитских языков, к которым его обычно относят (ср. хотя бы выделение четырех западнославянских языковых групп, с обособлением полабских наречий, у А. М. Селищева [19]), было сочтено допустимым выделить его в индивидуальную «группу». Это даст возможность уже в данной работе проверить тесноту его генетических связей с другими западнославянскими группами.

Упомянутая выше таблица типовых изоглоссных конфигураций при «семичленном» наборе идиомов выглядит следующим образом [латинскими буквами обозначены исследуемые группы славянских языков: (a) — болгарско-македонская, (b) — сербохорватско-словенская, (c) — чешско-словацкая, (d) — лужицкая, (e) — полабская, (f) — лехитская (польско-кашубско-словинская), (g) — восточнославянская (русско-белорусско-украинская)]:

Таблица 1

№	Типовая изоглоссная конфигурация							Число охватываемых групп (класс конфигурации)	Число ее лексических репрезентаций
	a	b	c	d	e	f	g		
1	+	+	+	+	+	+	+	7	243
2	+	+	+	+	+	+	−	6	1
3	+	+	+	+	+	−	+	6	4
4	+	+	+	+	+	−	−	5	0
5	+	+	+	+	−	+	+	6	601
6	+	+	+	+	−	+	−	5	24
. . . . .									
47	+	−	+	−	−	−	+	3	31
48	+	−	+	−	−	−	−	2	19
49	+	−	−	+	+	+	+	5	0
50	+	−	−	+	+	+	−	4	0
. . . . .									
125	−	−	−	−	−	+	+	2	83
126	−	−	−	−	−	+	−	1	66
127	−	−	−	−	−	−	+	1	483
(128)	−	−	−	−	−	−	−	(0)	(—)

Издание обоих упомянутых праславянских словарей — и московского, и краковского — находится в разгаре, и до полного их завершения должно пройти еще довольно много времени. На более продвинутом этапе находится работа московского коллектива авторов, доведшего публикацию своего словаря уже до 14-го выпуска (начата буква *L*). Материал «Этимологического словаря славянских языков» под ред. О. Н. Трубачева (далее сокращенно: ЭССЯ) и был предпочтен нами для предварительного лексикостатистического обследования.

В первых двенадцати выпусках ЭССЯ содержится 5 966 позиций, что почти в тридцать и шестьдесят раз соответственно превосходит объемы вариантов «базового словаря» Сводеша и почти в десять раз — объем списка понятий для таксономии европейских языков у Шайкевича. Можно надеяться, что пробные подсчеты по нашей формуле, сделанные на этом материале, будут обладать достаточной убедительностью, т. е. полученная картина родства будет близка к той, какая может быть выявлена на всем материале праславянских лексиконов, когда они завершатся. Надежду на это внушает то обстоятельство, что, как показывают статистические прикидки, уже с 7—8 выпусков ЭССЯ (около 4 000 праславянских лексем) примерные пропорции в силе родственных связей между отдельными упомянутыми семью группами славянских языков стабилизируются и в последующем, с возрастанием объема лексического материала, меняются в целом сравнительно незначительно.

Основной объект «нулевого цикла» работы — изоглосса в обобщенном представлении, т. е. фактически позиция в ЭССЯ — список рефлексов праславянской праформы, содержащийся в каждой словарной статье. В указанном словаре основанием межславянского сравнения и комплектования материала в пределах статьи является не корень, как это большей частью практикуется в этимологических лексиконах, а словообразовательная структура [20, с. 9; 25—27], и предметом рассмотрения становится не этимологическое гнездо, а праславянская лексема, т. е. межславянские лексические корреспонденции с полным формальным тождеством на уровне праформы (ср.: **glupiti*: словен. *glupíti*, чеш. редк. *hloupiti*, польск. диал. *głupić*, русск. *глупи́ть*; **košanica*: серб-хорв. *ко̀шаница*, русск. диал. *ко́шаница*, укр. *кошани́ця*, блр. диал. *кошаны́ця*; и под.). Однако принцип оперирования в границах одной словарной статьи только цельнолексемными соответствиями вряд ли может быть проведен с неукоснительной последовательностью, и составители ЭССЯ весьма далеки от лексикографического ригоризма, допуская иногда в наборе соответствий отступления от требования формального тождества корреспондирующих единиц и «максимально расчлененной подачи словника» [20, с. 9]. Наиболее обычный случай — включение в число корреспонденций заголовочной лексемы форм, находящихся на следующих ступенях деривации при незасвидетельствованности в данном языке прямого рефлекса праформы: укр. *дарбенний* в списке континуантов праслав. **darъba*, серб.-хорв. *дȅвуша*, *дјȅвуша*, русск. *дéвушка* как продолжения праслав. **dě̀vьxa*, словен. *gibežljȗv* в статье **gybežь* и т. п.; такие дополнения к спискам корреспонденций, свидетельствующие о былом существовании первообразных структур в идиомах, предшествующих современным языкам, вводятся в статью оборотом «Ср. сюда же...». Изредка встречаются и обратные случаи — когда статья посвящена анализу суффиксального образования, а «сюда же» подключается единичная нераспространенная форма, в самостоятельную словарную статью составителями не выделяемая (русск. диал. *гуть* в статье **gǫlьnъjь*, н.-луж. *gjarś* в статье **gъrtanъ*/**gъrtanь*). Морфологи-

ческие варианты могут в одних случаях объединяться в одну статью (как в **čemerъ/*čemera*, **xodъ/*xoda*, **gala/*galo*, **korьpa/*korьpo/*korьpъ* и т. д.), в других же — составлять разные словарные позиции, и если пары лексем **dara* — **darъ*, **doba* — **dobo* демонстрируют различия, восходящие еще к индоевропейскому состоянию, что отмечается в этимологической разработке, то разнесение лексических пар **ablъko* — **ablъkъ*, **brězga* I — **brězgъ* I, **brězga* II — **brězgъ* II, **buka* — **bukъ* II и др. в разные статьи подобным образом эксплицитно в тексте словаря не мотивируется. То же можно сказать и о морфологическом варьировании глаголов типа **čepati* — **čepiti*, **xlestati* — **xlestiti*, **děti* — **děvati* (члены пар составляют разные статьи), с одной стороны, и **gorniti* I (с включением серб.-хорв. *grȁnati* в ту же статью), **xarati* (с включением укр. *хáрити*), **dobyti/*dobyvati* и под., с другой. Могут объединяться в одну статью и словообразовательные омонимы, отмеченные в разных языках, ср., например, **agodьnica*, объединяющее, по-видимому, производные от **agoda* со словообразовательно вторичными случаями, непосредственно связанными с **agodica*; **gospodьnъ(jь)*, в котором слились дериваты от **gospodь* и **gospoda*).

Все это показывает, что «набор» лексических изоглосс, если последние исчислять, исходя из словника ЭССЯ, является до некоторой степени условностью. Другие чисто лексикографические решения из соображений удобства расположения материала в приведенных и аналогичных им случаях привели бы к констатации отличных изоглосс, часто иной мощности, что небезразлично для результатов нашей работы, поскольку разным по классу изоглоссным конфигурациям приписываются в нашей формуле разные «весовые» коэффициенты.

Состав подсчитываемых единиц, а вместе с ним (правда, едва ли сколько-нибудь значительно при таком объеме материала) и конечная картина родства зависят не только от способа подачи этого материала в словаре, но и от выбираемых этимологических решений. В статье **baxorъ* объединены примеры, часть которых может трактоваться как «своего рода экспрессивное имя деятеля **ba-x-orъ* от основы **ba-*», в то время как другие суть образования **bax-orъ* от глагола **baxati* с семантикой «внутренности, кишки», «колбаса», «что-либо толстое, круглое». Разграничить **baxorъ* I и **baxorъ* II до конца мешает серб.-хорв. *бáор*, совмещающее обе семантические линии. В случае признания этимологической связи глагола **kaniti* (при преобладании в южнославянских рефлексах значений «приглашать», «предлагать», «намереваться») с гнездом **kopъ*, **konati* «становится ясным отнесение сюда... русск. диал. *кáнить* „пятнать (при игре в пятнашки)“». Иные этимологические версии изменили бы состав рефлексов и изоглоссных конфигураций.

Наконец, нельзя быть до конца уверенным в полной исчерпанности списков рефлексов по языкам, в них могут оказаться лакуны, также влияющие на отнесение данного перечня соответствий к той или другой типовой изоглоссной конфигурации. Так, на наш взгляд, могут быть дополнены списки рефлексов в статьях **babьskъjь* (нет русск. *бáбский*, хотя тут же приведено укр. *бáбський* с русским переводом «бабий, б а б с к и й»), **bolgъ(jь)* (не учтен русский топоним *Бологóе*), **bъčelь(jь)* (польск. *pczeli*), **ěskravъjь* (укр. *яскрáвий*, если это не заимствование из польского), **gordьcь* (топоним *Gardiss* в Полабье), **govьnęnъ(jь)* (русск. *говняный*), **govьno* (блр. *гаўно́*), **guščerъ* (русск. фолькл. *ящер- гýщер*), **xarobura* (русск. диал. *хараборы* с фонетическими вариантами, фамилии вроде *Харабаров*, возможно, *Харабурда* и проч., относительно которых,

правда, могут возникать и тюркские ассоциации), **kolo* (полаб. *t'ülü*), **konь* (полаб. *t'ün*).

Однако к указанным лексикографическим и этимологическим решениям, принятым в ЭССЯ, мы отнеслись «некритично», как если бы они были единственно возможными. Об этих реальных или мнимых непоследовательностях, которые отнюдь не следует считать недостатками словаря, мы нашли необходимым упомянуть лишь затем, чтобы, во-первых, лишний раз подчеркнуть сложность всякой словарной работы, а во-вторых, показать важность привлечения массового материала для квантитативных оценок лингвистического родства. Потребность выбора между несколькими возможностями — ситуация более чем обычная в лексикографии, и только при значительном объеме материала спорные случаи могут стать статистически незначимыми, погашаясь мощно обнаруживающимися в словарном массиве тенденциями.

Тем не менее мы стремились к тому, чтобы лексический материал, используемый нами, был предельно чист в генетическом отношении. Для этого была осуществлена процедура сегрегационного анализа. Цель его — освобождение материала для статистических построений от результатов возможных послепраславянских взаимовлияний, от книжной лексики, потенциально поздних образований, сомнительных параллелей и т. п., словом, снятие всего, что может искажать синхронную картину генетической близости. Единственным и достаточным поводом для констатации небезупречности примера или реконструкции служили явно выраженные сомнения самих составителей словаря.

Прежде всего, из списков соответствий «вычеркивались» старославянские и церковнославянские лексические примеры. Старославянский, будучи в целом «болгарско-македонским» по происхождению, во-первых, не является продолжением живой народной языковой традиции; во-вторых, он не является генетически однородным, включая в свой лексический состав моравизмы и паннонизмы, не всегда, как можно судить, однозначно диагностируемые. Церковнославянская лексика любого извода большей частью ориентирована на восточный южнославянский ареал и, отличаясь преимущественно книжными путями распространения, также вносит аберрации в общую генетическую характеристику поздних языков. Разумеется, элиминация этих примеров при оперировании не отдельными языками, а их группами релевантна только тогда, когда болгарско-македонская или, скажем, восточнославянская группа, помимо старославянской или, соответственно, русскоцерковнославянской лексемы, другими примерами — из живых народных языков — в данном изоглоссном списке не отражена.

Далее, не включались в подсчет целиком те словарные статьи, в которых были найдены текстуально выраженные авторские сомнения в праславянской древности заголовочной лексемы (конкретные их формулировки очень многообразны и передают, надо полагать, разную степень сомнительности: «праславянская древность сомнительна», «древность проблематична», «момент хронологии неясен», «относительно позднее образование», «возможно, позднее», «м. б. местным новообразованием», «признаки вторичного образования» и т. п.). Изымались из списков соответствий слова живых языков, относительно которых у авторов ЭССЯ возникают подозрения в позднем, послепраславянском параллельном образовании или их книжной природе (например, зап.-укр. *багри́й*, русск. *бесча́дный*, чеш., слвц. *hrachor*, русск. *клевета́* и анал.). Некоторые слова западнославянских языков с префиксом *z-* двусмысленны в плане реконструкции:

с равным вероятием префикс может быть возведен и к *jьz-, и к *sъ- (см., например, *jьzbiti, *jьzbosti, *jьzdati и т. д.). При недостаточности семантических аргументов такие примеры «вычеркивались» из перечней соответствий, хотя, может быть, лучшим решением было бы опущение всей вокабулы целиком. Нерешительность составителей по отношению к формам из категории «сюда же» (о них см. выше), выражавшаяся в парентезе «возможно», «по-видимому» (ср., например, русск. диал. *баларýжина* — из *baro-lužina?, укр. *барúло* и др. в статье *bara; чеш. редк. *bluňkati* в статье *bl'uxati/*bl'ukati и под.), толковалась скорее в пользу сегрегации примеров.

В результате сегрегационных процедур из 5 966 праславянских реконструкций, составляющих словник первых двенадцати выпусков ЭССЯ (от *a до *kroma/*kromъ), для дальнейшего статистического анализа было оставлено 5 697 лексем, или 95,5% словника [3]. Это — больше четверти всего гипотетического объема праславянского лексикона, включая и его локальные, диалектные элементы.

Подсчет реконструированных лексем, приходящихся на каждую из семи постулированных групп славянских языков, в историческом отношении предположительно отождествляемых с позднепраславянскими диалектными областями, дает (по первым двенадцати выпускам ЭССЯ) следующие цифры:

		H	*H*(%)
(a)	болгарско-македонская	2686	47,15
(b)	сербохорватско-словенская	4073	71,49
(c)	чешско-словацкая	3685	64,68
(d)	лужицкая	1659	29,12
(e)	полабская	384	6,74
(f)	лехитская (без полабского)	2709	47,55
(g)	восточнославянская	4321	75,85

При включении полабского в лехитскую группу последняя характеризуется числом 2 753 лексемы (или 48,32% от общего количества отобранных к анализу позиций ЭССЯ). Если предположить, что проценты достаточно правдоподобно отражают реальную наследуемость праязыковой лексики разными группами, и экстраполировать их на гипотетический объем праславянского словаря в 20—22 тысячи лексем, то нетрудно выяснить, что всего отдельные группы сохраняют: (a) — 9,5—10 тыс., (b) — 14—15,5 тыс., (c) — 13—14 тыс., (d) — 6—6,5 тыс., (e) — 1,5 тыс., (f) — 9,5—10,5 тыс., (g) — 15—16,5 тыс. единиц праславянского лексического фонда. В свете этих оценок пассажи вроде «Слова, унаследованные из общеславянского (= праславянского.— *Ж. А.*) языка, не составляют и одного процента лексики современных языков» [21, с. 243] воспринимаются, пожалуй, как слишком решительная литота, если, конечно, исключить из поля зрения циклопические арсеналы узкотерминологической лексики вроде химической и фармакологической номенклатуры (ср. еще [22]).

---

[3] Любопытно отметить, что в первом выпуске словаря было «вычеркнуто» 14,4% позиций, во втором — уже 7,5%, в третьем — 3,8%, в последующих выпусках — в среднем 3,1%. Эта уменьшающаяся последовательность доли сегрегируемых вокабул в определенной мере отражает стабилизацию с ходом издания критериев отбора лексики для помещения ее в «Праславянский лексический фонд» самими составителями: напомним, что при сегрегации вокабул мы руководствовались текстуально выраженными сомнениями авторов относительно праславянской древности предлагаемых ими реконструкций.

Итоги статистических наблюдений над генетической близостью отдельных групп славянских языков (попарно) по данным лексики сведены в табл. 2. Последняя ее колонка содержит цифры, являющиеся конечной целью данного предварительного исследования, т. е. словный индекс родства.

Таблица 2

1	2	3	4	5	6	7	8	9	10
a ~ b	166	358	469	560	643	243	2439	1463,77	0,1338
a ~ c	19	143	406	552	642	243	2005	1122,91	0,1134
a ~ d	1	10	47	140	609	243	1050	489,48	0,1098
a ~ e	—	3	3	4	43	243	296	100,94	0,0979
a ~ f	9	38	153	488	640	243	1571	816,43	0,1122
a ~ g	50	276	461	548	643	243	2221	1289,12	0,1111
b ~ c	132	411	685	745	662	243	2878	1747,71	0,1164
b ~ d	14	64	134	321	629	243	1405	721,82	0,1068
b ~ e	2	10	11	20	63	243	349	132,41	0,0847
b ~ f	36	151	398	680	660	243	2168	1223,56	0,1109
b ~ g	286	538	730	742	663	243	3202	2006,80	0,1140
c ~ d	23	63	179	322	628	243	1458	750,22	0,1244
c ~ e	3	8	7	23	62	243	346	130,23	0,0920
c ~ f	62	210	408	678	659	243	2260	1296,76	0,1299
c ~ g	178	405	715	740	662	243	2943	1799,87	0,1130
d ~ e	2	2	7	10	29	243	293	101,15	0,1588
d ~ f	11	26	127	254	626	243	1287	643,05	0,1431
d ~ g	18	63	169	317	629	243	1439	746,47	0,1041
e ~ f	3	4	5	25	60	243	340	125,97	0,1211
e ~ g	4	5	9	26	63	243	350	132,42	0,0798
f ~ g	83	221	439	675	660	243	2321	1343,40	0,1148

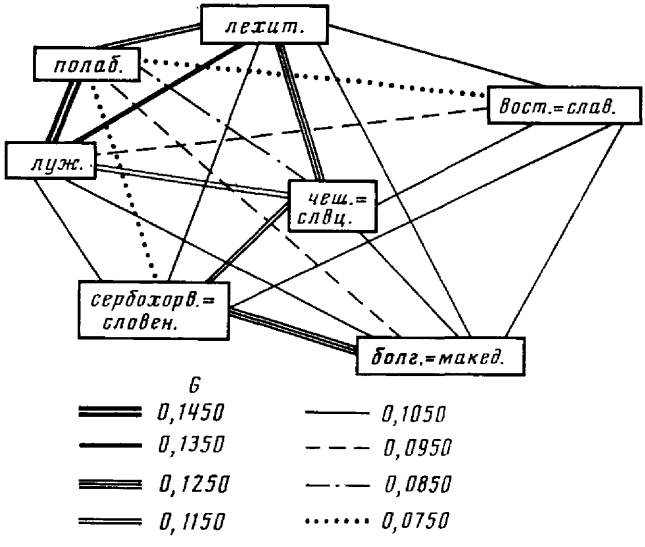
Пояснения к табл.: 1 — сравниваемые пары идиомов; 2—7 — число общих для них праславянских лексем, входящих в изоглоссные конфигурации, охватывающие от двух (эксклюзивные лексические связи) до семи (общеславянские слова) языковых групп; 8 — общее количество лексических корреспонденций между двумя данными группами; 9 — сумма связей, взвешенных с помощью шкалы поправочных коэффициентов (числитель формулы генетической близости); 10 — индекс генетической близости по данным лексикостатистики (G).

Следует подчеркнуть, что конкретное численное выражение индекса родства носит относительный характер, он показателен только в рамках предпринятого исследования и не сопоставим с цифрами, которые аналогичным способом могут быть получены для иных наборов идиомов и на ином по объему материале: существует сильная зависимость индекса генетической близости от объема праязыкового лексикона и длины шкалы поправочных коэффициентов, которая в свою очередь зависит от числа выделяемых языков или языковых групп.

Выявленные нашим исследованием соотношения в генетической близости между разными группами славянских языков можно наглядно передать схемой в виде графа, узлы которого символизируют изучаемые идиомы, а мощностью ребер изображается сила лексикостатистических связей праязыкового характера.

Подведем предварительные итоги работы.

Если исходить из абсолютных количественных данных о лексических связях между семью группами славянских языков (они содержатся в колонках 2—7 и 8 табл. 2), то выясняется, что наибольшее число лексических (цельнолексемных по подавляющему преимуществу) соответствий связывает сербохорватско-словенскую и восточнославянскую группы: 3 202



лексемы (или 56,21% от всего числа просчитанных позиций первых 12 выпусков ЭССЯ), из них 286 лексем (5,02%) представляют эксклюзивные связи этих групп — из сепаратных связей также максимум. Очень представительны по числу изолекс связи чешско-словацкой группы с сербохорватско-словенской и восточнославянской. Они характеризуются цифрами соответственно 50,52% и 51,66% от общего объема праславянского словаря, т. е. каждая из этих пар идиомов фигурирует более чем в половине всех позиций обсчитанного лексикона.

Наименьшие абсолютные цифры характеризуют лексические связи полабского языка с другими идиомами: число полабско-«иноидиомных» параллелей ни в одном случае не превышает 6,14% от объема праславянского словаря.

Однако эти данные никоим образом не отражают уровня генетической близости между отдельными языковыми группами. Абсолютное число лексических связей праязыкового характера между родственными идиомами связано с объемом праязыковой лексики, сохранившейся в каждом идиоме: чем пространнее праславянский словник какой-либо группы современных славянских языков, тем выше абсолютные показатели ее лексических корреспонденций с другими группами. Обращение к цифрам, полученным с помощью формулы генетической близости, рисует совершенно иную картину.

Из всех парных связей, которые устанавливаются между семью группами славянских языков с помощью предложенного нами метода, наиболее мощной оказывается статистическая связь между лужицкой и полабской группами (индекс генетической близости 0,1588). Следующая по силе связь — между лужицкой и «собственно» лехитской группами (0,1431). К сильным связям следует отнести родственные узы между двумя южнославянскими (болгарско-македонской и сербохорватско-словенской) и между чешско-словацкой и лехитской группами.

Самые низкие индексы родства отмечаются в парных связях полабского с восточнославянской (0,0798) и сербохорватско-словенской (0,0847) группами, чуть сильнее его родственные отношения с чешско-словацкой и болгарско-македонской группами. Статистические показатели родства

всех остальных пар идиомов составляют ряд средних по значению индексов в промежутке от 0,1041 до 0,1244.

Заслуживают внимания некоторые обстоятельства.

Выделение полабского языка в данном исследовании в самостоятельную группу дало возможность оценить его исконные лексические связи с другими группами славянских языков специально. Выясняется, что по индексу генетической близости, вычисляемому на базе лексикостатистики, более тесные узы связывают полабский не с лехитской (польско-кашубско-словинской) группой, как можно было априорно ожидать, а с лужицкой. Связи полабского с «собственно» лехитской группой по своей силе относятся к средним и являются даже несколько более низкими, чем статистические связи чешско-словацкой группы с лужицкой. Однако поскольку сохранность праславянской лексики в полабском языке чрезвычайно низка по сравнению с другими послепраславянскими идиомами, статистические суждения о генетических отношениях полабского не столь надежны, как в случаях с иными славянскими языками. Уточнить их может увеличение материала с дальнейшим изданием ЭССЯ.

Низкий индекс родства полабской и чешско-словацкой групп (0,0920) подтверждает сомнения в генетическом единстве западнославянских языков. Намного существеннее оказываются лексические связи чешско-словацкой группы с сербохорватско-словенской, которые подкрепляются и рядом общих фонетических явлений (южнославянско-чешско-славацкое изменение сочетания *tort*, южнославянско-(средне)словацкое изменение сочетания *ort-*, словенско-словацкая изоглосса утраты результатов второй палатализации в именном словоизменении...).

Опираясь на тезис О. Н. Трубачева о вторичной окцидентализации серболужицких языков [23—25], можно было ожидать сравнительно высоких показателей генетической близости лужицкой группы с южнославянскими, тем более что мнение О. Н. Трубачева основывается именно на лексических данных. Индексы родства лужицкой группы с болгарско-македонской и сербохорватско-словенской несколько выше, чем показатель близости с восточнославянской группой, однако этого, на наш взгляд, недостаточно, чтобы тезис О. Н. Трубачева получил весомое подтверждение. Нельзя, впрочем, считать, что он нашими данными опровергается. Упрочить или ослабить его могут лексикостатистические наблюдения, аналогичные настоящему, но осуществленные с более дробным членением привлеченных к исследованию идиомов — не групп, а отдельных славянских языков. От таких исследований можно ждать и уточнения суждений, например, о большей приближенности нижнелужицкого к польскому в противовес «чешской» ориентации верхнелужицкого ([26]; ср. [27—29]). Нужно лишь иметь в виду, что подобные штудии опираются на недифференцированное представление ранне- и позднепраславянских лексических и словообразовательных явлений. На значительное улучшение дел в изучении предполагаемых взаимных «переориентаций» диалектов праславянского языка в ходе его развития можно надеяться только тогда, когда мы будем увереннее разграничивать раннее и позднее в самой праславянской лексике.

Но и при существующем положении вещей обращение к полным праславянским словникам отдельных языков-потомков может конкретизировать наши понятия об отношениях взаимного родства внутри славянской языковой семьи. Немаловажную роль здесь должен сыграть строгий квантитативный анализ имеющегося лексического материала с помощью современных технических средств.

ЛИТЕРАТУРА

1. *Сводеш М.* Лексикостатистическое датирование доисторических этнических контактов (на материале племен эскимосов и североамериканских индейцев) // Новое в лингвистике. Вып. I. М., 1960.
2. *Сводеш М.* К вопросу о повышении точности в лексикостатистическом датировании // Новое в лингвистике. Вып. I. М., 1960.
3. *Gudshinsky S.* The ABC's of lexicostatistics (glottochronology) // Word. 1956. V. 12.
4. *Čejka M., Lamprecht A.* K otázce vzniku a diferenciace slovanských jazyků // Sborník prací Filosofické fakulty Brněnské university. 1963. A 11. S. 17 (отд. отт.).
5. *Арапов М. В., Херц М. М.* Математические методы в исторической лингвистике. М., 1974.
6. *Климов Г. А.* Об ареальной конфигурации протоиндоевропейского в свете данных картвельских языков // ВДИ. 1986. № 3. С. 151.
7. *Шайкевич А. Я.* Гипотезы о естественных классах и возможность количественной таксономии в лингвистике // Гипотеза в современной лингвистике. М., 1980.
8. *Buck C. D.* Dictionary of selected synonyms in the principal Indo-European languages. Chicago, 1949.
9. *Филин Ф. П.* Историческая лексикология русского языка: Проспект. М., 1984. С. 25.
10. *Lehr-Spławiński T., Kuraszkiewicz W., Sławski F.* Przegląd i charakterystyka języków słowiańskich. Warszawa, 1954. S. 24.
11. *Лер-Сплавинский Т.* Польский язык. М., 1954. С. 64.
12. *Ondruš Š.* Praslovanský základ slovenčiny v slovnej zásobe. // Stud. Acad. Slov., 1976. 5. S. 299—300.
13. Историческая типология славянских языков. Фонетика, словообразование, лексика и фразеология. Киев, 1986. С. 200.
14. *Лернер К. Б.* Статистические методы в историческом языкознании: Автореф. дис. ... канд. филол. наук. Тбилиси. 1973. С. 12.
15. *Furdal A.* Rozpad języka prasłowiańskiego w świetle rozwoju głosowego. Wrocław, 1961.
16. *Birnbaum H.* The dialects of Common Slavic // Ancient Indo-European dialects. Berkeley — Los Angeles, 1966.
17. [*Иванов В. В.*] Диалектные членения славянской языковой общности и единство древнего славянского языкового мира (в связи с проблемой этнического самосознания) // Развитие этнического самосознания славянских народов в эпоху раннего средневековья. М., 1982.
18. *Толстой Н. И.* Из истории славистики. Опыт карты праславянских диалектов Д. П. Джуровича. 1913 г. // Зборник Матице српске за филологију и лингвистику. 1984—1985. XXVII—XXVIII.
19. *Селищев А. М.* Славянское языкознание. Т. I. Западнославянские языки. М., 1941.
20. [Трубачев О. Н.] Этимологический словарь славянских языков (праславянский лексический фонд). Проспект. Пробные статьи. М., 1963.
21. *Булахов М. Г., Жовтобрюх М. А., Кодухов В. И.* Восточнославянские языки. М., 1987.
22. *Супрун А. Е.* Лексическая типология славянских языков. Минск, 1983.
23. *Трубачев О. Н.* Ремесленная терминология в славянских языках (этимология и опыт групповой реконструкции). М., 1966. С. 391—392.
24. *Трубачев О. Н.* О составе праславянского словаря (проблемы и результаты) // Славянское языкознание. VI Международный съезд славистов (Прага, август 1968 г.): Докл. советской делегации. М., 1968. С. 373.
25. *Трубачев О. Н.* Языкознание и этногенез славян. Древние славяне по данным этимологии и ономастики // ВЯ. 1982. № 5. С. 6.
26. *Mucke E.* Historische und vergleichende Laut- und Formenlehre der niedersorbischen (niederlausitzisch-wendischen) Sprache. Leipzig, 1891. S. 3.
27. *Шустер-Шевц Х.* Язык лужицких сербов и его место в семье славянских языков // ВЯ. 1976. № 6.
28. *Taszycki W.* Stanowisko języka łużyckiego // Symbolae grammaticae in honorem J. Rozwadowski. II. Kraków, 1928.
29. *Stieber Z.* Stosunki pokrewieństwa języków łużyckich. Kraków, 1934.