

уменьш
ен
в
вс
V(A, B)
/упр.
р гит
индекс
инк, то увеличивает на формуло
-словен. пред. уменьше сл. ~ Зет. слоб. зр
— предлоги несколько иные условия
а ТК. применительно к М. су,
зислам. Крайне сим — к лекс
инстай то статистика доходит, и

А. Ф. Журавлев

ЛЕКСИКО-
СТАТИСТИЧЕСКОЕ
МОДЕЛИРОВАНИЕ
СИСТЕМЫ
СЛАВЯНСКОГО
ЯЗЫКОВОГО
РОДСТВА



Российская академия наук
Институт славяноведения и балканистики

А. Ф. Журавлев

**ЛЕКСИКОСТАТИСТИЧЕСКОЕ
МОДЕЛИРОВАНИЕ
СИСТЕМЫ
СЛАВЯНСКОГО
ЯЗЫКОВОГО
РОДСТВА**



ИЗДАТЕЛЬСТВО
ИНДРИК
Москва 1994

ББК 81.2 Сл.
Ж 91

*Издание осуществлено
при финансовой поддержке
Международного Фонда «Культурная инициатива»*

Журавлев А. Ф.

Ж 91 Лексикостатистическое моделирование системы славянского языкового родства. М.: Издательство «Индрик», 1994. — 254 с.
ISBN 5-85759-006-X

Монография ставит своей задачей установление меры генетической близости между отдельными славянскими языками по данным праславянской лексики, сохраняемой каждым из них, и выяснение возможностей реконструкции на этой базе целостной картины взаимоотношений между диалектами позднепраславянского языка. Исследование праславянских изолексов осуществляется в книге с применением статистических методов при учете типологических характеристик отдельных славянских языков и их лексических составов. Книга рассчитана на лингвистов — специалистов по общему и славянскому языкознанию.

ББК 81.2 Сл.

ISBN5-85759-006-X

© А. Ф. Журавлев, 1994

От автора

Предлагаемая монография ставит своей задачей установление меры генетической близости между отдельными славянскими языками по данным праславянской лексики, сохраняемой каждым из них, и выяснение возможностей реконструкции на этой базе целостной картины взаимоотношений между диалектами позднепраславянского языка. Исследование праславянских изолексов осуществляется в книге с применением статистических методов (предложена формула генетической близости языков, разработан способ лексикостатистического выявления феномена языковой конвергенции и др.).

Особенностью работы является обращение не к коротким выборочным диагностическим перечням лексики типа стословного списка понятий, применяемого в глоттохронологических исследованиях, и не к отдельным группам лексики, ограниченным по каким-либо тематическим или формальным (словообразовательным, грамматическим и т. д.) признакам, а опора на сплошное статистическое обследование праславянского словаря. При этом количественному анализу подвергся доступный к началу работы праславянский лексический материал в его максимальном объеме (увеличенном привлечением дополнительных источников).

Лексикостатистический анализ проводится здесь с принятием во внимание типологических характеристик отдельных славянских языков и их лексических составов. При сопоставлении картины родственных взаимоотношений славянских языков, полученной на основе квантитативного анализа изолекс, с аналогичной картиной, которая восстанавливается по данным сравнительной фонетики, учитываются принципиальные различия в механизмах эволюции фонетического и лексического уровней.

В монографии предпринимается попытка реконструкции системы славянского языкового родства с отказом от традиционных дендроидных схем, которые не отражают подлинной природы отношений близости между языками.

Автор испытывает чувство признательности коллегам, помогавшим ему в процессе работы над книгой консультациями, замечаниями и советами. С рукописью книги знакомились Ю. С. Азарх, Ж. Ж. Варбот, Л. П. Крысин, В. З. Санников, С. М. Толстая, Н. И. Толстой, О. Н. Трубачев, А. Я. Шайкевич, † Д. Н. Шмелев. Основные ее положения или отдельные части обсуждались, кроме того, с О. Г. Гецовой, Л. Л. Касаткиным, Р. Ф. Касаткиной, Л. В. Куркиной, Х. Поповской-Таборской, В. А. Успенским. Все их соображения были приняты автором с благодарным вниманием.

Совершенно неоценимую помощь оказала автору М. В. Ломковская, без усилий которой по составлению компьютерных программ для необходимых вычислений, работа не могла быть осуществлена. Программы, сделанные Марией Владимировной, составляют ее интеллектуальную собственность и в настоящей книге не описываются.

Глава 1

Механизмы эволюции словаря и роль лексики в установлении уровней языкового родства

1. Сравнительная значимость фонетических, морфологических и лексических критериев в глоттогенетических исследованиях

В сравнительно-историческом языкознании если не общепринятым, то очевидно преобладающим можно считать мнение о том, что в процедурах установления относительной близости между разными родственными языками или диалектами одного и того же языка лексические и близкие им словообразовательные данные являются менее значимыми по сравнению со свидетельствами фонетики и морфологии.

В. Порциг, например, формулирует это так: «Фонетические и морфологические соответствия более доказательны для установления связей между диалектами, чем соответствия в словарном составе» [Порциг, 1964, стр. 91].

Т. В. Гамкрелидзе и Вяч. Вс. Иванов, объясняя построение своей работы, предваряют изоглосный анализ индоевропейских языковых фактов замечанием: «Значимость фонологических, морфологических (грамматических) и лексических критериев для установления характера языковых явлений, общих для родственных диалектов, определяет последовательность, в которой дается рассмотрение изоглосс, объединяющих древние индоевропейские диалекты. Ниже будут последовательно рассмотрены (а) морфологические, (б) фонологические и (в) лексические явления, позволяющие считать их пересекающимися разноуров-

невыми изоглоссами, на основании которых объединяются и классифицируются отдельные диалектные группы в пределах индоевропейского языка и в период их обособленного развития вплоть до исторически засвидетельствованного периода» (разрядка и буквенная нумерация наша. — А. Ж.) [Гамкрелидзе — Иванов, 1984, т. I, стр. 373]; ср. также: [Иванов, 1990, стр. 94].

Не так уж редко можно столкнуться с крайним нигилизмом в вопросе использования данных словаря в генетико-классификационных построениях. Например: «...опыт конкретного изучения живых диалектов современных индоевропейских языков показывает, что лексические критерии не могут быть положены в основу генетической классификации диалектов (и развившихся из них родственных языков). Лексика легко подвергается заимствованиям, степень ее устойчивости часто зависит от внешних и случайных причин. Генезис диалектов наиболее надежным образом выявляется при опоре на данные сравнительно-исторической фонетики и морфологии, лишь дополненные отдельными и хорошо проверенными данными лексики» (разрядка наша. — А. Ж.) [Широков, 1988, стр. 46–47]. Процитированный автор не является лексикологом, и он попросту, по инерции, присоединяется к точке зрения, освященной давней и авторитетной традицией (ср.: «Морфология есть наиболее устойчивый элемент языка» [Мейе, 1938, стр. 427]; «Словарный состав — самый неустойчивый элемент в языке» [Мейе, 1954, стр. 34]). Однако приведенные выше суждения В. Порцига и Т. В. Гамкрелидзе и Вяч. Вс. Иванова, разделенные тремя десятилетиями, представляются тем более показательными, что и в монографии первого и в фундаментальном труде вторых большую часть текстуальной площади занимает анализ именно лексических данных.

В этих суждениях, как и в большинстве высказываний подобного рода, не отмечается, что эффективность фонетических (фонологических), грамматических (в первую очередь морфологических) и лексических критериев в установлении степеней близости между родственными идиомами, в их генетической классификации может быть различной в зависимости от классификационного ранга этих идиомов (языковые семьи, группы языков, отдельные языки одной группы или диалекты одного языка). Преимущество того или иного языкового уровня в генетической классификации с оглядкой на возраст сравниваемых идиомов как самостоятельных таксономических величин в общем просматривается недостаточно ясно. Тем не менее оценки роли данных, относящихся к разным ярусам языковой

системы, как отправного материала в генетико-классификационных построениях заслуживают внимания.

Так, по мнению В. Я. Порхомовского, одного из немногих, кто ставит эту проблему, сравнительная ценность лексических, морфологических и фонетических критериев в генетической классификации языков неодинакова и варьирует в зависимости от временной глубины, на освоение которой посягает данная классификация. Занимаясь разработкой вопросов, связанных с классификацией различных групп и семей африканских языков, он приходит к выводу о том, что «при классификации достаточно близких друг другу языков или языковых групп, для которых предполагается сравнительно небольшой период времени, отделяющего их от общего праязыкового состояния, данные грамматики имеют определяющее значение. Напротив, для далеко разошедшихся языков и языковых групп, т. е. в случае существенно более глубоких хронологических диапазонов от праязыка до современных или засвидетельствованных памятниками языков, грамматические критерии утрачивают свое значение, и предпочтение следует отдавать лексике (и фонетике)» [Порхомовский, 1982, стр. 219].

В подтверждение этого высказываются соображения о большей, по сравнению с лексическим уровнем, структурированности и жесткости морфологии, из-за чего, под воздействием каких-то причин (допустимы и импульсы извне данной семьи родственных языков) «первоначально незначительные изменения могут вызвать полную перестройку всей грамматической системы языка, причем эти процессы могут приводить к существенно разным результатам в разошедшихся группах и отдельных языках, так что лексика окажется единственным источником, дающим достаточное количество материала для сравнительного анализа» [там же].

Однако роль лексических фактов в установлении меры близости между идиомами, а тем самым и их классификации или (первоначально) диалектного членения, может быть весьма высокой и при обращении к эпохам, сравнительно близким нашей. Так, занявшись поиском диалектных различий в области фонетики, которые проливали бы свет на сложение диалектной картины чешского языка и относились бы при этом к «доисторическому» периоду его развития, Я. Белич вынужден констатировать: «...видно, что мы оказываемся здесь весьма часто на довольно зыбкой почве. В области морфологии нельзя говорить, видимо, ни об одном различительном явлении на территории чешского языка, которое было бы старше первых записей на чешском языке; это же относится и к синтаксическим различиям... Напротив, остатки не-

которых древних различий, вероятно, существуют в области словообразования как в наборе словообразовательных средств, так и в степени употребительности некоторых из них. ...на территории чешского языка имеется поразительно много явно древних различий в области лексики, по крайней мере отчасти относящихся и к доисторической эпохе» [Белич, 1972, стр. 34].

Г. А. Климов, ища сбалансированный подход и полагая наиболее надежным «комплексный критерий, охватывающий разные стороны языковой структуры», из отдельно рассматриваемых разноуровневых критериев для установления степени взаимной близости между родственными идиомами и выработки историко-генетической классификации языков склонен выделять лексический критерий как относительно более надежный (см.: [Общее языкознание, 1973, стр. 18], со ссылкой на: [Сравнительная грамматика германских языков, т. I, 1962, стр. 36–37, 48 и след.]).

Вяч. Вс. Иванов, который защищает, как было показано выше, иную точку зрения, приводит в качестве иллюстрации ситуацию с японскими, китайскими и тибетскими числительными и отмечает возможность заимствования из одного языка в другой целых лексических групп¹ (фонетическое развитие тибетского языка привело к тому, что современные тибетские формы числительных оказываются ближе к японским, заимствованным из родственного тибетскому китайского языка, чем к древнетибетским): «...если бы древнетибетские формы не были известны, то прямое сравнение современных японских и тибетских числительных могло бы привести к ошибочным выводам относительно генеалогической классификации языков» [Иванов, 1990а, стр. 94]. Впрочем, здесь – неэксплицитно – речь скорее идет о «пороговых» проблемах, связанных с установлением самого факта генетического единства (являются ли данные два языка родственными, входящими в одну языковую семью или нет), то есть о том, что Г. А. Климов определяет как «синхронный аспект компаративистики» [Климов, 1990, стр. 20], чем собственно о местах, занимаемых данными языками в схеме, отражающей степень близости заведомо или «установлено» родственных языков.

¹ Ср.: «Устойчивость числительных объясняется, вероятно, такими специфическими причинами, как развитие экономики и обмена, известное индоевропейским народам с очень давнего времени, а не «естественными» или универсальными мотивами, общими для многих языков. Бывает, что числительные заимствуются из другого языка. Иногда даже в целях удобства или по иным причинам целые группы числительных могут заменяться другими группами числительных» [Бенвенист, 1963, стр. 40]. Ср. по этому поводу: [Леман, 1991, стр. 18–19].

На наш взгляд, проблему большей или меньшей эффективности использования различных — фонетических, грамматических, лексических — данных в установлении степеней языкового родства и при построении генеалогических классификаций внутри выделенных языковых семей вряд ли целесообразно пытаться решать в общем виде. Такому желанному решению, «найденному» однажды и навсегда, препятствует большое многообразие конкретных историко-лингвистических ситуаций. Сомнительно, что есть возможность подвергать это многообразие к неоспоримой иерархии уровней, служащих в заданной последовательности основаниями классификационных разбиений различной точности, как к некоему общему знаменателю, пригодному на все случаи. Может оказаться, что компаративные процедуры, ориентированные на определенную «шкалу предпочтительности» разных языковых уровней, доказав свою эффективность в исследовании одних групп языков (диалектов), продемонстрируют существенно меньшую разрешающую силу, будучи примененными к иным языковым группам и типологическим ситуациям.

Эти задачи могут решаться по-разному, в зависимости:

- от количества идиомов, подлежащих сравнению;
- от их таксономического ранга;
- от принадлежности сравниваемых языков тому или иному морфологическому типу;
- от степени временной удаленности рассматриваемых современных языков от состояния праязыкового единства;
- от того, происходило ли историческое развитие хотя бы одного из сравниваемых языков на каких-то этапах в относительной или полной изоляции от родственных языков (как, например, исландского, цыганского, калмыцкого, венгерского) или же они находились в состоянии непрерывных близкородственных контактов, которые могут приводить к нивелированию многих изначальных диалектных различий;
- от наличия в распоряжении исследователя сведений о древних периодах эволюции данных языков, в частности, от принадлежности языков к письменным (древлеписьменным) или бесписьменным (младописьменным) и т. д.

Морфологический критерий, отстаиваемый Вяч. Вс. Ивановым в качестве наиболее диагностичного при установлении диалектных отношений внутри праязыка, представляется малоприспособленным для исследования связей внутри семей языков с неразвитой флексией. Роль фонетических данных исключительно существенна для решения «пороговых» проблем, в «синхронном аспекте компаративистики», высока она и для выяснения

диалектного членения праязыка, но в этом случае кажутся достаточно убедительными и соображения о преимуществах морфологии перед фонетикой, когда приходится иметь дело с языками флективного типа (особенно «ультраграмматическими», по терминологии Ф. де Соссюра [Соссюр, 1977, стр. 165–166]). Значимость лексических фактов в подобных построениях может быть весьма существенной, однако, разумеется, при непременном условии, что заимствования надежно выявлены и исключены из процедур компарации.

Иногда отмечают, что значительная структурно-морфологическая близость между языками одной семьи (например, исключительно высокое грамматическое сходство тюркских языков) может быть препятствием к применению компаративных процедур к их материалу: «...при очень большом сходстве родственных языков в области грамматики трудно установить историческую перспективу грамматических соотношений между ними. Между тюркскими языками наблюдается больше лексических и семантических расхождений, чем фонетических и грамматических, это создает большую историческую перспективу лексических и семантических соотношений тюркских языков...» [Мусаев, 1975, стр. 4]; ср. классическую максиму: «Язык тем менее допускает сравнительное изучение, чем регулярнее его морфология» [Мейе, 1938, стр. 444].

Предпочтительность, так сказать, «очередность» тех или иных языковых уровней, к фактам которых должен обращаться компаративист, занявшись выяснением меры взаимной близости в группе родственных языков или восстановлением картины отношений между диалектами праязыка, в сущности, никогда не ясна а priori. Определить и утвердить ее применительно к конкретной историко-лингвистической ситуации может лишь сам анализ данных, относящихся к различным языковым уровням, результаты которого многократно корректируются последующими рассмотрениями проблемы.

И все же есть необходимость обрисовать особенности лексического уровня языка, данные которого служат исходным материалом для моделирования системы родства в группе языков с определением относительной силы их генетических связей, как источник для реконструкции картины диалектного членения праязыка, лежащего в основании данной группы и т. п., — в его отличии от иных языковых уровней, главным образом фонетического. Оценка сравнительной эффективности лексического критерия в лингвогенетических классификациях возможна лишь на основе принятия во внимание этих особенностей.

2. Количественная характеристика лексикона как базы генетико-классификационных построений

Наиболее очевидной чертой словарного материала как генетико-классификационной базы является его большой количественный состав. Объем лексических фактов, могущих быть положенными в основание реконструкции системы языкового родства, в сотни раз превосходит любые списки фонетических и грамматических явлений, которые используются для решения той же задачи. И это при том, что в указанных целях может использоваться только лексика, унаследованная современными (или засвидетельствованными в истории) языками от праязыкового состояния, но не возникшая на стадии существования самостоятельных языков, составляющих родственную группу или семью.

Встречающиеся в лингвистической литературе оценки объемов праязыковых лексиконов (для различных семей и групп) сильно колеблются, но всегда речь идет о тысячах лексем, при этом нетрудно заметить, что с продолжением исследований размеры как словарных составов праязыков, так и праязыковых словников отдельных языков-потомков, в таких оценках имеют тенденцию к возрастанию.

Для польского и чешского языков Т. Лер-Сплавинский и Т. Орлось считали возможным восстановить примерно по 1 700 праславянских лексем в каждом, см.: [Лер-Сплавинский, 1954, стр. 64; Орлось, 1958]. И. Попович предполагал, что в сербохорватском языке доступны реконструкции 4–5 тысяч слов праславянского происхождения [Попович, 1960, стр. 33]. Изданная к настоящему времени часть «Этимологического словаря славянских языков» под ред. О. Н. Трубачева — менее половины планируемого объема словаря — уже заметно перекрыла оценку И. Поповича, а цифры Т. Лер-Сплавинского и Т. Орлось уже превышены в два и два с половиной раза соответственно, так что по завершении издания ЭССЯ выяснится, что примерные оценки праязыковых словников отдельных славянских языков, данные этими двумя авторами, «скромнее» реальных показателей в 6–8 раз (наши гипотетические прикидки этого характера см. далее, в главе 5). Общая величина только доступного реконструкции праславянского словаря по современным весьма сдержанным подсчетам составляет никак не менее 20 тысяч лексем.

В своих попытках приблизительного определения объема лексиконов прауральского и прафинноугорского языков Петер Хайду называет сходные цифры: «Этимологий, отражающих лексику уральской и финноугорской эпох насчитывается до 1 200. Почти все из 1 200 возводимых к

празыку слов представляют неразложимые корни. Если, однако, принять, что по крайней мере половина (а возможно, и больше) праязыковых корней исчезла без следа или не выявлена, то в принципе число корней в праязыке можно оценить в 2 500...» [Хайду, 1985, стр. 259]; «...вероятно, не будет ошибкой предполагать, что объем праязыкового словаря — с учетом дериватов — достигал по меньшей мере 20 000—25 000 лексических единиц и даже, возможно, значительно превосходил это минимальное количество» [там же, стр. 260]. Конечно, не весь праязыковой лексикон может быть реконструирован, но тем не менее циффы выглядят очень внушительными.

Число же явлений, относящихся к фонетическому уровню и кладущихся в основу языковых классификаций, составляет обычно несколько десятков, реже несколько сотен, то есть на два-три порядка меньше.

Например, В. И. Перебийнос в вычислениях коэффициента родства славянских языков по фонетическим «дифференциальным признакам» («В качестве дифференциальных признаков были взяты различные фонетические процессы и явления: рефлексии индоевропейских звуков и звукосочетаний, развитие ударения, изменения в характере построения слога и т. п.») использует список в 180 единиц [Перебийнос, 1967, стр. 231]. В. Н. Чекман и О. С. Широков, подошедшие к отбору фонетических признаков в целях классификации славянских языков с несколько иных позиций, строят ее на списке из 38 фонетических процессов, приведших к образованию внутриславянских изоглоссов, см.: [Чекман — Широков, 1962, стр. 134—137].

Встречаются и более пространные реестры фонетических фактов, которые кладутся в основу генетических классификаций. Для дравидийских языков М. С. Андронов, например, перечисляет 344 фонетических изоглоссы: 77 в области вокализма и 267 в области консонантизма, см.: [Андронов, 1978, стр. 120—122, 145—154], правда, без их хронологической стратификации (то есть возможно, что далеко не все изоглоссы являются таковыми, если под изоглоссами понимать совместные, а не параллельные или разновременные, изменения, и, значит, конечная цифра может оказаться значительно меньшей; к тому же многие формулы отдельных фонетических рефлексаций могут быть обобщены — сведены к меньшему количеству типов звукоизменений, подобно тому как, например, славянские палатализации, толкуемые как изоглоссы, могут быть представлены в виде совокупностей более «мелких» процессов, определяемых с указанием на конкретные, «поименные» консонантные рефлексации, ср. упомянутую работу [Чекман — Широков, 1962]).

Вероятно, приблизительно сходную картину можно увидеть, обратившись к морфологии. Так, диалектное членение индоевропейского праязыка осуществляется Т. В. Гамкрелидзе и Вяч. Вс. Ивановым с привлечением шестнадцати грамматических изоглосс; с учетом буквенной индексации в их нумерации количество морфологических изоглосс исторических индоевропейских диалектов увеличивается здесь до 23, см.: [Гамкрелидзе – Иванов, 1984, т. I, стр. 393–394].

Нетрудно увидеть, что лексика дает возможность обнаружения гораздо большего количества сходжений и размежеваний между родственными языками и диалектами праязыка, из которых они развились, чем фонетический и морфологический уровни. Это означает, что лексический материал в принципе позволяет добиваться большей «прецизионности» при реконструкции картины родства благодаря большей сложности и дифференцированности «шкалы» степеней близости между отдельными языками.

Лексика (и словообразование, в той степени, в какой деривация имеет отношение к собственно лексикологической проблематике и в какой ее результаты сказываются на составе и облике словаря) дает чрезвычайно широкие возможности в исчислении и выборе изоглосс. Понимать это следует так, что факты, которые свидетельствуют о лексической близости между родственными языками и могут быть положены в основу реконструкции их исторических взаимоотношений, поставляются не только п о з и ц и я м и словаря. Они могут извлекаться также из многократных и многомерных разбиений словарного материала, из констатаций большого числа поверхлексемных, или, точнее, т р а н с л е к с е м н ы х, связей, которые позволяют установить различные межъязыковые сближения дополнительно к «чисто» лексемным изоглоссам. Количество изоглоссных сходжений, которые можно установить, исследуя диалектный или полный этимологический словарь, гораздо больше, чем число единиц, определяемое по заголовкам словарных статей. Это обстоятельство, вполне очевидное для любого, кто сколько-нибудь знаком с практикой составления словарей и их анализом, еще более наглядно проявляется на страницах лексических атласов, где для выделения различных типизируемых формальных модификаций подлежащих картографированию языковых единиц, которые включают в себя несколько структурно взаимосвязанных элементов, кроме основной конфигурации знака и его цвета обильно используется всевозможная диакритика, ср.: [Госсенс, 1969; Назарова, 1966; Назарова, 1974; За-кревская, 1974]. Это особенно характерно для славянской традиции лингвистического картографирования.

В качестве примера рассмотрим с этой точки зрения одну-единственную карту, с очень, подчеркнем, небольшой и компактной легендой, из Лексико-словообразовательной серии «Общеславянского лингвистического атласа».

На карте № 37 [ОЛЯ, Животный мир (I), стр. 98–99] представлены ареалы девяти славянских наименований щуки, *Esox lucius* L.: [1] *+ščika*², [2] *+ščika*, [3] *+ščipaka*, [4] *+ščipakъ*, [5] *+ščipakъ*, [6] *+ščipьль*, [7] *+ščipьль*, [8] *+kožušьkъ* и [9] *+kožušьka*. Собственно пока, при перечислении слов (или праформ), их семантика для исчисления изоглосс не играет большой роли (она существенна, правда, в процедуре этимологической идентификации диалектных формальных вариаций). При ограничении же данным списком, то есть при подключении семантического знаменателя – обоснования такой, а не иной группировки лексем, – возникают дополнительные возможности выделения формальных подгрупп внутри данного списка. Помимо перечисленных лексических (лексемных) изоглосс, он дает возможность выделить еще по крайней мере тринадцать (!) транслексемных сближений – изоглосс иного субстанциального наполнения. Все приведенные лексемы, связанные общностью семантики, объединяются в два этимологических гнезда, [10] одно из которых восходит к и.е. **skeu-* ‘острый’, см.: [Ильинский, 1917, стр. 204; Фасмер, т. IV, стр. 510]: лексемы [1] – [7]; [11] второе принадлежит к дериватам слав. **koža* (→ **kožuxъ*): лексемы [8] и [9]. [12] Лексемы [1] и [2] объединяет общий расширитель индоевропейского корня *-k-*. [13] Лексемы же [3] – [7] связываются общим расширителем *-p-*. Из них [14] лексемы [3] – [5] образованы с помощью суффикса *-ak(ъ)*, а [15] лексемы [6] и [7] – при помощи суффикса *-ьl(ъ)/-ьl(ъ)* (мы склонны считать их очень близкими вариантами одного и того же суффикса, вокализм которого подвергается колебаниям вследствие «гармонизации» с гласным корня). Различия в корневом вокализме служат для комбинирования изоглосс, объединяющих [16] формы [1], [3], [4] и [6] – с корневым *-u-* – и [17] формы [2], [5] и [7] – с корневым *-i-*. При этом здесь как частные изоглоссы выделяются объединения славянских корней [18] *+ščip-* (лексемы [5] и [7]) и [19] *+ščip-* (лексемы [3], [4] и [6]), а внутри последнего – еще и изоглосса [20] основы *+ščipak-* (лексемы [3] и [4]). Кроме того, в приведенном выше списке допустимо выделение названий щуки, принадлежащих [21] парадигме женского рода, с флексией *-a* (лексемы [1],

² Знаком + вместо астериска помечаются не праязыковые реконструкции, а междиалектные соответствия (любого возраста), ср.: [Толстой, 1969, стр. 21].

[2], [3] и [9]) и [22] парадигме мужского рода, с флексией -ѣ (лексемы [4] – [8]), то есть объединения элементов лексических рядов по морфологическому признаку. (Констатация еще одной «изоглоссы», объединяющей слова с суффиксами на -k-, то есть лексемы [3] – [5], [8] и [9], по-видимому, малоосновательна: общего механизма их образования не усмотреть). Итак, к девяти лексическим изоглоссам, формульно отраженным в легенде к карте ОЛА, добавляется еще тринадцать извлекаемых из того же материала изоглосс, не сводимых, однако, к «позициям» некоего наддиалектного (в данном случае) «словаря». Если же материал, представленный картой ОЛА, расширить не встретившимися эксплораторам атласа словами (см., например: [Коломиец, 1983, стр. 94]), то возможности исчисления лексемных, корневых, словообразовательных изоглосс, связанных со славянскими наименованиями щуки, увеличатся еще более.

Можно спорить о мере схоластичности подобных примеров и рассуждений, о том, все ли констатированные сближения и связи разумно привлекать к сравнительному исследованию в качестве лексико-словообразовательных изоглосс. Однако ясно, что в количественном отношении фонетические и морфологические изоглоссы конкурировать с лексическими не могут.

3. О мозаичности изолекс: есть ли порядок в хаосе?

Количественное превосходство лексических изоглосс над фонетическими и морфологическими, тем более значительное, чем менее удалены в прошлое исследуемые родственные идиомы, порождает одно немаловажное затруднение. Фонетические, морфонологические и грамматические явления в силу их относительной немногочисленности удобнее для использования как основания генетико-классификационных построений: система соответствий между языками обладает большей компактностью и обозримостью. Не столь значительное, по сравнению с лексическими, количество фонетических и морфологических изоглосс, даже если исследователь стремится к их максимальному выявлению и инвентаризации, позволяет реконструировать более простую, стройную и легче читаемую картину межъязыковых связей, особенно когда она обнаруживает различные более или менее четкие совпадения (пучки) изоглосс. Картина же, рисуемая лексическими изоглоссами, может представлять – и, видимо, обычно представля-

ет — собою существенно отличное зрелище, оставляя впечатление беспорядочного нагромождения линий.

Характерно в этом отношении признание О. Н. Трубачева, тем более что оно сделано на сравнительно раннем этапе обработки славянского лексического материала при подготовке «Этимологического словаря славянских языков»: «Вторым после этнонимов, а по количеству — даже первым лингвистическим знаком первоначальной этнолингвистической пестроты носителей славянских миграций к Западу, Югу и Северо-Востоку являются пестрые (и, как мы полагаем, не случайно пестрые) лексические изоглоссы, которые в массе выглядят как польско-болгарские, сербохорватско-украинские, полабско-великорусские... и т. п. соответствия, невероятно мозаичные и затрудняющие восстановление древней диалектной картины, которая предвляла языковую нивелировку новых этнолингвистических конгломератов Запада, Юга и Востока. Чем дальше двигается работа по праславянской реконструкции и этимологизации лексики для нового этимологического словаря славянских языков, тем более стойким делается впечатление мозаичности упомянутых внутриславянских изолек. Концепции вроде карпатско-полесского пояса диалектных изоглосс, возможно, увлекают воображение своей стройностью, но они, увы, попросту тонут в той праславянской сложности, которую дает сплошная реконструкция древнего состава лексики» [Трубачев, 1974, стр. 66] (пассаж о карпатско-полесском поясе отсылает к наблюдениям над того типа трансславянскими лексическими связями, который выявлен, например, в работах: [Гавацци, 1960; Толстой, 1968; Толстой, 1977а; Толстой, 1977б и др.]).

Проблема заключается в том, считать ли огромное численное превосходство изолек над фонетическими и грамматическими изоглоссами преимуществом или недостатком в глазах компаративиста. Не препятствует ли гигантское, не окидываемое единым взглядом количество лексических связей и расхождений между родственными языками выработке ясных представлений о диалектной структуре праязыка? Ведь некоторое число устойчиво повторяющихся по набору идиомов и этим поддерживающих, «усиливающих» друг друга изолек, которые могли бы стать основой для утверждения о большей генетической связанности одного данного языка с другим и тем самым, напротив, о меньшей его близости третьим языкам, может терять свою убедительность из-за существования множества не повторяющихся по конфигурации изолек, вкривь и вкось исчерчивающих карту родственных — в нашем случае славянских — диалектов. В состоянии ли менее многочисленные устой-

чивые лексические корреспонденции между языками «противостоять» хаотическому нагромождению иных соответствий в установлении уровней лингвистического родства?

Хочется надеяться, что дальнейший анализ лексических фактов славянских языков, представленный в настоящей работе, а именно базирующийся на статистическом подходе, может дать положительный ответ на этот вопрос.

4. Лексика: нестабильность или консерватизм?

Я не скажу, как все твердят давно,
Что, мол, в подлунной все обречено
Распаду...

Жюшан дю Белле. «Древности Рима»

Одной из чрезвычайно важных особенностей лексики, если она берется в качестве материала для реконструкции древних диалектных отношений, и при этом особенностью, естественно, оцениваемой негативно, считают ее неустойчивость. Мнение о нестабильности лексикона как составляющей языка имеет настолько широкое хождение, что приводит какие-либо цитаты в подтверждение его популярности нет ни малейшей необходимости.

Положение о большой подвижности лексической системы языка никоим образом нельзя отнести к заблуждениям. Основания для оценки лексического уровня как весьма динамичного более чем достаточны. И впрямь, будучи составляющей языка, наиболее тесно связанной с внелингвистической действительностью, и уровнем, наиболее податливым иноязычному воздействию, лексика испытывает постоянное давление внешних факторов. Воскресни авторы «Евгения Онегина» или «Мертвых душ», сегодняшний номер газеты показался бы им непереваримой абракадаброй. Примечательно, что из всех дисциплин, изучающих разные языковые уровни, именно лексикология пользуется наибольшим количеством терминов, в которых акцентируются семантические моменты 'времени', 'изменения', 'креации' и под.: «историзм», «архаизм», «словообразование», «производное», «неологизм», «деэтимологизация», «заимствование», «сужение значения», «перенос», «лексикализация» и мн. др., выражения вроде «количественный рост словаря», «обогащение лексикона», «лексические утраты» и т. д.

И все же подчеркивание только лишь изменчивости словаря — это весьма односторонняя точка зрения. Правильнее было бы определять лексику как парадоксальный языковой уровень. Наряду с очевидной подвижностью словарю свойствен и поразительный консерватизм. Он труднообъясним даже если прибегнуть к популярным аргументам вроде того, что стабильность является элементарным и аксиоматическим требованием, «предъявляющимся» коммуникативным системам, вне какого условия язык просто не может выполнять свое функциональное назначение, или на жизненную важность для коллектива носителей данного языка наиболее консервативных компонентов его словаря.

В самом деле, если сохранность таких древнейших слов, функционирующих в современном русском языке, как *дом, двор, сын, сестра, зерно, овес, мед, огонь, нить* и множество подобных, может объясняться исключительной социальной, культурной и хозяйственной ролью называемых ими реалий, то кажется сомнительным, чтобы хотя бы на каком-то этапе столь же высокой существенности для социума достигали означаемые таких слов, как *выдра, ясень* или *дрозд*, не менее древних и устойчивых, чем слова, входящие в первый перечень (ссылки на ритуально-культовую значимость тех или иных предметов, явлений и их имен (ср., например: [Гамкрелидзе — Иванов, 1984, т. II, стр. 529]), в определенной мере снижают свою объяснительную силу для поздних периодов существования этих слов, а также — и мы считаем это чрезвычайно важным замечанием — из-за возможной в подобных аргументах недооценки полидиалектного характера культуры, то есть излишней универсализации древнейших представлений, не всегда оправданной экстраполяции их на все локально-этнические традиции при более реалистичном предположении об их варьированности). С другой стороны, возможны десятки контрпримеров подвижности, неустойчивости обозначений как раз жизненно важных реалий или явлений, существенных для формирования и репродуцирования картины мира, ср. относительную новизну, неунаследованность от праславянского таких русских слов, как *женщина, мужчина, хозяин, власть, враг, пища, одежда, лошадь, воздух, длинный, средний, красный* (в цветовом значении), *сегодня* и т. д.

Многообразие причин, которые обуславливают консерватизм и, напротив, обновляемость словаря, настолько велико и противоречиво, что попытки свести их в более или менее стройную систему эксплицитных зависимостей не могут не оборачиваться сознательным игнорированием разрушающих убедительность этой системы фактов. Исчезновение в славянском

индоевропейских названий медведя и дуба с заменой их новообразованиями получает простое и добротное объяснение действием табу вследствие их особой культовой значимости. Но указание на эти мотивы обнаружит явную свою недостаточность, если пытаться объяснить причины различного сложения славянских судеб индоевропейских названий медведя или дуба, с одной стороны, и названий волка или ясеня, с другой. Рассуждения о роли табу в обновлении словаря всегда оставляют впечатление неполноты, «штучности», их как правило хочется снабдить оговоркой «в некоторых случаях». Применительно к лексикону какого-либо языка, рассматриваемому в какой-либо момент его развития, уместно напомнить Ф. де Соссюра, настойчиво говорившего, правда, по иному поводу, о «случайном характере всякого состояния» [Соссюр, 1977, стр. 118].

5. Гиперболизация фактора лексических утрат в изменении словаря

Тогда считать мы стали раны...

Лермонтов. «Бородино»

Столь же неоправданно гиперболизуемым является такой фактор изменения словаря, как естественная убыль слов вследствие их забвения. Суждения на этот счет обычно подкрепляются ссылками на исторические словари: не может не броситься в глаза большое количество свидетельствуемых ими слов, не доживших до сего дня, безвозвратно утраченных языком. Скажем больше: сложению представлений о значительности лексических потерь в эволюции языка исторические лексиконы способствовали более, чем другие научные экспликации языковых изменений. Поэтому обратим на них более пристальное внимание и позволим себе довольно обширное отступление.

Действительно, при сплошном, без пропусков, чтении исторического словаря вроде Срезневского, ДРС или СлРЯ XI–XVII вв. возникает впечатление, что прогуливаешься по аккуратным дорожкам некоего лексического некрополя. И, поражаясь его огромности, невольно склоняешься к хотя бы очень грубым сравнительным численным оценкам «живых» и «павших». Например, таким. В словнике упомянутого СлРЯ XI–XVII вв. к концу 18-го выпуска (последнее в нем слово – *преначальный*) накопилось 57 020 слов, не считая отсылочных. Середина буквы «П» – это приблизительно две трети конечного объема словаря

(примерно 68–70%?), и, значит, конечный объем СлРЯ XI–XVII вв. составит около 85 тысяч слов. Если даже четверть объема лексики, описанной в этом словаре, уже не функционирует, то лексические утраты, не включая в них исчезновение слов, возникших после XVII в., могут составить никак не меньше 20 тысяч единиц. Если же учесть утрату и более поздних лексических вхождений, в частности петровского и послепетровского периодов, словарные потрясения покажутся просто разрушительными.

Имеются и более конкретные оценки. Одна из них принадлежит В. З. Санникову, см.: [Санников, 1975, стр. 30 и след.]. Обследовав древнерусские юридические тексты XI–XIV вв., он установил, что из 1700 слов с частотой не менее 3 в современном литературном русском языке сохраняются 1189 слов, или 69,9% (критерием сохранности было выбрано включение слова в ССРЛЯ без помет «устаревшее», «областное», «просторечное»).

И все же скорбные ассоциации в данном случае обманчивы. Словарные утраты не настолько безумно многочисленны, как это может внушить не критичная перцепция собраний лексики прошлого.

Дело здесь вовсе не в том, что какая-то часть слов, исчезнувших из активного хождения, сохраняется в виде архаизмов, историзмов и проч., теплясь где-то на задворках современного словоупотребления – в сочинениях археологов (ср. *рало*, *требище*, *жалыник*, *дробница*, *колт*, *льянец*), историков (ср. *древляне*, *тиун*, *жолнер*, *лундыш*...) или исторических беллетристов. И даже не в том, что обращение к диалектным словарям может скорректировать неблагоприятный кладбищенский эффект исторического лексикона: некоторые лексемы оказываются в нем узкими регионализмами, находящимися за пределами литературного языка (носителями которого в большинстве своем являются читатели исторических словарей) и обнаруживаются живущими в тех или иных народных говорах до сей поры, нередко сохраняя прежние значения.

Главное все же в другом: значительная часть отмеченных в исторических лексиконах, но не сохранившихся в памяти народа слов, в сущности, с л о в а м и не являлась и никогда в этой самой п а м я т и не находилась. Добросовестные лексикографы помещают в исторические словари – и это, несомненно, правильно – все сегменты текста, имеющие формальные признаки самостоятельного слова: морфологические (флексия), графические («от пробела до пробела», если письменность пользуется пробелами, либо иные дистинкции) и т. п. Вопрос заключается в том, все ли такие сегменты равноправны друг с другом и должны расцениваться как принад-

лежность лексикона — не данного издания, но вообще словарного состава языка, «słownictwa»? Ответ должен быть, на наш взгляд, отрицательным. Большое число словарных позиций в исторических словарях (воспользуемся для примера тем же СлРЯ XI—XVII вв.) составляют разовые заимствования (ср.: *анесея* 'племянница', *дохиарь* 'хранитель казны', *друмь* 'дорога, улица', *евхия* 'молитва', *еландии* 'судно, корабль', *еродии* 'цапля или аист', *корида* 'вошь, клоп...'), варваризмы-«транслитераты» (ср.: *абдаль* 'дервиш', *агунь* 'мулла', *аммаль* 'носильщик', *гарипь* 'чужестранец', *гоппа* 'тип морского судна', *динь* 'китайская мера веса', *дюмендъ* 'сорт ткани', *зигия* 'род дерева', *импола* 'крытая улица для торговли', *кавкаль* 'сосуд с узким горлышком; чаша', *касаба* 'тип поселения (в Аравии)', *кидонать* 'род сладкого кушанья из плодов кидонии, айвы', *кумеркиарий* 'сборщик податей, налогов...'), кальки (ср.: *безвечерний* 'немеркнувший', *бездение* 'глубина', *благозельный* 'плодородный', *виночваньчий* 'виночерпий', *грозоокий*, *единовольство*, *звърубийца*, *зложизние*, *конеелень* 'разновидность лося или антилопы...'), с одной стороны, и ультрарегулярные образования, ограничивающиеся по сути только частеречной трансформацией (ср.: *бобный*, *велебности*, *вельблудский*, *голубный*, *гонзение*, *дружение I*, *дьяконовати/дьяконствовати*, *жезльный*, *земностный*, *ключничный*, *лучиество...*), с другой. Кальки и ультрарегулярные дериваты представляют собою, в сущности, не явления словаря, но явления текста.

К последним, без сомнения, относятся и многочисленные основосложения, возникающие либо в результате свертывания некоего предшествующего текста, инкорпорации, либо по высокопродуктивной модели аналогическим способом, независимо от того, служит образцом иноязычное или свое же слово (ср.: *басносоставецъ*, *благопогодный*, *вкуполатынный*, *воокругпредельный*, *всезвъриньственный*, *главоотяжство*, *гноегребецъ*, *капищезасъдатель*, *кривобраздникъ*, *лькаромазатель*, *матередосадитель*, *небоумный*, *новопрзязбиший*, *новоспротивный*, *ногобольние*, *пагуболаголивый*, *полутроебрачный...*; *любовеликоденьникъ*, *любодарливый*, *любоидольное*, *любоживотный*, *любомолвецъ*, *любоплачевный*, *любоплевельный*, *любопозорникъ*, *любосеребрьство*, *любоуцреждение...*). Окказиональность и высокая регулярность либо, часто, напротив, ощутимая натужность, неорганичность таких образований препятствуют признанию их полноценными словами, достоянием «народноязыктворца», подобно тому как вряд ли заслуживают помещения в словарь современного русского языка конструкции *бело-оранжево-фиолетовый*, *мозамбикско-новозеландский* или *южновеликорусско-белорусско-*

украинско-словацко-чешско-верхнелужицко-западнoслoвенский (последний пример, состоящий из двенадцати корневых морфем и достойный фиксации в «Русском Гиннесе», — не лабораторная выдумка, а подлинный случай: см. [Бернштейн, 1961, стр. 71]; перед ним меркнет даже пассаж А. Битова: «...с благосклонно-нежно-равнодушно-соблазнительно-независимо-тепло-человечно-мужской улыбкой» («Улетающий Монахов»))³.

Лексикографы-историки, не предаваясь в словарях (и, повторим, справедливо) соблазнительному теоретизированию насчет необходимости разграничивать лексику реальную, воспроизводимую и лексику потенциальную, производимую в ходе создания текста, делают акцент на результатах словообразовательных процессов, оставляя в стороне собственно креативный их характер. Мы должны привыкнуть к мысли о том, что минимальным текстом не только является сочетание лексем, но может являться и сочетание морфем, а словообразование должно рассматриваться, наряду с фразеологией, как область, промежуточная между словарем и текстом. Тема эта — скорее для историко-лексикологических монографий (см. об этом, в частности: [Сорокин, 1977; Мальцева, 1977; Шимчук, 1975]), чем для полных исторических словарей. Что до последних, то некоторые их участки уместнее сравнивать не с кладбищем, а с экспозицией экспериментальных образцов или даже музеем-лабораторией. Такая ассоциация кажется нам тем более оправданной, что к перечисленным типам несостоявшихся слов нужно добавить и обильно включаемые в словники исторических лексиконов в виде самостоятельных позиций отвергнутые впоследствии, непривившиеся фонетические варианты и грамматические формы, как бы вскрывающие изнутри сам процесс освоения заимствований и становления словаря (ср.: *анфилопъ* 'антилопа', *априлий* 'апрель', *бергместерь* 'бургомистр', *вирель* 'берилл', *голоберда* 'алебарда', *декамврй* 'декабрь', *ямвь* 'ямб', *иезовитъ* 'иезуит', *карвамиколось* 'карбункул', *мафиматикийский* 'математический'...).

Экскурс в лексикографические проблемы, с трактовкой многих словарных фиксаций как имеющих креативный, текстовый характер, как элементы дискурса, понадобился здесь затем, чтобы обосновать одно, может быть, несколько рискованное утверждение: исторический словарь

³ Удачным представляется решение, принятое с некоторых пор в лексикографических бюллетенях «Новое в русской лексике (Словарные материалы)», помещать подобную лексику не в основном корпусе выпуска, а в приложении «Список слов без описания» (ср.: [Новое в русской лексике, 1981, стр. 258—259]: *испуганно-предостерегающий*, *недосягаемо-суровый*, *пятьсотдвадцатикилометровый* и т. п.).

с его ориентацией на безысключительное запечатление всей встретившейся в древних источниках лексики дает искаженное представление, во-первых, уравнивая экспериментальные псевдослова с «натуральной» лексикой и, во-вторых, за их счет завышая объемы словарных составов предшествующих — исторически засвидетельствованных! — состояний языка. Реальный лексикон, составными элементами которого являются воспроизводимые лексические единицы, имеет несколько меньшие масштабы, а представления о роли словарных утрат в истории языка, как говорилось, имеют обыкновение сильно преувеличиваться.

В последнем приходится убеждаться с ростом числа диалектных словарей, которые оправдывают надежды на сохранение многих древних слов, ранее не попавших в поле зрения собирателей-диалектологов. Достаточно упомянуть в этой связи, что в опубликованных выпусках ЭССЯ, по нашим данным, каждая пятая позиция может быть дополнена не отмеченной в том или ином языке лексикой праславянского происхождения, см.: [Журавлев, 1990].

Не следует, как нам кажется, упускать из виду и возможности возвращения, воскрешения утраченной, казалось бы, лексики. Это может происходить вследствие не теряющих своей силы древних словообразовательных механизмов и благоприятных условий для их повторного «срабатывания» (такие условия могут усматриваться в сохраняющейся активности деривационной модели, наличии определенного количества аналогичных образований и т. п.). В наших, например, дополнениях к лексическим материалам ЭССЯ (подробно о них — ниже) в словарную статью **čьwъ(jь)* было включено в.-луж. *čьwу* 'нервы' [Трофимович, 1974, стр. 22]. На полях рукописи О. Н. Трубачев, читавший ее, сделал пометку «неологизм? (калька?)», находя причину для сомнения, по-видимому, в семантике современного лужицкого слова (в [ЭССЯ, вып. 4, стр. 137], праславянская реконструкция **čьwъ(jь)* осуществлена на основе сербохорватского, чешского и русскоцерковнославянского свидетельств в значениях 'внимание', 'чувствительный, чуткий' и 'существо, одаренное силою чувствования'). Предположение о семантическом калькировании представляется в данном случае достаточно правомерным (но не вполне ясно, как велика вероятность калькирования), однако важно, что в лужицком как бы хранилась возможность реализации именно этой праславянской формы, в глубине лексико-словообразовательного «котла» она латентно вызревала и ждала своего времени, в конце концов осуществившись, может быть, не без влияния внешнего толчка (калька). Но допустимо думать и о том, что форма, продолжающая праслав. **čьwъ(jь)*, была известна носителям идиума, пред-

ществовавшего современному верхнелужицкому языку, но из-за низкой ли ее употребительности, наличия более «сильных» синонимических возможностей или по иным причинам, на какое-то время оказалась призабытой, чтобы воскреснуть, когда возникла потребность в выражении нового (но, так сказать, «смежного», «соседнего») лексического значения, для заполнения вновь обнажившейся лексико-семантической лакуны.

6. Различия между механизмами эволюции фонетики и лексики: вытеснение и кумуляция

К отмеченной выше парадоксальности лексического состава языка, помещению в нем очевидного динамизма с удивительным консерватизмом, гораздо более глубоким, чем устойчивость других языковых уровней, приводит особый механизм развития словаря, принципиальнейшим образом отличный от механизмов эволюции звукового строя и грамматики. Если фонетика и, в меньшей степени, морфология, будучи системами с жесткой, в некоторых участках почти механической взаимообусловленностью компонентов, эволюционируют преимущественно путем вытеснения одних элементов и структур другими, то развитие словаря, системы более рыхлой, «мягкой», скорее даже конгломерата «мягких» систем, протекает главным образом за счет кумуляции элементов (слов).

Конечно, история любого языка дает примеры материального расширения состава фонетических единиц и усложнения системы фонологических оппозиций, то есть вытеснение одного звука другим не является единственно мыслимым типом фонетических изменений. Однако такому увеличению репертуара фонем положен, как известно, типологический предел: он не может превысить объема в несколько десятков элементов (максимум – 80 с небольшим в некоторых языках абхазо-адыгской группы, см.: [Климов, 1986, стр. 32–33]).

В лексической же системе таких пределов, как будто, не существует. При известной стабильности и ограниченности (1–4 тысячи лексем) активного словаря низкочастотная зона лексикона может разрастаться – за счет словообразования и заимствования извне – до сколь угодно больших размеров, и язык не испытывает при этом потребности в ее ограничении; коллапс словарю не угрожает.

Разумеется, жесткость и устойчивость даже такой языковой подсистемы, как фонологическая, относительна. Можно оспорить утверждение о

механической взаимообусловленности компонентов фонетической системы примерами проникновения в нее новых элементов (фонем), но не приводящего при этом к перестройке всей конструкции. Так, М. Я. Гловинская полагает, что состав согласных фонем заимствованных слов в современном русском литературном языке по сравнению с составом базовой фонетической системы включает в себя еще 22 фонемы (см.: [Гловинская, 1967, стр. 7]), то есть лексические заимствования расширяют консонантизм русского языка более чем в полтора раза. Однако, справедливо отмечается, это расширение фонемного репертуара происходит на языковой периферии (ср.: [Реформатский, 1966, стр. 96–109; Живов – Успенский, 1973, стр. 24–35; Журавлев, 1984, стр. 103]), куда относятся, помимо новых заимствований, ономастика, эмоциональная лексика, идеофоны, аббревиатуры и др., и не затрагивает центральных областей, которые характеризуются устойчивым тяготением к монотипизму.

Иным образом обстоит дело в лексике, где новые, в том числе заимствованные элементы (слова) способны оказываться в слое наиболее высокочастотных слов, то есть затрагивать сам центр лексической подсистемы, если ее рассматривать в динамическом аспекте. Например, заимствования из западноевропейских языков *армия, партия, машина* по своей встречаемости опережают слова исконного (праславянского и, далее, индоевропейского) происхождения *час, ночь, белый, старь, брат, сын, брат, сестра, нос, имя, ходить, солнце* и мн. др., входящие в фундаментальный словарь («основной словарный фонд языка»), а среди прилагательных 4-е место по частотности занимает свеженькое, по историческим меркам, новообразование-советизм *советский*, уступающее лишь адъективам *большой, новый и первый*, см.: [Частотный словарь, 1977, стр. 807–809]. Лексические новообразования и заимствования могут вызывать заметные смещения в исконном словарном составе, по крайней мере в употребительности отдельных лексем, но полное вытеснение новой лексикой прежней – явление сравнительно редкое, и наблюдать его можно, как правило, только в достаточно протяженном временном интервале, чтобы быть гарантированным от слишком поспешных констатаций. В обновлении словаря материальное вытеснение старых слов новыми играет в общем незначительную роль, которую иногда склонны сильно преувеличивать. Даже если наберутся сотни примеров подобного рода, они не смогут поколебать значения преобладающего пути изменения словарного состава – накопления, кумуляции лексического материала (о семантических изменениях в лексике мы здесь намеренно не говорим, имея в виду именно лексикон, состав единиц).

Лексический состав языка в силу кумулятивного принципа его формирования и меньшей значимости системных факторов в его организации, не подвержен таким большим потрясениям в своей эволюции, как грамматика и особенно фонетика. Изменения в фонетической системе языка часто носят характер цепной реакции, которая охватывает всю систему и изменяет принципы, обеспечивающие ее статическое равновесие, весьма радикальным образом. Лавинные фонетические процессы в славянских языках являют собою классический пример: действие закона открытого слога или падение редуцированных приводили к переменам в фонетической организации языка, которые нельзя оценить иначе как катастрофические.

Различия между лексикой и фонетикой в преобладающих механизмах исторических изменений (кумулятивный, с наслаиванием и «уплотнением» элементов, в одном случае, — и субститутивный, с вытеснением одного элемента другим, в ином) приводят к тому, что в глоттогенетических исследованиях, при установлении изоглосс, объектами отождествления, в зависимости от уровня языка, служат явления разной природы. Применительно к лексике и словообразованию (не семантике!) внимание сосредоточивается главным образом на собственно единицах (корнях, при более строгом анализе — лексемах, словообразовательных конструкциях с идентичным морфемным составом), применительно же к фонетическому уровню — главным образом на отношениях, соответствиях материальных единиц (звуков, звукотипов) и структур, присущих сравниваемым языкам, — регулярных корреспонденциях, позволяющих выявление закономерных переходов одних единиц в другие (сами единицы или их материальные характеристики — назальность, палатальность, придыхание, щелчок и т. д. — и структуры — количество фонем, ряды согласных, соотношение между вокализмом и консонантизмом, в системе и тексте, и т. д. — служат предметом рассмотрения в описательной фонетике и в типологии). Морфология в этом отношении занимает промежуточное положение: в качестве изоглосс, связывающих разные родственные языки, здесь рассматриваются и единицы — отдельные форманты и т. п. — и соответствия — в виде структур, выступающих в сходных или противопоставленных функциях.

Изменение в лексике состоит как правило в заполнении вновь сформированным элементом (появление новой лексемы путем словообразования или заимствования извне) пустоты, обнаружившейся при концептуальном освоении мира (иначе говоря, при возникновении потребности в номинации), реже — из потребностей экспре-

сии, но не заполнение пустоты, возникшей вследствие исчезновения некоего более раннего элемента. Такое изменение затрагивает — сравнительно незначительно — лишь ближайшие «окрестности» этого нового звена в рыхлой лексической системе: некоторые смысловые смещения у семантически близких слов, «захват» нового слова сетью деривационных возможностей данного языка, вовсе не непременно реализующихся, и т. п. В фонетике же происходит не заполнение «концептуальных» пустот, а замена одной единицы или структуры другою, нередко, как хорошо известно, с последующей перестройкой всей системы в целом, с полным обновлением ее облика, что обусловлено жесткостью ее организации. Недаром термин «строй» редко используется для характеристики лексикона (выражение «лексический строй» хочется поместить под астериском⁴), но стандартен в качестве обозначения фонетической и грамматической систем.

Вследствие относительно малого набора элементов и особенностей их организации большую роль в эволюции фонетической системы играют общетипологические моменты. Многие фонетические изменения протекают как бы по накатанным руслам, фонетические преобразования, происходящие в одних языках, по импульсам, характеру и результатам оказываются похожими на те, которые осуществляются в других (ср., например: [Серебренников, 1968, стр. 42 и след.; Серебренников, 1974, стр. 64–157]; специально о славянских: [Чекман, 1978]). Так, скажем, результаты множества славянских палатализаций и смягчений (передне- и особенно заднеязычных взрывных) в артикуляционном и акустическом отношении вполне аналогичны результатам подобных же процессов в индоевропейских и неиндоевропейских языках (свистящие, шипящие, в том числе аффрикаты, нередко с последующей дезаффрикацией). Всякий раз при решении вопроса о том, является ли сходное фонетическое изменение в родственных языках связывающей их изоглоссой, то есть совместной инновацией, исследователь стоит перед необходимостью взвешивать степень его специфичности и возможность участия в рассматриваемом процессе типологических факторов.

В лексике же, если, конечно, не иметь в виду звуко-символические образования простой структуры (впрочем, не столь уж многочисленные), типологические моменты или, точнее, возможность параллельного

⁴ В сущности, так и поступила О. С. Ахманова в своем «Словаре лингвистических терминов» [Ахманова, 1966, стр. 458], приведя это словосочетание не в числе толкуемых терминов, а в виде иллюстрации ко 2-му значению термина «строй»: в «Лингвистическом энциклопедическом словаре», судя по терминологическому указателю [ЛЭС, 1990, стр. 627–650], оно не встретилось ни разу.

образования приходится учитывать практически только в случаях выявления морфологически структурируемых цельнолексемных совпадений при чрезвычайно высокой активности участвующих в данном образовании аффиксальных элементов, что является существенным при сравнении материала заведомо близкородственных языков. Короче говоря, типологические импульсы при одинаковых инновационных фонетических трансформациях в разных языках в целом более вероятны, чем параллельность (несовместность возникновения) идентичных (или близких) по структуре лексических единиц.

Все изложенные на предыдущих страницах соображения должны, на наш взгляд, подкрепить суждения о высокой степени устойчивости лексики на фоне других языковых уровней. Очерченные особенности словаря в его отличии от грамматики и фонетики дают основания некоторым специалистам, в последнее время все чаще, высказывать мнение, согласно которому лексика является наиболее консервативным уровнем языка и потому самым надежным для установления степеней генетической близости.

Весьма категоричен в этом отношении В. Маньчак: «...прежде всего следует отбросить фонетические критерии, так как они наиболее обманчивы и ненадежны. Фонетика каждого языка меняется быстрее, чем его флексия и лексика, а фонетическое развитие имеет, как правило, настолько случайный характер, что нельзя делать на этом основании какие-либо выводы» [Маньчак, 1987, стр. 143] (мы находимся в некотором затруднении: не сделать ли в цитированном тексте поправку, заменив слово «случайный» на «неслучайный», что, если иметь в виду под последним обусловленность общетипологическими мотивами, лишь усилило бы смысл высказывания В. Маньчака); «Что же касается флексии, то она меняется медленнее, чем фонетика. Любой диалектолог знает, что дифференциация говоров в отношении флексии не так резка, как в отношении фонетики. Однако ошибочно мнение, что именно флексия представляет собой тот элемент языка, на основании которого можно решить, родственны данные языки или нет, а если да — то в какой степени» [там же, стр. 144]; «...следует четко констатировать, что, вопреки ошибочным мнениям многих лингвистов, наиболее медленно развивающимся элементом языка является не фонетика, не флексия, а лексика» [там же]; ср. еще: [Маньчак, 1986, стр. 491–493].

Удивляться решительности этих умозаключений не приходится. Они подготовлены всем предыдущим развитием сравнительно-исторического языкознания, как индоевропейского (ср., например: [Порциг, 1964; Золь-

та, 1960], где лексическим данным отводится роль решающего критерия при реконструкции древнего диалектного членения, и др.), так и специально славянского, см. известные книги Ф. П. Филина [Филин, 1962; Филин, 1972], многочисленные исследования О. Н. Трубачева и др. Последний, в частности, по поводу работы [Бирнбаум, 1966] пишет о «недооценке показаний лексики и чрезмерном доверии к фонетическим моментам» [Трубачев, 1968, стр. 368]. В целом весьма сочувственно к высокой оценке лексических показаний как генетико-классификационного критерия относится Г. А. Климов: «Среди отдельно взятых критериев группировки родственных языков морфологические признаки представляются обычно более надежными, чем фонетические, хотя необходимо подчеркнуть, что в основу соответствующей классификации они, как правило, ложились в совокупности с фонетическими или лексическими критериями (с последними, заметим, на наш взгляд, не столь уж часто. — А. Ж.), и что в случае изолирующих языков вопрос о них вообще снимается. Еще более надежным ориентиром в этом отношении часто признается степень лексической близости языков (особенно в условиях определенной тематической системности корнеслова), поскольку сравнение лексем предполагает наличие параллельного соотношения как в корневой, так и в словообразовательной морфеме... Действительно, именно в последнем отношении свою неповторимую специфику всегда обнаруживают наиболее близко родственные языки» [Климов, 1990, стр. 126].

Глава 2

Лексикостатистика как инструмент глотогенетического исследования

Быть точнее, чем циркуль,
точней, чем часы и линейка...
.....
Лишнее слово прибавь —
ничем не исправить ошибки.
Слово пропустишь одно —
в грех пред богами впадешь.

Феогнид (перевод С. Апта)

1. Статистическое моделирование в исторической лингвистике

Применение статистических методов для решения задач лингвистической компаративистики имеет уже почтенную историю. Если не считать громоподобного изобретения Моррисом Сводешем глоттохронологии (см.: [Сводеш, 1952; Сводеш, 1955; Лиз, 1953; Гуджинская, 1956; Хаймз, 1960]), произведшего сильное впечатление и на славистов, то в славистике ретроспективно наиболее, пожалуй, заметным событием была работа польского языковеда Яна Чекановского «Wstęp do historii Słowian» [Чекановский, 1927]. Я. Чекановский замечателен тем, что был одним из первых, если не первым, славистов, использовавших статистику для решения проблем языкового родства (см. еще: [Милевский, 1931; Лер-Сплавинский, 1930, стр. 130–137], оценку направления — в: [Поповская-Таборская, 1991, стр. 34]; до М. Сводеша и его продолжателей к этой теме обращались также А. Л. Кребер и К. Д. Кретъен при рассмотрении индоевропейских языков, Дж. Бонфанте, исследовавший индо-хеттскую проблему, см.: [Кребер — Кретъен, 1937; Бонфанте, 1946, и др.]). Об упомянутых первых опытах применения статистических подсчетов к историко-лингвистической проблематике писалось много, поэтому вдаваться здесь в их анализ мы считаем излишним.

Отношение к применению количественных методов в историческом языкознании, как и в языкознании вообще, было и остается очень разным — от энтузиастически-восторженного, с нередкой ссылкой на марксово утверждение, что наука только тогда, мол, обретает себя в этом качестве, когда связывает себя с математикой, или на исследовательский императив английского психолога и антрополога Фрэнсиса Гальтона, двоюродного брата Ч. Дарвина, «Где только есть возможность — считайте» (хотя тому же Ф. Гальтону принадлежит цитируемый отнюдь не реже саркастический афоризм о трех разновидностях лжи: извинительной лжи по необходимости, просто лжи, не находящей оправдания, и статистики), до абсолютно нигилистического, находящего свое обоснование в тезисе Николая Трубецкого насчет внеположности языка числу и мере [Трубецкой, 1960]. Энтузиастов не смущает известная маргинальность успехов математики на лингвистическом поприще, нигилисты же склонны вообще не замечать каких-либо достижений сторонников квантитативного исследования языка.

И критиками, и большинством трезвых сторонников использования статистических методов в историко-лингвистических штудиях этим методам отводится вспомогательная, подчиненная роль: верификация результатов неколичественного анализа, уточнение уже обнаруженных зависимостей и закономерностей. В целом это справедливо: большинство историко-лингвистических работ, где применена статистическая обработка численных данных, сводится именно к количественному подтверждению наблюдений, которые могли быть сделаны неквантитативным исследованием. К слову сказать, ничего унижительного ни для математики, ни для языкознания в этом нет: вспомогательную, подчиненную роль играют математические методы в применении к большинству наук, лингвистика здесь не исключение, ср.: [Колмогоров, 1988, стр. 8]. Но, как нам представляется, было бы все же ошибкой утверждать, что применение математического аппарата (точнее, статистических методов) в принципе не способно привести к получению новой информации о лингвистических объектах и обречено лишь конкретизировать и «оттенять» факты, найденные на других путях историко-лингвистического познания.

Помимо резонного сомнения во всемогуществе математики неприятие многими лингвистами использования математических методов в их специальности вызывается еще целым рядом причин. Не самая последняя из них — математическая неподготовленность большинства лингвистов, которая делает труднодоступными или даже вовсе недоступными им работы, в которых применен даже сравнительно простой математический аппарат: из школьного курса в памяти остаются в лучшем случае только численные

значение l до второго знака после запятой, знак радикала да понятие средней арифметической, и уже страница, на которой встречаются обозначения lg и Σ , внушает непреодолимый ужас. С другой стороны, в статистические исследования зачисляются работы, в которых содержатся простые подсчеты, действительно не вносящие ничего принципиально нового и только переводящие в относительно точную цифру интуитивно ясные соотношения, — случай самый частый, в котором чисто вспомогательная роль математики (в ее простейшей версии) предстает наиболее наглядно; не следует лишь относить подобные подсчеты к области статистики.

Паче же всего предубежденность против «математизации» лингвистических исследований может объясняться желанием — осозанным либо, скорее, бессознательным — уберечь от растления девическую сущность лингвистики. Между тем последняя, слава Богу, давно рассталась с девственностью, и обострять проблему «полезно ли применение статистики в языковедческих работах?» — значит отдавать прошлому уже ненужную дань.

Укреплять лингвистов в неприязни математики, хотя бы и простейшей, могут историко-лингвистические работы, выполненные с применением квантитативных методов, в которых из количественных наблюдений делаются неправомерные выводы и излишне смелые обобщения. Этот род работ, увы, не столь уж малочислен. Подробный разбор одной такой — во многих отношениях примечательной — работы [Стецюк, 1987], где механическая арифметика служит основой для «выявления» мест расселения древнеславянских племен, см. в статье: [Журавлев, 1991]. Неполный учет языков, образующих данную (в рассматриваемом случае славянскую) семью, опора только на факты двуязычных переводных словарей, то есть игнорирование диалектной лексики, отсутствие строгих критериев в установлении возраста обсчитываемых слов, нераспознавание поздних межславянских заимствований и заимствований извне, подсчет общих слов без принятия во внимание величины праязыковых словарей сравниваемых языков, то есть в абсолютном выражении, что приводит, например, к констатации большей «генетической» близости русского языка к сербохорватскому, чем к украинскому и белорусскому, механическое же, в произвольном масштабе, размещение полученной «схемы родства» на географической карте, мотивируемое ложным представлением о руслах рек как обязательных естественных границах древних этнических территорий и мн. др., — все это делает «графоаналитический метод» В. М. Стецюка совершенно непригодным для целей палеоэтногеографии и способно внушить лингвисту, недостаточно знакомому с возможностями статистики, лишь отрицательное отношение ко всякого рода количествен-

ным операциям на языковом (лексическом) материале, а тем более в глоттогенетических исследованиях.

Из сказанного о «графоаналитическом методе» целесообразно извлечь некоторые уроки, которыми следует руководствоваться в дальнейшем, изложив их здесь в намеренно догматическом ключе. Насколько можно судить по упомянутой работе, не все они являются столь уж очевидной банальностью.

Вера в безграничное могущество статистики неосновательна. Цифры способна служить сильным подспорьем в «измерении» отношений между языками, в том числе (и в нашем случае это наиболее существенно) родственными, но статистическое обследование языков — это не гарантия и даже не основа для правильных выводов об истории их носителей (соответствующих этносов).

Статистические данные о языках, в частности о лексике, никоим образом нельзя напрямую связывать с палеоэтногеографией. К решению проблем исторической лингвогеографии лингвостатистика может привлекаться только при учете множества иных факторов, оказывающих влияние на конкретную ландшафтную приуроченность изоглосс.

Лексикостатистика, как и количественные сведения, касающиеся других языковых уровней, не должна сводиться к арифметике, то есть представлять собою простой подсчет слов (форм, «явлений» и т. д.). Абсолютные цифры не отражают глубинных отношений между языками, поскольку зависят от множества внеязыковых моментов, ни характер действия, ни сам круг которых не «вычислим» вполне однозначно.

Генетическая близость языков по данным лексики может определяться только при обращении к праязыковому лексическому фонду. Все лексические межъязыковые схождения, относительно которых могут возникнуть подозрения в послепраязыковом происхождении, в результате более поздних этнических и языковых контактов, должны быть безусловно изъяты из статистического анализа глоттогенетической направленности.

Важнейшее правило глоттогенетических построений — безусловная ориентация на диалектный материал. Опора только на данные словарей литературных языков в решении проблем языковой «доистории» ведет к абберациям историко-лингвистического зрения и непоправимым перекосам в итоговой картине древнейших межъязыковых (диалектных на уровне праязыка) связей. Исключения из этого правила допустимы лишь по отношению к мертвым идиомам, известным только по письменным свидетельствам, когда у исследователя просто нет возможности выбрать между данными литературных языков и диалектов — живой народной речи.

Строгий отбор лексики к лингвостатистическим подсчетам в целях выявления степени родственной близости языков должен касаться генетической чистоты материала, но не должен ассоциироваться с его тематической фильтрацией. Тематический и семантический (равно как и грамматический — частеречный и проч.) отбор лексического материала, даже (а может быть, и в особенности) если он нацелен на выделение чрезвычайно важных в культурно-историческом отношении слов (хозяйственная, в частности ремесленная, терминология, ритуальная лексика, социальная номенклатура и т. п.), может привести к срабатыванию неравномерного и рассогласованного развития различных сторон культуры, вследствие чего в итоговой картине также могут случиться системные, хотя и не предсказуемые, перекосы. Всякого рода тематический отбор в компаративных исследованиях, где применяется лексикостатистика (ср.: [Историческая типология, 1986, стр. 197], где попарная близость славянских языков оценивается подсчетом прилагательных со значением размера), по существу противоречит элементарному требованию статистических наблюдений: наиболее достоверные результаты могут быть получены только с помощью анализа случайной выборки.

Излагая цели настоящей работы и выбранный способ их достижения, коснемся иных методов, применяемых компаративистикой, которые опираются на квантитативный анализ лексических данных.

2. Глоттохронология и близкие ей методы

...хотя врачи утверждают, будто я давно выздоровел, — до сих пор не могу с точностью и определенно судить ни о чем, что хоть в малейшей степени связано с понятием время. Мне представляется, у нас с ним, со временем, какая-то неразбериха, путаница, все не столь хорошо, как могло бы быть. Наши календари слишком условны и цифры, которые там написаны, ничего не означают и ничем не обеспечены, подобно фальшивым деньгам.

Саша Соколов. «Школа для дураков»

Лексикостатистический подход к оценке генетической близости языков, заявленный в названии настоящей работы, не должен отождествляться с методом глоттохронологии М. Сводеша. 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 М. Сводеша оказывается немного. Воспользуемся, однако, возможностью сравнения с глоттохронологией для того, чтобы отчетливее выявить особенности нашего подхода.

Метод М. Сводеша хорошо известен, но напомним, в чем он состоит. В каждом языке можно выделить так называемый «базовый словарь» (basic vocabulary), в который входят слова, обозначающие некие универсальные, общечеловеческие понятия, не изменяющиеся от языка к языку в зависимости от типа культуры (сомнения на этот счет см. в работах: [Хаймз, 1960, стр. 7; Лигети, 1971, стр. 30]). В разных авторских версиях этот список насчитывает 215, 200 и 100 элементов (в более ранней работе М. Сводеша с попыткой классификации 30 индейских языков фигурировал список в 165 единиц, не претендовавший, однако, на упомянутую универсальность, см.: [Лернер, 1973, стр. 12–13]). В конце концов был оставлен стословный список, в котором понятия отличались наименьшей зависимостью от культурного типа. В него вошли местоимения 'я', 'ты', 'мы', 'тот', 'весь' и др., числительные 'один', 'два' (уже числительные, обозначающие большие количества, могут отсутствовать в некоторых языках или быть культурными заимствованиями), названия частей тела 'рука', 'нога', 'голова', 'глаз', 'ухо', 'сердце', 'печень', 'хвост' и др., элементарных действий 'есть', 'пить', 'видеть', 'слышать', 'спать', 'умереть', 'идти', 'лететь', 'стоять' и др., цветовые обозначения 'белый', 'черный', 'красный', 'желтый', 'зеленый', размерные прилагательные 'большой', 'маленький', 'длинный', названия таких универсальных реалий, как 'солнце', 'луна', 'камень', 'вода', 'земля', 'огонь', 'дым', 'гора', 'дорога' и т. д. Эти понятия и обозначающие их слова отличаются чрезвычайной устойчивостью, скорость изменения базового словаря очень невелика. Тем не менее и он подвержен разрушению. Его эрозия может быть измерена в процентах сохранившейся лексики по отношению ко всему объему базового словаря за определенный период. Оказалось, что эта величина сходна для многих языков и колеблется около 86% за одно тысячелетие. Предполагается, что скорость распада базового словаря постоянна, и это дает основание, по аналогии с радиоактивным распадом и радиоуглеродным датированием, выработать формулу, с помощью которой можно измерять хронологическую глубину расхождения между двумя данными языками:

$$t(A,B) = \log c(A,B) / \log 2r, \quad (1)$$

где t — возраст дивергенции языков А и В в тысячелетиях, c — процент общих для них слов базового вокабулярия, r — индекс сохранности слов базового словаря за одно тысячелетие.

Впоследствии неоднократно предпринимались попытки усовершенствовать глоттохронологический метод или заменить его альтернативными подходами, см. например: [Арапов — Херц, 1974; Иванов, 1990б]. Одна из самых последних попыток этого рода — глоттохронология С. А. Старостина, см.: [Терентьев, 1991, стр. 174]. С. А. Старостин полагает, что коэффициент изменения базовой лексики составляет не 14%, как у Сводеша, а всего 5%, и выводит новую формулу:

$$t = \sqrt{\ln N(t) / -\lambda \times N(t) \times N_0}, \quad (2)$$

где t — время в тысячелетиях, λ — константа распада базового словаря, $N(t)$ — доля сохранившихся слов базового словаря в одном языке.

Следует отметить, что к стословному тесту М. Сводеша нередко прибегают для оценки взаимной близости отдельных языков и языковых групп (семей), без установки на непереносимое датирование их расхождения, ограничиваясь только указанием на процент корневых совпадений.

Одно из таких применений лексикостатистики описывается С. А. Старостиным; оно сообщено ему (письменно) С. Е. Яхонтовым. Яхонтов модифицировал список Сводеша, исключив из него понятия 'весь', 'течь', 'ноготь', 'пора', 'кусать', 'лежать', 'мы', 'перо', 'семя', 'теплый', но добавив словарные значения 'близкий', 'ветер', 'год', 'далеко', 'змея', 'короткий', 'соль', 'тонкий', 'тяжелый', 'червь'. Далее «он разбил 100-словный список на два: 35-словный и 65-словный, причем к 35-словному отнесены наиболее устойчивые слова. 35-словный список С. Е. Яхонтова выглядит следующим образом: «ветер, вода, вошь, глаз, год, дать, два, знать, зуб, имя, камень, кость, кровь, кто, луна, новый, нос, огонь, один, полный, рог, рука, рыба, собака, солнце, соль, ты, умереть, ухо, хвост, что, это, я, язык, яйцо». Если два языка действительно родственны, то процент совпадений в пределах списка 35 слов должен быть выше, чем процент совпадений в пределах остальной части 100-слового списка. Если же процент совпадений одинаков (или процент совпадений в пределах 65-слового списка больше, чем процент совпадений в пределах 35 слов) — то сходство между языками случайное (т. е. налицо либо случайные совпадения, либо результат активных контактов и заимствований). Эта процедура хорошо работает во многих известных нам случаях; в частности, она позволяет констатировать, что

между китайским и тайскими языками исконное родство отсутствует, несмотря на довольно большое количество совпадений в пределах 100-словного списка» [Старостин, 1991, стр. 59-60] ¹.

Отбор лексики в 35-словный список у С. Е. Яхонтова несколько отдаёт шаманством, во всяком случае он недостаточно мотивирован (не доказано, что остающиеся за его пределами слова измеримо менее устойчивы) и не объяснено, каким образом «нащупана» пропорция 35 : 65. Понятно, что она найдена экспериментально, но можно сомневаться в том, что результаты будут одинаково надежными для всех ситуаций. К тому же в условиях самого сравнения по процедуре С. Е. Яхонтова таится опасность констатации отсутствия родства между диалектами одного языка или между ближайше родственными языками, у которых на 100% совпадут оба списка — процент совпадений в 35-словном реестре получается не выше процента совпадений в 65-словном, как то требуется для установления исконого родства. Наш «прогон» русской и украинской лексики по методу С. Е. Яхонтова показал, что между этими двумя восточнославянскими языками отсутствуют отношения действительного родства! Следовательно, версия С. Е. Яхонтова пригодна лишь для языков, характеризующихся сравнительно отдаленным родством.

Цель настоящей работы несколько отлична от целей глоттохронологических исследований. Она состоит в выработке метода, основанного на лексикостатистических данных, с помощью которого можно «измерять» родство языков (точнее, идиомов — лингвистических объектов в принципе любого уровня генеалогической классификации, от говоров одного языка до родственных языковых семей, например, внутри ностратической гиперсемьи), и применении его к славянской ветви индоевропейской семьи. Это означает, что, в отличие от глоттохронологических исследований в духе М. Сводеша, мы не ставим перед собой задачи абсолютного датирования этапов «распада» праязыка (в нашем случае — праславянского). Если у М. Сводеша и сторонников его метода установление картины языкового

¹ Поправки С. Е. Яхонтова к диагностическому списку М. Сводеша — далеко не единственный пример растущего сознания необходимости дифференцированного подхода к лексике, кладущейся в основу глоттохронологических и — шире — классификационных построений: даже среди элементов базового словаря, например, наименований частей тела, проводится разграничение центральной (ядерной) и периферийной лексики (см., например: [Дерфер, 1978; Дерфер, 1981]), испытываются способы выявления устойчивых и неустойчивых слов (в частности, [Шайкевич, 1980]; см. ниже), оцениваются возможности учета их частотности, многозначности, вовлеченности в фразеологию и т. д. Уже сама широта «фронта» такого рода поисков свидетельствует о том, что метод Сводеша не в состоянии удовлетворить потребности компаративистики и в сущности бесповоротно устарел.

родства является хотя и необходимо важной, но лишь частью искомого результата (конечная цель все-таки хронологизация «распада»), то для нас главный предмет будет составлять именно картина родственных отношений языков без обязательного вписывания ее в точные хронологические рамки.

Отказ от датировки дивергентного развития славянских языков вызван прежде всего сохраняющимися сомнениями относительно доказуемости постулата о постоянстве скорости, с которой меняется базовый словарь (ср.: [Пуллиблэнк, 1972, стр. 203–206]); в формулировке Х. Бирнбаума: «представляется сомнительным, что любое допущение относительно декларируемой стабильности существенных языковых изменений <...> может быть доказано. Иными словами какая-либо аналогия с радиоуглеродным анализом, используемым для подтверждения определяемых величин, должна квалифицироваться как иллюзорная» [Бирнбаум, 1993, стр. 9]. Вяч. Вс. Иванов, однако, полагает, что «эмпирический вывод об одинаковости r (при достаточной протяженности лексич. развития языка, превышающей одно тысячелетие) представляет несомненный интерес для общей теории коммуникации, т. к. он указывает на наличие нек-рых обязат. условий, без соблюдения к-рых нарушилось бы взаимопонимание между членами коллектива, принадлежащими к разным возрастным группам» [Иванов, 1990б, стр. 109–110]). Но это не единственная причина уклонения от абсолютной хронологизации процессов «распада».

Метод М. Сводеша был подвергнут серьезной критике. Не повторяя многих замечаний в адрес глоттохронологии и не стремясь умножить ряды ее хулителей (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, то есть фактическая опора на давно скомпрометированную (ср. выразительное название работы [Хяккинен, 1984]: «Wäre es schon an der Zeit, denn Stammbaum zu fällen?») 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 basic vocabulary, как его определил М. Сводеш, непременно с самого начала един для всех компонентов (диалектов) праязыка, выделяющихся затем в самостоятельные лингвистические образования.

Но даже если признать унитаристскую концепцию безальтернативной, узвимость попыток датирования этапов дивергенции этим не устраняется,

поскольку в таком случае в упрек авторам глоттохронологических работ может быть поставлено пренебрежение тем обстоятельством, что «распад» праязыкового единства, как было бы логично предположить, начинается с низкочастотной и неустойчивой лексики и лишь какое-то время спустя затрагивает сам базовый словарь. Следовательно, полученные оценки времени «распада праязыка» (или, что то же, возраста языков-потомков) неизбежно окажутся заниженными, причем неясно, насколько. Глоттохронология сосредоточивается на списке всего в 100 или 200 наиболее устойчивых слов и начинает свой отсчет дифференциации родственных языков, когда процесс «распада» зашел слишком далеко, раз затронул базовую лексику. Конечно, это лишь диагностический список, и какие-то корреляции между его сохранностью и возрастом языков-потомков есть, но они слишком зыбки.

Предположение о возможном занижении возраста языков можно проиллюстрировать данными М. Чейки и А. Лампрехта, касающимися чешского и словацкого, а также двух серболужицких языков, см.: [Чейка — Лампрехт, 1963, стр. 10]. По их расчетам обе эти пары языков образовались (именно как пары) в результате дифференциации первоначальных чешско-словацкого и лужицкого единств в интервале между 1716 и 1858 годами. На середину этого временного отрезка, то есть наиболее вероятную, согласно методу, дату дифференциации, приходится деятельность Антона Бернолака по реформе литературного словацкого языка, а относительно лужицких языков высказываются обоснованные сомнения в их происхождении из общего пралужицкого корня! Исследование Чейки и Лампрехта проведено более тридцати лет назад, с тех пор стословные базовые словари славянских языков не претерпели заметных изменений, и будь это исследование осуществлено сегодня, указанный интервал сместился бы на 1747—1889 гг., что вызывало бы еще большее несогласие с результатами глоттохронологических измерений. В. Арндт объясняет эффект смещения даты расхождения языков к нашему времени тем, что глоттохронологические работы не учитывают явлений позднего взаимовлияния языков и конвергентных процессов (см.: [Арндт, 1959]; см. также: [Ри, 1958; Плат, 1965, стр. 220]).

К этому можно добавить, что оперируя понятиями «праязык», «распад», «автономное развитие языка» 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 Роберта Музиля. Это впечатление усиливается несколько гротескной формой представления результатов подсчета, принятой в глоттохронологии: начало дивергенции между чешским (!) и венгерским (!) языками (!) приходится на 6677 год до Р. Х. (плюс-минус 998 лет), а между русским (!) и финским (!) – на 6274 год до Р. Х. (плюс-минус 944 года), см.: [Чейка – Лампрехт, 1963, стр. 17]. Применительно к подобным выкладкам можно сказать, заостряя ситуацию, что специалисты по глоттохронологии не знают, что именно случилось, но достаточно уверенно судят о том, как давно это произошло.

Мало ясности вносят и такого рода теоретические поправки: «Ни один язык не существует без диалектных различий внутри его (него! – А. Ж.). 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» [Арапов – Херц, 1974, стр. 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 какой-то «порог» несогласованности? одинаков ли он будет 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 контrovers, сколько, быть может, к компетенции социолингвистики и даже этнопсихологии (впрочем, по вполне понятным причинам интересы социолингвистики до праязыка не простираются²). 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» [Арапов – Херц, 1974, стр. 27].

² 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 идентификации диалектов» [Климов, 1986б, стр. 151].

Ввиду отказа от датировок дивергенции и ограничения лишь задачей воссоздания картины родства наша работа оказывается по своему характеру гораздо ближе к попыткам количественной таксономии языков, предпринимаемых, например, А. Я. Шайкевичем, см.: [Шайкевич, 1980].

Методика, предложенная А. Я. Шайкевичем для измерения сходства (фактически генетического) между языками, отчасти напоминает процедуру классической глоттохронологии. Им рассматриваются два варианта методики. В простейшем из них сходство языков измеряется по формуле:

$$S_{AB} = 100 \sum x_i(AB) / N_A, \quad (3)$$

где x — лексический признак, общий обоим сравниваемым языкам А и В, который состоит в выражении понятия из заданного списка словами одного и того же корня; N_A — общее число выраженных признаков языка А (при условии, что $N_A < N_B$) (если понятие выражается в языке несколькими синонимами, то его выражению, учтенному в подсчетах по формуле, приписывается вес в зависимости от места i в списке синонимов: $1/i$; при разных весах общего корня в двух языках в общую сумму засчитывается меньший показатель).

От глоттохронологического метода квантитативный анализ А. Я. Шайкевича, помимо отсутствия в нем установки на определение возраста языков, отличается большей гибкостью и надежностью результатов. Эти качества обеспечиваются, во-первых, по сути, отказом от универсального списка в 100 понятий и увеличением его более чем в 6 раз (в качестве признаков берутся «627 понятий из первых 14 разделов» известного словаря избранных синонимов основных индоевропейских языков К. Бака: [Бак, 1949]), а во-вторых, внедрением в статистическую процедуру (второй вариант) операции апостериорного взвешивания. Последнее учитывает такие особенности признака, как мера неустойчивости (наиболее устойчивыми являются понятия, выраженные во всех сравниваемых языках одним корнем, наименее — понятия, в каждом языке выраженные только ему присущим корнем) и трудность заимствования (измеряется долей незаимствованных слов у данного признака)³.

³ К наиболее устойчивым и незаимствуемым словам здесь отнеслись числительные от 2 до 9, слова со значениями 'вода', 'солнце', 'мать', 'ухо', 'язык', 'ноготь', 'сердце', 'яйцо', 'лизать', 'соль', 'ядро', 'стоять', 'полный', 'один', 'сто', 'новый', 'ночь', 'месяц'. Слова со значениями 'мальчик', 'ребенок', 'скот', 'сильный', 'завтрак' и др. очень неустойчивы, а 'осел', 'мул', 'кошка', 'лев', 'верблюд', 'капуста', 'вино' и др. легко заимствуются из языка [Шайкевич, 1980, стр. 333—334].

Одним из итогов исследования А. Я. Шайкевича, весьма интересного в деталях и «мелких» наблюдениях, сделанных цепким глазом опытного статистика, явилась графическая схема, в высшей степени наглядно демонстрирующая группировку 34 языков Европы (индоевропейских – 6 или 7 ветвей, финноугорских и баскского) в генетические единства по силам обнаруженного сходства (стр. 335). Она отличается определенной убедительностью и дает автору возможность сделать некоторые любопытные нетривиальные заключения (например, «о необходимости повысить ранг таксона „кельтские языки“»: «следует считать их надгруппой внутри индоевропейских языков, включающей в себя гаэльскую и бриттскую группы», стр. 334). Нужно заметить, что в схеме на стр. 335 по недосмотру автора (или корректора) оказались не отраженными связи древнегреческого с латинским, литовского с древнеисландским, старославянского с древнеанглийским, древневерхнемецким и нидерландским, по своей силе входящие в тот же интервал показателей генетической близости, что и отмеченные в схеме связи латинского с готским и древневерхнемецким, а также между языками бриттской группы с языками гаэльской группы и языков балтийской группы с языками славянской группы (включая русский и латышский, индивидуальные связи которых в схеме также упущены). Интересно, что в парных связях этой мощности фигурируют прежде всего мертвые индоевропейские языки, включенные А. Я. Шайкевичем в схему, – древнегреческий, латинский, древнеирландский, готский, древневерхнемецкий, древнеанглийский, старославянский; группы современных индоевропейских языков, таким образом, связываются между собою относительно большей близостью друг к другу предшествующих языковых состояний, в чем можно увидеть изящное проникновение диахронии в ахроническую схему лингвистического родства (ср.: «...генетическое расстояние между древними языками различных групп в пределах одной семьи существенно меньше, чем между современными представителями тех же групп, и, следовательно, гораздо более очевидно» [Порхомовский, 1989, стр. 106]).

Вернемся, однако, к глоттохронологическим работам. В них проблемы датировки или возможные (либо же невозможные) способы материальной экспликации дивергентных явлений занимают нас не в первую очередь. Ограничившись лишь той частью глоттохронологических исследований, которая имеет своим результатом лежащую вне определенных хронологических ориентиров картину родственных взаимоотношений между языками, выраженных числом, обратим внимание на диапазон колебаний этих цифровых показателей в том случае, когда рассмотрению подвергается значительное количество достаточно близкородственных языков.

Для примера можно обратиться к работам по славянским языкам, которые представляют для нас специальный интерес. В упомянутом уже исследовании Мирека Чейки и Арношта Лампрехта⁴ анализируются в глоттохронологическом ключе одиннадцать современных славянских языков. Как и положено в работах по глоттохронологии, мерой родственной близости в паре языков является процент общей сохраненной ими лексики из стословного диагностического списка М. Сводеша. Наибольший процент сохранности (в парах «чешский & словацкий» и «верхнелужицкий & нижнелужицкий») — 95, наименьший (у пары «украинский & словенский») — 70. Колебания невелики — перепад всего в 25%. Из 26 «ступенек» на шкале родственной близости занятыми оказались 19. Между тем одиннадцать языков могут быть скомбинированы в 55 пар:

$$(n^2 - n) / 2 = (11^2 - 11) / 2 = 55.$$

Значит, в среднем на каждой ступеньке расположится по три пары языков; на деле равномерного распределения, конечно, не наблюдается: на «ступеньках», соответствующих 75% и 82%, уместилось по пяти пар, на «ступеньке» 78% — семь пар, а на 74% — даже восемь пар языков. Это означает, что глоттохронологический стословный список не дал возможности дифференцировать по силе генетической близости пары «чешский & болгарский», «словацкий & болгарский», «польский & болгарский», «верхнелужицкий & русский», «верхнелужицкий & украинский», «русский & болгарский», «русский & словенский» и «украинский & сербохорватский»; для всех этих пар языков констатируется одинаковый уровень генетического сходства. Опытный славист, пожалуй, не удовлетворится такой картиной: ему не будет хватать щекочущего исследовательское воображение разнообразия, которое к тому же и больше отвечает действительности.

Дифференцирующая сила диагностического списка М. Сводеша оказывается довольно невысокой. Происходит это, несомненно, по причинам сугубой его ограниченности: на полигоне в 100 слов не слишком разгуляешься. Если к равномоности отдельных связей генетически достаточно удаленных друг от друга языков (как, например, славянских с германскими: связи чешского и русского с английским и шведским

⁴ Подход Иштвана Фодора к глоттохронологическому анализу славянских языков (см.: [Фодор, 1961а; Фодор, 1961б]) весьма спорен и в отборе лексики сопоставляемых языков, и по своим результатам (кроме справедливых упреков ему в статье Чейки — Лампрехта, стр. 17—18, об исследованиях И. Фодора см. еще: [Арапов — Херц, 1974, стр. 27]).

определяются тождественными показателями в 34%, см.: [Чейка — Лампрехт, 1963, стр. 16]) относишься спокойно (больше того, пестрота связей в таком случае покажется даже подозрительной: с чего бы это русский ближе к романским итальянскому и испанскому, чем чешский?, см. [там же]), то недифференцированность «глоттохронологических» схождения в группе близкородственных языков воспринимается как следствие некоторой слабости метода.

Недостаточно дифференцирующим, к сожалению, оказывается и более пространный список семантических признаков у А. Я. Шайкевича. Те же славянские языки (у Шайкевича их рассмотрено пять: старославянский, сербохорватский, чешский, польский и русский) в своих отношениях генетического порядка демонстрируют унылое однообразие: лексические корреляции литовского с сербохорватским, чешским и русским характеризуются одинаковыми численными показателями, та же картина — для отношений латышского с сербохорватским, польским и русским языками. 627 понятий, лексические выражения которых послужили базой для таксономического разбиения индоевропейских языков у А. Я. Шайкевича, также дают довольно стертый портрет близкородственных языков, в то время как интуитивно такое постное сходство ощущается как следствие недостаточной исходной материальной базы. Возможное возражение, вроде того, что так и должно быть, поскольку речь идет о сходстве на уровне праязыков, следует отвести на том основании, что и праязык гораздо более адекватно реальности представлять себе как диалектно структурированное образование с неравномоными внутренними связями его компонентов.

Выход видится в значительном расширении исходной лексической базы исследования.

3. Расширение исходной базы

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, как и его позднейшие модификации, является вынужденным и в общем не скрываемым компромиссом.

Многие лингвисты, занимающиеся выявлением меры близости между разными идиомами (подсистемами или диалектами одного языка, самостоятельными языками, лингвистическими группировками разных уровней) на материале лексики, ощущают недостаточную эффективность различных тестов, назначенных оценить эту близость с помощью того или иного рода контрольных лексических списков, заданных заранее. В своем месте уже говорилось, что предварительное составление таких диагностических перечней вступает в известное противоречие с требованием статистики, согласно которому наиболее надежные результаты достигаются при обработке случайной выборки из предположительно однородного материала. Попытки преодолеть этот недостаток статистических методик, упомянутых выше, осуществляются в двух направлениях.

Одно из них опирается на анализ целого текста. В. Маньчак, наиболее решительный сторонник такого подхода (см.: [Маньчак, 1958; Маньчак, 1981]), заявляет прямо: «Для того, чтобы оценить, как представляются родственные отношения между данным языком и другими языками, ни в коем случае не следует изучать слов, отмеченных в словарях, а следует сопоставлять тексты» (разрядка наша. — А. Ж.) [Маньчак, 1987, стр. 144]: в этимологическом словаре румынского языка слова славянского происхождения явно преобладают над словами латинского корня, но было бы абсурдом относить на этом основании румынский язык к славянской группе. В текстах количество слов латинского происхождения достигает 80–90%, что с несомненностью доказывает романскую принадлежность румынского языка (аналогично обстоят дела с романским и германским компонентом в английском, с угрофинским и угрофинским компонентом в венгерском). С опорой на текст В. Маньчак устанавливает различную степень близости между некоторыми индоевропейскими языками. Анализ параллельных текстов Евангелия от Матфея позволяет близость польского и немецкого языков выразить числом 108, польского и литовского — 144, польского и болгарского — 306 (там же, стр. 146–147), а другое исследование, на материале переводов Евангелия от Луки, устанавливает различную силу связей готского языка: со старославянским — 157, литовским — 113, латинским — 67, чем, по мнению В. Маньчака, доказывается большая близость германских языков к славянским, чем к балтийским ([Маньчак, 1986]; с последним утверждением, пожалуй, не следует торопиться соглашаться, поскольку литовский, несмотря на свою знаменитую и образцовую архаичность, — все же современный язык, тогда как готский и старославянский принадлежат к языкам мертвым и находящимся ближе к собственным предшест-

вующим состояниям — германскому и славянскому праязыкам; ср. еще применение статистического анализа параллельных текстов в подходе к балто-славянской проблеме: [Маньчак, 1983]; ср.: [Поповская-Таборская, 1991, стр. 34–35, 94–97]).

Квантитативным наброскам В. Маньчака, усилия которого оставляют впечатление ломления в открытую дверь, все-таки недостает строгости — не в самих подсчетах (в добросовестности его сомневаться не приходится), а в отсутствии статистических оценок достоверности результатов, тем более необходимых, что результаты эти получаются автором путем просчитывания сравнительно небольших текстов (в отличие, например, от затрагиваемого ниже исследования В. З. Санникова).

К анализу текстов прибегают и в соображениях усовершенствования методики глоттохронологии (см.: [Терентьев, 1991, стр. 174] — о лексикостатистической методике, предложенной С. А. Старостиным).

Другой путь преодоления недостатков традиционных лексикостатистических методик, имеющих дело с диагностическими списками слов, состоит в обращении к сплошному статистическому анализу словаря. На наш взгляд, именно такой подход наиболее плодотворен и способен придать квантитативным оценкам лексической близости между идиомами неоспоримую точность, хотя он неизмеримо более трудоемок.

В качестве примера сошлемся на весьма тщательное исследование В. З. Санникова [Санников, 1975]. Рассматривая вопрос о степени близости восточнославянских письменных языков в определенный период их истории (XI–XVI вв.), В. З. Санников исходил из необходимости «сопоставления целых лексических систем, включающих все слова текста, а не только слова основного словарного фонда (basic vocabulary М. Сводеша)» (стр. 7; разрядка наша. — А. Ж.). Он полагает, что измерять степень близости числом или долей слов, общих для данных двух идиомов (предположим, языка грамот XV в. и языка грамот XVI в.), — недостаточно. Нужно учитывать частоту слова: «если какое-то слово является общим для грамот XV в. и грамот XVI в., но в первых встретилось 1–2 раза, а во вторых — 1 000 раз, то этот факт, бесспорно, не менее существен с точки зрения лексической близости, чем отсутствие в одной системе слова, которое в единичных случаях встретилось в другой системе» (стр. 10). Поэтому для характеристики степени лексической близости двух сопоставляемых систем более важной является доля общих словоупотреблений, которая определяется для каждого слова отдельно и равна меньшей из двух сопостав-

ляемых частот. Большие по объему тексты (В. З. Санниковым в основном эксперименте работы сопоставлялись лексические системы древнерусских юридических текстов XI–XIV вв., старорусских и западнорусских XV–XVI вв., отраженных в памятниках следующих суммарных объемов соответственно: около 70 тысяч, более 100 тысяч и более 100 тысяч словоупотреблений) обрабатываются по методу ранговой корреляции, где ранжируются не отдельные слова, упорядоченные по убыванию частот, а группы слов (50 или 100 единиц в группе). Используется формула коэффициента ранговой корреляции Спирмена

$$R = 1 - 6 \sum d^2 / (n^3 - n) \quad (4)$$

или, для групп лексических элементов,

$$R = 1 - 6m^2 \sum d^2 / (n^3 - n), \quad (5)$$

где R – коэффициент ранговой корреляции, n – число общих элементов, d – разность между рангами (порядковыми номерами), m – число элементов в одной группе.

Так, выясняется, что ранговая корреляция лексики старорусских текстов XV–XVI вв. с лексикой древнерусских текстов XI–XIV вв. составляет $R = 0,686$, а та же величина для западнорусских текстов XV–XVI вв. и древнерусских текстов определяется как $R = 0,576$ [Санников, 1985, стр. 162]. Простой и удобный для предварительных оценок существующих связей, коэффициент ранговой корреляции Спирмена все же уступает по ряду характеристик другим статистическим методам, в частности, методу Кендэла, более пригодного для детального и углубленного анализа связей (об этом см., например: [Методика и техника статистической обработки информации, 1968, стр. 246–247]), хотя, ссылаясь на Б. Л. ван дер Вардена [ван дер Варден, 1960], В. З. Санников придерживается иного мнения.

Заслуживают интереса и исследования, определяющие взаимную близость идиомов на основе статистического обследования сплошной выборки из массива лексических единиц неопределенно большого объема. Так, Н. И. Батожок [Батожок, 1986] строит свои подсчеты на выборке в 1 000 словарных единиц для установления меры близости между говорами всех двадцати районов Брянской области (по материалам 4-го вып. «Словаря брянских говоров», буква «Г»). Мера близости определяется в этой работе как

$$\alpha_{ij} = \varepsilon_{ij} - 1, \quad (6)$$

где относительная оценка близости говоров i -ого и j -ого районов

$$\varepsilon_{ij} = n_{ij} \times N / n_i \times n_j, \quad (7)$$

где, в свою очередь, N — объем выборки из словаря (= 1000), а n_i — число словарных единиц в i -ом районе, а n_{ij} — словарных единиц, общих для говоров i -ого и j -ого районов. В предложенных вычислениях весьма важно отнесение числа общих слов, отмеченных в двух данных говорах, к объемам дифференциальной (= диалектной, не фиксируемой в литературном языке, но не дифференциальной по отношению к другим диалектным подсистемам!) лексики, отмеченной в этих говорах. Этим последняя работа выгодно отличается от прямых сличений цифр в исследованиях, аналогичных злополучной статье В. М. Стецкока (см. выше).

Нетрудно заметить, что лексикостатистическим исследованиям, ставящим своей задачей выявление степени близости между разными идиомами и опирающимися на статистическую обработку целых лексических систем или сплошных выборок из них значительного объема, свойственна несравнимо большая тонкость и детализация в оценках этой близости, чем работам, которые основываются на коротких диагностических списках.

Использование ряда формул, которые можно встретить в лингвостатистической литературе и которые рекомендуются для измерения лексической близости языков (см.: [Тарнацкий, 1937, стр. 36; Левин, 1964, стр. 115; Соколовская, 1968, стр. 102, 111; Москович, 1968, стр. 133]), предполагает обращение к узким терминологическим группам и ориентировано в основном на установление сходств и различий в семантической организации этих групп (с возможной реинтерпретацией семантико-типологического сходства в глоттогенетическом плане). Применение этих формул требует строгого семантико-типологического анализа лексики на единообразных основаниях, что, при обширности привлекаемого материала и, как правило, неравнокачественности его источников, вряд ли осуществимо. Кроме того, как мы отмечали выше, в работах, аналогичных нашей, ограничение материала по тематическому принципу неудовлетворительно с точки зрения статистики, поскольку лишает выборку ее необходимейшего качества — случайности (в статистическом смысле).

Существуют и иные способы оценки близости между языками и диалектами, некоторые из них рассматриваются их авторами как промежуточные звенья в построении диалектных классификаций (в первую очередь необходимо указать на важные работы Н. Н. Пшеничновой по статистической группировке русских диалектов и опыты статистической классификации диалектов армянского языка, принадлежащие Г. Б. Джаукяну: [Пшеничнова, 1976, 1977, 1979, 1981, 1987; Джаукян, 1972 и др.]). Они пригодны и для оценок языковых сходств на лексическом материале, но не являются специально лексикологическими методиками, поэтому здесь мы рассматривать их не станем.

Глава 3

Поиск метода

1. Некоторые отправные положения

Господа, бросьте карты и станем считать! Я подам первый пример: дайте мне кусок мелу!..
Но что считать?.. Вот запятая!

Барон Брамбеус. «Арифметика»

Вначале — несколько отправных положений, на основе которых возможно построение лексикостатистической модели языкового родства.

Мы исходим из признания лексикостатистических методов наиболее объективным и надежным способом установления степени генетической близости между языками. Этот взгляд не является оригинальным. В тех же выражениях (и даже безапелляционнее: «единственным объективным»!) дает им оценку С. А. Старостин (см.: [Старостин, 1991, стр. 25]), однако у него лексикостатистика остается в рамках глоттохронологических процедур в духе М. Сводеша, к которым мы, как нетрудно, видимо, было заметить, относимся более скептически — по причинам, затронутым в предыдущей главе (прежде всего малый объем испытательного списка и вытекающая из этого невысокая надежность результатов, далее — опора на представление о генеалогическом древе).

Единственно надежной материальной базой глоттогенетических конструкций на лексикостатистической основе являются, как мы уже отмечали, праязыковые словники современных языков, то есть своды родственно-языковых лексических корреспонденций, возводимых к единым праязыковым реконструкциям. Таким образом, за пределы статистических процедур должны быть выведены все неунаследованные слова — лексические образо-

вания эпохи «после распада» праязыка, поздние (послепраязыковые) совместные и индивидуальные заимствования из неродственных языков (точнее, из языков, не входящих в рассматриваемую семью или группу) и, что необходимо подчеркнуть, результаты взаимного обмена лексикой между данными языками после распада их праязыкового единства, хотя бы эти лексемы в языке — источнике иррадиации и восходили к праязыковому фонду. Этим требованием предлагаемая модель отличается от модели А. Я. Шайкевича, который включал в компаративное рассмотрение позднюю и заимствуемую лексику (помимо смыслов, упомянутых на стр. 41, сноска, — 'картофель', 'шелк', 'хлопок', 'налог', 'магазин', 'карман'...), по каковой причине ему понадобилось ввести в анализ процедуру «апостериорного взвешивания» для усиления главной компоненты полученной меры сходства — генетической близости языков.

Важным мы считаем тезис, который, строго говоря, не является доказанным, а скорее относится к убеждениям, основанным на интуиции. Нам не известны способы его безупречного доказательства (возможно, они существуют), и в нашей модели мы отводим ему примерно такую же роль, какую играет в глоттохронологии М. Сводеша постулат о постоянстве скорости распада *basic vocabulary*. Мы полагаем, что количественные соотношения между праязыковыми словарями современных (и в меньшей степени — «исторических», известных лишь по памятникам) языков, точнее говоря, пропорции между лексикой, распространенной во всех сравниваемых языках (в нашем случае — общеславянской), лексикой, известной лишь каким-то определенным их группам, и, наконец, узкими праязыковыми лексическими регионализмами данного языка, в целом с достаточной степенью приближенности отражают реальные связи их протоидиомов — диалектов самого праязыка.

Во-первых, лексические утраты, как мы пытались показать, не являются неисследимыми и столь обширными, как это обычно полагают. Слова редко исчезают из словаря совершенно бесследно: очень большой пласт древней лексики может быть восстановлен путем анализа производных, фразеологии, данных диалектов и лексики памятников (если они, разумеется, есть). Количественные соотношения лексики, разной по характеру и по направлению ее внешних для данного протоидиома связей, в истории этого идиома затемняются не столько утратами, сколько временными наслоениями, безотчетно переносимыми на характеристику древних состояний языка.

Во-вторых, можно думать, что эрозия и обновление различных групп лексики со сходными типологическими характеристиками (универсаль-

ная — специализированная, общераспространенная — узколокальная, нейтральная — окрашенная и т. д.) в разных, но близкородственных языках протекают сходным образом и их пропорции в разных языках приблизительно одинаковы, более того — в общем остаются прежними, если только речь идет о сохранным (поддающемся реконструкции) праязыковом лексическом фонде, а не о пластах словаря, формирующихся в разных языках уже после разложения праязыкового единства.

В-третьих, мы собираемся иметь дело со славянскими языками. Как известно, в индоевропейской семье они являются самыми молодыми¹, и за 12—14 веков их развития «после распада» праславянского языка они просто не смогли измениться настолько, чтобы мы сильно сомневались в возможностях относительно полной реконструкции праязыкового лексического фонда. В силу своей молодости и, в основных деталях, ясности исторических преобразований, приведших к ее нынешнему состоянию, славянская языковая семья представляет собою исключительно удобный полигон для выработки и уточнения приемов компарации (в том числе и статистических), в некоторых отношениях даже более удобный, чем так же довольно молодая и хорошо засвидетельствованная семья романских языков. Именно эти обстоятельства сообщают славистике «режим наибольшего благоприятствования», приведший к особому ее положению среди отраслей сравнительно-исторического языкознания, особенно в том, что касается этимологии².

Конечно, положение разных славянских языков неодинаково, и не все они дают возможность полно восстановить лексику их протоидиомов на уровне праславянского существования. Прежде всего это касается языков мертвых. Старославянский никогда не был народно-разговорным языком, поэтому рассчитывать на то, что в старославянских текстах, весьма ограниченном объеме и тематикой, мы обнаружим достаточное количество праславянских лексем, составлявших словарь южнославян-

¹ Имеется в виду возраст самих отдельных языков, а не «возраст» славянского компонента индоевропейской семьи, который, разумеется, уходит в гораздо более глубокую древность.

² Следует, однако, с огорчением отметить, что слависты не всегда должным образом пользуются такими льготными условиями. В частности, не лучшим, на наш взгляд, образом обстоят дела с лингвогеографией и созданием региональных атласов: наряду с явным преимуществом в том отношении, что завершена работа над Общеславянским лингвистическим атласом, здесь есть и очень ощутимые пустоты, по крайней мере романстика и германстика значительно обходят славяноведение по количеству картографических изданий (см.: [Йордан, 1971, стр. 217—390; Мельникова — Сухачев, 1971; Сухачев, 1974, стр. 34—36]).

ских диалектов, на базе которых он был создан, попросту не приходится. Во многом это же относится и к другим письменным языкам, в частности, древнерусскому. Наши сведения о лексике полабского языка чрезвычайно скудны, и об этом можно только сожалеть.

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 лексическом массиве не менее двух десятков тысяч единиц, входящих в реконструируемый словарный состав праславянского языка, включая в него и праславянские диалектизмы. Оценки Т. Лер-Сплавинского, Фр. Копечного и др. в 1 700, 2 000, 9 000 единиц (ср.: [Лер-Сплавинский – Курашкевич – Славский, 1954, стр. 24; Лер-Сплавинский, 1954, стр. 64; Ондруш, 1976б, стр. 299–300; Историческая типология, 1986, стр. 200]) могут считаться безнадежно устаревшими. Таким образом, объемом материала, привлекаемого для статистического анализа, должен возрасти по сравнению со списком М. 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. Праязыковой лексический фонд – основа надежной статистики

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 взаимное родство славянских языков попарно не на крайне ограниченном материале «универсального» диагностического «базового» списка в 165–215–200–100 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. Они позволяют если не избежать полностью, то свести к мини-

муму опасность, от которой не гарантированы (более того — на которую почти обречены) подсчеты по методу М. Сводеша. Имеется в виду бессилие глоттохронологии в ее классическом варианте перед явлениями лексического взаимопроникновения в языках-потомках. Разумеется, и современный этимологический анализ и историческая лексикология конкретных языков не обеспечивают стопроцентной уверенности в том, что перед нами именно только унаследованная праязыковая лексика, а не следствие взаимовлияний эпохи «после распада» — результаты межъязыковых ареальных контактов, функционирования в составе надъязыковых объединений типа языковых союзов и т. п. Однако, во-первых, нынешнее состояние славянской этимологии следует оценивать очень высоко (что уже отмечалось), и в своем большинстве 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.

Поясним сказанное на примере. Праслав. **kradějь* отражено в укр. диал. *kradij* 'вор' и словин. *kražej* 'то же' (ЭССЯ, вып. 12, стр. 84). Наличие в старопольском языке производного *kradziejstwo* 'злодейство, воровство' (см.: [Славский, т. III, тетр. I(II), стр. 45]) позволяет заподозрить в украинском примере результат взаимодействия соседствующих лехитского и восточнославянского ареалов, хотя это подозрение остужается территориальной удаленностью украинской фиксации (лубенское). Но вот аналогичные опасения относительно укр. *клишáвий* 'косолапый' и белор. *клишáвы* 'то же'; белор. *лѣтацца* 'про период половой активности домашних животных'; белор. *лѣтáци* 'прошлогодний' имеют под собой большую основательность, ср. их территориальную близость соответственно к польск. стар. диал. *kliszawy* 'изуродованный, увечный', 'кривоногий', 'безрукий'; польск. диал. *latac* 'о половом влечении коз', словин. *lata się, la-tu-je się* 'проявлять половое влечение — об овцах'; ст.-польск. и польск. диал. *latosi* 'нынешнего года', словин. *latosi* 'прошлогодний' (см.: **klišati* [ЭССЯ, вып. 10, стр. 48]; **lětati* II [вып. 14, стр. 267]; **lětosnjь* [вып. 15, стр. 14]). Ареальная связь между восточнославянскими и лехитскими словами не несомненна, но довольно вероятна, не исключено, что современные ареалы этих образований — не результаты параллельного сохранения в лехитских и восточнославянских языках рефлексов праславянских форм, а итог лексического взаимодействия соседних групп родственных языков уже после разложения праязыкового единства. При возникновении подобных колебаний исследователь может либо включить примеры в подсчет, либо вычеркнуть

их из списка соответствий. Для глоттохронологии, с ее «малой» статистикой, выбор того или другого решения чреват существенными последствиями. Для «большой» же статистики, какая бы альтернатива ни была предпочтена, выбор окажет в целом незначительное влияние на конечный итог, тем более что ошибки и в ту и в другую сторону могут взаимно уравновесить друг друга.

Не следует думать, что подобные сомнения затрагивают только узкорегionalную лексику, а «универсальный» диагностический перечень М. Сводеша, в силу особой устойчивости его составляющих, от опасностей такого рода гарантирован. При глоттохронологическом анализе близкородственных русского и украинского языка исследователь, использующий стословный список заданных смыслов, может оказаться перед трудными дилеммами, решая, к расхождениям или, напротив, к сходствам между этими языками относить такие явления, как русск. *человек* – укр. *людина* (укр. *чоловік* – ‘мужчина’, ‘муж’, но и – реже – ‘человек’, ср. *чоловіче* ‘добрый человек’, звательная форма, и под.), русск. *хороший* – укр. *гарний, добрий* (гораздо реже укр. *хороший*), русск. *большой* – укр. *великий* (при русск. диал. *великий* ‘большой’ и «просвечивании» этого значения у прилагательного *великий* в устойчивых словосочетаниях вроде *велика Федора*, в образованиях с отрицанием – *невелик* и т. д.; при укр. форме сравнительной степени *більший, більш* и т. д.), русск. *красный* – укр. *червоний* (укр. *красний* ‘красный’ – диалектное), русск. *этот (эта, это)* – укр. *цей (ця, це)* (при наличии укр. *отой (ота, оте)*, возводимого к той же праславянской реконструкции **e to* (**e ta*, **e to*)) и под. Следует ли учитывать явно рецессивные формы и синонимы как продолжающие исконное сходство или же расценивать их как сохранившиеся (если не заимствованные) по причине поддержки, то есть все-таки влияния, близкородственного соседа, вследствие ареального контактирования? Решить это можно только опираясь на тщательнейшее историческое расследование. Учет или же, напротив, упущение одного-двух (!) словарных сходжений в глоттохронологическом метрировании может привести к значительным темпоральным смещениям: слишком уж куц предложенный М. Сводешем лексический тест. В глоттохронологической процедуре требуется решительный выбор между несколькими возможными вариантами, а такая решительность сильно сказывается на результате. При «большой» же статистике, которая предлагается в нашей методике определения уровней родства, зависимость конечного результата от предпочтенных в каждом конкретном случае альтернатив (в особенности таких, как варианты решения дилеммы «прямо унаследованное – появившееся вследствие ареальных контактов») далеко

не столь велика: спорные случаи как бы растворяются в массе более надежного материала, подвергаемого статистической обработке, и этим сильно смягчается воздействие допущенных погрешностей.

Необходимо сделать еще одно уточнение. Здесь мы оперируем словами разных языков как формальными единицами, обращаясь к их семантике лишь для констатации их этимологического тождества. Сколь бы далеко ни разошлись слова разных языков по значению (ср. семантику продолжений праслав. **brusъnica*: 'брусника', 'оспа', 'место, где добывается точильный камень'..., **grebja* 'тропинка', 'кочерга', 'уборка сена'..., **lazьba* 'поросшая травой долина', 'баня', 'выбирание меда из ульев'... и т. п.), семантические различия не являются препятствием для идентификации их праформ. Иначе поступают другие авторы, которые в поисках материала обращаются не к этимологическим (и диалектным), а к переводным словарям (см., например, упоминавшуюся работу: [Стецок, 1987]). При выборе переводных словарей в качестве источника для сравнения лексики разных языков, помимо того, что таким образом упускается огромный и чрезвычайно важный, если не сказать основной, словесный пласт, задается иное методологическое основание этого сравнения, а именно в поиске сходств и различий между родственными языками точкой отсчета назначается лексическая семантика.

Такой путь, однако, чреват серьезными уступками модернизации семантических отношений между (поздние) праславянскими диалектами и потерей множества реальных межславянских корреспондентий. Переводные словари предлагают не лексическое, но семантическое соответствие (либо, с учетом полисемии, несколько соответствий), и при этом вовсе не гарантируется материальное историческое тождество сопоставляемых лексем. Достаточно упомянуть классические, переходящие из одного учебного пособия по семаσιологии в другое, случаи позднейшего семантического расхождения таких праславянских лексических реконструкций, как **bystrъ(jь)*, **časъ, čьrstvъ(jь)*, **divigati, *gora, *neděl'a, *trudъ, *životъ* и мн. др.).

В подобных случаях вполне естественно ожидать возражения: семантические несовпадения как раз и свидетельствуют о генетическом расхождении, историческом удалении языков друг от друга — дивергенции, следовательно, их учет также существен в установлении меры родства: чем больше таких несовпадений, тем раньше разошлись языки и тем ниже уровень их генетической близости. Это так. Но кто сможет не обвиняясь установить четкую демаркационную линию между разошедшимися лексическими значениями и вариациями одного и того же значения, как определить теоретический порог, перейдя который значения могут признаваться семантическими?

ским несовпадением? 'Гора' и 'лес', в самом деле, далеко не одно и то же, но к какому из этих семантических полюсов «прислонить» значение 'гора, покрытая лесом'? К какому значению, 'корыто' или 'речное русло' ближе третье – 'желоб', если все они обозначаются одной и той же лексемой – **koryto* (с первым 'желоб' роднит физическая форма, материал и артификативность, со вторым – то, что оба являются ложем для текущей воды)? Сходство или, напротив, различие усматривать в значениях отражений праслав. **děťva* в разных славянских языках: 'дети, детвора' и 'личинки пчел' (да еще подключив в порядке развертывания «метафоры» – а это и не метафора! – пчелиную *матку*)? Переводные и этимологические словари склонны представлять лексико-семантические системы принципиально различным образом: первые – аналитично, в атомизированном состоянии, вторые – континуально. Этимологические словари, перечисляя лексические соответствия, дают при них целую шкалу тонких семантических взаимоперетеканий, четко разграничить которые на самостоятельные значения не удастся; во всяком случае, этого нельзя сделать априорно, в виде однажды сформулированного правила, которым можно было бы руководствоваться и впредь. Такого рода семантические проблемы приходится решать каждый раз заново, *ad hoc* (но – следует заметить – этимолога отсутствие универсальных критериев семантического тождества/различия отнюдь не огорчает; наоборот, осознание исторического единства далеких лексических значений может доставлять этимологу такое удовольствие и даже радость, какие, по-видимому, трудно испытать, занимаясь иными областями лингвистической науки). Таким образом, не обнаружив в переводном лексиконе семантического соответствия в виде этимологически тождественного слова другого языка, исследователь еще не вправе констатировать в данном пункте лексическое различие между родственными языками.

Вернемся, однако, к статистическим проблемам.

3. Выработка формулы количественного сравнения

Как выявить численную меру языкового родства?

Самый простой, на первый взгляд, способ определить относительную генетическую близость языков друг к другу на основании лексикостатистических сведений – сопоставить число общих данным двум языкам лексем, восходящих к восстановленному праязыковому словарному фонду, с аналогичными цифрами, характеризующими другие пары родственных языков: можно предположить, что идиомы А и В, имеющие в

своих словарях 1 000 общих для них праязыковых лексем, генетически менее близки, чем идиомы А и С, у которых число общих лексем, восходящих к праязыковой эпохе, допустим, 2 000. Это можно записать в виде формулы:

$$G(A,B) = V(A,B), \quad (8)$$

где *G* (лат. *gentilitas* 'родство') – показатель генетической близости, «степени родства», А и В – два данных идиома из всей совокупности исследуемых идиомов, *V* (лат. *verbum* 'слово') – число общих для двух данных идиомов слов. Таким образом, если $V(A,B) > V(A,C)$, то и $G(A,B) > G(A,C)$; если $V(A,B) > V(C,D)$, то $G(A,B) > G(C,D)$ и т. д.

Однако этот наиболее простой способ прямого сличения цифр одновременно наименее надежен и не пригоден для каких бы то ни было количественных сравнений глоттогенетической направленности. Он может работать только в том случае, когда все сравниваемые идиомы, находящиеся в установленном 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: праславянский словник чешского языка вдвое превышает объем аналогичного словника македонского языка, а список праславянских лексем русского – втрое больше, чем праславянский словник нижнедужицкого. Понятно, что количество общих праславянских слов, сохраненных, скажем, польским и полабским языками, заведомо должно быть меньшим, чем количество праславянских слов, общих польскому и русскому языкам, поскольку объем доступного реконструкции праславянского пласта в полабской 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.

Еще очевиднее это обстоятельство выявляется в случае, если подсчету подвергаются только эксклюзивные (сепаратные) связи данных языков с дальнейшим сличением полученных результатов для разных пар идиомов. Несколько забегаая вперед, приведем численные данные о лексических связях указанного рода. В интервале от А до середины буквы L латинского алфавита (объем первых пятнадцати выпусков ЭССЯ) близкородственные пары языков характеризуются следующими абсолютными показателями эксклюзивных лексических связей: болгарский и македонский – 14, верхнелужицкий и нижнелужицкий – 12, украинский и белорусский – 15, в то время как эксклюзивные лексические связи языков, более далеких друг от друга, такие как сербохорватско-русские, словенско-русские и чешско-русские, выражаются абсолютными цифрами 74, 32 и 30 слов соответственно. Причина таких резких несхождений – прежде всего именно в разнице объемов праславянского лексического наследия, выявляемого в каждом из этих языков: чем длиннее праславянские словники сопоставляемых идиомов, тем более вероятным оказывается нахождение в них общих, в том числе сепаратных, элементов.

Впрочем, ситуация с украинским и белорусским, праславянские словники которых достаточно велики, имеет еще одну – довольно парадоксальную – причину. Как ни странно, малое число объединяющих их эксклюзивных лексических связей праславянского происхождения объясняется как раз высокой степенью генетической близости между языками восточнославянской группы. «Незначительность» исключительных словарных корреспонденций праславянского характера между украинским и белорусским коренится в их близком родстве с третьим языком. Если прочие генетически наиболее тесные группировки славянских языков являются парными (болгарский – македонский, сербохорватский – словенский, чешский – словацкий, польский – кашубско-словинский; о полабском и паре серболужицких языков – особо), то восточнославянские языки составляют трицу, и отношения в этой подсистеме складываются иначе; тернарные связи имеют принципиально иную природу; нежели бинарные. Соседний великорусский язык с его гигантским праславянским словарем еще не успел настолько удалиться от белорусского и украинского, чтобы изоглоссы, связывающие два последних языка, сохранились в качестве сепаратных. Взаимная особенная близость трех языков, а не пар, как в других упомянутых случаях, делает двойственные исключительные связи между составляющими этой трицы менее вероятными, чем тройственные, отличающие их от всех прочих родственных языков. По этой же причине великорусско-белорусские (61) и великорусско-украинские (54; напомним, что цифирь

относится к 1–15 вып. ЭССЯ, а не ко всему праславянскому словарю) эксклозивные лексические изоглоссы выглядят скромнее, нежели сепаратные сербохорватско-великорусские лексические параллели.

Из этих наблюдений можно сделать вывод, что при всей чрезвычайной важности сепаратных сходжений между родственными языками им (сходжениям) нельзя придавать абсолютной значимости как свидетельствам особо тесных генетических отношений. Весьма существен фон, на котором разбивается статистический сценарий лингвистического родства. Прямые подсчеты, как бы они ни были соблазнительны, могут дать результаты, очень далекие от действительности и даже извращающие ее. Неучет фона также чреват неверными итогами: переоценкой относительно больших абсолютных показателей сепаратных связей между языками, для которых реконструируются пространные праязыковые словники, и, напротив, решительной недооценкой относительно меньших цифр, характеризующих эксклозивные параллели между языками, реконструируемое праязыковое лексическое наследие которых сравнительно невелико.

Устранение отмеченных искажений не представляет большой сложности. Очевидно, что показательны в данном случае числа лексических совпадений праязыкового происхождения для разных пар языков не в абсолютном выражении, а в отношении к количеству праязыковых лексем, сохраненных каждым сравниваемым языком вообще, то есть к объемам праязыковых словников обоих языков: для славянских языков А и В количество общих праславянских слов должно быть отнесено к произведению числа всех праславянских слов, обнаруженных в языке А, и числа всех праславянских слов, выявленных в языке В (в знаменателе вместо абсолютного числа всех праславянских лексем данного языка можно брать и его доленое выражение — отношение объема праславянского лексического пласта в данном языке к объему всего реконструированного этимологическим анализом праязыкового словаря в целом; выбор того или другого варианта регулируется только удобством масштаба численного выражения конечного результата).

Формула (8) усложняется:

$$G(A, B) = V(A, B) / H(A) \times H(B), \quad (9)$$

где G, А и В обозначают то же, что в формуле (8), а H (лат. *hereditas* — 'наследство') — число всех восстапавливаемых (или привлекаемых к анализу) единиц праязыкового лексического фонда, отмеченных в данном идиоме. Таким образом, если $V(A, B) = V(C, D)$, а произведение $H(A) \times$

$H(B)$ больше произведения $H(C) \times H(D)$, то $G(A, B) < G(C, D)$. Если же $H(A) \times H(B) = H(C) \times H(D)$, а $V(A, B) > V(C, D)$, то $G(A, B) > G(C, D)$.

Но и при таком подсчете полученная картина соотношений, приобретающая большее правдоподобие во многих деталях, в целом все же будет весьма далека от ожидаемой на основании интуитивных представлений. Причиной этому — неравноценность разных конкретных лексических корреспондентий 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 языки. В этом смысле слова **agoda*, **česati*, **gora*, **jъskra* или **lěvъ(jb)*, как бы они ни были интересны этимологу или семасиологу, особой ценности не имеют: будучи известными всем славянским языкам, сами эти формы ничего не дают нам для характеристики взаимной близости отдельных идиомов и их локальных групп (другое дело семантика — здесь во многих случаях обнаруживаются заслуживающие пристального внимания сближения и расхождения; однако по причинам, которые мы отчасти объяснили выше, а отчасти коснемся в другом месте, к анализу семантических изоглосс мы обращаться не будем, сосредоточившись лишь на формальных сходствах и различиях). В то же время такие слова, как **ǰ(ъ)věnъka* (болгарско-верхнелужицкая изоглосса: болг. стар. и диал. *dvěnki* мн. ч. 'два, двое', в.-луж. *dwěnka* 'два, двойка' [ЭССЯ, вып. 5, стр. 188; Журавлев, 1990, ч. 1, стр. 61]), **čьtokosъ* (болгарско-полесская изоглосса: болг. *чернокосъ*, диал. *чернокос* 'дрозд черный, *Turdus merula*', укр. диал. *чорнокос*, *чорнокис* 'то же', [ЭССЯ, вып. 4, стр. 154]), **kqtoriti*/**kqtriti* (сербохорватско-кашубско-словинская изоглосса: сербохорв. диал. *kútriti* 'слушать и ничего не говорить, помалкивать', словин. *-kqtořъs* в приставочных образованиях, [ЭССЯ, вып. 12, стр. 74]), **besědъ*/**besědъ* (сербохорватско-белорусская параллель: сербохорв. диал. *běcjed*, *běcĵed* 'беседа', белор. гидроним *Бэседзь/Бэседь* [ЭССЯ, вып. 1, стр. 213; Топоров — Трубачев, 1962, стр. 177; Журавлев, 1990, ч. 1, стр. 17]), **česnikъ* (словенско-украинская изоглосса: словен. *česnik* 'чеснок', укр. *часник*, *чосник* 'то же' [ЭССЯ, вып. 4, стр. 88]), **lęgota*/**lęgotъ* (словацко-северновеликорусская связь: словац. *l'ahota* 'низина (?)', сев.-русс. *лѣготь* 'низменное болотистое место', *лѣготина* 'низменное болотистое место, иногда поросшее кустарником, пригодное для косьбы', 'старое русло' [ЭССЯ, вып. 15, стр. 55–56]) и

мн. под. являются исключительно значимыми в выявлении уровней языкового родства, то есть древних отношений между праславянскими диалектами и самого диалектного членения праславянского языка.

Различия в ценности изолекс при определении степени генетической близости разных языков должны быть учтены в формуле, которая будет положена в основание лексикостатистического сравнения.

Введем рабочее понятие класса изоглоссы. Изоглоссные классы различаются по числу идиомов, охватываемых разными изоглоссами (сам список идиомов задается заранее и определяется во многом внелингвистическими обстоятельствами: исследователь сам решает, должен ли он включать в круг сравниваемых идиомов, например, полабский — как язык, о лексике которого известно гораздо меньше, чем о словарных составах других славянских языков, старославянский — как язык книжно-культурный, не народно-разговорный и т. д., должен ли он принимать в расчет диалектное членение языков, то есть выделять диалекты в самостоятельные идиомы и сравнивать с остальными на единых основаниях, и т. п.). Характеристика класса изоглоссы не включает в себя сведения о ее конкретной ландшафтной приуроченности.

В соответствии последовательности изоглоссных классов, выделяемых в зависимости от числа связываемых идиомов, должна быть поставлена шкала коэффициентов, позволяющих устранить или свести к минимуму неравнозначность показателей близости (ср. учет мощности изоглоссы в статистическом анализе ностратической лексики у М. Л. Палмайтиса: [Палмайтис, 1978]). Поправочные коэффициенты усиливают статистическую роль узких изоглосс и снижают роль изоглосс, охватывающих большое число языков, отодвигая эти изоглоссы на фоновые статистические функции. В статистический «фон» уходит огромное количество регулярных образований, существование которых в разных языках не свидетельствует об их особой генетической близости (например, прилагательные вроде **vodъnъ(jъ)*, диминутивы типа **golъvka* и т.п.), в то время как единичные нерегулярные образования (типа **bajadlo*, фиксирующегося только в сербохорватском и русском) приобретают в этой статистике особую значимость. Формула генетической близости должна преобразоваться в такую:

$$G(A, B) = \sum_2^n pV(A, B)_i / H(A) \times H(B), \quad (10)$$

где n — количество идиомов, подлежащих сравнению; p (лат. *pondus* 'вес') — коэффициент, зависящий от класса изоглоссы (количества охватываемых ею идиомов).

Выработка шкалы поправочных коэффициентов p для V в зависимости от принадлежности данной лексической связи между языками А и В тому или иному изоглоссному классу, представляет определенную трудность. Пробные подсчеты показали, что удовлетворительно работает шкала коэффициентов, представляющая собою последовательность натуральных чисел от 2 (минимальное число сравниваемых идиомов) до n , но расположенных по отношению к числам, выражающим лексические связи, в обратном порядке: число эксклюзивных лексических связей между двумя данными идиомами умножается на коэффициент $p = n$; число изоглосс, охватывающих два данные идиома и входящих в класс $i = 3$, умножается на $p = n - 1$; число изоглосс, связывающих два данные идиома и входящих в класс $i = 4$, — на $p = n - 2$ и т. д., до коэффициента $p = 2$, 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 (выявленного) праязыкового словаря отдельно в языках А и В. Результат и служит показателем генетической близости этих языков.

В конечном виде, таким образом, формула генетической близости языков А и В представляет собою равенство:

$$G(A, B) = \sum_2^n ((n + 2 - i) \times V(A, B)_i) / H(A) \times H(B). \quad (11)$$

4. О неудовлетворительности стандартных формул меры сходства объектов

...чего нет, того нельзя считать.

Екклесиаст 1, 15

Так ли уж остра потребность в новой формуле? Нельзя ли было прибегнуть к известным статистическим приемам?

В математической статистике выработано немало методик сравнения различных объектов с исчислимыми признаками. Большинство из этих принятых, стандартных методик используется и в лингвистике, для измерения близости, в том числе генетической, различных языков и их групп.

Мы, однако, пришли к необходимости выработки собственной формулы квантитативного сравнения. Вызвано это некоторой неудовлетво-

ренностью стандартными формулами, применяемыми для измерения близости между сходными объектами.

В частности, мы вынуждены были отказаться от использования формул, построенных на методе четырехклеточной корреляции, которые применяются во многих лингвистических работах (см.: [Чекановский, 1927; Кребер – Кретьен, 1937; Перебейнос, 1967; Чекман – Широков, 1962; Мартынов, 1965; Соколовская, 1969; Соколовская, 1973]; и др.) и в принципе могут быть использованы также в сравнительном анализе словарей. Напомним, что в этом методе исчисляемые признаки располагаются в виде квадратной матрицы:

	А	не А
В	<i>a</i>	<i>b</i>
не В	<i>c</i>	<i>d</i> ,

в которой символом *a* обозначается число признаков, общих для сравниваемых объектов А и В, символом *b* – число признаков, отмеченных у объекта А, но отсутствующих у объекта В, символом *c* – число признаков, отмеченных у объекта В, но отсутствующих у объекта А, символом *d* – число признаков, отсутствующих у обоих объектов. Признаками объектов (языков) в лингвистических квантитативных исследованиях можно считать языковые единицы самой различной природы, включая и слова, возводимые к общей праформе.

На основе приведенной матрицы вычисляются коэффициент ассоциации (или коэффициент связи)

$$Q = (ad - bc) / (ad + bc) \quad (12)$$

и коэффициент сопряженности (или коэффициент контингенции)

$$\Phi = (ad - bc) / \sqrt{(a + b)(a + c)(b + d)(c + d)}. \quad (13)$$

Обе эти формулы хорошо известны и дают неплохие результаты при об-счете с их помощью данных фонетики и грамматики. Примененные же к словарю, они дают результаты не столь высокого качества, что, впрочем, выявляется не всегда. Непременным условием применения формулы коэффициента ассоциации является небольшой разброс размеров списков положительных признаков у отдельных сравниваемых объектов. Фонетика это-

му условию обычно вполне удовлетворяет: объемы фонетических систем (в фонемах) колеблются в сравнительно небольших пределах, количество фонетических и грамматических изоглосс, кладущихся в основу статистического сравнения, также, как правило, близко в сравниваемых языках (количество совпадающих в данных двух языках признаков-изоглосс может быть при этом и резко различным для разных пар идиомов).

Иное дело словарь. Сильно забегаая вперед, сообщим некоторые наблюдения, которые мы сделали, анализируя данные праславянского лексикона. Формула коэффициента ассоциации Q (12) не применима в тех случаях, когда списки положительных признаков у сравниваемых объектов сильно расходятся в объемах. Для пар языков, у которых списки признаков малы, она дает чрезмерно высокие коэффициенты, для тех же пар, у которых списки признаков значительно больше, чем у других объектов, она склонна показатели связи сильно занижать. Так, будучи обчисленным по формуле коэффициента ассоциации, наш материал дал чрезвычайно высокие значения коэффициента для полабского языка со всеми остальными идиомами и в особенности для пар «полабский & старославянский», «полабский & македонский»; связи словенского с полабским или нижнелужицким по этой формуле оказываются чуть ли не в полтора раза сильнее, чем связи словенского с чешским или словацким. Дело здесь прежде всего в том, что объем праславянского лексического наследия, реконструируемого для полабского языка, исключительно невысок — в 10–11 раз, как говорилось, меньше, чем у русского и сербохорватского; малы праславянские словники и лужицких языков.

Причина такого эффекта формулы (12) кроется в самом ее устройстве. Отсутствие общего признака у данных двух языков она рассматривает тоже как общий признак, что не может не привести к искажению реальной картины отношений между идиомами: в компаративистике, как это неоднократно отмечалось, «доказательную силу имеют только положительные факты» [Мейе, 1954, стр. 28], ср. приведенные выше слова Екклесиаста. Когда размеры списков признаков у языков колеблются незначительно, как это происходит с фонетическими или грамматическими изоглоссами, смещения реальных соотношений малозаметны или даже трудновывявляемы, а значит, и не существенны. Когда же разница весьма ощутительна, как в случае с праславянскими словниками отдельных языков, формула Q теряет свою пригодность.

Что же касается коэффициента контингенции Φ (13), то низкие его значения признаются в статистике не слишком надежными (см.: [Методика и техника статистической обработки информации, 1968, стр. 239–

240]), в измерении же родства языков главный интерес представляют соотношения, выражаемые именно низкими числами, поскольку связи ближайшерадственных идиомов, отраженные в статистических выкладках высокими показателями, с достаточной адекватностью оцениваются на интуитивном уровне, без специального квантитативного анализа.

5. Пробная демонстрация методики

Вычисление уровня родства каждой пары идиомов по предложенной нами формуле G (11) несложно, однако нуждается в весьма трудоемком «нулевом цикле» работ.

Если количество сравниваемых идиомов n невелико и 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, то есть числа охватываемых им идиомов, в формуле, которую мы предлагаем, разным изоглоссным объединениям приписываются разные весовые коэффициенты. Предварительная разметка используемого словаря — отнесение списков межъязыковых («междиомных») соответствий справа от каждой заголовочной праформы к той или иной типовой изоглоссной конфигурации — и требует наибольших затрат труда.

Разметив словарь, то есть проставив против каждой словарной статьи номер типовой конфигурации, которой соответствует список континуантов праформы в заголовке статьи, нетрудно подсчитать, каким количеством праязыковых лексических 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 формулу (11) цифры.

Количество типовых изоглоссных конфигураций, которые обобщают конкретные изоглоссы с совпадающим 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 или нулями 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 воплощенными 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 можно было справиться вручную. Подобную работу можно поручить только компьютеру.

Продемонстрируем предлагаемую нами методику на меньшем числе идиомов — не отдельных славянских языков, а их группировок. Снижение числа идиомов (за счет не сокращения, но «обобщения», «укрупнения» их) до шести или семи, что вполне приемлемо для экспериментальной проверки формулы, дает $2^6 - 1 = 63$ или $2^7 - 1 = 127$ изоглоссных типов.

Для проверочной демонстрации статистических процедур славянские языки разбиты на шесть групп, отношения особо тесного родства внутри которых бесспорны и которые соответствуют наиболее распространенным (см.: [Фурдаль, 1961; Бирнбаум, 1966; Иванов, 1982 и др.]) представлениям о диалектном членении позднеславянского языка: восточная южнославянская (болгарско-македонская), западная южнославянская (сербохорватско-словенская), чешско-словацкая, лужицкая, лехитская и восточнославянская (русская в старой терминологии) языковые подгруппы. Учитывая несколько особое положение полабского языка в составе лехитских языков, к которым его обычно относят (ср. хотя бы выделение четырех западнославянских языковых подгрупп, с обособлением полабских наречий, у А. М. Селищева [Селищев, 1941] или четырехчленное деление западнославянской зоны, с отделением поморского и полабского от польского, у Ст. Рамулга (обзор различных взглядов на место в западнославянской группе поморского и полабского см.: [Хинце, 1962]), мы сочли допустимым выделить его в индивидуальную «подгруппу». Это дает возможность уже сейчас проверить тесноту его генетических связей с другими западнославянскими подгруппами.

Упомянутая выше таблица типовых изоглоссных конфигураций при «семичленном» наборе идиомов составлена на материале первых двенад-

цати выпусков ЭССЯ (5 968 словарных статей, из которых для обсчета оставлено 5 697 лексем; принципы отбора лексики будут изложены ниже) и выглядит следующим образом (латинскими буквами обозначены исследуемые группы славянских языков: (а) — болгарско-македонская, (b) — сербохорватско-словенская, (с) — чешско-словацкая, (d) — лужицкая, (e) — полабская, (f) — лехитская, (g) — восточнославянская):

Таблица 1

(Пояснения: 1 — номер типовой конфигурации; 2 — типовая изоглоссная конфигурация; 3 — класс изоглоссы (i, число охватываемых ею идиомов); 4 — число ее лексических репрезентаций.)

1		2		3		4		1		2		3		4			
a	b	c	d	e	f	g			a	b	c	d	e	f	g		
1	+	+	+	+	+	+	7	243	25	+	+		+	+	+	5	3
2	+	+	+	+	+		6	1	26	+	+		+	+		4	—
3	+	+	+	+	+	+	6	4	27	+	+		+		+	4	1
4	+	+	+	+	+		5	—	28	+	+		+			3	2
5	+	+	+	+		+	6	601	29	+	+		+	+		4	81
6	+	+	+	+		+	5	24	30	+	+		+			3	18
7	+	+	+	+		+	5	85	31	+	+				+	3	227
8	+	+	+	+			4	12	32	+	+					2	166
9	+	+	+		+	+	6	35	33	+		+	+	+	+	6	1
10	+	+	+		+	+	5	1	34	+		+	+	+	+	5	—
11	+	+	+		+	+	5	—	35	+		+	+	+	+	5	—
12	+	+	+		+		4	1	36	+		+	+	+		4	—
13	+	+	+		+	+	5	429	37	+		+	+	+	+	5	13
14	+	+	+		+		4	32	38	+		+	+	+		4	6
15	+	+	+			+	4	320	39	+		+	+		+	4	5
16	+	+	+				3	106	40	+		+	+			3	—
17	+	+		+	+	+	6	2	41	+		+	+	+	+	5	—
18	+	+		+	+	+	5	—	42	+		+	+	+		4	—
19	+	+		+	+	+	5	—	43	+		+	+		+	4	—
20	+	+		+	+		4	—	44	+		+	+			3	—
21	+	+		+	+	+	5	18	45	+		+		+	+	4	30
22	+	+		+	+		4	1	46	+		+		+		3	6
23	+	+		+		+	4	21	47	+		+			+	3	31
24	+	+		+			3	5	48	+		+				2	19

1								2											
1	2							3	4	1	2							3	4
	a	b	c	d	e	f	g			a	b	c	d	e	f	g			
49	+			+	+	+	+	5	—	89	+			+	+	+	4		2
50	+			+	+	+		4	—	90	+			+	+		3		1
51	+			+	+		+	4	—	91	+			+		+	3		2
52	+			+	+			3	—	92	+			+			2		2
53	+			+		+	+	4	2	93	+			+	+	+	3		68
54	+			+		+		3	1	94	+			+			2		36
55	+			+			+	3	4	95	+					+	2		286
56	+			+				2	1	96	+						1		188
57	+				+	+	+	4	1	97		+	+	+	+	+	5		7
58	+				+	+		3	—	98		+	+	+	+		4		1
59	+				+		+	3	1	99		+	+	+		+	4		1
60	+				+			2	—	100		+	+	+			3		—
61	+					+	+	3	13	101		+	+		+	+	4		72
62	+					+		2	9	102		+	+		+		3		11
63	+						+	2	50	103		+	+			+	3		25
64	+							1	54	104		+	+				2		23
65		+	+	+	+	+	+	6	21	105		+		+	+	+	4		—
66		+	+	+	+	+		5	—	106		+		+	+		3		2
67		+	+	+	+		+	5	2	107		+		+		+	3		2
68		+	+	+	+			4	2	108		+		+			2		3
69		+	+	+		+	+	5	191	109		+			+	+	3		132
70		+	+	+		+		4	30	110		+			+		2		62
71		+	+	+			+	4	50	111		+				+	2		178
72		+	+	+				3	27	112		+					1		177
73		+	+		+	+	+	5	13	113			+	+	+	+	4		—
74		+	+		+	+		4	1	114			+	+	+		3		1
75		+	+		+		+	4	1	115			+	+		+	3		—
76		+	+		+			3	4	116			+	+			2		2
77		+	+			+	+	4	236	117			+		+	+	3		8
78		+	+			+		3	59	118			+		+		2		11
79		+	+				+	3	215	119			+			+	2		18
80		+	+					2	132	120			+				1		41
81		+		+	+	+	+	5	1	121				+	+	+	3		—
82		+		+	+	+		4	—	122				+	+		2		3
83		+		+	+		+	4	3	123				+		+	2		4
84		+		+	+			3	1	124				+			1		6
85		+		+		+	+	4	15	125					+	+	2		83
86		+			+			3	5	126					+		1		66
87		+					+	3	26	127							+	1	468
88		+						2	14	128								0	(0)

Подсчет реконструированных лексем, приходящихся на каждую из семи постулированных подгрупп славянских языков, в историческом отношении предположительно отождествляемых с позднепраславянскими диалектными областями, дает (по первым двенадцати выпускам ЭССЯ) следующие цифры ($H(\%)$) – доля праславянской лексики данного идиома в объеме всего праславянского словаря):

	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 подгруппу последняя характеризуется числом 2 753 лексемы, или 48,32% от общего количества отобранных к анализу словарных позиций.)

Так же нетрудно теперь для всех возможных парных сочетаний идиом подсчитать количество общих праславянских лексических реконструкций, причем сделать это необходимо отдельно для разных изоглосных классов (от $i = 2$ до $i = 7$). Приписав каждой такой сумме соответствующий весовой коэффициент, суммировав эти промежуточные результаты и отнеся итог к произведению объемов праславянского лексического наследия в каждом идиоме – составляющем данной пары, то есть проделав все вычислительные операции, предусмотренные формулой (II), мы получаем индекс генетической близости G .

Итоги предварительных наблюдений над генетической близостью семи подгрупп славянских языков (попарно) по данным лексики (на материале первых двенадцати выпусков ЭССЯ) сведены в Таблицу 2. Последняя ее колонка содержит цифры, являющиеся целью данного предварительного обсчета, то есть условный индекс родства (численные значения G в ней из соображений удобства восприятия переведены в другой масштаб умножением на 1 000):

Таблица 2

(Пояснения: 1 – A,B; 2 – V(A,B)_i; 3 – $\Sigma V(A,B)_i$; 4 – $\Sigma pV(A,B)_i$;
5 – $G = \Sigma pV(A,B)_i / H(A) \times H(B)$.)

	1	2	3	4	5				
a,b	166	358	469	560	643	243	2439	10310	0,9424
a,c	19	143	406	552	642	243	2005	7641	0,7720
a,d	1	10	47	140	609	243	1050	3175	0,7125
a,e	—	3	3	4	43	243	296	664	0,6438
a,f	9	38	153	488	640	243	1571	5414	0,7441
a,g	50	276	461	548	643	243	2221	8918	0,7684
b,c	132	411	685	745	662	243	2878	12267	0,8173
b,d	14	64	134	321	629	243	1405	4809	0,7117
b,e	2	10	11	20	63	243	349	884	0,5652
b,f	36	151	398	680	660	243	2168	8334	0,7553
b,g	286	538	730	742	663	243	3202	14323	0,8138
c,d	23	63	179	322	628	243	1458	5092	0,8329
c,e	3	8	7	23	62	243	346	868	0,6134
c,f	62	210	408	678	659	243	2260	8909	0,8924
c,g	178	405	715	740	662	243	2943	12683	0,7965
d,e	2	2	7	10	29	243	293	674	1,0580
d,f	11	26	127	254	626	243	1287	4248	0,9452
d,g	18	63	169	317	629	243	1439	4990	0,6961
e,f	3	4	5	25	60	243	340	836	0,8036
e,g	4	5	9	26	63	243	350	882	0,5316
f,g	83	221	439	675	660	243	2321	9268	0,7918

Следует заметить, что конкретные численные выражения индекса родства носят релятивный характер, они показательны только в рамках предпринятого исследования и не сопоставимы с цифрами, которые аналогичным способом могут быть получены для других наборов идиомов и на ином по объему материале: существует сильная зависимость индекса генетической близости от объема словарной базы исследования (праязыкового лексикона или его части) и длины шкалы поправочных коэффициентов, которая в свою очередь зависит от числа выделяемых языков или языковых групп.

6. К выбору весового коэффициента

Предложенное здесь пробное вычисление индекса генетической близости для пар небольшого числа идиомов имеет еще одну задачу. В формулу (11) мы ввели весовой коэффициент

$$p = n + 2 - i. \quad (14)$$

В опубликованной ранее работе [Журавлев, 1988] мы пользовались коэффициентом

$$p = \lg (n + 2 - i). \quad (15)$$

Следовательно, конечная формула для определения G имела следующий вид:

$$G(A, B) = \frac{n}{2} (\lg(n + 2 - i) \times V(A, B)_i) / H(A) \times H(B). \quad (16)$$

Обе версии формулы, (11) и (16) в принципе дают очень близкие результаты (не в конкретном численном выражении индексов G для разных пар идиомов, а в их соотношении между собой в пределах списка значений G для каждого варианта). Здесь же мы попытаемся сравнить эти результаты с целью выяснения преимуществ той или другой шкалы весовых коэффициентов, чтобы в дальнейшем, при обчете отдельных языков (а не групп, как в данном случае), выбрать более предпочтительную.

В прежнем варианте подсчетов, с p по формуле (15), результаты аналогичных подсчетов были таковы (G также умножено на 1 000):

Таблица 3

A, B	$\Sigma pV(A, B)_i$	G	A, B	$\Sigma pV(A, B)_i$	G	A, B	$\Sigma pV(A, B)_i$	G
a, b	1463,77	0,1338	b, d	721,82	0,1068	c, g	1799,87	0,1130
a, c	1122,91	0,1134	b, e	132,41	0,0847	d, e	101,15	0,1588
a, d	489,48	0,1098	b, f	1223,56	0,1109	d, f	643,05	0,1431
a, e	100,94	0,0979	b, g	2006,80	0,1140	d, g	746,47	0,1041
a, f	816,43	0,1122	c, d	760,22	0,1244	e, f	125,97	0,1211
a, g	1289,12	0,1111	c, e	130,23	0,0920	e, g	132,42	0,0798
b, c	1747,71	0,1164	c, f	1296,76	0,1299	f, g	1343,40	0,1148

Напрямую сравнивать списки значений G , полученных разными способами, неудобно, поэтому нормируем их, отнеся каждую величину к среднему значению $\bar{G} = \Sigma G_i / n$ в данном списке:

Таблица 4

(Пояснения: 1 – A,B; 2 – G_i/\bar{G} по формуле (11); 3 – G_i/\bar{G} по формуле (16); 4 – $G_i/\bar{G}_{(11)} - G_i/\bar{G}_{(16)}$.)

1	2	3	4	1	2	3	4	1	2	3	4
a,b	1,221	1,175	+0,046	b,d	0,922	0,938	-0,016	c,g	1,032	0,992	+0,040
a,c	1,000	0,996	+0,004	b,e	0,732	0,744	-0,012	d,e	1,371	1,394	-0,023
a,d	0,923	0,964	-0,041	b,f	0,979	0,974	+0,005	d,f	1,225	1,256	-0,031
a,e	0,834	0,859	-0,025	b,g	1,054	1,001	+0,053	d,g	0,902	0,914	-0,012
a,f	0,964	0,985	-0,021	c,d	1,079	1,092	-0,013	e,f	1,041	1,063	-0,022
a,g	0,996	0,975	+0,021	c,e	0,795	0,808	-0,013	e,g	0,689	0,701	-0,012
b,c	1,059	1,022	+0,037	c,f	1,156	1,140	+0,016	f,g	1,026	1,008	+0,018

Из сравнения данных первых двух цифровых колонок несложно увидеть, что формула (11), включающая в качестве весовых коэффициентов ряд натуральных чисел, дает несколько более высокие значения G для пар идиомов, у которых отмечаются наибольшие величины H (объем праславянского лексического наследия в данном языке): сербохорватско-словенская, чешско-словацкая, восточнославянская подгруппы, — и, напротив, несколько более низкие значения G для сочетаний тех идиомов, у которых H мало: полабская, лужицкая подгруппы. Наоборот, по сравнению с ней формула (16), построенная на использовании логарифмической шкалы поправочных коэффициентов, дает более высокие значения G для пар идиомов с низкими величинами H и несколько более низкие значения G для пар идиомов с высокими показателями объема праславянского лексического наследия.

В целом же формулы (11) и (16), как было сказано, дают очень сходные результаты (не в конкретных цифрах — они заметно различаются, — а в соотношениях между величинами G в пределах одного и другого списков, что наглядно обнаруживается при сравнении их одинаковым образом нормированных вариантов). Предпочсть ту или иную версию формулы G довольно затруднительно ввиду отсутствия у какой-либо из них явных преимуществ перед другою.

В этом случае, пожалуй, можно отдать предпочтение той версии формулы (точнее, шкалы поправочных коэффициентов), которая дает ярче противопоставленные друг другу численные значения конечного искомого индекса. Выяснить это просто — с помощью общеизвестного коэффициента вариации

$$v = (s/x) \times 100\%, \quad (17)$$

где x — среднее значение величины, а s — среднее квадратическое (стандартное) отклонение, вычисляемое по формуле

$$s = \sqrt{\sum (x_i - \bar{x})^2 / n - 1}. \quad (18)$$

Подставляя в эти формулы вместо x_i значения G_i и \bar{G} вместо x , при $n = 21$ (число парных сочетаний, для которых вычисляются величины G), получаем $v = 16,42\%$ для версии формулы (11) и $v = 15,98\%$ для версии формулы (16). Следовательно, формула (11), использующая шкалу весовых коэффициентов в виде фрагмента натурального ряда, дает чуть более «рассредоточенные» значения индекса генетической близости. Ее мы и будем использовать в дальнейших вычислениях уже применительно к отдельным славянским языкам, а не их группам, как в данной пробной демонстрации методики.

Глава 4

Материал исследования.

Филологическая критика источника.

Оценка статистической достаточности материала

1. Выбор источника

...Кто говорит: «прекрасны оба» —
На нежный спрос: «который взять?»...

Некрасов. «Слезы и нервы»

В главе 2-й, при рассмотрении различных методик статистического анализа лексики с целью установления генетической близости между языками и конкретных работ в этом направлении, как нетрудно было заметить, мы довольно большое внимание уделяли проблеме количества и в особенности качества лексического материала, непосредственно кладущегося в основу классификационных и компаративных построений. Это вполне понятно, поскольку проблема материала в исследованиях, подобных настоящему, стоит необыкновенно остро.

Определение степени генетической близости между языками, осуществляемое на количественно недостаточном материале, приводит к весьма ненадежным итогам.

Ошибки же в хронологической стратификации лексики и упущения в ареальной характеристике отдельных слов, нераспознавание иноязычных заимствований или результатов взаимного лексического обмена между близкородственными идиомами в период их раздельного существования и неучет книжной природы тех или иных лексических образований будут иметь своим следствием изрядные перекосы в конечной картине языкового родства. К перекосам приводят также ограничения на тематических основаниях, в особенности если классификационно-статистиче-

ские наблюдения делаются на базе узких тематических групп и лексики, тесно связанной с тем или иным культурным типом и могущей испытывать сильное влияние извне.

Из сказанного вытекает, что в качестве источников материала в наших целях могут быть использованы собрания славянской лексики, возводимой к праязыковому состоянию и максимально свободные от результатов позднейшего взаимовлияния разделившихся языков, во-первых, и стремящиеся к ее исчерпывающему представлению, во-вторых.

В праславянской лексикографии таким условиям отвечают два упомянутых выше издания — «Этимологический словарь славянских языков. Праславянский лексический фонд» и «Słownik prasłowiański».

Имеются и другие современные словарные своды, описывающие лексику славянских языков и/или праславянское лексическое наследие в каждом из них в весьма солидных объемах. Однако лексикографические предприятия брненских ученых — «Etymologický slovník slovanských jazyků» [Копечный, 1973/1980] и «Základní všeslovenská slovní zásoba» [Копечный, 1964], а также «Vergleichendes Wörterbuch der slavischen Sprachen» германских славистов Линды Садник и Рудольфа Айцетмюллера [Садник — Айцетмюллер, 1963—] использованы в наших целях быть не могут.

Первое из них имеет своеобразное построение, при котором опубликованные к настоящему времени тома включают в рассмотрение только незначительную лексику — предлоги, союзы, частицы, местоимения, местоименные наречия и их составные элементы. Знаменательные слова пока еще ожидают своей очереди. Другой особенностью этого издания является отсутствие жесткой ориентации на праславянский лексический и морфемный инвентарь, в словаре анализируются не только праславянские, но и поздние служебные элементы («грамматические слова»), включая заимствования из неродственных языков.

Несколько ближе к тому, что интересует нас, второе чехословацкое издание (в сущности, представляющее собою некий пробный вариант будущей «полнозначной» части только что реферированного словаря). Однако оно целиком соответствует своему названию, имея дело с основным общеславянским лексическим инвентарем и оставляя за пределами своего внимания огромную массу праславянских локальных (диалектных) явлений, которые для наших целей представляют главный предмет. В силу этого обстоятельства статистический анализ материалов этого издания не может нарисовать реальной картины взаимоотношений между (поздне)праславянскими диалектами, легшими в

основу самостоятельных славянских языков или их групп¹. В силу того же обстоятельства объем «Základní všeslovanské slovní zásoby» относительно скромнен — около двух тысяч лексических единиц.

Словарь Садник и Айцетмюллера не может быть использован как по причине того же отсутствия ориентации на реконструкцию праславянского лексического фонда и привлечения огромного количества поздних образований и заимствований (то есть стремления к охвату всей лексики славянских языков), так и вследствие господствующего в словаре гнездового принципа расположения материала и крайне запутанной системы его (материала) представления, затрудняющей вычленение лексических изоглосс.

Публикация словарей, пригодных с нашей точки зрения к использованию в качестве источника материала, — ЭССЯ и Праславянского словаря — находится в самом разгаре, и до полного их завершения должно пройти еще довольно много времени. На более продвинутом этапе находится работа московского коллектива авторов, к настоящему времени доведшего публикацию своего словаря уже до 19-го выпуска (середина буквы «М»). Это было главной причиной нашего обращения именно к этому словарю.

Кроме большего объема изданной части словаря, ЭССЯ отличается от Праславянского словаря еще рядом качеств, которые сделали его более предпочтительным в наших глазах².

Нельзя не воспринять сочувственно принципиальное положение, которое лежит в фундаменте концепции ЭССЯ, — тезис об автономности праславянских состояний лексики славянских диалектов, при котором

¹ Любопытно, что почти буквально такая же характеристика указанного словаря имеется в отзыве О. Н. Трубачева, хотя рецензент не задавался, в отличие от нас, специальной целью оценивать издание именно с точки зрения возможностей и ожидаемых результатов его лексикостатистической обработки: «...для классификационно-статистических... обобщений... собранный здесь материал еще далеко не достаточен» [Трубачев, 1967, стр. 384].

² Достаточно полная сравнительная характеристика обоих праславянских лексиконов дана автором и редактором одного из них, О. Н. Трубачевым, в работе редкостного жанра, совмещающего в себе жанровые признаки обзора (параллельного чтения), рецензии и авторецензии [Трубачев, 1978], а еще раньше — в статьях Фр. Копечного [Копечный, 1976] и Ш. Ондруша [Ондруш, 1976а]. Фр. Копечный в указанной публикации имел возможность оценивать параллельно только 1-й том Праславянского словаря и 1-й выпуск ЭССЯ, в то время как О. Н. Трубачев сравнивал уже два первых тома («А» — начало «D») словаря польских ученых и четыре выпуска («А» — начало «D») московского словаря, так что наблюдения последнего обеспечены более широкой фактической базой. Рецензия О. Н. Трубачева на последовавший том краковского словаря [Трубачев, 1982] также содержала в себе элементы сопоставительного чтения. См. еще: [Шустер-Шевц, 1975; Мошинский, 1977; Ондруш, 1977].

важнейшим операционным понятием является понятие праславянского диалектизма, см.: [Трубачев, 1963а, стр. 23–25; Трубачев, 1963б, стр. 168; Трубачев, 1963в, стр. 159] ³. Строгая приверженность такому подходу в ЭССЯ обеспечивает ситуацию, когда словарный состав праславянского языка, трактуемого не как монолит (и притом изначально не монолит), является конечным результатом научного поиска (Праславянский словарь несколько не совпадает в этом пункте с ЭССЯ, прибегая к расписыванию не только диалектных, но и этимологических словарей, в которых праславянская реконструкция у ж е осуществлена).

В Праславянском словаре наблюдается реликтовое «просвечивание» гнездового принципа представления материала – в виде системы отсылочных статей. В ЭССЯ отсылочных позиций нет совсем, что знаменует полный отказ от гнездового принципа в пользу более адекватной установки на соответствие «словарная статья = отдельное слово».

Московский словарь отказывается от статейного анализа отдельных грамматических форм. Он стремится эксплицитно, в виде словарных статей, давать только лексикализованные (то есть нерегулярные) явления и формы, где наблюдается супплетивизм. Краковский словарь склонен многочисленным звеньям регулярных морфологических парадигм, в особенности в области глагольной морфологии, представлять в форме самостоятельных позиций словаря, уравнивая их тем самым со статьями чисто «лексическими». То же можно сказать и о регулярных (нелексикализованных) деминутивах.

Что касается полноты праславянского словника, то словарь польских ученых не обнаруживает равномерности по сравнению с ЭССЯ: если в первых томах он был осторожнее в констатации праславянской природы тех или иных формально возможных реконструкций, то дальше как будто начинает преобладать установка на восстановление лексического состава с заметным «запасом», с прихватыванием лишку, что приближает Праславянский словарь к типу сравнительного словаря.

В Праславянский словарь помещается довольно большое количество периферийной и проблематичной в отношении принадлежности праязыковому корпусу междометной и звукоподражательной лексики. ЭССЯ на этот счет выглядит гораздо более сдержанным. Правда, нам кажется, что в одном моменте О. Н. Трубачеву можно было бы возразить. Полемизируя по этому поводу с краковяками, он пишет: «Много... у них да-

³ Хотя, как нам кажется, оно не до конца прояснено.

но междометной лексики, причем ряд примеров с последовательностью звуков *si-*. Но праславянское *s* — звук строго позиционный, и для этой эпохи (приблизительно конец праславянской. — А. Ж.) его нельзя без оговорок абстрагировать от условий его возникновения — вторая палатализация *k* перед *ě* из дифтонга. Это делает сомнительными и непраславянскими в наших глазах все эти *cu!*, *cicati*, *cisъкъ*, *cukati*, *cipati*, *cirъ!*, *cirъkati*...» [Трубачев, 1978, стр. 9]. Аргумент строгой позиционности звука *s* в праславянском, как нам кажется, направлен немного мимо цели: о строгой обусловленности тех или иных явлений фонетики их позицией в слове (слоге и т. д.) можно говорить применительно к центру языковой системы; звукоподражания же и междометия составляют крайнюю периферию и лексики и фонетики. Написания *gm* и *тъфу*, конвенционально принятые современной русской орфографией, которая не стремится отразить все многообразие периферийной артикуляции, вовсе не безусловно соответствуют реально произносимым аффективным фонетическим комплексам (см.: [Никонов, 1965, стр. 127–131]), звуки, передаваемые этими написаниями, находятся за пределами стандартной («центральной») фонетической системы. Русская фонетика не включает в число нормальных явлений [ʔ] в сильной позиции, аффективное же, периферийное употребление языка допускает парадоксальное с «центральной» точки зрения совмещение ударности и редуциции (как бы сильной и слабой позиции одновременно?). До падения редуцированных и возникновения позиционной глухости согласных славянские языки не знали звука *f*; но речь идет только о «ядерной» фонетике: трудно вообразить, что праславяне не умели фыркать, фукать, шипеть и отплеываться и передавать эти аффективные проявления в соответствующей производной лексике. Тут можно, вслед за О. Н. Трубачевым, процитировать того же Фр. Копечного: «Разумеется, праславяне, как и большинство других людей, каким-то образом *ахали* и *ахкали* (само замечание вызвано находждением в Праславянском словаре междометия **achъ* и глагола **achъkati*. — А. Ж.), но вопрос в том, произносилось ли это *ах* с редуцированным» [Копечный, 1976, стр. 11]. В целом же установка составителей ЭССЯ на ограниченную подачу в нем периферийной междометной лексики и ономастопей вполне разумна и целесообразнее излишней толерантности краковских лексикографов.

В силу описанных отличий от Праславянского словаря, которые нам видятся определенными преимуществами, мы избрали в качестве основного материала для статистического обследования «Этимологический словарь славянских языков. Праславянский лексический фонд».

2. Филологическая критика источника

Любой языковой материал, кладущийся в основу лингвистического исследования, должен подвергаться оценке с точки зрения его соответствия характеру и целям предприятия, однородности, достаточности и т. д. Общая его характеристика, полученная сравнением ЭССЯ с аналогичными лексикографическими трудами, слишком кратка, чтобы с ее помощью можно было дать представление об особенностях использованных нами словарных данных. Избранный для нашего статистического обследования лексикон нуждается в более конкретной и подробной филологической критике (под критикой мы, конечно же, не имеем в виду непременно негативную квалификацию).

К моменту начала ввода в ЭВМ лексики, анализированной в ЭССЯ, свет увидел пятнадцатый выпуск этого лексикона (оканчивавшийся серединой буквы «L», вокабулой **lokačь*).

В первых пятнадцати выпусках ЭССЯ содержится 7 938 словарных позиций⁴, что почти в сорок и восемьдесят раз соответственно превосходит объемы вариантов basic vocabulary М.Сводеша и более чем в двенадцать раз объем списка понятий для квантитативной таксономии европейских языков у А. Я. Шайкевича. По всей видимости, это количество представляло собою предельную величину добротного, строго выверенного лексического материала, который к моменту начала нашей работы мог быть подвергнут статистическому анализу с целью определения меры взаимной генетической близости между отдельными славянскими языками в их полном составе. Можно с уверенностью утверждать, что подсчеты, опирающиеся на этот материал, будут обладать достаточной убедительностью, то есть полученная картина родства будет близка к той, какая может быть выявлена при обращении к полному объему праславянских лексиконов, когда их подготовка и публикация завершатся.

⁴ К сожалению, в подсчетах словарных статей, сделанных самими составителями ЭССЯ и приводимых в конце каждого его выпуска, допущены ошибки. Они могли вкратце как при самом механическом подсчете, так и при перераспределении материала между соседними выпусками в ходе редакционной подготовки словаря, уже после того как подсчеты произведены. Наш пересчет показал, что неверно определены объемы 1-го, 12-го, 13-го, 14-го и 15-го выпусков (следует соответственно указать: 467 словарных статей (вместо 466), 390 (вм. 389), 731 (вм. 730), 596 (вм. 599), 643 (вм. 641)). У авторов словаря суммарная величина словника первых пятнадцати выпусков составляет 7 936 слов. Последующие выпуски также не избежали ошибок в определении их объемов: для 16-го выпуска указано 623 статьи (следует — 619), а для 17-го — 569 (следует — 530, то есть объем здесь завышен на 39 единиц!).

Основной объект нулевого цикла работы — изоглосса в обобщенном представлении, то есть фактически позиция в ЭССЯ — список рефлексов праславянской формы, содержащейся в каждой словарной статье.

В указанном словаре основанием межславянского сравнения и комплектования материала в пределах статьи является не корень, как это большей частью практиковалось и продолжает практиковаться в этимологических лексиконах, а словообразовательная структура (см.: [Трубачев, 1963а, стр. 9, 25–27]), и предметом рассмотрения становится не этимологическое гнездо, а праславянская лексема, то есть межславянские лексические корреспонденции с полным формальным тождеством на уровне праформы (ср.: сербохорв. *běspутица* 'бездорожье': русск. *беспу́тица* 'то же': укр. диал. *безпутиця* 'плохая погода': белор. диал. *бяспуціца* 'беспорядок', 'нескладеха' < **bezputica*; чешск. *choroba* 'болезнь': словац. *choroba*: в.-луж. *khoroba*: н.-луж. *chóroba*: польск. *choroba*: словин. *xorùbǎ*, *xcoròba*, *xoròba*: русск. диал. *хвороба*, *хороба*: укр. *хвороба*, *хороба*: белор. *хвароба* 'то же' < **xvoroba*; болг. *ко̀но̀шпн*, *ко̀но̀штин* 'земляной сверчок': макед. *ко̀но̀штин* 'медведка': сербохорв. *к̀но̀штїн*, диал. *к̀но̀штин*, *к̀но̀штин* 'то же' < **konošćirь*; и под.).

Приведенные примеры отличаются предельной прозрачностью и даже «неинтересностью» соответствий. Иные же случаи требовали от авторов определенной смелости в утверждении исконного единства сильно различающихся поздних рефлексов, в усмотрении возможности дать им непротиворечивую реконструкцию. Ср. феноменальную в этом отношении статью **brǫščьl'anь*/**brǫščьlenь*, заголовком которой служит реконструкция славянских названий бересклета, плюща и некоторых других растений (айра, жимолости), формально столь пестрых в современных славянских языках и диалектах, что большинство авторов этимологических словарей вывести общую для них праформу считали невозможным: словен. *brěšec*, ст.-чешск. *břečtan*, *brsniel*, в.-луж. *bróstwonc*, н.-луж. *barstron*, русск. *бересклѣд*, *бересдрень*, *бурусклѣн*, *брухмѣля*, укр. *брусніна*, *бризлелїна*, белор. *брызгліна* — это лишь небольшая часть славянских слов, которые обязаны своим современным обликом многочисленным метатезам, упрощениям консонантных сочетаний, трансвокализациям, контаминациям и т. д. Тем не менее и в таких случаях допустимо говорить о цельнолексемных корреспонденциях.

Однако принцип оперирования в пределах одной словарной статьи только цельнолексемными соответствиями вряд ли может быть проведен с неукоснительной последовательностью: не слишком «удобная» языковая действительность, даже если внимание исследователя ограничивает-

ся формальными моментами, далеко не всегда без зазоров, деформаций и некоторых упрощений вписывается в избираемые схемы и реконструкции, обнаруживая — в порядке естественного «сопротивления материала» — такие свойства, как неопределенность, диффузность, пересекаемость форм и проч. Помимо того, лингвист (лексикограф) до известной степени должен быть озабочен и соображениями экономности описания. Поэтому, прокламируя требования полного формального тождества корреспондирующих единиц на уровне праславянской реконструкции и «максимально расчлененной подачи словника» [Трубачев, 1963а, стр. 9], составители ЭССЯ весьма далеки от лексикографического ригоризма и допускают иногда в наборе соответствий отступления от этих в общем довольно строгих правил.

Не имеющие словообразовательного характера, то есть не разрушающие лексемной структуры, фонетические вариации (вроде **driskati*/**dristati*, **kostl'avŕjъ*/**kostlavŕjъ* и под.) можно оставить в стороне.

Наиболее обычный и даже весьма частый в ЭССЯ случай отклонения от исповедуемого принципа цельнолексемного соответствия — включение в число корреспонденций заголовочной лексемы форм, находящихся на следующих ступенях деривации при незасвидетельствованности в данном языке прямого рефлекса праформы (с соблюдением того условия, что эти деривационно вторичные формы не относятся, по мнению составителей, к собственно праславянскому лексическому корпусу): болг. диал. *белавица* в списке континуантов праслав. **bělavŕ(jъ)*; укр. *дарбенний* в статье **darŕba*; сербохорв. *дѣвуша*, *дјѣвуша*, русск. *дѣвушка* как продолжения праслав. **děvuxa*; словен. *gibežljiv* в словарной статье **gybežъ*; русск. диал. *кльѣна*, попен *rgorŕiŕm* др.-русск. *Кльѣна* в качестве производных от заголовочного глагола **klypati* (*se*); чешск. *ločovina*, словац. *ločovina*, развивающие заголовочную реконструкцию **loča*/**ločъ*, и мн. др. Такие добавления к спискам корреспонденций, свидетельствующие о бытии существовании первообразных (для указанных дериватов) структур в идиомах, предшествующих современным языкам, (а возможно — о редкости более простых в деривационном отношении форм, не позволившей обнаружить их диалектологам-собираателям) вводятся в словарную статью стандартным оборотом «(ср.) сюда же...» или его стилистическими эквивалентами.

Изредка встречаются и обратные случаи — когда статья ЭССЯ посвящена этимологическому или словообразовательному анализу аффиксального (суффиксального в правиле) образования, а «сюда же» подключается единичная нераспространенная форма, в самостоятельную словар-

ную статью составителями не выделяемая (например, русск. диал. *гуть* в статье **gōtъnъjъ*, н.-луж. *gjarś* в статье **gōrtanъ/*gōrtanъ*).

Морфологические варианты могут в некоторых случаях объединяться в одну статью (как, например, в статьях **čemerъ/*čemera*; **d(ъ)vizъ, *d(ъ)vizъ*, **dvizъ/-etъ*; **gala/*galo*; **xodъ/*xoda*; **корьна/*корьно/*корьнъ* и т.п.), в других же — составлять разные позиции. При этом принятое авторами решение не всегда получает явное объяснение. Если пары лексем **dara* и **darъ*, **doba* и **dobo* и аналогичные, оформленные каждая в виде автономной словарной позиции, демонстрируют различия, восходящие еще к индоевропейскому состоянию, что отмечается в этимологической разработке, то раснесение лексических пар **ablъko* — **ablъkъ*, **brězga I* — **brězъgъ I*, **brězga II* — **brězъgъ 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, точнее, о том, что в одних случаях (типа **čepati* — **čepiti*, **xlestati* — **xlestiti*, **děti* — **děvati*, **lepati* — **lepiti* — **lepъiti*) разные статьи представляют глагольные парадигмы как словообразовательные, и их составляющие трактуются как самостоятельные лексемы, а в других случаях (типа **xъrčati/*xъrčiti*, **koposati/*kopositi*, **dryxati/*dryxъiti*, **dobyti/*dobyvati*, **gorniti I* с включением сербохорв. *granati* < **gornati* в ту же статью, **xarati* с включением укр. *xarъiti* < **xariti* и т.п.) объединение в пределах одной статьи форм, имеющих различающиеся реконструкции, заставляет предполагать, что у них авторы ЭССЯ формальные различия не отнесли к факторам, разрушающим единство лексемы. Случаи подобного объединения представлены в ЭССЯ заметно реже, чем оформленные грамматических вариаций в виде разных статей, объяснить их можно, по-видимому, иногда позднейшей лексикализацией некоторых рефлексов (что, как нам показалось, в общем отнюдь не очевидно), в иных же случаях — сравнительно небольшим материалом, кладущимся в основание данной праславянской реконструкции и — при его недостаточности — могущим навязывать читателю впечатление об исчерпанности ареальной характеристики таких «сублексем». Поскольку последнее все же нежелательно, подобные решения следует признать оправданными.

По той же причине в пределах одной словарной статьи могут описываться сильно различающиеся по суффиксальному оснащению дериваты (для уяснения позиции составителей ЭССЯ важна статья: [Варбот, 1976]), причем словообразовательные различия между ними могут выноситься и в заголовок статьи, а не только существовать в виде вариантов из категории «сюда же» при некоей «основной» праформе, напри-

мер, **drǫkolъ* (/b), **drǫkolъje*, **drǫkolъna*; **koničina*/**kon'ušina*; **ležišče*/**leževišče* и т.п. Не всегда это выглядит вполне убедительно. Таких случаев, правда, совсем немного.

Могут объединяться в одну статью и словообразовательные омонимы, отмеченные в разных языках, ср., например, **agodъnica*, объединяющее, по-видимому, производные от **agoda* со словообразовательно вторичными случаями, непосредственно связанными с **agodica*; **gospodъnъ(jb)*, в котором слились дериваты от **gcspodъ* и **gospoda* (составляющих предметы разных словарных статей).

Все сказанное ясно демонстрирует, что «набор» лексических изоглосс, если последние исчислять, исходя из словника ЭССЯ, является до некоторой степени условностью. Другие чисто лексикографические решения, принятые по соображениям удобства расположения материала в упомянутых или аналогичных им случаях, привели бы к констатации отличных изоглосс, часто иной мощности, что, как нетрудно показать, безразлично для результатов нашей работы, поскольку разным по классу изоглоссным конфигурациям приписываются в выработанной нами формуле разные весовые коэффициенты.

Состав подсчитываемых единиц, а вместе с ним (правда, едва ли сколько-нибудь значительно при таком объеме лексики) 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 **ba-x-orъ* с семантикой 'внутренности, кишки', 'колбаса', 'что-либо толстое, круглое' и т.п. от глагола **ba-xati*. Разграничить **baxorъ* I и **baxorъ* II до конца мешает сербохорв. диал. *bāor* «о ребенке», также 'внутренности животного', совмещающее в себе обе семантические линии: значение 'ребенок, карапуз' может быть вторичным по отношению к 'пузо, брюхо' (см.: [ЭССЯ, вып. I, стр. 136]; тюркское происхождение этих славянских слов, предполагаемое в работах: [БЕР, т. I, стр. 36; Семенова, 1983, стр. 164–165], — менее вероятно; в ЭССЯ не объяснено непривлечение в список соответствий польск. *bachór* 'большой живот, брюхо', а также *bachor* 'отродье, пашенок, убудок', которое семантически пересекается с сербохорватским, верхнелужицким и восточнославянскими примерами). Далее, в случае признания этимологической связи глагола **kaniti* (при преобладании в южнославянских рефлексах значений 'приглашать', 'предлагать', 'намереваться') с гнездом **konъ*, **konati* «становит-

ся ясным отнесение сюда... русск. диал. *кáнить* 'пятнать (при игре в пятнашки)>' [ЭССЯ, вып. 9, стр. 141].

Позволим себе привести несколько примеров предложенных в ЭССЯ этимологий и констатированных там гнездовых связей, которые вызывают у нас возражения (приведем замечания касательно только тех статей, где неверная, по нашему мнению, этимологизация повлекла за собою констатацию — в виде словарной статьи — особых изоглосс: нас здесь занимают не столько этимологии, — это сюжеты, интересные сами по себе, — сколько ареальные связи и межъязыковые параллели).

С нашей точки зрения, очень ненадежно выделение в качестве самостоятельной праславянской реконструкции **agľtǫjь/*jagľtǫjь* [ЭССЯ, вып. 1, стр. 53]. Многочисленные примеры ослабления смычки у *-d'*- в разных славянских языках заставляют предположить здесь варьирование праслав. **deǵľtǫjь < *deǵnǫti*, что находит подтверждение в параллелизме наименований сныги русск. *дѣглиця//яглиця* [Даль₂, т. I, стр. 512; т. IV, стр. 246, 672], укр. диал. *дѣглиця//яглиця* [ЕСУМ, т. 2, стр. 152–153]; см. подробнее: [Журавлев, 1990, ч. I, стр. 8–9] (к изложенной там этимологии и лексическим фактам можно добавить нижегор., яросл. *неяглый*, нижегор. *няглый* 'ленивый' [СРНГ, вып. 21, стр. 212, 331], из *не + яглый* 'сильный, тучный', восточнорусского же, по Далю, распространения, ср. сев.-русск., псков. *недяглый* 'слабый, болезненный, невзрачный, плохо растущий' [там же, стр. 39]).

ЭССЯ [вып. 3, стр. 111] фактически отказывается от этимологизации слова **bǫdolъ*, известного лишь по словацким данным (*bdol* 'улей'), сопровождая заголовочную реконструкцию, как и альтернативную **bьdolъ*, вопросительными знаками и замечанием в тексте статьи «Неясное слово. Возм., родственно **bьdla...*» (последнее — исключительно в значениях 'гриб', 'губка'). Между тем словацкое **bǫdolъ* 'улей' отношениями консонантной метатезы в корне несомненно связано со словацким же **dǫbolъ > dbol* в том же самом значении 'улей', относимым в ЭССЯ к гнезду с корнем **dǫb-*, ср. **dǫbrъ* [вып. 5, стр. 175–176]. Суть, однако, не столько в убедительности самой этимологизации, сколько в необходимости отождествления двух слов, описанных словарем в разных позициях.

Сомнительным представляется сближение русск. диал. *калмáтый* 'безрогий' и белор. *калмáты* 'косматый, лохматый' [ЭССЯ, вып. 13, стр. 189]. Севернорусск. *калмáтый* гораздо ближе семантически к севернорусск. же *комлáтый* 'комолый' [СРНГ, вып. 14, стр. 233], очевиднейшим образом связанному с *комóлый* 'безрогий', далее — с *комель* < праслав. **komьль* (последнее в ЭССЯ в отдельную праславянскую позицию не

выделено), см.: [Журавлев, 1990, ч. II, стр. 50]; аганье в севернорусской форме может объясняться процессами дезэтимологизации слова.

Фактически не доказанным остается сближение с в.-луж. *laćić* so 'медлить, выслеживать, идя следом' сербохорв. диал. *laćiti* 'производить обрезку виноградной лозы...' [ЭССЯ, вып. 14, стр. 8]. Последнее должно быть объединено на правах заимствования с болг. *лѣча*, диал. *лѣча* 'отделять, отлучать', 'очищать виноград, чеснок, лук, капусту от лишних веток и листьев' [БЕР, т. III, стр. 329; Журавлев, 1990, ч. II, стр. 56], восходящим к праслав. **lǫčiti*.

Продолжать перечислять спорные этимологизации в ЭССЯ нет большого резона. По-видимому, любой вдумчивый рецензент найдет в этом словаре неприемлемые для него этимологические решения (в чем нет никакого умаления его — именно как этимологического словаря — исключительно высоких достоинств). Следует лишь подчеркнуть, что иные этимологические версии, будь они отражены в ЭССЯ, изменили бы состав рефлексов и изоглоссных конфигураций, констатированных в некоторых статьях словаря, что опять-таки повлияло бы на статистическую оценку близости между отдельными славянскими языками. Правда, это влияние было бы столь незначительным, что им можно пренебречь.

Далеко не всегда мы позволяли себе менять состав формирующего данную изоглоссу лексического материала, даже если были полностью не согласны с развиваемой ЭССЯ этимологизацией или видели возможности иного распределения славянской лексики между словарными статьями. Обычно в спорных, с нашей точки зрения, моментах мы старались придерживаться подхода к словарю, который был положен в основу нашего статистического анализа, как к данности, относясь к указанным лексикографическим и этимологическим решениям скорее «некритично», как если бы они были единственно возможными (однако не непреложно; об отступлениях от такого принципиального подхода см. ниже).

Об отмеченных реальных или мнимых непоследовательностях, которые мы отнюдь не склонны считать недостатками словаря, мы нашли необходимым упомянуть затем, чтобы, во-первых, лишний раз подчеркнуть сложность всякой словарной работы, а во-вторых, показать важность привлечения массового материала для количественных оценок лингвистического родства. Потребность выбора между несколькими возможностями — ситуация более чем обычная в лексикографии, и только при значительном объеме материала спорные моменты могут стать статистически незначимыми, погашаясь мощно обнаруживающимися в словарном массиве тенденциями.

3. Уточнение изоглосс: непривлеченная лексика

Часовой. Ну что ж, сэр, всего четыре слова. Разве вы собиратель таких вот пустейших пустяков?

Шоу. «Смуглая леди сонетов»

По сравнению с поливариантностью лексикографических и этимологических решений на пропорции статистических связей в гораздо большей степени может оказать влияние неполнота привлеченного материала, чего избежать в словарях, подобных анализируемому, просто невозможно.

Об этом целесообразно сказать подробнее.

Ни к одному словарю, за исключением лексикографического описания конечного фиксированного текста или группы текстов, не применимо требование исчерпывающей полноты в ее буквальном понимании. Тем более это справедливо по отношению к словарям, которые сами в свою очередь создаются с опорой на другие словари. В полной мере это относится к лексиконам типа ЭССЯ, базирующегося на сплошном обследовании как литературных, так и, прежде всего, диалектных словарей всех языков славянской семьи. Составители ЭССЯ лишены на этот счет иллюзий. Содержащиеся в аннотациях на оборотах титульных листов почти каждого выпуска выражения «максимально полный охват» (вып. 3–7) и «исчерпывающее использование» (вып. 8–19), насколько можно понять, носят осознанный и извинительный декларативный характер – как указание на путеводный, но чрезвычайно трудно достижимый идеал.

Трезвое признание составителями возможности «лакун в том, что касается полноты сведений о словах и их ареалах» [ЭССЯ, вып. 1, стр. 6] сопровождается практическими усилиями по устранению обнаруженных пробелов, см., например, статью: [Меркулова, 1986].

Чудовищный объем предварительной черной работы, требующей огромного психологического напряжения и «многоканального» внимания, заведомая неполнота, неравнокачественность и, главным образом, открытость используемых источников (работа по выявлению и фиксации диалектной лексики в принципе не может быть завершена) делают пропуски в словарях, подобных ЭССЯ, неизбежными. К этим вполне объективным причинам, по которым оказываются непривлеченными в описания изоглосс слова тех или иных славянских диалектов, добавляются пропуски, имеющие случайный характер. К таковым, например, относятся отсутствие русск. *xata* в статье **xata*, белор. *што* в статье **što* и под. Наиболее же порази-

тельны случайные пропуски отражений праславянских слов в древнерусском языке. Курьезное отсутствие древнерусских примеров в статьях *berza, *bĕgati, *biti, *bitva, *blazniti (se), *bobrъ, *bogatъ(jb), *bogatъstvo, *bogъ, *bojati (se), *bokъ, *bolto, *božъjъ, *bratъ, *buditi, *byti, *četyre, *doma и некоторых других, наличие которых в древнерусских текстах не подлежит сомнению (см.: [СлРЯ XI–XVII вв.; ДРС; Срезневский]), можно объяснить, пожалуй, лишь эффектом «слона в кунсткамере», не отмеченного персонажем крыловской басни «Любопытный».

Посильный вклад в устранение лагун в своде лексических изоглосс, какковым, помимо прочего, является ЭССЯ, мы осуществили в упоминавшейся уже работе: [Журавлев, 1990, чч. I, II]. «Дополнения к лексическим материалам „Этимологического словаря славянских языков“» сложились как косвенный результат нашей работы по вводу материалов ЭССЯ в ЭВМ с целью определения взаимной генетической близости между отдельными славянскими языками на основе статистической обработки данных о лексических и словообразовательных изоглоссах.

Понятно, что одним из наиболее существенных предметов исследовательского внимания, едва ли не самоцелью в этом случае становится полнота описания ареалов отдельных слов. Этим и были обусловлены некоторые особенности наших «Дополнений» в их отличии от имеющихся аналогов (упомянутая статья В. А. Меркуловой, см. также: [Орел, 1987]).

В «Дополнениях» мы не ставили перед собой задачи расширения словника ЭССЯ, видя основную свою цель в «максимализации» представления уже выявленных изоглосс, то есть фактически тех списков континуантов лексических праформ, которые даются в каждой словарной позиции ЭССЯ после заголовка. Равным образом мы не стремились вносить дополнения, касающиеся неотмеченной семантики включенных в эти списки форм, если такие сведения не побуждали усомниться в правомерности объединения приведенных в данной статье славянских лексем в этимологическое гнездо. С другой стороны, мы стремились более детализованно представить такой крупный этноязыковой континуум, как территория великорусского языка, и в соответствии с этим вводили в «Дополнения» русские диалектные формы, если они отражены в ЭССЯ иллюстрациями лишь из одного наречия при более широкой реальной распространенности.

В «Дополнениях» включалась главным образом восточнославянская (русская, белорусская, украинская), болгарская и верхнелужицкая лексика. Источниками ее служили некоторые словари — преимущественно региональные [Акхимский словарь; Архангельский областной словарь; Народ-

нае слова; Народная лексіка; Народная словатворчасць; Слоўнік паўночна-заходняй Беларусі; Туровский словарь; Янкова и др.], ономастические [Веселовский Ономастикон; Жучкевич, 1974] и этимологические [БЕР; ЕСУМ; ЭСБМ; ESSJ], содержащие значительный диалектный материал, но также и литературные, например, [БРС; Трофимович, 1974 и др.], которые из-за времени своего выхода в свет могли быть учтены в опубликованных выпусках ЭССЯ лишь частично, некоторые несловарные издания [Демчук, 1988; Дзэндзелевский, 1987] и др., а также использованные в ЭССЯ словари, извлечения из которых не избежали случайных пропусков. Очевидно, что в качестве источника должен использоваться и «Общеславянский лингвистический атлас», публикация Лексико-словообразовательной серии которого началась с тома «Животный мир» [ЛСС ОЛА, вып. 1].

Сделанное нами пополнение списков континуантов славянских праформ коснулось около 1 600 словарных статей ЭССЯ.

Некоторые из добавлений позволили отклонить высказанные составителями сомнения в праславянской древности заголовочной праформы (**lěsati* в ЭССЯ реконструировано на основании единичного образования словин. *lasati* прилаг. 'лесистый' — с замечанием: «Праславянская древность сомнительна» [ЭССЯ, вып. 14, стр. 236]; привлечение дисконтактно-го в.-луж. *lěsaty* 'лесистый' [Трофимович, стр. 108] рассеивает сомнения), подозрения в книжной природе некоторых слов, отмеченных в поздних славянских языках (например: «...опущены отдельные формы от прилаг. **blōdъnъ*, которые могут быть заподозрены в книжном происхождении, как напр. русск. *блудный*» [ЭССЯ, вып. 2, стр. 128] — диал. *блудной* 'заблудившийся, сбившийся с дороги' [Архангельский областной словарь, вып. 2, стр. 36], отличающееся, кстати, от литературной формы флексийной акцентровкой, подтверждает народное существование слова в русском языке; белорусский топоним *Казярóгі/Козерóгі* [Жучкевич, 1974, стр. 166 — село Лоевского р-на] ослабляет подозрения в «литературном генезисе (калька с лат. *capricornus*) и книжном распространении (заимствовании) между слав. языками... ряда случаев названия» [ЭССЯ, вып. 12, стр. 22]), расширить ряды морфологических и морфонологических вариантов реконструированных праславянских лексем (**bolěti* — русск. диал. *болеть* 'становиться больше, расти, увеличиваться' — наряду с заголовочным праслав. **boliti*; **brujь* — белор. диал. *бруй* 'волнение, рябь', 'тот, кто мочится' — наряду с **bruja*; **čamьriti* — русск. диал. *чамрiть* 'накрапывать, моросить' — наряду с **čamьrěti*; **čučalo* — укр. диал. *чуцало* — наряду с **čučelo*/*čučela* (о большей точности возможной реконструкции **čučalo*/*čučala*, производного от глагола **čučati*, говорится и в самом ЭССЯ [вып. 4, стр. 127]); **ču-*

jadlo – в.-луж. *čujadlo* ‘щупальце’ – наряду с **čudlo*; **čьlpati* – в.-луж. *čowpać* ‘карабкаться’, ‘сползать, скользить’ – наряду с **čьlpti*; **dego* – белор. диал. *дзе́го* ср. р. ‘отрезанное лыко’ – наряду с **dega/*degь*; **gono* – в.-луж. *hona* множ. ч., род. п. *-ow* ‘нива, поля’ – наряду с **gonь*; **jъzskubati* – в.-луж. *zeskubac* ‘ощипать, ободрать’ – наряду с **jъzskubti*; **kolomqъ* – белор. *каламјць* ‘муть’ – наряду с **kolomqъ/*kolomqta*; **kostovaľ* (очевидно вторичная форма) – русск. диал. *костовал* ‘растение *Delphinium consolida* L., шпорник, живокость’ – наряду с **kostivaľ*, и мн. др., см.: [ЭССЯ – соответствующие позиции и Журавлев, 1990].

Главное, однако же, в том, что возможности пополнения исходного для определения различных изолексов материала оказались настолько велики (наши дополнения коснулись каждой пятой статьи опубликованных выпусков ЭССЯ), что эта супплементация может сказаться на конечной статистике межславянских связей и лексических параллелей. В особенности значимыми являются случаи, когда дополнения устанавливают двусторонние эксклюзивные изоглоссы, так как именно им приписывается в нашей формуле (II) наибольший весовой коэффициент, тем более, что эти констатации, как правило, возникали из нахождения соответствий формам обособленным, в ЭССЯ засвидетельствованным в единичных языках (единичные же фиксации, будучи учтенными в составе величины *H*, входящей в знаменатель формулы, способствуют – при значительном их количестве – уменьшению итоговой величины *G*, то есть статистическому удалению данного языка от остальных). Таковы, например, праславянские лексемы **bagnje* (единственному отмеченному в ЭССЯ нижнелужицкому свидетельству обнаруживается соответствие в белорусском языке), **bajunь* (украинская параллель к русскому свидетельству), **basnica* (чешская фиксация дополняется болгарской), **bazuriti* (русск. + белор., туровское), **berzujь* (русск. + белор., производный ойконим на Могилевщине), **bezplodъjь* (чешск. + русск.), **bezpqъ* (сербохорв. + русск.), **běľce* (сербохорв. + укр.), **běšeti* (русск. + укр.), **bitelь* (чешск. + укр.), **blošćica* (укр. + белор., туровское, сев.-зап. говоры), **bodadlo* (словен. + русск.), **boliti* (*/*bolēti*) (белор. + южнорусск.), **boltuxъ* (чешск. + укр.), **buxurь* (словац. + русск.), **bukarь* (в статье **bukariti*: болг. + русск.), **bъgadlo* (русск. + укр.), **bъrmiti* (русск. + укр., производное), **d(ъ)jъnъky* (болг. + в.-луж.), **dbrъbь* (полаб. + в.-луж.), **jat(ъ)jъje* (в статье **jat(ъ)ja/*jat(ъ)jo*: русск. + белор.), **katadlo* (русск. + укр., производное), **kouyka* (русск. + укр.), **kozvonь(kъ)* (русск. + белор.), **krqъtověja* (белор. + укр.), **krъbъ* (сербохорв. + русск., словообразовательно соотносительное), **lēsatъjь* (словин. +

в.-луж.), **lędzka* (русск. + белор., ойконим на Гродненщине), **lędъse* (чешск. + белор., ойконим) и др. (см. там же).

Говоря о пробелах в обрисовке словарем ареалов славянских слов, мы чрезвычайно далеки от желания представить это как изъяны ЭССЯ, а уж тем паче предъявить какие-то счета его составителям. Понимание гигантских трудностей, стоящих перед лексикографом, трудностей, воспетых поэтом, удерживает нас от каких-либо высказываний негативного толка в адрес авторов словаря. Эффект «слона в кунсткамере», о котором мы рассуждали выше, сказался и в наших дополнениях к материалам ЭССЯ: все перечисленные выше пропуски древнерусских слов, кроме **byti*, были не замечены вовремя и нами и оказались обнаруженными уже после того, как компьютер выдал окончательные цифры. Вводить эту лексику дополнительно и делать вычисления заново мы сочли необязательным и упоминаем об этом лишь из добросовестности. Влияние на окончательные результаты эти полтора-два десятка слов окажут ничтожнейшее: все они являются общеславянскими. Будучи отмеченными во всех или почти во всех исследуемых языках, они включаются в обсчет с самыми низкими весовыми коэффициентами. Лексика такого рода воспринимается нашей формулой как статистический фон, и ее — в небольшом количестве — учет или неучет принципиально ничего не изменит. Без всякого сомнения, тщательнейшая проверка лексического материала ЭССЯ позволит найти новые лакуны. Повторяем, избежать их в работе, подобной той, которую взяли на себя составители ЭССЯ, попросту невозможно. Очевидно несистемный, стохастический характер подобных пропусков не внушает опасений относительного их воздействия на надежность конечных статистических результатов.

4. Сегрегационный анализ

— ...А они нарочно в сочинениях сомнительные слова пишут, — все с Гротом приходится справляться.

Сологуб. «Мелкий бес»

Специального разговора требует еще один род вмешательства в материал ЭССЯ, которое мы осуществляли при его вводе в ЭВМ.

Речь идет о сегрегационном анализе, которому подвергалась содержащаяся в ЭССЯ лексика славянских языков. Эта процедура вызвана стремлением к тому, чтобы лексический материал, положенный в основу наших

расчетов, был предельно чист в генетическом отношении. Цель сегрегационного анализа – 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 вносить искажения в синхронную картину генетической близости (поздне)праславянских диалектов.

Достаточным поводом для того, чтобы констатировать небезупречность примера или реконструкции, служили явно выраженные сомнения самих составителей словаря.

Прежде всего не включались в подсчет целиком те 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», «может быть местным образованием», «(имеются) признаки вторичного образования» и т.п. Устные консультации с О. Н. Трубачевым подтвердили общий смысл всех этих замечаний.

Из перечней континуантов праформы вычеркивались церковнославянские лексические примеры (с пометами «русск.-цслав.», «сербск.-цслав.»). Церковнославянская лексика любого извода большей частью ориентирована на восточный южнославянский (болгаро-македонский) ареал и, отличаясь преимущественно книжными путями распространения, вносит aberrации в общую генетическую характеристику поздних языков, способствуя усилению статистических показателей генетической связи болгаро-македонской языковой зоны главным образом с западной южнославянской (сербохорватско-словенской) и восточнославянской зонами.

Если заголовочная праславянская реконструкция была осуществлена в ЭССЯ с опорой только на церковнославянские свидетельства (ср. **abrědb*, **abrědъje*, **apa?*, **askъdb*, **azъněnbъ*, **bugъ*, **bъtarъ I*, **bъtъ I*, **čestidlo*, **čajazъnъ* и т. д.), статья изымалась из подсчетов целиком.

Из списков соответствий изымались слова живых языков, относительно которых у авторов ЭССЯ возникают подозрения в позднем праславянском образовании или их книжной природе, например, зап.-укр. *багрий* 'темнокрасный' (**bag(ъ)ръ I*: «генезис зап.-укр. <...> для нас недостаточно ясен» [вып. 1, стр. 131]), русск. *бесчадный* 'бездетный' (**bezčędnъnъ(jъ)*/**bezčędnъna(ja)*: «русск. <...> возм., книжн.» [вып. 2, стр. 15]), русск. книжн. *довлеть* (**dovblěti*), чешск., словац. *hrachor* 'растение *Lathyrus*' (**gorxorъ*: «чеш., словц. <...> книжн. заимств. из ю.-слав.» [вып. 7, стр. 43]), русск.

клеветá, укр. диал. (!) клѣвотá (**kleveta*: «мнение о церковнославянском происхождении ряда слав. соответствий, в том числе русск. <...> можно было бы дополнить дальнейшим уточнением о чеш. (морав.) природе ст.-слав., цслав. клевета» [вып. 10, стр. 15]).

Специального внимания требовали статьи ЭССЯ, в которые был включен лексический материал поздних языков, вызывавший некоторые сомнения касательно его формальной (на уровне праславянской реконструкции) однородности.

Некоторые слова западнославянских языков с префиксом *z-* двусмысленны в плане реконструкции: нередко этот префикс с равным вероятием может быть возведен и к **ʒvz-* и к **zv-* (см., например, **ʒvzbŭiti*, **ʒvzbosti*, **ʒvzdati* и т. д.; несколько забегая вперед, отметим, что образования на **ʒvz-* вообще требуют известной осмотрительности, о чем мы еще скажем ниже). При недостаточности семантических аргументов такие примеры «вычеркивались» из перечня соответствий, в ряде случаев принималось даже более радикальное решение — опущение всей вокабулы целиком.

Учитывались содержащиеся в тексте словарных статей оговорки относительно отдельных славянских форм. Например, из перечня континуантов праформы **ablonŭnŭjъ*/*abolnŭnŭjъ* были вычеркнуты словен. *já-bolen* 'яблочный', которое «может продолжать также особое праслав. **ablŭnъ*, прилагательное на *-nŭjъ* от названия яблока без расширителя *-k-*», и проблематичное словин. *jablaní* 'die Apfel betreffend', «которое может представлять собой производное с суфф. *-ĕnŭjъ*/*-ĕnŭnŭjъ* от словинского же *jáblo* 'яблоко» [ЭССЯ, вып. 1, стр. 43–44]. Правка такого рода способствовала усилению принятого в словаре принципа комплектования статьи с учетом цельнолексемности соответствия.

Нерешительность составителей по отношению к формам из категории «сюда же» (см. о них выше), выражавшаяся в парентезах «возможно», «по-видимому» (ср. например русск. диал. *баларѹжина* 'лужа, скоп грязной воды', укр. диал. *барѹло* 'возвышенность в общем значении' и др. в статье **bara*; чешск. редкое *bluňkati* 'булькать' в статье **bl'uxati*/*bl'ukati*; болг. (Геров) *гулешъ* 'пескарь *Gobio*', *гѹлешъ* 'земляной червь' в статье **gul'ati*⁵; русск. диал. *люлѣйка* 'песня «веснянка», в которой встречается припев «люли»' в статье **l'ulĕja* с польским и словинским словами в значе-

⁵ Для последнего случая не упоминаем сербохорватское и русское наименования пескаря с корнем **gul-*, так как их изъятие или неизъятие не изменит состава констатируемой статьей **gul'ati* изоглоссы с «зиянием» на месте болгарского; впрочем, объединение болгарского, сербохорватского и русского в отдельной изоглоссной единице «наименование пескаря с корнем **gul-*, пусть и с сомнительной этимологией, было бы допустимо.

нии 'медлительный человек; недотепа' с достаточной, как нам кажется, вероятностью родства всех этих слов; и мн. др.), толковалась скорее в пользу сегрегации «возможных» примеров.

Естественно, что изымались из подсчета статьи, чрезвычайно немногочисленные, в заголовки которых помещались праформы, реконструированные с опорой не на зафиксированные славянские слова, а на косвенные свидетельства — отталкиваясь от отражений древнейшей славянской ономастической лексики в неславянском тексте, например, **xvalibudъ* (восстановленная праславянская форма, которую продолжает в виде глоссы у Прокопия Кесарийского *πομην προγριμνu personae Χιλβουδιος*), **lęditji* (гипотетическая праславянская форма западнославянского этнонима *Len-dizi* у Баварского географа, IX в.). Отнести эти формы к каким-то конкретным славянским языкам, между которыми выявляется степень взаимной близости, невозможно, в отличие, скажем, от близких к **lęditji* (и по корню, и, видимо, ономастиологически) форм этнонима **lędjane*/**lędĕne*, сохранившихся, кроме отражений *Λενζανῆνοι* и *Λενζενίνοις* у Константина Порфиrogenета (см.: [Ильинский, 1925; Хабургаев, 1979, стр. 186–187],) в польской и сербской ономастике, см.: [ЭССЯ, вып. 15, стр. 44].

Таким образом, начальная часть праславянского словаря, представленная в первых пятнадцати выпусках ЭССЯ и положенная в основу статистического обследования, была нами сокращена приблизительно на 430 словарных единиц, в основном за счет тех статей, которые либо в качестве заголовочной праформы имели реконструкцию, сомнительную в отношении ее праславянской древности (что выражалось в явной форме в тексте самой статьи), либо включали лексику, распространившуюся в славянских языках книжным путем как заимствования из церковнославянского. Это составило 5,4% словника ЭССЯ.

5. Расширение корпуса изоглосс

- Ты, пожалуйста, их перечти, — сказал Чичиков, — и сделай подробный реестрик всех поименно.
- Да, всех поименно, — сказал Манилов.

Гоголь. «Мертвые души»

Выше, когда речь шла о лексикографических альтернативах, в частности, о возможностях различной группировки славянских слов в статьях этимологического словаря (в данном случае имеются в виду прежде

всего его словообразовательные аспекты), а тем самым — об известной условности конечного списка праславянской лексики, мы отметили, что в подавляющем числе неоднозначных ситуаций принимали решения составителей ЭССЯ как данность. Однако авторы словаря демонстрируют, как было сказано, отсутствие лексикографического ригоризма (что привередливый рецензент расценил бы как непоследовательность) и нередко однотипные ситуации решают по-разному, не всегда аргументируя свой выбор⁶. Например, вариантная вокализация идеофонического по своей природе корня, морфологические различия по роду, морфонологические нюансы, словообразовательные модификации и т. д. то расцениваются как незначительные, то служат основанием для разнесения лексики, в остальных моментах идентичной, в разные словарные позиции.

Имея в виду эту компромиссную гибкость ЭССЯ, мы сочли возможным для себя использовать ее с тем, чтобы несколько расширить число словарных позиций, подлежащих статистическому анализу. Следуя некоторым имеющимся в ЭССЯ аналогиям, мы представляли списки континуантов заголовочных праформ в ряде словарных статей *divise*, если находили, что они дают этому достаточное основание. Тем самым усиливалась заявленная как принцип «максимально расчлененная подача словника» [Трубачев, 1963а, стр. 9]. Кроме того, эта операция давала возможность описанную в данной статье лексическую изоглоссу, по сути своей соответствующую этимологическому или словообразовательному гнезду, представить подробно, как совокупность более «мелких» изоглосс. При этом часто каждая из таких более узких «новых» изоглосс характеризовалась меньшим числом охватываемых идиомов. Последнее, как понятно, при нашем способе квантитативной обработки лексики позволяло сделать различия и сходства между славянскими языками статистически чуть более выпуклыми: чем массивнее изоглосса, тем менее она значима для определения меры близости между языками, и наоборот.

Собственно говоря, разделению на частные изоглоссы подвергался материал тех статей этимологического словаря, «антигнездовое» вмешательство в состав которых позволяло вычленивть более узкие межъязыковые связи. Правда, мы не злоупотребляли этими возможностями, поскольку возникала опасность некоторого нежелательного удревления более поздних сходств и различий между отдельными славянскими идиомами. Статей, материал которых мы представляли *divise*, в общем очень немного.

⁶ Впрочем, мы вполне допускаем, что аргументация такого рода имеется в большинстве случаев, но не была нами должным образом воспринята.

Перечислим их (в скобках после заголовочной праформы — ее адрес в ЭССЯ).

**brōkati* [3, стр. 53–54]. Выделено **brukati*, объединяющее восточнославянские и польский языки (итератив или экспрессивное изменение корневого вокализма).

**bukariti* [3, стр. 87–88]. Выделено, возможно, производящее **bukarъ*, по ЭССЯ, известное лишь в болгарском: диал. *bukár* 'кабан, хряк' (известно и русскому: диал. *bukárъ* 'насекомое, букашка', также прозвище [СРНГ, вып. 3, стр. 264]).

dadъbogъ* [4, стр. 182–183]. От старопольского и древнерусского примеров отделено сербохорв. *Dabog*, бог земли, образованное с усечением формы императива в первой части сложения до чистого корня (da-bogъ*).

**dętelina* [5, стр. 27]. От основного списка отделены русские и украинское названия клевера, продолжающие более сложное суффиксальное образование **dętel-ov-ina* (скорее всего позднее, послепраславянское).

**drabъ*/мн. **draby* [5, стр. 100–101]. Разделены слова со значениями 'телега', 'лестница', 'решетка', с одной стороны, и 'лохмотья', 'оборванец', с другой (последнее — в верхнелужицком и украинском), обычно рассматриваемые в этимологических словарях порознь. Даже если имеются основания для их этимологического сближения, различия в лексической семантике слишком значительны, чтобы делать это без специальных оговорок.

**dylъ* II [5, стр. 201]. Выделено гипотетически объясняемое с помощью того же корня **dylъda* — суффиксальное производное или, по ЭССЯ, результат частичной редупликации корня. Известно, кроме цитируемых русского и белорусского, также в украинском и польском.

**dъnevъnъ(jъ)* [5, стр. 212]. Выделен производящий для прилагательного глагол **dъnevati*, объединяющий болгарский, сербохорватский, польский и все восточнославянские языки (в ЭССЯ украинский и белорусский глаголы *dъnovати* и *dъnivати* не приведены).

Прилагательные от **elъša* 'ольха' (**elъšanъ(jъ)*, **elъšeвъ(jъ)* [6, стр. 25]). Наряду с ними выделяется **elъšatъ(jъ)*: в.-луж. *wólšaty* 'ольховый' [Трофимович, 1974, стр. 363], ср. еще производное русск. диал. *ольшати́ник* 'ольшаник' [СРНГ, вып. 23, стр. 193].

**gajъka* [6, стр. 86–87]. Отделено омонимичное **gajъka* (II), орнитологический термин (сербохорватский и русский примеры).

**gromaditi* [7, стр. 137]. Выделены формы с «экспрессивным удвоением -dd- и последующей диссимиляцией -zd-» — **gromazditi*, кроме

чешского и русского языков отмечаемые в польском, украинском, белорусском, см.: [ЕСУМ, т. 1, стр. 601].

**gyľb* [7, стр. 221]. Допустимо разделение названий снегирия и слов с другими значениями ('вздор', 'стая волков', 'гурьбой'), возможно, омонимичных.

jьz nenada*/jьz nenady* [9, стр. 9]. Выразительное несовпадение ареалов форм, продолжающих праслав. **jьz nenady*, **jьznenadъnъjь* (весь славянский юг, чешский и словацкий языки), с одной стороны, и **jьznenadъbskъjь* (польский, словинский, а также украинский и белорусский языки), с другой стороны, что наглядно продемонстрировано картой 2 на стр. 10, позволило нам при вводе материала в компьютер, описывать эти изоглоссы раздельно.

katěньnъ(j)ь*/katěнь(j)ь* [9, стр. 135–136]. Отделена более простая (и древняя) форма (с континуантами в старославянском, древнерусском и украинском) как представляющая другой, по сравнению с прилагательными с суффиксом *-нь(j)ь*, словообразовательный тип.

kľisati* [10, стр. 48]. Объединение в пределах одной статьи этимологического словаря глаголов со значениями 'прыгать', 'бежать', 'ходить' и существительных 'плохо пропеченный хлеб', 'шкварки', 'жаркое' и др. проблематично, что специально подчеркнуто самими составителями, поэтому выделение **kľisa*/kľisъ* с возможной этимологической апелляцией к **kľip-s* и т. д. оправдано (отделен болгаро-македоно-сербохорватский ареал).

**kolovatъ(j)ь* [10, стр. 148]. Разведены очевиднейшие омонимы — производное от **kolъ* < **kolti* (болгарский, польский, русский, белорусский примеры) и производное от **kolo*, **kolese* (чешский, верхнелужицкий, польский и украинский).

**kolъčatъ(j)ь* [10, стр. 167–168]. Аналогично: производное от **kolъce* < **kolo* и **kolъсь* < **kolъ*, что и отмечено в тексте статьи.

koničina*/kon'ušina* (10, стр. 183). Даже непонятно, почему производные от **konikъ* и **kon'uxъ* были собраны в одну статью. Относительно редкий случай, когда в ЭССЯ семантика ('клевер') как основание для комплектации материала в статью оказалась приоритетной по сравнению с формой.

**kosa* II [11, стр. 133–134]. Выделены в самостоятельные позиции производные **kosakъ* (с добавлением украинского отражения), **kosacъ* и **kosanъ/-nica* (из категории «сюда же»).

**kъdъkъ* (12, стр. 52). Параллельно-суффиксальное **kъdyľb*, сохранившееся в чешском *kunděl* 'кусочек' и приводимое после диэремы «ср.», отделено от польских слов.

*kǫžьbь/*kǫželь/*krǫželь [12, стр. 80–81]. Разведены в разные изоглосные единицы континуанты «наиболее авторитетной и исконной» формы с корнем *krǫg- (западнославянские и, в виде производного, сербохорватский) и вторичной, но также праславянской формы *kǫž-, контаминировавшей с *kǫdelь.

*kǫ(n) [13, стр. 170]. Вариант *kǫn, получивший «самостоятельное развитие в болг.-макед. области», рассматривается как индивидуальная изоглосса.

*kǫdun'a/*kǫdul'a [13, стр. 174–175]. Из общего массива выделены в индивидуальную изоглоссу формы с -n- (болгарский, македонский, сербохорватский и древнерусский языки) как более близкие к греческому оригиналу (κυδώνια).

*kyda/*kydja [13, стр. 252]. Разведены нейотированная и йотированная (кашубско-словинский, русский языки) формы.

*laziti [14, стр. 64–66]. Мы решились примеры с бортническим значением 'вынимать соты', 'выбирать мед' (сербохорватский, польский, древнерусский, русский, белорусский) выделить в самостоятельную праславянскую изоглоссу.

*lepenь/*lepen'ь/*lepeno [14, стр. 119–120]. Все три заголовочных варианта рассмотрены как самостоятельные изоглоссы.

*leža/*ležь/*leže (?) [14, стр. 156]. Вариант *leže отклонен как проблематичный в отношении праславянской древности, а первые два разграничены.

*ležetьka/*ležetьkь [14, стр. 166]. От -k-производных отделена производящая форма адъективирующего причастия *ležetь(jь), сохранившаяся в чешском, словацком и, в дополнение к материалу статьи, в верхнелужицком (ležaty 'лежащий') [Трофимович, 1974, стр. 107].

*ležišče/*ležeвиšče (?) [14, стр. 166–167]. Варианты разграничены (последний – словацко-кашубскословинская изоглосса).

*lěxa/*lěxь [14, стр. 184–185]. Словенско-великорусская переключка существительных мужского рода 'участок пашни', 'полоса', 'знак границы засеваемой полосы' дает основания для выделения частной изоглоссы.

*lějьka/*lějьkь [14, стр. 189–191]. Вариант в мужском роде имеет узкий серболужицко-польский ареал и отделен от остальных слов.

*lěka/*lěkь II [14, стр. 192–193]. Формы мужского и женского рода имеют неодинаковую географию и разграничены.

*lišajь [15, стр. 152–153]. Более древнее *liša, отмеченное как кашубско-болгаро-македоно-сербохорватская изоглосса [Поповская-Таборская, 1987, стр. 157], отделено от более позднего остального материала.

**iti* (*se*) [15, стр. 157]. Отграничен вторичный вариант огласовки **ljati*, «возможно, представляющей собой еще праславянский вариант».

lubbyca*/lubbyś* [15, стр. 187]. Вариант в женском роде отражен в статье только сербохорватским репрезентантом *lúnica* 'жена; возлюбленная'. Однако наличие в украинских диалектах наименования любистка *лѹбѣца* (см.: [ЕСУМ, т. 3, стр. 319] со ссылкой на словарь Желеховского) позволяет выделить на уровне формы украинско-сербохорватскую параллель.

ludina*/ludinъ* [15, стр. 192]. Мы решили развести варианты на *-a* и *-ъ*, хотя выразительных ареалов такое разграничение не дает.

ludižer'ъ*/ludožerъ* [15, стр. 193]. От основного материала статьи отделено «обособленное по структуре чеш. *lidožrout* 'людоед' со старым прич. наст. действ. **žbrqъ* во второй части сложения».

ludъ*/ludь*/**luda*/**ludo*/**ludьje* [15, стр. 194–197]. Сомнительные с точки зрения праславянской древности формы **luda* и **ludo* исключены из анализа. Остальные три формы введены отдельно, из них лишь форма **ludъ* является узкорегionalной (север и восток великорусской территории).

luxa*/luxъ*/**luxo*/**luša* [15, стр. 206–207]. Выделен вариант с *-š-*, представленный в русск. *люша* 'грязнуля', 'нечто мокрое' и, возможно, сербохорватский топоним *Luša*.

lūi*/lū'u* [15, стр. 208–209]. Формальные варианты междометия разграничены.

lutača*/lutakъ*/**lutaka* [15, стр. 220]. Разграничены варианты палатализованный и непалатализованный.

lutika*/lutikъ* [15, стр. 223–224]. Отделена форма женского рода, представленная только южнославянскими континuantами.

lobazъ*/lobozъ* (**loboza*), **lobuzъ*/**lobъzъ* [15, стр. 239–240]. Вариативность в вокализме суффикса послужила основанием для выделения четырех различных более узких изоглосс с несовпадающими составами охватываемых ими идиомов.

Нетрудно заметить, что чуть ли не половина статей, материал которых был нами введен в ЭВМ дифференцированно, принадлежит последним двум из обработанных нами выпусков ЭССЯ. Это объясняется тем, что начиная с 14-го выпуска варианты формы, описываемые в пределах одной статьи и отраженные в ее заголовке размещением их праславянских реконструкций через косую черту, группируются в отдельные абзацы. Такой способ представления вариантных форм, напоминающий, кстати сказать, манеру краковского Праславянского словаря и применя-

емый, правда, только в словарных статьях, написанных Ж. Ж. Варбот, нельзя не признать куда более удобным и наглядным, чем способ, принятый в первых тринадцати выпусках, при котором членение формально не всегда однородного материала внутри статьи проводилось только по одному основанию — принадлежности тому или иному языку (плюс отдельно формы «сюда же»). Можно лишь сетовать, что новый принцип группировки материала по двум его измерениям — конкретноречевой отнесенности и морфолого-парадигматическим признакам (реже — словообразовательному или морфонологическому варьированию) — введен в ЭСЯ не с самого начала и применяется не всеми его авторами.

В итоге — после изъятия из рассмотрения нескольких сотен статей, славянская древность заголовочных реконструкций которых вызывает сомнения, а также статей, материал которых подозревается в книжном происхождении, и разделения материала нескольких десятков статей на частные изоглоссы — образовался список из 7 557 позиций (лексических изоглосс), который и был подвергнут статистическому анализу.

6. Состав исследуемых идиомов (разные версии)

Перечень идиомов, степень генетической близости между которыми (попарно) выявляется в настоящей работе, в основном, естественно, совпадает с перечнем славянских языков, как они рассматриваются в нашем главном источнике — ЭСЯ (см. разделы библиографии в 1-м выпуске словаря [стр. 10–24]): старославянский, болгарский, македонский, сербохорватский, словенский; чешский, словацкий, верхнелужицкий, нижнелужицкий, полабский, польский, кашубско-словинский; русский, украинский, белорусский. Хронологические уточнения «др.» («древне-») и «ст.» («старо-») при названиях языков в списках континуантов заголовочных реконструкций (старославянский, конечно, не в счет) во внимание нами не принимаются, то есть, например, древнесловенская или старобелорусская лексика автоматически зачисляется в материал соответственно словенского и белорусского языков.

Однако такого рода автоматизм невозможен в случае древнерусского языка, поскольку он трактуется как предшествующее состояние не одного, а трех современных языков. Лексика древнерусского языка, который представлял собою «промежуточный праязык» и очевидное полидиалектное образование, не может быть в равной мере приписана всем трем языкам-потомкам. Поэтому древнерусский язык в нашем

перечне славянских идиомов целесообразно выделить в самостоятельную единицу.

Касательно древнерусского языка следует сделать еще одно уточнение. Его мы понимаем здесь как язык письменных памятников, относящихся к определенному месту и времени. Реконструкция лексики древнерусского языка в полном объеме и с использованием всех ее возможных источников в наши задачи не входит. К древнерусскому словарю мы относим только лексику, засвидетельствованную памятниками известной хронологической приуроченности и описываемую соответствующими словарями. Это значит, что лексика современных восточнославянских языков и диалектов как источник древнерусской лексической реконструкции не рассматривается, хотя, логически рассуждая, все праславянское, сохраненное современными великорусским, белорусским и украинским языками (и нашедшее свое место в ЭССЯ), конечно же, присутствовало в древнерусском языке, из которого они развились (и понимаемом в таком случае шире, чем язык, данный нам в письменных памятниках). При подключении фактов современных восточнославянских языков к числу источников реконструкции древнерусского словника как составной части праславянского лексикона, он (древнерусский словарь) представлял бы собою сумму лексики древнерусских памятников и всей праславянской (по крайней мере) лексики русского, белорусского и украинского (с явным численным перевесом в пользу именно этой последней), подобно тому, как полный праславянский лексикон восстанавливается в виде суммы праславянских словников отдельных поздних славянских языков. Это по статусу резко выделяло бы древнерусский из перечня славянских идиомов, рассматриваемых в настоящей работе.

Опора же только на древнюю письменную фиксацию лексики сближает статусы древнерусского и старославянского языков.

Перечень славянских идиомов может быть принят в двух версиях — с включением в него древних письменных старославянского и древнерусского и с их исключением из списка. Обе эти версии применены ниже в наших расчетах.

Другая оговорка касается пометы «диал.» («диалектное») в списках отражений заголовочной праформы в современных языках. Для всех языков, кроме русского, эти диалектные формы так же автоматически вливались в лексический корпус соответствующего языка.

Что же касается великорусского, то по отношению к его диалектной лексике введен двойной стандарт. Ввиду чрезвычайной обширности

русской территории мы сочли возможным (тем более что такая возможность предоставляется характером описания русской диалектной лексики в самом ЭССЯ) выделить внутри русского языка оба его наречия – северно- и южновеликорусское – в самостоятельные идиомы (лексика, не имеющая пометы «диал.» в таком случае приписывалась обоим наречиям, хотя иной раз здесь могли возникать немалые сомнения). Однако поскольку лексика среднерусских говоров, не составляющих самостоятельного наречия, при этом разделении повисает «неоприходованной», среднерусские говоры тоже были условно сочтены третьим «наречием». Границы наречий определялись по работам: [Русская диалектология, 1964; Захарова – Орлова, 1970]. Считаем необходимым отметить, что такой выбор не является безупречным. Многие диалектологи полагают традиционное членение, при котором полоса древнерусских говоров выглядит более узкой, правильнее отражающим реальные отношения между великорусскими диалектами, ср.: [Хабургаев, 1980, стр. 164]. Н. Н. Пшеничнова, предпринимающая пересмотр существующей классификации великорусских диалектов на основе статистического анализа данных, которые составляют базу «Диалектологического атласа русского языка», находит, что западный участок границы северновеликорусского наречия должен прочерчиваться южнее, чем это определено в работах школы Р. И. Аванесова, и захватывать часть говоров новгородской и гдовской групп (устное сообщение). Этот результат ближе к тому начертанию границ в этой зоне, который имеется на «Диалектологической карте русского языка в Европе» Московской Диалектологической Комиссии, см.: [Опыт диалектологической карты, 1915]. Не исключено, что принятие традиционного варианта или варианта Н. Н. Пшеничновой привело бы к несколько более точным статистическим наблюдениям и более адекватной итоговой картине.

Второй стандарт предусматривал толкование великорусского языка как целостного образования, без разделения его на наречия. Для этого материал всех трех крупных диалектных подразделений, как он вводился в ЭВМ в предыдущем варианте, с помощью несложной операции сливался компьютером в единый список слов.

В целях восстановления диалектных отношений в позднепраславянском языке было бы весьма желательным дифференцированное представление диалектной лексики не только русского, но и других славянских языков, в особенности языков со сложной эволюцией, с заметными конвергентными процессами в их возникновении (сербохорватский, сло-

вацкий и т. д.). Чем более дробным и пространным был бы их список, тем, по-видимому, тоньше и точнее рисовалась бы итоговая картина славянского языкового родства⁷. К сожалению, территориальные пометы в ЭССЯ даются только к русским диалектным формам. Устанавливать же конкретно-территориальную принадлежность диалектного материала иных языков по региональным лексиконам мы не имели возможности ввиду колоссального объема такого труда (по сути это было бы повторением всего предварительного, отборочного этапа работы коллектива составителей ЭССЯ). Поэтому диалектная («наречная») дифференциация лексики осуществлена у нас только для одного языка.

Таким образом, возникли следующие четыре варианта представления славянских идиомов.

(а) Максимальная версия списка славянских идиомов включала в себя восемнадцать единиц: четыре современных южнославянских языка плюс мертвый старославянский, семь западнославянских языков и шесть восточнославянских идиомов — древнерусский и три современных языка, из которых великорусский фигурирует в виде трех диалектных подразделений (версия получила обозначение SLAV). (б) Минимальную версию образовывали те же языки, с исключением книжного старославянского, древнерусского как промежуточного праязыка и представлением великорусского в виде целостного идиома, — всего четырнадцать языков (версия MIN). По шестнадцати идиомов фигурируют еще в двух «средних» версиях: (в) без старославянского и древнерусского, но с отдельно представленными великорусскими наречиями (версия CON), и (г) с представлением русского языка как целого и включением книжных старославянского и древнерусского (версия RUS).

Итак, объем обследованного нами фрагмента ЭССЯ — 7 557 словарных статей. Между поздними славянскими языками эта лексика, входящая в праславянский лексический фонд, распределена весьма неравномерно. Анализировать соотношения объемов праславянских словариков поздних славянских языков мы будем в следующих главах. Здесь же, в порядке предварения, приведем основные данные. Из указанного числа праславянских слов отмечено:

⁷ Ср. список славянских идиомов, классифицируемых с помощью количественного анализа фонетических изоглосс, у В. Н. Чекмана и О. С. Широкова ([Чекман — Широков, 1962]; реферат этой работы и сопоставление ее результатов с нашими см. у нас ниже).

– южнославянские языки: в старославянском – 1 124, в болгарском – 3 262, в македонском – 2 035, в сербохорватском – 4 568, в словенском – 3 519;

– западнославянские языки: в чешском – 4 264, в словацком – 2 933, в верхнелужицком – 1 895, в нижнелужицком – 1 574, в полабском – 452, в польском – 3 350, в кашубско-словинском – 1 683;

– восточнославянские языки: в древнерусском – 2 681, в великорусском – 4 956 (в северновеликорусском наречии – 4 168, в среднерусских говорах – 3 834, в южновеликорусском наречии – 3 833), в украинском – 3 905, в белорусском – 3 288 праславянских лексем.

7. Определение статистической надежности материала

Quaeris, quot, mihi basiationis
Tuae, Lesbia, sint satis superque.

Catullus. VII, 1–2

(Ты спрашиваешь, Лесбия, сколько твоих поцелуев было бы для меня достаточно и более чем достаточно. Катулл)

Является ли объем отобранного к анализу материала достаточным для того, чтобы наши наблюдения и выводы можно было считать статистически надежными?

Воспользуемся стандартными в статистике способами определения минимальной необходимой выборки. Частоты появления помет (языковых атрибуций) «старославянское», «болгарское», «македонское»... «украинское», «белорусское» в списках континуантов славянских праформ, содержащихся в ЭССЯ, довольно велики – от десятков на тысячу словарных статей для полабского языка до нескольких сотен на тысячу для языков «средних» и «крупных».

Рассчитаем объем минимальной выборки для полабского языка, у которого доступный реконструкции праславянский словарь наименьший. Пробные проверки частоты встречаемости полабских слов в словарных статьях ЭССЯ дают величины, колеблющиеся около 0,06. Расчет осуществляется по рекомендуемой (см.: [Пиотровский – Бектаев – Пиотровская, 1977, стр. 295]) формуле:

$$N_{\min} = ((z_p^2 (1 - f)) / \delta^2 f), \quad (19)$$

где N_{\min} — объем минимальной выборки (ориентированный в нашем случае на данные о встречаемости в статьях ЭССЯ полабских слов), f — частота (0,06) встречаемости признака, относительно которого определяется N , p — заданная надежность (95%⁸, для которой табличное значение квантиля z_p равно 1,96) (см. [там же, стр. 369]; ср.: [Носенко, 1981, стр. 61, 150–151]), δ — заданная относительная ошибка наблюдения (10%⁹).

Заданные величины являются стандартными: Б. Н. Головин, например [Головин, 1971, стр. 61], считает приемлемой для статистических подсчетов на лингвистическом материале надежность в 95%, а Пиотровский с соавторами [Пиотровский — Бектаев — Пиотровская, 1977, стр. 296–297] для δ рекомендует величины 0,1; 0,2 и даже 0,33 [Головин, 1971, стр. 57]: 5–10%, то есть 0,05–0,1).

Подставляя численные значения в формулу (19), получаем:

$$N_{\min} = 1,96^2 \times (1 - 0,06) / 0,1^2 \times 0,06 = 6018,5 \approx 6000,$$

то есть для того, чтобы при заданных δ и p (z) статистические наблюдения над праславянской лексикой полабского языка в ее отношениях с праславянской лексикой других языков по данным ЭССЯ были надежными, достаточна выборка из этого словаря объемом примерно 6 000 словарных статей. Обследованный объем словаря превосходит эту величину на четверть ($N_{\text{ЭССЯ, вып.1–15}} = 1,25 N_{\min(\text{полаб.})}$).

Если бы лексические материалы полабского языка были исключены из исследования, а все остальные словарные данные были сохранены, то наименьшей встречаемостью в ЭССЯ в таком случае характеризовалась бы праславянская лексика старославянского языка, и N_{\min} должно было бы выражаться, при сохранении заданных величин δ и p (z), числом 2 200 словарных статей, то есть почти в 3,5 раза меньшим, чем обработанный нами материал, и соответствующим объемом первых четырех выпусков словаря.

Таким образом, можно считать, что объем привлекаемого в нашем исследовании фрагмента ЭССЯ вполне достаточен для получения надеж-

⁸ Означает: при статистической однородности испытательных выборок в 95 случаях из 100 заданная ошибка не будет превзойдена.

⁹ Означает: относительная ошибка (определяемая как отношение абсолютной ошибки к выборочной средней частоте, выраженное в процентах или десятичной дробью) составляет 10% от средней частоты.

ных статистических сведений и последующих экстраполяций, — пользуясь переосмысленным выражением Саллюстия, *satis verborum*. Но даже если для полабского языка заданная относительная ошибка (0,10) покажется либеральной, то для всех прочих языков обследованный объем словаря ЭССЯ позволяет ввести более жесткие ее величины (от 0,05 для старославянского до 0,02 и менее для русского, сербохорватского и чешского языков при сохранении прежнего уровня надежности в 95%).

Глава 5

Из квантитативнотипологических наблюдений над лексикой славянских языков (праславянское наследие)

1. Необходимость типологических уточнений

Компаративист, занимающийся выяснением природы и силы связей в семье или группе родственных языков, как правило, имеет дело с неполным и зачастую довольно разнородным материалом, особенно если его внимание сосредоточено на фактах лексики и словообразования. Это обстоятельство существенно осложняет прямое сопоставление массивов данных, характеризующих каждый из родственных языков по отдельности.

Славянские языки являются в указанном отношении достаточно ярким примером. На одних языках в славянском мире говорят десятки (а в случае с русским — и более) миллионов людей, лексиконы этих языков складываются из многих сотен тысяч единиц; другие же (например, серболужицкие) обеспечивают коммуникацию нескольких десятков тысяч человек, и их словарные составы несравненно скромнее. Одни языки, привлекаемые компаративистом к анализу и сопоставлению, являются мертвыми, с конечным списком доступной исследованию лексики, другие — живыми, развивающимися и непрестанно увеличивающими свои словари. В одном случае (старославянский) мы имеем дело с языком исключительно книжным, другим языкам (скажем, полабскому) письменность как реализация социокультурных потребностей была вовсе неизвестна. С лексическими массивами одних языков мы знакомы исключительно по памятникам, и надежд

на значительное обогащение наших представлений об их лексике у нас чрезвычайно мало (оно возможно лишь в случае неожиданного обнаружения совершенно новых источников), материалы же других языков с каждым годом возрастают за счет новейших фактов, добываемых в диалектологических экспедициях, с изданием все новых и новых глоссариев¹, так что суждения об отсутствии того или иного слова или словообразовательной конструкции в таком-то языке или такой-то языковой группе постоянно рискуют устареть². И так далее: характеристические противопоставления «крупный — малый», «живой — мертвый», «письменный — устный», «литературный — диалектный» можно дополнить типологическими антитезами «с древней культурной традицией (скажем, болгарский) — относительно молодой (македонский)», «центральный (в плане языковой географии) — периферийный», «находящийся в зоне интенсивных родственноэтнических контактов — лишенный условий для взаимодействий подобного рода»...

Непринятие во внимание типологических несходств, о которых идет речь, может приводить к неверно осуществляемым сличениям данных (особенно количественных), относящихся к разным языкам, а отсюда — к смещенным результатам, а то и вовсе к ошибочным выводам, что мы уже пытались показать.

Простейший случай. Если оценивать взаимную близость между родственными языками, основываясь на прямом сравнении абсолютных лексических связей праязыкового возраста, то можно прийти к выводу о том, что русский и сербохорватский языки находятся в гораздо более тесных генетических отношениях, чем верхне- и нижнелужицкие языки между собою: в первых пятнадцати выпусках ЭССЯ эксклюзивные лексические и словообразовательные связи последней пары языков характе-

¹ Так может обстоять дело не только с живыми диалектами. Достаточно неожиданное и чрезвычайно радующее исключение из общего порядка вещей представляет собой ситуация с исследованием древненовгородского диалекта, восстанавливаемого на основании берестяных грамот. К настоящему времени раскопано не более 2% грамот, лежащих в земле Новгорода. Каждый год Новгородская археологическая экспедиция находит в среднем 19 новых грамот, число их уже заметно превышает 700, совокупная длина найденных грамот составляет около 11 тысяч словоупотреблений, а объем словаря — около 2,5 тысяч лексем (см.: [Зализняк, 1988а, стр. 93, 94, 99]). По данным на конец 1993 г. количество грамот, раскопанных в Новгороде, достигло 753, а вместе с грамотами из Старой Руссы, отражающими тот же диалект, — 779, длиной в 12 тысяч словоупотреблений.

² Чтобы убедиться в этом, предлагаем читателю сопоставить предварительный список «праславянских диалектизмов русского языка», включенный в проспективный доклад О. Н. Трубачева на V Международном съезде славистов [Трубачев, 1963б] или списки соответствий в пробных статьях Проспекта «Этимологического словаря славянских языков» [Трубачев, 1963а] с материалами реализующегося ЭССЯ: разница более чем заметная.

ризируются числом всего в 12 изоглосс (цельнолексемных соответствий), в то время как исключительные цельнолексемные связи между сербохорватским и русским языками определяются на данном участке праславянского словаря 74-мя лексико-словообразовательными параллелями. Вряд ли кто-либо из славистов согласится с тем, что на какой-бы то ни было шкале измерения сербохорватский и русский языки генетически в шесть раз ближе друг к другу, чем два серболужицких языка. Очевидно, что правильным использованием этих цифр было бы их соотнесение с объемами списков праславянских лексем в словарях каждого из языков, то есть учет одного из весьма важных типологических признаков исследуемых идиомов — величины доступного реконструкции праязыкового лексикона, который сохраняется в каждом из них.

Еще один пример, который должен показать необходимость осторожности в использовании массовых данных в глоттогенетических исследованиях. В ЭССЯ [вып. 9, стр. 8–105] реконструируется 305 праславянских лексем с начальным **jъz-* (и сочетаний с предлогом **jъz*, позднее слившихся в одно слово). 129 из них, или 42,3%, в списках соответствий заголовочным праформам содержат старославянские иллюстрации. Это не может не обратить на себя внимание: концентрация старославянской лексики в данном участке словаря почти в три раза выше, чем в словаре в целом (14,9%, по первым пятнадцати выпускам). Если сюда добавить и те словарные статьи, где приводятся, в случае отсутствия старославянских примеров в перечнях континуантов, иллюстрации с пометами «церковнославянское» (преимущественно русского извода), то число словарных статей с начальным **jъz-*, включающих лексику культовых языков, увеличивается до 204, или 66,9%, что также значительно выше «нормы».

Конечно, эти цифры пока еще ничего не говорят о «неестественности», «ненародности» форм на **jъz-*: старославянский и церковнославянский как его продолжение могут оказаться попросту очень склонными именно к древней лексике с этим префиксом в силу особых их потребностей, удобства таких конструкций для передачи определенных (и существенных с точки зрения сфер функционирования этих языков) типов значений и т. п., ср.: [Львов, 1976, стр. 83, 84]. Но объяснять описанную ситуацию только лишь повышенным влечением старославянского языка к тем или иным деривационным схемам и моделям, по-видимому, недостаточно. Во всяком случае имеют право на существование подозрения в том, что сюда могли попасть и образования (в том числе и послепраславянские), распространившиеся в народной речи под влиянием книжных и культовых сфер языка, тем более что на позднее распространение во многих случаях,

помимо сугубой деривационной и семантической прозрачности слов, может указывать и преобладающая в примерах абстрактность денотата (ср.: **ъbzbytkъ*, **ъzходъ*, **ъzкусъ*, **ъzvѣтъ* и др., без старославянских параллелей — **ъzguba*, **ъzkaza*, **ъzprava* и т. д.). Но так или иначе, использование в качестве исходного материала для лингвогенетических построений некоторых групп лексики, где пропорции в распределении ее по языкам книжным и «народным» сильно нарушены, требует большой осмотрительности: внешняя архаичность форм еще не является гарантом их оригинальности для данной языковой традиции.

Обширный материал новейших этимологических изданий и прежде всего ЭССЯ, даже если ограничиться только опубликованными его выпусками, уже сейчас достаточен для формирования предварительных представлений о типологических особенностях лексиконов современных славянских языков в их отношении к праязыковому лексическому наследию.

2. Объем праславянского лексического корпуса

Сколько их!..

Пушкин. «Бесы»

Прежде всего необходимо уточнить, что речь идет не о словарном «запасе» некоего цельного идиома, из которого («запаса») потомки праславянского языка — современные славянские языки — «извлекли» или «донесли», «сохранили» лишь некоторую его часть. Единными и цельными этот идиом и его лексический корпус никогда не были. Словарь праславянского языка может реконструироваться только как теоретико-множественная сума словарных составов, как объединение множеств праславянских «*słownictw*» более поздних продолжений, то есть славянских языков, включая «несовременные» старославянский и полабский.

15-й вып. ЭССЯ оканчивается серединой буквы «L» (словом **lokašь*), вышли еще четыре выпуска, готов к изданию 22-й, работа над следующими томами включает пополнение картотеки, поэтому технически праславянский словник еще представляет незавершенный список. Каков будет его окончательный объем, точно сказать пока трудно.

Ф. П. Филин определяет его следующим образом: «В семи опубликованных (к моменту подсчетов. — А. Ж.) выпусках словаря О. Н. Трубачева содержится 3 771 словарная статья на буквы А — С... Отыскивая в словаре М. Фасмера слова, засвидетельствованные только в восточнославян-

ских языках и диалектах (включая древние письменные источники), я обнаружил их около 150, а в статьях на буквы, соответствующие буквам напечатанных выпусков словаря О. Н. Трубачева, их оказалось около 25, причем распределение восточнославянизмов по сравниваемым статьям обоих словарей более или менее равномерно. Если указанная пропорция будет сохраняться и дальше, то в словаре О. Н. Трубачева, вероятно, должно быть около 22 626 словарных статей, соответственно и лексических праславянизмов. Значит, в праславянском языке, по современным данным, было что-то около 22 000 слов» [Филин, 1983, стр. 276–277].

Подсчеты Ф. П. Филина следует признать весьма приблизительными и, скорее всего, занижающими итоговую цифру. Во-первых, в семи выпусках ЭССЯ не 3 771, а 3 838 слов (уже из этого, при верности вычисленной пропорции, конечное число составит более 23 000 слов). Во-вторых, словарь Фасмера, которым пользовался Филин для выявления необходимых численных соотношений между праславянской и восточнославянской лексикой, не рассматривает, в отличие от ЭССЯ, словообразовательно прозрачные явления. В диапазоне «А – С» латинского алфавита относительно высокую продуктивность обнаруживают только славянские префиксы *bez- и *do-, за пределами этого диапазона находятся такие высокопродуктивные начальные элементы, как *jbz-, *na-, *ne-, *ob-, *orz-, *ot-, *pa-/po-, *per-, *pri-, *pro-, *sz-, *u-, *vz(n)-, *vy-, *za-, сложения с которыми, в силу их прозрачности, в этимологическом словаре Фасмера в огромном большинстве своем места не нашли. Не отражаются у Фасмера, как правило, и двукорневые сложения, в то время как composita составляют значительную часть праславянского словаря. К тому же практика показывает, что длительные и объемные лексикографические предприятия завершаются дополнительными списками слов, пропущенных при составлении предыдущих томов или выпусков. Следовательно, реально доля деривационно непрозрачных слов, заслуживших помещения в этимологическом лексиконе М. Фасмера, гораздо ниже, чем это вычисляется с опорой на начальные участки славянских словарей.

Целесообразно привести здесь количественные прикидки польских лексикографов. В предисловии в 1-му тому краковского Праславянского словаря сообщается, что «litery A – B przedstawiają jakaś 1/10–1/12 część całości» [Праславянский словарь, т. I, стр. 10]. В этом диапазоне в Праславянском словаре помещена 1 021 статья, включая отсылочные. Следовательно, конечный объем словаря может составить 10–12,5 тысяч лексем. Однако, имея в виду характер следующих томов краковского словаря, можно уверенно сказать, что эта оценка является сильно заниженной.

Допустимо думать поэтому, что конечная цифра, характеризующая величину словарного состава праславянского языка, получится близкой к 25 тысячам единиц, если не значительно более. Экстраполяции, осуществляемые в следующем параграфе, мы будем строить, исходя из гипотетических объемов праславянского словаря, полученных как нами, так и Ф. П. Филиным.

3. Объемы праславянских словариков отдельных славянских языков

...с бедным словарем
Ты более всего нуждаешься в защите...

Виктор Кривулин

Эти сведения, пока гипотетические, несложно получить, экстраполировав на тот или иной предполагаемый суммарный объем праславянского словаря данные о предъявленности каждого славянского языка в списках континуантов заголовочной праформы в уже опубликованной части ЭССЯ.

В таблицу 15 (см.) включены данные обо всех славянских языках, как они представлены в максимальной версии нашего списка (SLAV).

Экстраполяции, произведенные в таблице 15, показывают, что объемы доступных реконструкции праславянских словариков живых славянских языков, включая в их состав, разумеется, и диалектную лексику, колеблются от 4–5 тысяч лексем (нижнелужицкий, кашубский языки) до 15–16 тысяч лексем (русский язык). В свете этих оценок пассажи вроде «Слова, унаследованные из общеславянского (= праславянского. – А. Ж.) языка, не составляют и одного процента лексики современных языков» [Булахов – Жовтобрюх – Кодухов, 1987, стр. 243] воспринимаются, пожалуй, как слишком решительная литота, если, конечно, вывести из поля зрения циклопические арсеналы узкотерминологической лексики вроде химической или фармакологической номенклатуры. Принимая величину лексикона современного русского языка равной примерно 500 000 слов, несложно определить, что праславянское лексическое наследие в нем составляет около 3%. В кашубско-словинском лексиконе, объем которого в 60 000 слов «можно считать едва ли заниженным» (см.: [Супрун, 1983, стр. 11]), к праславянскому наследию относится до 10% словарных единиц.

Таблица 5

(Пояснения: 1 — название идиома; 2 — количество праславянских лексем в данном идиоме по 1–15 выпускам ЭССЯ; 3 — их доля в общем объеме обследованной лексики (в процентах); 4 — гипотетический объем праславянского словника данного идиома при экстраполяции на суммарный объем праславянского словаря (а) в 22 000 лексем и (б) в 25 000 лексем.)

1	2	3	4	
			(а)	(б)
ст.-слав.	1124 сл.	14,87%	3,2 тыс.сл.	3,7 тыс.сл.
болг.	3262	43,17	9,5	10,8
макед.	2035	26,93	5,9	6,7
с.-хорв.	4568	60,45	13,3	15,1
словен.	3519	46,57	10,2	11,6
чешск.	4264	56,42	12,4	14,1
словац.	2933	38,31	8,5	9,7
в.-луж.	1895	25,08	5,5	6,3
н.-луж.	1574	20,83	4,6	5,2
полаб.	452	5,98	1,3	1,5
польск.	3350	44,33	9,8	11,1
кашуб.	1683	22,27	4,9	5,6
др.-русск.	2681	35,48	7,8	8,9
русск.	4956	65,58	14,4	16,4
сев.-в.-рус.	4168	55,15	12,1	13,8
ср.-рус.гов.	3834	50,73	11,2	12,7
юж.-в.-рус.	3833	50,72	11,2	12,7
укр.	3905	51,67	11,4	12,9
белор.	3288	43,51	9,6	10,9

По тому, какую долю всей праславянской лексики «сохраняет» каждый славянский язык, их можно условно разделить на четыре группы:

1) «крупные» языки — охват праславянского словаря в 50% и более: великорусский (оба наречия также относятся к «крупным» идиомам), сербохорватский, чешский и украинский;

- 2) «средние» — от 30% до 50%: словенский, польский, белорусский, болгарский, словацкий и древнерусский;
- 3) «малые» — менее 30%: македонский, верхнелужицкий, кашубско-словинский, нижнелужицкий и старославянский;
- 4) «реликтовый» — полабский.

Мы не случайно прибегли здесь к обозначениям членов полученного ряда, указывающим на «величину» языка. Точнее, речь идет о величине всего словарного состава каждого данного языка, а не только его праславянского фонда. Достаточно надежных сведений о размерах лексиконов отдельных славянских языков (за исключением, пожалуй, мертвых старославянского и полабского — языков с конечным и в общем очень небольшим числом известных текстов; собственно, сами термины «старославянский» и «полабский» определяются со ссылкой на жестко дефинированные корпуса источников, см: [Вайан, 1952, стр. 19–22; ван Вейк, 1957, стр. 42–49; Цейтлин, 1977; Супрун, 1987; Олеш, 1959, 1962, 1967]) не имеется, хотя предпринимались, и неоднократно, попытки их оценки по косвенным данным (из последних по времени см., например, [Супрун, 1983]). Тем не менее очевидно, что лексиконы русского и сербохорватского, включая и диалектную лексику, несравненно пространнее, чем словарные составы македонского, нижнелужицкого или кашубского.

Хорошо просматривается положительная корреляция между долей праславянского лексикона, приходящейся на словарный состав данного языка, и (оцениваемым, к сожалению, на глазок) объемом всего его лексикона в целом.

Наши данные существенно отличаются от соотношений, которые определены другими исследователями (см.: [Лер-Сплавинский, 1938; Лер-Сплавинский, 1954; Орлось, 1958; Попович, 1960; Радева, 1960] и др.)³.

Окончательно рухнули представления, совершенно неосновательные, но державшиеся на удивление долго, вплоть до настоящего времени, — о старославянском языке как самом богатом «общеславянским достоянием» [Срезневский, 1866, цит. по: Виноградов, 1977, стр. 44].

Ф. Копечный (см.: [Копечный, 1964, стр. 53]) полагает, что наилучшим образом праславянский лексический фонд сохраняется в чешском и словацком языках, в то время как у нас по объему праславянского фонда

³ Очень соблазнительно отметить совпадение наших «предсказаний» относительно объема праславянского словника нижнелужицкого языка с данными О. Н. Трубачева: последний называет для нижнелужицкого цифру «не менее 4 500 слов (примерно 10% слов из словаря Муки)» [Трубачев, 1963в, стр. 159].

чешский уступает русскому и сербохорватскому, а словацкий — всем восточнославянским, южнославянским, кроме македонского, а из западнославянских, помимо чешского, — польскому, то есть очень далек от того качества, которое, по мнению Ф. Копечного, проявляется в нем «наилучшим образом».

Т. Г. Линник [Историческая типология, 1986, стр. 201–202] приходит к выводу, что наибольшей сохранностью праславянская лексика отличается в сербохорватском, словенском, чешском и словацком (отметим выразительный «ареал!»), наименьшей — в верхнелужицком и белорусском, остальные языки составляют промежуточную зону. Выводы Т. Г. Линник, однако, хотя и получены на основе количественного анализа данных ЭССЯ, неубедительны из-за зильных ограничений в отборе лексического материала: подсчеты проводились только на массиве лексики с начальным *b-*, и при этом совершенно неоправданно «учитывались только слова литературных языков, данные диалектов игнорировались как несущественные для литературных языков». Остается загадочной цель подсчетов с такой селекцией материала. Если сопоставлению подвергаются литературные языки, можно ли считать для их типологии степень сохранения ими праславянской лексики очень уж важным параметром? (Косвенным образом, однако, через связь с ранней/поздней историко-культурной специализацией идиома, наличие индивидуальной праславянской лексики в нем может служить таким типологическим параметром, см. ниже, п. 3 этой же главы).

Столь же недостоверными нужно признать установление меры лексической «консервативности» того или иного языка путем обращения к частотным словарям [Историческая типология, 1986, стр. 202–204]: во-первых, в попытках этого рода дело опять-таки касается литературных языков, а во-вторых, речь здесь должна идти не о сохранении лексики или ее «исчезновении», а о ее перемещении по шкале употребительности, что, согласимся, несколько иная материя. Кorteж языков, определенный Т. Г. Линник с опорой на фреквентные списки: «словацкий — чешский — украинский — русский — польский» (первый член последовательности — наиболее «консервативен», «утраты» праславянской лексики в последнем — почти катастрофичны; анализировались частотные словари только этих пяти языков), — по нашим данным, имеет другой порядок: русский — чешский — украинский — польский — словацкий (см. выше).

Вообще же следует, несколько отвлекаясь в сторону теоретического уточнения позиций, заметить, что термины «степень сохранности», «ут-

раты» и под. часто скрывают за собою неизжитый рецидив стремительно устаревающей концепции диалектной монолитности праязыка и потому далеко не всякий раз удобоприменимы в описании отношений между словарем праславянского языка и праславянским словарем какого-либо данного (современного) славянского языка. Ведь отсутствие в данном языке того или иного праславянского слова, реконструируемого на основании данных других языков, не всегда означает лексическую утрату: то слагаемое (диалект) праязыка, из которого развился данный идиом, вовсе не обязательно характеризовалось наличием этого слова в его лексиконе. Даже весьма консервативный в лексическом отношении язык может быть отмеченным сравнительно незначительной долей от суммарного количества праязыковой лексики. Употребляя термины «сохранность», «утрата» и т. п., видимо, нужно делать оговорку касательно той лексической (под)системы, от которой ведется отсчет⁴.

4. Специфичность праславянского лексического наследия отдельных славянских языков (объемы дифференциальных словарей)

Индивидуальность лексической физиономии каждого языка обуславливается наличием в его словарном составе слов, которые не имеют прямых соответствий в других родственных языках. Конечно, своеобразие лексической системы языка создается не только индивидуальной лексикой. Не менее важны межъязыковые различия в семантике контингуантов одной и той же праязыковой лексемы (соответствий), в структуре лексико-семантических парадигм, сочетаемости и по множеству иных измерений. Однако отправным моментом в установлении специфичности лексикона какого-либо языка является обращение именно к «материальной» его стороне — выявление дифференциального слоя лексики. Наиболее существенно для компаративистских построений указание на такие индивидуальные лексемы, восходящие к праязыковой древности.

⁴ Наши упреки одному из авторов монографии «Историческая типология славянских языков» не ограничиваются лишь сказанным. Попутно можно отметить неточности в историко-лингвистической квалификации конкретной славянской лексики. Так, Т. Г. Линник безосновательно, на наш взгляд, выводит за пределы унаследованного праславянского лексического фонда многие слова современных языков (например, польск. *dziedzina*, *storień*, русск. *большой*, *письмо*, укр. *жінка*, *людина*, *лише*, *обличчя* и др., см., в частности, соответствующие позиции ЭССЯ).

Доля дифференциальной лексики в праславянском словнике каждого славянского языка может служить количественной мерой его лексической индивидуальности.

В таблице 6 приводятся статистические сведения о таких словах в каждом языке, извлеченные из обследованных материалов 1–15 выпусков ЭССЯ.

Таблица 6

(Пояснения: 1 — название идиома; 2 — абсолютное число индивидуальных праславянских слов в словаре данного идиома; 3 — доля индивидуальной лексики данного языка во всем массиве праславянской лексики, фиксируемой лишь в единичных языках (в процентах); 4 — доля индивидуальной лексики данного языка от количества такой индивидуальной лексики в данной языковой группе (южнославянской, западнославянской, восточнославянской; в процентах); 5 — доля индивидуальной лексики в праславянском лексическом слое данного языка (в процентах).)

1	2	3	4	5
ст.-слав.	28	3,7	12,8	2,49
болг.	35	4,6	16,0	1,07
макед.	0	0	0	0
с.-хорв.	111	14,7	50,7	2,43
словен.	45	6,0	20,5	1,28
всего в группе	219	29,0	100,0	
чешск.	125	16,6	49,0	2,93
словац.	17	2,3	6,7	0,58
в.-луж.	30	4,0	11,8	1,58
н.-луж.	13	1,7	5,1	0,83
полаб.	5	0,7	2,0	1,11
польск.	46	6,1	18,0	1,37
кашуб.-словин.	19	2,5	7,5	1,13
всего в группе	255	33,8	100,0	
др.-русск.	48	6,4	17,1	1,79
русск.	189	25,1	67,5	3,82
сев.-в.-рус.	52	6,9	18,6	1,25
ср.-рус. гов.	23	3,1	8,2	0,60
юж.-в.-рус.	19	2,5	6,8	0,50
укр.	31	4,1	11,1	0,79
белор.	12	1,6	4,3	0,36
всего в группе	280	37,1	100,0	
Всего	754	100,0		

Наибольшую лексическую специфичность по отношению к праславянскому лексическому фонду обнаруживает русский язык: почти четыре процента его праславянской лексики не встречается ни в одном другом славянском языке. При этом составляющие его идиомы (оба наречия и тем более полоса среднерусских говоров) по отношению к праславянскому наследию сами по себе характеризуются невысокой степенью своеобразия (как оно понимается в данном контексте) за счет тесной генетической близости друг к другу: индивидуализирующими лексически единицами для русского языка как целого принимаются и те, которые могли фиксироваться во всех трех выделяемых диалектных поясах, в то время как для наречий, взятых в качестве независимых идиомов, они (слова) индивидуализирующими уже не являются. Поэтому сумма дифференциальных праславянских лексем для наречий русского языка меньше абсолютного показателя лексической специфичности этого языка как целого.

Наименее индивидуализированным выглядит македонский язык. В интервале «*a* — *lokačь*» он не имеет выявленной «собственной» лексики, возводимой к праславянскому состоянию. Это, без сомнения, объясняется весьма поздним выделением македонского в самостоятельный язык (к тому же, как известно, самостоятельный его статус до сих пор оспаривается официальной болгаристикой).

Будучи упорядоченным по убыванию доли дифференциальной лексики в праславянском словнике каждого данного языка, список славянских идиомов имеет следующий вид:

1) русский, 2) чешский, 3) старославянский, 4) сербохорватский, 5) древнерусский, 6) верхнелужицкий, 7) польский, 8) словенский, 9) (северновеликорусское наречие), 10) кашубско-словинский, 11) полабский, 12) болгарский, 13) нижнелужицкий, 14) украинский, 15) (среднерусские говоры), 16) словацкий, 17) (южновеликорусское наречие), 18) белорусский, 19) македонский.

Пренебрегая некоторыми деталями, в выведенной последовательности можно усмотреть определенную тенденцию. К началу списка тяготеют идиомы, которые можно охарактеризовать такими чертами, как достаточно раннее диалектное обособление, раннее осознание себя носителями данного идиома как выделенной этнической общности, давняя историко-культурная традиция, включающая собственную письменность, и т. п. В противовес тому к его концу сосредоточиваются лингвистические образования более позднего с историко-культурных позиций характера: словацкий, ставший сложившимся литературным языком только в середине

XIX в., белорусский, окончательно оформившийся в национальный литературный язык к XX в., упомянутый уже в этом качестве македонский с его всего лишь полувековым «литературным стажем», великорусские наречия, не преодолевшие порога языковой самостоятельности и на роль более высокую, чем территориальный субязык, не претендующие. Иными словами, ранняя историко-культурная специализация идиома с необходимостью предполагает высокую долю индивидуальной лексики в сохраняемом им праязыковом лексическом наследии.

Взгляд на индивидуальность праславянского лексического наследия в идиомах более крупного масштаба — языковых группах — дает следующие показатели: в первых пятнадцати выпусках ЭССЯ исключительно в южнославянских языках отмечено 475 слов (или 6,3% всего праславянского лексикона в целом), исключительно в западнославянских языках — 483 слова (6,4%), исключительно в восточнославянских языках — 654 слова (8,7%). Долевое выражение числа слов, известных только какой-либо одной группе, к объему всего праславянского лексикона, у нас отличается от цифр, полученных путем обобщения первых семи выпусков ЭССЯ Ф. П. Филиным. Его данные на этот счет выражаются показателями 11,2%, 5,4% и 6,8% соответственно [Филин, 1984, стр. 33]. В точности своей цифри сомневался и сам Ф. П. Филин («в чем-то я мог ошибиться — работа эта весьма трудоемкая» [там же]), и сомнения эти были вполне небеспочвенны: в его арифметических вычислениях обнаруживаются прямые ошибки, не исправленные издателями его посмертной книжки об исторической лексикологии русского языка, — неверно высчитан процент праславянской лексики, встречающейся только в восточнославянских языках (6,8% нужно исправить на 7,2%). К тому же, как мы отмечали выше, Ф. П. Филин неточно установил объем обследованных им выпусков ЭССЯ; следовательно, все его проценты заведомо недостоверны.

Таким образом, индивидуальные части праславянских словарей южно- и западнославянской групп примерно одинаковы по объемам. Список индивидуальной праславянской лексики восточнославянской группы пространнее каждого из них приблизительно в 1,3–1,4 раза.

Глава 6

Картина родства славянских языков по данным лексикостатистики

1. Абсолютные показатели статистических связей

— Сядем на это бревно у дороги, — говорю я, — и забудем о бесчеловечности и развращенности поэтов. В длинных столбцах удостоверенных фактов и общепринятых мер и весов — вот где надо искать красоту.

О. Генри. «Справочник Гименя»

Выше мы уже неоднократно имели повод заметить, что абсолютное численное представление количества изоглосс, связывающих те или иные идиомы, не может служить мерой их генетической близости. Тем не менее прежде, чем привести данные индекса генетической близости, как он вычислен по предложенной формуле (11), сообщим данные о лексических связях праязыкового характера между славянскими языками в абсолютном выражении (сопроводив их, однако, сведениями о доле этого числа изолекс в общем объеме праславянского словника одного из языков, составляющих данную пару).

Пояснения к таблице 7: в первой колонке — названия идиомов (двубуквенные сокращения: «сс» — старославянский, «бг» — болгарский, «мк» — македонский, «сх» — сербохорватский, «сн» — словенский, «чш» — чешский, «сц» — словацкий, «вл» — верхнелужицкий, «нл» — нижнелужицкий, «лб» — полабский, «пл» — польский, «кс» — кашубско-словинский, «др» — древнерусский, «ср» — северновеликорусское наречие, «цр» — среднерусские говоры, «юр» — южновеликорусское наречие, «ук» — украинский, «бр» — белорусский), левая колонка цифр — количество всех лексических изоглосс (по первым 15-ти выпускам ЭССЯ), в которых участвуют два данных idioma, правая колонка цифр — процент изолекс с участием двух данных идиомов по отношению к объему праславянского словника первого из них (А).

Таблица 7

A	& B	V(A,B)	% от H(A)	A	& B	V(A,B)	% от H(A)
ст.-слав.	& бг	841	74,8	болг.	& сс	841	25,8
	мк	648	57,7		мк	1846	56,6
	сх	992	88,3		сх	2752	84,4
	сн	856	76,2		сн	2204	67,6
	чш	854	76,0		чш	2258	69,2
	сц	683	60,8		сц	1738	53,3
	вл	540	48,0		вл	1193	36,6
	нл	495	44,0		нл	1004	30,8
	лб	220	19,6		лб	354	10,9
	пл	750	66,7		пл	1846	56,6
	кc	473	42,1		кc	1038	31,8
	др	890	79,2		др	1586	48,6
	вр	914	81,3		вр	2435	74,6
	(ср	865	77,0)		(ср	2168	66,5)
	(цр	818	72,8)		(цр	2043	62,6)
	(юр	816	72,6)		(юр	2039	62,5)
	ук	801	71,3		ук	2118	64,9
	бр	668	59,4		бр	1775	54,5
макед.	& сс	648	31,8	с.-хорв.	& сс	992	21,7
	бг	1846	90,7		бг	2752	60,2
	сх	1890	92,9		мк	1890	41,4
	сн	1570	77,1		сн	2964	64,9
	чш	1516	74,5		чш	2974	65,1
	сц	1242	61,0		сц	2194	48,0
	вл	877	43,1		вл	1470	32,2
	нл	874	42,9		нл	1236	27,1
	лб	298	14,6		лб	398	8,7
	пл	1257	61,8		пл	2384	52,5
	кc	779	38,3		кc	1295	28,3
	др	1127	55,4		др	2003	43,8
	вр	1604	78,8		вр	3229	70,7
	(ср	1456	71,5)		(ср	2812	61,6)
	(цр	1383	68,0)		(цр	2634	57,7)
	(юр	1375	67,6)		(юр	2628	57,5)
	ук	1433	70,4		ук	2699	59,1
	бр	1202	59,1		бр	2247	49,2

Продолжение таблицы 7

A	& B	V(A,B)	% от H(A)	A	& B	V(A,B)	% от H(A)
словен.	сс	856	24,3	чешск.	сс	854	20,0
	бг	2202	62,6		бг	2258	53,0
	мк	1570	44,6		мк	1516	35,6
	сх	2964	84,2		сх	2974	69,7
	чш	2519	71,6		сн	2519	59,1
	сц	1917	54,5		сц	2587	60,7
	вл	1341	38,1		вл	1615	37,9
	нл	1308	37,2		нл	1356	31,8
	лб	382	10,9		лб	406	9,5
	пл	1985	56,4		пл	2586	60,6
	кс	1125	32,0		кс	1374	32,2
	др	1659	47,1		др	1826	42,8
	вр	2574	73,1		вр	3010	70,6
	(ср	2293	65,2)		(ср	2618	61,4)
	(цр	2144	60,9)		(цр	2468	57,9)
	(юр	2137	60,7)		(юр	2515	59,0)
ук	2265	64,4	ук	2703	63,4		
бр	1875	53,3	бр	2264	52,1		
словац.	сс	683	23,3	в.-луж.	сс	540	28,5
	бг	1738	59,3		бг	1193	63,0
	мк	1242	42,3		мк	877	46,3
	сх	2194	74,8		сх	1470	77,6
	сн	1917	65,4		сн	1341	70,8
	чш	2587	88,2		чш	1615	85,2
	вл	1330	45,3		сц	1330	70,2
	нл	1162	39,6		нл	1168	61,6
	лб	362	12,3		лб	333	17,6
	пл	2034	69,3		пл	1408	74,3
	кс	1181	40,3		кс	935	49,3
	др	1405	47,9		др	1003	52,9
	вр	2243	76,5		вр	1493	78,8
	(ср	1979	67,5)		(ср	1360	71,8)
	(цр	1889	64,4)		(цр	1282	67,7)
	(юр	1929	65,8)		(юр	1329	70,1)
ук	2142	73,0	ук	1420	74,9		
бр	1837	62,3	бр	1254	66,2		

Продолжение таблицы 7

А	Б	V(A,B)	% от H(A)	А	Б	V(A,B)	% от H(A)
н.-луж. Б	сс	495	31,4	полаб. Б	сс	220	48,7
	бг	1004	63,8		бг	354	78,3
	мк	874	55,5		мк	298	65,9
	сх	1236	78,5		сх	398	88,1
	сн	1308	83,1		сн	382	84,5
	чш	1356	86,1		чш	406	89,8
	сц	1162	73,8		сц	362	80,1
	вл	1168	74,2		вл	333	73,7
	лб	325	20,6		нл	325	71,9
	пл	1247	79,2		пл	391	86,5
	кс	876	55,7		кс	326	72,1
	др	915	58,1		др	333	73,7
	вр	1274	80,9		вр	404	89,4
	(ср	1176	74,7)		(ср	382	84,5)
	(цр	1123	71,3)		(цр	375	83,0)
(юр	1158	73,6)	(юр	381	84,3)		
ук	1223	77,7	ук	381	84,3		
бр	1094	69,5	бр	346	76,5		
польск. Б	сс	750	22,4	кашуб. Б	сс	473	28,1
	бг	1846	55,1		бг	1038	61,7
	мк	1257	37,5		мк	779	46,3
	сх	2384	71,2		сх	1295	76,9
	сн	1985	59,3		сн	1125	66,8
	чш	2586	77,2		чш	1374	81,6
	сц	2034	60,7		сц	1181	70,2
	вл	1408	42,0		вл	935	55,6
	нл	1247	37,2		нл	876	52,0
	лб	391	11,7		лб	326	19,4
	кс	1432	42,7		пл	1432	85,1
	др	1664	49,7		др	942	56,0
	вр	2605	77,8		вр	1357	80,6
	(ср	2285	68,2)		(ср	1221	72,5)
	(цр	2178	65,0)		(цр	1175	69,8)
(юр	2258	67,4)	(юр	1207	71,7)		
ук	2409	71,9	ук	1281	76,1		
бр	2097	62,9	бр	1164	69,2		

Продолжение таблицы 7

A	&	B	V(A,B)	% от H(A)	A	&	B	V(A,B)	% от H(A)
др.-рус.	&	сс	890	33,2	русск.	&	сс	914	18,4
		бг	1586	59,2			бг	2435	49,1
		мк	1127	42,0			мк	1604	32,4
		сх	2003	74,7			сх	3229	65,2
		сн	1659	61,9			сн	2574	51,9
		чш	1826	68,1			чш	3010	60,7
		сц	1405	52,4			сц	2243	45,3
		вл	1003	37,4			вл	1495	30,1
		нл	915	34,1			нл	1274	25,7
		лб	333	12,4			лб	404	8,2
		пл	1664	62,1			пл	2605	52,6
		кс	942	35,1			кс	1357	27,4
		вр	2294	85,6			др	2294	46,3
		(ср)	2117	79,0)			ук	3298	66,5
		(цр)	2007	74,9)			бр	2957	59,7
		(юр)	2006	74,8)					
		ук	1919	71,6					
		бр	1719	64,1					
(сев. -рус.	&	сс	865	20,8	(ср.-рус.	&	сс	818	21,3
		бг	2168	52,0			бг	2043	53,3
		мк	1456	34,9			мк	1383	36,1
		сх	2812	67,5			сх	2634	68,7
		сн	2293	55,0			сн	2144	55,9
		чш	2618	62,8			чш	2468	64,4
		сц	1979	47,5			сц	1889	49,3
		вл	1360	32,6			вл	1282	33,4
		нл	1176	28,2			нл	1123	29,3
		лб	382	9,2			лб	375	9,8
		пл	2285	54,8			пл	2178	56,8
		кс	1221	29,3			кс	1175	30,6
		др	2117	50,8			др	2007	52,3
		цр	3409	81,8			ср	3409	88,9
		юр	3264	78,3			юр	3260	85,0
		ук	3076	73,8			ук	2787	72,7
		бр	2601	62,4)			бр	2527	65,9)

Продолжение таблицы 7

A	и	B	V(A,B)	% от H(A)	A	и	B	V(A,B)	% от H(A)
(юж.	и	сс	816	21,3	укр.	и	сс	801	20,5
-рус.		бг	2039	53,2			бг	2118	54,2
		мк	1375	35,9			мк	1433	36,7
		сх	2628	68,6			сх	2699	69,1
		сн	2137	55,8			сн	2265	58,0
		чш	2515	65,6			чш	2703	69,2
		сц	1929	50,3			сц	2142	54,9
		вл	1329	34,7			вл	1420	36,4
		нл	1158	30,2			нл	1223	31,3
		лб	381	9,9			лб	381	9,8
		пл	2258	58,9			пл	2409	61,7
		кс	1207	31,5			кс	1281	32,8
		др	2006	52,3			др	1919	49,1
		ср	3264	85,2			вр	3298	84,5
		цр	3260	85,1			(ср	3076	78,8)
		ук	2923	76,3			(цр	2787	71,4)
		бр	2654	69,2)			(юр	2923	74,9)
							бр	2697	69,1

Продолжение таблицы 7

A	и	B	V(A,B)	% от H(A)	и	B	V(A,B)	% от H(A)
белор.	и	сс	668	20,3	и	лб	346	10,5
		бг	1775	54,0		пл	2097	63,8
		мк	1202	36,6		кс	1164	35,4
		сх	2247	68,3		др	1719	52,3
		сн	1875	57,0		вр	2957	89,9
		чш	2264	68,9		(ср	2601	79,1)
		сц	1827	55,6		(цр	2527	76,9)
		вл	1254	38,1		(юр	2654	80,7)
		нл	1094	33,3		ук	2697	82,0

Наименьшее число лексических изоглосс праславянского происхождения (220) связывает старославянский и полабский языки, которые в списке славянских идиомов, упорядоченном по убыванию объемов праславянского лексического наследия, занимают последние места. Наибольшее число праславянских изолекс связывают северновеликорусское наречие со среднерусскими говорами (3 409), а если представлять великорусский язык как цельное образование, — великорусский с украинским (3 298 общих слов праславянского происхождения, или 43,6% от числа всех праславянских лексем). По величине своих праславянских словников украинский и великорусский, как и его крупные диалектные составляющие, стоят в начале упомянутого списка.

Даже не слишком пристальный анализ абсолютных цифр, рисующих отношения в парах славянских идиомов в количественном аспекте (левая колонка таблицы 7), выявляет сильную зависимость этих показателей от объемов праславянских словников отдельных языков. Вычисление же коэффициента корреляции

$$r_{xy} = (\Sigma(x_i - \bar{x})(y_i - \bar{y})) / \sqrt{\Sigma(x_i - \bar{x})^2 \times \Sigma(y_i - \bar{y})^2} \quad (20)$$

между абсолютным числом изоглосс, связывающих данный славянский идиом с остальными членами славянской языковой семьи, (x), и величинами, характеризующими размеры поддающегося реконструкции праславянского лексического наследия остальных славянских идиомов, (y), например, для старославянского языка в качестве «данного» дает значения $r = +0,912$, для сербохорватского — $r = +0,963$, для нижнелужицкого — $r = +0,859$, для польского — $r = +0,972$ и т. д. Эти значения коэффициента корреляции чрезвычайно высоки (близки к единице), что полностью отклоняет возможность использования абсолютных численных показателей лексической близости между идиомами как основы для моделирования системы славянского языкового родства.

Не обнаруживает большой информативности и относительное, в процентах к величине праславянского лексического наследия данного языка, выражение его лексических связей с другими идиомами (правая колонка таблицы 7). Правда, некоторые наблюдения, касающиеся особых отношений между отдельными языками и группами языков, могут быть сделаны и на этой довольно скудной статистической базе.

Приведем пример такого рода наблюдений.

Сопоставляя правые колонки двух соседних фрагментов («периодов») таблицы, описывающих ближайшеразродственные идиомы A_1 и A_2 (в ка-

честве таковых можно выделить пары: болгарский & македонский, сербохорватский & словенский, чешский & словацкий, верхне- & нижнелужицкий, польский & кашубско-словинский, северно- & южновеликорусский, украинский & белорусский), несложно увидеть отчетливую и устойчивую закономерность: из данных (A_1 , A_2) двух ближайшеродственных языков с третьим языком (В) связан большим процентом «сохраненных» им праславянских лексем тот, у которого меньше объем праславянского лексического наследия H . Например, с сербохорватским болгарский язык объединяет 84,4% его, болгарского, праславянской лексики, тогда как македонский, праславянский словарь которого существенно меньше, чем у болгарского, связан с сербохорватским 92,9% праславянского слоя своего словаря. В северновеликорусских связях словацкого участвуют 67,5% его праславянской лексики, в то время как у ближайшеродственного ему чешского языка, праязыковое лексическое наследие которого обширнее, в параллелях с северновеликорусским языком участвуют только 61,4% его праславянских слов. В описанной ситуации нет ничего парадоксального, как может показаться на первый, не слишком внимательный, взгляд. Объяснение этой закономерности просто. Чем больше праславянский словарь какого-либо языка, тем больше в нем узкодialeктных праславянских лексических образований; напротив, чем он короче, тем большую его долю составляют слова широко распространенные, общеславянские, и тем выше доля слов, пересекающаяся с праславянской лексикой других языков.

Закономерность, согласно которой в отношениях с третьим языком участвует больший процент праславянского словаря того из двух данных ближайшеродственных языков, который характеризуется меньшим объемом словаря, выдерживается на протяжении всей таблицы весьма последовательно и имеет фактически единственное исключение. Оно касается украинского и белорусского языков в их отношениях к южнославянским (и чешскому). Список праславянской лексики украинского (3 905 слов в первых пятнадцати выпусках ЭССЯ) заметно превосходит числом своих элементов аналогичный список для белорусского (3 288 слов). При равновесном распределении сил связей ожидался бы более высокий процент лексики, участвующей в связях с языками южнославянской группы, как и любыми другими, у белорусского. Однако мы сталкиваемся с нарушением ожидаемого равновесия «в пользу» украинского. Это может свидетельствовать об особых исторических взаимоотношениях между протоукраинским компонентом праславянского языка и праславянскими диалектами, легшими в основу южнославянской ветви (и чешского языка); следует за-

метить, что эти «особые взаимоотношения» с южнославянскими мы здесь вправе констатировать для украинского только на фоне белорусско-южнославянских связей, как бы нам ни хотелось большего. (К такого же рода отступлениям от указанной закономерности как будто можно отнести и чуть-чуть большую связанность с южнославянскими языками среднерусских говоров по сравнению с южновеликорусским языком. Но, во-первых, мы не готовы признать заметную дифференцированность «протосреднерусского» и протоюжнорусского в рамках позднепраславянского языка, а во-вторых, реестры праславянских лексем в среднерусских говорах и южнорусском одинаковы по объему — 3 834 и 3 833 слова соответственно, — следовательно, крайне малые отклонения «в пользу» среднерусских говоров могут носить случайный характер).

2. Индекс генетической близости

— Продолжайте, мистер Пратт, продолжайте, — говорит миссис Сэмпсон, — ваши идеи так оригинальны и успокоительны. По-моему, нет ничего прелестнее статистики.

О. Генри. «Справочник Гименя»

Полагая вопрос об использовании абсолютных цифровых показателей лексических связей между родственными идиомами в лингвогенетических построениях окончательно закрытым, приведем результаты статистического обследования лексики славянских языков с целью установления их взаимной близости попарно, осуществленного по формуле (11).

Подсчеты, как уже говорилось (см. гл. 4, п. 6), мы делали в четырех версиях, различающихся между собою составом исследуемых идиомов. Мы, естественно, заинтересованы в том, чтобы результаты подсчетов в разных версиях были соизмеримыми. Поскольку конечное численное выражение индекса генетической близости G зависит, помимо прочего, от длины списка идиомов, подвергнутых статистическому сравнению, решено было представить численные результаты нормированными — отнесенными к средней в каждой версии величине G . (Средние значения G таковы: в версии SLAV, максимальной по количеству охватываемых анализом идиомов, — 1366,218; в версии RUS, не дифференцирующей великорусский язык, — 1281,387; в версии CON, из которой исключены старославянский и древнерусский языки, — 1187,769; нако-

нец, в самой короткой по составу идиомов версии MIN – 1092,259). В таком случае численные значения нормированных индексов генетической близости колеблются вокруг единицы, которой равна нормированная указанным образом величина G среднее. Возможное нормирование численных данных путем их отнесения не к среднему, а к максимальному (во всех версиях – между двумя серболужицкими языками) или к минимальному (в версиях RUS и MIN – между полабским и великорусским языками, в версиях SLAV и CON – между полабским и северновеликорусским наречием) значениям индекса генетической близости, несколько хуже, так как выделяет указанные пары идиомов из всего перечня попарных объединений и ставит их в исключительное положение, делая их стабильными (равными единице), не колеблющимися в зависимости от размеров списка идиомов.

Удобства ради результаты статистического анализа в разных версиях сведены в одну таблицу.

Таблица 8

(Пояснения: 1 – пары идиомов (A,B), 2–5 – индексы генетической близости G в разных версиях: 2 – $G(A,B)_{SLAV}$, 3 – $G(A,B)_{RUS}$, 4 – $G(A,B)_{CON}$, 5 – $G(A,B)_{MIN}$).

	1	2	3	4	5
старославянский	ѣ бг	1,0912	1,0913	—	—
	мк	1,1878	1,1918	—	—
	сх	1,0189	1,0141	—	—
	сн	1,0433	1,0347	—	—
	чш	0,8262	0,8223	—	—
	сц	0,8035	0,8082	—	—
	вл	0,8673	0,8746	—	—
	нл	0,9189	0,9277	—	—
	лб	1,0748	1,1091	—	—
	пл	0,8363	0,8396	—	—
	кс	0,7982	0,8078	—	—
	др	1,4999	1,5033	—	—
	вр	—	0,8041	—	—
	ер	0,8512	—	—	—
	цр	0,8438	—	—	—
	юр	0,8350	—	—	—
ук	0,7981	0,8109	—	—	
бр	0,7011	0,7209	—	—	

Продолжение таблицы 8

		1	2	3	4	5
болгарский	&	мк	1,7274	1,6940	1,7331	1,7113
		сх	1,2674	1,2434	1,2756	1,2605
		сн	1,2023	1,1819	1,2050	1,1930
		чш	0,9952	0,9794	0,9911	0,9819
		сц	0,9807	0,9702	0,9690	0,9653
		вл	0,8968	0,8883	0,8782	0,8753
		нл	0,8228	0,8191	0,8041	0,8062
		лб	0,7347	0,7447	0,7252	0,7433
		пл	0,9318	0,9260	0,9249	0,9266
		кс	0,8461	0,8410	0,8301	0,8310
		др	0,9707	0,9780	—	—
		вр	—	0,9452	—	0,9580
		ср	0,9344	—	0,9355	—
		цр	0,9308	—	0,9304	—
		юр	0,9204	—	0,9185	—
		ук	0,9672	0,9698	0,9623	0,9744
бр	0,8941	0,9073	0,8844	0,9077		
македонский	&	сх	1,2619	1,2385	1,2663	1,2516
		сн	1,2430	1,2222	1,2417	1,2294
		чш	0,9321	0,9197	0,9236	0,9174
		сц	0,9884	0,9800	0,9740	0,9725
		вл	0,9105	0,9088	0,8873	0,8922
		нл	0,8644	0,8647	0,8419	0,8486
		лб	0,8785	0,8892	0,8631	0,8829
		пл	0,8653	0,8628	0,8556	0,8603
		кс	0,8846	0,8775	0,8628	0,8613
		др	0,9682	0,9765	—	—
		вр	—	0,8703	—	0,8795
		ср	0,8759	—	0,8743	—
		цр	0,8791	—	0,8763	—
		юр	0,8602	—	0,8551	—
		ук	0,9103	0,9138	0,9021	0,9144
		бр	0,8342	0,8486	0,8223	0,8463
сербохорватский	&	сн	1,2963	1,2683	1,3042	1,2851
		чш	1,0355	1,0149	1,0355	1,0214
		сц	0,9765	0,9611	0,9674	0,9580
		вл	0,8788	0,8686	0,8642	0,8595
		нл	0,8089	0,8045	0,7944	0,7958
		лб	0,6643	0,6671	0,6569	0,6660
		пл	0,9547	0,9442	0,9522	0,9490
		кс	0,8605	0,8494	0,8489	0,8434
		др	0,9741	0,9775	—	—
		вр	—	0,9891	—	1,0058

Продолжение таблицы 8

	1	2	3	4	5	
	сх &	ср	0,9580	—	0,9625	—
		цр	0,9466	—	0,9491	—
		юр	0,9334	—	0,9345	—
		ук	0,9604	0,9596	0,9585	0,9668
		бр	0,8904	0,8994	0,8845	0,9032
словенский	&	чш	1,0770	1,0569	1,0737	1,0603
		сц	1,0581	1,0408	1,0470	1,0361
		вл	1,0064	0,9943	0,9884	0,9825
		нл	0,9334	0,9281	0,9167	0,9182
		лб	0,8081	0,8135	0,7971	0,8104
		пл	0,9546	0,9455	0,9481	0,9458
		кс	0,8865	0,8785	0,8711	0,8691
		др	0,9609	0,9631	—	—
		вр	—	0,9425	—	0,9531
		ср	0,9360	—	0,9345	—
		цр	0,9225	—	0,9194	—
		юр	0,9071	—	0,9020	—
		ук	0,9873	0,9861	0,9821	0,9898
		бр	0,8949	0,9066	0,8847	0,9062
чешский	&	сц	1,3622	1,3350	1,3525	1,3331
		вл	1,1001	1,0832	1,0798	1,0689
		нл	1,0291	1,0166	1,0103	1,0041
		лб	0,7375	0,7406	0,7291	0,7391
		пл	1,1629	1,1461	1,1573	1,1485
		кс	0,9931	0,9829	0,9756	0,9718
		др	0,8973	0,9004	—	—
		вр	—	0,9449	—	0,9517
		ср	0,9087	—	0,9042	—
		цр	0,9091	—	0,9022	—
		юр	0,9285	—	0,9204	—
		ук	1,0273	1,0254	1,0190	1,0262
		бр	0,9595	0,9689	0,9481	0,9676
словацкий	&	вл	1,1847	1,1684	1,1552	1,1450
		нл	1,1693	1,1571	1,1416	1,1363
		лб	0,8282	0,8388	0,8122	0,8315
		пл	1,1994	1,1840	1,1834	1,1758
		кс	1,1493	1,1386	1,1226	1,1190
		др	0,8551	0,8643	—	—
		вр	—	0,9147	—	0,9136
		ср	0,8875	—	0,8760	—
		цр	0,9016	—	0,8877	—

Продолжение таблицы 8

		1	2	3	4	5
	сц &	юр	0,9278	—	0,9129	—
		ук	1,0841	1,0812	1,0679	1,0736
		бр	1,0290	1,0385	1,0111	1,0308
верхнелужицкий	&	нл	1,9349	1,9061	1,8965	1,8782
		лб	1,1545	1,1662	1,1274	1,1501
		пл	1,1232	1,1123	1,0980	1,0939
		кс	1,2744	1,2629	1,2362	1,2318
		др	0,8287	0,8387	—	—
		вр	—	0,8162	—	0,8085
		ср	0,8205	—	0,8029	—
		цр	0,8109	—	0,7906	—
		юр	0,8556	—	0,8358	—
		ук	0,9459	0,9477	0,9233	0,9326
		бр	0,9401	0,9504	0,9145	0,9335
нижнелужицкий	&	лб	1,3477	1,3591	1,3207	1,3448
		пл	1,1383	1,1287	1,1142	1,1119
		кс	1,3829	1,3752	1,3415	1,3423
		др	0,8681	0,8818	—	—
		вр	—	0,7853	—	0,7780
		ср	0,8021	—	0,7845	—
		цр	0,8043	—	0,7850	—
		юр	0,8428	—	0,8243	—
		ук	0,9090	0,9161	0,8878	0,9022
		бр	0,9146	0,9315	0,8895	0,9160
полабский	&	пл	0,8972	0,8999	0,8857	0,8967
		кс	1,2961	1,3131	1,2639	1,2938
		др	0,8209	0,8417	—	—
		вр	—	0,6354	—	0,6359
		ср	0,6593	—	0,6516	—
		цр	0,6893	—	0,6811	—
		юр	0,6975	—	0,6900	—
		ук	0,6909	0,7039	0,6807	0,7018
		бр	0,6900	0,7131	0,6758	0,7084
польский	&	кс	1,3826	1,3648	1,3625	1,3533
		др	0,9968	1,0078	—	—
		вр	—	0,9944	—	1,0019
		ср	0,9608	—	0,9567	—
		цр	0,9825	—	0,9751	—
		юр	1,0262	—	1,0188	—
		ук	1,1219	1,1251	1,1134	1,1274
		бр	1,1137	1,2275	1,1017	1,1278

Продолжение таблицы 8

		1	2	3	4	5
кашубско-словинский	&	др	0,8732	0,8874	—	—
		вр	—	0,8405	—	0,8355
		ср	0,8245	—	0,8104	—
		цр	0,8475	—	0,8309	—
		юр	0,8741	—	0,8501	—
		ук	0,9553	0,9575	0,9373	0,9476
		бр	0,9907	1,0021	0,9684	0,9895
древнерусский	&	вр	—	1,1260	—	—
		ср	1,1661	—	—	—
		цр	1,1715	—	—	—
		юр	1,1582	—	—	—
		ук	1,0553	1,0783	—	—
		бр	1,0929	1,1244	—	—
великорусский	&	ук	—	1,2011	—	1,2180
		бр	—	1,2697	—	1,2859
северновеликорусский	&	цр	1,5265	—	1,5386	—
		юр	1,4265	—	1,5386	—
		ук	1,1639	—	1,4301	—
		бр	1,2254	—	1,2241	—
среднерусские говоры	&	юр	1,5448	—	1,5540	—
		ук	1,2055	—	1,2052	—
		бр	1,2839	—	1,2811	—
южновеликорусский	&	ук	1,2691	—	1,2679	—
		бр	1,3744	—	1,3711	—
украинский	&	бр	1,3745	1,3981	1,3695	1,4101

Итак, наибольшей теснотой генетических связей характеризуются два серболужицких языка, наименьшей — полабский в отношении с великорусским и, в других версиях, с его северным наречием. Чрезвычайно высоким значением индекса генетической близости отличаются болгарский и македонский языки, далее с ощутимым отрывом от двух наиболее тесных пар идут пары идиомов «среднерусские говоры & южновеликорусское наречие», «северновеликорусское наречие & среднерусские

говоры» и «старославянский & древнерусский». Следующие пары – «северно- & южновеликорусское наречия», «украинский & белорусский», «южновеликорусское наречие & белорусский», «польский & кашубско-словинский», «нижнелужицкий & кашубско-словинский», «чешский & словацкий», «нижнелужицкий & полабский», «полабский & кашубско-словинский», «сербохорватский & словенский»...

Длинные списки цифр не наглядны и трудны для восприятия. Представим результаты схематически – в виде графов, узлы которых символизируют исследуемые идиомы, а сила связей между последними изображается мощностью ребер.

См. схемы 1–4 (стр. 135–138).

3. Обсуждение результатов

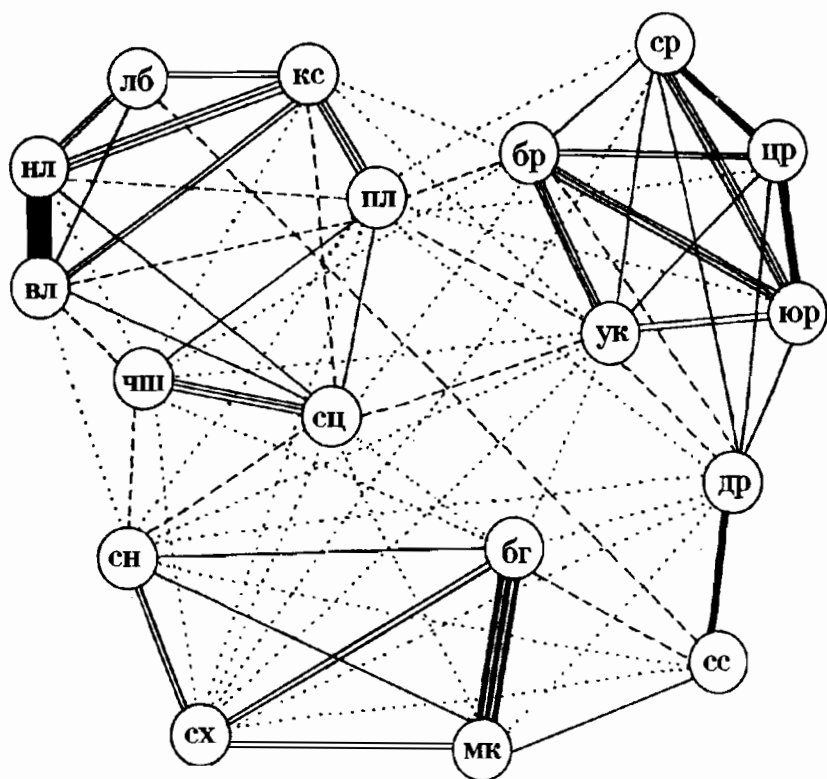
Сравнение графов, отражающих отношения между славянскими языками в разных версиях их состава, показывает, что колебания в силе генетических связей в отдельных парах языков при небольших изменениях состава исследуемых идиомов происходят в малозначительных пределах. В целом соотношения между силами связей в разных языковых парах и их группировках сохраняются. Поэтому, описывая лексикостатистические связи отдельных славянских языков, мы не будем специально указывать, в какой именно версии они проявляются.

а. Южнославянские языки

Старославянский язык, по данным лексикостатистики, из южнославянских языков теснее всего связан с македонским языком, несколько слабее его связи с болгарским. В более сильном тяготении старославянского к македонскому, по-видимому, отражается его сложение главным образом на базе солунского диалекта. Связи старославянского с остальными языками нынешней южнославянской группы также выше средней силы попарных связей в славянской семье в целом.

Значительно выше, чем с языками южнославянской группы, индекс лексической близости старославянского и древнерусского. Это может объясняться двояким образом.

Во-первых, лексика книжного древнерусского (естественно, отраженного памятниками полнее обиходного народного языка) находится в большой зависимости от лексического состава литературного церковно-



$G (\geq)$

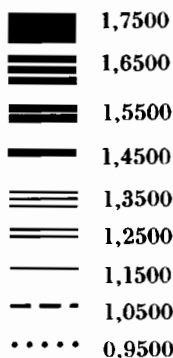
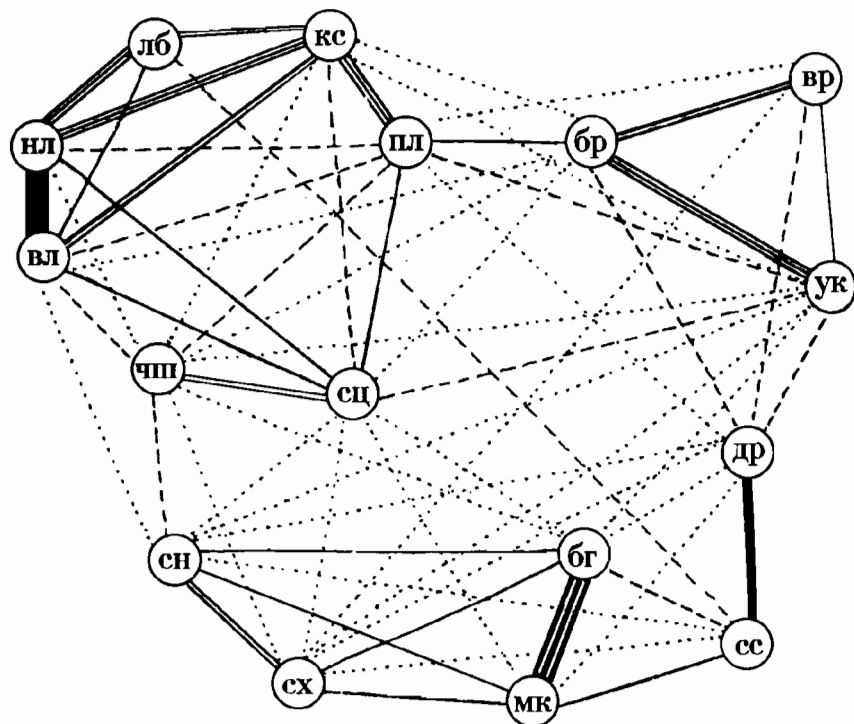


Схема 1

Сила генетических связей (G)
 между идиомами славянской семьи
 по данным лексикостатистики
 (версия SLAV)



$G (\geq)$

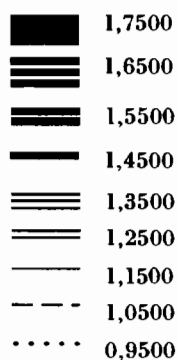
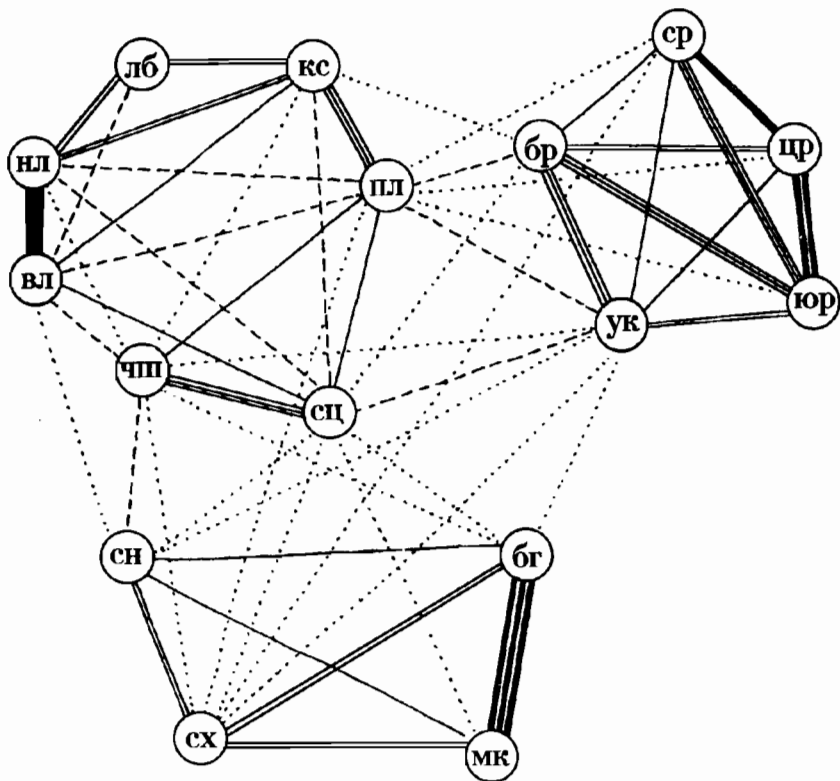


Схема 2

Сила генетических связей (G)
 между идиомами славянской семьи
 по данным лексикостатистики
 (версия RUS)



$G (\geq)$

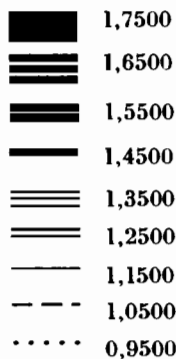
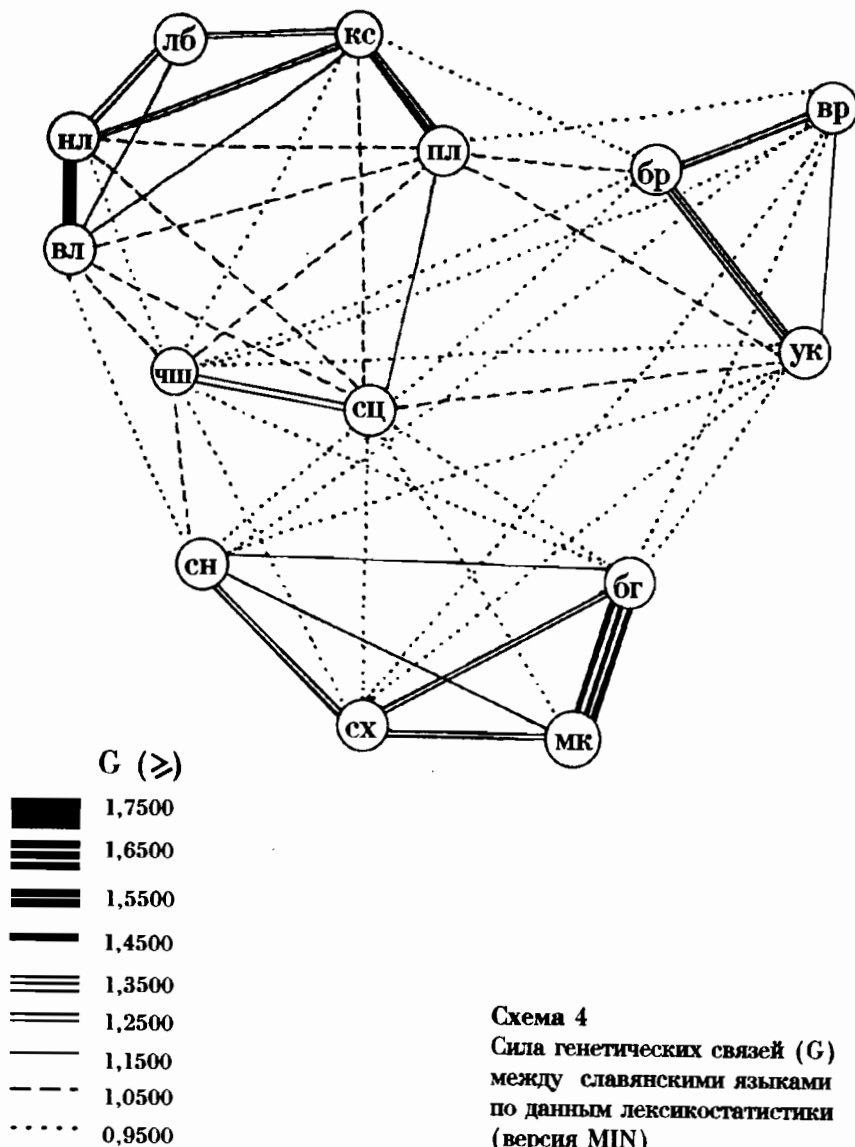


Схема 3
Сила генетических связей (G)
между идиомами славянской семьи
по данным лексикостатистики
(версия CON)



славянского языка, которым непосредственно продолжается старославянский. Однако, поскольку и сами составители ЭССЯ, и мы процедурой сегрегационного анализа (см. главу 4) стремились уменьшить потенциальное послепраславянское влияние одного идиома на другой, можно думать, что, хотя указанная причина и не была полностью устранена, все же не она является определяющей в данном случае.

И старославянский, и древнерусский – во-вторых – относятся к языкам со сравнительно небольшими объемами праславянского лексического наследия в каждом из них. Поэтому в их праславянских словниках выше, чем у языков с высокими значениями *H*, доля общеславянской лексики и, напротив, меньше доля лексики индивидуальной и узко-региональной. Для сравнения приведем следующие цифры. Слова, распространенные не менее, чем в четырнадцать (рубеж выбран условно: любой другой дал бы принципиально сходную картину) из восемнадцати идиомов, входящих в список славянских языков в версии SLAV, занимают 45,6% праславянского словника старославянского языка и 28,9% праславянского словника древнерусского языка, в то время как в болгарском, например, они составляют 25,2%, в сербохорватском – 19,1%, в северновеликорусском наречии – 20,8%. Редкие слова дают обратную картину. Лексемы, известные не более чем пяти идиомам из восемнадцати, в объеме праславянского словника старославянского языка составляют 13,9%, древнерусского – 17,5%, тогда как в тех же болгарском, сербохорватском и северновеликорусском они составляют соответственно 22,2%, 30,0% и 22,9%. По причине меньшей индивидуализированности праславянского наследия в старославянском и древнерусском эти языки и демонстрируют высокую лексическую близость, выразившуюся в наибольшем для них обоих значении *G*.

Можно полагать, что эффект большей взаимной близости древних состояний языка (древних родственных языков) по сравнению с мощностью их связей с более поздними языками, в том числе и с языками – прямыми потомками, является особенностью любых лексикостатистических процедур. В качестве аналога сошлемся на некоторые результаты упоминавшейся квантитативной таксономии европейских языков А. Я. Шайкевича. По его данным, древнеанглийский язык обнаруживает большую лексикостатистическую близость к древневерхненемецкому, средневерхненемецкому и даже готскому, чем к современному английскому, см.: [Шайкевич, 1980, стр. 328–329, схема – стр. 325].

Вероятно, той же природы относительно высокий – выше среднего – индекс лексикостатистического сходства старославянского с мертвым по-

лабским. Собственно, здесь существен не возраст или степень «витальности» последнего языка, а плохая сохранность его словаря. В более чем скромном по объему праславянском словнике полабского в долевом отношении преобладает общеславянская лексика: широко распространенная и редкая лексика (их условные пороговые характеристики — те же, что и для вышеприведенных цифр) составляют соответственно 68,8% и 7,7% всего праславянского списка. Относить сравнительно высокий показатель лексикостатистического сходства между старославянским и полабским на счет иных причин, помимо незначительной лексической индивидуализированности обоих языков, пока нет достаточных оснований. Вряд ли свидетельством какой-либо общности судеб полабского и болгарско-македонской подгруппы, к которой принадлежит и старославянский язык, являются аналогии в развитии φ в зависимости от предшествующего согласного и в отсутствии утраты редуцированных в предупредном слоге (см.: [Селищев, 1941, стр. 426]) или некоторые сближающие моменты изменения глагольных флексий в полабском и восточно-македонских говорах [там же, стр. 443], хотя, как это будет показано ниже, лексикостатистическое тяготение полабского к болгарско-македонской группе несколько заметнее, чем его чешско-словацкие, сербохорватско-словенские и восточнославянские ориентации.

Болгарский и македонский языки обнаруживают лексикостатистическое сходство на уровне праславянского наследия, принадлежащее к самым высоким парным связям. И тот и другой статистически ближе к сербохорватскому, чем к словенскому, однако связи со словенским у македонского выражаются числом несколько более высоким, чем у болгарского. Следующие по значимости для болгарского — связи с чешским, словацким, древнерусским и украинским, но все они несколько ниже среднего значения G для славянской семьи в целом. Тот же список — и для македонского, лишь с переменной мест чешского и словацкого: связи с последним для македонского более значимы (разница, впрочем, невелика). Наименьший индекс сходства праславянского словника болгарского языка — с праславянским лексическим фондом полабского. У македонского языка наименьшие связи отмечаются с белорусским языком. Если же говорить о внеюжнославянских лексикостатистических притяжениях этих двух ближайшерадственных языков в более общем плане, то у болгарского ярче выражены восточнославянские связи, македонский же, напротив, больше тяготеет к западнославянским.

Сербохорватский и словенский языки. Их праславянские лексические связи между собою приблизительно такие же, как

между праславянскими словниками великорусского и белорусского языков, чуть слабее — с болгарским и македонским. Из неюжнославянских языков наиболее близки они к чешскому, немного слабее связи со словацким. Как и следовало ожидать, лексикостатистические связи с чешским и словацким у словенского языка выражаются более высокими показателями, чем у сербохорватского. Заметны различия между сербохорватским и словенским в связях с лужицкими языками: у словенского они выше, причем с верхнелужицким в версии SLAV даже чуть выше средней попарных связей в славянской семье в целом. Весьма низким показателем определяются статистические отношения сербохорватского с полабским, полабские же лексические связи словенского языка выглядят более значительными, хотя и для словенского они самые низкие среди всех его парных связей с остальными языками. В целом словенский язык ближе к западнославянской группе, чем сербохорватский, что подтверждает оценки на этот счет у других исследователей (ср.: «...сербохорватский язык также сохраняет следы древних связей с западнославянской территорией, но не столь многочисленные, как словенский» [Штибер, 1972, стр. 309], хотя, как мы видели ранее и будет показано ниже, при численном представлении лексики, связывающей данный язык с идиомами других групп, но не находящей соответствий в идиомах той же славянской ветви, чисто количественно, в абсолютном выражении, параллели сербохорватского с языками западнославянской зоны преобладают над теми же связями словенского, за исключением связей с нижнелужицким языком). В отношениях с русским языком и его наречиями по отдельности картина обратная, хотя и не так ярко выражена. Что же касается украинского и белорусского языков, то здесь опять-таки «преимущество» имеет словенский перед сербохорватским. Вполне объяснимо, что и у того и у другого языка лексикостатистические связи на уровне праславянского словаря с украинским языком существенно выше, чем с белорусским.

Несколько слов о южнославянской группе отдельно. В обследованном участке ЭССЯ она имеет 5 591 слово, или 74,0% праславянского лексического корпуса. Л. В. Куркина, много и плодотворно занимающаяся изучением праславянских лексических связей южнославянских языков и в особенности словенского (библиографию ее работ на эти темы, хотя и неполную, см.: [Куркина, 1987]) считает, что общеславянскими или находящими соответствия по крайней мере в одном из языков западной и восточной групп являются примерно 70–80% древнейших южнославянских слов, реконструируемых для праславянского состояния, см.: [Курки-

на, 1973, стр. 17]. По нашим подсчетам, к таковым относится (в указанном диапазоне ЭССЯ) 5 121 слово, или 91,6% всей южнославянской лексики праславянского возраста. (Возможно, расхождения в оценках вызываются некоторыми различиями в отборе лексики: мы, исходя из позиций ЭССЯ, применяем более дробный отсчет, Л. В. Куркина принимает во внимание и корневые, а не только цельнолексемные, соответствия). При этом общими для обеих южнославянских подгрупп являются 3 240 слов (или 58,0%); 467 слов (8,4%) встречаются в болгарско-македонской подгруппе, но не отмечены в сербохорватско-словенской, 1 884 слова (33,7%) известны сербохорватско-словенской подгруппе, но отсутствуют в болгарском и македонском.

б. Западнославянские языки

Чешский и словацкий языки. Между собою эти два языка весьма близки. В отношениях с остальными славянскими языками они ведут себя по-разному. Южнославянские статистические связи чешского (за исключением македонских) несколько сильнее, чем у словацкого; то же можно сказать и об их связях с древнерусским и великорусским языком (как целым образованием, так и рассматриваемым «понаречно»). Что же касается других западославянских языков, украинского и белорусского, то к ним обнаруживает большую близость словацкий, иной раз отличаясь от чешского очень заметно. В кругу западославянских языков ближе всего к чешскому и словацкому является польский. Самые низкие показатели характеризуют их связи с полабским — не только из западославянских, но вообще из всех славянских идиомов. Среди южнославянских к ним ближе прочих словенский, среди восточнославянских — украинский. Если говорить о восточнославянской группе в целом, то связи с нею у словацкого языка несколько сильнее, чем у чешского.

Серболужицкие языки. Эта пара языков, что уже говорилось, характеризуется наибольшей теснотой лексикостатистических связей среди всех возможных пар славянских идиомов, выделяемых в качестве самостоятельных лингвистических образований. Естественно, что теснее всего они связаны с западославянскими языками, при этом по убыванию величины индекса близости к серболужицким языкам они располагаются в следующем порядке: кашубско-словинский, полабский, словацкий, польский, чешский. У верхнелужицкого выше, чем у нижнелужицкого, показатели связи с чешским и словацким языками, у ниж-

нелужицкого — с языками лехитской подгруппы (об этом, с опорой на данные иных языковых уровней, ср. еще у Э. Муки: [Мука, 1891, стр. 3]). Это обстоятельство можно толковать как явное отражение того положения, которое позволило Здзиславу Штиберу отнести к суждению Витольда Ташицкого о сильно «лехизированной» сущности нижнелужицкого (см.: [Ташицкий, 1928]) как к «излишне осторожному» [Штибер, 1934, стр. 90–91]. З. Штибер был склонен отождествлять нижнелужицкий с лехитскими языками. В тяготениях к южнославянской группе оба серболужицких языка отдают «предпочтение» словенскому, однако связи обоих языков с другим западным южнославянским языком, сербохорватским, — самые слабые из их южнославянских «пристрастий». Возможно, что это косвенно говорит о негомогенности сербохорватско-словенской подгруппы. Из восточнославянских языков наиболее сильные статистические связи у верхнелужицкого — с украинским (далее с белорусским), у нижнелужицкого — с белорусским (далее с украинским), из наречий русского языка «предпочтение» обоими языками отдается южновеликорусскому. В целом внезападнославянские связи верхнелужицкого — и в отношениях с южнославянскими, и в отношениях с восточнославянскими — выглядят более значительными, чем у нижнелужицкого. Нижнелужицкий как бы плотнее вписан в западнославянскую группу, чем его ближайший родственник.

Лехитские языки. Статистические связи на уровне праславянской лексики между польским и кашубско-словинским языком являются очень высокими, по своей силе они приблизительно уравниваются со связями между украинским и белорусским или чешским и словацким. Статистический анализ «Słownika języka pomorskiego czyli kaszubskiego» Стефана Рамулга показал, что лексико-словообразовательные отличия от польского языка задевают свыше 20% кашубского лексикона (см.: [Лещак, 1990, стр. 41]), что сильнее различий, которые констатируются нами. Весьма сильными представляются и связи между кашубско-словинским и полабским, будучи соизмеримыми, например, с лексикостатистической близостью сербохорватского и словенского. Однако отношения между польским и полабским описываются куда более скромным по величине индексом лексикостатистической близости, сравнимым с индексами таких пар, как болгарский или словенский с белорусским, нижнелужицкий с украинским, то есть характерными для парных объединений идиомов, принадлежащих разным современным славянским языковым группам. Это может наталкивать на мысль о генетическом неединстве подгруппы, определяемой термином «лехитские язы-

ки». Впрочем, такая мысль в какой-то степени охлаждается данными о лексикостатистическом поведении полабского с языками иных групп, например, со старославянским, о чем было сказано ранее.

Неодинаково складываются отношения отдельных лехитских языков с иными группами и подгруппами идиомов в славянском мире. В рамках западнославянской группы кашубско-словинский и полабский сильнее тяготеют к серболужицким языкам, а польский — к чешско-словацкой подгруппе. При этом, если сопоставляемые с лехитскими языками иные западнославянские пары рассматривать «покомпонентно», каждый лехитский язык ближе к нижнелужицкому, чем к верхнелужицкому, и к словацкому, чем к чешскому. (О схождениях между полабским и нижнелужицким, кроме общей лексики, могут говорить и факты пегативного порядка, например, совместное несохранение или префикса **pra-*, см.: [Борьсь, 1975, стр. 68].) Если цифры, описывающие статистические связи полабского с южнославянскими языками (исключая, пожалуй, сербохорватский), выше, чем показатели его же связей с восточнославянской группой (и те и другие, надо сказать, весьма низки), то польский и кашубско-словинский во внезападнославянских связях тяготеют более к восточным соседям, чем к южнославянским языкам. Рассматривая незападнославянские группы в их отношениях с лехитскими языками не целиком, а подробно, нетрудно обнаружить, что полабский сильнее связан с восточной (болгарско-македонской) подгруппой южнославянской ветви, в то время как польский и кашубско-словинский ближе к западной (сербохорватско-словенской) подгруппе, чем к болгарскому и македонскому языкам. С продвижением с севера на юг в пределах восточнославянского континуума, связи польского и кашубско-словинского с его составляющими усиливаются: слабее всего они с северновеликорусским наречием, сильнее всего — с украинским для польского и белорусским для кашубско-словинского. Сходную картину демонстрирует и полабский, но она выглядит очень стертой. Связи полабского языка с великорусским в целом, а из его наречий порознь — с северновеликорусским, — самые, как отмечалось, низкие во всей сети межславянских попарных связей.

в. Восточнославянские языки

При представлении русского языка целостным образованием восточнославянская группа рисуется наиболее компактной из трех славянских групп: наименьшая парная лексикостатистическая связь внутри нее,

русско-украинская (подтверждающая имеющиеся суждения на этот счет, ср.: «наиболее далеки друг от друга русский и украинский языки» [Кузнецов, 1966, стр. 44]), выражена более высоким индексом, чем наименьшие связи между компонентами южнославянской и западнославянской групп. Однако эта компактность расшатывается, если сосредоточить внимание на внутренних связях восточнославянской группы при представлении великорусских наречий отдельными идиомами: наименьшая по индексу лексикостатистическая близость между украинским языком и северновеликорусским наречием оказывается меньшей, чем самая слабая связь внутри южнославянской группы (болгарско-словенская; относительно слабыми внутриюжнославянскими связями старославянского языка мы здесь пренебрегаем ввиду его особого статуса), и таким образом южнославянская группа становится более плотным образованием, чем две другие славянские группы. Но нужно сказать, что при толковании отдельных диалектных комплексов, составляющих, например, сербохорватский язык (кайкавица, чакавица, штокавица, торлацкие говоры), как самостоятельных идиомов, чего осуществить в рамках настоящей работы мы были не в состоянии, могло бы случиться, что южнославянская языковая группа выглядела лексикостатистически гораздо более рыхлой, подобно тому как это и происходит с восточнославянской группой.

Древнерусский язык. К нему из современных восточнославянских языков ближе прочих русский, а из его наречий — северновеликорусское. Наиболее удален от древнерусского украинский. За пределами восточнославянской группы (потомков древнерусского) наибольшую близость к нему показывает книжный старославянский. Из живых не-восточнославянских языков самое большое лексикостатистическое сходство с древнерусским обнаруживает польский, но остальные западнославянские языки демонстрируют лексическую близость к древнерусскому существенно меньшую, чем языки, принадлежащие южнославянской ветви (связи последних с древнерусским численно очень сходны).

Русский язык. Северно- и южновеликорусские наречия на уровне праславянской лексики характеризуются статистической связью между собою менее мощной, чем связи двух серболужицких языков и болгарского с македонским. Выделив, возможно не с абсолютной корректностью, среднерусские говоры в отдельный идиом и численно описав его лексикостатистические связи с основными великорусскими наречиями, мы получим более крупные цифры (большую для связи среднерусских говоров с южнорусским наречием), но и они не достигают уровня лек-

сикостатистической близости двух самых тесных пар за пределами восточнославянской группы. Оба великорусских наречия лексически ближе к белорусскому языку, чем к украинскому, при этом, естественно, цифры, относящиеся к связям южнорусского, выше чисел, характеризующих внутривосточнославянские связи северновеликорусского.

За пределами восточнославянской группы с лексикой северновеликорусского наречия наибольшее статистическое сходство обнаруживает праславянский словарь сербохорватского языка, достаточно высоки и показатели связей последнего с другими восточнославянскими языками. Южновеликорусское наречие демонстрирует наибольшую близость своего праславянского лексикона за пределами восточнославянской группы с лексикой польского языка. Перепад значений индекса лексической близости для западнославянских языков с наречиями русского очень велик (от высоких у польского до крайне низких у полабского, далее — у серболужицких), причем он больше в индексах южновеликорусского наречия, чем северновеликорусского.

Русский язык как целое на уровне праславянской лексики ближе к польскому языку, однако его связи с сербохорватским слабее лишь немного, а если оценивать его тяготения к группам и подгруппам восточнославянских языков, то более «притягательными» нужно назвать южнославянскую и — внутри нее — западную (сербохорватско-словенскую) группировки.

Украинский и белорусский языки на уровне праславянской лексики ближе друг к другу, нежели между собой польский и кашубско-словинский, чешский и словацкий, а тем более сербохорватский и словенский. Внутривосточнославянские лексические связи описываются более высокими индексами G у белорусского, чем у украинского.

За пределами восточнославянской группы к ним наиболее близким оказывается польский язык, далее следуют: для украинского — словацкий, чешский, словенский и болгарский, для белорусского — словацкий, кашубско-словинский, чешский и верхнелужицкий; наименьшие вневосточнославянские их связи — с полабским. В целом южнославянские и западнославянские связи украинского приблизительно сбалансированы, у белорусского преобладает тяготение к западнославянским языкам.

У украинского выше, чем у белорусского, значения индекса G в отношениях к южнославянским, чешскому и словацкому, белорусский дает более высокие цифры для связей с лужицкими и лехитскими языками.

По отношению к русскому языку во вневосточнославянских связях белорусский ведет себя так же, как и по отношению к украинскому, то

есть дает более высокие цифры в связях с лехитскими и лужицкими языками. Иные соотношения обнаруживаются при сопоставлении индивидуальных связей русского и украинского: за исключением связей с сербохорватским все вневосточнославянские лексикостатистические связи украинского выражаются более крупными числами. Это можно назвать экстравертностью украинского языка — в противоположность лексически более «автохтонному» восточнославянскому облику белорусского.

4. Обобщенное представление межславянских лексикостатистических связей отдельных языков

Предыдущие цифровые таблицы описывали индивидуальные лексические связи между славянскими языками, относящиеся к праславянскому лексическому наследию. В комментариях к ним мы неоднократно прибегали к обобщению этих связей, ориентируя взгляд не на отдельные языки, а на их группировки. Как нам кажется, имеет смысл представить эти обобщенные сведения в более точном выражении — цифрами, усредняя численные показатели связей данного языка с отдельными языками других группировок. Такое представление статистических показателей позволит с большей наглядностью описать место каждого языка в славянской семье. Они сведены в таблицу 9.

Устройство таблицы таково. Для каждого языка сообщаются два рода цифр, помещенных в две колонки. Левая из них (а) включает средние величины, полученные сложением нормированных индексов близости данного языка к идиомам тех или иных групп и подгрупп и последующим отнесением к числу идиомов, составляющих эти группы и подгруппы (в качестве таковых определяются: южнославянская группа, включающая болгарско-македонскую и сербохорватско-словенскую подгруппы; западнославянская группа, составляемая чешско-словацкой, лужицкой и лехитской подгруппами; восточнославянская группа; последняя — при рассмотрении групповых тяготений великорусских наречий раздельно — условно разбивается на великорусскую и украинско-белорусскую подгруппы). Строка, соответствующая группе и подгруппе, к которой относится данный язык (его название стоит в заголовке данного фрагмента таблицы), включает усредненные данные, характеризующие отношения данного языка с остальными идиомами этой группы или подгруппы, то есть если подгруппа образуется двумя языками, то в левой

колонке попросту повторяется индекс их лексической близости из таблицы 8. Правая колонка (б) составляется числами, представляющими собою результат отнесения (путем деления) чисел левой колонки к средней величине лексикостатистических связей данного языка со всеми идиомами – составляющими славянской языковой семьи в целом (эта средняя, \bar{G} , сообщается при заголовке фрагмента – названии данного языка).

Поскольку усредняться должны только однородные величины, таблица 9 составлялась на основе данных версий CON и MIN, которые не включали статистической информации о старославянском и древнерусском языках, отличных по статусу от прочих рассматриваемых идиомов. Усреднение данных версии CON, представляющей русский язык в виде трех диалектных комплексов, из соображений экономии места проведено только для восточнославянских идиомов.

Таблица 9

Обобщенные лексикостатистические связи
отдельных славянских языков

	(а)	(б)	(а)	(б)
	болгарский (1,0103)		македонский (1,0052)	
южнослав.	1,3883	1,3741	1,3974	1,3902
болг.-макед.	1,7113	1,6939	1,7113	1,7024
сербохорв.-словен.	1,2267	1,2142	1,2405	1,2341
западнослав.	0,8757	0,8668	0,8907	0,8861
чеш.-словац.	0,9736	0,9637	0,9449	0,9400
лужицк.	0,8407	0,8321	0,8704	0,8659
лехитск.	0,8336	0,8251	0,8602	0,8558
восточнослав.	0,9467	0,9370	0,8801	0,8755
	сербохорватский (0,9820)		словенский (1,0138)	
южнослав.	1,2657	1,2889	1,2358	1,2190
болг.-макед.	1,2560	1,2790	1,2112	1,1947
сербохорв.-словен.	1,2851	1,3087	1,2851	1,2676
западнослав.	0,8704	0,8864	0,9461	0,9332
чеш.-словац.	0,9897	1,0078	1,0482	1,0339
лужицк.	0,8276	0,8428	0,9503	0,9374
лехитск.	0,8195	0,8345	0,8751	0,8632
восточнослав.	0,9586	0,9762	0,9497	0,9368

Продолжение таблицы 9

	(а)	(б)	(а)	(б)
	чешский (1,0148)		словацкий (1,0531)	
южнослав.	0,9952	0,9807	0,9830	0,9334
болг.-макед.	0,9496	0,9358	0,9689	0,9200
сербохорв.-словен.	1,0408	1,0256	0,9970	0,9467
западнослав.	1,0442	1,0290	1,1234	1,0668
чеш.-словац.	1,3331	1,3137	1,3331	1,2659
лужицк.	1,0365	1,0214	1,1406	1,0831
лехитск.	0,9531	0,9392	1,0421	0,9896
восточнослав.	0,9818	0,9675	1,0060	0,9553
	верхнелужицкий (1,0655)		нижнелужицкий (1,0602)	
южнослав.	0,9024	0,8469	0,8422	0,7944
болг.-макед.	0,8837	0,8294	0,8274	0,7804
сербохорв.-словен.	0,9210	0,8644	0,8570	0,8083
западнослав.	1,2613	1,1838	1,3029	1,2289
чеш.-словац.	1,1069	1,0389	1,0702	1,0094
лужицк.	1,8782	1,7627	1,8782	1,7716
лехитск.	1,1586	1,0874	1,2663	1,1944
восточнослав.	0,8915	0,8367	0,8654	0,8163
	полабский (0,8773)		польский (1,0553)	
южнослав.	0,7756	0,8841	0,9304	0,8722
болг.-макед.	0,8131	0,9268	0,8934	0,8466
сербохорв.-словен.	0,7382	0,8414	0,9474	0,8978
западнослав.	1,0427	1,1885	1,1300	1,0708
чеш.-словац.	0,7853	0,8951	1,1621	1,1012
лужицк.	1,2474	1,4219	1,1029	1,0451
лехитск.	1,0952	1,2485	1,1250	1,0660
восточнослав.	0,6820	0,7774	1,0857	1,0288
	кашубскословинский (1,0376)		русский (0,9404)	
южнослав.	0,8512	0,8204	0,9491	1,0093
болг.-макед.	0,8461	0,8154	0,9187	0,9769
сербохорв.-словен.	0,8562	0,8252	0,9794	1,0415
западнослав.	1,2187	1,1745	0,8464	0,9000
чеш.-словац.	1,0454	1,0075	0,9326	0,9917
лужицк.	1,2870	1,2404	0,7932	0,8435
лехитск.	1,3235	1,2755	0,8244	0,8766
восточнослав.	0,9242	0,8907	1,2519	1,3312

Продолжение таблицы 9

	(а)	(б)	(а)	(б)
	украинский (1,0142)		белорусский (0,9948)	
южнослав.	0,9613	0,9478	0,8908	0,8954
болг.-макед.	0,9444	0,9312	0,8770	0,8815
сербохорв.-словен.	0,9783	0,9646	0,9047	0,9094
западнослав.	0,9588	0,9454	0,9534	0,9583
чеш.-словац.	1,0499	1,0352	0,9992	1,0044
лужицк.	0,9174	0,9046	0,9247	0,9295
лехитск.	0,9256	0,9126	0,9419	0,9468
восточнослав.	1,3140	1,2956	1,3480	1,3550

Показатели статистических связей отдельных идиомов с целыми группами и подгруппами, полученные в версии перечня славянских идиомов CON, приведем, как было сказано, только для составляющих восточнославянской группы.

Таблица 9а

Обобщенные лексикостатистические связи
восточнославянских языков

	(а)	(б)	(а)	(б)
	севернорусский (0,9901)		среднерус. говоры (1,0071)	
южнослав.	0,9267	0,9360	0,9188	0,9123
болг.-макед.	0,9049	0,9139	0,9033	0,8969
сербохорв.-словен.	0,9485	0,9580	0,9342	0,9276
западнослав.	0,8266	0,8349	0,8361	0,8302
чеш.-словац.	0,8901	0,8990	0,8949	0,8886
лужицк.	0,7937	0,8016	0,7878	0,7822
лехитск.	0,8162	0,8244	0,8290	0,8232
восточнослав.	1,3394	1,3528	1,3947	1,3849
руск.	1,4833	1,4991	1,5463	1,5354
укр.-белор.	1,1945	1,2064	1,2431	1,2343

Продолжение таблицы 9а

	(а)	(б)	(а)	(б)
	южнорусский (1,0190)		украинский (1,0295)	
южнослав.	0,9025	0,8857	0,9512	0,9239
болг.-макед.	0,8868	0,8703	0,9322	0,9055
сербохорв.-словен.	0,9182	0,9011	0,9703	0,9425
западнослав.	0,8646	0,8485	0,9471	0,9200
чеш.-словац.	0,9166	0,8995	1,0434	1,0135
лужицк.	0,8300	0,8145	0,9055	0,8796
лехитск.	0,8530	0,8371	0,9105	0,8844
восточнослав.	1,4058	1,3796	1,2519	1,2160
русск.	1,4920	1,4642	1,2127	1,1780
укр.-белор.	1,3195	1,2949	1,3695	1,3303
	белорусский (1,0154)			
южнослав.	0,8690	0,8558		
болг.-макед.	0,8533	0,8404		
сербохорв.-словен.	0,8846	0,8712		
западнослав.	0,9299	0,9158		
чеш.-словац.	0,9796	0,9647		
лужицк.	0,9020	0,8883		
лехитск.	0,9153	0,9014		
восточнослав.	1,3114	1,2915		
русск.	1,2921	1,2725		
укр.-белор.	1,3695	1,3487		

Различие между левой (а) и правой (б) колонками таблицы 9 (9а) состоит в том, что первая рисует «привязанности» каждого идиома как бы с точки зрения системы родственных славянских идиомов в целом, а во второй точке отсчета служит сам данный язык. В колонке (а) отражается сила реальных статистических связей данного языка или наречия с различными группами славянских языков; данные, содержащиеся в колонке (б), позволяют сопоставлять пропорции между разными языками в распределении сил различных их связей, как бы уравнивая сами данные языки. Когда указанная при названии данного языка средняя величина его лексикостатистических связей со всеми остальными славянскими идиомами близка к единице (как, например, у македонского

или белорусского), разница между цифрами левой и правой колонок незначительна, когда же средняя заметно отклоняется от единицы (русский, лужицкие и особенно полабский языки), цифры разных левых колонок затруднительно сравнивать между собою, если нас интересуют «тяготения» языков, рассматриваемых «изнутри», как самодовлеющие сущности. Если в общем контексте межславянских связей (колонка (а)) отношения русского и украинского языка с сербохорватско-словенской подгруппой описываются весьма близкими числами (соответственно 0,9794 и 0,9783), то исходя из баланса статистических связей каждого данного языка (колонка (б)) мы можем констатировать, что для великорусского языка они значат больше, чем для украинского (и описываются иными цифрами: соответственно 1,0415 и 0,9646). Наоборот, по оценкам левой колонки, ориентирующей на среднюю силу связей во всей славянской семье, скажем, верхнелужицкий и полабский демонстрируют различную степень своей «привязанности» к западнославянской группе (1,2613 у верхнелужицкого и 1,0427 у полабского), в то время как оценка этих отношений «изнутри» самих этих языков показывает, что мера их «вписанности» в западнославянский языковой континуум приблизительно одинакова (1,1838 и 1,1885 соответственно).

Имея в виду именно оценки ориентаций «изнутри» самих языков, то есть данные колонки (б) таблицы 9, и выделив шесть упомянутых языковых подгрупп, составим еще одну табличку, которой будут отражены «предпочтения», обнаруживаемые разными языками, в ранжированном виде (см. таблицу 10).

При подобном сличении степеней близости между языками не может не броситься в глаза одно обстоятельство. Из общей равновесной картины, когда язык наибольшее тяготение обнаруживает к остальным идиомам собственной подгруппы, выпадают два лехитских языка. У полабского на первом месте оказываются не польский вместе с кашубско-словинским, а лужицкая подгруппа, а у польского — не кашубско-словинский в совокупности с полабским, а чешско-словацкая подгруппа. Это возникает по причине невысокой статистической связи между польским и полабским языками (при очень внушительном показателе статистической близости праславянских словников полабского и кашубско-словинского), из-за которой средняя величина попарных показателей лексической близости к языкам собственной подгруппы оказывается ниже, чем средняя степеней близости к языкам, составляющим другую подгруппу. У польского средняя величина его индекса лексического сходства с чешским и словацким языками, сильно уступая индексу связи с кашубско-словинским, оказывается

Таблица 10

	бг/мк	сх/сн	чш/сц	луж.	лех.	всл.
болг.	1	2	3	5	6	4
макед.	1	2	3	5	6	4
сербохорв.	2	1	3	5	6	4
словен.	2	1	3	4	6	5
чешск.	6	2	1	3	5	4
словац.	6	5	1	2	3	4
в.-луж.	6	4	3	1	2	5
н.-луж.	6	5	3	1	2	4
полаб.	3	5	4	1	2	6
польск.	6	5	1	3	2	4
кашуб.-словин.	6	5	3	2	1	4
русск.	4	2	3	6	5	1
сев.-в.-рус.	3	2	4	6	5	1
ср.-рус. гов.	3	2	4	6	5	1
юж.-в.-рус.	4	2	3	6	5	1
укр.	4	3	2	6	5	1
белор.	6	5	2	4	3	1

превосходящей средней его внутрилехитских связей в небольшой мере, у полабского же лужицкие «предпочтения» существенно выше внутрилехитских, хотя и уступают полабско-кашубско-словинской статистической связи. Описанная ситуация может найти объяснение в неомогенности лехитской подгруппы. Многочисленные совпадения в фонетической эволюции полабского и поморского при достаточно больших отличиях фонетического развития полабского от польских исторических процессов в сочетании с неравновесностью внешних лексикостатистических связей польского, кашубско-словинского и полабского, сильное его статистическое тяготение в сторону серболужицких языков (особенно нижнелужицкого) наталкивают на мысль о вторичности лехитской подгруппы, о том, что в данном случае отчетливо просматриваются результаты ее сложения путем конвергенции.

Из той же таблицы видно, что болгарско-македонские статистические связи для польского и кашубско-словинского находятся на последнем месте в их инославянских ориентациях, в то время как для полабского они уступают лишь лужицким, опережая чешско-словацкие, сербохорватско-словенские и восточнославянские. Похожая ситуация отмечается лишь в чешско-словацкой подгруппе, где сербохорватско-словенские ориентации в це-

лом для чешского гораздо важнее, чем для словацкого. Разнонаправленность внешних связей составляющих этой подгруппы, разница в лексикостатистических «предпочтениях» чешского и словацкого также может наталкивать на мысль о значительности моментов конвергенции в их истории, как между самими чешским и словацким, так и в их отношениях с другими славянскими языками (об этом см. ниже).

5. Эксклюзивные лексические связи

Значительный интерес для установления тесноты генетических связей между языками представляют их исключительные соответствия. Приводя далее доступную на нынешнее время статистику таких эксклюзивных лексических и словообразовательных связей, наблюдаемых в кругу славянских языков, следует, как кажется, оговорить некоторые в общем немаловажные моменты.

С исторической точки зрения разворачивающиеся во времени взаимоотношения между родственными языками укладываются, как известно, в схему двух противонаправленных движений — дивергенции и конвергенции. И то и другое может приводить в конце концов к существованию исключительных связей между двумя языками.

Эксклюзивные связи, являющиеся следствием дивергентных процессов, — это результаты языковой эрозии (архаизмы, реликты, «переживания» и т. д.), то есть «остатки» некогда более широкого ареала какого-либо лингвистического явления. Парадоксальность таких связей состоит в том, что языковое явление (фонетическая структура, закономерность фонетического преобразования, артикуляционный тип, словообразовательная конструкция, отдельная лексема, морфологический формант, синтаксическая схема и т. д.), исчезая в одних языках, способствует их отдалению от тех языков, где оно сохраняется. Но — в то же время — удерживаясь в небольшом числе языков, оно вызывает своим следствием увеличение их статистически оцениваемой близости друг к другу именно за счет редкости архаизма и, значит, большей его значимости в сложном сплетении разнообразных черт и характеристик конкретных языков, то есть приводит к эффекту, обратному по отношению к «стандартным» результатам дивергенции.

Исключительные связи, возникшие как последствия конвергентных процессов, — новообразования. Однако и инновации различаются по своей природе. Одни из них возникают в результате параллельного образо-

вания на разных территориях при сходных условиях (часто такие явления называют, не различая степени их конкретности, «типологическими» связями, что не совсем удачно, так как типология скорее должна иметь дело со сходствами и различиями структур, а не конкретных языковых единиц), другие — как результат волнового распространения из единого центра, вследствие контактного положения разных языков. Именно последние и признаются в большинстве посвященных этим явлениям работ наиболее доказательными при установлении тесного исторического взаимодействия языков (см., например: «Они (соответствия. — А. Ж.) только в том случае будут доказывать наличие тесных связей между языками, если окажутся в них общими новшествами» [Порциг, 1964, стр. 84]); иным же иногда даже отказывают в праве называться изоглоссами (ср.: «...я не спешу сказать „изоглоссы“, потому что для последних допускают нередко происхождение в порядке совместной инновации...» [Трубачев, 1963в, стр. 155]; подробнее см.: [Курилович, 1958, стр. 15–18; Макаев, 1964, стр. 25–33, 56–57; Макаев, 1971, стр. 9 и др.]).

Далеко не всегда эксклюзивные связи в лексике несут на себе приметы того пути, на каком они сложились. В изданной части ЭССЯ, например, зарегистрировано очень немного польско-«ближайшевозточнославянских» (= украинско-белорусских) лексико-словообразовательных изоглосс. Если можно с достаточной уверенностью говорить о сравнительно позднем и локальном характере таких соответствий, как праслав. **lahmanъ* (польск. *lachman* 'лохмотья, тряпка', укр. *láxmán* 'отрепье, лоскутья, рубище; оборвыш', диал. *лахмань* 'старая, рваная одежда', белор. диал. *láxmán*, множ. *лахмань* 'тряпье, лохмотья', см. ЭССЯ, вып. 14, стр. 17) или **linovišče* (польск. *linowisko*, укр. *линóвище*, белор. диал. *линóвишча*, *линóвишчэ* 'змеиный выползок', 'рачий панцирь, сброшенный при линьке', см. [ЭССЯ, вып. 15, стр. 111–112]), — о первом в силу исключительной редкости суффикса *-manъ*, «который связан отношениями чередования гласных с и.-е. суф. *-men-*», а также имея в виду колебания в акцентовке украинского и белорусского диалектных слов, которые могут указывать на их заимствованный характер, о втором — ссылаясь на сложный суффикс *-ov-išče*, к тому же вызвавший диспалатализацию корневого *-n'-* в глаголе **lin'ati* (более «ожиданными» были бы формы ***linišče* или ***lin'evišče*), то гораздо труднее судить о степени древности, а тем самым и о характере эксклюзивной связи («переживание» или совместное новообразование в границах позднепраславянского существования) в случаях с существительными **belnъ* или **lędъсь*. Праслав. **belnъ* (в ЭССЯ [вып. 1, стр. 187] регистрируется

только старопольск. *bleń* 'белена, *Huoscyanus niger* L.', мы добавляем сюда укр. диал. *бѣлень* 'белена' и белор. диал. *бѣлень* 'горечь', 'белена', см.: [ЭСУМ, т. 1, стр. 165; Туровский словарь, т. 1, стр. 51]; грамматический род польского и украинского слов неясен, белорусское — муж. р.) является словообразовательно-грамматическим вариантом к праслав. **belnъ*, известному несравнимо шире — во всех языках южнославянской группы, чешском, словацком, польском и русском языках в тех же и других ('порочное вожделение', 'греза, несбыточная мечта') значениях. Праслав. **lędъьь* (в ЭССЯ [вып. 15, стр. 51] включаются продолжения польск. *lędziewicz*, укр. *лядвѣц* 'горох, вика', видимо, сюда же должен быть присоединен белорусский топоним *Лядец*, с. Столинского р-на, см. [Жучкевич, 1974, стр. 217]) — производное с суффиксом *-ььь* от **lędo*, главным образом 'пустошь', тогда как в других языках (чешском, словацком, верхнелужицком, русском, украинском) известно производное с тем же суффиксом и в близких — ботанических — значениях от прилагательного **lędnъjъ* 'произрастающий на заброшенных полях' — **lędnъььь*. На то, что **belnъ* и **lędъььь*, скорее всего, не являются совместными польско-«ближайшевозточнославянскими» изоглоссами-новообразованиями, может указывать их чрезвычайно высокая формальная регулярность: морфолого-словообразовательное варьирование финалей *-ъ/-ь...* (далее также *...-a/-o/-y*), отражающее более раннюю систему именных основообразующих вокалических элементов *-û/-ĩ* (*-ā/-ō/-ū*), наблюдается в славянских именах как весьма свободное, почти не знающее морфонологических и семантических ограничений, а суффикс *-ььь*, если судить по материалам ЭССЯ, принадлежит к самым высокоупотребительным деривационным средствам в праславянском словообразовании. Так что у последних двух имен существительных более основательны причины считаться (особенно для **belnъ*) остатками прежних ареалов или (особенно для **lędъььь*) параллельными образованиями, чем совместными инновациями. Но подобные суждения в большинстве случаев вынуждены строиться если не на зыбких посылках, то на не достаточно точных для совершенной уверенности данных, и нередко могут быть опровергнуты или по меньшей мере оспорены.

Безоговорочное предпочтение совместных инноваций при доказательстве особо тесных генетических отношений между языками кажется нам все-таки излишне жестким подходом, отдающим ортодоксальностью. Эксклюзивные связи иной природы также могут в какой-то мере расцениваться как свидетельствующие о генетической близости языков. Параллельное образование, особенно с не самыми высокопродуктивными

деривационными элементами, может говорить о наличии в данных двух идиомах сходных предпосылок для возникновения именно этой лексико-словообразовательной конструкции. А сужение древнего широкого ареала до исключительного парного соответствия также может быть истолковано в пользу присутствия неких общих у данных языков условий для сохранения в них архаичной структуры, об общем для них в данном случае каком-то механизме иммунитета к утратам. Разумеется, силой особенной доказательности генетической тесноты эти факты не обладают, и основная аргументация лингвиста-историка должна сосредоточиваться не на них, но все же сбрасывать их как вовсе не значащие также нельзя.

Исходя из этих соображений, мы не взяли на себя смелость по собственному усмотрению сортировать материал эксклюзивных лексических и словообразовательных связей, представленный в ЭССЯ, на «архаичный», «совместноинновационный» и «параллельноинновационный», с тем чтобы посылно «усовершенствовать» статистику, полагая, во-первых, что сегрегация послепраславянских совместных образований уже осуществлена составителями этого лексикона (хотя и не всегда, согласно добросовестным оговоркам в тексте ЭССЯ, абсолютно надежно), а во-вторых, итоги собственно праславянских процессов, отраженные в нем, — наиболее объективная данность, с которой следует считаться: *hic Rhodus, hic salta*.

Численные данные об эксклюзивных лексических связях по первым пятнадцати выпускам ЭССЯ приведены ниже для каждого языка в отдельности (их имена набраны вразрядку). Цифра в скобках после имени идиома — количество исключительных парных связей у данного языка вообще; далее — их процент в общем объеме праславянского словника данного языка. За двоеточием — сведения о вторых языках: название и абсолютный численный показатель его эксклюзивных лексических связей с первым языком.

Старославянский (27; 2,40%):

др.-русск. — 9, сербохорв., словен., чешск. — по 4, болг., укр. — по 2, польск., белор. — по 1, остальные — 0.

Болгарский (113; 3,46%):

сербохорв. — 37, русск. — 17, макед. — 14, чешск. — 13, укр. — 7, словен. — 6, сев.-в.-рус. — 5, словац., в.-луж., польск., кашуб.-словин., юж.-в.-рус. — по 4, ст.-слав. — 2, н.-луж., др.-русск., белор. — по 1, полаб. — 0.

Македонский (22; 1,08%):

болг. — 14, сербохорв. — 4, чешск. — 2, др.-русск., русск. — по 1, остальные — 0.

Сербохорватский (268; 5,87%):

русск. — 74, словен. — 50, болг. — 37, чешск. — 34, сев.-в.-рус. — 25, польск. — 15, др.-русск., укр. — по 10, в.-луж., кашуб.-словин., юж.-в.-рус. — по 8, словац. — 7, белор. — 5, ст.-слав., макед. — по 4, н.-луж. — 2, полаб. — 0.

Словенский (135; 3,84%):

сербохорв. — 50, русск. — 32, чешск. — 19, сев.-в.-рус. — 11, укр. — 8, болг. — 6, ст.-слав., словац., в.-луж., польск., юж.-в.-рус. — по 4, др.-русск. — 2, н.-луж., белор. — по 1, остальные — 0.

Чешский (235; 5,51%):

словац. — 54, польск. — 35, сербохорв. — 34, русск. — 30, словен. — 19, др.-русск. — 15, болг. — 13, сев.-в.-рус. — 10, укр., белор. — по 9, н.-луж. — 7, ст.-слав., юж.-в.-рус. — по 4, макед., в.-луж., полаб. — по 2, кашуб.-словин. — 0.

Словацкий (95; 3,24%):

чешск. — 54, русск. — 12, сербохорв. — 7, болг., словен., польск. — по 4, н.-луж., кашуб.-словин., юж.-в.-рус., укр. — по 3, сев.-в.-рус. — 2, белор. — 1, остальные — 0.

Верхнелужицкий (46; 2,43%):

н.-луж. — 12, сербохорв. — 8, русск. — 7, болг., словен. — по 4, польск., сев.-в.-рус. — по 3, чешск., кашуб.-словин. — по 2, полаб., др.-русск., юж.-в.-рус., укр., белор. — по 1, остальные — 0.

Нижнелужицкий (34; 2,15%):

в.-луж. — 12, чешск. — 7, русск. — 5, словац., польск., сев.-в.-рус. — по 3, сербохорв. — 2, болг., словен., кашуб.-словин., др.-русск., укр., белор. — по 1, остальные — 0.

Полабский (5; 1,11%):

чешск. — 2, в.-луж., польск., русск. — по 1, остальные — 0.

Польский (107; 3,19%):

чешск. — 35, сербохорв., кашуб.-словин. — по 15, русск. — 11, укр. — 5, болг., словен., словац., юж.-в.-рус. — по 4, в.-луж., н.-луж., др.-русск., белор. — по 3, ст.-слав., полаб., сев.-в.-рус. — по 1, макед. — 0.

Кашубско-словинский (47; 2,79%):

польск. — 15, русск. — 9, сербохорв. — 8, болг., укр. — по 4, словац., сев.-в.-рус. — по 3, в.-луж. — 2, н.-луж., белор. — по 1, остальные — 0.

Древнерусский (115; 4,29%):

русск. — 61, сев.-в.-рус. — 17, чешск. — 15, сербохорв. — 10, ст.-слав. — 9, укр. — 6, белор. — 5, польск. — 3, словен. — 2, болг., макед., в.-луж., н.-луж., юж.-в.-рус. — по 1, остальные — 0.

Русский (375; 7,57%):

сербохорв. — 74, др.-русск., белор. — по 61, укр. — 54, словен. — 32, чешск. — 30, болг. — 17, словац. — 12, польск. — 11, кашуб.-словин. — 9, в.-луж. — 7, н.-луж. — 5, макед., полаб. — по 1, ст.-слав. — 0.

[Северновеликорусское наречие (110; 2,64%):

сербохорв. — 26, др.-русск. — 17, словен., юж.-в.-рус. — по 11, чешск., белор. — по 10, укр. — 8, болг. — 5, в.-луж., н.-луж., кашуб.-словин. — по 3, словац. — 2, польск. — 1, остальные — 0.

Южновеликорусское наречие (53; 1,38%):

сев.-в.-рус. — 11, белор. — 9, сербохорв. — 8, болг., словен., чешск., польск., укр. — по 4, словац. — 3, в.-луж., др.-русск. — по 1, остальные — 0.]

Украинский (125; 3,20%):

русск. — 54, белор. — 15, сербохорв. — 10, чешск. — 9, словен., сев.-в.-рус. — по 8, болг. — 7, др.-русск. — 6, польск. — 5, кашуб.-словин., юж.-в.-рус. — по 4, словац. — 3, ст.-слав. — 2, в.-луж., н.-луж. — по 1, остальные — 0.

Белорусский (105; 3,19%):

русск. — 61, укр. — 15, сев.-в.-рус. — 10, чешск., юж.-в.-рус. — по 9, сербохорв., др.-русск. — по 5, польск. — 3, ст.-слав., болг., словен., словац., в.-луж., н.-луж., кашуб.-словин. — по 1, остальные — 0.

Полученные данные, разумеется, нельзя использовать «буквально». Они со всей убедительностью свидетельствуют о том, что абсолютные численные показатели лексических связей между языками не могут служить индексацией их взаимной близости. Тем не менее и в таком виде приведенные сведения достаточно наглядны.

Укрупнение масштаба — обращение не к отдельным языкам, а к трем славянским группам — дает следующие цифры (излишне оговаривать, что они не являются результатом суммирования вышеприведенных данных: в связях описываемого рода участвует и не по одному языку с каждой стороны).

В 1—15 выпусках ЭССЯ между южнославянской и западнославянской группами отмечено 632 лексические связи (или 8,4% всего праславянского лексикона), в которых не участвуют восточнославянские язы-

ки; между южнославянской и восточнославянской группами, без участия западнославянских – 816 лексических связей (10,8%); западнославянские и восточнославянские языки связываются 809 праславянскими лексемами (10,7%), которых не знают языки южнославянской группы. Цифры аналогичного свойства, приводимые Ф. П. Филиным (соответственно 6,9%, 12,4% и 8,3% – см.: [Филин, 1984, стр. 33]), как мы показывали, недостоверны.

На первый взгляд, отношения южно- и западнославянской групп характеризуются более сильными лексическими связями праславянского происхождения, чем их отдельные отношения с восточнославянской группой. Однако сказанное выше об отдельных языках следует распространять и на их группировки. Мы знаем, что праславянское лексическое наследие в восточнославянском русском языке превосходит размеры праславянских словарей южно- и западнославянских языков, что отражается, в частности, в приведенных выше данных об объемах индивидуальной части праславянских словарей разных подгрупп. Мы знаем, с другой стороны, что число лексических связей, восходящих к праязыковому состоянию между любыми идиомами обнаруживает сильную зависимость от общих размеров праязыкового наследия в каждом из этих идиомов. Следовательно, более высокие абсолютные показатели индивидуальных связей восточнославянской группы с южно- и западнославянской группами сравнительно со связями между двумя последними нужно отнести на счет именно большего объема праславянского лексического списка для восточнославянских языков.

Поэтому количественные различия в парных связях между целыми группами славянских языков в общем можно считать маловыразительными.

6. Отдельные славянские языки вне ближайшеродственных связей

Для адекватных представлений о генетических и типологических взаимоотношениях между славянскими языками полезно располагать сведениями о статистике лексико-словообразовательных связей отдельных языков за пределами круга их ближайшеродственных идиомов – подгруппы и, далее, современной группы.

В обследованном нами объеме 1–15 выпусков ЭССЯ славянские языки характеризуются такими показателями связей, в которых не участву-

ют идиомы той же подгруппы (для каждого наречия русского языка ближайшеродственными условно определены другое наречие с «примыкающими» к нему среднерусскими говорами, для украинского — белорусский, а для белорусского — украинский, наряду с вариантами отнесения к ближайшеродственным обоим остальным восточнославянским языкам, хотя такое решение может быть оспорено; наличие слова в древнерусском языке в статистических сведениях относительно современных восточнославянских идиомов в расчет здесь не принималось):

старославянский — вне связей с болгарским и македонским — 266 слов (или 23,67% от объема праславянского словника старославянского языка),

- болгарский — без македонского — 1 416 (43,41%),
- македонский — без болгарского — 189 (9,29%),
- сербохорватский — без словенского — 1 604 (35,11%),
- словенский — без сербохорватского — 555 (15,77%),
- чешский — без словацкого — 1 677 (39,33%),
- словацкий — без чешского — 346 (11,80%),
- верхнелужицкий — без нижнелужицкого — 727 (38,36%),
- нижнелужицкий — без верхнелужицкого — 406 (25,79%),
- полабский — без остальных лехитских — 45 (9,96%), без остальных лехитских и лужицких — 27 (5,97%),
- польский — без кашубско-словинского — 1 918 (57,25%),
- кашубско-словинский — без польского — 251 (14,91%),
- великорусский — без украинского и белорусского — 864 (17,43%),
- северновеликорусское наречие — без южновеликорусского наречия и среднерусских говоров — 541 (12,98%),
- южновеликорусское наречие — без северновеликорусского наречия и среднерусских говоров — 350 (9,13%),
- украинский — без белорусского — 1 208 (30,93%), без белорусского и русского — 413 (10,58%),
- белорусский — без украинского — 591 (17,97%), без украинского и русского — 140 (4,26%).

Расширив круг ближайшеродственных языков до группы (то есть южно-, западно- и восточнославянской ветвей в целом), получим следующие данные:

Таблица 11

(Пояснения: 1 — язык, 2 — количество праславянских слов данного языка, не отмеченных в других языках той же группы (сюда же входит число индивидуальных слов, фиксируемых только в данном языке), 3 — то же, в процентах к объему праславянского словника данного языка, 4 — количество изолекс, связывающих данный язык с языками других групп, но в которых не принимают участия другие языки той же группы (то есть данные графы 2 за вычетом индивидуальной праславянской лексики), 5 — то же, в процентах к объему праславянского словника данного языка, 6 — количество лексико-словообразовательных связей данного языка за пределами своей группы — с другими группами по отдельности (соответственно — только с южными, только с западными и только с восточными славянскими языками).

1	2	3	4	5	6		
					южн	зап	вост
старославянский	83	7,38%	55	4,89%	—	10	22
болгарский	270	8,28	235	7,20	—	58	60
македонский	22	1,08	22	1,08	—	3	6
сербохорватский	795	17,40	684	14,97	—	141	197
словенский	382	10,86	337	9,58	—	75	75
чешский	775	18,18	650	15,24	169	—	126
словацкий	162	5,52	145	4,94	34	—	37
верхнелужицкий	98	5,17	68	3,59	22	—	13
нижнелужицкий	56	3,56	43	2,73	8	—	16
полабский	16	3,54	11	2,43	1	—	4
польский	398	11,88	352	10,51	43	—	116
кашуб.-словинский	106	6,30	87	5,17	24	—	26
древнерусский	256	9,55	208	7,76	65	37	—
русский	864	17,43	675	13,62	227	140	—
сев.-рус.нареч.	240	5,76	188	4,51	70	39	—
южнорус.нареч.	97	2,53	78	2,03	27	20	—
украинский	350	8,96	319	8,17	79	76	—
белорусский	117	3,56	105	3,19	22	33	—

Цифры, отражающие изоглоссы какого-либо славянского языка, в которых не принимают участия другие языки той же группы, обнаружи-

вают высокую скоррелированность с объемами праславянского лексического корпуса в отдельных языках. Этой зависимости не устраняет и долевое, процентное по отношению к объему праславянского словника каждого языка, выражение его лексических связей за пределами собственной подгруппы.

Интерес здесь вызывают данные, говорящие о «предпочтениях», которые демонстрируют разные языки в «ориентациях» за пределами своей группы, — сведения, содержащиеся в трех колонках последней графы таблицы 11.

Южнославянские языки.

Связи старославянского языка с восточнославянской группой, в которых не принимают участия остальные южнославянские языки, количественно явно значительнее, чем такого же свойства его связи с западнославянскими языками. Вызывается это, можно думать, прежде всего высокой долей лексических и словообразовательных совпадений с древнерусским языком. Но даже и за вычетом эксклюзивных лексических связей с древнерусским, которые можно объяснить непосредственным влиянием на последний со стороны старославянского, число параллелей с восточнославянскими оказывается больше числа связей с западнославянскими языками.

Заметен перевес восточнославянской «ориентации» во внешних связях сербохорватского языка. Внешние «ориентации» болгарского и словенского языков приблизительно уравновешены. О македонском судить затруднительно в силу статистической ненадежности имеющихся данных (слишком малой величины абсолютных показателей македонско-западнославянских и македонско-восточнославянских совпадений в праславянской лексике).

Западнославянские языки.

Чешский и верхнелужицкий явно склоняются в своих внешних связях к южнославянским языкам. Польский и нижнелужицкий, напротив, отдают «предпочтения» связям с восточнославянской группой. «Склонности» словацкого и кашубско-словинского выражены неярко, с некоторым перевесом словацкого- и кашубско-восточнославянских изоглосс. Данные, касающиеся полабского языка, статистически ненадежны.

Восточнославянские языки.

Древнерусский и современный русский язык в заметно большей мере ориентированы на южнославянскую группу. У украинского языка южнославянские параллели, в которых не принимают участия остальные восточнославянские языки, почти уравновешиваются западнославянскими. Бело-

русский же язык имеет очевидный обратный уклон – к сходству в праславянском лексиконе с языками западнославянской группы.

В целом картина межславянских лексических связей на уровне праславянского наследия, складывающаяся на основе статистики исключительных связей данного языка с языками других славянских групп, весьма близка той, которая рисуется таблицей 9, теми ее строками, в которых представлены величины связи отдельных идиомов с иными группами славянских языков.

Глава 7

О возможности лексикостатистического обнаружения явления языковой конвергенции

1. Конфронтационная статистика разнонаправленных изоглоссных связей ближайшеродственных языков как способ выявления феномена конвергенции

Выше, комментируя данные лексикостатистического обследования славянских языков по предложенной нами формуле (11), мы упоминали о возможном отражении в цифрах, описывающих генетическую близость между языками, результатов процесса вторичного сближения языков в ходе их эволюции.

Имеются ли статистические способы обнаружения конвергентных тенденций в истории языков? Этот вопрос имеет существенное значение, поскольку феномен конвергенции как мощный фактор возникновения материальной и структурной близости между языками отвлеченно признается, пожалуй, всеми языковедами-диахронистами, в том числе и применяющими статистические методы, однако в специальных лингвостатистических работах диахронической направленности почти не обсуждается. Более того, почти всегда цифры, полученные с помощью статистического анализа изоглосс, будь они фонетическими, морфологическими или лексическими (словообразовательными), безоговорочно объявляются данными, отражающими ситуацию, которая сложилась в результате дивергентных процессов, и тем самым являющимися прямым мерилем родства. Тенденция интерпретировать итоги статистического подсчета только как свидетельство последовательного распада праязыко-

вого единства наиболее отчетливо выступает в глоттохронологических штудиях: согласно глоттохронологической концепции, чем выше цифра, демонстрирующая сходство между языками, тем позже они разошлись друг от друга и тем теснее их изначальное родство. Сомнения в адекватности подобных суждений и возможные соображения о роли такого фактора языкового родства, как конвергенция, деликатно отводятся в сторону. Подоплека такой деликатности очевидна: сама постановка вопроса о конвергенции как одной из причин особой близости языков уязвляет самое существо лингвостатистических диахронических построений, базирующихся в конечном счете на концепте родословного древа. Признать важность фактора вторичного сближения в подобном случае значит признать слабость своего метода, точнее говоря, — признать, что результаты статистического исследования выявляют не прямо меру генетической близости, то есть изначального родства, а что-то несколько отличное от нее.

Между тем статистика изоглосс дает, как нам представляется, возможность обнаружения роли конвергентных процессов в итоговой картине статистической близости языков.

Примечательно, что намек на такую возможность содержится в работе, автора которой нельзя упрекнуть в открытой приверженности количественным подходам к решению проблем языковой истории, хотя он, можно заметить, и не чуждается некоторых элементарных подсчетов. Речь идет о статье О. Н. Трубачева о праславянских лексических диалектизмах лужицких языков [Трубачев, 1963в].

С основанием сомневаясь в прошлом существовании единого «промежуточного» праязыка, развившегося впоследствии в современную серболужицкую подгруппу славянских языков, О. Н. Трубачев пишет: «В высшей степени интересно складывается на нашем материале картина нижнелужицко-верхнелужицких отношений, а именно: на девять случаев, когда праславянский лексический диалектизм известен и нижнелужицкому и верхнелужицкому..., приходится двадцать один случай, когда верхнелужицкий не участвует в праславянском диалектизме и слово известно только в нижнелужицком... Мы склонны видеть в этом нечто большее, нежели простую случайность.

Иными словами, именно на древних элементах лужицкой лексики, которые связаны подчас надежными соответствиями в отдельных диалектах славянских и других индоевропейских, явственно рисуется столь глубокая граница, отделяющая нижнелужицкий от верхнелужицкого. Древние диалектные связи и рубежи в лексике, выявляемые этимологи-

чески, очевидно, вполне годятся как объективное средство при изучении истории языковых отношений, многократно перекрытых и сглаженных позднейшими напластованиями. На совершенно ином материале мы приходим к концепции Штибера о так называемом „праљужицком“ языке: „Język prałużycki nigdy nie istniał“» [Трубачев, 1963в, стр. 172, со ссылкой на: Штибер, 1934, стр. 93].

Понятие праязыкового (праславянского) лексического диалектизма у О. Н. Трубачева если и не эзотерично, то во всяком случае достаточно трудноформализуемо, по крайней мере для использования его в квантитативном анализе праславянской лексики современных славянских языков. На некоторых понятиях, применяемых О. Н. Трубачевым, лежит печать неокончательной проясненности, и возникает ощущение, что вполне свободно пользоваться этими понятиями может только сам автор. Впрочем, порицать О. Н. Трубачева за эксплуатацию интуиции, весьма, надо сказать, острой и изощренной, в наши намерения совсем не входит: роль интуитивных моментов в этимологии и компаративной лексикологии действительно исключительно велика. Воспользоваться намеком О. Н. Трубачева на возможность методики выявления конвергенции путем статистического анализа изоглосс можно, попытавшись упростить понятие праязыкового диалектизма с целью его формализации.

По Трубачеву, праславянский «лексический диалектизм, или диалектную лексему, можно понимать в общем достаточно широко с допущением на правах самостоятельных слов лексикализации словообразовательных и семантических актов... методику изучения праславянских лексических диалектизмов должна отличать важность определения и характеристики корневой морфемы, в частности, ее этимологических индоевропейских связей. Словообразовательное оформление, разумеется, здесь также важно, но преимущественный объект этой последней — скорее общие элементы праславянского словаря, реконструируемого в связи с этим с максимальной доступной словообразовательной конкретностью» [Трубачев, 1963в, стр. 157]. Из этого и цитированных выше объяснений понятно лишь, что праславянский лексический диалектизм имеет ограниченное распространение в языках-потомках, прочие же характеристики — относительная простота основы, наличие индоевропейских этимологических и словообразовательных соответствий, разорванный ареал и т. д. — хотя и крайне, так сказать, желательны как указания на древний возраст лексемы, но не жестко обязательны в качестве признаков праславянского диалектизма.

Гораздо лучше формализуется понятие праславянского диалектизма, как оно принято в конкурирующем с ЭССЯ краковском Праславянском словаре. Там диалектизмом признается праславянская лексема, если она не встречается по меньшей мере в языках одной современной ветви (южно-, западно- или восточнославянской). Ср.: «Z gotowych prawie 900 hasel na litery A B blisko 400 uznajemy za wyraźne dialektyzmy... Materiał ten wyodrębnia z jednej strony północ Słowiańszczyzny od południa, z drugiej zaś wskazuje na łączność języków południowo- i wschodniosłowiańskich, a także zachodnio- i południowosłowiańskich. Słabiej zaznacza się odrębność leksykalna grup zachodniej, południowej i wschodniej» [Праславянский словарь, т. 1, стр. 10]. Такое понимание диалектизма проще, но одновременно прямолинейнее и грубее. Содержание понятия праславянского лексического диалектизма у создателей краковского словаря, явно соотносясь с современным трихотомическим состоянием славянского языкового мира, оказывается тем самым сильно модернизированным и не учитывает более тонких возможностей. По логике краковской трактовки понятия диалектизма, например, исключительная словенско-словацко-украинская изоглосса, охватывающая языки всех трех современных славянских ветвей, в праславянские диалектизмы зачислена не будет, с чем согласиться, пожалуй, трудно.

Поэтому откажемся от понятия праславянского лексического диалектизма — во многом интуитивного у О. Н. Трубачева¹ и чересчур прямолинейного у авторов Праславянского словаря — и примем во внимание слова с более или менее ясными операционально определенными приметами. В качестве такой приметы мы изберем узкую распространенность праславянского слова в поздних славянских языках (одном, двух, трех и т. д. — до произвольно назначенного, но, разумеется, сравнительно невысокого предела) — с тем чтобы далее в праславянских лексиконах каждого из языков, составляющих ближайшеродственную пару, выделить слова двух категорий: отмеченные в них обоих — и отмеченные в каждом из них поодиночке. Именно формульное противопоставление

¹ Мы готовы принять утверждение, что в понятии праславянского лексического диалектизма нет ничего загадочного и вообще «ничего такого», что могло бы смутить, тем более что сам О. Н. Трубачев охотно цитирует Ф. П. Филина, подсчитавшего, что в праславянском языке более половины (!) всего словарного состава — регионализмы, то есть праславянские диалектные слова, хотя тут же делается оговорка, что «этимолог кладет в основу своего понимания древнего лексического диалектизма более фундаментальные отличия (разные этимологии — разные слова), чем оттенки значений и отклонения в словообразовании, и что в дальнейшем величины будут сбалансированы в сторону уменьшения...» [Трубачев, 1985, стр. 9].

объединяющих и разводящих данные языки узких, «маломощных» изоглосс может составить суть статистической методики, способной контрастно представить пары языков с изначальной и вторичной (усиленную конвергенцией) близостью.

2. Численные данные и итоги наблюдений

Выделим состав ближайшеродственных пар: болгарский & македонский, сербохорватский & словенский, чешский & словацкий, верхнелужицкий & нижнелужицкий, польский & кашубско-словинский, русский & украинский, русский & белорусский, украинский & белорусский. В пары с ближайшеродственными языками не включены старославянский и древнерусский (вопрос о мере их конвергентности с болгарским или македонским и, соответственно, с поздними восточнославянскими языками не имеет смысла), а также полабский (количественные данные о его участии в узких изоглоссах отличаются, по отмеченным ранее причинам, весьма невысокой статистической надежностью). Наряду с парами ближайшеродственных языков мы включили – для сопоставления – и две пары языков, заведомо не являющихся ближайшеродственными (болгарский & польский, чешский & украинский).

Приведем численные сведения о распределении праславянской лексики между составляющими отдельных пар. В первых двух колонках таблицы 12 помещены цифры, определяющие количество лексем, встречающихся в данных языках порознь (но, за исключением первой строки, известные и иным языкам): $V(A, \bar{B})$ и $V(\bar{A}, B)$; в третьей колонке сообщается число лексем, общих обоим языкам: $V(A, B)$. Строки таблицы соответствуют максимальной мощности рассматриваемых изоглосс (каждая следующая строка включает в себя и количество изоглосс меньшей мощности): первая строка ($m = 1$) содержит численные данные о словах, известных только какому-либо одному языку, вторая строка ($m = 2$) – сведения о словах, входящих в изоглоссы, охватывающие не более двух языков, третья ($m = 3$) – не более трех языков и т. д. Мы решили ограничиться предельной мощностью изоглоссы в семь языков (из шестнадцати), но для сравнения напомним и цифры, касающиеся всего объема праславянской лексики в данном языке (под $m = 16$). Подсчеты сделаны на материале первых пятнадцати выпусков ЭССЯ. Приводятся сведения, полученные путем подсчетов, осуществленных на базе файла RUS.

Таблица 12

m	$V(A, \bar{B})$	$V(\bar{A}, B)$	$V(A, B)$	m	$V(A, \bar{B})$	$V(\bar{A}, B)$	$V(A, B)$
а) А – болгарский, В – македонский				б) А – сербохорватский, В – словенский			
1	35	0	—	1	111	45	—
2	137	8	14	2	329	130	50
3	283	33	74	3	589	213	185
4	438	52	163	4	819	282	392
5	583	72	262	5	994	358	605
6	728	99	364	6	1146	418	825
7	883	116	494	7	1284	459	1078
...
16	1416	189	1846	16	1604	555	2964
в) А – чешский, В – словацкий				г) А – верхнелужицкий, В – нижнелужицкий			
1	125	17	—	1	30	13	—
2	306	58	54	2	64	39	12
3	525	90	144	3	98	65	31
4	740	138	253	4	148	95	51
5	962	188	391	5	208	131	75
6	1135	225	561	6	273	158	103
7	1305	261	757	7	347	194	159
...
16	1677	346	2587	16	727	406	1168
д) А – польский, В – кашубско-словинский				е) А – русский, В – украинский			
1	46	19	—	1	189	31	—
2	138	51	15	2	510	102	54
3	315	83	34	3	787	193	229
4	499	112	64	4	1011	295	457
5	699	152	110	5	1192	366	719
6	888	169	167	6	1331	432	961
7	1089	191	240	7	1436	495	1257
...
16	1918	251	1432	16	1658	607	3298

Продолжение таблицы 12

m	$V(A, \bar{B})$	$V(\bar{A}, B)$	$V(A, B)$	m	$V(A, \bar{B})$	$V(\bar{A}, B)$	$V(A, B)$
ж) А – русский, В – белорусский				з) А – украинский, В – белорусский			
1	189	12	—	1	31	12	—
2	503	57	61	2	141	102	15
3	788	100	231	3	299	212	126
4	1023	140	440	4	474	297	289
5	1240	188	665	5	618	380	479
6	1413	224	878	6	742	444	664
7	1565	251	1124	7	874	490	891
...
16	1999	331	2957	16	1208	591	2697
и) А – болгарский, В – польский				к) А – чешский, В – украинский			
1	35	46	—	1	125	31	—
2	145	149	4	2	350	147	9
3	332	310	23	3	587	363	63
4	543	491	56	4	814	602	159
5	736	690	107	5	1023	784	310
6	907	859	183	6	1157	908	497
7	1085	1026	290	7	1294	1034	730
...
16	1416	1504	1846	16	1561	1202	2703

Количественное соотношение изоглосс, охватывающих оба данных языка, и изоглосс, в которых данные языки участвуют порознь, можно представить в виде формулы:

$$q = V(A, B) / m \sqrt{V(A, \bar{B}) \times V(\bar{A}, B)}. \quad (21)$$

Различия в численных значениях условной величины q для разных пар языков могут быть интерпретированы как различия в степени изначальной генетической близости этих языков: чем выше значение q в строке с данным значением m (то есть при принятии во внимание изо-

глосс данной мощности), тем вероятнее исконная близость между языками этой пары; чем ниже значение q в строке с данным значением m , тем вероятнее роль конвергенции в эволюции языков, обнаруживающих значительную тесноту генетических связей (то есть языков с высоким значением индекса G). Пары языков, характеризующиеся низким значением показателя генетической близости G и не затронутые процессами взаимного сближения, также должны демонстрировать низкие значения q .

Приведем расчеты по данной формуле для упомянутых пар ближайших родственников языков и, ради сравнения, для двух пар языков, не состоящих в ближайшем родстве.

Таблица 13

m	бг & мк	сх & сн	чш & сц	вл & нл	пл & кс
1	0	0	0	0	0
2	0,211	0,121	0,203	0,120	0,089
3	0,255	0,174	0,221	0,129	0,070
4	0,270	0,204	0,198	0,108	0,068
5	0,256	0,203	0,184	0,091	0,067
6	0,226	0,199	0,185	0,083	0,072
7	0,220	0,201	0,185	0,084	0,075
...
(16)	(0,223)	(0,196)	(0,212)	(0,134)	(0,129)

m	бг & мк	сх & сн	чш & сц	вл & нл	пл & кс
1	0	0	0	0	0
2	0,118	0,180	0,063	0,013	0,020
3	0,196	0,274	0,167	0,024	0,045
4	0,218	0,275	0,197	0,030	0,069
6	0,211	0,260	0,193	0,034	0,081
7	0,213	0,256	0,195	0,039	0,090
...
(16)	(0,205)	(0,227)	(0,199)	(0,079)	(0,123)

Для наглядности преобразуем таблицу 13 в график (см. схему 5).

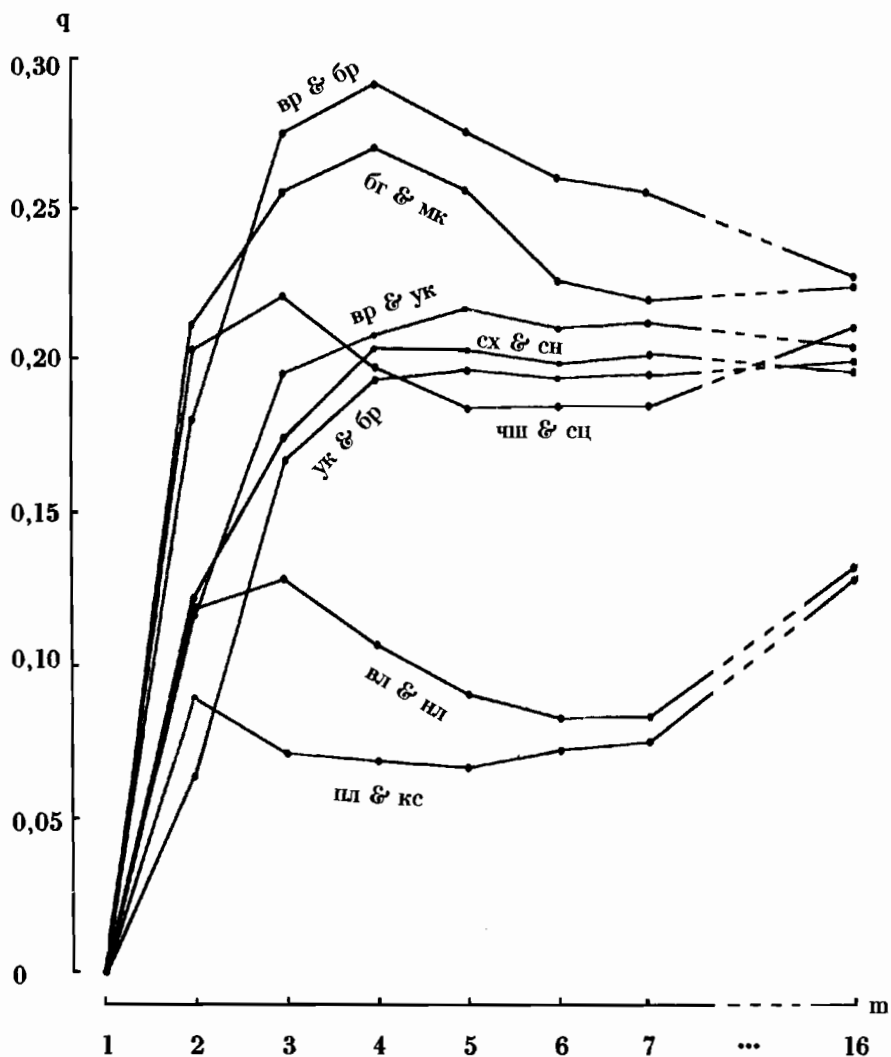


Схема 5

Соотношение общей и несовпадающей лексики (q) в праславянских словниках ближайшперодственных славянских языков (к разграничению дивергенции и конвергенции в сложении таксономических единиц)

Легко увидеть, что пары ближайшерадодственных языков весьма отчетливо разделяются на две категории. Первую из них составляют пары «болгарский & македонский», «сербохорватский & словенский», «чешский & словацкий» и парные объединения в кругу восточнославянских языков. Ко второй категории относятся пары лужицких и лехитских (польского и кашубско-словинского) языков. Для языковых пар первой категории отмечаются, начиная с порога $m = 3$, более высокие значения q в каждом классе изоглосс (наибольшие они у пар «великорусский & белорусский» и «болгарский & македонский»). Применительно же к языковым парам, составляющим вторую категорию, наблюдаются более низкие значения q .

Разумно предположить, что высокие значения индекса генетической близости G для языковых пар первого типа обусловлены изначальной их теснотой, развитием образующих эти пары идиомов на базе общих («промежуточных») праединств, а общность языков в парах второго типа вызвана их сближением в ходе эволюции, то есть первоначально идиомы, составляющие пары второго типа, были не столь близки друг другу и периода особого «промежуточного» праязыка, общего только для них и отличного от других родственных группировок, не переживали.

Тезис о несуществовании в прошлом особого пралужицкого единства [см.: Щерба, 1915, стр. 191; Штибер, 1934; Шустер-Шевц, 1959, стр. 582; Трубачев, 1963в; Базилович — Верещагин, 1970; ср., однако: Шахматов, 1916б; Летч, 1965] нашей статистикой разнонаправленных «маломощных» изоглосс подтверждается. Несхождения между двумя серболужицкими языками ярко выявляются в распределении лексем, которые связывают их только с южнославянскими и только с восточнославянскими языками (см. таблицу II в предыдущей главе): число лексических изоглосс, связывающих верхнелужицкий язык только с южнославянской группой и в которых не участвует никакой другой западнославянский язык, включая нижнелужицкий, в 1,7 раз превосходит число таких же изоглосс, связывающих его только с восточнославянской группой; у нижнелужицкого же соотношение таких изоглосс прямо противоположное: численность нижнелужицко-восточнославянских лексико-словообразовательных параллелей вдвое больше численности нижнелужицко-южнославянских лексических связей.

С еще большим резонансом, судя по квантитативным оценкам, сделанным в настоящей главе, мы можем говорить о нереальности особого лехитского праязыка, под которым понимался бы язык, лежащий в основе по крайней мере польского и кашубско-словинского (данные о полаб-

ском, как мы сказали, ненадежны): особенная близость польского и поморского, остатком которого являются современные кашубские говоры, вызвана конвергентными процессами (историю проблемы и различные мнения на этот счет см. в работах: [Рамулт, 1893; Бодуэн де Куртэнэ, 1897; Лоренц, 1902; Нитш, 1905; Розвадовский, 1915; Лоренц, 1925; Лер-Славинский, 1934; Селищев, 1941; Лер-Славинский, 1954; Лер-Славинский, 1956; Нитш, 1956; Штибер, 1956; Селищев, 1968; Поповская-Таборская, 1987 и др.]).

Из соображений корректности следует заметить, что сказанное отнюдь не означает, что в языковых парах, которые составляют первую категорию — с более высокими значениями q , явлений конвергенции не наблюдается. Речь идет лишь о конвергенции как главном, решающем факторе образования современных языковых таксономических общностей. Кроме того, по причинам слишком «крупного» членения славянских идиомов в словаре, служившем основным источником нашего материала (членение практически останавливалось на уровне языка — в нынешнем статусе — и не опускалось, за исключением русского, до диалектных идентификаций лексики), мы не имели возможности проверить излагаемую здесь методику на более низких таксономических единицах, — ср., например, проблемы, возникающие в связи с историей кайкавщины или торлацкого диалекта в контексте истории сербохорватского языка. В связи с упоминанием последнего уместно процитировать такое мнение: «...основные диалекты сербохорватского языка являются не столько «веточками» одной и той же ветви, сколько самостоятельными языковыми ветвями (возможно, отражающими различные миграционные пути прошлого), которые лишь позднее сблизились, т. е. подверглись конвергенции, не слившись, однако, полностью» [Бирнбаум, 1985, стр. 38].

В наш график не включены кривые, которые отражали бы статистику совместных и розных изоглосных связей у языков, не входящих в ближайшеразветвленные пары (то есть языков с взаимно невысокими значениями индекса генетической близости G). Они могли бы составить третью группировку языковых пар, кривые для которых на графике располагались бы еще ниже, чем линии, характеризующие лужицкие и лехитские языки. В таблице для сопоставления приведены числа, относящиеся только к двум из таких возможных парных объединений («болгарский & польский», «чешский & украинский»). Но подобных пар насчитывается больше сотни, и, будь соответствующие кривые помещены в графике, нижняя его часть покрылась бы густой чернотой: наши пробные оценки показали, что все эти пары склонны давать по изложенной

здесь методике чрезвычайно сходные пропорции разнонаправленных изоглоссных связей. нарастающих с повышением предельной мощности изоглосс.

Этот третий тип языковых пар не очень четко отделяется на графике от типа, к которому относятся пары лехитских и лужицких языков, что лишь укрепляет уверенность в правильной квалификации последних как приобретших значительное сходство (высокий индекс G) в результате конвергентных процессов.

Обращает на себя внимание также тенденция к сближению кривых графика с увеличением предельной мощности (m) подключаемых к расчетам изоглосс, иначе говоря, со снижением их специфичности. Различия между парами, для которых более вероятными являются предположения о первоначальном единстве составляющих их языков, и парами, которые сложились в результате аттрактивных движений, с включением все менее специфичных изоглосс становятся менее сильными. Однако дистанция между двумя первыми типами языковых пар сохраняется.

3. Об условиях эффективности предлагаемой методики

Необходимо заметить, что предлагаемая здесь методика статистического обнаружения конвергенции как фактора образования таксономических общностей может считаться эффективной только при том условии, что конкретные изоглоссы, на которых основываются расчеты, описываются достаточно большими числами. Практически это означает, что она может быть применена главным образом к лексическому материалу. Аналогичные расчеты на базе фонетических и морфологических изоглосс праславянского периода, исчисляемых для любой пары славянских языков не более чем двумя десятками, дадут картину, статистически слишком уязвимую.

Следовательно, косвенный результат наших наблюдений в этой области заключается в выводе о том, что статистический анализ фонетических и грамматических изоглосс, относящихся к более или менее глубокой древности, — инструмент довольно грубый. До нас доходят запечатления только очень сильных и широких фонетических и морфологических изменений, от мелких и частных процессов постоянного диалектообразования мощные волны последующих грандиозных фонетических и морфологических катастроф оставляют слишком мало следов. Лексика же, не будучи жесткой системой, подверженной тотальному выравнива-

нию, способна долго хранить память о прежних диалектных размежеваниях и единствах. Однако поспешно абсолютизировать преимущества лексического материала для статистического анализа глоттогенетической направленности мы не склонны, понимая, что обращение к иным языковым семьям и историко-типологическим языковым ситуациям может потребовать относительно подобных суждений существенных оговорок.

Возникает вопрос, можно ли на основе статистики узких разнонаправленных изоглосс выработать какой-нибудь добавочный коэффициент — поправку «на конвергентность», с тем чтобы, внося его в нашу формулу генетической близости (11), получить скорректированную, «окончательно» точную меру изначального языкового родства?

Сразу же скажем, что такого коэффициента нам выработать не удалось. Более того, возможность его установления представляется нам сомнительной. Тому имеется несколько причин.

Во-первых, пытаясь оценить возможное влияние конвергентных процессов на искомую нашей статистикой степень генетической близости языков, мы опирались не на понятие праязыкового лексического диалектизма, а на аппроксимативное по отношению к нему понятие узкой («маломощной») изоглоссы. Введение различных для разных языковых пар коэффициентов поправки «на конвергентность» требовало бы большей строгости в установлении таких коэффициентов. Более щепетильный отбор узких изоглосс — «заведомых» диалектизмов праязыкового периода — непременно сократит количество расцениваемого в этом качестве лексического материала, а значит, и понизит его статистическую достоверность.

Во-вторых, даже если мы решим опираться в этих поисках на понятие «маломощной» изоглоссы, то не вполне ясно, где провести демаркационную линию между необходимо малой и чрезмерно большой мощностью. В настоящей работе этот рубеж мы устанавливали в известной степени произвольно, полагаясь на здравый смысл, что для строгих количественных оценок не является достаточным основанием. Высокий предел мощности изоглосс, на котором мы остановимся в расчетах, будет способствовать сглаживанию различий между парами идиомов в итогах анализа, понижение же предела сократит привлекаемый к выработке коэффициента «на конвергентность» лексический материал и снизит его статистическую надежность. К тому же при повышении предельной мощности изоглоссы в нашей формуле (21) могут, пусть не так заметно, сказываться величины объемов праславянских словариков отдельных идиомов: все-таки в ее числителе находятся прямые, неотносительные

числа $V(A, B)$, а они, как мы знаем, сильно зависят от величин H соответствующих языков. Разумеется, при низком значении предельной мощности изоглосс m влияние общих объемов праславянской лексики данных идиомов малосущественно, но опять-таки — как определить, где оно начинает сказываться? По-видимому, для разных пар языков этот порог будет разным.

В-третьих, сведения о праславянском пласте лексиконов некоторых языков довольно скудны (а о некоторых — речь о лексике большинства диалектов полабских славян — и вовсе отсутствуют), в то время как наша статистика узких изоглосс, и мы это уже отмечали, применима лишь к обширным их спискам. Это означает, что для одних случаев поправочный коэффициент будет статистически более обоснованным, для других — менее. При различающейся, хоть, как мы полагаем, и незначительно, степени надежности величин G , характеризующих разные пары языков, мы будем корректировать их также «разнонадежными» коэффициентами, что, очевидным образом, большой точности нашим статистическим оценкам не прибавит.

Итак, мы считаем, что предложенная в настоящей главе статистическая методика позволяет обнаружить роль конвергенции в формировании современных языковых таксономических единств и даже приблизительно (в относительных показателях) измерить ее, но из-за известных квазитипологических различий между отдельными славянскими идиомами и некоторой неопределенности, конвенциональности понятия праязыкового лексического диалектизма выработка коэффициентных поправок к формуле генетической близости затруднительна.

Глава 8

Проблема древненовгородского диалекта sub specie лексикостатистики

Новгородцы, быв всегда старшими сынами России, вдруг отделились от братьев своих; быв верными подданными князей, ныне смеются над их властью... и в какие времена?

Карамзин. «Марфа-посадница,
или Покорение Новгорода»

1. Проблема древненовгородского диалекта в современных историко-диалектологических исследованиях

Публикацией обширного исследования А. А. Зализняка «Наблюдения над берестяными грамотами» [Зализняк, 1984], практически воспроизведенного двумя годами позднее под названием «Новгородские берестяные грамоты с лингвистической точки зрения» [Зализняк, 1986а], а затем его докладами «Значение новгородских берестяных грамот для истории русского и других славянских языков» на заседании бюро Отделения литературы и языка АН СССР [Зализняк, 1988а] и «Древненовгородский диалект и проблемы диалектного членения позднего праславянского языка» [Зализняк, 1988б] внимание славистов было вновь привлечено к крупной проблеме, порождаемой довольно многочисленными диалектными особенностями идиома, отраженного в группе памятников, которая относится к древнейшим фиксациям славянской речи (точнее, второй после старославянского, но первому из письменных свидетельств бытовой, живой славянской речи).

Ряд важных черт связывает древненовгородский диалект с западнославянскими языками: — отсутствие палатализации задненебных в сочетаниях **kv*, **gv* в позиции перед *ě*, *i*, *ь*; — сохранение взрывного элемента в древних сочетаниях **tl*, **dl*, с передвижкой смычки к заднему небу: *kl*, *gl*; — совпадение рефлексов **tj* и **dj* с *k* и *g*, «находящимися „в позиции второй палатализации“ (т. е. перед *ě* или *i*)», с тем различием,

что в западнославянских эволюция этих рефлексов пошла дальше; — реализация фонемы *ě* как относительно открытой гласной, подобно ситуации в польском; — устранение сочетаний *вл'*, *мл'*: «Не исключено, что тенденция к устранению сочетаний *вл'*, *мл'* находится в некоторой отдаленной связи с отсутствием таких сочетаний (кроме позиции начала морфемы) в древних диалектах западнославянской зоны»; — морфологическое оформление существительных с суффиксом *-ък-* от имен собственных *о-*склонения как *masculina*, подобно западнославянскому оформлению таких имен, а не как *neutra*, что наблюдается в диалектах Юго-Западной Руси и в южнославянских языках; — совпадение родительного падежа единственного числа женского рода с дательным и местным падежами единственного числа в адъективном и местоименном склонении; — другие особенности, см.: [Зализняк, 1988б, стр. 166–175].

Чертами, объединяющими диалект, который отражается новгородскими берестяными грамотами, с западнославянской зоной, он обязан, по мнению А. А. Зализняка, севернокривичскому компоненту. Другой компонент, влившийся в древненовгородский диалект, — говоры ильменских словен — стоит ближе к остальным восточнославянским идиомам, которые в свою очередь обнаруживают много общего с языками южнославянской зоны и особенно сербохорватско-словенской подгруппой.

Древненовгородский, таким образом, представляет тип смешанного диалекта, развившегося в зоне контактов двух групп говоров с разными генетическими характеристиками. Последние, собственно, заключаются не в изначальной принадлежности древненовгородского диалекта в целом или одного из его компонентов какой-либо восточнославянской диалектной области, а в особой его архаичности, отражении им состояния, предшествовавшего окончательному размежеванию восточных, западных и южных славян (см.: [Зализняк, 1988а, стр. 99; ср.: Трубецкой, 1987а, стр. 166–167]). Древненовгородский диалект, как считает А. А. Зализняк, разрушает традиционные устойчивые представления о моногенезе восточнославянской языковой группы [Зализняк, 1988б, стр. 176].

Сама проблема происхождения древненовгородского диалекта далеко не нова.

Одна из наиболее крупных работ на эту тему — вышедшая посмертно статья (по сути целая монография) Д. К. Зеленина «О происхождении северновеликорусов Великого Новгорода» [Зеленин, 1954], где проблема рассматривается в полемике прежде всего с работами А. А. Шахматова о возможном ляхском влиянии на сложение восточнославянских диалектов и этносов (см.: [Шахматов, 1907; 1911; 1913; 1915]), далее — П. А. Рас-

торгуева (см.: [Расторгуев, 1927]) и др. Она решается у Зеленина выдвиганием тезиса о переселении на берега Ильменя и Волхова несколькими партиями выходцев из области восточнобалтийских (поморских) славян, переселении, случившемся (или случавшемся) «так рано, что до летописца XI в. дошли лишь глухие предания» [Зеленин, 1954, стр. 95]. Для обоснования этого тезиса Д. К. Зеленин рассматривает многочисленные, по преимуществу фонетические, параллели между западно- и севернорусскими диалектами, включая говоры Сибири, с одной стороны, и западнославянскими (польским, поморским, полабским) языками, с другой. Среди аргументов Зеленина — и лексические факты, почерпнутые как из живых говорів, большей частью сибирских, так и из древних письменных памятников, в частности, из Первой Новгородской летописи, проанализированной Н. М. Петровским (см.: [Петровский, 1922, стр. 370 и след.]) с точки зрения наличия в древненовгородском западнославянском лексического компонента.

К настоящему времени сложилась довольно обширная лингвистическая литература, касающаяся проблемы генетических связей между севером западнославянской зоны и западом восточнославянской области (см.: [Васильев, 1907; Дурново, 1929; Ларин, 1960; Горнунг, 1963; Мжельская, 1963; Роспанд, 1965; Мокиенко, 1969а; Мокиенко, 1969б; Филин, 1972; Ланг, 1975; Толстой, 1977; Толстой — Толстая, 1979; Хабургаев, 1979 — с библиографией; Хабургаев, 1980; Пивторак, 1984 — с библиографией, в том числе исторических работ]; из последних работ: [Николаев, 1988—1989, 1990, 1993]; о славянских связях Новгорода по данным археологии см.: [Янин, 1992]). Исследования А. А. Зализняка, таким образом, дают новый и богатый материал для размышлений на достаточно старую тему, без привлечения которого — теперь это вполне очевидно — решение проблем восточнославянского глоттогенеза попросту неосуществимо.

Возможны ли лексикостатистические измерения внешних ориентаций древненовгородского диалекта, которые могли бы пролить свет на его происхождение?

2. Изоглосные лексические связи древненовгородского диалекта

В поисках ответа на этот вопрос обратимся к данным, которые можно извлечь из Словоуказателя к новгородским берестяным грамотам [Зализняк, 1986б].

В Словоуказатель помещена лексика всех найденных к моменту его создания берестяных грамот Новгорода (№ 1–614) и Старой Руссы (14 грамот, отражающих тот же диалект), а также Пскова, Смоленска, Витебска, Мстиславля, Твери; всего 645 грамот. Объем Словоуказателя – примерно 2 250 лексем. Приблизительно 840 из них – имена собственные (сюда не включаются производные – отойконимические – наименования жителей), причем большинство из них не имеет соответствий в апеллятивном слое лексики и/или является заимствованиями. Таким образом, праславянский слой лексикона древненовгородских берестяных грамот ожидается довольно скудным по величине.

К лексикостатистическому анализу по испытанным выше процедурам мы можем привлечь лишь ту часть праславянской лексики древненовгородского диалекта, которая может быть возведена к праславянским реконструкциям, осуществленным в первых пятнадцати выпусках ЭССЯ. Сколь ни интересна и ярка была бы лексика, остающаяся за обозначенными пределами, ее статистическое исследование на фоне инославянского материала нецелесообразно ввиду неполноты наших представлений о ее корреспонденциях в языках разных групп и подгрупп славянской семьи. Оно будет возможно лишь по выходе всех выпусков полного этимологического словаря славянских языков.

Сопоставление лексики, представленной Словоуказателем, с материалами обследованной части ЭССЯ позволило выделить 378 праславянских слов, которые могли быть использованы в статистическом анализе праславянского словаря древненовгородского диалекта с целью выявления его квантитативных лексических связей с другими славянскими идиомами. Это составляет всего лишь около 5% всего объема части праславянского лексического фонда, подвергшейся статистической обработке. Величина *N* у новгородского диалекта по берестяным грамотам оказалась весьма низкой, меньшей даже, чем у полабского языка. Это означает, что статистические наблюдения над праславянской лексикой диалекта древнего Новгорода будут отличаться не слишком высокой надежностью.

Именно по этой причине мы нашли нерациональным специальное вычленение лексики *ранних* берестяных грамот, относящихся, по Зализняку (см.: [Зализняк, 1986а, стр. 91]), к домонгольскому периоду и составляющих 37% всех имеющихся грамот: такой отбор сделает статистику еще менее надежной.

Для обрисовки изоглоссных связей древненовгородского диалекта важно указание на наличие в его словарном составе праславянских по происхождению слов, не встречающихся в других восточ-

нославянских идиомах. К таковым относятся лексемы **dorgobqđь* (Дорогобудъ, личное имя; свидетельство ценно тем, что до него фиксировалось только производное **dorgobqđьь*, включенное в ЭССЯ [вып. 5, стр. 75] с отражениями в сербской, чешской и древнерусской ойконимии, из восточнославянских *Дорогобуужей* упомянут лишь смоленский и не упоминается волынский, название которого встречается в Лаврентьевской летописи раньше смоленского — под 1084, 1097 и 1100 гг.), **jaroměьь*/**jaromirьь* (Аромиръ, личное имя; отмечено в болгарском, сербохорватском, старочешском, старопольском антропонимиконах, ЭССЯ [вып. 8, стр. 176]), **kleščьь* 'лещ' (*клещьь*, которому А. А. Зализняк придает особенную значимость как доказательству переклички древненовгородского с «очевидными соответствиями в западнославянских языках» [Зализняк, 1986а, стр. 121] — старопольском, словинском, нижнелужицком, далее — с иным рефлексом инициальной консонантной группы: *dleščьь* — в старочешском; см. также: [Усачева, 1977, стр. 118–119; ЭССЯ, вып. 14, стр. 144]). Правильное толкование др.-новг. *клещьь* предложено в работе: [Куза — Медынцева, 1974, стр. 222–223]. Строго говоря, древненовгородское слово не является одиноким на восточнославянской территории, ср. производные *клещійнец*, *клещійнци*, *клещинец* 'род рыболовной снасти', уральский фразеологизм *клещ на уду*, недоброе пожелание рыболову [СРНГ, вып. 13, стр. 292–293] топонимы *Клещино (озеро)* и под. (ср. *Плещеево озеро*). «Западнославянским» в нем следует считать отражение *tl- > kl-*, подобно спорадическому переходу среди нных *-tl-*, *-dl-* в *-kl-*, *-gl-* в севернолехитских говорах [Зализняк, 1986а, стр. 121]), **krqрььь* (*крупььь* 'мелкий'; отмечено еще в болгарском, старочешском, польском и словинском, а также в церковнославянском [ЭССЯ, вып. 13, стр. 27]), **kyselььь*/**kyselьььь* (*кисельььь* 'кислый'; соответствия — во всех южнославянских, чешском и польском, также в русскоцерковнославянском [ЭССЯ, вып. 13, стр. 271] — в отличие от господствующих на восточнославянской территории рефлексов формы **kys(ь)льььь*), хотя производные от формы с *e/ě* в суффиксе в восточнославянских языках, конечно, отмечаются, ср. русск. *кисель*, *-еля*, укр. *кисіль*, русск. диал. *киселица*, *киселяй* [СРНГ, вып. 13, стр. 226, 228])¹.

¹ При принятии в расчет только данных ЭССЯ, в этот список должно попасть и личное имя *Добровитъ* (продолжения **dobrovitьь* отмечены в ЭССЯ лишь для сербохорватского, словенского и старочешского), однако донские говоры знают прилагательное *добровитый* 'хороший' [СРНГ, вып. 8, стр. 77].

Если иметь в виду и слова, известные по древнерусским памятникам, но не свидетельствуемые живыми восточнославянскими языками, то этот список должен быть пополнен лексемами **a ti* союз (*ати*, *ать*, далее *ати но*, *ать но*, *ать ти*; ср. [Зализняк, 1986а, стр. 163–164]; также – в чешском, старопольском [ЭССЯ, вып. 1, стр. 40]), **bebrъ* (*бебръ* 'бобр', также – в сербскоцерковнославянском, болгарском, словенском, верхнелужицком [ЭССЯ, вып. 1, стр. 174]; но возможно также отнесение к вариантной праформе **bbbrъ*, рефлектирующей в сербохорватском и древнерусском [ЭССЯ, вып. 3, стр. 158]), **boguslavъ* (в виде производного *Богуславъ*, отражения производящего – в болгарском, старочешском, польском, личные имена [ЭССЯ, вып. 2, стр. 161]), **borislavъ* (*Бориславъ*, личное имя; также – в болгарском и старочешском [ЭССЯ, вып. 2, стр. 203]), **čędъ* (*чадь*, также – в старославянском и сербохорватском [ЭССЯ, вып. 4, стр. 104]) и, возможно, иными (об упомянутых здесь именах собственных можно, впрочем, заметить, что они не обязательно отсутствуют на восточнославянской территории в поздние времена как продолжения исконного ономастического репертуара и существуют как заимствования из других славянских именников).

Вместе с соответствиями в других восточнославянских идиомах древненовгородский диалект имеет следующие корреспонденции только с южнославянскими языками (но не с западнославянскими): **ba-janъ* (*Боканъ*; болг. *баян*, несклоняемое прилагательное, фолькл. – исключительно редкое, см.: БЕР, т. 1, стр. 38), **blędъ* (*бладъ*), **bl'usti* (*блюсти*), **čьгъпница* (в виде производного *чермничъничи*, от названия ткани), **gostъba* (*гостъба*, *гозба*), **dorъ* (*доръ* 'земля, расчищенная под пашню'), **kačovъ(ь)* (*каковъ*), **konъčati* (в префиксальном производном *доконъчати*, с *a*-тематизацией в отличие от формы на *-iti*, продолжаемой и западнославянскими), **koževъnikъ* (*кожевникъ*).

Напротив, вместе с другими восточнославянскими отражениями древненовгородский имеет корреспонденциями только в западнославянских (но не в южнославянских) языках **bęgt'i* (в префиксальном производном *побъчи*), **brtъ* (*бортъ*), **galqza/*galqzъ* (*Голуза* или *Голузъ*; с иным, однако, вокализмом корня, как в чешск. *holeska* 'ветка', ср. укр. *голузъ* 'сук, ветвь', см.: [ЭССЯ, вып. 6, стр. 95–96]), **gor'eslavъ/*gorislavъ* (в виде производного *Гориславичъ*, отчество), **gostęta* (*Гостыта*), **ъbstъbъka* (*устебка*), **ъzvętati* (в чешском, древнерусском (русскоцерковнославянском) и русских диалектах отмечается только глагол на *-iti*, *a*-тематизация – только в новгородском?; существительное **ъvъčъ* отмечается в южно- и восточнославянской группах,

но не в западославянских языках, см. [ЭССЯ, вып. 9, стр. 94–95]), **korъbka* (коробка), **laditi* (*se*) (ладитисл), **legati* (*se*) II (в суффиксальном производном Лагачь, прозвище).

Наконец, случаи, когда лексические изоглоссные (праславянского происхождения) связи древненовгородского диалекта замыкаются только восточнославянским ареалом: **bersto* ср. р. (бересто 'берестяная грамота'), **četvъrgъ* (четвергъ), **deševъ(jь)* (дешевыи), **domažirъ* (Домажиръ, личное имя, производные Домажировъ, Домажировичъ; к имеющимся в ЭССЯ [вып. 5, стр. 69] восточнославянским продолжениям **domažirъ*, вероятно, следует добавить белорусский топоним Дóмжерицы/Дóмжарыцы [Жучкевич, 1974, стр. 105]: «В основе названия древнерусское имя Домажир»), **godъjъ* (гожиш), **gъlъkъ* (глекъ), **kolъ/*kolobъ* (**kolobъ*?) (в виде суффиксального производного *колобья* 'ком, пригоршня'), **kororye* (Копорья жен. р., нынешнее Копорье в Санкт-Петербургской губ. (Ленинградской обл.)), **koręga/*koręka* (в виде суффиксального производного *каракула* 'род железного инструмента', ср. *каракули*, яросл. *каракуля* 'железные навозные вилы', см.: [Фасмер, т. II, стр. 192]), **košъjъ* (Коциш, прозвище), **kriti/*krynъti* 'купить' (*крити*; форму **krynъti* А. А. Зализняк считает лексикографической фикцией, идущей от Срезневского, см.: [Зализняк, 1984, стр. 115; 1986а, стр. 174]; к древнерусскому слову, имеющему далекие индоевропейские – индоарийские, греческие, кельтские, а в производных и балтийские – связи, но изолированному на собственно славянской почве, О. Н. Трубачев привлекает укр. диал. *кринѹтис'а* 'взяться, схватиться', белор. (*да*)*кранѹцца* 'дотронуться' [ЭССЯ, вып. 13, стр. 74–75], что, на наш взгляд, выглядит скорее гипотетичным, чем вполне доказанным; если же, однако, супплементации Трубачева признать правомерными, то реконструкция Срезневским глагольной формы на *-nъ* получает новые аргументы), **kъltъkъ* (колткы 'серьги'; производящее **kъltъ* отражено, по мнению О. Н. Трубачева, см. [ЭССЯ, вып. 13, стр. 192], в чешск. *klut* 'выбоина на дороге'; производящий глагол **kъltati* известен польскому в значениях 'резать, кромсать; молоть; крутить, мешать', 'качаться' [там же, стр. 190]).

Разумеется, выход за пределы алфавитного диапазона *a – lo-*, которым вынужденно ограничено наше исследование, даст множество других лексических примеров, иллюстрирующих разнонаправленные связи древненовгородского диалекта, связи, тянущиеся еще с праславянского времени, по крайней мере, с позднепраславянского периода. Например, А. А. Зализняк указывает др.-новг. *рути* 'подвергать конфискации,

секвестру», которому «соответствует большое словообразовательное гнездо в словенском. Ни в каком другом славянском языке такого гнезда нет» [Зализняк, 1986а, стр. 170]; *провадити* с соответствиями в украинском и польском [там же, стр. 176]; безйотовую основу *-въта-* в глаголе *извътати* 'заявить о правонарушении', имеющую соответствия в виде единичных реликтов в сербохорв. (черногор.) *вијѣтати* 'обещать', словен. *obětati* 'обещать' [там же, стр. 177]; *тъгъдъ* с прямым соответствием в др.-чешск. *thed* 'тогда' [там же, стр. 189]; союз *та* 'да, и' с параллелями прежде всего в украинском (впрочем, как отмечает А. А. Зализняк, южнорусские и украинские факты делают вероятным — но не более — предположение о южнорусском происхождении берестяной грамоты № 109) [там же, стр. 190]; личное имя *Сторонька*, возможно, гипокористическое от незасвидетельствованного в восточнославянских **Сторониславъ* с соответствием в др.-польск. женском имени *Stronislawa* [там же, стр. 216]. Добавим еще личное имя *Лудьславъ* с параллелью в ст.-чешск. *Ludislav* (см. [ЭССЯ, вып. 16, стр. 167]). Однако мы останемся в обозначенных алфавитных границах, не будучи уверенными в достаточной полноте наших представлений об ареальных свойствах слов вне указанного диапазона, без чего попытки статистического анализа этой лексики обесцениваются.

Вне всякого сомнения, словарь ЭССЯ, с которым мы сличаем древненовгородскую лексику, неполон и будет в дальнейшем расширяться за счет новых источников и вновь обнаруженных межславянских параллелей. По-видимому, и в обследованном фрагменте древненовгородского Словоуказателя можно найти лексику, которая не связывается с имеющимися в ЭССЯ праславянскими реконструкциями, но может претендовать на праславянскую древность. Мы относим к такой потенциально праславянской лексике **bratilo* (ср. др.-новг. *братиловичь*, отойконимическое наименование, ср. укр. диал. *брати́ло* 'брат' [ЕСУМ, т. 1, стр. 246]), **dědilo* (др.-новг. *Дѣдила*, личное имя, ср. болг. *Дедил*, личное имя, 1491 г. [БЕР, т. 1, стр. 472]; далее укр. диал. *дєдильница* 'сньгъ'² [ЕСУМ, т. 2, стр. 86], белор. топоним *Дєдiловици/Дзядзiлавичы* [Жучкевич, 1974, стр. 95]; см. еще: [Роспонд, 1979, стр. 12]; предполагается, что имя *Дѣдило* — сокращение от композит типа **dědoslavъ*, **dědomirъ*, **dědomilъ*), **bratęta* (*Братата*, личное имя), **budo-*

² Втяжение украинского названия сныти в круг производных от **děd-* вторично; изначально этот ботанический термин принадлежит корню **dęgl-*, см. ранее у нас о **dęglъjъ*/**(j)aglъjъ* [Журавлев, 1990, ч. 1, стр. 8]: *дѣглиця* и под.

та, менее вероятное **bqdota* (*Будота*, личное имя; функционирование корня как первого компонента ономастических сложений ср. **budimirъ*, **budislavъ*, **budigojъ*, **budimilъ*), **dětja* (*дѣтъя*, собирательное), **domaslavъ* (*Домаславъ*, личное имя), **gadъka* (кроме укр. *gádka*, относительно которого можно предполагать западнославянское влияние, ср. польск. *gadka*, чешск. *hádka*, — также др.-новг. *гадка* 'предположение, ожидание, надежда'), **gostiměръ* (ср. производное др.-новг. *Гостьмерици*, топоним), **kojъ* (*Кои*, личное имя; возможно, к праслав. **kojiti* 'вскармливать (молоком матери)', 'укрощать' и др., **kojěta* — ст.-чешск. *Kojata*, ст.-польск. *Kojěta*, личные имена [ЭССЯ, вып. 10, стр. 113]; не исключена, однако, связь с **kyjъ*) и т. д. Но, не имея полной славянской картины распространения этих слов, мы сочли невозможным включать в статистическую обработку древненовгородского лексического материала лексику, отсутствующую в списке реконструкций ЭССЯ.

3. Лексикостатистический анализ и его результаты: негомогенность восточнославянской языковой группы

С о ф и я
Смесь языков?
Ч а ц к и й
Да, двух, без этого нельзя ж!

Грибоедов. «Горе от ума»

Оценка статистической близости древненовгородского диалекта с другими славянскими идиомами на материале праславянского лексического наследия, осуществленная по нашей формуле (II), дает такие результаты (перед индексом близости *G* сообщаются числа лексических корреспонденций в абсолютном выражении $V(A, B)$ и их доля в процентах в объеме праславянского словника древненовгородского диалекта; величина $G \times 1000$ дается в ненормированном виде) (см. таблицу 14).

Весьма низкие абсолютные числа лексических схождений древненовгородского диалекта с иными родственным идиомами, резкое отличие общего объема доступных исследованию праславянских лексических фактов в новгородском от аналогичных характеристик других языков дают заметно перекошенную картину его статистических связей. Происходит это не из-за несовершенства формулы определения лексикостатистической близости (хотя мы далеки от утверждений о ее безупречности), а

Таблица 14

A	&	B	V(A,B)	%	G
др.-новг.	&	ст.-слав.	219	57,9	2,3725
		болг.	295	78,0	1,2506
		макед.	232	61,4	1,3338
		сербохорв.	338	89,4	1,1258
		словен.	292	79,2	1,1247
		чешск.	326	86,2	1,1546
		словац.	274	72,5	1,1536
		в.-луж.	237	62,7	1,4156
		н.-луж.	234	61,9	1,6438
		полаб.	117	31,0	2,2124
		польск.	309	81,7	1,2998
		кашуб.-словин.	216	57,1	1,4477
		др.-русск. ³	342	90,5	2,0534
		сев.-в.-рус.	341	90,2	1,2402
		ср.-рус. гов.	336	88,9	1,3479
		юж.-в.-рус.	335	88,6	1,3217
укр.	339	89,7	1,3231		
белор.	306	81,0	1,3485		

по причинам, которые мы затрагивали раньше, обсуждая высокие показатели связи между полабским и старославянским: малый объем праязыкового наследия в словаре какого-либо языка предполагает бóльшую в нем долю слов, отличающихся значительным распространением (в нашем случае — общеславянских), и меньшую — узколокальных и острохарактерных (праславянских диалектизмов). Со старославянским языком, например, северновеликорусское наречие связывается 20,8% своей праславянской лексики, древнерусский язык, объем праславянского словника которого в полтора раза меньше, — 33,2%; у древненовгород-

³ Древненовгородский диалект — «часть» древнерусского языка и, строго говоря, данные Словоуказателя (который мог стать источником ЭССЯ только с 16-го выпуска) следовало автоматически влить в материалы, относящиеся к рубрике «др.-русск.», в качестве дополнений; тогда значения V и % в строке «др.-русск.» выглядели бы как 378 и 100. Соответственно, гораздо выше оказалось бы в этой строке и значение коэффициента родства (по нашим прикидкам, оно возросло бы на 0,5—0,6 и превысило бы значение G для древненовгородского и старославянского, то есть оказалось бы у рассматриваемого здесь диалекта максимальным). Однако в данном случае, имея в виду проблему, о которой сказано выше, мы предпочли возможность независимого рассмотрения древненовгородского диалекта в его связях с другими идиомами.

ского же, с его минимальным H , — 57,9%. С болгарским — соответственно 52,0%, 59,2% — и 78,0%; и т. д. Сам характер общей лексики не может не сказаться на статистических результатах. Языки с невысокими показателями праязыкового лексического наследия почти неизбежно будут давать завышенные значения индекса родства между собою, которые интуитивно воспринимаются как завышенные.

Именно поэтому наибольшие величины G у древненовгородского обнаруживаются в его отношениях со старославянским и полабским языками.

Сходным образом в данном случае ведет себя и формула коэффициента сопряженности Φ (формула (13)). Несколько более правдоподобную картину для древненовгородского рисует коэффициент ассоциации Q (формула (12)), хотя и здесь просматривается эффект зависимости его величины от общего объема праславянских словников сравниваемых языков: за пределами восточнославянского круга наибольшие значения коэффициента Q древненовгородский диалект дает в парах с сербохорватским и чешским языками, то есть с теми, у которых значения H наибольшие. Но уже само существенное расхождение результатов, полученных с помощью разных (но в случае Φ и Q довольно близких) методик статистического анализа (расхождение, обратим внимание, не в конкретных величинах, но в их соотношениях между собою в пределах каждого индекса), указывает на очень невысокую степень надежности этой статистики (см. таблицу 15).

Попытаемся улучшить наши результаты, вернее, извлечь из них более определенную информацию.

Сравним поведение древненовгородского диалекта с поведением других выделяемых у нас восточнославянских идиомов в их лексикостатистических связях с невостоочнославянскими языками, просчитанных по одной и той же формуле.

Для этого воспользуемся таблицей 14. Исключив из расчетов данные, касающиеся полабского и старославянского языков (первого — ввиду крайней ненадежности вычисленного индекса G , второго — принимая во внимание его отличный от остальных языков статус), определим среднюю величину \bar{G} для пар, в которые с древненовгородским диалектом сопрягаются невостоочнославянские языки. К этой полученной величине \bar{G} отнесем конкретные величины G таблицы 14, то есть сделаем расчеты по формуле

$$G(A,B) / \bar{G}(A,B), \quad (21)$$

где A — древненовгородский диалект, B — невостоочнославянские языки (за исключением старославянского и полабского).

Таблица 15

A	&	B	Q	Ф
др.-новг.	&	ст.-слав.	0,8356	0,2855
		болг.	0,7938	0,1970
		макед.	0,7622	0,2056
		сербохорв.	0,8703	0,1896
		словен.	0,7686	0,1820
		чешск.	0,8401	0,1880
		словац.	0,7566	0,1905
		в.-луж.	0,6437	0,1830
		н.-луж.	0,7076	0,2451
		полаб.	0,8174	0,2448
		польск.	0,8287	0,2079
		кашуб.-словин.	0,7393	0,2079
		др.-русск.	0,9339	0,2844
		русск.	0,9170	0,1930
		сев.-в.-рус.	0,8912	0,2064
		ср.-рус. гов.	0,8858	0,2146
		юж.-в.-рус.	0,8829	0,2136
		укр.	0,8924	0,2147
		белор.	0,8235	0,2077

Аналогично сделаем пересчет для остальных восточнославянских идиомов, исходя из данных нашей основной таблицы 8 (в версии SLAV) и подставляя в А формулы (21) вместо древненовгородского диалекта другие восточнославянские идиомы.

Сведенные вместе, результаты пересчета выглядят следующим образом:

Таблица 16

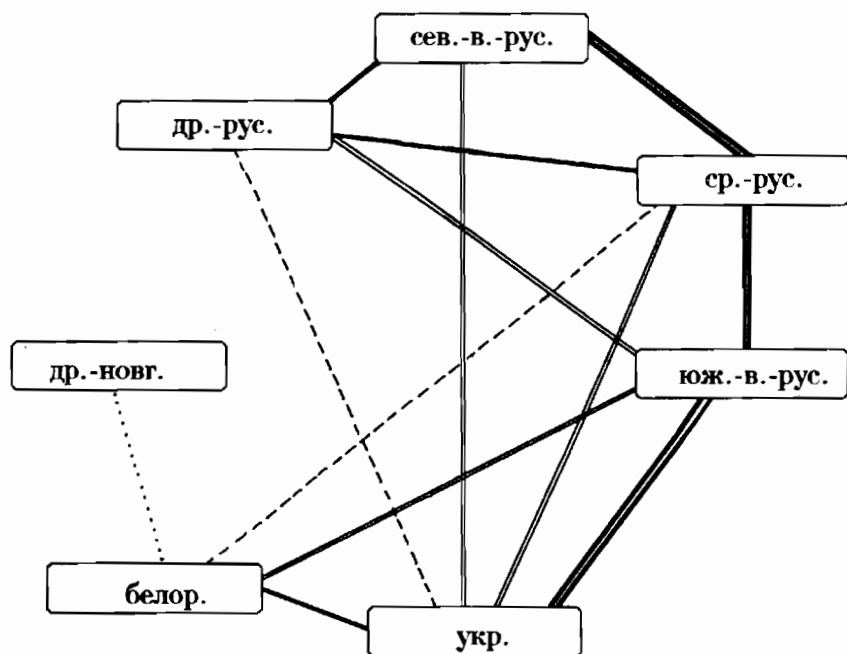
	др.-новг.	др	ср	цр	юр	ук	бр
болг.	0,966	1,056	1,049	1,042	1,014	0,980	0,945
макед.	1,030	1,053	0,983	0,984	0,948	0,922	0,882
с.-хорв.	0,869	1,060	1,075	1,059	1,028	0,973	0,941
словен.	0,868	1,045	1,051	1,033	0,999	1,000	0,946
чешск.	0,892	0,976	1,020	1,017	1,023	1,041	1,014
словац.	0,891	0,930	0,996	1,009	1,022	1,098	1,088
в.-луж.	1,093	0,901	0,921	0,908	0,943	0,959	0,994
н.-луж.	1,269	0,994	0,900	0,900	0,929	0,921	0,967
польск.	1,004	1,084	1,079	1,100	1,131	1,137	1,177
кашуб.	1,118	0,950	0,926	0,949	0,963	0,968	1,047

Из таблицы 16 видно, что отношения древненовгородского диалекта с языками южно- и западнославянской групп несколько иные, чем у остальных восточнославянских идиомов. «Стандартный» [Зализняк, 1988а, стр. 94] древнерусский язык, северновеликорусское наречие и среднерусские говоры в своих внешних лексикостатистических связях отдают предпочтение южнославянским; если к южнославянским присовокупить близкие к ним во многих отношениях чешский и словацкий языки, то этой группе перед лужицко-лехитской будут отдавать предпочтение и южнорусское наречие с украинским языком. Белорусский язык обнаруживает преимущество своих западнославянских связей перед южнославянскими. У древненовгородского диалекта наблюдается очевидное тяготение к западнославянской зоне, главным образом, к серболужицкой и лехитской подгруппам, что напоминает вневосточнославянские «пристрастия» белорусского языка.

Меру сходства в поведении восточнославянских идиомов в лексикостатистических связях с южно- и западнославянскими языками можно установить с помощью корреляции. Вычисление коэффициента корреляции (по формуле (20) между вертикалями таблицы 16 приводит к следующим соотношениям. Древненовгородский диалект со всеми остальными выделяемыми здесь восточнославянскими языками и диалектами дает отрицательную корреляцию (наименьшее отрицательное значение, $r = -0,011$, — с белорусским языком, наибольшее, $r = -0,808$, — с северновеликорусским наречием). Все же остальные идиомы коррелируют между собой положительно (за исключением пар «древнерусский & белорусский» и «северновеликорусский & белорусский», у обеих пар фиксируются низкие значения отрицательной корреляции: $r = -0,134$ и $r = -0,125$ соответственно). Наиболее сильная положительная корреляция наблюдается между северновеликорусским наречием и среднерусскими говорами ($r = +0,975$), самая слабая ($r = +0,290$) — между среднерусскими говорами и белорусским языком.

Сходства между восточнославянскими идиомами в распределении их лексикостатистических тяготений к языкам южно- и западнославянской групп могут быть изображены схематически (см. схему 6).

Схема изображает восточнославянскую группу довольно компактным единством, лишь древненовгородский диалект выглядит на ней инородным телом (корреляция с белорусским языком, так же как и с прочими восточнославянскими идиомами, является отрицательной, но только наименьшей по модулю численного выражения). Конечно, обособленность древненовгородского диалекта здесь сильно заострена принятием во внимание лишь внешних ориентаций, но это и было целью описанной процедуры.



	$r (\geq)$
▬▬▬	+ 0,850
▬▬	+ 0,675
▬▬▬	+ 0,500
▬	+ 0,325
- - -	+ 0,150
.....	- 0,025

Схема 6

Коррелятивные связи (r) между восточнославянскими идиомами в их вневосточнославянских тяготениях (к проблеме древненовгородского диалекта)

Последнее обстоятельство — обращение здесь лишь к внешнеславянским лексикостатистическим связям — является, по-видимому, главной причиной поразительного, на первый взгляд, факта: крайне низкой (точнее говоря, чрезвычайно высокой отрицательной) корреляции между показателями внешних тяготений для древненовгородского диалекта и северновеликорусского наречия. Принятие в расчет собственных лексикостатистических связей между указанными идиомами (по коэффициенту ассоциации, Q , они являются значительными), разумеется, сильно умалило бы парадоксальность этой картины. Кроме того, древний новгородский диалект — это хотя и важнейшая, но не единственная база, на которой формировалось северновеликорусское наречие. Да и складывалось оно уже в основном в относительно поздний период, когда диалект новгородских земель прошел этап значительной конвергенции с ильменско-словенскими говорами, отличавшимися от «стандартного» древнерусского языка гораздо менее, чем «ранний» новгородский (эффект прослеживающейся по берестяным грамотам конвергенции специально отмечается А. А. Зализняком (см.: [Зализняк, 1988а, стр. 97; Зализняк, 1988б, стр. 165]) как один из главных сюрпризов лингвистического изучения новгородской бересты). Возможно, что на полученной картине в какой-то мере сказалось и обстоятельство, отмеченное в гл. 4, п. 6, — оперирование у нас данными северновеликорусского наречия с южной границей, как она определена К. Ф. Захаровой и В. Г. Орловой, то есть севернее, чем в Опытте диалектологической карты, 1915.

Допустимо думать, что различия между «стандартным» древнерусским языком и его потомками, с одной стороны, и древненовгородским диалектом, с другой, выявленные корреляционным анализом данных о внешних статистических «предпочтениях», отражают негомогенность восточнославянской языковой группы. Просматривающиеся связи древненовгородского диалекта с языками других славянских групп, прежде всего западной, могут трактоваться как свидетельство особой архаичности севернокривичских говоров, легших в его основу. Для суждений же о «западнославянском» генезисе древненовгородского диалекта, ввиду заметного количества лексических переключек с южнославянскими языками (см. выше), серьезных оснований, с нашей точки зрения, не имеется. Оценивая взгляды Д. К. Зеленина на «ляшские» особенности некоторых русских диалектов, Н. И. Толстой и С. М. Толстая справедливо отмечают: «Возможно, что некоторые языковые черты на восточнославянской территории, воспринимавшиеся как западнославянские („ляшские“), являются просто общеславянскими ар-

хаизмами» [Толстой – Толстая, 1979, стр. 82]. Новгородская земля, как известно, наряду с Припятским Полесьем, Псковщиной, русским Севером и др., включается в число архаических зон славянского языкового мира (см.: [Толстой, 1977, стр. 55]).

С утверждениями об архаичности древненовгородских языковых фактов в общеславянском контексте, как нам кажется, не вступают в противоречие соображения, касающиеся балтийской ретроспективы (см.: [Топоров, 1991]) и представляющие еще один важный аспект древненовгородской проблемы.

Глава 9

Сопоставление с результатами глоттохронологии

...это все же была стыдливость в самом истинном смысле; притом, мадам, не в отношении слов, ибо, к несчастью, он располагал крайне ограниченным их запасом, — но в делах...

*Стерн. «Жизнь и мнения
Тристрама Шенди, джентльмена»*

1. Глоттохронологическая модель: численные данные

Сравнение результатов лексикостатистического анализа праславянского словаря с результатами глоттохронологического исследования славянских языков, входящее в задачи настоящей работы, требует приведения данных глоттохронологии в эксплицитном и полном представлении. Воспользуемся итогами глоттохронологических наблюдений в уже неоднократно упоминавшейся статье М. Чейки и А. Лампрехта [Чейка — Лампрехт, 1963].

В их статье сообщается четыре ряда цифр: процент совпадения слов в каждой паре языков (c); время их дивергентного развития (в тысячелетиях, t); дата начала дивергенции; d ips — условные единицы генетического расстояния между двумя данными языками, d (единица расстояния введена С. Гуджинской и А. Родригесом)¹. Числа последних трех рядов — приблизительные, то есть с определенным «довеском» после знака «плюс-минус». Мы воспроизведем эти данные не целиком, а с исклю-

¹ Для нас осталось неясным, почему Чейка и Лампрехт нашли единицу d ip «úspěšným rokem», который способен устранить некоторые из недостатков глоттохронологии: при определении этой единицы $d = 14t$ производится просто-напросто механический коэффициентный перевод некоторых единиц (t) в другой численный масштаб — и не более.

чением цифр, сообразных с конечной целью глоттохронологических исследований, то есть только процент совпадения слов у языков данной пары и математические расстояния между ними (dips, без доверительного «довеска»).

$c = 70\%$ ($d = 17,1$): укр. & словен.,

$c = 71\%$ ($d = 16,4$): н.-луж. & болг.,

$c = 72\%$ ($d = 15,7$): русск. & сербохорв.,

$c = 73\%$ ($d = 15,1$): в.-луж. & болг., укр. & болг.,

$c = 74\%$ ($d = 14,4$): чешск. & болг., словац. & болг., польск. & болг., в.-луж. & русск., в.-луж. & укр., русск. & болг., русск. & словен., укр. & сербохорв.,

$c = 75\%$ ($d = 13,8$): чешск. & укр., н.-луж. & русск., н.-луж. & укр., н.-луж. & сербохорв., белор. & словен.,

$c = 76\%$ ($d = 13,1$): чешск. & русск., словац. & русск.,

$c = 77\%$ ($d = 12,5$): польск. & сербохорв., в.-луж. & сербохорв., н.-луж. & словен., словен. & болг.,

$c = 78\%$ ($d = 11,9$): в.-луж. & белор., словац. & укр., русск. & польск., укр. & польск., белор. & сербохорв., белор. & болг., в.-луж. & словен.,

$c = 79\%$ ($d = 11,3$): чешск. & белор., н.-луж. & белор., словац. & словен., польск. & словен.,

$c = 80\%$ ($d = 10,7$): чешск. & сербохорв., словац. & сербохорв.,

$c = 82\%$ ($d = 9,5$): чешск. & словен., словац. & белор., в.-луж. & польск., белор. & польск., сербохорв. & болг.,

$c = 83\%$ ($d = 8,9$): н.-луж. & польск.,

$c = 84\%$ ($d = 8,3$): чешск. & польск., словац. & польск.,

$c = 86\%$ ($d = 7,2$): русск. & укр., русск. & белор., сербохорв. & словен.,

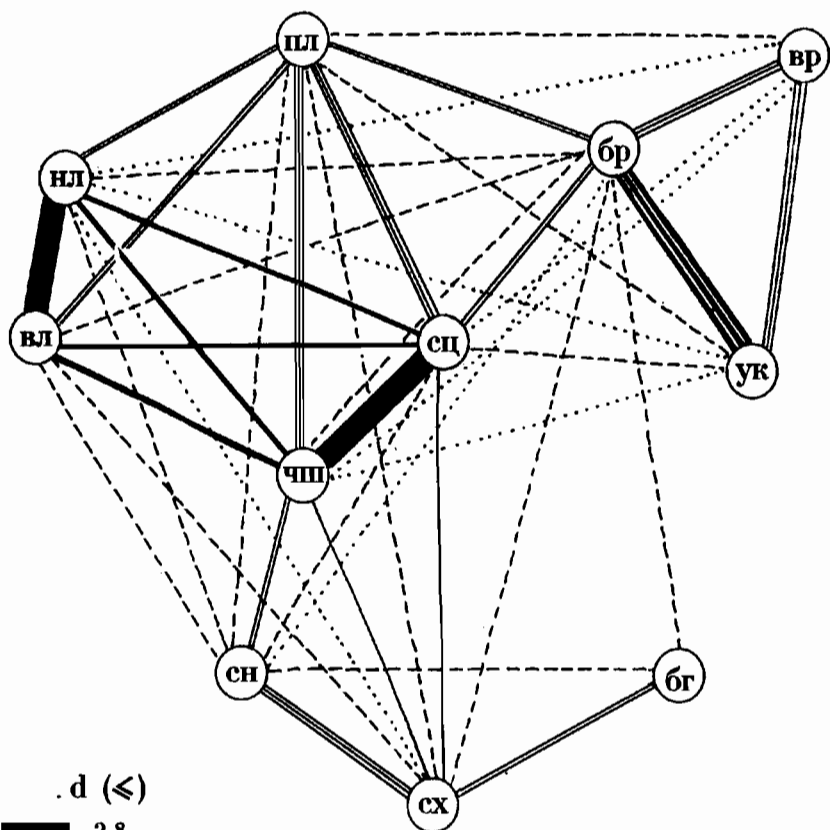
$c = 87\%$ ($d = 6,7$): чешск. & н.-луж., словац. & н.-луж.,

$c = 88\%$ ($d = 6,1$): чешск. & в.-луж., словац. & в.-луж.,

$c = 92\%$ ($d = 4,0$): укр. & белор.,

$c = 95\%$ ($d = 2,5$): чешск. & словац., н.-луж. & в.-луж. [Чейка — Лампрехт, 1963, стр. 9–10].

Схематически картина отношений между одиннадцатью славянскими языками, как они установлены глоттохронологическим исследованием Чейки и Лампрехта, выглядит следующим образом:



$d (\leq)$

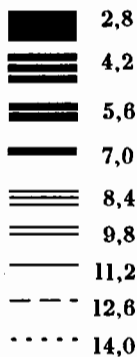


Схема 7

Генетические связи (dips)
 между славянскими языками
 по данным глоттохронологии
 (по Чейке – Лампрехту)

2. Сходства и различия

Имея в виду, что в исследовании М. Чейки и А. Лампрехта глоттохронологическому анализу славянские языки были подвергнуты в неполном составе, целесообразно сравнивать с их данными результаты нашего лексикостатистического исследования в версии с наименьшим (14 элементов) списком идиомов (версия MIN).

Мы не станем здесь развивать кригические взгляды на сам метод глоттохронологии, не только на проблематичность диахронических ее постулатов (опора на понятие родословного древа и представление языковой эволюции как исключительно дивергентных процессов), но и на некоторые ее слабости в освещении синхронической картины взаимоотношений между родственными языками (недостаточная дифференцирующая потенция стословного теста, при которой значительно нивелируются различия в силе связей между разными парами языков). Об этом мы писали в главе 2, к которой и отсылаем ради полноты сопоставления.

Обе полученные картины, и наша, и глоттохронологическая (вне диахронических моментов), обнаруживают как большое сходство, что не удивительно, так и достаточно серьезные расхождения, что тоже не является неожиданностью.

К моментам сходства можно отнести довольно хорошо просматривающуюся отчлененность трех больших групп (южно-, западно-, восточнославянской) друг от друга. И в той и в другой картине максимальной величиной отличаются статистические связи между двумя серболужицкими языками (минимальные связи труднее сравнивать, поскольку схема Чейки – Лампрехта не учитывает, в отличие от нашей, македонский, полабский и кашубско-словинский языки). За исключениями, которые следует отметить, когда речь будет идти о различиях, количественная иерархия связей, их соотношения в грубом измерении «сильнее – слабее» предстают во многом близкими (ср.: в обеих полученных картинах сербохорватско-словенские связи сильнее болгарско-словенских, украинско-белорусские – сильнее связей этих языков с великорусским, словацко-украинские сильнее чешско-украинских и т. п.).

Интереснее расхождение.

Прежде всего бросается в глаза разительное несходство в связях двух подгрупп западнославянской ветви – серболужицкой и чешско-словацкой. В отличие от положения, которое рисуется нашим лексикостатистическим анализом, связи чешского и словацкого в глоттохронологической схеме по своей силе принадлежат, наряду со связями между верхнелужицким и

нижнелужицким языками, к максимальным. К тому же сама серболужицкая подгруппа чрезвычайно близка чешско-словацкой. Обе эти подгруппы, по исследованию Чейки – Лампрехта, образуют единство, превосходящее своей спаянностью любые парные, а тем паче более крупные по числу компонентов, объединения языков, кроме пары «украинский & белорусский». В нашей же схеме связи этих подгрупп относятся к средним по силе (парные связи чешского и словацкого с лужицкими несколько больше единицы в нормированной шкале, см. таблицу 8).

Высокие попарные связи языков этих подгрупп с польским в схеме по Чейке – Лампрехту делают западнославянскую группу весьма компактной (правда, уступающей в этом отношении восточнославянской группе). Однако следует помнить, что Чейка и Лампрехт не включили в свое исследование полабского и кашубско-словинского. Окажись полабский и поморский в сфере анализа чехословацких ученых, общий «портрет» западнославянской группы мог бы оказаться иным, и в значительной степени. Но даже при вычлениении этих двух последних из нашей схемы (в версии MIN, как, впрочем, и во всех других версиях), группа из оставшихся пяти языков не достигает меры компактности, получающейся при применении глоттохронологической методы.

В противоположность тому, что было сказано о западнославянской группе, как она предстает у Чейки – Лампрехта, южнославянская группа выглядит в их данных весьма неплотным образованием. Словенско-болгарская статистическая связь в глоттохронологической схеме сильно отличается от других внутрiryжнославянских связей и характеризуется той же мощностью, что и словенско-нижнелужицкие статистические отношения. По своей силе эта связь языков одной и той же южнославянской группы уступает шестнадцати (!) парным связям между языками, принадлежащими разным группам. В нашей же схеме болгарско-словенская связь обнаруживается как довольно высокая, ее по силе не превосходит ни одна связь между языками разных групп. Компактность южнославянской группы, определяемая уровнем самой слабой парной связи внутри нее, в наших данных выше аналогичного показателя для западнославянской группы.

Иначе, по сравнению с результатами глоттохронологического анализа Чейки – Лампрехта, выглядит в нашем исследовании на фоне других групп и восточнославянская группа. Будучи у нас немногим более компактной, чем южнославянская, у Чейки – Лампрехта восточнославянская языковая группа предстает массивным монолитом, где русско-украинские связи одинаковы по мощности с русско-белорусскими (отметим,

что по нашим статистическим оценкам белорусский язык находится на меньшем расстоянии от великорусского, чем украинский).

Хотя, как мы сказали, все три современные группы славянских языков достаточно хорошо отделяются друг от друга в обеих схемах, глоттохронологическая модель рисует их отношения более тесными. К сильным межгрупповым связям в ней относятся статистические связи словенского с чешским, польского с белорусским и словацкого с белорусским (такой же силы, как внутригрупповые польско-верхнелужицкая и чуть слабее польско-нижнелужицкой). В нашей схеме расстояния между компонентами современной славянской трихотомии оказываются большими.

В целом, нетрудно заметить по схемам, у Чейки и Лампрехта южнославянская и восточнославянская группы отстоят друг от друга дальше, чем у нас. В цифрах это выражается следующим образом.

По нашим данным, из трех славянских языковых групп ближе прочих друг к другу находятся южно- и восточнославянская, средний индекс генетической близости между ними (в версии MIN) составляет $\bar{G} = 0,9338$. За ними следует пара «западнославянская группа & восточнославянская группа», отношения между которыми выражаются средним индексом генетической близости $\bar{G} = 0,9195$. Наконец, связи между южной и западной группами славянских языков выражаются усредненным показателем $\bar{G} = 0,8958$.

По Чейке – Лампрехту (пересчет наш), наименьшее расстояние между тремя группами отмечается у западославянских и восточнославянских (в среднем $\bar{d} = 12,4$), наибольшее расстояние – между южнославянскими и восточнославянскими (усредненное $\bar{d} = 14,3$), отношения между южной и западной ветвями – промежуточные между этими обозначенными пределами (среднее межгрупповое $\bar{d} = 12,8$).

Впрочем, пересчет средних в наших данных, с исключением, для уравнивания условий сопоставления, отсутствующих у Чейки и Лампрехта македонского, полабского и кашубско-словинского, приводит к перемене общего пейзажа: на первом месте оказываются связи, как и у чехословацких ученых, западославянских языков с восточнославянскими (среднее $\bar{G} = 0,9661$), далее следуют лексикостатистические связи южной и восточной групп (среднее $\bar{G} = 0,9517$), южнославянско-западославянские же отношения остаются, как и прежде, на последнем статистическом месте (среднее $\bar{G} = 0,9388$).

При более детальном сопоставлении межгрупповых лексикостатистических отношений по данным глоттохронологии и нашим оценкам выявляются расхождения в «предпочтениях» инославянских связей, демон-

стрируемых отдельными языками (наши сведения на этот счет представлены в таблицах 9, 9а и 10). Расчеты, сделанные на основании данных Чейки — Лампрехта, показывают, что западнославянские связи сербохорватского (в среднем 77,8% совпадающих слов в стословном списке Сводеша) и словенского (в среднем — 79%) преобладают над восточнославянскими (в среднем — соответственно 74,7% и 73,0%). Русский и украинский языки, по этим расчетам, также ориентированы скорее на западнославянские языки (соответственно в среднем 75,8% и 76,0%), чем на южнославянские (средний процент совпадений соответственно 73,3% и 72,3%). Наши же сведения, касающиеся внешних статистических ориентаций этих четырех языков, рисуют их, пусть иной раз с незначительным перевесом, прямо противоположным образом.

Данные М. Чейки и А. Лампрехта, как видим, находятся в некотором несогласии с устойчивыми представлениями о большей близости южной и восточной ветвей славянских языков, которую часто объясняют первоначальным диалектным расколом праславянского языка на западную и восточную области, последняя из которых развилась далее в современные южнославянскую и восточнославянскую языковые группы.

Вполне, однако, может оказаться, что не согласующаяся с традиционными представлениями (но стоящая, таким образом, ближе к представлениям Ф. Мареша и А. Фурдаля о большей связанности севернославянских — западных и восточных — языков) картина внешних «предпочтений» отдельных членов славянской языковой семьи за пределами собственной группы, воссоздаваемая на основании сведений Чейки — Лампрехта, обусловлена упомянутым отсутствием в числе обследованных ими идиомов трех языков, два из которых традиционно входят в лехитскую подгруппу западнославянской ветви.

Внешние ориентации западнославянских языков, по Чейке и Лампрехту, совпадают с тем, что получено нашим лексикостатистическим анализом.

Учитывая склонность приверженцев глоттохронологического метода выводить картину праязыковых диалектных отношений прямо из выявленных подсчетами лексических сходжений между языками (притом на крайне ограниченном материале), игнорируя проблему конвергенции как фактора языковой эволюции, к выводам М. Чейки и А. Лампрехта, на наш взгляд, следует относиться с большой осторожностью.

Глава 10

Сопоставление с результатами статистического определения родства на материале сравнительной фонетики

Математики Тлёна утверждают, что сам процесс счета изменяет количество и превращает его из неопределенного в определенное. Тот факт, что несколько индивидуумов, подсчитывая одно и то же количество, приходят к одинаковому результату, представляет для психологов пример ассоциации или хорошего упражнения памяти.

Борхес. «Тлён, Укбар, Orbis tertius»

1. Две фоностатистические модели.

Численные данные

Попытки установить меру родства между отдельными славянскими языками на основе статистического препарирования данных исторической фонетики предпринимались неоднократно. Инициатива в этом отношении, как отмечалось, принадлежала Я. Чекановскому, но он объединял фонетические явления с морфологическими изоглоссами.

Мы имеем возможность сравнить наши результаты с итогами статистического анализа специально фонетических изоглосс в кругу славянских языков, осуществленного в двух работах.

а. Модель В. И. Перебейнос

Статья В. И. Перебейнос «Использование статистических методов в типологических исследованиях (на материале славянских и германских языков)» представляет собою текст доклада к X Международному съезду лингвистов [Перебейнос, 1967].

С самого начала заявим о своем резко критическом отношении к данной работе. В. И. Перебейнос привлекает к сравнению только девять славянских языков – по три от каждой из современных групп. Из рассмотрения и сравнения с другими идиомами исключены македонский, верхне- и нижне-

лужицкий, полабский и кашубско-словинский. При неполном списке анализируемых языков не может быть восстановлена цельная картина взаимоотношений между языковыми группировками внутри славянской семьи, и материалы статьи приобретают характер не более чем иллюстрации к излагаемой методике (это ощущение усиливается действительно чисто иллюстративным характером данных, касающихся германских языков, из которых автор затрагивает только отношений между английским, немецким, голландским и шведским).

Автор избегает пунктуального перечисления фонетических изоглосс, обсчет которых послужил базой статистических оценок близости между языками. Указывается, что к анализу привлечены фонетические преобразования, список которых составляет 180 «дифференциальных признаков» («рефлексы индоевропейских звуков и звукосочетаний (? фонем и их сочетаний. — А. Ж.), развитие ударения, изменения в построении слога и т. п.» [Перебийнос, 1967, стр. 231]). Показательно, что из приводимых в виде примера трех «дифференциальных признаков» (1. **ǣ* > **o*: *соль*, *ось*; 12. **ĕ* > **o*: *мёд*, *нёс*; 32. **x*' > **š*: *тишина*) два, собственно говоря, дифференциальными признаками не являются, относясь к общеславянским фонетическим явлениям, рефлексы которых не выявляют внутри славянского языкового пространства праязыковой диалектной расчлененности. Невозможность оценить факты, взятые в качестве базы исследования, не внушает большого доверия к конечным численным оценкам родства.

В. И. Перебийнос ссылается на формулы коэффициентов ассоциации и сопряженности (Q и Ф, (12) и (13) из приводимых в настоящей работе), однако приводит результаты обработки своего материала лишь по одной из них, не сообщая по какой именно; по конкретным величинам коэффициентов можно заключить, что использовалась формула коэффициента ассоциации. Они выглядят так (данные таблицы см.: [Перебийнос, 1967, стр. 233]):

болгарский & с.-хорв. 0,85, словен. 0,81, чешск. 0,72, словац. 0,72, польск. 0,60, русск. 0,65, укр. 0,62, белор. 0,52;

сербохорватский & словен. 0,94, чешск. 0,73, словац. 0,83, польск. 0,52, русск. 0,61, укр. 0,73, белор. 0,60;

словенский & чешск. 0,77, словац. 0,78, польск. 0,52, русск. 0,51, укр. 0,53, белор. 0,37;

чешский & словац. 0,93, польск. 0,82, русск. 0,37, укр. 0,51, белор. 0,32;

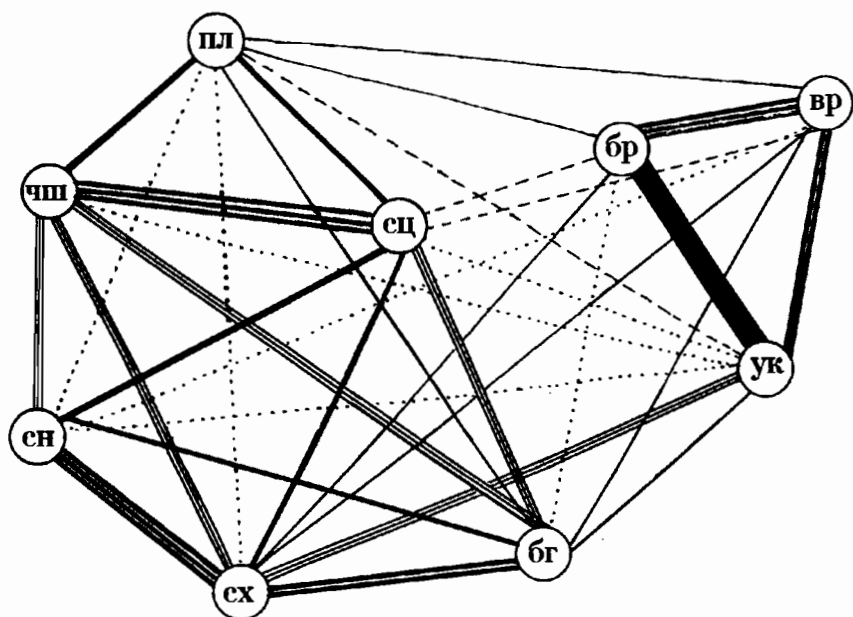
словацкий & польск. 0,80, русск. 0,58, укр. 0,52, белор. 0,59;
 польский & русск. 0,60, укр. 0,54, белор. 0,64;
 русский & укр. 0,89, белор. 0,95;
 украинский & белор. 0,97.

Графическое представление системы родства, статистически определенной В. И. Перебейнос, дает картину, представленную на схеме 8 (см.).

Данные В. И. Перебейнос изображают славянские языки разделенными в первую очередь на две большие группы, одна из которых соответствует современной восточнославянской ветви, а другая объединяет языки южно- и западнославянских ветвей. При большей, по сравнению с чешским и словацким, удаленности польского от южнославянских языков, особенно от сербохорватского и словенского, связи южно- и западнославянской групп рисуются здесь чрезвычайно сильными. Восточнославянская же группа, если не считать неожиданно высокой связи украинского с сербохорватским на фоне остальных внешних связей восточнославянских языков, достаточно далеко отстоит от обеих неюжнославянских группировок.

Усреднение численных показателей родства между славянскими языками по Перебейнос дает такие значения \bar{Q} : южнославянско-западнославянское фонетическое родство — $\bar{Q} = 0,69$, южнославянско-восточнославянское родство — $\bar{Q} = 0,57$, западнославянско-восточнославянское родство — $\bar{Q} = 0,52$ (то есть усредненный коэффициент родства между южно- и западнославянскими языками, по Перебейнос, превосходит $\bar{Q} = 0,67$ для всей славянской семьи в целом).

Взаимоотношения между тремя группами славянских языков в том виде, в каком их представляют статистические расчеты В. И. Перебейнос, вступают в противоречие с господствующими концепциями межгрупповых славянских отношений — с концепцией западного и восточного (легшего в основу дальнейших восточно- и южнославянской групп) первоначальных славянских диалектов (ср.: [Шахматов, 1916, стр. 42–43; Бернштейн, 1961, стр. 68–69; и др.]), с одной стороны, и с концепцией первоначального деления праславянского языка на южный и северный (впоследствии распавшийся на западно- и восточнославянскую группы) диалекты (ср.: [Фурдаль, 1961; Мареш, 1969]), с другой. В этом отношении они скорее приближаются к первоначальному диалектному членению праславянского языка по воззрениям Ежи Налепы (см.: [Налепа, 1968]).



$Q (\geq)$



Схема 8

Система славянского языкового родства (в единицах Q) по данным статистического анализа исторической фонетики (по Перебийнос)

Разумеется, картина, возникающая из фоностатистических оценок, предложенных в рассматриваемой работе, отчасти обусловлена неполнотой списка славянских идиомов, привлеченных к сравнению, однако столь же несомненно, что в ней сказался и характер отбора изоглосс. Не имея их полного описания, мы не в состоянии дать более подробный комментарий к полученной картине, а тем более сравнивать ее в деталях с результатами лексикостатистического анализа.

б. Модель В. Н. Чекмана — О. С. Широкова

Более основателен подход к классификации славянских языков на базе статистики фонетических изоглосс, который предложен в статье В. Н. Чекмана и О. С. Широкова «Квантитативный метод Крубера¹ и классификация славянских языков» [Чекман — Широков, 1962]. Статья Чекмана — Широкова — не законченное квантитативное исследование, а скорее четкая и продуманная его программа (конечные индексы родства приведены в этой работе лишь для нескольких пар славянских языков в качестве иллюстрации). Однако весьма внятное изложение метода и исходных данных позволяет довести количественный анализ до конца самому читателю без каких-либо опасений насчет возможных неточностей и разночтений.

Подход В. Н. Чекмана и О. С. Широкова имеет два неоспоримых достоинства, качественно отличающих его от неубедительного подхода, примененного в предыдущей анализированной работе: широкий и дифференцированный список обследуемых идиомов и четкий, относительно короткий список кладущихся в основу их классификации фонетических изоглосс.

В работе Чекмана — Широкова предлагается сравнивать восемнадцать славянских идиомов. Русский язык представлен в этом списке двумя наречиями (северно- и южновеликорусским); болгарский — также двумя диалектными объединениями: восточноболгарским и западноболгарским комплексами; в сербохорватско-словенской подгруппе выделены четыре диалектных группировки: штокавское наречие, чакавское наречие, объединенные кайкавское наречие и южнословенские диалекты и, наконец, каринтийский с некоторыми другими севернословенскими диа-

¹ Крубера в заголовке статьи — А. Л. Крёбер, диграф *oe* в фамилии которого (*Kroeber*), очевидно, при расшифровке конспективных записей авторов был ошибочно воспринят как латинское *u*.

лектами; в лехитской группе рассматриваются южнопольские диалекты, вместе кашубский, словинский и некоторые севернопольские диалекты и, наконец, полабский язык. Без дополнительного членения рассматриваются украинский, белорусский, македонский, чешский, словацкий, верхнелужицкий и нижнелужицкий языки.

Эти идиомы классифицируются на основании 38 фонетических изоглосс, перечисляемых в приблизительном хронологическом порядке с I в. по XIV в. по Р. Х.:

1. а) палатализация $kv, gv > cv, dzv$;
 б) вторая и третья палатализация $x > s$;
 в) $tj, dj > \check{c}, d\check{z}$;
2. а) вторая и третья палатализация;
 б) $tj, dj > c, dz$;
3. $tl, dl > l$;
4. $-jons > -(j)\bar{e}$;
5. $-jons > -(j)\bar{e}$;
6. $tort, tolt > t\bar{o}r\bar{t}, t\bar{o}l\bar{t}$;
7. а) $tort, tolt > t\bar{a}r\bar{t}, t\bar{a}l\bar{t}$;
 б) совпадение рефлексов r и r' слоговых;
8. $ort, olt > \bar{a}r\bar{t}, \bar{a}l\bar{t}$ под циркумфлексом;
9. $telt > tolt$;
10. последовательное проведение закона открытых слогов (включая и все сочетания типа $(t)ort, (t)olt$);
11. $torot, tolot > trot, tlot$;
12. $p^l, b^l, m^l > p', b', m'$ на стыке морфем;
13. $on > \bar{u}$ -носоевое;
14. $on > \bar{q}$;
15. а) $\bar{e}_1, \bar{e}_2 > \hat{e}$;
 б) утрата носовых;
 в) $dz > z$;
16. $\bar{e}_1, \bar{e}_2 > \bar{a}$;
17. $\bar{e} > 'a$;
18. несохранение r -слогового;
19. $\check{c} (, d\check{z}) > h (, \check{h})$;
20. $d\check{z} > j$;
21. $\check{c}, d\check{z} > \check{s}t, \check{z}d$;
22. $s\check{c}, zd\check{z} > \check{s}t, \check{z}d$;

23. заменительное удлинение;

24. $g > \gamma$;

25. $t', d' > c', dz$;

26. аканье;

27. $r > \check{r}$;

28. $\check{c}r > cr$;

29. $-vj, -vj > oj$;

30. $-vj, -vj > i(j)$;

31. $\check{c} > c$;

32. установление неподвижного ударения

(фонетические преобразования, помещенные под одним номером и различающиеся с помощью буквенной индексации, имеют идентичное распределение по исследуемым в этой работе славянским идиомам).

Чекман и Широков, в отличие от Перебийнос, не отнесли к достойным отражения в списке изоглосс, который кладется в основу статистического определения родства в кругу славянских языков, тех явлений, которые носят общеславянский — в буквальном значении — характер; главным образом речь идет о рефлексации индоевропейских фонем в славянском, что, насколько можно судить по упомянутым весьма немногочисленным примерам изоглосс («дифференциальных признаков»), у Перебийнос составляет едва ли не большую часть ее 180-элементного списка. Если бы авторы задались целью сравнить статистические уровни родственных связей внутри славянской семьи с аналогичными показателями для других языковых семей, то такие общеславянские изоглоссы, конечно же, были бы необходимы в базовом материале сравнения: взаимные тяготения языков в границах разных семей могут сильно различаться. Если же, как это и наблюдается у Чекмана и Широкова, сравнение касается составляющих только одной семьи, то непривлечение обильного материала одинаковой по всем языкам фонетической рефлексации выглядит вполне оправданным: использование всего, что дает история позднеиндоевропейской и раннепраславянской фонетических систем, ослабило бы выпуклость статистических противопоставлений одних связей другим (хотя чисто арифметически это легко устранимо).

Указанные фонетические явления распределены по идиомам, согласно Чекману — Широкову, как показано в таблице 17 (дополнительные к нашим двубуквенным сокращениям названий языков и диалектов: «вб» —

Таблица 17

	Вост.-слав.				Южнослав.						Зап.-слав.							
	ср	юр	ук	бр	вб	зб	мк	шт	чк	кк	кр	сц	чш	вл	нл	юп	сп	лб
1	+	+	+	+	+	+	+	+	+	+	+							
2												+	+	+	+	+	+	+
3	+	+	+	+	+	+	+	+	+	+								
4	+	+	+	+								+	+	+	+	+	+	+
5					+	+	+	+	+	+	+							
6	+	+	+	+										+	+	+	+	+
7					+	+	+	+	+	+	+	+	+					
8					+	+	+	+	+	+	+	+						
9	+	+	+	+													+	+
10	+	+	+	+		+		+	+	+	+	+	+			+		
11														+	+	+	+	
12					+	+	+					+	+	+	+	+	+	+
13	+	+	+	+				+	+			+	+	+	+			
14					+	+	+									+	+	+
15	+	+	+	+				+	+	+	+	+	+	+				
16					+											+	+	+
17	+	+	+	+								+	+	+				
18	+	+	+	+	+	+								+	+	+	+	+
19							+	+	+									
20									+	+	+							
21					+	+												
22					+	+		+				+	+					
23			+											+		+	+	
24		+	+	+								+	+	+				
25				+										+	+	+	+	
26		+		+														
27													+			+	+	
28							+	+										
29	+	+																
30			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
31															+			+
32												+	+	+	+	+		

восточноболгарский, «зб» — западноболгарский, «шт» — штокавский, «чк» — чакавский, «кк» — кайкавский и южнословенский, «кр» — каринтийский и некоторые другие севернословенские, «юп» — южнопольский, «сп» — кашубский, словинский и некоторые севернопольские).

Квантитативная обработка фонетических изоглосс осуществлялась Чекманом и Широковым с помощью неоднократно упоминавшейся нами популярной в статистических работах формулы коэффициента сопряженности (13), который здесь интерпретируется как показатель степени родства:

$$R = (ac - bc) / \sqrt{(a + b)(a + c)(b + d)(c + d)}.$$

Разброс значений R (= Φ в символической записи, приводимой у нас) в формуле сопряженности — от -1 до $+1$. Так как в статье Чекмана — Широкова имеются только выборочные сведения о степенях родства между отдельными славянскими языками, мы приведем итоги наших расчетов полностью:

северновеликорусский & юр +0,898, ук +0,789, бр +0,728,
вб -0,101, зб +0,009, мк -0,128, шт +0,248, чк +0,293, кк +0,276,
кр +0,212, сц +0,031, чш +0,031, вл +0,140, нл +0,119, юп -0,212,
сп -0,170, лб -0,041;

южновеликорусский & ук +0,787, бр +0,842, вб -0,185,
зб -0,077, мк -0,203, шт +0,148, чк +0,197, кк +0,191, кр +0,132,
сц +0,042, чш +0,090, вл +0,148, нл +0,031, юп -0,287, сп -0,248,
лб -0,107;

украинский & бр +0,842, вб -0,077, зб +0,035, мк -0,091,
шт +0,255, чк +0,305, кк +0,300, кр +0,244, сц +0,148, чш +0,148,
вл +0,361, нл +0,140, юп -0,083, сп -0,032, лб +0,009;

белорусский & вб -0,119, зб -0,011, мк -0,129, шт +0,206,
чк +0,258, кк +0,259, кр +0,205, сц +0,100, чш +0,100, вл +0,312,
нл +0,215, юп -0,119, сп -0,069, лб -0,024;

восточно-болгарский & зб +0,890, мк +0,666, шт +0,356,
чк +0,293, кк +0,388, кр +0,326, сц -0,077, чш -0,185, вл -0,402,
нл -0,322, юп -0,101, сп -0,059, лб -0,078,;

западно-болгарский & мк +0,672, шт +0,464, чк +0,449,
кк +0,449, кр +0,439, сц +0,031, чш -0,077, вл -0,402, нл -0,322,
юп -0,101, сп -0,170, лб -0,041;

македонский & шт +0,578, чк +0,544, кк +0,484, кр +0,415, сц -0,091, чш -0,203, вл -0,395, нл -0,335, юп -0,278, сп -0,206, лб -0,093;

штокавский & чк +0,841, кк +0,744, кр +0,690, сц +0,255, чш +0,148, вл -0,277, нл -0,100, юп -0,510, сп -0,577, лб -0,497;

чакавский & кк +0,896, кр +0,846, сц +0,197, чш +0,090, вл -0,231, нл -0,143, юп -0,482, сп -0,541 лб -0,427;

кайкавский и южнословенский & кр +0,944, сц +0,191, чш +0,081, вл -0,248, нл -0,170, юп -0,394, сп -0,470, лб -0,367;

каринтийский (севернословенский) & сц +0,241, чш +0,132, вл -0,203, нл -0,128, юп -0,335, сп -0,436, лб -0,338;

словацкий & чш +0,894, вл +0,468, нл +0,356, юп +0,031, сп -0,139, лб +0,009;

чешский & вл +0,468, нл +0,356, юп +0,140, сп -0,053, лб +0,009;

верхнедужицкий & нл +0,789, юп +0,464, сп +0,410, лб +0,243;

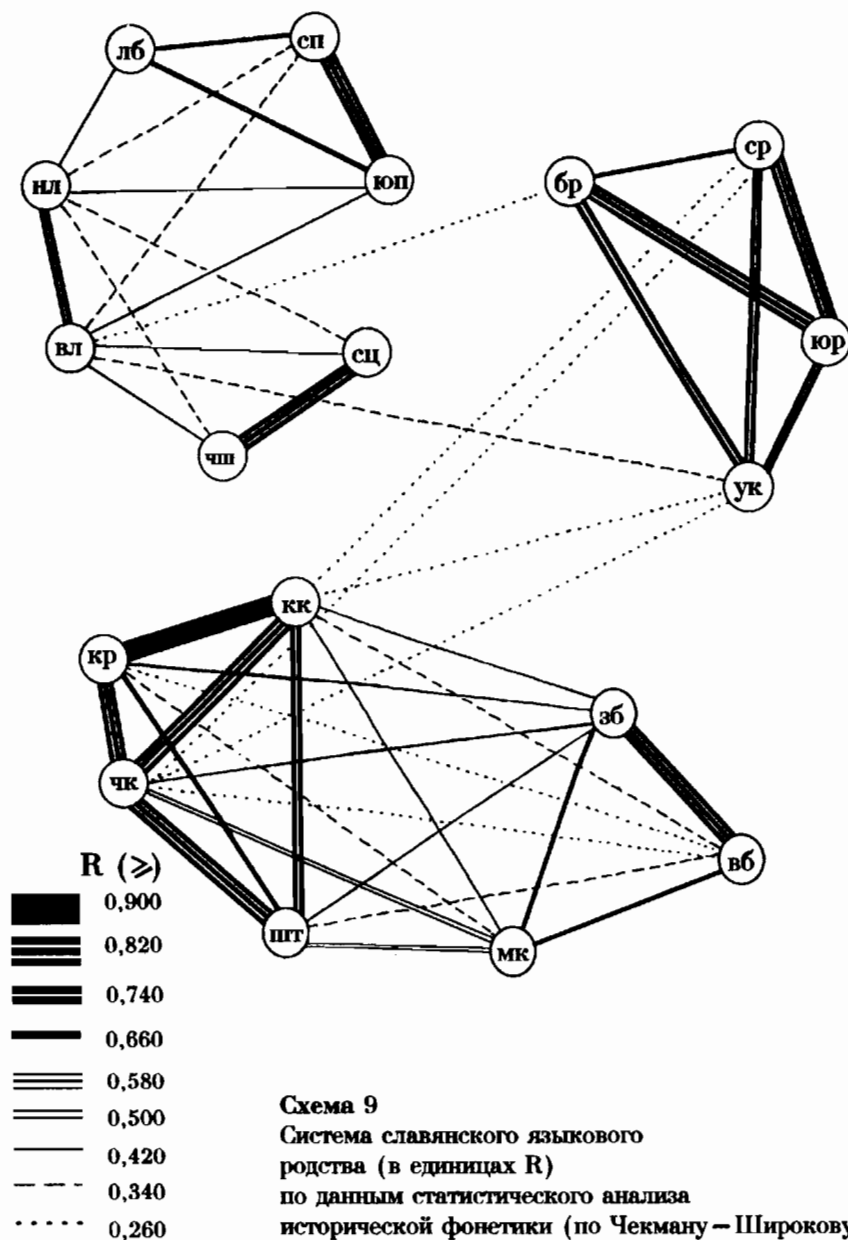
нижнедужицкий & юп +0,449, сп +0,388, лб +0,434;

южнопольский & сп +0,834, лб +0,553;

севернопольский и кашубско-словинский & лб +0,715.

Снова прибегнем к графической экспликации полученных числовых данных (см. схему 9).

В кругу отобранных В. Н. Чекманом и О. С. Широковым идиомов по данным исторической фонетики самая сильная связь, как видим, характеризует диалекты словенского языка: южнословенские (объединяемые с сербохорватской кайкавщиной, исторически тесно связанной со словенским языком) и севернословенские. Весьма высокими показателями генетической близости отмечены далее: северное и южное наречия великорусского языка; чакавский и кайкавский диалекты сербохорватского языка; чешский и словацкий языки; восточная и западная группы болгарских диалектов... Самые слабые статистические связи наблюдаются между группами диалектов сербохорватского языка (без кайкавщины) и идиомами, составляющими лехитскую подгруппу западнославянских языков, особенно у штокавищины и чакавищины в их отношениях с севернопольскими и поморскими говорами.



2. Дискретность фоностаτισической – «вязкость» лексикостаτισической картин родства: возможные причины различий

Л и р

...Знайте: разделили

Мы на три части королевство наше...

Шекспир. «Король Лир»

Модель Чекмана – Широкова представляет славянские языки отчетливо разделенными на группы (и подгруппы) с очень высокими показателями связей внутри подгрупп и очень слабыми показателями связей между идиомами, входящими в разные группы. Главное, что бросается в глаза на схеме, отражающей отношения между славянскими языками по данным фонетики в статистической интерпретации Чекмана – Широкова, – резко выраженная «островная» структура системы родства, с сильным обособлением одних группировок идиомов от других. Этой особенностью результаты статистического определения степеней родства на материале фонетики отличаются от результатов аналогичного анализа лексических изоглосс у нас.

Различия в указанном отношении между двумя сравнивающимися здесь моделями могут быть точнее выражены с помощью цифр.

Мы выделим шесть языковых славянских подгрупп (те, которые выделяли ранее). Нормируем показатели родства по обоим моделям, соотнеся интервал от минимального значения соответствующего показателя ($R_{\min} = -0,577$ у Чекмана – Широкова и $G_{\min} = 0,6516$ в версии CON – наиболее близкой по составу к чекмановско-широковскому набору языков и диалектов – у нас) до его максимального значения ($R_{\max} = +0,944$ и $G_{\max} = 1,8965$) с интервалом от 0 до 1. В этом случае средняя величина внутриподгрупповых коэффициентов по Чекману – Широкову ($\bar{R}_{\text{inter}} = +0,793$) примет значение 0,901, а средняя межподгрупповых коэффициентов родства ($\bar{R}_{\text{exter}} = +0,025$) примет значение 0,396. В нашей лексикостаτισической же модели аналогичные показатели ($\bar{G}_{\text{inter}} = 1,3650$ и $\bar{G}_{\text{exter}} = 0,9396$) будут выражены числами 0,573 и 0,231 соответственно. Разница между нормированными внутриподгрупповыми и межподгрупповыми средними в фоностаτισической модели составит 0,505, в то время как в лексикостаτισической модели перепад между нормированными внутри- и межподгрупповыми средними будет равным всего 0,342, то есть в полтора

раза меньше. Таким образом, членение на группы в фоностатистической модели славянского языкового родства является намного более рельефным.

Ответ на вопрос об источниках различий между показаниями фоностатистической и лексикостатистической моделей может искаться в двух направлениях. Одной из возможных причин выявленного положения могут быть расхождения в моделях, иными словами в системах описания фактов и их квантитативного препарирования, вторую — различия в характере самого материала, то есть фонетических изоглосс в одном случае и лексических в другом.

Можно предполагать, что чисто технические различия, существующие между двумя моделями, которые теоретически способны детерминировать различия в результатах, вряд ли играют здесь сколько-нибудь серьезную роль. К таким техническим различиям мы отнесли бы расхождения в наборе идиомов, устройство формул количественного сравнения, несовпадение шкал, на которых располагаются вычисленные индексы и т. п.

Различия в наборе минимальных таксонов не слишком значительны: в нашей модели более дробно берется русский язык (три диалектных единицы у нас и две у Чекмана и Широкова), в рассматриваемой фоностатистической модели более расчлененно выступает сербохорватско-словенская подгруппа (четыре идиома против двух в нашей модели) и болгарский язык (две диалектные группировки против цельного идиома у нас); остальные десять языков в принципе полностью совпадают в обеих моделях.

Из-за исключительно высоких индексов родства между компонентами сербохорватско-словенской подгруппы в фоностатистической модели шкала показателей оказывается сильно выгнутой в области высоких значений и относительно сжатой в области низких значений. Однако сильное понижение верхнего предела в шкале, с объединением высоких, но далеко отстоящих друг от друга значений индекса родства в фоностатистической модели, то есть преимущественная ориентация шкалы на средние значения индекса, не устраняет эффекта значительной дискретности подгрупп в итоговой схеме родства.

Различия же в действии самой формулы индекса родства (коэффициент сопряженности Φ (R) у Чекмана — Широкова и наш индекс генетической близости G) можно выявить, пересчитав исходные фоностатистические данные по другой формуле. Пересчет материала Чекмана — Широкова по нашей формуле (11), результаты которого мы

здесь не приводим из соображений экономии места, делает связи внутри восточнославянской и западнославянских подгрупп более слабыми по сравнению со связями внутри сербохорватско-словенской подгруппы и — шире — южнославянской группы (или, что равнозначно, усиливает связи внутри южнославянской группы), но также не устраняет заметной обособленности севернославянских подгрупп друг от друга.

Как более серьезные должны быть квалифицированы, по-видимому, те потенциальные причины различий в общем характере сравниваемых картин родства, которые можно усмотреть в отборе изоглосс. Чекман и Широков кладут в основание своей статистической генеалогической классификации славянских языков всего 38 фонетических изоглосс, тогда как В. И. Перебейнос, как мы видели, использует список из 180 изоглосс, и ее конечная картина родственных отношений внутри славянской семьи получается более «вязкой» (межгрупповые индивидуальные связи могут быть в ее модели более сильными, чем внутригрупповые, ср. индексы родства для болгарского и словенского, относящихся к одной большой группе, с одной стороны, и для сербохорватского и словацкого, принадлежащих разным современным группам, с другой). Однако, насколько можно понять, большую, если не большую, часть списка Перебейнос составляют изоглоссы, устанавливающие соотношения между праславянской и позднеиндоевропейской фонетическими системами и, таким образом, не служащие выявлению различий между праславянскими диалектами, то есть по отношению к сопрягаемым языкам скорее интегральные, чем дифференциальные, признаки. Составляя «эфир», в котором растворены дифференциальные признаки в собственном смысле, они, интегральные изоглоссы, легко вычлняются из этого списка и, с учетом возможностей шкалирования, преследующего целью именно выявление статистических сближений и расхождений между анализируемыми идиомами, в принципе картину отношений менять не должны. Расширение списка изоглосс, положенных в основу статистического анализа у Чекмана и Широкова, возможно главным образом за счет явлений, по преимуществу ограниченных ближайше-родственными идиомами (например, $*(j)e- > o-$: $*ezero > *ozero$; $*ju- > u-$: $*jutro > *utro$; $*u > i$: $*l'udъ > *lid(ъ)$ и т. д.). Привлечение частных фонетических изоглосс, во-первых, будет усугублять имеющуюся картину, способствуя статистическому отдалению групп друг от друга, и, во-вторых, при относительно малом объеме материала, потребует разработки процедуры «взвешивания» изоглосс для их разграничения на

более и менее существенные с точки зрения генеалогической классификации².

Некоторую роль в создании эффекта сильной дискретности языковых групп и подгрупп в общей картине родства у Чекмана и Широкова, по-видимому, играет особенность их списка фонетических явлений, которую можно определить как дублирование изоглосс: в коротком, менее четырех десятков позиций, списке фонетических преобразований двойные, противопоставленные рефлексии некоторых сочетаний, находящиеся в отношениях дополнительного географического распределения (например, **-ons* после *j > -ǰ* : *-e* или же **-vj(b) > -oj* : *-ij*) представлены самостоятельными позициями. При использовании статистической формулы, базирующейся на принципе четырехклеточной корреляции, в котором отсутствие общего положительного признака трактуется как наличие общего признака, можно было бы ограничиться предъявлением только одной из двух дополнительно распределенных рефлексий.

Тем не менее мы склонны думать, что главная причина сильной дискретности итоговой картины языкового родства у Чекмана и Широкова по сравнению с нашими результатами кроется не в особенностях метода, примененного исследователями, а в самом материале, то есть в различиях между лексикой и фонетикой и принципиально различных путях их исторического развития, о которых мы писали в главе 1.

Если искать языковым преобразованиям аналогии в геологической эволюции, то изменения в лексике уместно сравнивать с оседанием пород, а фонетические трансформации — с мощными тектоническими процессами. Лексикон, огромный в сравнении с другими языковыми уровнями по числу элементов, рыхлый в своей организации и эволюционирующий на основе кумулятивного принципа, обладает способностью долго сохранять старые диалектные сближения и размежевания, корреспонденции в относительно далеких идиомах и т. д. Отсюда — пестрота лексических связей праязыкового характера, столь многочисленных, что это приводит к известной «вязкости» схемы, опирающейся на эти связи (ср. цитированный выше пассаж О. Н. Трубачева о мозаичности славянских лексических изоглосс, затрудняющей историческую интерпрета-

² О трудности и недостаточной теоретической разработанности вопроса о большей или меньшей классификационной релевантности фонетических явлений свидетельствует хотя бы тот простой факт, что Чекман и Широков исключили из своего списка изоглосс восточнославянский переход **e- > *o-*, в то время как другие исследователи находят его архиважным (см.: [Толстой, 1977, стр. 41]).

цию диалектной картины). В противоположность тому компактная, с жестким сцеплением относительно малочисленных элементов, фонетическая система подвержена лавинным перестройкам, радикально меняющим ландшафт, стирающим следы разнообразных диалектных вариаций и результаты «мелких» диалектообразующих процессов. Относительная немногочисленность фонетических корреспонденций между языками и их тотальность (для демонстрации чуть ли не всех праславянских фонетических явлений достаточно относительно короткого текста-интерлингов, как то может подтвердить любой преподаватель сравнительной грамматики) делают такую же схему чрезвычайно стройной и четко вычленяющей объекты сравнения – праязыковые диалекты и формирующиеся на их основе самостоятельные языки (ср. подобные схемы в работах: [Фурдаль, 1961; Бирнбаум, 1966; Иванов, 1982, стр. 225]). Статистическая интерпретация фонетических корреспонденций между родственными языками приводит к сложению весьма сходной дискретной картины, что мы и имеем на примере графической экспликации квантитативных оценок близости славянских языков по данным сравнительной фонетики в модели Чекмана – Широкова.

3. Детальное сопоставление: несогласованность фонетической и лексической эволюции

Общий сопоставительный взгляд на итоги анализа выявляет как много общего, так и заметные различия. Из расхождений поверхностного свойства отметим, что в лексикостатистической модели наибольшая близость обнаруживается между серболужицкими языками, тогда как в фоностатистической – между чешским и словацким, которые характеризуются «умеренной» лексикостатистической близостью. Из сходств фоностатистической и лексикостатистической картин родства нельзя обойти вниманием низкие показатели внутренней связанности (измеряемой либо усредненными показателями близости внутри данной группировки, либо минимальным значением индекса связи между ее составляющими) языков западнославянской группы по сравнению с двумя другими, а внутри нее – языков, образующих лехитскую подгруппу. Квантитативные оценки и фонетического и лексического родства заставляют констатировать ярко выраженную негомогенность западнославянской группы (на одном уровне) и лехитской подгруппы (на другом). Полученные статистические показатели, на наш взгляд, с очевидностью подтверждают мнение

Н. С. Трубецкого: «„Западнославянский“ оказывается... чисто географическим понятием» [Трубецкой, 1987, стр. 194].

Более детальное сопоставление результатов фоностатистического и лексикостатистического анализа системы славянского языкового родства путем непосредственного столкновения численных показателей связи затрудняется несходством шкал, на которых они располагаются: теоретически от 0 до неопределенной величины в нашей модели (реально – между 0,6593 и 1,9349 в «нормированной» версии SLAV) и от –1 до +1 для корреляционной формулы, которой пользовались Чекман и Широков (реально – от –0,577 до +0,944). Устранение этого неудобства не составляет большой проблемы. Разброс значений и в той и в другой шкалах нетрудно разместить в интервале от 0 до 1 с помощью пересчета по простейшей формуле

$$(G - G_{\min}) / (G_{\max} - G_{\min})$$

или, для фоностатистических данных,

$$(R - R_{\min}) / (R_{\max} - R_{\min}),$$

где max и min – пометы для максимального и минимального значений обоих индексов. Нет необходимости приводить здесь переписанные таким образом численные данные соответствующих их сводов.

Рассчитаем указанным способом не индивидуальные фоностатистические связи между славянскими языками, а средние показателей связей каждого данного языка с целыми подгруппами. Ограничимся лишь внешними связями, не учитывая отношений с ближайшеродственными идиомами, то есть представителями одной и той же из шести постулированных подгрупп. Полученные данные можно соотнести с аналогичными сведениями относительно лексикостатистических связей, извлеченными из таблицы 9 (и 9а) и пересчитанными тем же образом (см. таблицу 18).

Цифры в таблице 18, на наш взгляд, говорят сами за себя. Однако в качестве наиболее существенных итогов сравнения отметим наблюдения, которые позволяют судить о том, что лексическая и фонетическая эволюции позднепраславянских диалектов, приводящие к образованию самостоятельных языков, протекают несогласованно, и идиомы, демонстрирующие значительную близость по данным статистики фонетических изоглосс, вовсе не непременно будут отличаться высокими показателями лексикостатистического сходства.

Таблица 18

	Вост.-слав.		Болг.-макед.		С.-хорв.-словен.	
	phon.	lex.	phon.	lex.	phon.	lex.
ср	—	—	0,298	} 0,391	0,700	} 0,492
юр	—	—	0,361		0,673	
ук	—	—	0,469	0,434	0,778	0,490
бр	—	—	0,428	0,322	0,736	0,368
вб	0,394	} 0,438	—	—	0,841	} 0,900
зб	0,500		—	—	0,947	
мк	0,378	0,327	—	—	1,000	0,923
шт	0,718	} 0,457	0,962	} 0,949	—	—
чк	0,766		0,962		—	—
кк	0,759		0,937		—	—
кр	0,703	0,442	0,892	0,875	—	—
сц	0,586	0,536	0,467	0,474	0,725	0,521
чш	0,600	0,496	0,361	0,442	0,442	0,593
вл	0,743	0,346	0,124	0,333	0,279	0,395
нл	0,633	0,303	0,196	0,240	0,380	0,289
юп	0,342	0,667	0,356	0,349	0,095	0,439
сп	0,385	0,400	0,371	0,271	0,021	0,288
лб	0,471	0	0,493	0,217	0,117	0,093

Так, отвлекаясь от степени значимости конкретных связей отдельных языков в их совокупности, можно заметить, что фонетические связи сильнее лексических у сербохорватского, словенского, обоих лужицких и полабского языков — с восточнославянской группой, у чешского языка и (южно)польских диалектов — с лужицкой подгруппой, у верхнелужицкого и полабского — с чешско-словацкой, у полабского — с болгарско-македонской.

Напротив, лексические связи в их статистическом измерении преобладают над фонетическими в отношении польского и кашубско-словинского языков (южно- и севернопольских диалектов) — с сербохорватско-словенской подгруппой, южнопольских диалектов — с восточнославянскими языками, а также, в меньшей мере, в отношении верхнелужицкого — с болгарско-македонской подгруппой и лехитских языков (кроме, как было отмечено, полабского!) — с чешско-словацкой подгруппой.

Продолжение таблицы 18

	Чеш.-словац.		Лужицк.		Лехитск.	
	phon.	lex.	phon.	lex.	phon.	lex.
ср	0,541 } 0,575 }	0,414	0,636 } 0,597 }	0,184	0,373 } 0,304 }	0,235
юр						
ук	0,654	0,608	0,753	0,389	0,460	0,403
бр	0,608	0,524	0,766	0,401	0,442	0,430
вб	0,384 } 0,489 }	0,482	0,161 } 0,161 }	0,262	0,471 } 0,410 }	0,251
зб						
мк	0,369	0,435	0,158	0,311	0,325	0,295
шт	0,706	0,506	0,329	0,241	0	0,227
чк	0,650 } 0,643 }		0,330 }		0,044 }	
кк			0,309		0,114	
кр	0,691	0,605	0,351	0,443	0,153	0,319
сц	—	—	0,910	0,758	0,479	0,595
чш	—	—	0,910	0,586	0,542	0,448
вл	0,964	0,701	—	—	0,871	0,788
нл	0,856	0,642	—	—	0,923	0,966
юп	0,593	0,794	0,353	0,696	—	—
сп	0,418	0,601	0,897	1,000	—	—
лб	0,520	0,171	0,838	0,935	—	—

Если показатели связей отдельных языков с целыми подгруппами ранжировать, как это было сделано с лексикостатистическими данными в таблице 10, то можно установить, что в отношениях болгарского с лехитской подгруппой, словенского — с восточнославянской, словацкого — с сербохорватско-словенской, (южно)польского — с болгарско-македонской, северно- и южнорусского — с лужицкой, белорусского — с сербохорватско-словенской и лужицкой более значимо фонетическое сходство, тогда как лексическое сходство играет большую роль в отношениях обеих групп болгарских диалектов — с чешско-словацкой подгруппой, словацкого языка — с лехитской подгруппой, (южно)польских диалектов — с чешско-словацкой подгруппой, северновеликорусского наречия — с болгарско-македонской подгруппой, украинского и белорусского языков — с чешско-словацкой и лехитской подгруппами.

Усреднение фоностатистических индексов, очерчивающих внешние связи отдельных языков, показывает большую близость между восточ-

нославянской и южнославянской группами, что совпадает с результатами аналогичных измерений, сделанных на лексикостатистической базе:

Таблица 19

(Пояснения: (а) — усредненные данные по модели Чекмана — Широкова, (б) — пересчет тех же фоностатистических данных по нашей формуле (II)).

	(а)	(б)
южнослав. & восточнослав.	+0,095	0,4827
западнослав. & восточнослав.	+0,028	0,4060
южнослав. & западнослав.	—0,190	0,2646

В целом бóльшая близость восточно- и южнославянских языков не дает еще, однако, весомых оснований судить о наличии в «начале» диалектного раскола праславянского языка единого восточного диалекта, разделившегося впоследствии на восточнославянскую и южнославянскую ветви и противопоставленного западному, давшему затем чешско-словацкую, лужицкую и лехитскую подгруппы современных языков. Признанию монолитности или хотя бы выделенности раннего «восточного» диалекта праславянского языка препятствует неравновесность южнославянских статистических связей восточнославянских языков. Их связи с сербохорватским и словенским гораздо мощнее, чем с болгарским и македонским. Кроме того, болгарско-македонские связи восточнославянских языков уступают и чешско-словацким (а фоностатистические — также и лужицким).

Следовательно, усреднение межгрупповых статистических показателей с целью установления относительной близости между такими крупными таксонами в границах славянской языковой семьи, как три современные ее ветви, — операция, которая подвергает реальные отношения в кругу родственных языков сильной их нивелировке, и потому малооправданная.

Представления о первоначальном западно-восточном — «вертикальном» — расчленении праславянского языка (как, впрочем, и о «горизонтальном» — северно-южном) должны уступить место другим, более сложным, многомерным, но вместе с тем и более адекватным исторической реальности.

Глава II

К реконструкции диалектного членения позднепраславянского языка

Оперирование в предыдущем анализе исключительно праславянской лексикой, сохраняющейся словарными составами современных славянских языков, заставляет расценивать отношения близости между последними, реконструируемые на этой основе, как непосредственное продолжение системы отношений между диалектами праславянского языка.

Нам нельзя обольщаться относительно возможностей проникновения выбранным путем в значительные хронологические глубины. Словарь, который служил нам источником праславянской лексики, — ЭССЯ — ориентирован на эпоху «конца существования праславянского языка» [Трубачев, 1963а, стр. 13]. Это находит свое выражение прежде всего во внешних моментах — «в применяемой фономорфологической реконструкции, в транскрипции» [там же, стр. 13–14]. Не столь очевидное, но для нас наиболее важное выражение указанной хронологической приуроченности словаря-источника — значительное преобладание в его словнике производной лексики, сформированной главным образом уже к концу праславянского периода, лексики, демонстрирующей развитую систему прозрачных аффиксальных средств, принадлежность большинства ее составляющих к четким, для того периода живым и продуктивным деривационным моделям и т. д.

На первый взгляд, в нашем распоряжении имеются возможности хронологического углубления итоговой картины родства. Они лежат на пути отбора к статистическому анализу лексики, которая могла бы относиться к более ранним эпохам в развитии праславянского языка, то есть в операциях, аналогичных технологическим процессам обогащения полезных ископаемых. Это могли быть словарные статьи, в этимологических разделах которых заголовочной праформе ставятся в соответствие внеславянские, индоевропейские цельнолексемные (реже) или корневые (чаще) параллели. Из той праславянской лексики, для которой индоевропейских корреспондентов в словаре не находится, к подсчету могли быть взяты слова, не членимые или плохо членимые на славянской почве.

Однако подобного рода селекция резко сузила бы объем привлеченного к статистическому анализу лексического материала. Следствием этого стала бы существенно менее высокая достоверность результатов. Исходя из этих соображений мы отказались от попыток «обогащения» древней лексики, ограничившись более надежным позднеславянским «срезом».

Говоря о том, что отношения близости между славянскими языками, восстанавливаемые на основании статистического анализа исключительно праславянской лексики, продолжают систему отношений между позднеславянскими диалектами, мы должны, однако, отдавать себе отчет в достаточной условности таких суждений. Линейный принцип соответствия между современными идиомами — группами и подгруппами языков, самостоятельными языками, крупными диалектными массивами типа северновеликорусского наречия, юго-западных украинских говоров («галицкого» наречия) или сербохорватской штокавщины — и протоструктурами в праславянском, когда современному украинскому языку соответствует некий относительно цельный «протоукраинский» компонент, современному словацкому — «протословацкий» и т. д., не вполне точно отражает историческую реальность (весьма определенно об этом — [Осипова, 1994]) и поэтому требует замены качественно иными представлениями.

Если классификации языков и картины диалектного членения обычно ставят своей целью создание иерархии непересекающихся таксонов (которая легко переписывается в древовидную схему с последовательным ветвлением), то при опоре на статистические данные о связях между языками становится возможным построение схем родства с отказом от непеременимой ориентации на дендроидные структуры, то есть с частичным н а л о ж е н и е м различных образующихся идиомов («квазидиалектов» праязыка) друг на друга. Упомянутые «квазидиалекты» вычлениваются при построении таких схем не с опорой на перечисление качеств

венных признаков – диалектообразующих изоглосс (хотя какое-то, как правило небольшое, количество изолекс и может быть приписано данной «квазидиалектной» единице) или на их сгущения (пучки), а на основании учета сил взаимного статистического притяжения поздних родственных идиомов, входящих в данный конституируемый круг.

В традиционных классификациях попарные отношения между идиомами, находящимися на одной и той же ступени последовательного членения, но принадлежащими разным ветвям классификационного древа (например, западнославянско-восточнославянские и западнославянско-южнославянские связи; лехитско-сербохорватскословенские и лужицко-болгарскомакедонские связи; кашубско-болгарские и нижнелужицко-белорусские), как бы молчаливо признаются эквивалентными друг другу. Это уравнивание связей, как кажется, не является адекватным действительной картине отношений внутри славянского языкового мира. Системе родственных идиомов искусственно навязывается не сказать априорная, безосновательная, но все же очень негибкая и неудобная иерархия таксономических единиц: «первый уровень членения» (языковые группы), «второй уровень членения» (подгруппы), «третий» (отдельные языки), «четвертый» (наречия, диалектные группы), «пятый» и т. д.

Устройство предлагаемого же варианта схемы славянского языкового родства, который, строго говоря, не относится к классификациям, как они обычно понимаются, может показать сравнительную статистическую значимость специальных «далеких» межславянских корреспонденций, несовпадающие силы этих связей для идиомов, занимающих одни и те же ступени в традиционной неквазитативной иерархической схеме. Один и тот же язык в предлагаемом варианте схемы родства одновременно включается в группировки разного состава и разной компактности (которая измеряется силой минимальной статистической связи в конкретной группировке), см. схемы 10 и 11. Количество таких вхождений не задается априорной иерархией, а определяется заданной «подробностью» шкалы минимальных сил связей: чем меньше установленный волей исследователя перепад между соседними ступенями шкалы, тем больше возможных группировок различной компактности и тем сложнее и дифференцированнее итоговая картина. В нашей работе выбор исходных для статистического обследования идиомов, в основном совпадающих с современными славянскими языками, определялся, во-первых, стремлением к целостному охвату славянской языковой семьи, во-вторых же – степенью детализованности представления ее состава в наиболее полном сейчас своде славянских изолекс – ЭССЯ).

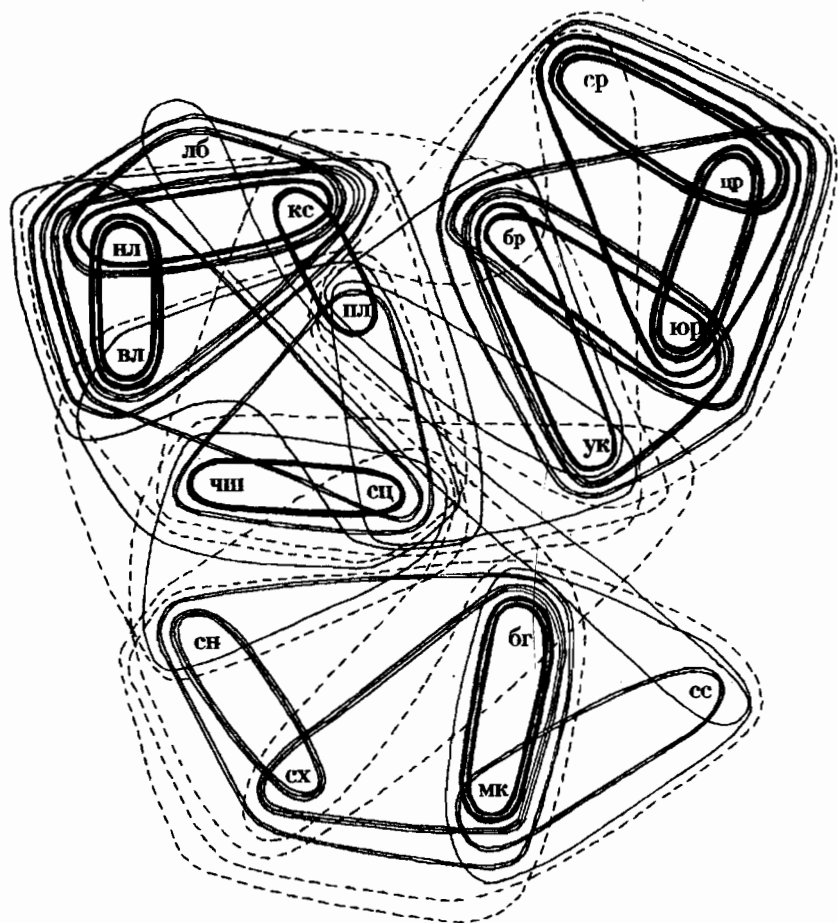


Схема 10

Аппроксимативная схема диалектных объединений
в позднеславянском по данным лексикостатистики
(версия SLAV)

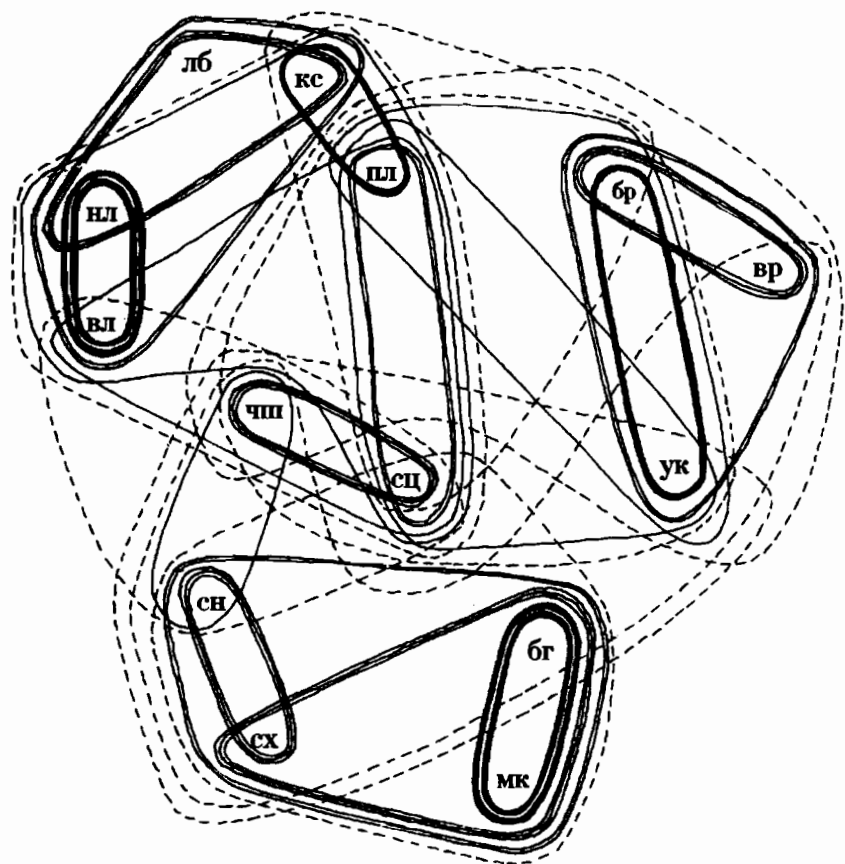


Схема 11

Аппроксимативная схема диалектных объединений
в позднеславянском по данным лексикостатистики
(версия MIN)

Так, например, словацкий язык при установленном пороге образования группировки языков в $G = 1,30$ объединяется только с чешским языком — (а) «сц + чш»; при пороге $G = 1,15$ он одновременно входит в группировки (б) «сц + чш + пл» и (в) «сц + вл»; с понижением порога вхождения до $G = 1,00$ словацкий включается уже в пять группировок с одинаковой мерой компактности: (г) «сц + чш + вл + нл + пл», (д) «сц + вл + нл + пл + кс», (е) «сц + чш + сн», (ж) «сц + чш + пл + ук» и (з) «сц + пл + ук + бр»¹. Таким образом, особенностью получаемой таксономической картины является ее не дискретный, а континуальный характер.

Собственно, именно принципиальной установкой на континуальность картины наш подход и отличается от процедур в рамках количественной таксономии, которые преследуют целью подтверждение или отклонение гипотез о «естественных классах» языков. А. Я. Шайкевич об этом пишет следующее: «Если изучаемые (методами квантитативной таксономии. — А. Ж.) объекты распадутся на такие группы, внутри которых мера сходства будет велика, а между группами — существенно ниже, гипотеза о естественных классах получит весомое подкрепление. При этом мы сможем установить и естественные уровни классификации — те объединения, которые в биологии называются таксонами (ср. класс, отряд, семейство, род, вид). Постепенно понижая порог меры сходства, мы сможем определить критические точки, при которых нарушается стабильность полученных групп и начинается их быстрое объединение. Эти точки и будут соответствовать переходу от одного так-

¹ Выше мы отмечали, что этим группировкам, хотя они постулируются учетом сил взаимного притяжения их компонентов, могут соответствовать некоторые совокупности лексики, описываемой именно этими ареальными характеристиками. Так, для примера, группировке (а) в обследованном фрагменте ЭСЯ отвечают 54 эксклюзивных изолексы: **ašutъnъjъ*, **bel'ati/běi'ati*, **bolgati*, **boligojъ*, **bornъnъ(jb)*, **bridlica*, **brusivo*, **brusičъ*, **byrgъ*, **čadičilo*, **čarati II*, **čarba II*, **čečeriiti*, **čen'ixati*, **čugati*, **čъněi*, **čъrnělъ*, **čъvan'ati*, **druždžii*, **dux(ъ)na*, **dъrstъnъjъ* и т. д.; группировке (б) — 21 изолекса: **askъni*, **astriti*, **berъnъjъ*, **bezpečъnъjъ* (восточнославянские и болгарское слова — заимствования), **bezprēmъ(jb)*, **brъdnъi*, **byčati*, **čevěi*, **čъtmakъ* и т. д.; группировке (ж) — 9 изолекс: **bagnъnъjъ*, **bylkoati*, **bylko*, **čъmъrъ/čъmyrъ*, **dučъkъ*, **klakъ*, **kričati*, **kъjana/kъjanъ/kъjanъ*, **kъzъnъjъ/kъzъno*; группировке (е) — 7 изолекс: **avorъje*, **čabrai (sq)*, **čepъriiti (sq)*, **xaběi*, **xabnъi*, **kupežъ*, **ičivo*; в то же время группировкам (д) и (з) — только по одной изолексе: соответственно **a bo* (восточнославянские *або* — заимствования из западнославянских; в чешском слово не отмечено, что лишний раз и довольно выразительно свидетельствует о сложности отношений в западнославянской группе) и **ičivъ III*. Однако признаком, объединяющим языки в «квазидиалекты» праславянского, является все же не набор исключительных изоглосс, а мера их взаимной близости, определяемой на основании квантитативного анализа всего объема лексических фактов.

сона к другому (например, от лингвистической группы к лингвистической семье)» [Шайкевич, 1980, стр. 325]. В цитированном рассуждении ключевыми для понимания задач, которые ставит перед собой А. Я. Шайкевич, являются выражения «распадутся на... группы», «критические точки», «нарушается стабильность... групп», а также «получит весомое подкрепление». Цель количественной таксономии в таком преломлении — прежде всего поиск в статистической структуре некоторого множества лингвистических объектов подтверждения или опровержения существующих классификаций, которые оперируют дискретными конечными величинами (таксоны). Но это лишь одна из возможных задач квантитативных построений на лингвистическом материале. Далекие частные связи между идиомами рассматриваются при таком подходе именно как частные, главный предмет здесь — непересекающиеся классы объектов. Мы же полагаем, что для сравнительной лингвистики не менее, если не более, важными являются как раз частные связи, которые «работают» не на основной результат диалектной эволюции праязыка — формирование различных более или менее определенных групп языков, а как бы даже препятствуют ему. Таксономический анализ старается отвлечься от «далеких» языковых связей и параллелей, толкуя их скорее как помеху достижению главной цели. Мы же, напротив, считаем, что в этих «помехах» сосредоточен наибольший интерес компаративиста, в них можно найти ключи к пониманию некоторых немаловажных моментов исторических судеб родственных языков.

Существует большой соблазн прямо интерпретировать полученную картину языковых связей в диахроническом плане. При этом как будто можно руководствоваться допущением о том, что наиболее компактные объединения языков, характеризующиеся наибольшей силой статистической связи (естественно, двучленные, вроде болгарско-македонского или северновеликорусско-средневеликорусского) являются единствами самыми тесными в генетическом отношении, и им может быть приписано позднее разделение. Для единств с меньшей компактностью, уже не обязательно двучленных, разумно предположить более значительную удаленность в прошлое, и т. д. Таким образом, ахроническая схема языкового родства типа тех, что воспроизведены у нас в схемах 10 и 11, переписываются в последовательность «срезов», где для получения каждого следующего из них постепенно снижаются пороги меры статистической близости идиомов, с возникновением на каком-то шаге системы пересекающихся кругов («квазидиалектов» праязыка), до тех пор, пока все отображаемые схемой поздние идиомы не сольются в один круг. Эту по-

следовательность «срезов» и можно, на первый взгляд, толковать как ретроспективное изображение диалектного дробления праязыка.

Однако такой прямолинейный взгляд сильно упрощает и обедняет реальную историю языковой семьи. Он рисует диалектную эволюцию праязыка подобной тому, от чего мы пытаемся уйти, то есть напоминающей в сущности древовидные ее модели, которые предполагают, во-первых, бездиалектность праязыка в начальном его состоянии и не учитывают, во-вторых, феноменов вторичного сближения языков, конвергентных процессов. Буквальное восприятие подобной схемы как прямого отображения эволюции праязыка, следовательно, нужно расценить как довольно вульгарное. Тем не менее в качестве некоторой приближенной статистической экспликации дивергентных процессов на лексическом уровне она кажется нам допустимой.

С понижением порога образования группировок («квазидиалектов» праславянского языка при диахронической реинтерпретации) они разрастаются по количеству охватываемых членов (поздних идиомов, от системы которых ведется отсчет) и постепенно утрачивают отчетливость противопоставления друг другу, отличаясь между собою одним-двумя составляющими. Принятие в качестве такого порога самой слабой статистической связи в списке всех возможных пар идиомов выделяет в с ю совокупность рассматриваемых родственных языков и делает возможной трактовку этой совокупности как единого идиома (цельного, однако, только при ограничении лишь внутриславянскими лексикостатистическими связями: чем дальше в глубь славянской языковой истории, чем слабее при этом углублении рисуются связи между постулированными или отобранными к анализу поздними идиомами, тем более значимыми в глазах диахрониста должны становиться связи внешние – балтийские, германские, итальянские, иранские и т. д., опять-таки неравномошные для разных поздних славянских языков, что как нам представляется, уже само по себе должно снимать, хотя бы теоретически, проблему диалектной монолитности праславянского языка; последняя тема, впрочем, по понятным причинам в монографии совершенно не затрагивается).

Разумеется, как и всякая схема, полученная картина является только аппроксимацией действительного положения вещей: чем сильнее статистические связи между языками, тем вероятнее их группировки как отображения протоединств, реконструируемые применительно к поздним периодам развития праславянского языка (с оговорками относительно выявленных случаев конвергенции типа вторичного единства серболужицкой подгруппы), чем они слабее, тем большей условностью становятся получаемые

«квазидialeкты», относимые к более ранним периодам славянской языковой истории. Однако на этом этапе лексикостатистических исследований диахронического плана существенным представляется не столько достижение высокой степени соответствия между кабинетной глоттогенетической конструкцией и исторической реальностью, сколько уточнение самих механизмов языковой дивергенции.

Становление и эволюция диалектов праславянского языка по данным лексики выглядят иными, чем в зеркале сравнительной фонетики: не резкое диалектное размежевание с четкими линиями разломов, а постепенное и медленное «расползание» в стороны с долгим сохранением следов прежней близости. Эффект крутого разлома возникает при принятии в соображение только диахронических изменений на фонетическом уровне со свойственной им «отменой», стиранием предыдущего состояния языковой системы. Преимущественно кумулятивный принцип эволюции словаря предшествующих состояний вовсе не устраняет, но, обнаруживаясь в наложении вновь возникающих диалектообразующих изолиний на систему существующих диалектных размежеваний, тем самым способствует ее усложнению во времени. Диалектная эволюция по данным лексики и лексикостатистики не вписывается в многочисленные простые «первоначальные» дихотомические, трихотомические и тетрахотомические модели диалектообразования в праславянском.

Излишне говорить, что лексикостатистическая картина лингвистического родства не отрицает, но только дополняет представления о славянском глоттогенезе, сложившиеся на основании изучения иных языковых уровней.

Использованная литература

- Акчимский словарь: Словарь говора д. Акчим Красновишерского района Пермской области. (Акчимский словарь). Вып. 1—, Пермь, 1984—.
- Андронов, 1978: *М. С. Андронов*. Сравнительная грамматика дравидийских языков. М., 1978.
- Арапов — Херц, 1974: *М. В. Арапов, М. М. Херц*. Математические методы в исторической лингвистике. М., 1974.
- Ардт, 1961: *W. W. Arndt*. The performance of glottochronology in Germanic // *Language*, vol. 35, 1959.
- Архангельский областной словарь: Архангельский областной словарь. Вып. 1—, М., 1980—.
- Ахманова, 1966: *О. С. Ахманова*. Словарь лингвистических терминов. М., 1966.
- Базилевич — Верецагин, 1970: *Л. И. Базилевич, Е. М. Верецагин*. Славистические проблемы в научном наследии Л. В. Щербы // *Советское славяноведение*, 1970, № 1.
- Бак, 1949: *C. A. Buck*. Dictionary of selected synonyms in the principal Indo-European languages. Chicago, 1949.

- Батожок, 1986: *Н. И. Батожок*. Диалектный словарь как источник лингвогеографического изучения региона // Диалектное слово в лексикографическом аспекте. Межвузовский сборник научных трудов. Л., 1986.
- Белич, 1972: *Я. Белич*. Доисторические диалектные различия в области чешского языка // Русское и славянское языкознание. К 70-летию члена-корреспондента АН СССР Р. И. Аванесова. М., 1972.
- Бенвенист, 1963: *Э. Бенвенист*. Классификация языков // Новое в лингвистике. Вып. III. М., 1963.
- БЕР: Български етимологичен речник. Т. I—III. София, 1971—1986.
- Бернштейн: *С. Б. Бернштейн*. Болгарско-русский словарь. Изд. 2. М., 1975.
- Бернштейн, 1961: *С. Б. Бернштейн*. Очерк сравнительной грамматики славянских языков. М., 1961.
- Бирнбаум, 1966: *Н. Birnbaum*. The dialects of Common Slavic // Ancient Indo-European dialects. Berkeley; Los Angeles, 1966.
- Бирнбаум, 1985: *Х. Бирнбаум*. О двух основных направлениях в языковом развитии // Вопросы языкознания, 1985, № 2.
- Бирнбаум, 1987: *Х. Бирнбаум*. Праславянский язык. Достижения и проблемы в его реконструкции. М., 1987.
- Бирнбаум, 1993: *Х. Бирнбаум*. Генетические и типологические методы внешнего сравнения языков // Вопросы языкознания, 1993, № 4.
- Бодуэн де Куртенэ, 1897: *И. А. Бодуэн де Куртенэ*. Кашубский «язык», кашубский народ и «кашубский вопрос» // Журнал Министерства народного просвещения, 1897, апрель-май.
- Бонфанте, 1946: *G. Bonfante*. Indo-Hittite and areal linguistics // American Journal of Philology. Vol. LXII, 1946, No 4.
- Борысь, 1975: *W. Boryś*. Prefiksacja imienna w językach słowiańskich. Wrocław; Warszawa; Kraków; Gdańsk, 1975.
- БРС: Беларуска-рускі слоўнік. Выд. 2. Т. 1—2. Мінск, 1988—1989.
- Булахов — Жовтобрюх — Кодухов, 1987: *М. Г. Булахов, М. А. Жовтобрюх, В. И. Кодухов*. Восточнославянские языки. М., 1987.
- Вайан, 1952: *А. Вайан*. Руководство по старославянскому языку. М., 1952.

- Варбот, 1976: *Ж. Ж. Варбот*. Вариантность суффиксальной структуры в однокоренных славянских именах и реконструкция праславянского лексического фонда // Вопросы языкознания, 1976, № 6.
- ван дер Варден, 1960: *Б. Л. ван дер Варден*. Математическая статистика. М., 1960.
- Васильев, 1907: *Л. Л. Васильев*. О случае сохранения общеславянской группы *-dl-* в одном из старых наречий русского языка // Русский филологический вестник, 1907, № 4.
- ван Вейк, 1957: *Н. ван Вейк*. История старославянского языка. М., 1957.
- Веселовский Ономастикон: *С. Б. Веселовский*. Ономастикон. Древнерусские имена, прозвища и фамилии. М., 1974.
- Виноградов, 1977: *В. В. Виноградов*. Об изучении общего лексического фонда в структуре славянских языков // Там же.
- Гавацци, 1960: *M. Cavazzi*. Zapadno-panonski slavenski pojas u davnini // Etnografia Polska. Т. III. Wrocław, 1960.
- Гамкрелидзе — Иванов, 1984: *Т. В. Гамкрелидзе, В. В. Иванов*. Индоевропейский язык и индоевропейцы. Реконструкция и историко-типологический анализ языка и протокультуры. Т. I—II. Тбилиси, 1984.
- Геров: *Н. Геровъ*. Речник на българския език. Ч. I—5. Пловдив, 1895—1904. Ч. 6. Допълнение. Събрал, наредил и изтълкувал Т. Панчев. Пловдив, 1908. Фототипно издание. София, 1975—1978.
- Гловинская, 1967: *М. Я. Гловинская*. Фонологическая подсистема редких слов в современном русском литературном языке (на материале заимствований). Автореферат диссертации на соискание ученой степени кандидата филологических наук. М., 1967.
- Гловинская, 1971: *М. Я. Гловинская*. Об одной фонологической подсистеме в современном русском литературном языке // Развитие фонетики современного русского языка. Фонологические подсистемы. М., 1971.
- Головин, 1971: *Б. Н. Головин*. Язык и статистика. М., 1971.
- Горнунг, 1963: *Б. В. Горнунг*. Из предьстории образования общеславянского языкового единства. М., 1963. (Доклады советской делегации. V Международный съезд славистов (София, сентябрь 1963)).

- Госсенс, 1969: *J. Goossens*. Strukturelle Sprachgeographie. Eine Einführung in Methodik und Ergebnisse. Heidelberg, 1969.
- Гуджинская, 1956: *S. C. Gudshinsky*. The ABC's of lexico-statistics (glottochronology) // *Word*. Vol. 12, 1956.
- Даль: *В. И. Даль*. Толковый словарь живого великорусского языка. Изд. 2. Т. I—IV. М.; СПб., 1880—1882 (1978—1980).
- Демчук, 1988: *М. О. Демчук*. Слов'янські автохтонні особові імена в побуті українців XIV—XVII ст. Київ, 1988.
- Дерфер, 1978: *G. Doerfer*. Grundwort und Sprachmischung. Eine Untersuchung an Hand von Körperteilbezeichnungen. Stuttgart, 1978.
- Дерфер, 1981: *Г. Дерфер*. Базисная лексика и алтайская проблема // Вопросы языкознания, 1981, № 4.
- Джаукян, 1972: *Г. Б. Джаукян*. Многопризнаковая статистическая классификация армянских диалектов // Вопросы языкознания, 1972, № 4.
- Дзедзелевский, 1987 (Дзедзєлівський): *Й. О. Дзедзєлівський*. Програма для збирання матеріалів до Лексичного атласу української мови. Вид. 2. Київ, 1987.
- ДРС: Словарь древнерусского языка (XI—XVII вв.). Т. I—. М., 1988—.
- Дурново, 1929: *Н. Н. Дурново*. Несколько замечаний по вопросу об образовании русских языков // Известия ОРЯС, т. II, 1929, кн. 2.
- ESSJ — см.: Копечный, 1973/1980.
- ЕСУМ: Етимологічний словник української мови. Т. I—3. Київ, 1982—1989.
- Живов — Успенский, 1973: *В. М. Живов, Б. А. Успенский*. Центр и периферия в языке в свете языковых универсалий // Вопросы языкознания, 1973, № 5.
- Журавлев, 1984: *А. Ф. Журавлев*. Иноязычные заимствования в русском просторечии (фонетика, морфология, лексическая семантика) // Городское просторечие. Проблемы изучения. М., 1984.
- Журавлев, 1988: *А. Ф. Журавлев*. Лексикостатистическая оценка генетической близости славянских языков // Вопросы языкознания, 1988, № 4.
- Журавлев, 1990: *А. Ф. Журавлев*. К уточнению представлений о славянских изоглоссах. Дополнения к лексическим материалам

«Этимологического словаря славянских языков». Ч. I—II. М., 1990.

- Журавлев, 1990а: А. Ф. Журавлев. Поморский («протокашубскословинский») в кругу позднеславянских диалектов (по данным лексикостатистики) // Поморські слов'яни. Тези конференції до 120-річчя з дня народження М. В. Бречкевича. 25—26 жовтня 1990 р. (Тернопільські славистичні історико-філологічні читання. Рік II). Тернопіль, 1990.
- Журавлев, 1990б: А. Ф. Журавлев. К критике некоторых методов определения древнеславянских племенных ареалов // Закономерности языковой эволюции. Всесоюзная научная конференция. Тезисы докладов. Рига, 1990.
- Журавлев, 1991: А. Ф. Журавлев. К проблеме расселения древних славян. (О так называемом «графоаналитическом методе») // Вопросы языкознания, 1991, № 2.
- Жучкевич, 1974: В. А. Жучкевич. Краткий топонимический словарь Белоруссии. Минск, 1974.
- Закревская, 1974: Я. В. Закревская. Словообразование существительных *potina* іосі (к проблеме ареального изображения) // Общеславянский лингвистический атлас. Материалы и исследования. 1971. М., 1974.
- Зализняк, 1984: А. А. Зализняк. Наблюдения над берестяными грамотами // Вопросы русского языкознания. Вып. V. История русского языка в древнейший период. М., 1984.
- Зализняк, 1986а: А. А. Зализняк. Новгородские берестяные грамоты с лингвистической точки зрения // ВЛЯнин, А. А. Зализняк. Новгородские грамоты на бересте (из раскопок 1977—1983 гг.). Комментарии и словоуказатель к берестяным грамотам (из раскопок 1951—1983 гг.). М., 1986.
- Зализняк, 1986б: А. А. Зализняк. Словоуказатель к берестяным грамотам // Там же.
- Зализняк, 1988а: А. А. Зализняк. Значение новгородских берестяных грамот для истории русского и других славянских языков // Вестник Академии наук СССР, 1988, № 8.
- Зализняк, 1988б: А. А. Зализняк. Древненовгородский диалект и проблемы диалектного членения позднего праславянского языка // Славянское языкознание. X Международный съезд славистов. София, сентябрь 1988 г. Доклады советской делегации. М., 1988.

- Захарова — Орлова, 1970: *К. Ф. Захарова, В. Г. Орлова*. Диалектное членение русского языка. М., 1970.
- Зеленин, 1954: *Д. К. Зеленин*. О происхождении северновеликорусов Великого Новгорода // Доклады и сообщения Института языкознания АН СССР, 1954, вып. 6.
- Зольта, 1960: *G. R. Solta*. Die Stellung des Armenischen im Kreise der indogermanischen Sprachen. Wien, 1960.
- Иванов, 1982: [*Вяч. Вс. Иванов*]. 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.
- Иванов, 1990а: *Вяч. Вс. Иванов*. Генеалогическая классификация языков // Лингвистический энциклопедический словарь. М., 1990.
- Иванов, 1990б: *Вяч. Вс. Иванов*. Глоттохронология // Там же.
- Ильинский, 1917: *Г. А. Ильинский*. Славянские этимологии // Русский филологический вестник. Т. 78, вып. 1/2, 1917.
- Ильинский, 1925: *Г. А. Ильинский*. Кто были Λευξάνῶνι Κοινητοῦ Константина Багрянородного? // *Slavia*. Т. IV, 1925, s. 2.
- Историческая типология, 1986: Историческая типология славянских языков. Фонетика, словообразование, лексика и фразеология. Киев, 1986.
- Йордан, 1971: *Й. Йордан*. Романское языкознание. Историческое развитие, течения, методы. М., 1971.
- Климов, 1986а: *Г. А. Климов*. Введение в кавказское языкознание. М., 1986.
- Климов, 1986б: *Г. А. Климов*. Об ареальной конфигурации протоиндоевропейского в свете данных картвельских языков // Вестник древней истории, 1986, № 3.
- Климов, 1990: *Г. А. Климов*. Основы лингвистической компаративистики. М., 1990.
- Колмогоров, 1988: *А. Н. Колмогоров*. Математика // Математический энциклопедический словарь. М., 1988.
- Коломиец, 1983: *В. Т. Коломиец*. Происхождение общеславянских названий рыб. К IX Международному съезду славистов. Киев, 1983.

- Копечный, 1964: *Základní všeslovanská slovní zásoba*. Brno, 1964.
- Копечный, 1973/1980: *Etymologický slovník slovanských jazyků. Slova gramatická a zájmena*. Sv. 1. Předložky. Koncové partikule. Praha, 1973; Sv. 2. Spojky, částice, zájmena a zájmena adverbia. Praha, 1980.
- Копечный, 1976: *Ф. Копечный*. О новых этимологических словарях славянских языков // *Вопросы языкознания*, 1976, № 1.
- Кребер — Кретьен, 1937: *A. L. Kroeber, C. D. Chrétien*. Quantitative classification of Indo-European languages // *Language*. Vol. 13, 1937, No 2.
- Куза — Медынцева, 1974: *А. В. Куза, А. А. Медынцева*. Заметки о бережтяных грамотах // *Нумизматика и эпиграфика*. Т. XI. М., 1974.
- Кузнецов, 1966: *П. С. Кузнецов*. Восточнославянские языки // *Языки народов СССР*. Т. 1. Индоевропейские языки. М., 1966.
- Курилович, 1958: *Е. Курилович*. О балто-славянском языковом единстве // *Вопросы славянского языкознания*. Вып. 3. М., 1958.
- Куркина, 1972: *Л. В. Куркина*. Словенско-восточнославянские языковые связи // *Этимология*. 1970. М., 1972.
- Куркина, 1973: *Л. В. Куркина*. К сравнительной характеристике лексического состава южнославянских языков // *Кузнецовские чтения*. 1973. История славянских языков и письменности. М., 1973.
- Куркина, 1976: *Л. В. Куркина*. Изоглосные связи южнославянской лексики. (Материалы к проблемам славянского этногенеза) // *Вопросы этногенеза и этнической истории славян и восточных романцев*. М., 1976.
- Куркина, 1987: *Л. В. Куркина*. Диалектная структура праславянского языка по данным южнославянской лексики. Автореферат диссертации на соискание ученой степени доктора филологических наук. М., 1987.
- Лант, 1975: *H. G. Lunt*. On the language of Old Rus: some questions and suggestions // *Russian linguistics*, 1975, No 3/4.
- Ларин, 1960: *Б. А. Ларин*. Историческая диалектология русского языка в курсе лекций акад. А. А. Шахматова и наши задачи // *Ученые записки ЛГУ*, № 267. Серия филологических наук, вып. 52. Л., 1960.

- Левин, 1964: Ю. И. Левин. Об описании системы лингвистических объектов, обладающих общими свойствами // Вопросы языкознания, 1964, № 4.
- Лернер, 1973: К. Б. Лернер. Статистические методы в историческом языкознании. Автореферат диссертации на соискание ученой степени кандидата филологических наук. Тбилиси, 1973.
- Леман, 1991: В. П. Леман. Новое в индоевропейских исследованиях // Вопросы языкознания, 1991, № 5.
- Лер-Сплавинский, 1930: T. *Lehr-Splawiński*. Zagadnienie pokrewieństwa językowego // Biuletyn Polskiego Towarzystwa Językoznawczego, II, 1930.
- Лер-Сплавинский, 1934: T. *Lehr-Splawiński*. O narzeczach słowian nadbałtyckich // Kaszubi, kultura ludowa i język. Toruń, 1934.
- Лер-Сплавинский, 1938: T. *Lehr-Splawiński*. Element prasłowiański w dzisiejszym słownictwie polskim // Studia historyczne ku czci Stanisława Kutrzeby. II. Kraków, 1938.
- Лер-Сплавинский, 1954: T. *Лер-Сплавинский*. Польский язык. М., 1954.
- Лер-Сплавинский, 1956: T. *Lehr-Splawiński*. O dawnych narzeczach słowian Pomorza Zachodniego i ziem przyległych // Konferencja Pomorska (1954). Prace językoznawcze. Warszawa, 1956.
- Лер-Сплавинский — Курашкевич — Славский, 1954: T. *Ler-Splawiński*, W. *Kuraszkiewicz*, F. *Slawski*. Przegląd i charakterystyka języków słowiańskich. Warszawa, 1954.
- Летч, 1965: R. *Lötzsch*. Einheit und Gliederung des Sorbischen. Berlin, 1965.
- Лещак, 1990: О. В. Лещак. Структурно-функциональный статистический анализ «Słownika języka pomorskiego czyli kaszubskiego» С. Рамулта // Поморські слов'яни. Тези конференції до 120-річчя з дня народження М. В. Бречкевича. 25—26 жовтня 1990 р. (Тернопільські славістичні історико-філологічні читання. Рік II). Тернопіль, 1990.
- Лигети, 1971: Л. *Лигети*. Алтайская теория и лексикостатистика // Вопросы языкознания, 1973, № 3.
- Лиз, 1953: R. B. *Lees*. The basis of glottochronology // Language. Vol. 29, 1953.
- Лоренц, 1902: F. *Lorentz*. Das gegenseitige Verhältniss der sogenannten lechischen Sprachen // Archiv für slavische Philologie, Bd. XXIV, 1902.

- Лоренц, 1925: *F. Lorentz. Geschichte der pomoranischen (kaschubischen) Sprache.* Berlin; Leipzig, 1925.
- ЛСС ОЛА, вып. 1 — см.: ОЛА. Животный мир.
- Львов, 1976: *А. С. Львов. Праславянский слой старославянской лексики // Вопросы языкознания.*, 1976, № 2.
- ЛЭС, 1990: Лингвистический энциклопедический словарь. М., 1990.
- Макаев, 1964: *Э. А. Макаев. Проблемы индоевропейской ареальной лингвистики.* М.; Л., 1964.
- Макаев, 1971: *Э. А. Макаев. Проблема индоиранского языкового единства // Вопросы языкознания*, 1971, № 3.
- Мальцева, 1977: *И. М. Мальцева. Лексическая окказиональность в языке XVIII в. // Проблемы исторической лексикографии.* Л., 1977.
- Маньчак, 1958: *W. Mańczak. Problem klasyfikacji genealogicznej języków słowiańskich // Z polskich studiów slawistycznych. Prace językoznawcze i etnogenetyczne na IV. Międzynarodowy Kongres Slawistów w Moskwie 1958.* Warszawa, 1958.
- Маньчак, 1981: *W. Mańczak. Praojczyzna Słowian.* Wrocław, 1981.
- Маньчак, 1983: *W. Mańczak. Zagadnienie wspólnoty bałto-słowiańskiej // Z polskich studiów slawistycznych. Seria 6. Językoznawstwo. Prace na IX. Międzynarodowy Kongres Slawistów w Kyjowie 1983.* Warszawa, 1983.
- Маньчак, 1986: *W. P. Mańczak. Germanic and other Indo-European languages // Linguistics across historical and geographical boundaries. In honour of Jacek Fisiak on the occasion of his 50th birthday.* Berlin etc., 1986 (vol. I. Linguistic theory and historical linguistics).
- Маньчак, 1987: *В. Маньчак. О балто-славянских отношениях // Балто-славянские исследования.* 1985. М., 1987.
- Мареш, 1969: *F. V. Mareš. Diachronische Phonologie des Ur- und Frühslavischen.* Munich, 1969 (= Slavistische Beiträge, Bd. 40).
- Мареш, 1980: *F. V. Mareš. Die Tetrachotomie und doppelte Dichotomie der slavischen Sprachen // Wiener slavistisches Jahrbuch.* Bd. 26. Wien, 1980.
- Марков, 1913: *А. А. Марков. Пример статистического исследования над текстом «Евгения Онегина», иллюстрирующий связь испытаний в цепь // Известия императорской Академии наук.* Серия VI, т. X, 1916, № 4.

- Мартынов, 1965: В. В. Мартынов. Проблема славянского этногенеза и методы лингвогеографического изучения Припятского Полесья // Советское славяноведение, 1965, № 4.
- Махек (Machek): V. Machek. Etymologický slovník jazyka českého. Praha, 1971.
- Мейе, 1938: А. Мейе. Введение в сравнительное изучение индоевропейских языков. М.; Л., 1938.
- Мейе, 1954: А. Мейе. Сравнительный метод в историческом языкознании. М., 1954.
- Мельникова — Сухачев, 1971: Лингвистические и этнографические атласы и карты. Сост.: Т. Н. Мельникова и Н. Л. Сухачев. Л., 1971.
- Меркулова, 1986: В. А. Меркулова. О дополнениях к словарям праславянских словарей // Этимология. 1984. М., 1986.
- Методика и техника статистической обработки информации, 1968: Методика и техника статистической обработки первичной социологической информации. М., 1968.
- Мжельская, 1963: О. С. Мжельская. О лексических связях псковских говоров с западными славянскими языками (слово *скорлуна*) // Вестник Ленинградского университета, 1963, № 14, вып. 3.
- Милевский, 1931: N. Milewski. Zachodnia granica pomorskiego obszaru językowego w wiekach średnich // Slavia Occidentalis, X, 1931
- Мокиенко, 1969а: В. М. Мокиенко. Об одной псковско-западнославянско-литовской изоглоссе (*багно*) // Вопросы теории и истории языка. Сборник статей памяти проф. Б. А. Ларина. Л., 1969.
- Мокиенко, 1969б: В. М. Мокиенко. Ареальный анализ местной географической терминологии и его интерпретация // Советское славяноведение, 1969, № 5.
- Морозов, 1915: Н. А. Морозов. Лингвистические спектры // Известия Императорской Академии наук. Отделение русского языка и словесности, т. XX, 1915, кн. 1—4.
- Москович, 1968: В. А. Москович. Из полесской терминологии цветообозначений (опыт типологического сравнения семантического поля цветообозначений в полесских говорах с говорами других славянских и контактирующих с ними языков) // Полесье. (Лингвистика. Археология. Топонимика). М., 1968.

- Мошинский, 1977: *L. Moszyński*. Dwa nowe słowniki etymologiczne języka gdańskobiałogórskiego // *Rocznik Slawistyczny*, 35, 1975.
- Мука, 1891: *E. Mucke*. Historische und vergleichende Laut- und Formenlehre der niedersorbischen (niederlausitzisch-wendischen) Sprache. Leipzig, 1891.
- Мусаев, 1975: *К. М. Мусаев*. Лексика тюркских языков в сравнительном освещении. (Западнокыпчакская группа). М., 1975.
- Назарова, 1966: *Т. В. Назарова*. Відбиття структурно-генетичних зв'язків у Атласі української мови // *Українська лінгвістична географія*. Київ, 1966.
- Назарова, 1976: *Т. В. Назарова*. О картографировании комплексных лингвистических единиц // *Общеславянский лингвистический атлас. Материалы и исследования*. 1971. М., 1974.
- Налепа, 1968: *J. Nalepa*. Słowiańszczyzna północno-zachodnia. Podstawy jedności i jej rozpad. Poznań, 1968.
- Народнае слова: Народнае слова. Мінск, 1976.
- Народная лексіка: Народная лексіка. Мінск, 1977.
- Народная словатворчасць: Народная словатворчасць. Мінск, 1979.
- Николаев, 1988—1989: *С. Л. Николаев*. Следы особенностей восточнославянских племенных диалектов в современных великорусских говорах. I. Кривичи // *Балто-славянские исследования*. 1986; 1987. М., 1988; 1989.
- Николаев, 1990: *С. Л. Николаев*. К истории племенного диалекта кривичей // *Советское славяноведение*, 1990, № 4.
- Николаев, 1993: *С. Л. Николаев*. Место племенного языка кривичей в общеславянском континууме // *Истоки русской культуры (археология и лингвистика)*. М., 1993.
- Никонов, 1965: *В. А. Никонов*. Происхождение русского *зм* // *Этимология*. [1964]. Принципы реконструкции и методика исследования. М., 1965.
- Нитш, 1905: *K. Nitsch*. Stosunki pokrewieństwa języków lechickich // *Materiały i prace Komisji Językowej Akademii Umiejętności*, III, 1905.
- Нитш, 1956: *K. Nitsch*. Historia badań nad dialektami północnej Polski // *Konferencja Pomorska (1954)*. *Prace językoznawcze*. Warszawa, 1956.
- Новое в русской лексике, 1981: Новое в русской лексике. Словарные материалы — 78. М., 1981.

- Носенко, 1981: *И. А. Носенко*. Начала статистики для лингвистов. М., 1981.
- Общее языкознание, 1973: *Общее языкознание. Методы лингвистических исследований*. М., 1973.
- ОЛА. Животный мир: Общеславянский лингвистический атлас. Серия лексико-словообразовательная. Вып. I. Животный мир. М., 1988.
- ОЛА. МиИ. 1971: *Общеславянский лингвистический атлас. Материалы и исследования*. 1971. М., 1974.
- Олеш, 1959: *R. Olesch. Vocabularium Venedicum von Christian Hennig von Jessen*. Köln; Graz, 1959.
- Олеш, 1962: *R. Olesch. Juglers Lüneburgisch-Wendisches Wörterbuch*. Köln; Graz, 1962.
- Олеш, 1967: *R. Olesch. Fontes linguae Dravaeno-Polabicae minores et Chronica Venedica* J. P. Schultzi. Köln; Graz, 1967.
- Ондруш, 1976а: *Š. Ondruš. Tři slovanské etymologické kompendia // Slavia*. 45. Praha, 1976.
- Ондруш, 1976б: *Š. Ondruš. Praslovanský základ slovenčiny v slovnej zásobe // Stud. Akad. Slov.* 1976, 5.
- Ондруш, 1977: *Š. Ondruš*. (Рец. на: *Этимологический словарь славянских языков. Праславянский лексический фонд. Вып. 3*) // *Slavia*. 46. Praha, 1977.
- Опыт диалектологической карты, 1915: [*Н. Н. Соколов, Н. Н. Дурново, Д. П. Ушаков*.] Опыт диалектологической карты русского языка в Европе с приложением очерка русской диалектологии // *Труды Московской диалектологической комиссии*. М., 1915, вып. 5.
- Орел, 1987: *В. Э. Орел*. К реконструкции праславянского словарного состава // *Советское славяноведение*, 1987, № 5.
- Орлось, 1958: *T. Z. Orłoś. Element prasłowiański w dzisiejszym słownictwie czeskim // Studia z Filologii Polskiej i Słowiańskiej*. 3. Warszawa, 1958.
- Осипова, 1994: *М. А. Осипова*. Праславянская лексика юго-запада Украины: словообразовательный аспект, 1994 (рукопись).
- Палмайтис, 1978: *М. Л. Палмайтис*. Праязык — генетическая или контактная общность? (Анализ данных словаря В. М. Иллича-Свитыча) // *Вопросы языкознания*, 1978, № 1.

- Перебейнос, 1967: *В. И. Перебейнос*. Использование статистических методов в типологических исследованиях (на материале славянских и германских языков) // Проблемы языкознания. Доклады и сообщения советских ученых на X Международном конгрессе лингвистов. (Бухарест, 28. VIII — 2.IX.1967). М., 1967.
- Петровский, 1922: *Н. М. Петровский*. О новгородских «словенах». По поводу книги: А. А. Шахматов. Древнейшие судьбы русского племени... // Известия ОРЯС, т. XXV (1920). Пг., 1922.
- Пивторак, 1988: *Г. П. Пивторак*. Формування і діалектна диференціація давньоруської мови. (Історико-фонетичний нарис). Київ, 1988.
- Пиотровский — Бектаев — Пиотровская, 1977: *Р. Г. Пиотровский, К. Б. Бектаев, А. А. Пиотровская*. Математическая лингвистика. М., 1977.
- Плат, 1965: *У. Плат*. Математическая лингвистика // Новое в лингвистике. Вып. IV. М., 1965.
- Попович, 1960: *J. Popović*. Geschichte der serbokroatischen Sprache. Wiesbaden, 1960.
- Поповская-Таборская, 1987: *Н. Popowska-Taborska*. Z problematyki badawczej nawiązań leksykalnych (na materiale kaszubsko-południowosłowiańskim) // См. след.
- Поповская-Таборская, 1987а: *Н. Popowska-Taborska*. Szkice z kaszubszczyzny. Dzieje. Zabytki. Słownictwo. [Gdańsk, 1987].
- Поповская-Таборская, 1991: *Н. Popowska-Taborska*. Wczesne dzieje Słowian w świetle ich języka. Wrocław; Warszawa; Kraków, 1991.
- Порхомовский, 1982: *В. Я. Порхомовский*. Проблемы генетической классификации языков Африки // Теоретические основы классификации языков мира. Проблемы родства. М., 1982.
- Порхомовский, 1989: *В. Я. Порхомовский*. О принципах сравнительно-исторического изучения бесписьменных и младописьменных языков // Сравнительно-историческое изучение языков разных семей. Реконструкция на отдельных уровнях языковой структуры. М., 1989.
- Порциг, 1964: *В. Порциг*. Членение индоевропейской языковой области. М., 1964.
- Праславянский словарь: Słownik prasłowiański. Tł. I—. Wrocław; Warszawa; Kraków; Gdańsk, 1974—.

- Пуллиблэнк, 1972: *E. G. Pulleyblank*. How rapidly do languages change? // *Langues et techniques. Nature et société*. T. 1. Approche linguistique. Paris, 1972.
- Пшеничнова, 1976: *Н. Н. Пшеничнова*. Об использовании элементов дискретной математики в классификациях говоров // *Лингвистическая география, диалектология и история языка*. Ереван, 1976.
- Пшеничнова, 1977: *Н. Н. Пшеничнова*. Применение метода таксономического анализа в классификации говоров // *Диалектологические исследования по русскому языку*. М., 1977.
- Пшеничнова, 1979: *Н. Н. Пшеничнова*. Мера специфичности и некоторые вопросы классификации частных диалектных систем // *Общеславянский лингвистический атлас. Материалы и исследования*. 1977. М., 1979.
- Пшеничнова, 1981: *Н. Н. Пшеничнова*. Некоторые способы группирования объектов, применяемые в диалектологии // *Проблемы структурной лингвистики*. 1979. М., 1981.
- Пшеничнова, 1987: *Н. Н. Пшеничнова*. О классификации частных диалектных систем вероятностно-статистическим методом // *Русские диалекты. Лингвогеографический аспект*. М., 1987.
- Радева, 1963: *S. Radewa*. Element prasłowiański w dzisiejszym słownictwie bułgarskim // *Studia z Filologii Polskiej i Słowiańskiej*, 4. Warszawa, 1963.
- Рамулт, 1893: *St. Ramułt*. Słownik języka pomorskiego czyli kaszubskiego. Kraków, 1893.
- Расторгуев, 1927: *П. А. Расторгуев*. К вопросу о ляхских чертах в белорусской фонетике // *Труды постоянной комиссии по диалектологии русского языка*. Вып. IX. Л., 1927.
- Реформатский, 1966: *А. А. Реформатский*. Неканоничная фонетика // *Развитие фонетики современного русского языка*. М., 1966.
- Ри, 1958: *J. A. Rea*. Concerning the validity of lexicostatistics // *International Journal of American Linguistics*, vol. 24, 1958.
- Розвадовский, 1915: *J. Rozwadowski*. Stosunek języka polskiego do innych słowiańskich // *Encyklopedia Polska Akad. Umiejętn.*, II, Kraków, 1915.
- Роспонд, 1965: *С. Роспонд*. Структура и классификация древневосточнославянских антропонимов (имена) // *Вопросы языкознания*, 1965, № 3.

- Роспонд, 1979: *С. Роспонд*. *Miscellanea onomastica rossica // Восточно-славянская ономастика. Исследования и материалы*. М., 1979.
- Русская диалектология, 1964: [*Р. И. Аванесов, С. В. Бромлей, Л. Н. Булатова* и др.]. *Русская диалектология*. М., 1964.
- Садник — Айцетмюллер, 1963: *L. Sadnik, R. Aitzetmüller*. *Vergleichendes Wörterbuch der slavischen Sprachen*. Wiesbaden, 1963—.
- Санников, 1975: *В. З. Санников*. Очерк восточнославянской сравнительно-исторической лексикологии (лингвостатистический аспект). М., 1975.
- Санников, 1985: *В. З. Санников*. О степени лексической близости древнерусской, старорусской, староукраинской и старобелорусской письменной речи // *Восточные славяне. Языки, история, культура*. К 85-летию академика В. И. Борковского. М., 1985.
- Сводеш, 1952: *V. Swadesh*. *Lexico-statistic dating of prehistoric ethnic contacts, with special reference to North American Indians and Eskimos // Proceeding of the American Philosophical Society*. Vol. 96, 1952 (русс. пер.: *М. Сводеш*. Лексикостатистическое датирование доисторических этнических контактов (на материале племен эскимосов и североамериканских индейцев) // *Новое в лингвистике*. Вып. I. М., 1960).
- Сводеш, 1955: *V. Swadesh*. *Towards greater accuracy in lexicostatistic dating // International Journal of American Linguistics*. Vol. 21, 1955 (русс. пер.: *М. Сводеш*. К вопросу о повышении точности в лексикостатистическом датировании // *Новое в лингвистике*. Вып. I. М., 1960).
- Селищев, 1941: *А. М. Селищев*. *Славянское языкознание*. Т. I. Западно-славянские языки. М., 1941.
- Селищев, 1968: *А. М. Селищев*. Введение в сравнительную грамматику славянских языков // *А. М. Селищев*. *Избранные труды*. М., 1968.
- Семенова, 1983: *Т. Ф. Семенова*. К вопросу о путях проникновения тюркизов в западноукраинские говоры. I // *Славянское и балканское языкознание. Проблемы лексикологии*. М., 1983.
- Серебренников, 1968: *Б. А. Серебренников*. Об относительной самостоятельности развития системы языка. М., 1968.
- Серебренников, 1974: *Б. А. Серебренников*. Вероятностные обоснования в компаративистике. М., 1974.

- Славский: *F. Sławski. Słownik etymologiczny języka polskiego. T. I—*. Kraków, 1952—.
- Слоўнік паўночна-заходняй Беларусі: Слоўнік беларускіх гаворак паўночна-заходняй Беларусі і яе паграніччя. Т. I—.
- Мінск, 1979—.
- СлРЯ XI-XVII вв.: Словарь русского языка XI—XVII вв. Т. I—.
- М., 1975—.
- Соколовская, 1968: *А. С. Соколовская. Опыт определения лексической близости диалектов (на материале названий одежды и обуви Припятского Полесья) // Полесье. (Лингвистика. Археология. Топонимика). М., 1968.*
- Соколовская, 1973: *А. С. Соколовская. Статистический метод исследования говоров // День Артура Озола. Актуальные вопросы диалектологии. Материалы научной конференции. Рига, 1973.*
- Сорокин, 1977: *Ю. С. Сорокин. Что такое исторический словарь? // Проблемы исторической лексикографии. Л., 1977.*
- Соссюр, 1977: *Ф. де Соссюр. Курс общей лингвистики // Ф. де Соссюр. Труды по языкознанию. М., 1977.*
- Сравнительная грамматика германских языков, 1962—1966: Сравнительная грамматика германских языков. Т. 1—4. М., 1962—1966.
- Срезневский: *И. И. Срезневский. Материалы для Словаря древнерусского языка Т. I—III. СПб., 1893—1903.*
- Срезневский, 1866: *Мнения о словаре славянских наречий А. Б. Шлейхера и И. И. Срезневского. СПб., 1866 (Сборник статей, читанных в Отделении русского языка и словесности императорской Академии наук, т. I, № 2).*
- СРНГ: Словарь русских народных говоров. Вып. 1—.
- Л., 1965—.
- ССРЛЯ: Словарь современного русского литературного языка. Т. 1—17. М.; Л., 1950—1965.
- Старостин, 1991: *С. А. Старостин. Алтайская проблема и происхождение японского языка. М., 1991.*
- Стецюк, 1987: *В. М. Стецюк. Определение мест поселения древних славян графоаналитическим методом // Известия АН СССР. Серия литературы и языка, 1987, № 1.*
- Супрун, 1983: *А. Е. Супрун. Лексическая типология славянских языков. Минск, 1983.*

- Супрун, 1987: А. Е. Супрун. Полабский язык. Минск, 1987.
- Сухачев, 1974: Н. Л. Сухачев. Лингвистические атласы и карты // Проблемы картографирования в языкознании и этнографии. Л., 1974.
- Тарнацкий, 1937: J. Tarnacki. Studia porównawcze nad geografią wyrazów (Polesie — Mazowsze). Warszawa, 1937.
- Ташицкий, 1928: W. Taszycki. Stanowisko języka łużyckiego // Symbolae grammaticae in honorem J. Rozwadowski. Т. II. Kraków, 1928.
- Терентьев, 1991: В. А. Терентьев. [Информация о конференции «Лингвистическая реконструкция и древнейшая история Востока», 1989 г.] // Вопросы языкознания, 1991, № 1.
- Толстой: И. И. Толстой. Сербскохорватско-русский словарь. Изд. 4. М., 1976.
- Толстой, 1968: Н. И. Толстой. Об изучении полесской лексики. (Предисловие редактора) // Лексика Полесья. Материалы для полесского диалектного словаря. М., 1968.
- Толстой, 1969: Н. И. Толстой. Славянская географическая терминология. Семасиологические этюды. М., 1969.
- Толстой, 1977а: Н. И. Толстой. О соотношении центрального и маргинального ареалов в современной Славии // Ареальные исследования в языкознании и этнографии. Л., 1977.
- Толстой, 1977б: Н. И. Толстой. Уз проблем словенских лексичких изоглоса. Српскохрватска лексика у општесловенском оквиру // Научни састанак слависта у Вукове дане. 6. Св. 1. Београд, 1977.
- Толстой — Толстая, 1979: Н. И. Толстой, С. М. Толстая. Д.К.Зеленин — диалектолог // Проблемы славянской этнографии (к 100-летию со дня рождения члена-корреспондента АН СССР Д. К. Зеленина). Л., 1979.
- Топоров, 1991: В. Н. Топоров. О кривичском элементе и кривичской ретроспективе // Славистика. Индоевропеистика. Ностратика. К 60-летию со дня рождения В.А.Дыбо. Тезисы докладов. М., 1991.
- Топоров — Трубачев, 1962: В. Н. Топоров, О. Н. Трубачев. Лингвистический анализ гидронимов Верхнего Поднепровья. М., 1962.
- Трофимович, 1974: К. К. Трофимович. Верхнелужицко-русский словарь. М.; Бауцен, 1974.

- Трубачев, 1963а: [О. Н. Трубачев.] Этимологический словарь славянских языков (праславянский лексический фонд). Проспект. Пробные статьи. М., 1963.
- Трубачев, 1963б: О. Н. Трубачев. О составе праславянского словаря. (Проблемы и задачи) // Славянское языкознание. Доклады советской делегации. V Международный съезд славистов (София, сентябрь 1963). М., 1963.
- Трубачев, 1963в: О. Н. Трубачев. О праславянских лексических диалектизмах сербо-лужицких языков // Сербо-лужицкий лингвистический сборник. М., 1963.
- Трубачев, 1967: О. Н. Трубачев. [Рец. на кн.: «Základní všeslovanská slovní zásoba». Вгпо, 1964] // Этимология. 1965. Материалы и исследования по индоевропейским и другим языкам. М., 1967.
- Трубачев, 1968: О. Н. Трубачев. О составе праславянского словаря (проблемы и результаты) // Славянское языкознание. VI Международный съезд славистов (Прага, август 1968 г.). Доклады советской делегации. М., 1968.
- Трубачев, 1974: О. Н. Трубачев. Ранние славянские этнонимы — свидетели миграции славян // Вопросы языкознания, 1974, № 6.
- Трубачев, 1978: О. Н. Трубачев. Этимологический словарь славянских языков и Праславянский словарь (опыт параллельного чтения) // Этимология. 1976. М., 1978.
- Трубачев, 1982: О. Н. Трубачев. [Рец. на кн.: «Słownik prasłowiański». Т. III (*davъtъ* — *dobъrati*). Wrocław; Warszawa; Kraków; Gdańsk, 1979] // Этимология. 1980. М., 1982.
- Трубачев, 1985: О. Н. Трубачев. Праславянская лексикография // Этимология. 1983. М., 1985.
- Трубецкой, 1960: Н. С. Трубецкой. Основы фонологии. М., 1960.
- Трубецкой, 1987а: Н. С. Трубецкой. О звуковых изменениях русского языка и распаде общерусского языкового единства // Н. С. Трубецкой. Избранные работы по фонологии. М., 1987.
- Трубецкой, 1987б: Н. С. Трубецкой. Возникновение общих западнославянских особенностей в области консонантизма // Там же.
- Туровский словарь (Тураўскі слоўнік): Тураўскі слоўнік. Т. 1— Мінск, 1982—.

- Усачева, 1977: *В. В. Усачева*. Ареальная характеристика названий леща в славянских языках // *Общеславянский лингвистический атлас. Материалы и исследования*. 1975. М., 1977.
- Фасмер: *М. Фасмер*. Этимологический словарь русского языка. Т. I—IV. М., 1964—1973.
- Филин, 1962: *Ф. П. Филин*. Образование языка восточных славян. М.; Л., 1962.
- Филин, 1972: *Ф. П. Филин*. Происхождение русского, украинского и белорусского языков. Историко-диалектологический очерк. Л., 1972.
- Филин, 1983: *Ф. П. Филин*. Проблемы исторической лексикологии русского языка (древний период) // *Славянское языкознание. IX Международный съезд славистов. Доклады советской делегации*. М., 1983.
- Филин, 1984: *Ф. П. Филин*. Историческая лексикология русского языка. Проспект. М., 1984.
- Фодор, 1961a: *Fodor I.* A glottochronologia érvényessége a szláv nyelvek anyaga alapján // *Nyelvtudományi közlemények*. Vol. 63, 1961, № 2.
- Фодор, 1961b: *Fodor I.* The Validity of glottochronology on the basis of the Slavonic languages // *Studia Slavica*. T. VII, 1961, № 4.
- Фурдаль, 1961: *A. Furdal*. Rozpad języka prasłowiańskiego w swietle rozwoju głosowego. Wrocław, 1961.
- Хабургаев, 1979: *Г. А. Хабургаев*. Этнонимия «Повести временных лет» в связи с задачами реконструкции восточнославянского глоттогенеза. М., 1979.
- Хабургаев, 1980: *Г. А. Хабургаев*. Становление русского языка (пособие по исторической грамматике). М., 1980.
- Хайду, 1985: *П. Хайду*. Уральские языки и народы. М., 1985.
- Хаймз, 1960: *D. H. Hymes*. Lexicostatistics so far // *Current Anthropology*. Vol. 1, 1960, No 1.
- Хинце, 1962: *Fr. Hinze*. Das Pomoranische Wörterbuch // *Slavia*, 1962, ses. 3.
- Хяккинен, 1984: *K. Hakkinen*. Wäre es schon an der Zeit, denn Stammbaum zu fällen? Theorien über die gegenseitigen Verwandtschaftsbeziehungen der finnisch-ugrischen Sprachen // *Ural-Altäische Jahrbücher*. N.[eue] F.[olge]. Bd. 4. Wiesbaden, 1984.

- Цейтлин, 1977: *Р. М. Цейтлин*. Лексика старославянского языка. Опыт анализа мотивированных слов по данным древнеболгарских рукописей X—XI вв. М., 1977.
- Частотный словарь, 1977: Частотный словарь русского языка. М., 1977.
- Чейка — Лампрехт, 1963: *М. Čejka, A. Lamprecht*. K otázce vzniku a difereenciace slovanských jazyků // Sborník praci filosofické fakulty Brněnské university. 1963, A 11 (отдельный оттиск).
- Чекановский, 1927: *J. Czekanowski*. Wstęp do historii Słowian. Lwów, 1927. (Переиздание: То же. Poznań, 1957).
- Чекановский, 1932: *J. Czekanowski*. Różnicowanie się dialektów prasłowiańskich w świetle kryterium ilościowego // Sborník prací I. Sjezdu slovanských filologů v Praze 1929. Praha, 1932.
- Чекман, 1978: *В. Н. Чекман*. Типологические аспекты фонетических изменений в праславянском. Минск, 1978.
- Чекман — Широков, 1962: *В. Н. Чекман, О. С. Широков*. Квантитативный метод Крубера и классификация славянских языков // Всесоюзная конференция по славянской филологии. 17—22 декабря 1962 года. Программа и тезисы докладов. Л., 1962.
- Шайкевич, 1980: *А. Я. Шайкевич*. Гипотезы о естественных классах и возможность количественной таксономии в лингвистике // Гипотеза в современной лингвистике. М., 1980.
- Шахматов, 1907: *А. А. Шахматов*. Южные поселения вятичей // Известия Академии наук. СПб., 1907, № 16.
- Шахматов, 1911: *А. А. Шахматов*. Древние ляхские поселения в России // Славянство, 1911, № 4/6.
- Шахматов, 1913: *А. А. Шахматов*. К вопросу о польском влиянии на древнерусские говоры // Русский филологический вестник. Т. 69, 1913, № 1.
- Шахматов, 1915: *А. А. Шахматов*. Очерк древнейшего периода русского языка. Пг., 1915.
- Шахматов, 1916а: *А. А. Шахматов*. Введение в курс истории русского языка. Ч. 1. Пг., 1916.
- Шахматов, 1916б: *А. А. Шахматов*. Заметки по истории звуков лужицких говоров. По поводу книги Л. В. Щербы «Восточнолужицкое наречие» // Известия ОРЯС. Т. XXI, кн. 2, 1916.
- Шимчук, 1975: *Э. Г. Шимчук*. О потенциальных и окказиональных словах в исторических словарях // Проблемы славянской исто-

- рической лексикологии и лексикографии. Тезисы конференции. Октябрь 1975 г. Москва. Вып. 3. Теория и практика исторической лексикографии. М., 1975.
- Широков, 1988: *О. С. Широков*. Реконструкция праязыковых изоглосс общеиндоевропейского диалектного континуума // Сравнительно-историческое изучение языков разных семей. Теория лингвистической реконструкции. М., 1988.
- Штибер, 1934: *Z. Stieber*. Stosunki pokrewieństwa języków łużyckich. Kraków, 1934.
- Штибер, 1956: *Z. Stieber*. Stosunek kaszubszczyzny do dialektów Polski łądowej // Konferencja Pomorska (1954). Prace językoznawcze. Warszawa, 1956.
- Штибер, 1972: *З. Штибер*. О древних словенско-западнославянских языковых связях // Русское и славянское языкознание. К 70-летию члена-корреспондента АН СССР Р.И.Аванесова. М., 1972.
- Шустер-Шевц, 1959: *H. Schuster-Šewc*. Sprache und ethnische Formation in der Entwicklung des Sorbischen // Zeitschrift für Slawistik. Bd. IV, N. 4, 1959.
- Шустер-Шевц, 1975: *H. Schuster-Šewc*. [Рец. на кн.: «Этимологический словарь славянских языков. Праславянский лексический фонд». Вып. 7.] // Zeitschrift für Slawistik. Bd. 20, 1975.
- Щерба, 1915: *Л. В. Щерба*. Восточнолужицкое наречие. Т. 1. Пг., 1915.
- Эллергрод, 1959: *A. Ellegård*. Statistical measurement of linguistic relationship // Language. Vol. 35, 1959.
- ЭСБМ: Этымалагічны слоўнік беларускай мовы. Т. 1.— Мінск, 1978—.
- ЭССЯ: Этимологический словарь славянских языков. Праславянский лексический фонд. Вып. 1.— М., 1974—.
- Янин, 1992: *В. Л. Янин*. Древнее славянство и археология Новгорода // Вопросы истории, 1992, № 10.
- Янкова: *Т. С. Янкова*. Дыялектны слоўнік Лоеўшчыны. Мінск, 1982.

Оглавление

<i>От автора</i>	3
------------------------	---

Глава 1

Механизмы эволюции словаря и роль лексики в установлении уровней языкового родства

1. Сравнительная значимость фонетических, морфологических и лексических критериев в глоттогенетических исследованиях	5
2. Количественная характеристика лексикона как базы генетико-классификационных построений	11
3. О мозаичности изолексов: есть ли порядок в хаосе?	15
4. Лексика: нестабильность или консерватизм?	17
5. Гиперболизация фактора лексических утрат в изменении словаря	19
6. Различия между механизмами эволюции фонетики и лексики: вытеснение и кумуляция	24

Глава 2

Лексикостатистика как инструмент глоттогенетического исследования

1. Статистическое моделирование в исторической лингвистике	30
2. Глоттохронология и близкие ей методы	34
3. Расширение исходной базы	44

Глава 3

Поиск метода

1. Некоторые отправные положения	50
2. Празыковой лексический фонд — основа надежной статистики	53

3. Выработка формулы количественного сравнения	57
4. О неудовлетворительности стандартных формул меры сходства объектов	63
5. Пробная демонстрация методики	66
6. К выбору весового коэффициента	72

Глава 4

Материал исследования. Филологическая критика источника. Оценка статистической достаточности материала

1. Выбор источника	75
2. Филологическая критика источника	80
3. Уточнение изоглосс: непривлеченная лексика	87
4. Сегрегационный анализ	91
5. Расширение корпуса изоглосс	94
6. Состав исследуемых идиомов (разные версии)	100
7. Определение статистической надежности материала	104

Глава 5

Из количественно-типологических наблюдений над лексикой славянских языков (праславянское наследие)

1. Необходимость типологических уточнений	107
2. Объем праславянского лексического корпуса	110
3. Объемы праславянских словарей отдельных славянских языков	112
4. Специфичность праславянского лексического наследия отдельных славянских языков (объемы дифференциальных словарей)	116

Глава 6

Картина родства славянских языков по данным лексико-статистики

1. Абсолютные показатели статистических связей	120
2. Индекс генетической близости	128
3. Обсуждение результатов	134
а. Южнославянские языки	134
б. Западнославянские языки	142
в. Восточнославянские языки	144
4. Обобщенное представление межславянских лексико-статистических связей отдельных языков	147
5. Эксклюзивные лексические связи	154
6. Отдельные славянские языки вне ближайших родственных связей	160

*Глава 7***О возможности лексикостатистического обнаружения явления языковой конвергенции**

1. Конфронтационная статистика разнонаправленных изоглоссных связей ближайшерадственных языков как способ выявления феномена конвергенции 165
2. Численные данные и итоги наблюдений 169
3. Об условиях эффективности предлагаемой методики 176

*Глава 8***Проблема древненовгородского диалекта sub specie лексикостатистики**

1. Проблема древненовгородского диалекта в современных историко-диалектологических исследованиях 179
2. Изоглоссные лексические связи древненовгородского диалекта 181
3. Лексикостатистический анализ и его результаты: неомогенность восточнославянской языковой группы 187

*Глава 9***Сопоставление с результатами глоттохронологии**

1. Глоттохронологическая модель: численные данные 195
2. Сходства и различия 198

*Глава 10***Сопоставление с результатами статистического определения родства на материале сравнительной фонетики**

1. Две фоностатистические модели. Численные данные 202
2. Дискретность фоностатистической — «вязкость» лексикостатистической картин родства: возможные причины различий 213
3. Детальное сопоставление: несогласованность фонетической и лексической эволюции 217

*Глава 11***К реконструкции диалектного членения позднеславянского языка 222***Использованная литература 231*

Научное издание

Утверждено к печати

Институтом славяноведения и балканистики РАН

Анатолий Федорович Журавлев

**Лексикостатистическое моделирование
системы славянского языкового родства**

Директор издательства – *С. Григоренко*

Главный редактор – *Н. Волочаева*

Выпускающий редактор – *О. Климанов*

Корректурa авторская

Оформление оригинал-макета и верстка

Н. Волочаевой

Оформление обложки

С. Григоренко

ЛР № 070644
выдан 26 октября 1992 г.
Формат 60×84 1/16
Гарнитура «Бодони»
Печать офсетная
16,0 п. л.
Тираж 1 000 экз.
Заказ № 892

Отпечатано с оригинал-макета
в Московской типографии № 2 ВО «Наука»
121099, Москва, Г-99, Шубинский пер., д. 6

8) "О возможности лексикостатистики. обобщение
 "списки". \sqrt{V}
 Для возможности в языках (подчеркнуто) конкурентных
 языковых изоляционных черт лексикостатистики
 $V(A, B)$

ЛЕКСИКОСТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

А. Ф. Журавлев

Особенностью работы является обращение не к коротким выборочным диагностическим перечням лексики типа стословного списка понятий, применяемого в глоттохронологических исследованиях, и не к отдельным группам лексики, ограниченным по каким-либо тематическим или формальным (словообразовательным, грамматическим и т. д.) признакам, а опора на сплошное статистическое обследование праславянского словаря. При этом количественному анализу подвергся доступный к началу работы праславянский лексический материал в его максимальном объеме (увеличенном привлечением дополнительных источников).

Лексикостатистический анализ проводится здесь с принятием во внимание типологических характеристик отдельных славянских языков и их лексических составов. При сопоставлении картины родственных взаимоотношений славянских языков, полученной на основе количественного анализа изолекса, с аналогичной картиной, которая восстанавливается по данным сравнительной фонетики, учитываются принципиальные различия в механизмах эволюции фонетического и лексического уровней.

$\sqrt{V(A, B)}$
 для
 с. / у
 ар
 у ш
 кануб-
 сердох-
 слово
 введение

...решения. изоляцион - икативный