

Российская академия наук
Институт русского языка

На правах рукописи

Анатолий Федорович ЖУРАВЛЕВ

**ЛЕКСИКОСТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
СИСТЕМЫ СЛАВЯНСКОГО ЯЗЫКОВОГО РОДСТВА**

10.02.01 — Русский язык

10.02.03 — Славянские языки

**Автореферат диссертации
на соискание ученой степени
доктора филологических наук**

Москва, 1992



Работа выполнена в Отделе современного русского языка Института русского языка РАН.

Официальные оппоненты:

действительный член РАН,
доктор филологических наук Н.И.Толстой,
член-корреспондент РАН,
доктор филологических наук О.Н.Трубачев,
доктор филологических наук В.З.Санников.

Ведущее учреждение — кафедра общего и славянского языкознания Белорусского государственного университета.

Защита состоится • 18 • 06 1992 г.
в 14 час. на заседании специализированного совета
Д 002.19.01 при Институте русского языка РАН.
(Москва, ул. Волхонка, 18/2).

С диссертацией можно ознакомиться в библиотеке
Института русского языка РАН (Москва, ул. Волхонка,
18/2)

Автореферат разослан • 14 • 05 1992 г.

Ученый секретарь
специализированного совета

В.Н.Белусов

Диссертация 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 позднеславянского языка. Исследование праславянских изолексов осуществляется в диссертации с применением статистических методов (предложена формула индекса 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, ограниченных по каким-либо тематическим или формальным (словообразовательным, частеречным и т.п.) признакам, а опора на сплошное статистическое обследование праславянского словаря. При этом количественному анализу подвергается доступный в настоящее время праславянский лексический материал в его максимальном объеме (увеличенном привлечением дополнительных источников). Лексикостатистический анализ проводится с учетом типологических характеристик отдельных славянских языков и их лексических составов. При сопоставлении картины родственных взаимоотношений языков, полученной на основе количественного исследования изолексов, с аналогичной картиной, восстанавливаемой по данным сравнительной фонетики, учитываются принципиальные различия в механизмах эволюции лексического и фонетического уровней.

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

Диссертация состоит из двенадцати глав, объединенных в две части. В первой части (гл. 1—5) рассматриваются теоретические предпосылки работы, во второй части (гл. 6—12) — результаты статистического анализа праславянской лексики современных славянских языков.

В главе 1 («Механизмы эволюции словаря и роль лексики в установлении языкового родства») содержится попытка пересмотра если и не общепринятой, то преобладающей в компаративистике точки зрения, согласно которой лексика является самым неустойчивым языковым уровнем, а потому в процедурах установления относительной близости между разными родственными языками (или диалектами одного и того же языка) лексические и словообразовательные данные обладают меньшей значимостью по сравнению с данными фонетики и грамматики (главным образом морфологии).

Проблему большей или меньшей эффективности использования различных — фонетических, грамматических, лексических — данных в установлении степеней языкового родства и при построении генеалогических классификаций 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 подверстать это многообразие к неоспоримой иерархии языковых 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, продемонстрируют существенно меньшую разрешающую силу, будучи примененными к иным языковым группам и типологическим ситуациям. Подобные задачи могут решаться по-разно-

му, в зависимости от таксономического ранга идиомов, подлежащих сравнению; от их количества; от их принадлежности тому или иному грамматическому типу (морфологический критерий малопригоден для выявления степеней родства изолирующих языков); от возраста сравниваемых языков, т.е. их удаленности от состояния праязыкового единства; от эволюции хотя бы одного из сравниваемых языков на том или ином ее этапе в относительной или полной изоляции от родственных языков; от наличия в распоряжении лингвиста сведений о древних периодах развития языков, в частности, от их принадлежности древле- или младописьменным (бесписьменным) и т.д. Определить и утвердить предпочтительность тех или иных уровней при выяснении меры взаимной близости между родственными языками и восстановлении картины отношений между диалектами праязыка в конкретной историко-лингвистической ситуации может лишь сам анализ разноуровневых данных, результаты которого многократно корректируются последующими рассмотрениями проблемы. Однако есть необходимость обрисовать особенности лексического уровня, кладущегося в основу классификационных и глоттогенетических построений, в его отличии от иных уровней, прежде всего фонетического.

Наиболее очевидной чертой лексики как генетико-классификационной базы является ее большой количественный состав. Объем лексических фактов, могущих лечь в основание реконструкции системы родства, в сотни раз превосходит любые списки фонетических и грамматических явлений — и это при том, что в указанных целях может быть использована только праязыковая лексика, сохраненная современными языками. Общая величина доступного реконструкции праславянского словаря составляет никак не меньше 20 тысяч лексем (см. ниже, гл. 6), тогда как праязыковые фонетические и морфологические изоглоссы насчитывают не более нескольких десятков единиц. Следовательно, лексика дает возможность обнаружения гораздо большего количества сходжений и размежеваний между родственными языками и диалектами праязыка, из которых они развились, чем фонетика и морфология, а благодаря большей сложности и дифференцированности «шкалы» степеней близости между языками — достижения большей «прецизионности» при ре-

конструкции картины родства. Огромный объем лексических изоглосс, измеряемый числом позиций праязыкового словаря, дополняется значительным количеством транслексемных изоглосс, извлекаемых из того же материала, но не сводимых к словарным позициям (корневых, формантных, морфонологических и т.п., в диалектологических атласах обычно выражаемых разветвленной диакритикой при картографическом знаке). Все это делает лексический изоглоссный материал несопоставимым по объему с данными фонетики и грамматики.

Другой чертой словаря, отличающей его как от фонетической, так и от грамматической систем, является нежесткий характер его организации, что прямо связано с величиной и открытостью списка его элементов.

Эволюция лексической системы, в соответствии с указанными особенностями словаря, осуществляется иными, по сравнению с фонетикой, механизмами. Если фонетика и, в меньшей степени, морфология, будучи системами с жесткой, в некоторых участках почти механической взаимобусловленностью компонентов, эволюционируют преимущественно путем вытеснения одних элементов и структур другими, то развитие словаря, системы более рыхлой, скорее даже конгломерата рыхлых систем, протекает главным образом за счет к у м у л я ц и и элементов (слов). Изменения в фонетической системе часто носят характер цепной реакции, которая охватывает всю систему и изменяет принципы, обеспечивающие ее статическое равновесие, весьма радикальным образом, ср. коренные перестройки славянской фонетики под действием закона открытого слога или вследствие падения редуцированных. Эволюция словаря протекает гораздо более спокойно, и даже ее этапы, расцениваемые как чрезвычайно бурные, например, в истории русского языка — петровская и советская эпохи, редко затрагивают центральные участки огромного лексического корпуса.

Различия между лексикой и фонетикой в преобладающих механизмах исторических изменений 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 е д и н и ц а х (корнях,

лексемах, словобразовательных конструкциях с исторически идентичным морфемным составом), применительно же к фонетическому уровню — главным образом на отношениях соответствия материальных единиц и структур — регулярных корреспонденциях, позволяющих выявление закономерных переходов одних единиц в другие. Морфология в этом отношении занимает промежуточное положение. Изменение в лексике состоит как правило в заполнении вновь сформированным элементом некоей пустоты, обнаружившейся при концептуальном освоении мира (при возникновении потребности номинации), реже — из потребностей экспрессии. В фонетике же происходит не заполнение «концептуальных» пустот (их нет), а замена одной единицы или структуры другою, нередко с последующим перестроением всей системы в целом, что обусловлено жесткостью ее организации. Вследствие относительно малого набора элементов и особенностей их взаимосвязей большую роль в эволюции фонетической системы играют общетипологические моменты, что приводит к известному сходству результатов некоторых частных фонетических процессов в разных языках. В лексике типологические моменты или, точнее, возможность параллельного образования, играют существенно меньшую роль.

Сказанное позволяет судить о высокой степени устойчивости лексики на фоне других языковых уровней, что служит некоторым исследователям поводом утверждать, что лексика является наиболее консервативным уровнем и потому самым надежным для установления степеней генетической близости между родственными языками. В силу кумулятивного принципа эволюции и неподверженности лавинным перестройкам лексика способна гораздо дольше, чем фонетика и морфология, сохранять следы древних диалектообразующих сближений и расхождений.

Из-за огромного количественного превосходства лексических изоглосс над фонетическими и грамматическими картина межъязыковых связей, построенная на лексических данных, выглядит не столь стройной, но значительно более сложной и труднее читаемой по сравнению с аналогичной картиной, воссоздаваемой на базе фонетических и грамматических изоглосс, оставляя впечатление мозаичности и даже хаоса. Немногие лексические изоглоссы, образующие пучки, попросту

тонут в нагромождении изолексов, исчерчивающих карту славянских языков вкривь и вкось. Следует думать, что справиться с этим нагромождением, увидеть в хаотичности изолексов порядок, определенные тенденции, устойчивые взаимные сближения и отталкивания между языками и диалектами может статистический анализ данных праславянского словаря.

В главе 2 («Лексикостатистика как инструмент глоттогенетического исследования») содержится сжатый критический обзор направления, базирующегося на применении лексикостатистики к области сравнительно-исторического языкознания.

Использование статистических методов для решения задач компаративистики имеет уже почтенную историю (в славистике начало применения статистики к истории языка связывается с именем Я.Чекановского, см. его работу «Wstęp do historii słowian», 1927). Однако не всеми языковедами-диахронистами лингвостатистика признается полноценным и эффективным методом. Укреплять лингвистов в неприятии квантитативных методов, хотя бы и простейших, могут историко-лингвистические работы, выполненные с применением статистики, в которых из количественных наблюдений делаются неправомерные выводы и излишне смелые обобщения. Этот род работ не столь малочислен, поэтому было сочтено возможным сделать подробный разбор одной такой работы, где, на наш взгляд, сконцентрированы все ошибки, которые могут быть допущены при привлечении лексикостатистики к решению проблем исторической лингвистики и смежных дисциплин. Полнота разбора оправдывается тем, что анализ такого рода дает возможность проще и нагляднее обозначить пределы компетенции квантитативных методов применительно к истории языка и его носителей и извлечь более основательные уроки из чужого неудачного опыта интерпретации количественных данных.

Извлеченные из критического анализа крайне неудачной работы В.М.Стецюка «Определение мест поселения древних славян графоаналитическим методом» (ИАН СЛЯ, 1987, № 1) некоторые уроки изложены в главе в намеренно догматическом ключе: не все они, судя по рассмотренной работе, являются столь уж очевидной банальностью.

— Вера во всемогущество статистики неосновательна.

Число способно служить сильным подспорьем в «измерении» отношений между родственными языками, но статистическое обследование языков — еще не гарантия и даже не основа для правильных выводов об истории их носителей (соответствующих этносов). Статистические данные о языках, в частности о лексике, 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 изоглоссы.

— Лексикостатистика, как и квантитативное исследование других языковых уровней, не должна сводиться к арифметике, т.е. 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 аберрациям историко-лингвистического зрения и непоправимым перекосам в итоговой картине древнейших межъязыковых (диалектных на уровне праязыка) связей.

— 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. Тематическая и семантическая (равно как и грамматическая — частеречная и проч.) селекция лексического материала, даже (а может быть, в особенности) если она нацелена 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 фактора 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 применяются количественные методы, по существу противоречит элементарному требованию статистических наблюдений: наиболее достоверные результаты могут быть получены только с помощью случайной выборки.

Из конкретных квантитативных методов, применяемых в компаративистике, наибольшую известность получила глоттохронология М.Сводеша (которую часто, неправомерно ограничивая семантический объем термина, называют лексикостатистикой). Сравнение процедур, предлагаемых в нашей работе, с глоттохронологическими напрашивается само собой — хотя бы для того, чтобы отчетливее выявить особенности нашего подхода.

Основная цель метода Сводеша — абсолютное датирование этапов распада праязыка, при котором установление картины языкового родства с определением неравномоности отдельных межъязыковых связей — лишь промежуточный результат. Нужно отметить, что глоттохронологическая методика (в традиционной версии Сводеша или модифицированная, ср. версии С.Гуджинской и А.Родригеса, С.А.Старостина, С.Е.Яхонтова и др.) нередко используется для оценки взаимной близости отдельных языков и языковых групп (семей) без установки на непременно датирование их расхождения, с указанием только процента корневых совпадений как меры родства. О методе глоттохронологии много писали, рассматривая в качестве его недостатков прежде всего малый объем *basic vocabulary* и недоказанность постулата о постоянстве скорости его эрозии. К другим, и чрезвычайно существенным, недостаткам метода глоттохронологии следует отнести исходное — как правило молчаливое — допущение диалектной монолитности праязыка, т.е. фактически опора на давно скомпрометированное понятие родословного древа; принятие во внимание лишь дивергентных процессов и «замалчивание» вторичного сближения, конвергенции как одного из факторов лексической близости языков; отказ от обсуждения критериев языковой самостоя-

тельности, вследствие чего разделяющиеся идиомы выступают на страницах глоттохронологических работ бесплотными сущностями, языками «без свойств»; отчетливо осязаемое в результатах применения традиционной версии метода (например, у И.Фодора, у М.Чейки и А.Лампрехта в глоттохронологическом исследовании славянских языков) занижение реального возраста самостоятельных языков; вытекающая из ограниченности 100-словного диагностического теста невысокая разрешающая способность глоттохронологической «оптики», вследствие чего близкородственные языки характеризуются чрезвычайно однообразными показателями сил связей. Преодоление этих внутренних пороков глоттохронологии как метода составляет необходимое звено в совершенствовании лексикостатистического подхода к компаративистским проблемам.

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. Методика А.Я.Шайкевича предполагает обращение не к 100-словному списку Сведеша, а к значительно превосходящим его по объему спискам понятий (627 понятий по словарю индоевропейских синонимов К.Вака); кроме того, вводится процедура апостериорного взвешивания признаков, выявляющая меру устойчивости/неустойчивости различных групп лексики в истории языка. Эти особенности методики А.Я.Шайкевича обеспечивают значительно большую гибкость анализа и надежность результатов. Однако и в данном случае применения лексикостатистики дифференцирующая сила лексического теста является, на наш взгляд, недостаточной. Возможное возражение, состоящее в том, что «коллективный портрет» близкородственных языков и должен оказаться довольно стертым, поскольку речь идет о сходстве на уровне праязыка, следует отклонить на том основании, что и праязык гораздо более адекватно реальности представлять как диалектно структурированное образование с неравномоными внутренними связями его компонентов. Выход видится в значительном расширении исходной лексической базы исследования.

Преодоление недостатков количественных методик, основанных на использовании ограниченных по объему контрольных лексических списков, осуществляется

в двух направлениях. Одно из них опирается на анализ целого текста (например, работы В. Маньчака по установлению меры близости между языками с помощью статистического исследования параллельных текстов). К анализу текстов прибегают и в соображениях усовершенствования методики глоттохронологии (С. А. Старостин). Другой путь — обращение к сплошному статистическому анализу словаря. На наш взгляд, именно такой подход наиболее плодотворен и способен придать количественным оценкам лексической близости между языками неоспоримую точность, хотя он намного более трудоемок. В виде примера можно сослаться на выполненное В. З. Санниковым тщательнейшее статистическое исследование целых лексических систем, отраженных в древнерусских юридических текстах XI—XIV вв., старорусских и западнорусских юридических текстов XV—XVI вв. по памятникам суммарным объемом ок. 70 тыс. словоупотреблений для первой группы и более 100 тыс. — для каждой из двух других групп. Заслуживают интереса и исследования, определяющие взаимную близость языков (или диалектов одного языка) на основе статистического обследования сплошной выборки из массива лексических единиц неопределенно большого размера. Нетрудно обнаружить, что количественным исследованиям, опирающимся на статистическую обработку целых лексических систем или сплошных выборок из них значительного объема, свойственна несравнимо большая тонкость и детализация в оценках близости между языками, чем работам, которые основываются на коротких диагностических списках лексики.

В г л а з е 3 («Поиск метода») излагаются отправные положения, на основе которых возможно построение лексикостатистической модели языкового родства, и предлагается формула, с помощью которой выявляется численная мера генетической близости языков.

Мы исходим из признания лексикостатистических методов наиболее объективным способом установления степени генетической близости между языками. Единственно надежной материальной базой глоттогенетических построений на лексикостатистической основе являются праязыковые словники современных языков, т.е. своды родственноязыковых лексических коррес-

пондений, возводимых к единым праязыковым реконструкциям. Таким образом, за пределы статистических процедур должны быть выведены все неунаследованные слова — лексические образования эпохи «после распада» праязыка, поздние (послепраязыковые) сорместные и индивидуальные заимствования из третьих языков и, что следует 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. Можно предполагать, что квантитативные соотношения между праязыковыми словниками современных языков, точнее говоря, пропорции между различными пластами лексики в зависимости от степени ее распространенности (лексика, известная всем сравниваемым языкам, каким-либо их группам и, наконец, узкие праязыковые регионализмы данного языка), в целом с достаточной мерой приближенности отражают реальные связи их протоидиомов — диалектов самого праязыка.

С предельной строгостью картина языкового родства лексикостатистическими методами может быть воссоздана на основе квантитативного анализа в сей праязыковой лексики, сохраняющейся в языках-потомках. Это означает, что применительно к славянским языкам объем материала, привлекаемого для статистического анализа, по сравнению со списком Сводеша должен возрасти на два порядка, что значительнейшим образом повысит надежность результатов. Для современных средств вычисления этот объем, разумеется, отнюдь не является чрезмерным; вопрос состоит лишь в том, чтобы исследователь располагал подобным материалом.

Публикация создаваемых в Москве и Кракове двух словарей, ставящих своей целью реконструкцию праславянского лексического фонда в его полном объеме, делает нашу задачу выполнимой, только с той оговоркой, что в основу количественного анализа сейчас может быть положен не весь праславянский словарь, а сплошная выборка из него, равная объему изданной его части.

Использование этимологического словаря, ориентированного на предельно полный охват и статейное перечисление праязыковой лексики, обнаруживаемой в современных языках, принципиально важно для рабо-

ть подобного рода. Оно дает сильное преимущество перед работами в духе глоттохронологии, поскольку сводит к минимуму возможное влияние на статистические результаты явлений лексического взаимопроникновения в языках-потомках: во-первых, при нынешнем, достаточно доброкачественном, состоянии славянской этимологии и истории языков послепрязыковые лексические перемещения из языка в язык в своем большинстве выявляются более или менее надежно; во-вторых, при оперировании лексическими массивами в несколько тысяч единиц возможна, и статистически вполне корректна, элиминация сомнительных случаев. При возникновении сомнений относительно древности или исконного характера какого-либо слова его включение или невключение в подсчеты для глоттохронологии с ее «малой» статистикой (не следует думать, что 100-словный список Сводеша в силу его особенной устойчивости гарантирован от смнений такого рода) выбор того или другого решения чреват существенными последствиями: учет или неучет только нескольких слов может сдвинуть даты этапов дивергенции на столетия. Для «большой» же статистики, какие бы альтернативы в оценках древности и исконности лексики ни были предпочтены, на конечный итог выбор окажет в целом незначительное влияние: спорные случаи как бы растворяются в массе более надежного материала, подвергаемого статистической обработке, и этим сильно смягчается воздействие допущенных погрешностей.

К формуле, с помощью которой выявляется численная мера родства, предъявляется несколько важных требований.

Статистический анализ не должен, как было сказано, сводиться к простой арифметике — определению количества слов, связывающих данную пару языков ($V(A, B)$), с дальнейшим его сличением с аналогичным показателем для другой пары. Такой способ, нередко, как мы можем отметить, встречающийся в «лингвостатистических» работах, не пригоден для каких бы то ни было количественных сравнений глоттогенетической направленности, поскольку эти цифры находятся в сильнейшей прямой зависимости от объема праязыкового словника, реконструируемого для данного языка. Прямое сличение абсолютных показателей лексических связей «покажет», что болгарский язык генетически ближе к северновеликорусскому наречию, нежели к

македонскому языку, а родство польского и сербохорватского языков намного глубже, чем родство польского с кашубско-словинским. Следовательно, численные показатели лексически: связей праязыкового характера между данными двумя языками должны быть отнесены к общим объемам восстанавливаемых праязыковых словарей данных языков (т.е. ... / $N(A) \cdot N(B)$).

Количественное сравнение близости языков в различных их парных объединениях не должно замыкаться выделением сепаратных (эксклюзивных) изолексов. Ограничение лишь сепаратными лексическими связями приводит к срабатыванию той же зависимости результатов от объемов праязыковых словарей данных языков: абсолютные эксклюзивные лексические связи русского и сербохорватского языков преобладают над количеством русско-белорусских, русско-украинских и сербохорватско-словенских эксклюзивных изолексов, что резко противоречит имеющимся представлениям о взаимной близости между упомянутыми языками. Ближайшие родственные пары, составляемые болгарским и македонским или украинским и белорусским, по числу сепаратных изолексов уступают связям русского языка с сербохорватским, словенским и чешским. Кроме того, ситуация с украинским и белорусским, праславянские словники которых достаточно велики, имеет еще одну — довольно парадоксальную — причину. Как ни странно, малое число объединяющих их эксклюзивных связей объясняется как раз высокой степенью генетической близости между языками восточнославянской (под)группы. «Незначительность» исключительных словарных корреспондентий праславянского характера между украинским и белорусским коренится в их близком родстве с третьим языком. Отношения в подсистеме, где преобладают тернарные связи, имеют иную природу, чем отношения в парных подсистемах. Соседний великорусский язык с его гигантским праславянским словарем еще не успел настолько удалиться от белорусского и украинского, чтобы в изоглоссах, связывающих два последние языка, стали внушительными по количеству сепаратные, бинарные связи. Взаимная особенная близость трех языков, а не пар, как в других подгруппах, делает двойственные исключительные связи между составляющими данной (под)группы менее вероятными, чем тройственные, отличающие их от всех прочих родственных языков.

Из этого следует вывод, что при всей чрезвычайной важности сепаратных сходжений их нельзя абсолютизировать как свидетельства особо тесных генетических отношений. Весьма существен фон, на котором разгрыбаются сценарий лингвистического родства. Неучет лексического фона чреват переоценкой относительно высоких абсолютных показателей сепаратных связей между языками, для которых реконструируются пространные праязыковые словники, и, напротив, репутительной недооценкой относительно меньших цифр, отражающих эксклюзивные параллели между языками со сравнительно небольшим праязыковым лексическим наследием.

Однако обращение ко всей лексике, а не только к узким изоглоссам, требует разграничения разных их видов. Ценность лексических изоглосс в установлении степени генетической близости языков находится в обратной зависимости от числа охватываемых ими языков: наиболее показательными в отношении родственной близости являются исключительные связи между языками, наименее — общеславянские (в буквальном смысле) слова. Следовательно, в формулу, с помощью которой будет определяться мера родства языков по данным лексики, должны быть введены коэффициенты (р), которые устраняли бы перекосы, возникающие из уравнивания значимости неравномоощных изоглосс. В качестве шкалы весовых коэффициентов мы предлагаем последовательность натуральных чисел от 2 (минимальное число сравниваемых идиомов) до n (максимальное), но связываемых с числами, выражающими лексические связи, в обратном порядке, т.е. весовой коэффициент $p = n + 2 - i$. (В диссертации рассматривается возможность использования и других шкал весовых коэффициентов).

В конечном виде предлагаемая формула генетической близости языков по данным лексикостатистики представляет собой равенство:

$$G(A, B) = \sum[(n+2-i) \cdot V(A, B)_i] / H(A) \cdot H(B),$$

где G — показатель «степени родства», A и B — два данных идиома из всей совокупности исследуемых идиомов, V — число общих для двух данных идиомов слов, восходящих к праязыковому лексическому фонду, H — число всех восстанавливаемых (или привлекаемых к анализу) единиц праязыкового лексического

фонда, отмеченных в данном языке (объем его праязыкового лексического наследия в целом), n — общее число сравниваемых идиомов, i — мощность изоглоссы (число охватываемых ею идиомов).

Пробные применения некоторых стандартных статистических формул меры сходства объектов, в частности, опирающихся на метод четырехклеточной корреляции (коэффициент ассоциации Q , коэффициент сопряженности Φ), показали свою недостаточную эффективность. Они рассматривают отсутствие общего признака у данных двух объектов как их общий признак, что при большом перепаде размеров списков положительных признаков у отдельных сравниваемых объектов (ср. величины N у лужицких или кашубского языков, с одной стороны, и у русского, чешского или сербохорватского, с другой) приводит к смещению результатов; к тому же низкие значения коэффициента Φ признаются в статистике не слишком надежными, тогда как именно они представляют главный интерес в итоговой картине языкового родства.

В главе 4 («Материал исследования. Филологическая критика источника. Оценка статистической достаточности материала») главное внимание уделяется проблеме количества и особенно качества лексического материала, непосредственно кладущегося в основу классификационных и компаративных построений на статистической базе.

Выбор источника материала для лексикостатистического определения степеней генетической близости языков в славянской семье осуществлялся по сути между двумя конкурирующими лексикографическими изданиями — краковским «Słownikiem prasłowiańskim» под редакцией Ф.Славского и московским «Этимологическим словарем славянских языков» под редакцией О.Н.Трубачева (бременский «Etymologický slovník slovanských jazyků» с предварительным вариантом массива полнозначной лексики «Základní všeslovanská slovní zásoba», а также «Vergleichendes Wörterbuch der slavischen Sprachen» Л.Садник и Р.Айцетмюллера по целому ряду причин нашим источником служить не могут). Большая продвинутость издания московского словаря (ЭССЯ), а также ряд качеств, концептуально отличающих его от краковского предприятия (отказ от расписывания материалов этимологических словарей

отдельных языков с уже осуществленной праславянской реконструкцией, отказ от гнездового принципа в пользу установки «словарная статья = отдельное слово», непривлечение в словарь отдельных звеньев регулярных морфологических парадигм, осторожность в констатации праславянской природы тех или иных формально возможных реконструкций, ограниченное привлечение периферийной междометной и ономатописической лексики и др.), сделали его более предпочтительным в наших глазах.

К моменту начала ввода материала ЭССЯ в ЭВМ свет увидел 15-й вып. этого словаря, оканчивавшийся вокабулой **lokašь* (примерно треть всего праславянского лексического списка). Это составляет 7938 словарных позиций, что почти в 80 раз превосходит *basic vocabulary* М.Сводеша и более чем в 12 раз объем списка понятий для количественной таксономии европейских языков у А.Я.Шафкевича. Без сомнения, это предельная величина добротного, строго выверенного праславянского лексического материала, который к моменту начала нашей работы мог быть подвергнут статистическому анализу.

Основной объект нулевого цикла работы — и з о г л о с с а в обобщенном представлении, т.е. фактически позиция в ЭССЯ — список рефлексов праславянской формы. Основанием межславянского сравнения и комплектации материала в границах словарной статьи ЭССЯ является не корень, как это обычно практикуется в этимологических лексиконах, а словообразовательная структура, т.е. предметом рассмотрения является не этимологическое гнездо, а праславянская лексема. Однако принцип цельнолексемного соответствия не может быть проведен с неукоснительностью в силу естественного «сопротивления материала» (диффузность, пересекаемость форм, безусловность некоторых этимологий и т.п.). Отклонения от исповедуемого принципа цельнолексемного соответствия (включение в число корреспондентов заголовочной лексемы форм, находящихся на следующих ступенях деривации: укр. *дарбенний* в статье **darьba*...; подключение нераспространенной формы, не выделяемой в самостоятельную статью, к анализу аффиксального образования: н.-луж. *gjarś* в статье **gьrtanъ*/**gьrtanь*...; объединение морфологических вариантов: **čemerъ*/**čemera*, **dryxati*/**dryxnъti*... — при их разнесении в разные по-

зиции в иных аналогичных случаях: *brězga I, II — brězǝ I, II, *xlestati — *xlestiti...; объединение различных по суффиксальному оснащению дериватов: *leži.čē/*leževisče...; неразделение словообразовательных омонимов: статья *agoda, совмещающая производные от *agoda и словообразовательно вторичные образования от *agodica...; и др.) демонстрируют, что «набор» лексических изоглосс, если последние исчислять, исходя из словника ЭССЯ, является до некоторой степени условностью. Другие чисто лексикографические решения привели бы к констатации отличных изоглосс, часто иной мощности, что небезразлично для конечных результатов квантитативного исследования, поскольку разным по мощности изоглоссам в нашей формуле приписываются разные весовые коэффициенты. В ЭССЯ имеются, с нашей точки зрения, спорные этимологии, уточнение которых (например, **aglъjь/**jaglъjь: русск. диал. яглый можно отнести к «незакономерным» продолжениям *degnoti, ср. параллелизм названий сныти укр. диал. яглиця/дяглиця; с.-хорв. lăčiti правомернее не связывать с в.-луж. tačic so, а относить к праслав. *lqčiti как заимствование из болгарского или македонского; и др.) могло бы, в весьма незначительной, правда, мере, повлиять на статистическую оценку близости языков. Однако в большинстве случаев мы избегали вмешательства в этимологические решения, принятые в ЭССЯ, как если бы они были единственно возможными.

Серьезнее дела обстоят с лакунами в списках славянских рефлексов заголовочной праформы, имеющих в каждой словарной статье и представляющих по сути описание ареалов отдельных слов. Неполнота привлеченного материала может оказать влияние на статистические итоги работы. Обнаруживаемые лакуны мы не склонны относить к недостаткам словаря, поскольку пропуски в лексиконах, подобных ЭССЯ, неизбежны из-за огромного объема материала, заведомой неполноты, неравнокачественности и, главным образом, открытости используемых источников диалектной лексики. Стремясь получить максимально достоверные статистические результаты, мы сочли возможным внести в лексические материалы ЭССЯ дополнения, источниками которых послужили около трех десятков изданий диалектной и литературной лексики, в основном словарей, преимущественно восточнославянских

(русского, белорусского и украинского), болгарского и верхнечужицкого языков. Посильный вклад в устранение лагун в своде лексических изоглосс, каковым помимо прочего является ЭССЯ, осуществлен в нашей работе «К уточнению представлений о славянских изоглоссах. Дополнения к лексическим материалам «Этимологического словаря славянских языков» (чч. I, II. М., 1990), которая дана приложением к настоящей диссертации.

Сделанное нами пополнение списков континуантов славянских праформ коснулось около 1600 словарных статей ЭССЯ (то есть каждой пятой статьи). Некоторые из них позволили отклонить высказанные составителями сомнения в праславянской древности заголовочной праформы (*berь, *lěsatъь...), подозрения к книжной природе отдельных позднеславянских рефлексов (*blg-dьnzь, *kozegogъ...), расширить ряды формальных вариантов праславянских лексем (*bruъь к *bruja, čujadlo к *čudlo, *kostovalъ (очевидно вторичная форма) к *kosti-valъ, *ъzskubati к *ъzskubti...), выявить двучленные эксклюзивные изоглоссы, статистически наиболее значимые в установлении степени близости между языками (*bagnьje, *bajunъ, *basn'ca, *bazuriti, *berzuъь, *bezplodъь, *bezpqtъ, *bělъce, *běšěti, *bit'elъ, *bloščica, *bodadlo, *boliti/*bolěti, *boltuxъ, *buxyръ, *bukarъ, *bzgadlo, *bъrmiti, *d(ъ)věnzьky, *dъrъzъь, *jatъьъje, *katadlo, *kovьka, *kozvonъ(kъ), *krq̇tověja, *krъzъь, *lěsatъь, *lędzьka, *lędzьce и др.).

Еще один род вмешательства в дагные ЭССЯ представляет сегрегационный анализ, обращение к которому вызвано стремлением к тому, чтобы положенный в основу расчетов материал был предельно чист в генетическом отношении. Цель сегрегационных процедур — освобождение материала от результатов возможных послепраславянских взаимовлияний, книжной лексики, потенциально поздних образований, сомнительных параллелей и т.п. — всего, что может вносить искажения в синхронную картину генетической близости (чоздне)праславянских диалектов. Достаточным позодом для констатации небезупречности примера или реконструкции служили явно выраженные сомнения самих составителей ЭССЯ. Прежде всего изымались из подсчетов статьи, в которых выражаются сомнения в праславянской древности заголовочной лексемы; вычеркивались церковнославянские примеры

(«русс.-цслав.», «сербск.-цслав.») как лексика книжных путей распространения (**abrědъ*, **askqđъ*, **ѡjazpъ*...), книжные слова живых языков (укр. *багрий*, русск. *бесчадный*, *клевсѣта*, чешск. *hracňor*...), слова, двуслынные в плане реконструкции (например, западнославянские слова с префиксом *z-*, который может быть возведен и к **ъz-*, и к **sz-* (**ъzbiti*, **ъzboosti*, **ъzdati*...)) и т.д. Изданная часть праславянского словаря сокращена таким образом примерно на 430 единиц (5,4% словника ЭССЯ).

Кроме того, в некоторых случаях устранялась непосредственность составителей относительно принципа цельнолексемного соответствия, и материал ряда статей, объединяющих случаи вариантной вокализации идеофонического корня, морфологические варианты, словообразовательные модификации и т.п., был представлен *divise*, чем усиливался принцип «максимально расчлененной подачи словаря» и выявлялись более «мелкие» изоглоссы, вычлениющие более узкие межъязыковые связи (**orъkati* > ... — **bryhati*, **dadъbogъ* > ... — **da-bogъ*, **gromaditi* > ... — **gromazditi*, **lišajъ* > ... — **liša* и т.п.). Подобное расширение корпуса лексических изоглосс, по видимому, позволило представить различия между славянскими языками статистически несколько более выпуклыми (впрочем, разделением материала таких статей мы не злоупотребляли).

После изъятия из рассмотрения нескольких сотен статей с лексикой сомнительной древности, лексикой книжной природы, а также некоторого расширения корпуса изолекс, образовался список из 7557 позиций (лексических изоглосс), который и был подвергнут статистической обработке.

Перечень идиомов, степень генетической близости между которыми (попарно) выявляется в диссертации, естественно, совпадает с перечнем славянских языков, как они рассматриваются в ЭССЯ: старославянский (ср), болгарский (бг), македонский (мк), сербохорватский (сх), словенский (си), чешский (чш), словацкий (сл), верхнелужицкий (вл), нижнелужицкий (ил), полабский (лб), польский (пл), кашубско-словинский (ке), русский (вр), украинский (ук), белорусский (бр). Хронологические уточнения «древне-» и «старо-» («древнесловенский», «старобелорусский») в расчет не принимались. Исключение сделано, конечно, для старославянского и древнерусского; последний, будучи

предком трех современных языков, выделен у нас в самостоятельную единицу (др). Относительно древнерусского следует также сделать оговорку, что, строго говоря, вся лексика праславянского происхождения, наблюдаемая в современных восточнославянских языках, наличествовала и в древнерусском, если последний понимать как их предшествующее состояние. Здесь же древнерусский трактуется только как язык письменных памятников, приуроченных к определенному времени и месту, что сближает его статус со статусом старославянского языка.

В ЭССЯ имеются пометы «диал(ектное)», которыми в подавляющем большинстве и ограничивается локальная паспортизация региональной лексики всех языков, — но за исключением русского: данные словаря-источника дают возможность представить огромный великорусский этноязыковой континуум более детально, чем остальные идиомы. В его составе мы выделяли два наречия (ср и кр) и, условно, пояс среднерусских говоров (др)

Состав идиомов, подвергавшихся лексикостатистическому анализу, рассматривается в работе в четырех версиях: 1) SLAV — наибольшей по количеству идиомов — с включением древнерусского и старославянского и дифференцированным представлением русского; 2) CON — то же — с исключением книжных древнерусского и старославянского, 3) RUS — с их включением, но представлением русского в виде цельного идиома, 4) MIN — то же — с исключением мертвых книжных. Расчеты делались для каждой версии состава славянских идиомов отдельно.

Объемы праславянского лексического наследия в отдельных идиомах, по данным 1—15 выпусков ЭССЯ, составляют:

сс — 1124 лексемы, бг — 3262, мк — 2035, сх — 4568, си — 3519, чш — 4264, сц — 2933, вл — 1895, ил — 1574, лб — 452, ил — 3350, ке — 1683, др — 2681, вр — 4956 (ср — 4168, кр — 3384, юр — 3833), ук — 3905, бр — 3288 праславянских лексем.

Статистическая надежность привлекаемого лексического материала определялась при помощи расчета объема минимальной выборки для языка с наименьшим доступным реконструкции праславянским словариком (полабского). Пробные проверки частоты встре-

чаемости полабских слов в статьях ЭССЯ дали величины около 0,06. По формуле

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

(см.: Пиотровский — Бектаев — Пиотровская, 1977, с. 295), где N — объем выборки. f — частота (0,06) встречаемости признака, относительно которого определяется N , p — заданная надежность (95%, для которой табличное значение квантиля $z_p = 1,96$), δ — заданная относительная ошибка наблюдения (10%), установлено, что наименьшая необходимая выборка в нашем случае составляет $N_{\min} = 6018$ слов; иначе говоря, объем привлеченного к статистическому анализу материала (7557 изолекс) превосходит минимальную требуемую величину (у нас $N = 1,25N_{\min}$) и может считаться вполне надежным.

В главе 5 («Порядок ввода материала в ЭВМ и его машинной обработки») описываются этапы компьютерного исследования праславянской лексики согласно условиям нашего эксперимента. Работа выполнена на компьютере ЕС-1036 Института русского языка РАН. Серия программ, необходимых для этого, подготовлена ст. науч. сотр. канд. физ.-мат. наук М.В. Ломковской и составляет ее интеллектуальную собственность.

В главе 3 («Из количественно-типологических наблюдений над лексикой славянских языков (праславянское наследие)») показана необходимость типологических уточнений характеристик привлекаемых к анализу языков. Компаративист, занимающийся выяснением природы и силы связей в языковой семье, как правило, имеет дело с неполным и зачастую довольно разнородным материалом, что существенно осложняет прямое сопоставление массивов данных, относящихся к каждому из родственных языков по отдельности. Непринятие во внимание типологических различий между языками (например, объемов их полных праславянских словарей, книжной или устно-диалектной природы, степени изолированности от контактов с родственными языками и т.д.) может приводить к неверно осуществляемым сличениям данных (особенно количественных), а отсюда — к смещенным или даже ошибочным выводам относительно их истории.

Обширный материал ЭССЯ уже сейчас позволяет

сделать оценки величины праславянского лексического корпуса в целом и языковых словников отдельных славянских языков.

Определяя объем праславянского лексического корпуса, необходимо иметь в виду, что речь не идет о словарном «запасе» некоего цельного идиома: словарь праславянского языка, который некогда не был бездиалектным монолитом, может реконструироваться только как теоретико-множественная сумма словарных составов, как объединение множеств праславянских «*słownictw*» более поздних продолжений. Имеющиеся оценки этой величины либо совершенно устарели (Т.Дер-Силавинский, Ф.Копечный и др.), либо основаны на неполном учете особенностей материала, избранного в качестве теста, и к тому же усугублены арифметическими ошибками (Ф.П.Филин и др.). По нашим прикидкам, объем праславянского словаря может достигать 25000 лексем (распространенных крайне неравномерно и в значительной своей массе представляющих языковые регионализмы).

Сведения об объемах праславянских словников отдельных славянских языков, пока гипотетические, несложно получить путем экстраполяции на тот или иной предполагаемый суммарный объем праславянского словаря данных о предьявленности каждого языка (%) в списках континуантов заголовочных праформ в уже опубликованной части ЭССЯ. Экстраполяция на суммарный объем в 25 тысяч лексем дает следующие показатели:

сс (14,87%) — 3,7 тыс. слов, бг (43,17%) — 10,8 тыс., мк (26,93%) — 6,7 тыс., сх (60,45%) — 15,1 тыс., си (46,57%) — 11,6 тыс., чш (56,42%) — 14,1 тыс., сц (38,31%) — 9,7 тыс., вл (25,08%) — 6,3 тыс., нл (20,83%) — 5,2 тыс., лб (5,98%) — 1,5 тыс., пл (44,33%) — 11,1 тыс., кс (22,27%) — 5,6 тыс., др (35,48%) — 8,9 тыс., вр (65,58%) — 16,4 тыс., ср (55,15%) — 13,8 тыс., юр (50,72%) — 12,7 тыс., ук (51,67%) — 12,9 тыс., бр (43,51%) — 10,9 тыс. праславянских слов.

По доле праславянской лексики, сохраненной в каждом славянском языке, их можно условно разделить на группы:

- 1) «крупные» языки — с охватом более 50% прасла-

вянского словаря: вр (оба наречия также относятся к крупным идиомам), сх, чш, ук;

2) «средние» — от 30% до 50%: си, пл, бр, бг, сц, др;

3) «малые» — менее 30%: мк, вл, кс, гл, сс;

4) «реликтовый»: лб.

Просматривается положительная корреляция между долей праславянского лексикона, приходящейся на словарный состав данного языка, и объемом всего его лексикона в целом (оценки величин лексикона отдельных идиомов см. в работах А.Е. Супруна по лексической типологии славянских языков). Наши данные об объемах праславянских словников современных славянских языков существенно отличаются от соотношений, определенных другими исследователями.

Выявление дифференциального слоя праславянской лексики в каждом языке дает возможность оценить меру специфичности его праславянского лексического наследия. В 1—15 выпусках ЭССЯ отмечено (после указания абсолютного числа индивидуальных праславянских слов приводится, в процентах, их доля во всем массиве праславянской лексики, фиксируемой лишь в единичных языках, и их доля, также в процентах, в праславянском словнике данного языка):

сс — 28 (8,7; 2,49), бг — 35 (4,6; 1,07), мк — 0, сх — 111 (14,7; 2,42), ск — 45 (6,0; 1,28), чш — 125 (16,6; 2,93), сц — 17 (2,3; 0,58), вл — 30 (4,0; 1,58), гл — 13 (1,7; 0,83), лб — 5 (0,7; 1,11), пл — 46 (6,1; 1,37), кс — 19 (2,5; 1,13), др — 48 (6,4; 1,79), вр — 189 (25,1; 3,82), ср — 52 (6,9; 1,25), юр — 19 (2,5; 0,50), ук — 31 (4,1; 0,79), бр — 12 (1,2; 0,36).

Наибольшую лексическую специфичность по отношению к праславянскому фонду обнаруживает русский язык, наименьшую — македонский. Если список славянских идиомов упорядочить по убыванию доли дифференциальной лексики в праславянском словнике каждого языка, то в нем можно усмотреть определенную тенденцию: к началу списка тяготеют идиомы, которые характеризуются достаточно ранним диалектным обособлением, ранним осознанием себя носителями данного идиома как выделенной этнической общности, давней историко-культурной традицией, включая собственную письменность, и т.п. Напротив, к его концу сосредоточиваются лингвистические образования более позднего с историко-культурных позиций харак-

тера: словацкий, белорусский, македонский, великорусские наречия, не ставшие самостоятельными языками. Иными словами, ранняя историко-культурная специализация идиом с необходимостью предполагает высокую долю индивидуальной лексики в сохраняемом им праязыковом наследии.

Индивидуальность целых славянских групп выражается такими показателями: 6,3% всего праславянского лексикона составляет исключительно южнославянская лексика, 6,4% — исключительно западнославянская, 8,7% — исключительно восточнославянская.

В главе 7 («Картина родства славянских языков по данным лексикостатистики») приведены результаты статистического анализа праславянской лексики по предложенной нами формуле и их интерпретация.

Информативность абсолютных показателей лексических связей между языками (количество общих для них изоглосс) невелика: они обнаруживают чрезвычайно высокую скоррелированность с величинами N у обоих сравниваемых языков. Весьма наглядно зависимость абсолютных показателей лексических связей между языками от величин их праязыковых словников выступает при обращении к эксклюзивным параллелям — материалу, который чаще всего привлекают для доказательства особых отношений между данными идиомами в прошлом, как свидетельства их тесных контактов. Наибольшее количество слов праславянского происхождения, связывающих данный язык с каким-либо другим языком, — у русского (375 по 1—15 вып. ЭССЯ), величина N которого самая высокая из всех славянских языков, далее идут сербохорватский (268 слов), чешский (235), словенский (135), украинский (125), т.е. наиболее «крупные»; замыкают список «малые» старославянский, македонский и «реликтовый» полабский. Неудивительно, что сильнейшие сепаратные связи в абсолютном численном выражении характеризуют именно русский и сербохорватский (74 лексемы из 7557, рассмотренных в нашем анализе), хотя они принадлежат и разным славянским группам. В то же время белорусский и украинский, близкое родство которых сомнения не вызывает, на том же пространстве связываются всего 15-ю исключительными изолексами, а болгарский и македонский — 14-ю. Из этого следует сделать вывод, что при

нестатистических оценках взаимной близости данных идиомов, при констатации их особых отношений в прошлом необходимо учитывать их квантитативнотипологические особенности: большое количество словарных сходжений между языками не непременно указывает на наличие в их истории интенсивного эксклюзивного взаимодействия. В подсобных случаях постулирование особо тесных взаимоотношений языков нуждается в дополнительных аргументах, находимых в исследовании лексики неквантитативными методами.

Не слишком информативным является и относительное, в процентах к величине праславянского лексического наследия данного языка, выражение его лексических связей с другими идиомами. Тем не менее и по этим данным можно сделать некоторые наблюдения, касающиеся исторических взаимоотношений между языками. Так, обнаруженная нашими наблюдениями закономерность, согласно которой в связях с третьим языком участвует больший процент праславянского словника того из двух данных ближайшеразродственных языков, который характеризуется меньшим объемом словника, выдерживается очень последовательно, но имеет одно исключение. Оно касается украинского и белорусского языков в их отношениях к южнославянским и чешскому. При равновесном распределении сил связей ожидался бы более высокий процент лексики, участвующей в связях с языками южнославянской группы; у белорусского, объем праславянского лексического наследия N которого меньше, чем N украинского языка. Однако закономерность нарушена «в пользу» украинского. Это может свидетельствовать о более тесных исторических взаимоотношениях между протоукраинским компонентом праславянского языка и праславянскими диалектами, легшими в основу южнославянской ветви и чешского языка.

Главный массив численных данных, полученных в диссертации и на которые опираются ее основные выводы, — индекс генетической близости (G). Этот индекс, как отмечалось, рассчитывался для четырех версий состава славянских идиомов. Результаты расчетов в разных версиях различаются между собою в некоторых деталях, однако в общем оказываются весьма сходными. Здесь приводятся значения индекса G для версии SLAV (с максимальным составом, т.е. представляющей русский язык в виде трех идиомов):

старославянский & бг 1,0912, мк 1,1878, сх 1,0189, си 1,0433; чш 0,8262, сц 0,8035, вл 0,8673, нл 0,9189, лб 1,0748, пл 0,8363, кс 0,7982; др 1,4999, ср 0,8512, цр 0,8438, юр 0,8350, ук 0,7921, бр 0,7011;

болгарский & мк 1,7274, сх 1,2674, си 1,2023; чш 0,9952, сц 0,9807, вл 0,8968, нл 0,9228, лб 0,7347, пл 0,9318, кс 0,8461; др 0,9707, ср 0,9344, цр 0,9308, юр 0,9204, ук 0,9672, бр 0,8941;

македонский & сх 1,2619, си 1,2430; чш 0,9321, сц 0,9884, вл 0,9105, нл 0,8644, лб 0,8785, пл 0,8653, кс 0,8846; др 0,9682, ср 0,8759, цр 0,8791, юр 0,8602, ук 0,9103, бр 0,8342;

сербохорватский & си 1,2968; чш 1,0355, сц 0,9765, вл 0,8788, нл 0,8039, лб 0,6643, пл 0,9147, кс 0,8605, др 0,9741, ср 0,9580, цр 0,9466, юр 0,9334, ук 0,9604, бр 0,8904;

словенский & чш 1,0770, сц 1,0581, вл 1,0064, нл 0,9334, лб 0,8071, пл 0,9546, кс 0,8865; др 0,9609, ср 0,9360, цр 0,9225, юр 0,9071, ук 0,9873, бр 0,8949;

чешский & сц 1,3622, вл 1,1001, нл 1,0291, лб 0,7375, пл 1,1629, кс 0,9931; др 0,8973, ср 0,9087, цр 0,9091, юр 0,9285, ук 1,0273, бр 0,9595;

словацкий & вл 1,1847, нл 1,1693, лб 0,8282, пл 1,1994, кс 1,1403; др 0,8551, ср 0,8875, цр 0,9013, юр 0,9278, ук 1,0841, бр 1,0290;

верхнелужицкий & нл 1,9349, лб 1,1545, пл 1,1232, кс 1,2744; др 0,8287, ср 0,8205, цр 0,8109, юр 0,8556, ук 0,8459, бр 0,9401;

нижнелужицкий & лб 1,3477, пл 1,1383, кс 1,3829; др 0,8681, ср 0,8021, цр 0,8043, юр 0,8428, ук 0,9090, бр 0,9146;

полабский & пл 0,8972, кс 1,2961; др 0,8209, ср 0,6593, цр 0,6893, юр 0,6975, ук 0,6909, бр 0,6900;

польский & кс 1,3828; др 0,9968, ср 0,9008, цр 0,9825, юр 1,0262, ук 1,1219, бр 1,1137;

кашубско-словинский & др 0,8732, ср 0,8245, цр 0,8475, юр 0,8741, ук 0,9553, бр 0,9907;

древнерусский & ср 1,1661, цр 1,1715, юр 1,1582, ук 1,0553, бр 1,0929;

северновеликорусский & цр 1,5265, юр 1,4265, ук 1,1639, бр 1,2254;

среднерусские говоры & юр 1,5443, ук 1,2055, бр 1,2839;

южновеликорусский & ук 1,2691, бр 1,3744;

украинский & бр 1,3745.

Графическую интерпретацию статистических связей между славянскими идиомами см. на прилагаемой к автореферату схеме I.

Южнославянская группа.

Высокие значения G для старославянского в его связях с древнерусским должны объясняться не столько лексическим влиянием одного языка на другой, сколько сравнительно низкими у них значениями H и связанной с этим невысокой индивидуализированностью праславянского лексического наследия как в одном, так и в другом. Вероятно, той же природы неожиданно сравнительно высокий индекс лексикостатистического сходства старославянского с полабским языком: в скромном праславянском словнике последнего в долевом отношении преобладает общеславянская лексика.

Сравнение лексикостатистических связей болгарского и македонского с третьими языками демонстрирует большую яркость восточнославянских связей первого и западнославянских связей — второго. Сравнение в том же плане сербохорватского и словенского обнаруживает устойчиво более сильные связи последнего с языками западнославянской ветви, у сербохорватского же — с русским (но не украинским и белорусским).

В целом южнославянская группа на обследованном участке словника ЭССЯ имеет 5591 лексему (74,0% всего праславянского лексического корпуса). Из этого числа общеславянскими или находящими соответствия по крайней мере в одном из языков западной и восточной групп являются 91,6% всей южнославянской лексики праславянского возраста, что отличается, например, от оценок Л.В.Куркиной (70—80%). Общими для обеих южнославянских подгрупп являются 58,0% южнославянской лексики, 8,4% встречаются в болгарско-македонской, но не отмечены в сербохорватско-словенской подгруппе, 33,7% известны сербохорватско-словенской, но отсутствуют в болгарском и македонском.

Западнославянская группа.

Чешский и словацкий языки, характеризуясь значительной лексической близостью на праславянском уровне, ведут себя по-разному в отношениях с остальными славянскими языками. Южнославянские статистические связи чешского (за исключением македонских) несколько сильнее, чем у словацкого; то же можно сказать и об их связях с древнерусским и велико-

русским. Что же касается других западнославянских языков, украинского и белорусского, то к ним обнаруживается большую близость словацкий. Из южнославянских к чешскому и словацкому ближе прочих словенский, из восточнославянских — украинский.

Серболужицкие языки отличается наибольшая теснота лексикостатистических связей среди всех возможных пар славянских идиомов. Однако не следует, на наш взгляд, приписывать эту близость их генетическому единству (выводимости из общего языка-предка), которое нередко оспаривается. Отношения с третьими языками у лужицких складываются несходно: верхнелужицкий демонстрирует более высокие показатели связи с чешским и словацким, в то время как нижнелужицкий сильно тяготеет к лехитским языкам. Внезапнославянские связи верхнелужицкого выглядят гораздо более выпуклыми, чем у нижнелужицкого, который более плотно вписан в западнославянскую группу по сравнению с верхнелужицким.

Ситуация с лехитскими языками (высокие связи кашубско-словинского и с польским и с полабским при существенно более скромном показателе лексикостатистической близости между двумя последними языками) наталкивает на мысль о генетическом неединстве этой подгруппы. Неодинаково складываются отношения лехитских языков с иными языковыми группировками. В рамках западнославянской группы кашубский и полабский сильнее тяготеют к лужицким языкам, польский — к чешскому и словацкому. Если польский и кашубско-словинский во внезапнославянских связях сильнее тяготеет к восточнославянским языкам, то у полабского над восточнославянскими преобладают южнославянские статистические связи (хотя их показатели очень низкие вообще). При этом полабский сильнее связывается с болгарско-македонской подгруппой, чем с сербохорватско-словенской, тогда как южнославянские тяготения польского и кашубского дают обратную картину.

Восточнославянская группа.

Из современных восточнославянских языков со своим предком, древнерусским, слабее остальных связан украинский, теснее всех — русский (при представлении же его в виде трех идиомов — северновеликорусское наречие).

Два наречия русского языка связаны между собою менее мощной связью, чем, например, болгарский с македонским или два лужицких языка. Оба они лексически ближе к белорусскому, чем к украинскому языку. За пределами восточнославянской группы у северновеликорусского наречия самые сильные лексикостатистические связи — с сербохорватским языком, у южновеликорусского — с краславянской лексикой польского языка. Как целое русский язык испытывает большее тяготение к южнославянской группе, а внутри нее — к западной (сербохорватско-словенской) подгруппе.

Взаимная лексикостатистическая близость украинского и белорусского языков выражается более высокими цифрами, чем в парах польский — кашубско-словинский, чешский — словацкий и тем более сербохорватский — словенский. Внутривосточнославянские связи у белорусского статистически более значимы, чем для украинского. У украинского языка внешние (южнославянские и западнославянские) связи приблизительно сбалансированы, у белорусского преобладают западнославянские. Почти все вневосточнославянские связи украинского выражаются более крупными числами, чем у русского языка. Сравнение его с поведением белорусского и русского во вневосточнославянских тяготениях также демонстрирует «экстравертность» украинского языка в противоположность более «автохтонному» (восточнославянскому) лексическому облику ближайших соседей.

Ближайшеродственные идиомы, входящие в одну языковую подгруппу, нередко обнаруживают, как можно было заметить, различия в «предпочтениях», отдаваемых во внешних лексикостатистических связях тем или иным языкам или их группировкам, находящимся в сравнительно отдаленном родстве. Достаточно показательными в этом отношении являются статистические сведения о лексических связях данного языка, в которых не принимают участия другие языки той же группы. В целом эти сведения подтверждают суждения, сделанные на основе наблюдений над индексом G. У сербохорватского языка в лексических связях, в которых не участвуют другие южнославянские языки, восточнославянская «ориентация» (197 лексем в первых пятнадцати выпусках ЭССЯ) преобладает над западнославянской (141 лексема). Внешние «ориентации» болгарского и словенского языков (приблизитель-

но сбалансированы. О македонском судить трудно в силу статистической недостаточности такого материала. Чешский и верхнелужицкий языки в своих внезападнославянских связях явно склоняются к южнославянской группе (169 лексем против 126 у первого, 22 против 13 у второго), польский и нижнелужицкий, напротив, — к восточнославянской (116 против 43 и 16 против 8 соответственно). «Склонности» словацкого и кашубско-словинского выражены неярко. Древнерусский и великорусский языки обнаруживают большее тяготение к южнославянским языкам, чем к западнославянским (65 против 37 у древнерусского, 227 против 140 у великорусского). У украинского южнославянские параллели, в которых не участвуют другие восточнославянские языки, почти уравниваются западнославянскими (79 и 76), белорусский же больше схож в праславянском лексиконе с западнославянской группой (33 лексемы против 22 белорусско-южнославянских лексических параллелей).

В главе 8 («О возможности лексико-статистического обнаружения явления языковой конвергенции») содержится попытка решить проблему, имеющую весьма 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 количественные методы исследования, однако в специальных лингвостатистических работах почти не обсуждается. Более того, почти всегда цифры, полученные с помощью статистического анализа изоглосс, безоговорочно объявляются данными, отражающими ситуацию, которая сложилась в результате неуклонных дивергентных процессов, и тем самым являющимися прямым мерилом родства. В особенности это характерно для работ в рамках глоттохронологического направления: согласно концепции глоттохронологии, чем выше численное выражение сходства между языками, тем позже они разошлись и тем теснее их и з н а ч а л ь н о е родство. Сама постановка вопроса о конвергенции как факторе родства уязвляет самую суть диахронических построений, базирующихся в конечном счете на концепте родословного древа и последовательного распада праязыковых единств.

Возможность лексикостатистического обнаружения роли конвергентных процессов в истории данных языков или группировок, на наш взгляд, имеется в фронтационной статистике разнонаправленных изоглосных связей языков, отмеченных высоким значением индекса родства (С).

Сравним две пары славянских языков, составляющие которых (пар) связаны между собою высокими показателями родства: болгарский & македонский и польский & кашубско-словинский. Для каждой из этих пар выясним число общих для них праславянских слов ($V(A,B)$), число слов, встречающихся в одном, но неизвестных другому ($V(A,-B)$), и число слов, известных второму, но отсутствующих в первом ($V(-A,B)$), причем будем различать эти изоглосные связи по числу всех охватываемых ими идиомов, а не только двух данных (m , от 2 до произвольно назначенной, но невысокой, мощности, например, до 7 идиомов):

бг & мк

пл & кс

m	$V(A, B)$	$V(A, -B)$	$V(-A, B)$	$V(A, B)$	$V(A, -B)$	$V(-A, B)$
2	14	137	8	15	138	51
3	74	233	33	34	315	83
4	163	438	52	64	499	112
5	262	583	72	110	699	152
6	364	728	99	167	888	169
7	494	883	116	240	1089	191
...

Количественное соотношение изоглос, охватывающих оба данных языка, и изоглос, в которых они участвуют порознь, можно представить в виде формулы:

$$q = V(A, B) / m \sqrt{V(A, -B) \cdot V(-A, B)}$$

Различия в численных значениях условной величины q для разных пар могут быть интерпретированы как различия в степени изначальной генетической близости языков: чем выше значение q в строке с данным значением m , тем вероятнее исконная близость между языками этой пары; чем q ниже, тем вероятнее роль конвергенции в их эволюции. Последовательность значений q для ближайшеродственной пары бг & мк при значениях m от 2 до 7 выглядит следующим образом:

0,211 — 0,255 — 0,270 — 0,256 — 0,206 — 0,220;
 для пары пл & кс:

0,089 — 0,070 — 0,068 — 0,067 — 0,072 — 0,075.

Легко увидеть несходство этих числовых последовательностей. Из прочих пар, составляемых ближайшерадственными славянскими языками (сх & зн, чш & сц, вл & нл, вр & ук, вр & бр, ук & бр) низкие значения q , сходные с тем, что наблюдается в соотношениях польского и кашубско-словинского, характерны для пары сербослужичских языков; высокие значения q — для остальных. Языки, заведомо не составляющие ближайшерадственных пар (с низкими значениями индекса G), дают значения q , сходные с теми, которые наблюдаются у пар вл & нл и пл & кс, или еще более низкие. Разумно предположить, что лексическая близость пары серболужицких языков и пары польский — кашубский вызвана вторичным сближением, в то время как у остальных ближайшерадственных пар высокие значения G обусловлены главным образом теснотой их связей (происхождением из одного, промежуточного, праязыка).

Изложенная методика может считаться эффективной только при условии, что опирается на конкретные изоглоссы, описываемые достаточно большими числами. Практически это означает, что она применима главным образом (а может быть, только) к лексическому материалу. Следовательно, косвенный результат наблюдений в этой области состоит в том, что статистика фонетических и грамматических изоглосс более или менее глубокой древности — инструмент довольно грубый: от мелких и частных процессов постоянного диалектообразования мощные волны последующих грандиозных фонетических и морфологических катастроф оставляют слишком мало следов. Лексика же, будучи рыхлой системой, не подверженной тотальному нивелированию, способна дольше хранить память о прежних диалектных размежеваниях и единствах, хотя поспешно абсолютизировать преимущества лексики как объекта статистического анализа глоттогенетической направленности и не следует.

В главе 9 («Проблема древновгородского диалекта *sub specie* лексикостатистики») методика, разрабатываемая в диссертации, применена к решению вопроса, в последнее время обострившегося в связи с лингвистичес-

ким исследованием А.А.Зализняком новгородской бересты. По мнению А.А.Зализняка, будучи связанным в севернокривичском своем компоненте с севером западнославянской зоны многими общими явлениями в области фошетики, грамматики, словообразования и лексики, а также — в другом своем компоненте, ильменско-словенском, — обнаруживая, как и остальные восточнославянские идиомы, много общего с южнославянской зоной, прежде всего сербохорватско-словенской, древне-новгородский диалект, отраженный берестяными грамотами, разрушает устойчивые представления о моногенезе восточнославянской языковой группы.

Сравнение данных Словоуказателя к новгородским берестяным грамотам с лексическими материалами изданной части ЭССЯ выявило различные пласты праславянской лексики в словнике древне-новгородского диалекта. Среди них — слова, связывающие древне-новгородский с западно- и южнославянскими языками, но не встречающиеся в других восточнославянских идиомах (**dobrovitz*, **dorgobqđz*, **jaromirz*, **kleščь* 'лещ', **krqřz* (*ju*), **kyselz* (*ju*)...), не отмеченные в живых восточнославянских языках (**a ti*, **bebrz*, **boguslavz*, **borislavz*, **čęđь*...), вместе с соответствиями в других восточнославянских идиомах — корреспонденции с южнославянскими, но не с западнославянскими (**bajanz*, **blęđь*, **bl'usti*, **čьrьmьni ca*, **gostьba*, **dorz*, **kakovz* (*ju*), **konьčati*, **koževьnikъ*...), с западнославянскими, но не южнославянскими (**bęgt'i*, **ьrtь*, **gaiqza*/*galqzь*, **gorislavz*, **gostęta*, **ьstьbьzka*, **ьz-vętati*, **korbьzka*, **laditi* (*sę*), **legati* (*sę*) II...), только в восточнославянских (**bersto*, **četvьrgz*, **deševz* (*ju*), **domazirz*, **godььь*, **gzьkъ* **kol(o)bz*, **koporyje*, **koręka*, **koš-čьь*, **kriti*, **kzltьkъ*...).

Однако в берестяных грамотах отмечено всего лишь около 5% праславянских слов, заключенных в диапазоне обследованных нами выпусков ЭССЯ. Индексы родства (G) древне-новгородского диалекта в парах с другими славянскими идиомами, вычисленные на этой статистической основе, не могут считаться достаточно надежными, так же как и коэффициенты Q и Ф.

Можно попытаться извлечь из этих результатов более определенную информацию. Отнеся величины индекса G для древне-новгородского диалекта с восточнославянскими языками к их средней величине G и проделав ту же операцию с данными индекса G, характеризующих остальные восточнославянские идиомы,

вычислим меру сходства между составляющими восточнославянской группы в их поведении по отношению к восточнославянским языкам с помощью коэффициента корреляции.

Древнерусский язык, три главных диалектных единицы великорусского и украинский язык коррелируют между собою по указанному признаку положительно, достигая максимума ($r = +0,975$) в паре северновеликорусский & среднерусские говоры; белорусский положительно коррелирует с украинским, южновеликорусским наречием и среднерусскими говорами, отрицательно — с северновеликорусским наречием и древнерусским языком; древненовгородский диалект со всеми восточнославянскими идиомами коррелирует по указанному признаку отрицательно.

Таким образом, восточнославянская группа в заданном нами составе идиомов изображается довольно компактным единством, лишь древненовгородский диалект выглядит инородным телом (см. схему II). Конечно, обособленность древненовгородского диалекта в результатах этого опыта сильно заострена принятием во внимание лишь внешних «ориентаций», но это и было целью описанной процедуры.

Допустимо думать, что различия между «стандартным» (Зализняк) древнерусским языком и его потомками, с одной стороны, и древненовгородским диалектом, с другой, выявляемые лексико-статистическими процедурами, отражают неомогенность 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 (а тем самым — отчасти — и северновеликорусского наречия) ввиду заметного количества лексических переключек с южнославянскими языками, серьезных оснований не имеется.

В главе 10 («Сопоставление с результатами глоттохронологии») картина славянского языкового родства, полученная на основании лексико-статистического анализа праславянского словаря, сравнивается с аналогичной картиной, воссозданной М.Чейкой и А.Лампрехтом (1963) в рамках глоттохронологических

процедур, т.е. с опорой на стословный диагностический список М.Сводеша.

Наблюдаемые сходства результатов глоттохронологического и нашего моделирования системы родства славянских языков довольно банальны.

Интереснее расхождения. У Чейки — Лампрехта математическое расстояние между чешским и словацким языками принадлежит к наименьшим, является равным расстоянию между двумя лужицкими языками, тогда как в нашей модели связи чешского и словацкого выглядят гораздо более слабыми. Близость чешско-словацкой подгруппы к серболужицкой в глоттохронологической модели столь велика, что указанные четыре языка образуют единство, превосходящее своей спаянностью любые парные, а тем паче более крупные по числу компонентов объединения языков, кроме пары украинский — белорусский. Это сообщает западнославянской группе высокую степень компактности (правда, уступающую компактности восточнославянской группы). Однако у Чейки — Лампрехта не учитывались полабский и кашубско-словинский языки. Будь последние включены в анализ, возможно, мера единства западнославянской группы была бы ближе к тому, что получено нами. В противоположность нашим результатам, южнославянская группа рисуется глоттохронологическими измерениями весьма чеплотным образованием — ввиду очень слабой связи между словенским и болгарским (шестнадцать парных связей между языками, принадлежащими разным группам, оказываются у Чейки — Лампрехта сильнее, чем статистическая связь двух этих языков, входящих в одну и ту же группу).

Усреднение коэф.циентов показывает, что в глоттохронологической картине наиболее тесно связаны между собою западнославянская и восточнославянская группы (среднее расстояние $\bar{d} = 12,4$), самая слабая связь — между южнославянской и восточнославянской (среднее $\bar{d} = 14,3$), южная и западная ветви характеризуются средним расстоянием $\bar{d} = 12,8$. В наших данных наибольшая близость отличает южнославянскую и восточнославянскую группы (среднее $\bar{G} = 0,9338$), далее следует связь западнославянской и восточнославянской групп (среднее $\bar{G} = 0,9195$), слабая — связь между южной и западной группами (среднее $\bar{G} = 0,8958$). Близость результатов Чейки — Лампрехта концепциям Ф.Мареша и А.Фурдаля с констатацией

особого севе́рославянского единства, противостоящего раннему праславянскому диалекту, легшему в основу южнославянских языков («горизонтальное» членение), возможно, объясняется отсутствием в глоттохронологической модели чехословацких ученых двух лексических языков — полабского и гашубско-словинского: их включение в анализ могло привести к сложению отличных результатов. Эта причина, а также недостатки методологического характера, рассмотренные при критическом анализе глоттохронологической концепции в гл. 2, заставляют относиться к итогам работы Чейки — Лампрехта с большой осторожностью.

В главе 11 («Сопоставление с результатами статистического определения родства на материале сравнительной фогетики») ставится вопрос о мере согласованности лексической и фонетической эволюции славянских языков и способах ее выявления. Из имеющихся статистических моделей славянского языкового родства, построенных на материале сравнительной фогетики, наиболее подходящей для сопоставления с лексикостатистическими данными, полученными нами, представляется модель В.Н.Чекмана — О.С.Широкова (1962).

По составу обследуемых идиомов фоностатистическая модель Чекмана — Широкова достаточно близка нашей (более детализованно, чем у нас, рассматриваются сербохорватско-словенская подгруппа и болгарский язык). В основу статистики положен список из 38 изоглосс, оразивших славянские фонетические процессы «с *i* в. по XIV в.» (от палатализации *kv, gv > cv, dzv* и упрощения *tl, dl > l* до прояснения редуцированных во флексиях прилагательного и стабилизации акцента). Квантитативная обработка материала осуществлена применением коэффициента сопряженности Φ (у авторов он обозначен как коэффициент корреляции R).

Фоностатистическая 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. Резко выраженная «островная» структура системы славянского языкового родства, с сильным обособлением одних группировок идиомов от других, отличается от лексикостатистической картины родства, рисуя-

щей отношения славянских языков гораздо более «вязкими». Перепад между средней внутрigrупповых индексов родства и средней межподгрупповых индексов в фоностатистической модели составляет 0,505, тогда как аналогичный перепад в лексикостатистической модели (численные данные нормированы таким же способом) описывается величиной 0,342, т.е. в полтора раза меньше: членение на группы в модели Чекмана — Широкова является намного более рельефным.

Причины указанного различия следует искать не столько в системах описания фактов и их квантитативного препарирования (расхождения в наборе идиомов, устройство формул количественного анализа, несовпадение шкал индексации, частичное дублирование фонетических изоглосс — учет некоторых дополнительно распределенных противопоставленных рефлексаций и т.д.), сколько в самом материале — в несходстве между лексикой и фонетикой и различных механизмах их эволюции (о чем писалось в гл. 1). Лексикон, огромный в сравнении с другими языковыми уровнями по числу составляющих элементов, нежесткий в своей организации и эволюционирующий на основе кумулятивного принципа, может отражать последствия весьма древних междиалектных взаимодействий, консервируя словарные корреспонденции в относительно далеких идиомах. Отсюда — пестрота лексических связей праязыкового характера, столь многочисленных, что это приводит к отяжеленной «вязкости» схемы, отображающей эти связи. В противоположность тому компактная, с жестким сцеплением относительно малочисленных элементов, фонетическая система подвержена лавинным самореорганизациям, радикально меняющим ее облик, стирающим следы разнообразных диалектных вариаций и результаты «мелких» диалектообразующих процессов. Относительная немногочисленность фонетических корреспонденций между языками и их тотальность делают такую же схему очень стройной и четко вычлняющей объекты сравнения — (поздние) праязыкские диалекты и формирующиеся на их основе самостоятельные языки. Статистическая интерпретация фонетических корреспонденций между родственными идиомами приводит к сложению весьма сходной дискретной картины, что и обнаруживается в фоностатистической модели Чекмана — Широкова.

Из важных сходств в результатах фоно- и лексикос-

татистического моделирования необходимо отметить низкие показатели внутренней связанности западнославянской языковой группы по сравнению с двумя другими, а внутри нее — языков, образующих лехитскую подгруппу. Квантитативные оценки и фонетического и лексического родства заставляют констатировать ярко выраженную неомогенность западнославянской группы (на одном уровне) и лехитской подгруппы (на другом).

Сравнение данных обеих моделей, возможное при их сведении к однотипным шкалам, показывает, что идиомы, демонстрирующие значительную близость по данным статистики фонетических изоглосс, вовсе не непременно будут характеризоваться высокими показателями лексикостатистического сходства, и наоборот. Из этого следует вывод о том, что лексическая и фонетическая эволюции праславянских диалектов, приводящие к образованию самостоятельных языков, протекают несогласованно. В отношении болгарского языка с лехитской подгруппой, словенского — с восточнославянской, словацкого — с сербохорватско-словенской, польского — с болгарско-македонской, северно- и южновеликорусского — с лужицкой, белорусского — с сербохорватско-словенской и лужицкой более значимо статистическое фонетическое сходство. Лексикостатистическое сходство выражено сильнее фоностатистического в отношении болгарского — с чешско-словацкой подгруппой, словацкого — с лехитской, польского — с чешско-словацкой, северновеликорусского — с болгарско-македонской, украинского и белорусского языков — с чешско-словацкой и лехитской подгруппами.

Усреднение фоностатистических индексов, очерчивающих внешние связи отдельных языков, показывает большую близость между восточнославянской и южнославянской группами (среднее $\bar{R} = + 0,095$), чем между составляющими двух других пар (зап.-слав. & вост.-слав.: среднее $\bar{R} = + 0,028$; юж.-слав. & зап.-слав.: среднее $\bar{R} = - 0,190$). Это совпадает с результатами аналогичных измерений на лексикостатистической базе. Однако в целом такая картина еще не дает несомненных оснований судить о наличии в начале диалектного «раскола» праславянского языка единого восточного диалекта, разделившегося впоследствии на восточнославянскую и южнославянскую ветви и противопоставленного западному. Признанию монолитности или хо-

та бы выделенности раннего «восточного диалекта» праславянского языка препятствует неравновесность южнославянских статистических связей восточнославянских языков. Связи последних с болгарским и македонским существенно слабее связей с сербохорватско-словенской зоной (и уступают также чешско-словацким, а фоностатистически — и лужицким связям).

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.

В заключительной главе 12 («К реконструкции диалектного членения позднепраславянского языка») попытка представить целостную картину взаимоотношений между протоидиомами современных славянских языков (в границах позднепраславянского состояния) предпринимается на иных, по сравнению с лингвистической традицией, основаниях.

Оперирование исключительно праславянскими словниками современных языков позволяет представить отношения взаимной близости между ними как непосредственное продолжение системы отношений между диалектами праславянского языка, сложившейся к началу его «распада» (осторожно говоря, 2-я пол. I тыс. по Р.Х.), — время, на которое преимущественно ориентирован основной источник работы — ЭССЯ. Однако линейный принцип соответствия между современными идиомами (группами и подгруппами языков, языками, крупными диалектными массивами типа русских наречий) и протоструктурами в праславянском, когда современному украинскому языку в праславянском соответствует некий относительно цельный «протоукраинский» компонент, современному словацкому — «протословацкий» и т.д., не вполне точно отражает реальность и должен быть заменен 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 неперменной ориентации на дендроидную структуру, т.е. с частичным наложением различных образующихся идиомов («квазидиалектов» праязыка) друг на друга. Упомянутые «квазидиалекты» вычлениаются не с опорой на перечисление 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 входящих в данный круг поздних родственных идиомов.

В традиционных классификациях попарные отношения между идиомами, находящимися на одной и той же ступени последовательного членения, но принадлежащими разным ветвям классификационного древа (например, полабско-македонские или нижнелужицко-белорусские), как бы молчаливо уравниваются друг с другом, что, как кажется, не адекватно реальной картине связей внутри славянского языкового мира. Системе родственных идиомов искусственно навязывается не сказать априорная, 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.

Предлагаемый же вариант схемы славянского языкового родства, который, строго говоря, не относится к классификациям, может показать статистическую значимость специальных «далеких» межславянских корреспонденций, несопадающие силы этих связей для идиомов, занимающих одни и те же ступени в традиционной некантитативной иерархической схеме. Один и тот же язык в предлагаемом варианте схемы одновременно включается в группировки разного состава и разной компактности (которая измеряется силой минимальной статистической связи в конкретной группировке), см. схему III. Количество таких вхождений не задается априорной иерархией, а определяется заданной «подробностью» шкалы минимальных сил связей: чем меньше перепад между соседними ступенями шкалы, тем больше возможных группировок и тем сложнее и дифференцированнее конечная картина. Так, например, словаций язык при установленном пороге образования группировки язы-

ков в $G = 1,30$ объединяется только с чешским языком; при пороге $G = 1,15$ он одновременно входит в группировки «сц + чш + пл», «сц + вл»; с понижением порога вхождения до $G = 1,00$ словацкий включается уже в пять группировок с одинаковой мерой компактности (минимальной силой связи внутри них): «сц + чш + вл + нл + пл», «сц + вл + нл + пл + кс», «сц + чш + си», «сц + чш + пл + ук» и «сц + пл + ук + бр». Таким образом, особенностью получаемой таксономической картины является ее не дискретный, а континуальный характер.

С понижением порога образования группировок («квазидиалектов» праславянского языка при диахронической реинтерпретации) они разрастаются по количеству членов (поздних идиомов, от которых ведется отсчет) и постепенно утрачивают отчетливость противопоставления друг другу, отличаясь между собою одним-двумя составляющими. Принятие в качестве такого порога самой слабой связи в списке всех возможных пар идиомов выделяет в сию совокупность рассматриваемых родственных идиомов и делает допустимой трактовку этой совокупности как цельного идиома (цельного, однако, только при ограничении лишь внутриславянскими 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 идиомов, что, как нам представляется, уже само по себе должно снимать проблему диалектной монолитности праславянского; последняя тема, впрочем, по понятным причинам в диссертации совершенно не затрагивается).

Разумеется, как и всякая схема, эта картина является лишь аппроксимацией реального положения вещей: чем сильнее статистические связи между идиомами, тем вероятнее их группировки как протоединства, реконструируемые для поздних периодов развития праславянского языка (с оговорками относительно выявленных случаев конвергенции типа единства сербо-лужицкой подгруппы), чем они слабее, тем большей усложностью становятся получаемые «квазидиалекты», относимые к более ранним периодам. Однако на этом

этапе лексикостатистических исследований диахронического плана существенным представляется не столько достижение высокой степени соответствия между глотогенетической конструкцией и исторической реальностью, сколько уточнение самих механизмов дивергенции. Эволюция праславянского языка по данным лексики рисуется иной, чем ее представляют данные сравнительной фонетики: не резкое диалектное размежевание с четкими линиями разломов, а постепенное и медленное «расползание» в стороны с долгим сохранением следов прежней близости.

Излишне, однако, говорить, что лексикостатистическая картина родства не «отменяет», но только дополняет представления, сложившиеся на основании анализа иных языковых уровней, главным образом фонетического.

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

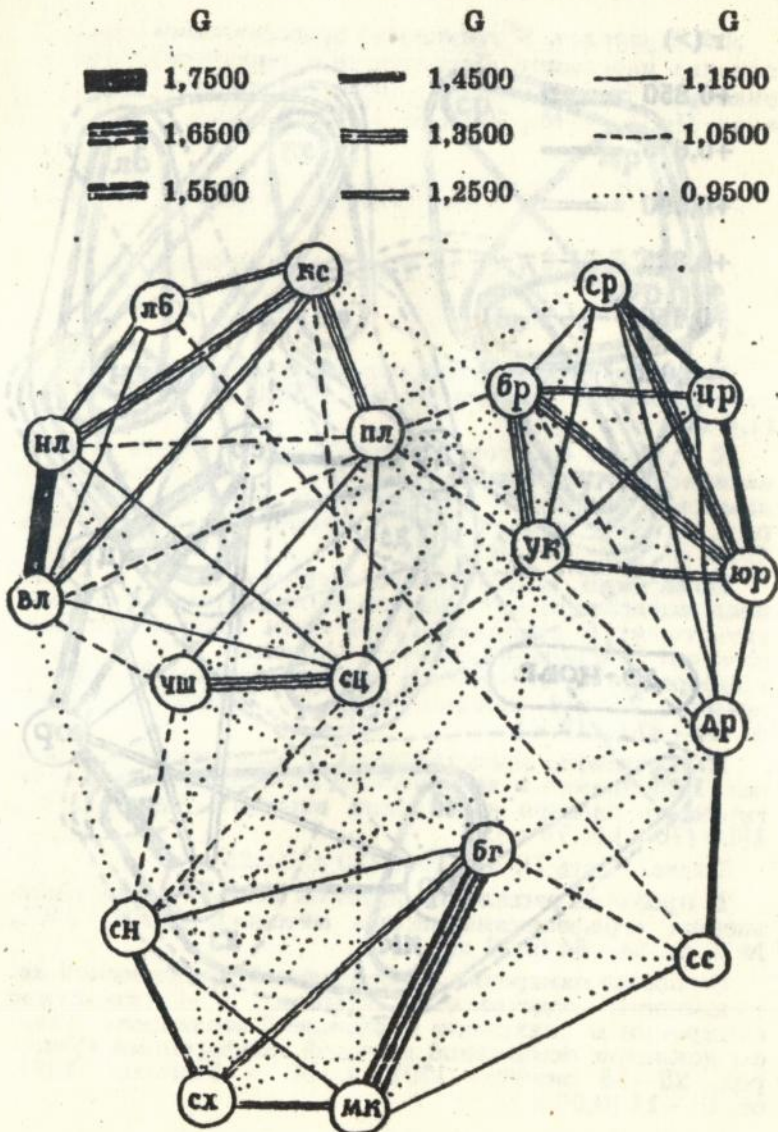






Схема II. Коррелятивные связи (r)
 между восточнославянскими идиомами
 во вневосточнославянских тяготениях
 (к проблеме древненовгородского диалекта)


$r (>)$


+0,850 

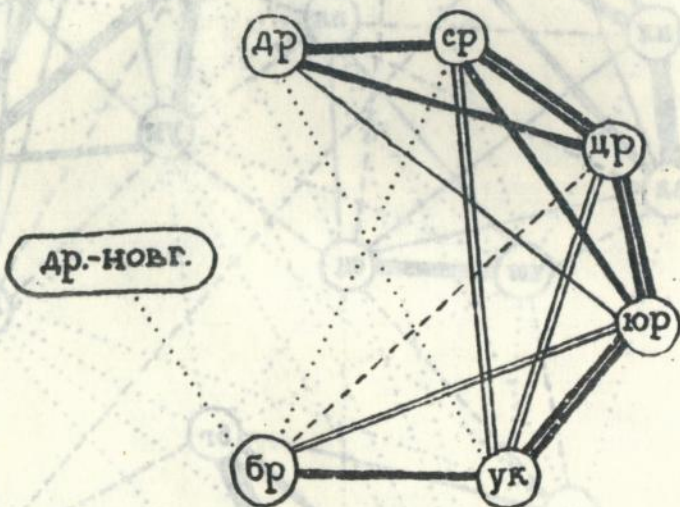
+0,675 

+0,500 

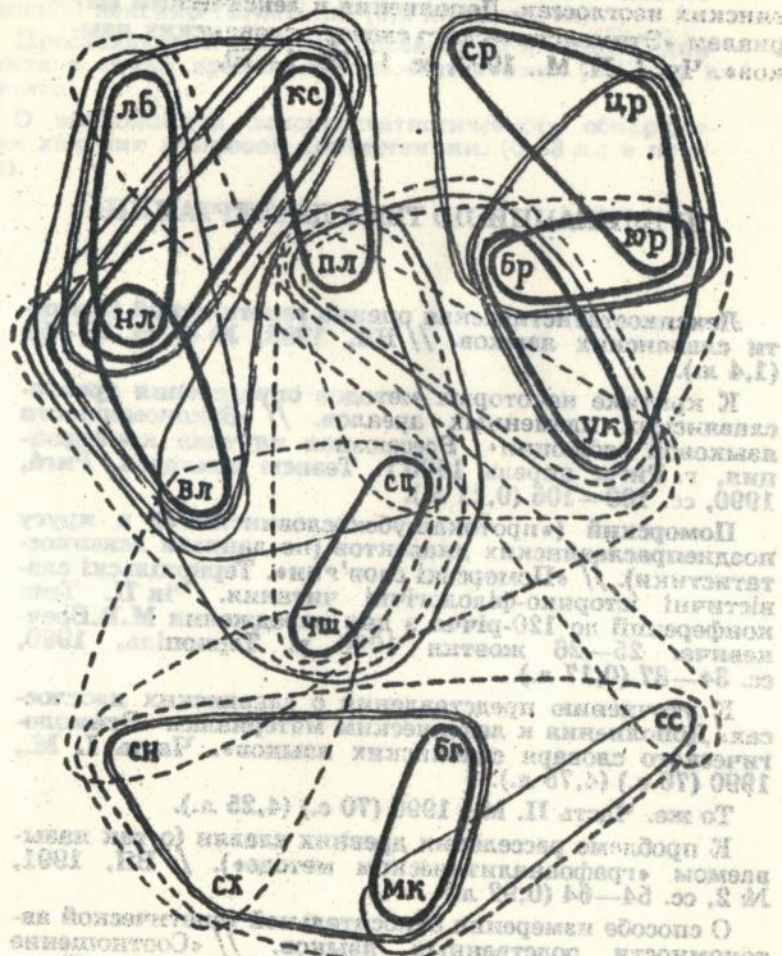
+0,325 

+0,150 

-0,025 



**Схема III. Аппроксимативная схема
диалектных объединений в позднераславянском
по данным лексикостатистики
(версия SLAV, с исключением древнерусского языка)**



ПРИЛОЖЕНИЕ К ДИССЕРТАЦИИ:

А.Ф. Журавлев. К уточнению представлений о славянских изоглоссах. Дополнения к лексическим материалам «Этимологического словаря славянских языков». Ч. I, II. М., 1990, сс. 1—76, 1—70.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ:

Лексикостатистическая оценка генетической близости славянских языков. // ВЯ, 1988, № 4, сс. 37—51 (1,4 л.).

К критике некоторых методов определения древнеславянских племенных ареалов. // «Закономерности языковой эволюции». Всесоюзная научная конференция, г. Рига, апрель 1990 г. Тезисы докладов. Рига, 1990, сс. 103—106 (0,11 л.).

Поморский («протокашубскословинский») в кругу позднеславянских диалектов (по данным лексикостатистики). // «Поморські слов'яни». Тернопільські славістичні історико-філологічні читання. Рік II. Тези конференції до 120-річчя з дня народження М.В.Брежневича. 25—26 жовтня 1990 р. Тернопіль, 1990, сс. 34—37 (0,17 л.).

К уточнению представления о славянских изоглоссах. Дополнения к лексическим материалам «Этимологического словаря славянских языков». Часть I. М., 1990 (76 с.) (4,75 л.).

То же. Часть II. М., 1990 (70 с.) (4,25 л.).

К проблеме расселения древних славян (о так называемом «графоаналитическом методе»). // ВЯ, 1991, № 2, сс. 54—64 (0,92 л.).

О способе измерения относительной генетической автономности родственных языков. // «Соотношение синхронии и диахронии в языковой эволюции». Тезисы докладов Всесоюзной научной конференции (Ужгород, 23—25 октября 1991 г.). М. — Ужгород, 1991, сс. 10—11 (0,06 л.).

Из квантитативнотипологических наблюдений над лексикой славянских языков (праславянское наследие). // ВЯ, 1992, № 2 (1,38 л.).

Заметки на полях «Этимологического словаря славянских языков». // «Этимология. 1988—1990». М., 1992 (0,87 л.).

Украинский и белорусский языки в свете праславянской лексикостатистики. (0,2 л.; в печати).

Праславянский словник древненовгородского диалекта с точки зрения лексикостатистики. (1,13 л.; в печати).

О возможности лексикостатистического обнаружения явления языковой конвергенции. (0,83 л.; в печати).

Получено в печать 24.02.92
Цена 475
Тираж 100 экз.

Издательство ГИИ «Византизм и славянство» Адрес: 119435, Москва, Савиновский пер., д. 23

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

В. И. Савинский. К вопросу о славянских диалектах // Вестник Ленинградского университета. 1950. № 1. С. 11-15.

Подписано в печать 24.02.92 Офсетная печать Формат 60x84/16
Тираж 100 экз. Заказ 425 Цена договорная

Ротапринт ГП «Газавтоматика» Адрес: 119435, Москва, Савинская наб., 25

467021

AB 25.574

212

212