Д. В. Деопик, М. А. Членов

ТОПОНИМИЯ И ЯЗЫК

(К ПРОБЛЕМЕ ВЫЯВЛЕНИЯ СУБСТРАТНЫХ ТОПОНИМИЧЕСКИХ АРЕАЛОВ)

Язык, как известно, считается одним из важнейших признаков этноса. Важность языковых данных для этнографической характеристики существующих народов несомненна. Не меньшее значение имеют данные языка и для исследований в области этногенеза. При этом различаются два случая: 1) когда язык исчезнувшего народа сохранился в доступных для нас письменных памятниках, 2) когда таковые памятники не сохранились или не существовали вообще.

Во втором случае, к которому относится наше исследование, можно выделить ряд источников для выяснения языковой принадлежности исчезнувшей языковой общности. К этим источникам относятся: иноязычный субстрат в лексике языка современного населения, отдельные слова, сохранившиеся в письменных источниках других народов, и главным образом топонимия.

Исследование лингвистического комплекса по данным топонимии может быть опять-таки представлено в виде двух основных направлений: 1) когда сохранились представители той языковой семьи, к которой принадлежал исчезнувший народ; в этом случае оправдано применение этимологических исследований (в таком положении находятся, например, исследователи, занимающиеся этногенезом населения Севера Европейской части СССР, где сохранился богатый финно-угорский субстрат в топонимии); 2) когда нет возможности выдвинуть какую-либо обоснованную гипотезу о происхождении исчезнувших народов (классическим примером такого случая являются этруски). Но существуют и другие, более сложные случаи, когда не сохранилось никаких сведений о природе (этнической, языковой) субстратного населения, кроме самого факта заселенности территории в прошлом.

В таком случае немаловажные результаты можно получить при помощи формантно-рядового анализа топонимов, т. е. установления и картографирования устойчивых звукосочетаний, характерных для определенного топонимического ареала. Наиболее сложные проблемы возникают в том случае, когда восстановление такого рода топонимических формантов, по крайней мере на ранней ступени исследования, затруднено или невозможно. В таком положении, как нам представляется, единственной реальной характеристикой исчезнувшего языка будет частотность распределения фонем и их сочетаний. В случае концентрации системно отклоняющихся от современного языка и достаточно многочисленных частотных характеристик можно высказать предположение о существовании на данной территории некой языковой общности и о пределах ее распространения.

При изучении этногенетических процессов на территории Индонезии чаще всего приходится сталкиваться с последней возможностью. Поэто-

му основное внимание авторов было сосредоточено именно на ее детальном анализе.

Теоретически можно представить, что установленный ареал системно отклоняющихся частотных характеристик отражает не столько влияние исчезнувшего субстратного языка, сколько специфику распределения фонем в топонимии, которая может определяться различными факторами, или что этот ареал отражает как первое, так и второе. С определенностью ответить на этот вопрос можно, только сопоставив частотный фонемный состав конкретного живого языка и географических названий, бесспорно созданных носителями этого языка. В качестве гипотезы можно утверждать, что частотный фонемный состав в обоих случаях одинаков. Проверке этой гипотезы и посвящена данная работа. В случае ее подтверждения можно будет говорить, что устойчивые отклонения частотности фонем в определенном ареале есть следствие влияния субстратного языка, сфера распространения которого, таким образом, приблизительно устанавливается. Доказательство данной гипотезы позволит восстановить некоторые лингвистические характеристики полностью исчезнувшего языка. Возможное влияние протосубстратного языка предполагается незначительным. После выявления отдельных ареалов системных отклонений восстановленные лингвистические характеристики послужат основанием для проведения сопоставлений и сравнительного анализа.

Проверка гипотезы о единстве фонемного состава языка и топонимии требует принципиально иной методики, чем этимологический или формантно-рядовой анализ, которые пока преобладают в работах по топонимике. Исследования предлагаемого типа, опирающиеся на массовый (в идеале — весь имеющийся) материал, по необходимости базируются на статистических методах. Особенно важна такая методика исследования для стран, где этимологический анализ затруднен в результате существования бесписьменных языков и вследствие их малой изученности. Это — страны Юго-Восточной Азии, Африки, Латинской Америки.

Следует оговориться, что при работе с субстратными топонимами мы встречаемся с проблемой фонетической адаптации, т. е. с переозвучиванием современным населением топонимов чуждого ему лингвистически субстратного населения. Этот факт делает практически невозможной полную реконструкцию фонемного состава субстратного языка, так как процесс фонетической адаптации является необратимым (например, в случае Херсонес>Корсунь при отсутствии каких-либо дополнительных источников восстановление начального «Х» было бы невозможным). Часть фонем, характерных для субстратной топонимии, заимствуется победившим языком, существование (но не звуковая природа) остальных может быть установлено путем статистического анализа топонимии; они реализуются в виде отклонений частот соответствующих фонем от их частоты в победившем языке.

Эксперимент по проверке изложенной выше гипотезы был поставлен на материале современного малайского языка (включая обе существующие литературные формы — малайзийскую и индонезийскую) и топонимии территорий, населенных малаеязычными народами. При выборе из всех языков островной Юго-Восточной Азии именно малайского мы руководствовались следующими соображениями.

1. Это — наиболее распространенный язык данной области, уже в средневековье ставший языком общения различных народов Индонезии и с тех пор постоянно расширявший сферу своего распространения.

2. Этому языку соответствуют наиболее древние районы компактного расселения одноязычного населения (из тех районов подобного рода, для которых можно исторически проследить древность расселения). Реально предположить, что именно в таких районах помехи, создаваемые

субстратом, будут минимальными. Соответственно, наибольшая вероятность выделения субстратных топонимических ареалов при помощи устойчивых отклонений в частотных характеристиках будет иметь место в районах сравнительно недавнего заселения.

3. На этом языке создана обширная, легко доступная у нас литература.

Выбор языка обусловил и выбор территории, с которой был взят топонимический материал. Это районы, населенные народами, считающими малайский язык и его диалекты родным, т. е. п-ов Малакка, восточное побережье Суматры от Медана до Палембанга, Центральная Суматра, архипелаги Риау-Лингга, о-ва Бангка и Биллитон, Анамбас и Натуна, а также малайские районы Калимантана (включая территории бывших султанатов Понтианак, Котаварингин, Кутей и Булунган). Спорным может показаться включение в эту территорию области расселения минангкабау, так как некоторые исследователи склонны считать минангкабау отдельным языком. Однако несомненно, что из всех языков суматранской группы минангкабау наиболее близок к малайскому и находится как раз на неясной и расплывчатой грани между языком и диалектом языка. В связи с этим мы сочли возможным условно рассматривать минангкабау как диалект малайского языка с расчетом на то, что существенные различия между этими двумя языками, если они существуют, неизбежно выявятся в ходе эксперимента.

На данном этапе исследования мы отказались от анализа диалектальных различий внутри малайского языка, так как наиболее существенные частотные характеристики малайского языка должны быть общими и для его диалектов в пределах общепризнанной зоны распространения малайского языка.

Нами была сделана выборка из 721 топонима с указанной территории на основании карт атласа («Atlas untuk sekolah landjutan», Djakarta, 1957) и 857 слов текста. Текст представлял собой три связных отрывка из следующих книг:

1) отрывок из книги Х. Б. Ясина «Kesusasteraan Indonesia dimasa Djepang»;

2) отрывок из романа известного индонезийского писателя Т. С. Алишахбана «Lajar terkembang» (цитируется по книге «Образцы современной индонезийской прозы», Л., 1964, стр. 24—25);

3) отрывок из сказки «Kanchil dan anak-anak memerang», любезно предоставленный нам Е. В. Кочановым.

Последний отрывок был издан на малайзийской литературной форме малайского языка. Эти три текста отражают важнейшие жанры литературы: научный, художественный и фольклорный и поэтому могут рассматриваться как грубая выборка языка, достаточная по объему для фонологических исследований и близкая по объему к выборке из топонимов.

Выборка топонимов из всего имеющегося материала географических названий подчиняется требованиям, обычно предъявляемым статистикой к такого рода операциям. Проблема репрезентативности выборки, т. е. того, насколько она характеризует весь массив, сводится, во-первых, к определению ее необходимого объема, и, во-вторых, к обеспечению ее случайного характера.

Около 4000 фонемных сегментов в топонимии и приблизительно столько же в текстах — достаточно большие выборки для статистического анализа. Напомним, что отнесение выборки к большим или малым определяется не ее отношением к размерам генеральной совокупности, а ее абсолютной величиной. При анализе данных по районам мы работали с малыми выборками, теория которых в достаточной степени разработана. В данной связи следует упомянуть, что при достаточно больших процентных значениях тех или иных фонем полученные данные

пригодны для характеристики генеральной совокупности даже при выборках порядка двадцати топонимов.

Поскольку в большинстве работ по частотному анализу фонемного состава различных языков предполагается нормальное распределение величин экспериментально наблюденных частот тех или иных фонем вокруг некоторой постоянной для данного языка частоты (вероятности), мы предполагаем нормальное распределение и для малайского языка. Фонетический состав топонимов — явление лингвистическое, нормы распределения вероятностей вокруг средней здесь должны быть аналогичны. Большой объем выборок, а также то, что для выводов использовались лишь относительно частые фонемы, позволили нам использовать готовые таблицы доверительных интервалов для биномиального распределения [1], что значительно облегчило расчеты. Правомерность сведения к биномиальному распределению основана на его сближении с нормальным при достаточно больших числах. Для данных достаточно грубых расчетов сходимость частоты к вероятности считалась бесспорной, так как необходимость поправок к отклонениям появляется только при более строгом подходе. Естественно, что предполагается проверить сходимость на ряде дополнительных выборок.

Проведенная статистическая обработка материала показала, как будет видно далее, достаточность взятого материала для доказательства соответствия частотного фонемного состава языка и его топонимии. В то же время необходимость такого исследования очевидна, так как оно, во-первых, дает единственную возможность определять субстратные ареалы топонимов, связанных с языками иных семей, во-вторых, может послужить базой реконструкции фонемного состава субстратных языков, и, в-третьих, может быть применено к анализу родственных, но достаточно отличных от малайского языков, например, яванского.

Фонологический анализ топонимов поставил нас перед проблемой, общей для многих ономастических задач — проблемой записи топонимов в какой-либо общепринятой транскрипционной системе. Таковая запись еще никем не делалась. Но поскольку большинство топонимов Индонезии вошло в современную картографию через современные индонезийский и малайзийский варианты малайского языка и поскольку эти варианты чрезвычайно близки, мы сочли возможным записать выбранные нами топонимы средствами специально разработанной нами (применительно к графемам указанных источников) транслитерационной системы, приближающейся по своему характеру к фонематической. В дальнейшем транслитерационные единицы (графемы) условно воспринимались нами как фонемы соответствующего языка. Авторы отдавали себе отчет в том, что такой подход не давал возможности судить о звуковой (фонетической) природе данных фонем. Это тем не менее не мешало работе, так как акустическая субстанция не являлась объектом нашего исследования. Основанием для условной идентификации графем и фонем служит знаковый характер языка в целом и семиотическая (фонологическая) природа фонем в частности. Следует отметить, что фонологический анализ на основании исследования графем не является чем-либо новым ни в лингвистике, ни в топонимике [2]. При работе с бесписьменными или младописьменными языками (каковыми является большинство языков Индонезии) такой подход правомерен и плодотворен.

Для статистической обработки материала данных выборок были использованы 80-колонные перфокарты, обработка которых велась на

---

[1] Л. Н. Большов, Н. В. Смирнов, Таблицы математической статистики, М., 1965, табл. 5.2.

[2] См. например: Ю. К. Лекомцев, Дистрибуция фонем и генерация слогов (на материале графических слогов классического тибетского языка), сб. «Вопросы структуры языка», М., 1964, стр. 8.

счетно-аналитических машинах: табуляторе типа Т-5-М, сортировке и перфораторе. Сам неопределенный характер задачи затруднял ее моделирование для обработки на электронно-вычислительных машинах, поэтому анализ был проведен на счетно-аналитических устройствах. Статистическому анализу подвергались следующие элементы: фонемы, слоги[3] и межслоговые двухфонемные стыки; каждый из них анализировался в зависимости от типа топонима, от района, в котором данный топоним находится, от возможности заимствования данного топонима из какого-либо неиндонезийского языка, от величины слова и от позиции данного элемента в слове; изучалась и зависимость от одного или нескольких указанных факторов. Эта методика позволила проводить анализ основных интересующих нас показателей как в зависимости друг от друга, так и в зависимости от внелингвистических факторов. В итоге получился следующий макет перфокарты, в котором вся информация о топонимах была записана цифровым кодом[4]:

1—2 колонки — номер района. Весь ареал был нами разбит на 19 районов в соответствии с историко-административным делением этой области. Цифровые обозначения возрастают с севера на юг.

01 Перлис
02 Пенанг
03 Перак
04 Тренггану
05 Келантан
06 Паханг
07 Селангор
08 Негри-Сембилан
09 Малакка
10 Джохор
11 Центральная Суматра
12 Кутей-Булунган
13 Западный Калимантан
14. Бангка-Биллитон
15 Палембанг
16 Джамби
17 Северная Суматра
18 Риау
19 Кедах

3—5 колонки — порядковый номер. Нумерация сквозная, в пределах соответствующего района (например, «Палембанг, Баюнглинчир» будет кодироваться как — «15001», где 15 — номер района (Палембанг), а «001» — порядковый номер топонима данного района).

6 колонка — тип топонима. Учитывая цели исследования, мы в данной работе не предполагали использования дробной классификации типов, а также отказались от привлечения топонимов, обозначающих крупные топонимические объекты. Была использована следующая градация:

1 — название населенного пункта
2 — название острова
3 — гидронимы (исключая морские объекты)
4 — оронимы
5 — мысы
6 — морские проливы и заливы

---

[3] Результаты анализа слогов опускаются, поскольку методика их обработки принципиально не отличается от работы с фонемами, а объем статьи не позволяет изложить конкретные результаты.

[4] Колонка перфокарты может нести информацию, записанную цифровым кодом (от 0 до 9). Выбор количества колонок определялся соответственно числом знаков, необходимых для цифровой кодировки данного элемента. Например, поскольку количество фонем превышает 10, каждая из них записывалась двухзнаковым кодом (а=03), занимающим две колонки.

7—8 колонки — количество фонем в топониме. Цифра кода соответствует количеству фонем (например, «07» — семь фонем, «12» — двенадцать фонем).

9 колонка — количество слогов в топониме.

10 колонка — количество корневых и служебных морфов в топониме:

1 — один корневой морф (*км*)
2 — 2 *км*
3 — 3 *км*
4 — 4 *км*
5 — 1 *км* и 1 *см* (служебный морф)
6 — 1 *км* и 2 *см*
7 — 2 *км* и 1 *см*
8 — 2 *км* и 2 *см*
9 — 3 *км* и 1 *см*
0 — 3 *км* и 2 *см*

11 колонка — факт заимствования данного топонима (адресация слова к тому или иному языковому комплексу):

0 — незаимствованный топоним
1 — 1 элемент — малайский, 1 — индийский
2 — 1 элемент — малайский, 1 — из других индонезийских языков
3 — из других индонезийских языков
4 — 1 элемент малайский, 1 — из прочих языков (например, кхмерского, арабского, европейских, персидского)
5 — 1 элемент — из других индонезийских языков, 1 — индийский
6 — заимствование из индийских языков (индоевропейской или дравидийской семьи)
8 — заимствование из европейских языков
9 — прочие и неизвестные.

В процессе работы выяснилось, что подобная градация заимствований не оправдала себя и дальнейшее применение ее нецелесообразно. Прежде всего, часто затруднительно выделить чисто малайские топонимы из числа тех, которые могут относиться к другим индонезийским языкам. Кроме того, каждый отдельный вид заимствования обычно представлен слишком малым количеством топонимов для проведения статистического анализа. И, наконец, анализ фонетического состава каждого из отдельных видов заимствований не входит в цели исследования. Вероятно, наилучшим способом фиксации заимствований для подобных целей является дихотомическая запись, при этом топонимы, относящиеся к другим индонезийским языкам, должны быть включены в число незаимствованных.

12—31 колонки — запись двух фонем, находящихся на межслоговых стыках, в порядке расположения слогов (например, Байунглинчир — Bajuŋ-lint'ir: его межслоговые стыки будут записаны следующим образом: a — j — 03—93, где 03 — a, 93 — j, затем ŋ — l — 77—10, где 77- ŋ, 10-l и т. д.).

32—55 колонки — запись первого слова топонима по слогам: на каждый слог отводилось 6 колонок, для записи слова применялась система фиксированной записи слогов, при которой каждый слог, независимо от его длины, начинался с определенной колонки; в редком случае 4-фонемного состава слога первые две фонемы кодировались одним специальным обозначением. Например, слог «slim»условно разбивался не на 4, а на 3 элемента, причем фонемы «s» и «l» кодировались одним знаком. s кодируется как 22, l как 10, а sl как 98.

56 колонка — запасная.

57—80 колонки — запись второго слова для двухсловных топонимов, широко распространенных на территории Индонезии. Немногочисленные 5-сложные неразложимые топонимы, состоящие из одного слова, записывались таким образом, что 5-й слог писался в графе, отведенной первому слогу второго слова.

Примером может служить следующая кодовая запись:

Читается она так: «15» — Палембанг, «001» — порядковый номер внутри района, «1» — населенный пункт, «11» — количество фонем в топониме, «4» — количество слогов в топониме, «2» — два корневых морфа, «0» — незаимствованный, «0393» — первый межслоговой стык: a — j, «7710» — второй межслоговой стык: ŋ — l, «0299» — третий межслоговой стык: n — t′. «00000000» — запасные места для межслоговых стыков более длинных топонимов. «06» — b, «03» — a, 00 — запасное место для третьей фонемы первого слога, «93» — j, «30» — u, «77» — ŋ, «10» — l, «12» — i, «02» — n, «99» — t′, «12» — i, «11» — r, 0000000...0 запасные колонки для второго слова, в данном случае отсутствующего. В результате читаем название топонима — город Байунглинчир — Bajuŋlint′ir.

По этому же принципу был составлен макет перфокарты для слов текста. В колонках 1—2 вместо кода района ставился номер соответствующего текста, в колонке 6 — вместо типа топонима отмечалась принадлежность данного слова к определенной синтаксической категории.

Каждый топоним и слово текста были нанесены на отдельную перфокарту.

Первоначально весь массив карт подвергается обработке на сортировке в соответствии с каждой конкретной задачей. Например, ставится задача — выбрать все незаимствованные топонимы и расположить их в определенном порядке (в силу технических удобств не совпадающем с алфавитом) по первой фонеме и в соответствии с типом топонима. В целом такая операция занимает на сортировке полторы-две минуты для массива в 721 перфокарту. В результате материал группируется по трем признакам (при массиве около 800 топонимов решение подобной задачи на руках требует значительно большего количества времени) и определяется корреляционная связь между ними. Количественная оценка этой связи получается после того, как отсортированный массив перфокарт обрабатывается на табуляторе. Эта операция занимает еще 6—7 минут. Полученные количественные оценки (например, преобладание такой-то фонемы в таком-то районе или группе районов среди незаимствованных топонимов) печатаются табулятором в виде табуляграммы, причем подсчет может производиться одновременно по нескольким признакам. Табуляграмма представляет собой стандартный список фонем с указанием их численности на всей территории, их вхождения в тот или иной район, их численности в данном районе, их принадлежности к тому или иному типу топонима и численность их по типам топонимов. Прибегая к другим видам сортировки и коммутации на табуляторе, мы можем получать и другие характеристики: число слогов — территория — длина слова и пр.

Полученные количественные характеристики не могут сами по себе служить основанием для той или иной интерпретации, поскольку численность топонимов в разных районах различна и статистическая достоверность количественных данных также неодинакова.

Для сопоставлений необходимо оперировать отношениями (в нашем случае процентами), а не абсолютными данными. Однако сами по себе проценты еще не являются статистически значимыми, поскольку при выборке из 4 топонимов и из 200 топонимов 50% в первом и втором случаях обладают совершенно различной степенью достоверности применительно к оценке общей совокупности топонимов той или иной территории.

Поэтому оказалось необходимым ввести в исследование понятие доверительного интервала или допустимой ошибки. Доверительный интервал представляет собой те пределы, в которых может оказаться истинный процент данной фонемы при исследовании всех без исключения топонимов данного района. Например, при нашей выборке в райо-

не Центральной Суматры оказалось 17,6% топонимов с инициалью «b». Доверительный интервал показывает, что при привлечении всех топонимов этого района (сколько бы их ни было) эта величина не может быть меньше, чем 11,3%, и больше, чем 23,3% (с вероятностью=0,95).

Для определения доверительных интервалов мы применяли таблицы доверительных интервалов биномиального распределения Л. Н. Большова и Н. В. Смирнова, уже упомянутые выше.

Выявление значимых различий как между топонимией и текстом в целом, так и между топонимами отдельных районов и групп районов производилось нами посредством сопоставления доверительных интервалов процентных долей одной и той же фонемы в одной и той же позиции в различных совокупностях.

Когда наибольший (верхний) доверительный интервал одной из сравниваемых групп оказывался меньше наименьшего (нижнего) доверительного интервала другой группы, считалось, что процент данной фонемы в данной группе *значимо выше* ее процента во второй группе (и наоборот) этот случай обозначался знаком ++ (или ——).

Когда доверительные интервалы частично совпадают, но наблюденное значение одной из групп меньше нижнего доверительного интервала другой группы, *то различие отмечалось,* но до увеличения статистической базы *не считалось значимым* и обозначалось знаками + или —. Подобные данные имеют значение для дальнейших исследований, а также могут быть использованы при группировке районов.

Когда не имелось и такого различия, сопоставляемые доли считались практически *одинаковыми* и обозначались знаком 0.

Выделенные таким образом значимые отклонения и совпадения и являются тем материалом, интерпретация которого может дать надежные выводы для тех или иных построений в области топонимики, истории, этнографии или лингвистики. Напомним, что редко встречающиеся фонемы не использовались для выводов.

Применительно к нашей задаче это значило определить, имеются ли существенные различия в распространении фонем, слогов и межслоговых стыков в малайской топонимии и тексте, чтобы, как отмечалось выше, иметь надежную базу для реконструкции домалайских фонетических комплексов на территории, занятой ныне малайцами. Это означало выявление групп, не имеющих существенных различий и имеющих таковые. Применительно ко второй группе необходимо было выяснить, является ли причиной этих различий специфика топонимии или некоторые нефонетические закономерности, учет которых мог бы устранить эти различия.

Предполагая, что специфика текста, с одной стороны, и топонимов, с другой, усложняет сопоставление фонемного состава того и другого, мы попытались проверить это обстоятельство, составив соответствующую таблицу частот (табл. 1). Как видно, частотный состав по инициалям показал значительные отличия, хотя даже здесь из 25 фонем 12 не показали никаких отличий, 3 дали несущественные отличия, и лишь 10— существенные (причем среди этих последних имеются как фонемы с высокой частотой встречаемости, так и с низкой). Это вполне естественно, так как и топонимы, и текст обладают определенными специфическими элементами, частотность которых выше, чем частота употребления корневых морфов. Сравнение суммарных (для топонимов и для текста) частот конкретных фонем становится возможным после учета этих элементов. Строгое вычленение этих элементов требует дальнейшей разработки на общетеоретическом уровне; применительно же к нашей задаче вопрос решался следующим образом.

1. В состав индонезийских топонимов часто входят обозначения характера географического объекта («вода», «река», «гора» и т. д.) и слова, характеризующие качество географического объекта («новый», «ста-

Таблица 1

Частота инициальных фонем в топонимии и тексте (без учета специфики каждой из двух групп)

Фонема	Топонимия			Текст			Оценка
	min	%	max	min	%	max	
b	12,8	15,6	18,2	6,2	8,1	10,0	++
s	11,7	14,4	16,9	8,8	10,9	13,0	+
k	10,5	13,2	15,5	5,4	7,2	9,0	++
t	9,5	11,9	14,3	5,9	7,7	9,5	++
m	8,9	11,2	13,5	10,3	12,6	14,9	0
p	7,5	9,7	11,9	7,6	9,6	11,6	0
l	3,8	5,5	7,2	1,8	2,9	4,0	+
a	3,5	5,1	6,7	4,7	6,4	8,1	0
r	2,2	3,6	4,8	0,1	0,5	1,0	++
g	0,8	1,8	2,8	0,1	0,4	0,7	++
d′	0,8	1,8	2,8	0,6	1,4	2,2	0
d	0,7	1,7	2,7	10,9	13,2	15,5	−−
n	0,3	1,0	1,7	0,1	0,7	1,3	0
i	0,1	0,8	1,5	5,0	6,7	8,4	−−
t′	0,1	0,8	1,5	0,1	0,6	1,1	0
u	0,1	0,6	1,2	0,1	0,5	1,0	0
w	0	0,3	0,7	0	0,2	0,5	0
ə	0	0,3	0,7	0,1	0,6	1,1	0
ń	0	0,1	0,3	0	—	0,2	0
o	0	0,1	0,3	0,6	1,4	2,2	−−
h	0	0,1	0,3	2,6	3,8	5,1	−−
ń	0	0,1	0,3	0	—	0,2	0
j	0	—	0,2	3,0	4,4	5,8	−−
e	0	—	0,2	0	0,2	0,5	−
f	0	—	0,2	0	0,1	0,3	0

Таблица 2

Частота инициальных фонем топонимии и текста (с учетом специфики каждой из групп)

Фонема	Топонимия				Текст				Оценка
	n	min	%	max	n	min	%	max	
b	54	10,6	14,2	17,8	31	9,5	14,2	17,8	0
s	53	10,4	13,9	17,4	18	4,5	8,2	11,9	+
k	50	9,8	13,1	16,4	12	2,4	5,5	8,6	++
t	49	9,5	12,9	16,3	24	6,8	11,0	15,2	0
m	40	7,4	10,5	13,6	11	2,1	5,0	7,9	+
p	42	7,8	11,0	14,2	18	4,5	8,2	11,9	0
l	23	3,6	6,0	8,4	16	3,8	7,3	10,8	0
a	11	1,2	2,9	4,6	13	2,7	5,9	9,1	0
r	15	1,9	3,9	5,9	3	0	1,4	3,0	+
g	7	0,4	1,8	3,2	3	0	1,4	3,0	0
d′	9	0,8	2,4	4,0	2	0	0,9	2,2	+
d	11	1,2	2,9	4,6	14	3,1	6,4	9,7	+
n	5	0,1	1,3	2,5	1	0	0,5	1,5	0
i	1	0	0,3	0,9	5	0,3	2,3	4,3	−
t′	3	0	0,8	1,7	2	0	0,9	2,2	0
u	4	0	1,0	2,0	4	0	1,8	3,6	0
w	—	0	—	0,6	—	0	—	0,9	
ń	1	0	0,3	0,9	—	0	—	0,9	0
o	—	0	—	0,6	10	1,8	4,6	7,4	−−
h	1	0	0,3	0,9	26	7,5	11,8	16,1	−−
j	—	0	—	0,6	—	0	—	0,9	
e	—	0	—	0,6	2	0	0,9	2,2	−
f	—	0	—	0,6	—	0	—	0,9	
ŋ	—	0	—	0,6	—	0	—	0,9	
ə	2	0	0,5	1,2	—	0	—	0,9	0
	Σ 381				Σ 219				

рый», «высокий», «любимый» и т. д.). Обе эти категории мы называем детерминативами в случае, если на взятом как выборка массиве они имеют достаточно большую частоту (5 случаев в препозиции и 3 случая в постпозиции). Различие в частоте — следствие особенностей индонезийской топонимии, в которой постпозиции встречаются реже.

2. Для текста мы считали необходимым учет различных служебных морфов и других явлений, которые практически не встречаются в топонимии. К ним в первую очередь относятся: глагольные и субстантивные аффиксы, фонетическое изменение основы при присоединении аффикса (топонимы не пользуются аффиксами), местоимения и предлоги.

И то и другое имеет смысл выделять лишь при наличии значительной частоты, которая могла бы повлиять на исход данного эксперимента, что мы и **делали.**

Вычленение из текста и из топонимов указанных элементов позволяет привести те и другие к сопоставимому виду, сформировать сопоставимый частотный фонемный комплекс. Эти комплексы, лишенные каждый своей специфики, являются наиболее подходящим объектом для проведения сравнений.

Далее оба комплекса были приведены к сопоставимому виду при помощи изложенной выше процедуры. В итоге была получена табл. 2. Из нее видно, что частоты большинства фонем топонимического комплекса и текста или совпадают (11 из 21), или отличаются несущественно (7 из 21); только три фонемы из исходного списка несовпадающих в табл. 1 сохранились (h, k, o).

Для сопоставлений были взяты инициальные фонемы: объем статьи не позволяет ввести данные по остальным фонемам, хотя технически это сделать несложно. Наименее обусловленные законами фонетических сочетаний в языке, не ассимилирующиеся предшествующей фонемой инициали кажутся наиболее представительной группой для первого эксперимента. Правда, по предварительным данным, сопоставление фонем во второй позиции в слоге, где фонемы в основном представлены редкими гласными, дает меньше различий в топонимике и тексте по сравнению с инициалями. То же самое можно сказать и о финалях слога, на фонемный состав которого накладываются определенные фонетические ограничения, суть которых будет изложена ниже. Вместе с тем на выбор инициалей в качестве первого примера оказал влияние также тот факт, что инициали очень часто связаны с определенными морфологическими элементами (префиксы и т. д.), что, на первый взгляд, должно вызывать немалые расхождения при сравнении двух исследуемых комплексов. Возможно, этим и объясняются указанные расхождения, сохранившиеся после обработки комплексов (табл. 2), хотя исход уже проведенной обработки показал продуктивность предложенной процедуры и позволяет надеяться, что ее улучшение снимет и эти расхождения. Возможно, эти расхождения частично могут быть объяснены и дефектами при первичной записи, в частности, неверным определением исконности или заимствованности данного слова или топонима[5]. Не исключена возможность интерпретации расхождений на основе специфики того или иного района.

Представилось целесообразным проанализировать также состав двухфонемных межслоговых стыков[6], поскольку предполагалось, что они менее подвержены фонетической ассимиляции, чем внутрислоговые сочетания, и поэтому в них следует ожидать большей сохранности суб-

[5] Это вызвано отсутствием квалифицированных этимологических словарей малайского языка и топонимии.

[6] Разделение на слоги в спорных случаях производилось в соответствии со схемой, предложенной в статье А. П. Павленко, Предварительные данные интонографического исследования звуков индонезийского языка, в сб. «Языки Юго-Восточной Азии», М., 1967, стр. 117—125.

Таблица 3

Частоты первого межслогового стыка в топонимии и тексте*

Стык**	Топонимия			Текст	Оценка	Стык**	Топонимия			Текст	Оценка
	min	%	max	%			min	%	max	%	
m—p	2,2	4,0	5,8	1,2	+	i—a	0	0,7	1,5	1,8	—
a—b	0,7	2,0	3,3	0,8	0	i—d	0	0,2	0,4	3,9	—
a—k	0	0,9	1,8	2,0	—	i—n	0	0,2	0,4	1,4	—
a—t	3,2	5,3	7,4	3,2	0	i—j	0	0,4	1,0	4,3	—
a—p	0,2	1,3	2,4	2,0	0	u—b	0,8	2,2	3,6	0,8	0
a—l	1,4	3,1	4,8	3,7	0	u—m	0,1	1,1	2,1	—	+
a—r	1,4	3,1	4,8	1,8	0	u—l	0,7	2,0	3,3	1,2	0
a—d	0,5	1,7	2,9	2,4	0	u—a	5,0	7,5	10,0	1,6	+
a—n	0,1	1,1	2,1	3,0	—	u—d'	0	0,6	1,3	1,2	0
a—i	1,8	3,5	5,2	—	+	u—h	0	—	0,6	1,0	—
a—w	0,1	1,1	2,1	—	+	u—ŋ	1,6	3,3	5,0	—	+
a—ń	0	—	0,6	1,2	—	o—r	0	—	0,6	1,6	—
a—h	0	0,6	1,3	1,0	0	ŋ—k	1,9	3,7	5,5	0,2	+
a—j	0,3	1,5	2,7	—	+	ə—b	0,5	1,8	3,1	2,2	0
a—mb	0,1	1,1	2,1	—	+	ə—s	0	0,9	1,8	1,4	0
r—s	0	—	0,6	1,4	—	ə—k	0,2	1,3	2,4	1,8	0
r—t	0	—	0,6	1,6	—	ə—t	0	0,6	1,3	1,0	0
n—t	1,3	2,9	4,5	0,4	+	ə—m	0,1	1,1	2,1	3,4	—
n—d'	1,3	2,9	4,5	0,4	+	ə—l	2,6	4,6	6,6	2,2	+
n—d	1,6	3,3	5,0	2,6	+0	ə—r	2,4	4,4	6,4	1,4	+
i—s	0	—	0,6	1,0	—	ə—n	0,1	1,1	2,1	2,2	—
i—k	0	0,4	1,0	1,2	—	ə—n'	0	—	0,6	2,2	—
i—t	0	—	0,6	6,7	—	ə—ń	0	0,6	1,3	4,1	—
i—l	0,5	1,7	2,9	1,0	0	ə—mb	0,3	1,5	2,7	—	+

* В таблице в целях экономии места представлены данные только по тем межслоговым стыкам, частота которых хотя бы в одном из сопоставляемых комплексов равна или превышает 1%.

** Различие констатировалось только в том случае, если доверительный интервал процентов в топонимии не перекрывал наблюденное значение процента по тексту.

стратных элементов. Если даже в этом случае расхождения будут незначительны, то возможность реконструкции элементов фонетики языка-субстрата по данным остаточной топонимии представляется реальной.

Проведенный эксперимент показал, что частоты распределений межслоговых стыков не образуют системы (табл. 3). Не только при сравнении топонимии и текста, но даже и при сравнении стыков в словах текста (между первым и вторым слогом, с одной стороны, и между вторым и третьим слогом — с другой) в пределах одного массива не наблюдается соответствия частот конкретных стыков. Не останавливаясь на возможных интерпретациях, выходящих за рамки компетенции авторов, укажем, что частотные характеристики в рамках данного эксперимента использованы быть не могут, поскольку стыки, более многочисленные, чем фонемы, имеют на данном массиве малую частоту. Сами же по себе стыки могут быть объектом качественных сопоставлений и в качестве таковых представляют определенный интерес. Уже сам набор стыков оказался весьма единообразным и устойчивым как для топонимии, так и для текста, что свидетельствует о том, что в данном случае топонимия и язык в равной степени отражают существующие языковые закономерности. В то же время выделены списки стыков, специфичных для топонимов, и такие же списки для текста. Дальнейшая работа, видимо, уменьшит этот список за счет редких пока стыков; в то же время специфичность принадлежности некоторых из них очевидна. Как один из способов сравнительной характеристики стыков можно предложить их оценку через «валентность» финалей слога. «Валентностью» в данном случае называется способность финали предыдущего слога сочетаться с определенными инициалями следующего слога. По

«валентности» оба комплекса показывают значительное сходство, что проявляется прежде всего в стабильном сравнительно небольшом наборе фонем в первой позиции межслогового стыка. Ограниченность набора — следствие известных закономерностей языка; существенно, что эти закономерности в одинаковой степени действуют и в языке и в топонимии. Как пример приведем список фонем, которые не могут быть финалью слога ни в малайском языке, ни в малайской топонимии: k, b, p, l, d, w, t′, d′, ń f, j.

Появление одной из фонем данного списка в топонимии какого-либо района может быть лишь следствием субстратных или суперстратных влияний.

* * *

Изложенный эксперимент был проведен с целью выяснения степени соответствия частотных фонологических показателей малаеязычной топонимии и малайского текста. Статистическая обработка массового материала на счетно-аналитических машинах дает возможность утверждать, что:

1) частоты отдельных фонем (в соответствии с их позицией в слоге и в слове) в общих чертах совпадают в топонимии и в языке;

2) частоты межслоговых стыков не являются устойчивыми и по-разному выражены не только в сопоставляемых комплексах, но и внутри каждого комплекса;

3) сам же набор стыков является относительно стабильным и в топонимии и в тексте;

4) расхождения частот отдельных фонем и отличия в выборе межслоговых стыков между конкретным исследуемым языком и топонимией заставляют предположить иноязычное происхождение той части топонимии, где наблюдаются эти расхождения;

5) в случае невозможности отнесения данных различий к чему-либо, кроме субстрата (т. е. если исключено наличие адстрата или суперстрата), представляется возможным приблизительно определить ареал распространения данного языка (что особенно важно для исчезнувших языков) и дать некоторые приблизительные оценки его фонетической структуры.

Поскольку достаточное для наших целей сходство частотного фонемного состава топонимии и языка, ее создавшего, может считаться установленным, в дальнейшем будет достаточно сравнивать топонимию какой-то замкнутой территории с единым языком и какую-то часть того же топонимического комплекса, считая средние для данной территории частотные показатели надежно отражающими язык этой территории, а региональные или просто резкие конкретные отклонения от нее — следами субстрата[7]. Итоги данной работы позволяют утверждать, что выявленные при помощи подобной методики фонетические особенности того или иного топонимического комплекса не могут быть отнесены за счет фонетической специфики топонимии внутри создавшего ее языка, а являются следствием иноязычных влияний.

[7] При помощи подобной процедуры были выявлены субстратные этнические ареалы на территории Центральных Молукк. См. М. А. Членов, Очерки по этнической истории народов Центральных Молукк (Индонезия) (автореф. дисс.), М., 1969, стр. 17—20.

TOPONYMY AND LANGUAGE (TOWARDS THE PROBLEM OF REVEALING SUBSTRATUM TOPONYMIC AREAS)

In reconstructing the past geographical distribution of peoples an important part is played by toponymy, especially where written sources are few, as in Indonesia. The reconstruction of past areas of population distribution is possible both at the semantic and the phonetic level. But in order to prove the existence of a certain area of constant phonetic deviations of the toponymy from the language of the people at present inhabiting the territory under investigation, it is necessary first to prove another basic thesis that the phonetic regularities of the toponymy originating with the speakers of a certain language are identical with the regularities of the phoneme frequencies of the language itself.

A statistical analysis of phoneme frequencies in the modern Malay language and in the toponymy of Malaya and the Malayan regions of Sumatra shows that the phoneme frequencies are identical; this permits ethnic reconstructions, at least in the Indonesian linguistic area, based upon the phonetic composition of the toponyms.