

## ВОПРОСЫ ОБЩЕГО ЯЗЫКОЗНАНИЯ

УДК 81:004.9

### ПРОБЛЕМЫ И МЕТОДЫ ЛИНГВИСТИЧЕСКОЙ ФИЛОГЕНИИ

*В.Д. Соловьев*

#### Аннотация

Статья носит обзорный характер. В ней кратко представлено новое направление исследований – лингвистическая филогенетика, возникшая немногим более 10 лет назад. Приведены формулировки некоторых из наиболее интересных полученных результатов. Основное внимание уделено методологическим проблемам. В аспекте изучения языковой эволюции лингвистическая филогенетика трактуется как дополнение к традиционному сравнительно-историческому методу. Обсуждаются перспективы данного направления, в первую очередь в сверхглубокой реконструкции языкового родства.

**Ключевые слова:** типологические базы данных, языковая эволюция, филогенетические алгоритмы, языковое родство.

---

#### Введение

Традиционно под филогенией (филогенетикой, филогеномикой) понимается реконструкция деревьев эволюции биологических видов. Уже несколько десятилетий для реконструкции эволюционных деревьев используются строгие математические методы и алгоритмы. В этом направлении существует обширная литература (см. [1]), поток которой значительно возрос с появлением методов расшифровки геномов. Однако древовидная модель эволюции характерна и для естественных языков, что обуславливает возможность применения развитых в эволюционной биологии методов также и в исторической лингвистике.

Пионером в этой области является, видимо, Т. Варноу, начавшая использовать методы филогении для реконструкции эволюции языков еще в 1997 г. [2]. Место генома занимает описание языка в виде строки значений признаков. Признаки могут быть как лексическими (когнаты), так и грамматическими. Применение филогенетических алгоритмов в лингвистике стало возможным после появления обширных типологических и лексических баз данных, в которых достаточно много языков описывается единообразным способом, допускающим применение математических методов. Типологические базы данных, такие, как WALS [3] и «Языки мира» [4], содержат описания языков по множеству грамматических параметров. Например, база данных «Языки мира», созданная в Институте языкознания РАН, включает описание 315 языков по 3821 параметру. Эти

базы данных могли бы называться грамматическими, однако в силу уже установившейся традиции в данной работе мы будем называть их типологическими.

Актуальность этих исследований для лингвистики определяется тем, что традиционными лингвистическими методами не удается получить решение многих проблем языковой эволюции. Сравнительно-исторический метод достаточно хорошо работает на уровне групп родственных языков, разошедшихся несколько тысяч лет назад. Принято считать, что реконструировать языковое родство на глубину более 10 тыс. лет этим методом невозможно. Впрочем, и в этом временном горизонте решены далеко не все проблемы. Например, остается неясным, в какой последовательности формировались ветви индоевропейской семьи. Не удается построить эволюционное дерево для папуасских языков. Остро ощущается нехватка новых идей и методов.

В данной работе будут описаны основные идеи филогенетики и приведены примеры ее применения в лингвистике, в том числе с использованием базы данных «Языки мира».

## 1. Филогенетические алгоритмы

Исходной целью разработки и применения филогенетических алгоритмов было восстановление эволюционных деревьев. Однако применять эти методы в лингвистике следует с определенной осторожностью. Дело в том, что по сравнению с эволюцией видов живых организмов эволюция языков включает не только механизм дивергенции, но и механизм конвергенции в форме заимствования признаков.

В итоге получаемые деревья языков отражают степень их близости – типологической, если используются типологические базы данных, и лексической, если применяются лексические базы данных. В этих деревьях совмещается информация двух типов – о генеалогическом родстве языков и об ареальных контактах.

В филогенетике разработаны многочисленные алгоритмы, применяемые в зависимости от конкретных задач и используемых данных. Выделяется два основных класса алгоритмов: основанные на расстояниях и основанные на символьных описаниях.

**2.1. Методы, основанные на расстояниях.** Первоочередной задачей является определение расстояния между языками в базе данных. Почти во всех работах используется расстояние Хемминга (иначе *Эвклидово расстояние* или *городская метрика* [4]). Расстояние между языками определяется как число признаков, по которым эти языки различаются. Для заданной группы языков совокупность расстояний между ними образует квадратную матрицу. Филогенетические алгоритмы являются, по сути, некоторым вариантом кластерного анализа, при котором множество анализируемых объектов разбивается на кластеры (группы) близких между собой объектов.

Первыми и до сих пор популярными филогенетическими алгоритмами являются UPGMA (Unweighted pair grouping method of agglomeration) и NJ (Neighbor joining). Они доступны через Интернет, в том числе в составе пакета PAUP\* [5]. Общая идея алгоритма UPGMA (если пренебрегать техническими деталями) состоит в следующем. Из рассматриваемого множества языков выбираются два

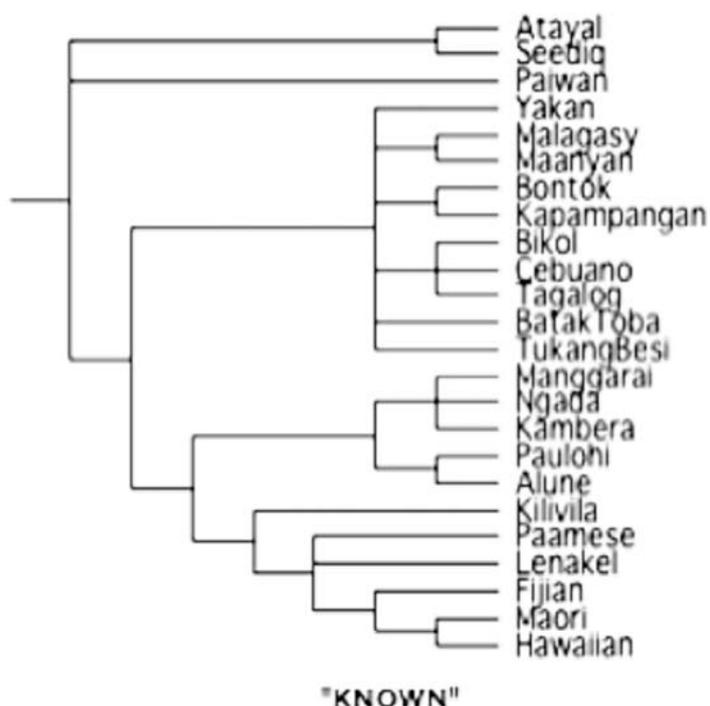


Рис. 1. Общепринятое дерево эволюции австронезийских языков

наиболее близких друг к другу (то есть с наименьшим расстоянием). Они объединяются в один кластер. По специальным формулам рассчитываются расстояния между этим кластером и остальными языками множества. В итоге получается новая матрица расстояний между объектами (отдельными языками и кластерами), к которой вновь применяется указанная процедура. Так продолжается до тех пор, пока не останется единственный кластер. В результате получается последовательность объединений наиболее близких языков, которую можно представить в виде бинарного дерева.

Приведем примеры работ, в которых применялись эти алгоритмы.

А. Саундерс [6] изучал возможности филогенетических алгоритмов для получения генеалогической классификации австронезийских языков. Он использовал базу данных, содержащую как лексические, так и грамматические признаки. На рис. 1 показано дерево эволюции австронезийских языков, установленное классическим сравнительно-историческим методом. На рис. 2 – дерево эволюции австронезийских языков, построенное с помощью алгоритма NJ.

Можно видеть, что во многих аспектах построенное алгоритмом NJ дерево совпадает с принятым в современной компаративистике. Полное совпадение на данный момент просто невозможно, так как сравнительно-исторический метод не привел к построению полного дерева эволюции со всеми промежуточными вершинами. Практически все деревья, получаемые в компаративистике, неполны – многие праязыки делятся одновременно на три и более потомка, то есть деревья содержат большую неопределенность в порядке отщепления ветвей. В то же время математические алгоритмы стремятся решить задачу «до конца», выдавая бинарное дерево, описывающее эволюционный процесс полностью.

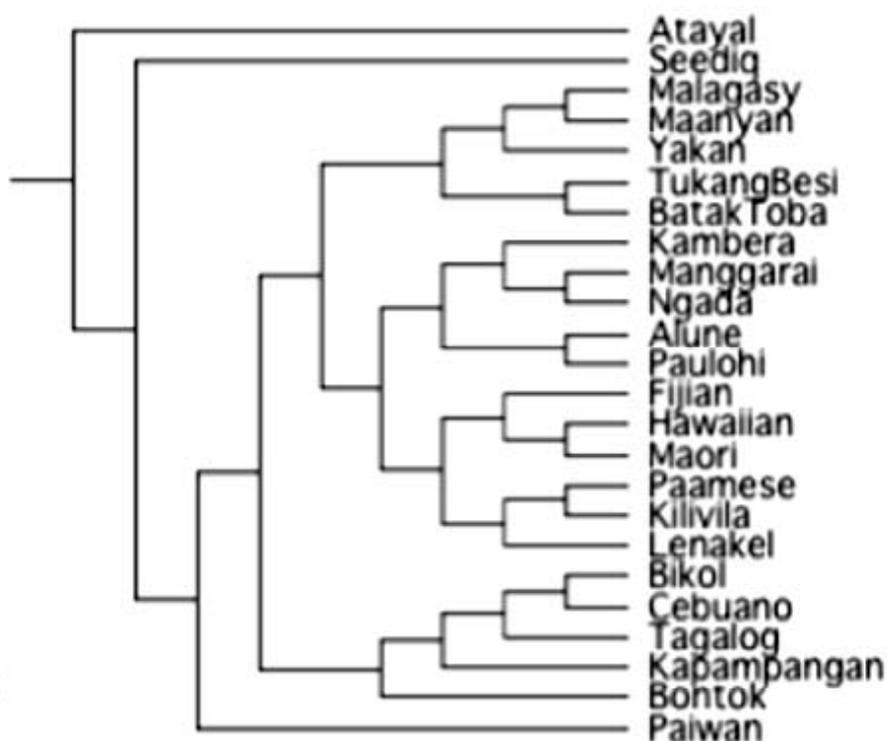


Рис. 2. Дерево эволюции австронезийских языков, построенное с помощью филогенетического алгоритма

А. Саундерс показал, что наилучшие результаты (то есть дерево, наиболее близкое к установленному в компаративистике) получаются при использовании комбинированных данных – лексических и грамматических.

Методологический анализ возможностей филогенетических алгоритмов был предпринят в [7] с использованием базы данных WALS. В этой работе выбрано 6 пар языков американских индейцев из 6 твердо установленных семей (полу жирным выделено название семьи): **Athapaskan**: Slave, Navajo; **Chibchan**: Rama, Ika; **Aymaran**: Jagaru, Аумара; **Uto-Aztecan**: Comanche, Yagui; **Otomanguean**: Chalcatongo Mixtec, Lealao Chinantec; **Carib**: Carib, Hixkaryana.

В итоге (рис. 3) алгоритм NJ правильно установил только 4 пары языков. После опробования различных вариантов алгоритмов авторы делают следующие основные выводы: 1) лучшие результаты дают алгоритмы Bayes, Neighbor-joining и NeighborNet, не адекватен лингвистическим задачам UPGMA; 2) проблемы WALS – пропуски в данных; 3) необходимо разработать методику оценки надежности филогенетических реконструкций; 4) правильный выбор набора признаков (например, выбор стабильных признаков) способствует получению лучших результатов.

Более поздние исследования [8] подтвердили предположение, что причиной получения неудачного дерева является неполнота описаний языков в WALS. Кроме того, открытым остается вопрос с выбором лучших алгоритмов. В [7] неправомерно отклонен алгоритм UPGMA, он даже вообще не рассматривался.

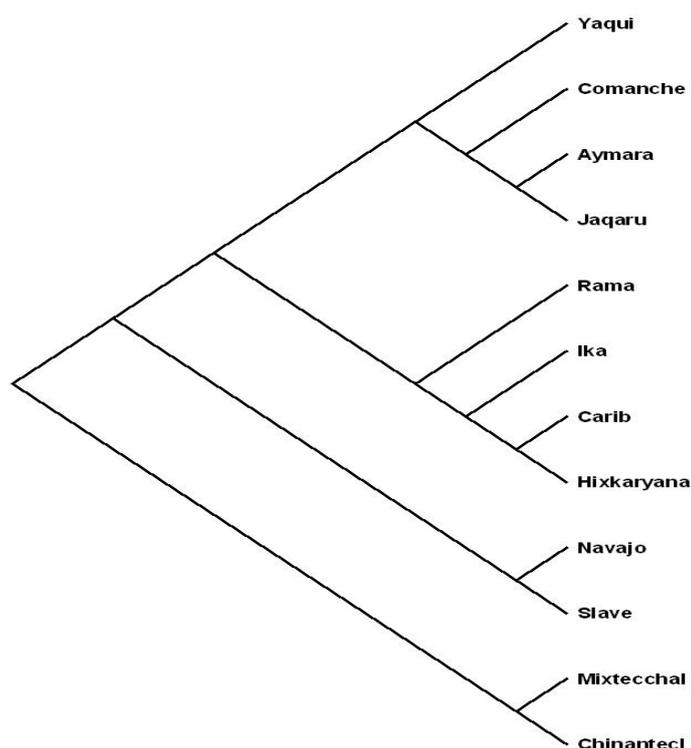


Рис. 3. Классификация индейских языков с помощью алгоритма NJ

Между тем в работе [9] показано, что в данном случае как раз UPGMA дает лучший результат. Таким образом, проблема выбора лучшего алгоритма пока не решена.

Сложной является и проблема выбора оптимального набора признаков. В [7], ограничившись лишь семнадцатью наиболее стабильными признаками, авторы получили примерно те же результаты, что и для полного набора признаков. Однако идея ограничиться небольшим набором стабильных признаков является спорной. Нестабильные признаки, разумеется, приносят шум в филогенетическое дерево, однако они несут и некоторую полезную информацию, которая будет потеряна в случае отбрасывания этих признаков. В [10] на тестовой выборке языков из WALS показано, что лучшие результаты получаются при использовании 100 признаков с наиболее высоким рангом стабильности из 141 в WALS. При использовании 25 и 50 лучших признаков результаты получались хуже. Таким образом, лишь относительно небольшое число признаков (41) являются скорее вредными, чем полезными при применении филогенетических алгоритмов.

Возможным направлением исследований в задаче поиска лучших алгоритмов является применение метрик, отличных от метрики Хемминга. Кроме метрик, описанных в главе 2, в литературе по кластерному анализу рассматривались и иные метрики. Обнадеживающие результаты дает использование идеи  $\lambda$ -компактности [11] и соответствующей  $\lambda$ -метрики. Определение  $\lambda$ -метрики можно найти в [11], здесь оно в силу своего технического характера опускается.

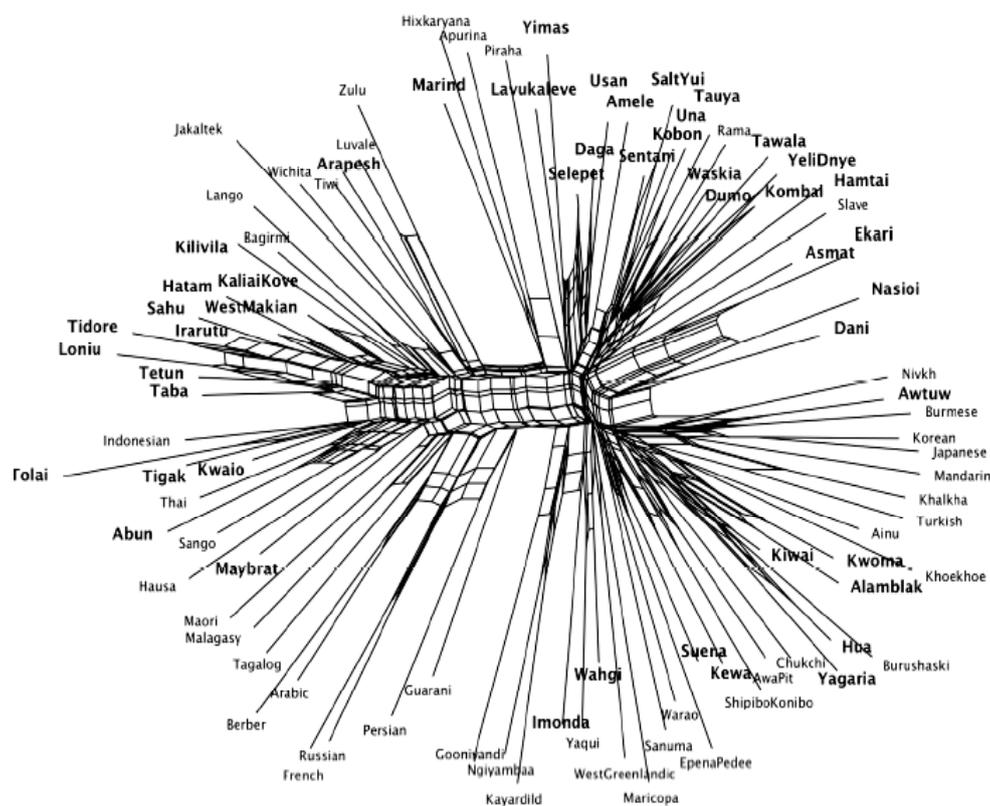


Рис. 4. Распределение языков Новой Гвинеи и остального мира

В примере с индейскими языками, как показано в [9], применение алгоритма UPGMA с  $\lambda$ -метрикой позволяет получить точное разбиение на 6 пар языков. Более того, применение  $\lambda$ -метрики систематически дает лучшие результаты для разных алгоритмов, разных наборов признаков и разных множеств языков. Следует отметить, что в этом исследовании выборки языков брались и из WALS, и из БД «Языки мира». В последние годы часто используется алгоритм NeighborNet, который строит не деревья, а более наглядные «звезды», в которых близкие языки располагаются близко друг к другу.

В [12] изучалось типологическое разнообразие языков Новой Гвинеи. Поставлен вопрос: насколько велико типологическое разнообразие языков этого региона по сравнению с языками всего мира? Для исследования выбрано 48 языков из 48 традиционно выделяемых языковых семей Новой Гвинеи. В каждой семье выделен язык из числа наиболее полно описанных в WALS. Для сравнения их с языками остального мира случайным образом выделено 48 семей и в каждой из них также выбрано по одному (опять-таки наиболее полно описанному) языку. Затем к описаниям выбранных 96 языков применен алгоритм NeighborNet. Результат приведен на рис. 4. Полужирным шрифтом выделены языки Новой Гвинеи.

Хорошо видно, что языки Новой Гвинеи не образуют компактную группу, а рассредоточены практически по всему языковому пространству. Другими

словами, степень типологического разнообразия языков Новой Гвинеи вполне сопоставима с типологическим разнообразием языков всего мира.

Как видим, форма представления результатов алгоритмом NeighborNet позволяет получить наглядное представление о степени типологического разнообразия языков различных групп. Следует отметить, что визуальным восприятием дело, разумеется, не ограничивается. За этими картинками стоят строгие математические выкладки, которые могут быть предъявлены для обоснования результатов.

Приведем пример исследования, проведенного на материале базы данных «Языки мира». Одной из нерешенных проблем исторической лингвистики является проблема происхождения юкагирского языка. Согласно [13] юкагирский язык состоит в (отдаленном) родстве с уральскими языками. На древе эволюции проекта «Вавилонская башня» (<http://starling.rinet.ru>) он помещается близко к чукотско-камчатским языкам. В [3] он считается изолятом (не входящим ни в одну семью).

Для проверки этих гипотез построим филогенетическое дерево для следующего множества языков: {чукотский, корякский, керекский, ительменский, ненецкий, селькупский, марийский, мокшанский, финский, эстонский, юкагирский}. Первые 4 из них относятся к чукотско-камчатским языкам, причем ительменский относится к южным чукотско-камчатским, остальные – к северным чукотско-камчатским [3]. Следующие шесть языков являются уральскими. Ненецкий и селькупский языки принадлежат к самодийской ветви, остальные четыре – к финно-угорской. В этой последней ветви финский и эстонский языки относятся к финно-пермской подветви финно-угорской ветви [4]. Из базы данных «Языки мира» взяты все чукотско-камчатские языки и, в целях упрощения рисунка, часть уральских – наиболее известные уральские языки из различных ветвей.

На рис. 5 показано дерево, построенное по данным «Языки мира» алгоритмом NJ с применением  $\lambda$ -расстояний.

Как видим, классификация уральских и чукотско-камчатских языков точно соответствует установленной в лингвистике. Юкагирский же язык не включается ни в ту, ни в другую группу, то есть подтверждается гипотеза о том, что это изолят. Разумеется, данное исследование не является окончательным решением проблемы родства юкагирского языка, но может служить дополнительным аргументом в пользу одной из точек зрения.

На древе эволюции в проекте «Вавилонская башня» возникновение юкагирского языка датируется 3000 г. до нашей эры. Согласно полученным данным его возникновение нужно удалить в прошлое по крайней мере на 4 тыс. лет – до момента распада общего предка уральских и чукотско-камчатских языков.

Существование двух больших типологических баз данных ставит вопрос о том, какая из них более пригодна для исследований с применением филогенетических алгоритмов.

Такое сопоставительное исследование было проведено в [8]. Для сравнения выбрано тестовое множество из 38 языков, присутствующих в обеих базах данных и достаточно полно описанных в WALS. После применения алгоритма NJ получены следующие два дерева – рис. 6 и 7.

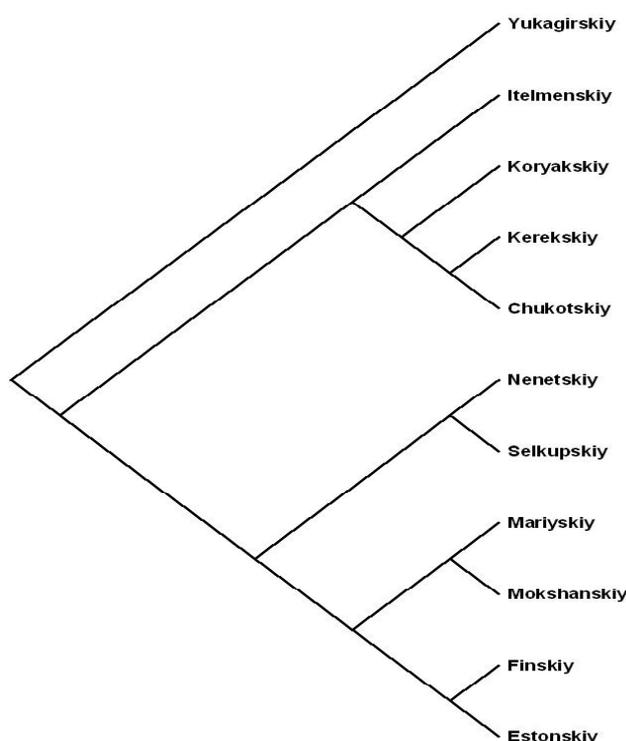


Рис. 5. Дерево эволюции отобранных языков, построенное алгоритмом NJ с  $\lambda$ -метрикой

Дерево, построенное на основе базы «Языки мира», содержит и генеалогический, и ареальный сигналы, которые можно анализировать дальше. В дереве же WALS объединение языков в группы не может быть объяснено ни их родством, ни заимствованиями. Так, вместе в одну группу со славянскими оказались объединены коми-зырянский и исландский. Складывается впечатление, что данные WALS слишком зашумлены в первую очередь отсутствием необходимой информации, так что использование WALS в этих целях проблематично и потребует как минимум разработки новых математических методов анализа в условиях характерной для WALS недостаточности данных.

**2.2. Методы, основанные на символьных описаниях.** Эти методы используют полное описание языков из базы данных в виде строки значений признаков и обеспечивают реконструкцию вместе с деревом эволюции и всех праязыков, то есть сопоставляют каждой промежуточной вершине дерева строку значений признаков гипотетического праязыка, соотносимого с этой вершиной. Основной идеей является минимизация числа мутаций – изменений значений признаков при переходе по ребру от предка к потомку. Задача построения эволюционного дерева с минимальным числом мутаций известна под названием «Максимальная бережливость». Она оказалась вычислительно трудна – NP-полна [1], что означает, что она не может быть решена точно в сколько-нибудь нетривиальном случае за приемлемое время даже с использованием суперкомпьютеров. Поэтому на практике применяются приближенные алгоритмы вроде MP (Maximum Parsimony), реализованные в пакете PAUP\*.

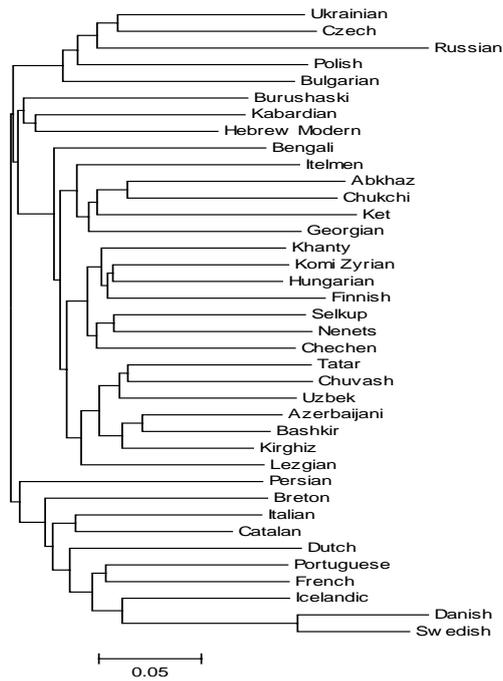


Рис. 6. Дерево тестовой выборки 38 языков, построенное на основе базы данных «Языки мира»

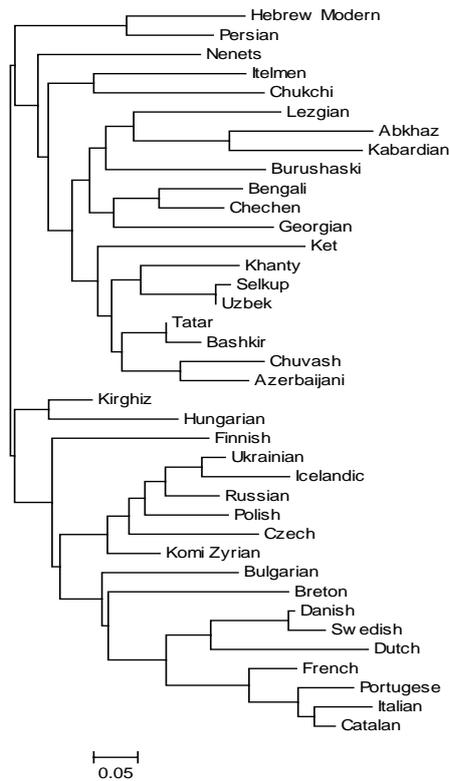


Рис. 7. Дерево тестовой выборки 38 языков, построенное на основе WALS

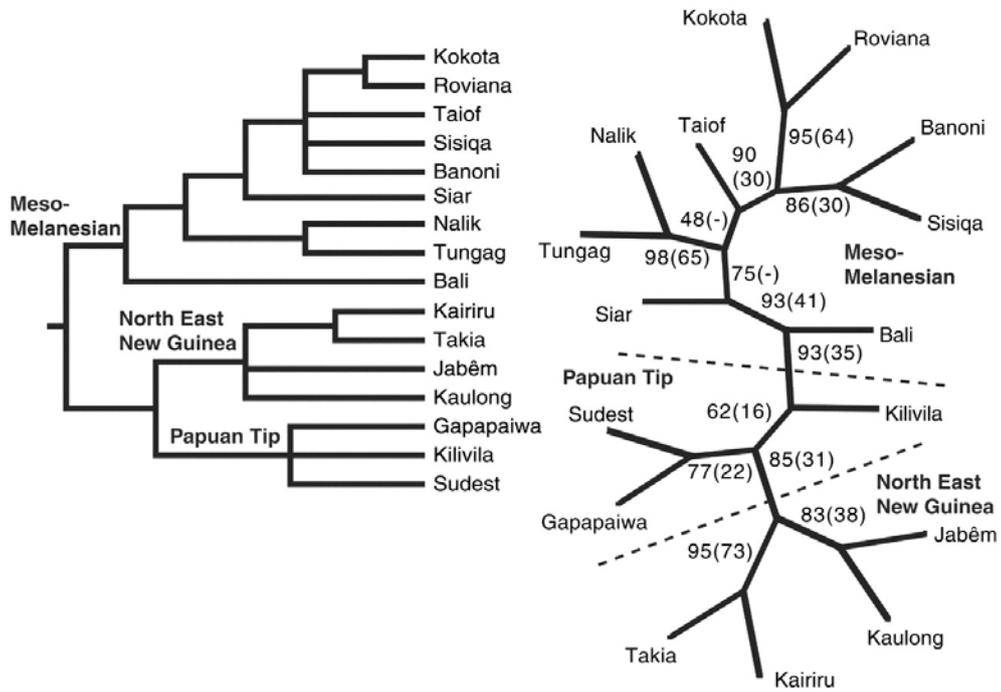


Рис. 8. Классификация австронезийских языков: слева – известная, справа – полученная алгоритмом МР

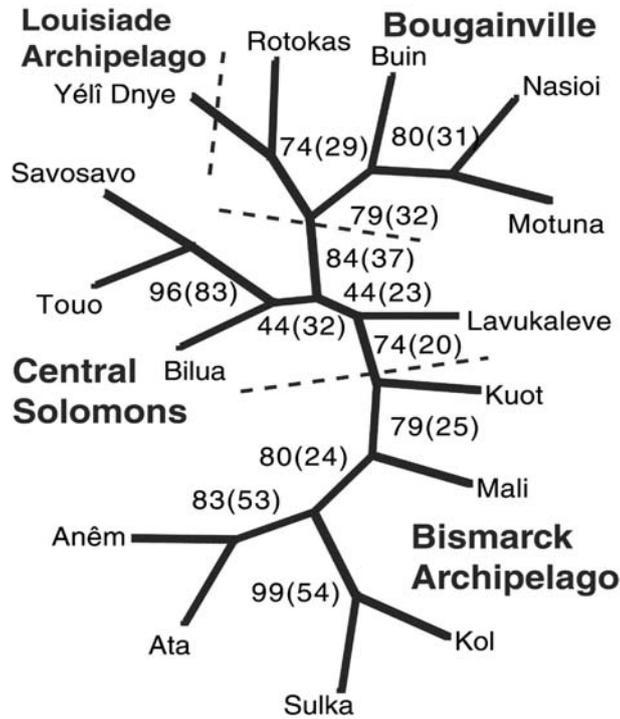


Рис. 9. Классификация папуасских языков

В [14] алгоритм МР был применен для реконструкции эволюции папуасских языков – задача, остающаяся нерешенной сравнительно-историческим методом. Перед применением МР к папуасским языкам авторы [14] сначала демонстрируют эффективность этого алгоритма на примере австронезийских языков, для которых существует надежно установленная таксономия. На рис. 8 представлен полученный результат. По техническим причинам дерево, построенное алгоритмом МР, представлено как некорневое. Видно, что алгоритм МР дает результат, достаточно близкий к правильному дереву.

Классификация папуасских языков представлена на рис. 9. Выделенные группы языков хорошо согласуются с распределением языков по архипелагам. Авторы не рассматривают полученный результат как окончательный. Для дальнейшего уточнения может быть применен сравнительно-исторический метод.

В рамках символического подхода можно учитывать различные аспекты эволюции. Реалистичными являются следующие два ограничения: исключение обратных мутаций, то есть возврата к предшествовавшему состоянию ( $a \rightarrow b \rightarrow a$ ), и параллельной эволюции, то есть одной и той же мутации  $a \rightarrow b$  на ребрах из разных ветвей эволюционного дерева. Чаще всего эти ограничения вводятся при работе с лексическими базами данных [15].

Если на филогенетическом дереве выполняются оба эти условия, то оно называется совершенной филогенией [15]. К сожалению, задача нахождения совершенной филогении также NP-полна [15].

Впрочем, на реальных данных редко можно получить совершенную филогению, поэтому актуальной является задача нахождения лучшего приближения, то есть нахождения такого филогенетического дерева, чтобы для наибольшего числа признаков выполнялись сформулированные выше ограничения.

Наиболее успешным применением этих методов является исследование эволюции индоевропейской семьи языков в работе [15]. Традиционными лингвистическими средствами не удастся получить описание последовательности выделения различных ветвей этой семьи: германских, балто-славянских и т. д. В этой работе, в частности, получено подтверждение нескольких гипотез: индо-хеттской, итало-кельтской, греко-армянской.

Исследование эволюции индоевропейской семьи языков, проведенное в [16] на основе иных теоретических принципов, дало иные результаты. Вообще большинство работ по применению филогенетических алгоритмов проводилось на индоевропейской семье языков. Испытывались различные модели. В [15] алгоритмы были усложнены за счет учета заимствований и перехода от чисто древовидной к сетевой модели языковой эволюции.

### 3. Краткое обсуждение филогенетических методов

Сравнивая традиционные лингвистические методы реконструкции эволюционных деревьев и описанные здесь математические методы, обратим внимание на резкое увеличение скорости построения деревьев. Для описания основных семей потребовались десятилетия труда сотен лингвистов. Сейчас же это можно сделать (разумеется, при наличии описаний языков в виде баз данных) за долю секунды, запустив соответствующий филогенетический алгоритм. Описанные методы могут оказаться полезными для быстрого построения гипотетических

классификаций малоизученных языков Новой Гвинеи и некоторых других регионов земного шара.

Проблема глубокой реконструкции языковой эволюции очень сложна и, как принято считать, не может быть решена чисто лингвистическими методами. В последние годы стали использоваться методы анализа генома различных народностей. Применение, по аналогии с биологией, математических и компьютерных методов для построения лингвистических филогенетических деревьев представляется весьма перспективным. Однако эта задача оказалась вычислительно крайне сложна – во многих вариантах она NP-полна. Заимствования признаков составляют дополнительную трудность. Весь взаимосвязанный комплекс проблем – создание обширных баз данных описаний языков, построение быстрых алгоритмов, выбор метрик и наборов признаков – находится на грани современных научных и технологических возможностей. Для продвижения в этой области требуются концентрированные согласованные усилия математиков, лингвистов, специалистов по компьютерным технологиям.

Важным аспектом исследований является не только описание эволюции языков, но и открытие базовых принципов, определяющих ход лингвистической эволюции. В основе различных используемых математических моделей и алгоритмов лежат различные эксплицитно сформулированные предположения о механизмах эволюции, такие, как минимизация числа мутаций. Систематическое тестирование на разном материале различных моделей и алгоритмов позволит лучше понять движущие силы языковой эволюции.

Центральной идеей установления адекватности филогенетических алгоритмов является проверка их результатов на группах языков с точно известным эволюционным деревом. В той или иной мере эта мысль проводится в работах [7, 14]. Представляется целесообразным установление некоторой точки отсчета – шаблона, на котором могут тестироваться различные методы. Наиболее близко к реализации этой идеи подошли авторы [8], предложившие тестовый набор из 38 языков различных семей.

Наконец, есть надежда на получение филогенетическими методами с совместным использованием грамматических и лексических данных более глубокой реконструкции языкового родства. Основоположник применения типологии для изучения эволюции языков Дж. Николс считает [17], что: 1) ограничения сравнительно-исторического метода вряд ли удастся преодолеть; 2) типология может помочь проникнуть значительно глубже в прошлое, чем сравнительно-исторический метод; 3) для получения лучших результатов с помощью типологии требуется лучшее понимание таких аспектов, как стабильность, независимость признаков, скорость изменений.

### Заключение

В статье дан краткий обзор нового направления лингвистических исследований – применения филогенетических алгоритмов к большим лингвистическим базам данных. Истоки этого направления можно отнести к классическим работам Дж. Николс (см. [17]). В современной версии оно начало развиваться с работы Т. Варноу [2]. Использование больших типологических баз данных берет свое начало с публикации WALS в 2005 г. Несмотря на короткую историю,

в этой области уже опубликованы десятки работ, в том числе в таких ведущих журналах, как “Science” и “Language”.

К настоящему времени можно считать, что завершается первый поисковый этап исследований. Основные его итоги, на наш взгляд, следующие.

1. Показано, что адекватным инструментом анализа больших лингвистических баз данных и построения эволюционных деревьев являются филогенетические алгоритмы. Это дает в руки лингвистов совершенно новый инструмент исследований, который является дополнительным к сравнительно-историческому методу. Он также может быть полезен в типологии и ареальной лингвистике.

2. Лучшие результаты получаются при комбинировании различных данных – лексических и грамматических. В целом вопрос о выборе баз данных, алгоритмов, метрик, наборов признаков и т. д. остается открытым. При интерпретации получаемых результатов главной проблемой является дифференциация генетического и ареального сигналов.

3. Использование нового подхода наиболее перспективно при сверхглубокой генеалогической реконструкции и для быстрой предварительной классификации большого числа языков, родственные отношения между которыми пока не установлены традиционными способами.

Работа выполнена в рамках Аналитической ведомственной целевой программы «Развитие научного потенциала высшей школы (2009–2010 годы)» Федерального агентства по образованию РФ (проект № 2.2.1.1/6944).

### Summary

*V.D. Solovyev. Problems and Methods of Linguistic Phylogeny.*

The paper is a survey of a new linguistic field, linguistic phylogeny, which appeared only 10 years ago. Some of the most interesting results are discussed. Main attention is paid to methodological problems. Linguistic phylogeny is presented as a complement to the traditional comparative-historic method. The perspectives of this field are discussed, in the first place, for deep reconstruction of the language relationships.

**Key words:** typological databases, language evolution, phylogenetic algorithms, linguistic affinity.

### Литература

1. *Гасфилд Д.* Строки, деревья и последовательности в алгоритмах. Информатика и вычислительная биология. – СПб.: Невский диалект, 2003. – 654 с.
2. *Warnow T.* Mathematical approaches to comparative linguistics // Proc. Natl. Acad. Sci. USA. – 1997. – V. 94. – P. 6585–6590.
3. The World Atlas of Language Structures / Eds. M. Haspelmath, V. Dryer, D. Gil, B. Comrie. – Oxford: Oxford Univ. Press, 2005. – 695 p.
4. *Поляков В.Н., Соловьев В.Д.* Компьютерные модели и методы в типологии и компаративистике. – Казань: Казан. гос. ун-т, 2006. – 207 с.
5. *Swofford D.* PAUP\*: Phylogenetic analysis under parsimony (and other methods). Version 4.0. Sinauer Associates. – Sunderland, 1997. – 285 p.
6. *Saunders A.* Linguistic Phylogenetics of the Austronesian Family: Ph. D. Thesis. – Swarthmore: Swarthmore College, 2005. – 198 p.

7. *Wichmann S., Saunders A.* How to use typological database in historical linguistic research // *Diachronica*. – 2007. – V. 24, No 2. – P. 78–96.
8. *Polyakov V., Solovyev V., Wichmann S., Belyaev O.* Using WALS and Jazyki Mira // *Linguistic Typology*. – 2009. – V. 13, No 1. – P. 135–165.
9. *Соловьев В.Д.* Задачи и методы лингвистической филогенетики // Труды конф. «Знания. Онтологии. Теории». – Новосибирск: ИМ СО РАН, 2007. – С. 229–235.
10. *Albu M.* Quantitative Analyses of Typological Data: Dr. dissertation. – Universität Leipzig, 2007. – 214 p.
11. *Загоруйко Н.Г.* Прикладные методы анализа данных и знаний. – Новосибирск: Ин-т математики СО РАН, 1999. – 270 с.
12. *Comrie B., Cysouw M.* New Guinea through the eyes of WALS. *Language and Linguistics in Melanesia*. 2009. – URL: <http://email.eva.mpg.de/~cysouw/publications.html>, свободный.
13. Языки мира. Палеоазиатские языки / Ред. А.П. Володин. – М.: Индрик, 1997. – 231 с.
14. *Dunn M., Terrill A., Reesink G., Foley R.A., Levinson S.C.* Structural phylogenetics and the reconstruction of ancient language history // *Science*. – 2005. – V. 309, No 5743. – P. 2072–2075.
15. *Nakhlen L., Ringe D., Warnow T.* Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages // *Language*. – 2005. – V. 81. – P. 382–420.
16. *Atkinson Q., Nicholls G., Welch D., Gray R.* From words to dates: water into wine mathematic or phylogenetic inference? // *Trans. Philolog. Soc.* – 2005. – V.103, No 2. – P. 193–219.
17. *Nichols J.* Typology in the service of classification. – 2009. – URL: [http://aalc07.psu.edu/papers/jn\\_typol\\_class3.pdf](http://aalc07.psu.edu/papers/jn_typol_class3.pdf), свободный.

Поступила в редакцию  
07.04.09

---

**Соловьев Валерий Дмитриевич** – доктор физико-математических наук, профессор кафедры теоретической кибернетики Казанского государственного университета.  
E-mail: [Valery.Solovyev@ksu.ru](mailto:Valery.Solovyev@ksu.ru)