

Юрий Алексеевич ТАМБОВЦЕВ

*Сибирский Независимый Институт, Новосибирск*

## ЛИНГВИСТИЧЕСКАЯ ТАКСОНОМИЯ: КОМПАКТНОСТЬ ЯЗЫКОВЫХ ПОДГРУПП, ГРУПП И СЕМЕЙ

В данной работе более полно рассматривается понятие компактности языковых подгрупп, групп, семей, общностей и союзов. В таком понимании понятие компактности было введено автором в научный обиход языкознания еще в 1986 году (см.: Тамбовцев 1986). Следует отметить, что в таком виде, как оно трактуется здесь, «компактность» обычно используется в естественных, а не гуманитарных науках, хотя это понятие даётся нами без излишней математической формализации. Мы трактуем его, прежде всего, как «плотный, без промежутков, сжатый, убористый, занимающий небольшой объем» (Кондаков 1975, 254; Ожегов 1970; СИС 1964, 314).

Более конкретно можно заметить, что в данном случае под «компактностью» понимается то, в какой степени более или менее плотно расположены языки в исследуемых языковых подгруппах, группах, семьях и т.д. Иными словами, каково суммарное расстояние между языками внутри исследуемых таксонов (кластеров), под которыми мы понимаем в данной работе языковые подгруппы, группы, семьи, общности и союзы (Sprachbund). В систематике животных или растений под таксоном понимают вид, род, семейство или класс.

Расстояние между языками измеряется на основе величины частоты встречаемости данной лингвистической (фонологической) характеристики в каком-либо языке. В этом смысле введённое нами понятие **компактности** по-видимому, совершенно не отличается от смысла этого термина, когда он употребляется в некоторых лингвистических работах. Как нам кажется, примерно в таком же смысле, как и мы, о компактности говорится в некоторых работах по языкознанию, написанных В. В. Мартыновым, Т. В. Гамкрелидзе, Вяч. Вс. Ивановым и другими. К сожалению, они не объясняют это понятие конкретно и детально, вероятно, употребляя этот термин в общепринятом логическом смысле, то есть в таком, который отражён в словарях. В этих работах идёт речь, например, о «**компактном языковом массиве**» (Мартынов 1983,6) или «о **компактности территориального распространения языков**» (Гамкрелидзе, Иванов 1984, 44). Также часто прибегают к употреблению термина «**компактный**» и в других гуманитарных науках. Так, в археологии он используется в близком

смысле, например, говорится о «**компактных скоплениях древних вещей**» (Третьяков 1970, 81), «**компактности территорий народностей**» (Третьяков 1970, 3) или «**компактном расселении**» (Третьяков 1982, 118).

На наш взгляд, необходимо подчеркнуть, что мы применяли (Тамбовцев 1983, 1986) и применяем более определённое понятие «**компактности**» потому, что мы, прежде всего, придерживались определения компактности, принятой в распознавании образов и статистических методах классификации (Аркадьев, Браверман 1964, 29–34; Бонгард 1967, 30–31; Васильев 1969, 16–18; Дидэ 1985, 65–71; Загоруйко 1972, 21; 1978, 26; Елисеева, Рукавишников 1977, 9–14; 1982, 6–24; Миркин 1976, 114–146; 1980, 232–238). Очень близко к нашему смыслу понятия о компактности, стоит понятие о «**сгущении**» (Миркин 1980, 234–236), которое используется и в лингвистике (Апресян 1966, 72–79). Мы считаем, что в нашей трактовке понятие «**сгущение**» и «**компактность**» можно не различать, что подтверждается определениями, которые даны ниже. Так, «**сгущением** называется такое множество объектов, в котором его любая внутренняя связь превышает внешние связи» (Миркин 1980, 235). Далее приведём определение компактности в понимании разных учёных, занимающихся разработкой методов распознавания образов.

Впервые о **компактности** (в том смысле, как употребляем это понятие мы) заговорили А. Г. Аркадьев и Э. М. Браверман, которые считают, что **компактным** можно назвать такое множество, которое «удовлетворяет следующим требованиям:

- а) число граничных точек мало по сравнению с их общим числом;
- б) любые две внутренние точки множества могут быть соединены достаточно плавной линией, проходящей только через точки того же множества, и, как следствие;
- в) почти каждая внутренняя точка имеет в достаточно обширной окрестности только точки того же множества» (Аркадьев, Браверман 1964, 28).

Разрабатывая эти идеи в своих последующих работах, А. Г. Аркадьев и Э. М. Браверман отмечают, что при наличии компактности объекты внутри группы должны быть тесно связаны между собой, и, наоборот, объекты разных групп должны быть расположены далеко друг от друга (Аркадьев, Браверман 1971, 27).

В. И. Васильев считает, что компактное множество точек в пространстве предполагает:

- 1) всегда возможен плавный переход от одного изображения к другому внутри данного образа так, что и все промежуточные изображения будут восприниматься как изображения этого же образа; и наоборот, от изображений одного

образа нельзя плавно перейти к элементам другого без возникновения изображений с неопределённой принадлежностью;

2) при малой деформации изображений в любом направлении они не выходят за пределы данного образа. Заметим, что далее В.И. Васильев полностью приводит требования Аркадьева и Бравермана, которые даются выше (В а с и л ь е в 1969, 16–17).

М. М. Бонгард замечает, что компактность расположения точек предполагает, что они лежат в пространстве кучно (Б о н г а р д 1967, 30–31).

Н. Г. Загоруйко полагает, что в распознавании образов «наиболее распространённым является предположение о «компактности», т.е. о том, что реализация одного и того же образа отображается в геометрически близких друг к другу точках выборочного пространства или что генеральная совокупность распределена по одному из известных унимодальных законов (например нормальному, характеристики которого (первые и вторые моменты) оцениваются по обучающей выборке)» (З а г о р у й к о 1972, 21).

И. Т. Турбович, В. Г. Гитис и В. К. Маслов принимают ограничение, что исходные описания объектов различных классов являются компактным семейством функций (Т у р б о в и ч и д р. 1971, 22–36).

И. И. Елисеева и В. О. Рукавишников рассуждает о компактности и **сгущении** как о местах «наибольшей концентрации точек в пространстве», то есть как о «скоплениях» точек (Е л и с е е в а, Р у к а в и ш н и к о в 1977, 40).

Б. Т. Миркин пишет, что «множество можно назвать **компактным** только тогда, когда средняя внутренняя связь между точками в каждой группе превышает среднюю внешнюю между разными группами» (М и р к и н 1980, 233).

В конце концов, большинство учёных сходится на таком определении, которое даётся в «Энциклопедии кибернетики», где гипотеза компактности рассматривается как предположение о том, что подмножество распознаваемых изображений одного класса является в определённом смысле простым. Понятие простоты может быть конкретизировано по-разному. Например, классы изображений называются компактными, если они могут быть отделены друг от друга с помощью гиперплоскостей или когда каждый класс изображений может быть представлен в виде объединения некоторого числа выпуклых множеств. В ряде исследований критерий компактности отражает представление о том, что сходство изображений одного класса должно быть больше, чем сходство изображений разных классов (ЭК 229).

С понятием компактности тесно связано понятие таксона, которое подробнее мы будем рассматривать при построении фонологических расстояний. Если сам язык считать внутренней системой, как ее понимает Ю. А. Шрейдер, то

схемы языковых классификаций можно понимать как внешние системы. Фактически под внешней системой Ю. А. Шрейдер понимает некоторый класс объектов, которые обладают некоторой общей сущностью. Более того, он считает, что эти объекты и могут объединяться в силу того, что они образуют естественную систему. Развивая эту мысль, С. Г. Кордонский дополняет ее тем, что общая сущность может быть приписана как внешним, так и внутренним системам, а естественным образом члениться может и внешняя система (Кордонский 1983, 141). Для языкознания важным является тот факт, что языковые классификации могут рассматриваться как естественные системы. Мы будем считать, что большая компактность указывает именно на естественность классификации каких-либо языковых групп. Наоборот, уменьшение компактности может показывать на неестественное объединение языков в данную группу.

Объединяя языки в группы и семьи, мы вычленим эти языки из среды всех остальных языков. Как верно замечает Г. П. Мельников, идя от среды к объекту, субъект обнаруживает объект как нарушение качественного свойства, как нарушение неоднородности качества (Мельников 1978, 22). В этом смысле компактность может служить показателем целостности. Фактически величина компактности показывает границы допустимых отклонений каждого из языков, входящих в какую-либо языковую группу. Г. П. Мельников называет это границами допустимого размытия функциональных состояний. Перенос его замечание на языковую группу или семью как целостный объект, можно говорить о пределах изменений, при которых данный объект остается в границах устойчивого функционирования (Мельников 1978, 55). Другими словами, устойчивость языковой группы как единого объекта состоит в том, чтобы влиять на языки, у которых значения частоты встречаемости отклоняются от средней величины, характерной для данной группы. Эти отклонения не должны превышать допустимых величин. Если эти отклонения больше допустимых у одного языка, то мы должны говорить, что он не принадлежит данной группе. Если отклонения больше у большинства языков данной группы, то мы должны констатировать, что фактически группы нет, а есть множество отдельных языков, которые не составляют группу как таковую. Фактически тогда можно говорить лишь о случайном механическом смешении языков, но никак не о группе, или о том, что можно назвать естественным классом.

На наш взгляд, степень компактности языковой группы является показателем меры системности данной группы как целостного объекта. По мнению Г. П. Мельникова, любой объект может рассматриваться как система, если оговорена степень его системности (Мельников 1978, 68).

В данной главе компактность групп и семей будет рассматриваться с точки зрения частоты встречаемости в языке следующих групп согласных :

1) губные; 2) переднеязычные; 3) среднеязычные; 4) заднеязычные; 5) сонорные; 6) смычные; 7) щелевые; 8) звонкие.

Частота встречаемости в потоке речи указанных выше групп согласных измерялась в процентах по отношению ко всем фонемам в звуковой цепочке. Следовательно, в данном случае неважно, будет ли частота гласных отображена в явном виде при измерении компактности, в неявном виде она фактически всегда присутствует.

Как уже отмечалось выше, всё многообразие языков мира, с точки зрения функционирования в их звуковой цепочке определённых консонантных групп, находится в пределах довольно узкого интервала. Именно поэтому, значения, например, коэффициента губности (или смычности, или звонкости), могут совпадать у языков из совершенно разных семей, которые к тому же имеют совершенно разный морфологический строй. В связи с этим, нами предлагается не ограничивать рассуждение о типологических (но не генетических!) расстояниях, т.е. величин этих коэффициентов как средства показа межъязыковых расстояний только внутрисемейным уровнем. В качестве эталона сравнения типологического сходства звуковых цепочек на межсемейном уровне можно применять как **показатель компактности** (диффузности), так и показатель евклидова расстояния, рассуждения о котором можно в следующей главе данной работы. Для сравнения величин компактности различных языковых групп необходимо выработать определенные методы. Мы предлагаем несколько таких методов. Прежде всего, мы будем применять **коэффициент вариации**, который соответствует требованию соизмеримости сравнения данных. Кроме того, необходимо обратить внимание читателя на то, что принцип соизмеримости не будет выдерживаться, если будут неверно определены указанные выше консонантные группы или будут слишком малы выборки, о чём подробно говорилось в предыдущей главе. В связи с этим, можно заявить, что нами полностью соблюдался принцип соизмеримости, при этом надёжность полученных нами измерений обеспечивалось правильным применением статистических методов (величиной доверительного интервала, формой распределения и т.д.). Принцип соизмеримости данных отмечается как один из самых главных в любом научном исследовании, хотя обычно в лингвистических исследованиях он в явном виде не оговаривается. В то же время в точных науках и даже в такой гуманитарной науке как философия на него обращается много внимания (Друянов 1982, 28–52; Загоруйко 1981, 32–34; Никифоров 1984, 92–93; Петров 1974, 6; 1975, 147–150; 1982, 29–35; Пфанцгаль 1976, 30–33).

При этом следует заметить, что даже сопоставимые данные должны сравниваться адекватно, иначе неизбежны ошибки (Хургин 1979, 101–115).

Посчитать все необходимые данные по частоте встречаемости фонем по разным языкам мира нам не представилось возможным, поэтому часть данных мы взяли из литературы, где не были указаны доверительные интервалы их частотного функционирования. Из предыдущего опыта нашей работы мы в состоянии грубо прикинуть надёжность этих данных, исходя из объёма их выборки. При сравнении объектов (в данном случае языков) должен выдерживаться единый принцип и сравнение должно осуществляться по универсальным параметрам (Кондаков 1971, 151). На наш взгляд таким универсальным параметром может служить компактность (или ее обратный эквивалент – диффузность). Она определяется двумя способами: 1) как величина, обратная коэффициенту «Т», полученного по величине «хи-квадрат»; и 2) как величина, обратная значению коэффициента вариации, который сильно зависит от величины среднего квадратического отклонения (СКО). К сожалению, среднее квадратическое отклонение сильно зависит от частоты употребления согласных какой-либо группы в потоке речи и от количества фонем в этой группе. Нам нужно употреблять величину, которая бы меньше зависела от указанных причин. Поэтому, в связи с этим, необходимо обсудить применимость коэффициента вариации и «хи-квадрат», которые свободны от этого недостатка. Они позволяют сравнивать совершенно разнородные величины. Для расчета коэффициента вариации, необходимо рассмотреть алгоритм подсчета среднего квадратического отклонения (СКО), который является базой для расчёта коэффициента вариации. Как мы покажем ниже, СКО является более устойчивым, но и более тонким инструментом, чем просто средняя величина или её размах.

Следует еще раз отметить, что для базы расчета нами берётся СКО, куда входит именно статистическая средняя (более подробно об этом см. предыдущие главы), а не мода или медиана, так как она имеет ряд преимуществ перед ними. Одно из них состоит в том, что средняя обеспечивает большую точность и надёжность при сравнениях (Хургин 1979, 105).

Кроме того, средняя предпочтительнее в связи с тем, что она лучше, чем мода или медиана, представляет распределение (Павловский 1967, 55). Далее нужно отметить, что среди числовых характеристик средняя занимает особое положение вследствие того, что вокруг неё группируются все возможные значения случайной величины (Вентцель 1964, 85). Более подробно об этом высказывается В. И. Романовский, который считает среднее квадратическое отклонение (СКО) мерой рассеяния и изменчивости. Он отмечает, что значения некоторой величины в статистическом распределении могут быть более или

менее рассеяны около их средней. Это рассеяние указывает на большую или меньшую изменчивость величины и измеряется обычно средним квадратичным отклонением значений от их средней. При этом, чем больше СКО, тем более рассеяны около средней её значения, то есть тем более она изменчива, и наоборот, с её уменьшением рассеяние и изменчивость величины уменьшаются (Романовский 1947, 50–53). СКО может служить мерой оценки параметра. Интересно, что С. Кульбак отмечает, что наилучшей оценкой является та из оценок параметра, которая обладает меньшей величиной дисперсии (Кульбак 1967, 400).

Попробуем показать более детально сущность СКО. Рассмотрим, почему ни наибольшее расхождение (размах), ни средняя величина частоты в языковой семье или группе не являются самодостаточными при их сравнении. Для того, чтобы лучше понять, каким образом располагаются языки внутри групп (семей, общностей и т.д.), необходимо вычислить СКО. Поясним это графически для простоты на одномерном случае, хотя все выводы будут сохранять свою силу и для многомерных случаев. Можно теоретически представить несколько случаев. Легко увидеть, что при одинаковом числе членов в группе и при одинаковом размахе (расхождении) получается одинаковая средняя и средний размах, но СКО, а, следовательно, и компактность двух взятых групп различны:

$$A) \left[ \begin{array}{cccccccc} \_ & \_ & \_ & \_ & \_ & \_ & \_ & \_ \end{array} \right] B) \left[ \begin{array}{cccccccc} \_ & \_ & \_ & \_ & \_ & \_ & \_ & \_ \end{array} \right]$$

Это самый сложный случай, где верно измерить разницу между группами, можно только по величине компактности, так как размах, средний размах и средняя арифметическая будут одинаковы. На практике часто встречается такая ситуация, при которой количество членов в группах и их размах различны. Эти случаи будем считать более простыми, хотя без вычислений СКО увидеть насколько различается их компактность трудно даже там. Если при одинаковом количестве членов в группе размах её больше, то ясно, что компактность её меньше:

$$A) \left[ \begin{array}{cccccccc} \_ & \_ & \_ & \_ & \_ & \_ & \_ & \_ \end{array} \right] B) \left[ \begin{array}{cccccccc} \_ & \_ & \_ & \_ & \_ & \_ & \_ & \_ \end{array} \right]$$

И наоборот, если при одинаковом размахе в группах разное число членов, то в той группе, где он меньше, компактность меньше:

$$A) \left[ \begin{array}{cccccccc} \_ & \_ & \_ & \_ & \_ & \_ & \_ & \_ \end{array} \right] B) \left[ \begin{array}{cccccccc} \_ & \_ & \_ & \_ & \_ & \_ & \_ & \_ \end{array} \right]$$

Мы рассмотрели все возможные ситуации, которые показывают нам, что с помощью коэффициента вариации (КВ), который основан на СКО, можно изучать и сравнивать между собой подгруппы, группы, семьи, общности и языковые союзы, в которых обычно различное количество членов и разное значение размахов. Ещё раз заметим, что в этих случаях невозможно провести анализ

без вычисления СКО, которое показывает меру рассеяния, т.е. диффузности (Д), достаточно хорошо. Компактность (К) обратна рассеянию или диффузности (Д). Фактически, компактность и диффузность связаны между собой как две стороны одной монеты и не могут существовать друг без друга. Заметим, что величины «хи-квадрат» (ХК) и коэффициента вариации (КВ), как меры рассеяния, обратны компактности, то есть  $D = KB$  (или  $ХК$ , в зависимости от того, в чем ее измерять), а  $K = 1/D$ , – где  $D$  величина диффузности. В дальнейшем нам пригодится понятие диффузности при обсуждении правильности таксономии или кластеризации, когда кроме оттенков «рыхлости», «разбросанности» мы будем использовать ещё один оттенок в понятии «диффузный», а именно – «смешанный». Имеется в виду тот факт, что при смешении двух разнородных групп диффузия общей группы повышается, а следовательно, её компактность уменьшается. В этом случае компактность будет служить признаком, по которому можно судить об однородности объединяемых групп. Величину диффузности (компактности) фактически можно верно измерить только коэффициентом вариации и коэффициентом  $T$  потому, что абсолютные величины частоты встречаемости наших признаков различны: например, частота встречаемости губных в финно-угорских языках 10,90%, а сонорных – в два раза больше – 24,10%.

С помощью величины компактности можно понять, принадлежит ли какой-нибудь данный язык какой-либо группе или семье языков. Мы можем проверить подобие звуковой картины какого-либо языка группе каких-либо языков. В данном случае сравним звуковую картину монгольского языка с аналогичными картинками тюркских языков. Почему для сравнения с монгольским языком выбрана семья тюркских языков, а не семья финно-угорских или австроазиатских, или австронезийских языков? В принципе наш метод позволяет сравнивать любой язык с любой группой языков. В то же время, нельзя не учитывать тот факт, что многие языковеды считают монгольские языки близкими к тюркским языкам. О. Бетлинк, В. Радлов, и другие тюркологи считали, что якутский язык близок монгольскому потому, что первоначально он был языком неизвестного типа, затем он стал языком монгольского типа, а затем подвергся тюркизации. И хотя в дальнейшем эта идея была отвергнута большинством тюркологов, тем не менее, необходимо объяснить огромное число монгольских элементов в якутском. Е. И. Убрятова полагает, что этого нельзя объяснить, не приняв гипотезу о том, что якутский язык имеет монгольский субстрат, так как тюрки обычно не заимствуют глаголы (Убрятова 1969, 29 – 30).

Н. Н. Широбокова отмечает наличие якутско-монгольских контактов в 12–15 веках и даже вплоть до 16 века (Широбокова 1980, 140–141). Сравнивая якутскую и бурятскую лексику, она приходит к выводу о том, что якутский сильно повлиял на бурятский язык (Широбокова 1980, 147).

Исследования схожести звуковых цепей тюркских языков и монгольского помогут решению «Алтайской гипотезы». Нельзя не согласиться с В. И. Рассадным, который замечает, что в долгой дискуссии «за» и «против» алтайской языковой семьи аргументы как «за», так и «против» уже давно повторяются. Нужны новые данные, которые бы поколебали чашу весов в ту или иную сторону (Р а с с а д и н 1988, 104–108). Наш метод определения компактности дает такие данные. Мы исходим из предположения того, что, если какой-либо язык поместить в какую-либо группу языков, то он либо повысит, либо понизит ее компактность. Если язык похож по своим характеристикам распределения групп согласных на языки данной группы, то он не понижает ее компактность. Другими словами не повышает ее диффузность. Довольно часто компактность группы повышается, а ее диффузности понижается. И, наоборот, язык, имеющий непохожие характеристики повышает диффузность данной группы. При этом, чем более непохож данный язык на языки данной группы, тем больше он повышает ее диффузность. Следовательно, компактность группы понижается.

Обычно языки, которые являются генетически близкими, имеют схожие звуковые цепочки. Особенно схожи их согласные. Генетически близкие языки обнаруживают свое звуковое сходство (Б а р а н н и к о в а 1973, 277–280):

Русский:	<i>мать брат сестра дети внук</i>
Украинский:	<i>мати брат сестра діти внук</i>
Белорусский:	<i>маці брат сястра дзеці унук</i>
Болгарский:	<i>майка брат сестра деца внук</i>
Сербский:	<i>мати, мајка брат сестра дете унук</i>
Чешский:	<i>mater, matka bratr sestra dite, deti vnuk</i>
Польский:	<i>maï, matka brat siostra dzieci wnuk</i>

Примерно такую же схожую звуковую картину мы находим и в других генетически родственных языках.

Русские:	<i>отец дочь вода лето хороший</i>
Немецкий:	<i>Vater Tochter Wasser Sommer gut</i>
Английский:	<i>father daughter water summer good</i>
Голландский:	<i>vader dochter water zomer good</i>
Шведский:	<i>fader dotter vatten sommar got</i>
Русские:	<i>лошадь корова кот земля большой носить</i>
Латинский:	<i>caballus vacca cattus terra grandis portare</i>
Французский:	<i>cheval vache chat terre grand porter</i>
Итальянский:	<i>cavallo vacca gatto terra grande portare</i>
Испанский:	<i>caballo vacca gato tierra grande portar</i>

## Тюркская языковая семья

Русские:	<i>мать отец дочь лошадь земля нет</i>
Татарский:	<i>ана ата кыз ат жир йук</i>
Азербайдж.:	<i>ана ата гыз ат (ер) йох</i>
Казахский:	<i>ана ата кыз ат жер жок</i>
Туркменский:	<i>ана ата кыз ат ер- йок</i>
Узбекский:	<i>она ота киз от ер йук</i>
Башкирский:	<i>ана ата кыз ат ер йук</i>
Хакасский:	<i>ине аба хызы ат чир чох</i>
Кумандинский:	<i>эне аба кыс ат чер чок</i>
Алтай-кижи:	<i>эне ада кызым ат тер тьок</i>
Якутский:	<i>ийэ ага кыс ат сир суох</i>
Тувинский:	<i>авай ачай кызым ат чер чок</i>

Мы видим, что частота встречаемости групп согласных и даже отдельных согласных почти одинакова.

Исходя из этого, мы, в какой-то степени, можем судить о звуковой близости данного языка данной группе, имея в виду типологическую схожесть. В то же время известно, что языковые контакты уподобляют один язык другому, поэтому необходимо их отслеживать.

Начнем наш анализ компактности с наиболее компактных языковых семей. Ими оказались монгольская, самодийская и тюркская языковые семьи. Далее перейдем к анализу тунгусо-маньчжурской, финно-угорской, палеоазиатской, индоевропейской и других языковых семей. В финно-угорской и индоевропейской семьях, в которых представлены хорошо выделяемые группы и подгруппы, рассмотрим также и их. Далее проанализируем такие языковые общности как уральскую, алтайскую, урало-алтайскую и балканский языковой союз.

Создав монгольскую империю, монголы несколько веков имели огромную власть над тюркскими племенами. Таким образом, монгольский язык не мог не повлиять на тюркские языки. Именно поэтому, более логично сравнивать монгольские языки с тюркскими, а не какими-то другими.

Прежде всего, рассмотрим функционирование губных согласных в тюркских языках и монгольском. Интересно узнать, какой из тюркских языков употребляет губные согласные с такой же частотностью, как и монгольский. Какие тюркские языки используют меньше губных согласных, чем монгольский, а какие, наоборот, больше?

Используют ли тюркские языки губные согласные с таким же постоянством, как финно-угорские или тунгусо-маньчжурские? Мы отвечает на эти и на неко-

торые из этих вопросов здесь и осветили их в своих предыдущих работах (напр., Тамбовцев 1983).

В монгольском языке встречаемость губных согласных в потоке речи составляет 7,52% по отношению ко всем фонемам звуковой цепочки. В тюркских языках встречаемость губных колеблется в пределах от 5,98% до 12,80%. Размах колебаний составляет 6,82%, а средняя величина размаха – 0,49%. Меньше всего губных использует алтайский язык, а больше всего – каракалпакский. Можно построить следующий упорядоченный ряд по убыванию частоты встречаемости губных согласных в тюркских языках: каракалпакский (12,80%) – турецкий (10,41%) – уйгурский (9,83%) – азербайджанский (9,66%) – узбекский (9,42%) – кумандинский (9,22%) – барабинский татарский (9,04%) – туркменский (8,50%) – киргизский (8,43%) – казанский татарский (8,03%) – казахский (7,99%) – хакасский (7,82%) – якутский 6,10% – алтайский (9,66%). Таким образом, монгольский (7,52%) займет свое место между хакасским и якутским языками. Это показывает большую схожесть дистрибуции в монгольском, хакасском и якутском языках, нежели других тюркских языках. Возвращаясь к сказанному выше, мы можем заметить, что подобная схожесть дистрибуции может говорить о типологической схожести, которая в свою очередь может быть вызвана генетической близостью, или же субстратом, о котором говорилось выше. В то же время, монгольский язык достаточно отличается по дистрибуции своих губных согласных от турецкого или каракалпакского, которые имеют похожую дистрибуцию. Такая же схожая дистрибуция наблюдается у других соседних языков: турецкого и уйгурского, уйгурского и узбекского, узбекского и кумандинского и т.д. Если рассмотреть все монгольские языки, которые были нами обработаны, т.е. монгольский (7,52), бурятский (7,67) и калмыцкий (6,65%), то можно заметить, что все они вписываются в интервал 5,98%–12,80%. В то же время, их разброс намного меньше – 1,02%. Средний разброс тоже меньше (0,34%), чем в тюркских языках (0,49%). Далее мы проанализируем компактность тюркских и монгольских языков более тонкими методами, но даже такой грубый метод как подсчет среднего разброса значений в той или иной группе уже говорит о многом. Введение монгольского языка в группу тюркских языков понижает средний разброс значений до 0,46%. Для того, чтобы понять в принципе, большой ли это разброс нужно проанализировать разбросы значений в других группах языков, например, в уральской общности, которая включает семью финно-угорских и самодийских языков. В этих языках губные встречаются в интервале от 7,71% до 13,72%. Среднее значение разброса намного меньше, чем в тюркских языках – 0,30%. Оно даже меньше, чем в смешанной группе монгольских и тюркских языков – 0,40%. Объединение монгольских и тунгусо-маньчжурских языков (7,52%–12,46%) дает большой средний разброс

(0,70%), что может говорить о меньшей схожести звуковых цепочек монгольских и тунгусо-маньчжурских языков.

Средний разброс палеоазиатских языков (6,43%–11,34%) такого же уровня (0,70%). Это может говорить о значительном типологическом расхождении этих языков. Исследованные нами славянские языки функционируют в пределах от 6,43% до 11,34% при меньшем среднем разбросе (0,52%). Известно, что славянские языки имеют значительное типологическое сходство, поэтому можно считать, что группы языков с такими же (или меньшими) значениями среднего разброса, коэффициента вариации или коэффициента  $T$  могут считаться достаточно сходными типологически. Коэффициент вариации губных согласных тюркских языков 19,60%. Он больше аналогичного коэффициента финно-угорских (14,42%), тунгусо-маньчжурских (15,57%) или монгольских (7,56%) языков. В то же время, он меньше коэффициента (2,44%) самодийских языков. Из этого следует, что компактность тюркских языков по дистрибуции губных согласных довольно низкая. Подсчет коэффициента  $T$  показывает, что его значение (0,210) достаточно далеко от единицы. Напомним, что если значение коэффициента  $T$  меньше единицы, то группа может считаться однородной. Более подробную информацию о коэффициенте  $T$  можно найти в главе, где обсуждается критерий «хи-квадрат», на основе которого построен коэффициент  $T$ . Однородность всего ряда тюркских губных говорит о том, что ни одно его значение не может считаться выпадающим из этого ряда. При рассмотрении других групп согласных (например, среднеязычных) мы увидим, что коэффициент  $T$  хорошо показывает неоднородность ряда. В данном случае однородность ряда может говорить о компактности ряда. Заметно, что нарушение однородности ряда ведет к снижению его компактности. Так, финно-угорские языки, которые показывают большую компактность по дистрибуции губных, показывают и большую однородность ряда (0,140). Вводя монгольский язык в группу тюркских языков мы тем самым повышаем однородность ряда с 0,197 до 0,192 и повышаем компактность с 19,60% до 19,45% (Следует помнить об обратной зависимости: чем однороднее и компактнее ряд, тем меньше значение).

Не следует думать, что введение любого языка в тюркскую группу улучшает ее однородность и компактность. Возьмем для примера мансийский язык из финно-угорской семьи, который может послужить примером языка, на который тюркские и монгольские языки не оказали сильного влияния. Поместив его в группу тюркских языков, мы заметим, что увеличится диффузность этой группы (23,50%) и увеличится ее разнородность (0,290). Следовательно, в отличие от монгольского языка, мансийский язык не принадлежит этой генеральной совокупности, т.е. в данном случае группе тюркских языков.

Далее перейдем к анализу компактности тюркских, монгольских и других языков по функционированию в них переднеязычных согласных.

Тюркские языки, также как и другие языки мира, чаще других групп согласных используют переднеязычные согласные. В тюркских языках они используются в интервале от 32,35% до 40,24%. Размах составил 7,89%, а средний размах – 0,564%. В монгольском языке переднеязычные согласные употребляются с частотой 36,57% из всех фонем звуковой цепочки. Как и в случае с губными согласными, переднеязычные монгольские согласные попадают в этот интервал. Введение монгольских данных уменьшает средний процент переднеязычных с 0,564% до 0,526%. В уральских языках размах составил 11,99% (24,79%–36,78%) при среднем 0,600. Даже по этому грубому индексу видно, что тюркские языки более компактны. Монгольские и тунгусо-маньчжурские языки колеблются в пределах 17,31%–36,57%; с размахом  $R = 19,26\%$ ; и средним размахом  $MR = 2,75$ . Еще менее компактную группу представляют собой палеоазиатские языки: 20,01%–36,74%;  $R = 16,64$ ;  $MR = 2,38$ . Достаточно большой разброс показывают и славянские языки: 27,42%–35,46%;  $R = 8,04$ ;  $MR = 1,01$ . Нельзя не заметить, что прикидка по разбросу и среднему разбросу является достаточно грубой. Она может применяться тогда, когда необходимо быстро понять улучшает или ухудшает введение какого-либо языка в какую-либо группу языков компактность этой группы. Тем не менее, более точные методы, в частности, коэффициент вариации и коэффициент  $T$  показывают ту же тенденцию. Введение монгольского языка уменьшает диффузность по  $KB$  с 6,03% до 5,91%, а по  $T$  с 0,074 до 0,072.

Коэффициент вариации ( $KB$ ) переднеязычных (6,03%) в несколько раз меньше, чем губных (19,45%). Та же тенденция отмечается и коэффициентом  $T$  (0,074 против 0,192). Это означает, что переднеязычные согласные во всех тюркских языках встречаются более равномерно, чем губные согласные. Значение  $KB$  и  $T$  в финно-угорских языках в несколько раз выше:  $KB=11,61\%$ ,  $T=0,264$ .

Включив монгольский язык в упорядоченный ряд тюркских языков, мы получаем следующую цепочку по убыванию значений: алтайский (40,24%) – хакасский (36,99%) – азербайджанский (36,61%) – монгольский (36,57%) – киргизский (36,27%) – казанский татарский (35,52%) – казахский (35,09%) – барабинский татарский 34,60%) – туркменский (34,50%) – каракалпакский (34,40%) – узбекский (34,34%) – турецкий (33,50%) – якутский (32,77%) – кумандинский (32,47%) – уйгурский (32,35%). Следовательно, по дистрибуции переднеязычных согласных монгольский ближе азербайджанскому и киргизскому. Интересно заметить, что по дистрибуции губных якутский язык ближе казахскому и алтайскому, а по дистрибуции переднеязычных он ближе турецкому и куман-

динскому. Показателем того, что монгольские переднеязычные функционируют аналогично переднеязычным в тюркских языках может послужить увеличение компактности (уменьшение диффузности), измеряемое КВ и Т, что показано выше. Введение в группу тюркских языков данных по мансийскому языку приводит к повышению диффузности как по КВ (7,13%), так и по Т (0,100).

Среднеязычные (палатальные) согласные в тюркских языках распределены в пределах 5,22%–0,74%. Следует заметить, что данные по частоте встречаемости среднеязычных в кумандинском языке резко выдаются вон из данного ряда (14,32%). Пик значений вызван тем, что в отличие от других тюркских языков, в кумандинском языке не одна среднеязычная фонема, а несколько. Это установила И. Я. Селютина своими экспериментально-фонетическими исследованиями. Она же подчеркивает, что кумандинцы резко отличаются от южных алтайцев по антропологическому типу. В то время как южные алтайцы относятся к центральноазиатскому типу, кумандинцы вместе с рядом угорских и самодийских народов, относятся к уральскому типу, монголоидность которого выражена слабее (С е л ю т и н а 1983, 4).

Вероятно, это вызвано субстратными явлениями в кумандинском языке, сблизив его с самодийскими и финно-угорскими, в фонемном инвентаре которых обычно несколько среднеязычных согласных. Если мы исключим кумандинский из группы тюркских языков, то ее размах резко уменьшится с 13,58% до 4,48%. Так же резко уменьшится и средний с 0,97% до 0,37%. Таким образом, даже по такому грубому индексу как средний размах видно, что кумандинский язык имеет распределение среднеязычных, которое резко отличается от других тюркских языков. Так, убирая кумандинский из группы тюркских, мы тем самым увеличиваем ее компактность. Ниже мы приведем более точные значения КВ и Т. В уральской языковой общности среднеязычные изменяются в интервале от 10,46% до 1,33%; с размахом  $R = 9,13$ ; и средним размахом  $MR = 0,46$ . Если мы объединим в одну группу монгольские и тунгусо-маньчжурские языки, то интервал составит 10,51%–1,48%. Заметим, что в палео-азиатских языках этот интервал имеет значение 11,81%–1,01%; а в славянских 9,31%–1,45%.

Диффузность тюркских языков с включением кумандинского наибольшая –  $КВ = 204,32\%$ . При этом коэффициент Т ( $T = 3,40$ ) показывает полное отсутствие однородности. Следовательно, кумандинский язык ведет себя в группе тюркских как инородный элемент. Убрав кумандинский язык, мы получаем значительное снижение диффузности ( $КВ = 41,57\%$ ,  $T = 0,279$ ).

Упорядоченный ряд по среднеязычным имеет следующий вид:

Кумандинский (14,32%) – якутский (5,22%) – туркменский (4,10%) – азербайджанский (3,97%) – казанский татарский (3,19%) – киргизский (3,07%) – узбекский (2,84%) – турецкий (2,81%) – каракалпакский (2,65%) – казахский

(2,53%) – алтайский (2,54%) – барабинский татарский (2,13%) – монгольский (1,48%) -хакасский (1,14%) – уйгурский (0,74). Здесь, так же как и в других группах согласных, монгольский язык оказывается между тюркских языков, а не на краю, как кумандинский. Ближайшими соседями монгольского языка являются барабинский татарский и хакасский. Резко бросается в глаза то, что кумандинский язык сильно оторвался от других тюркских языков: ближайший его сосед – якутский – намного ближе к туркменскому, чем к кумандинскому. Туркменский, в свою очередь, ближе к азербайджанскому, азербайджанский к языку казанских татар и т.д.

Очень важно рассмотреть частоту встречаемости среднеязычных фонем с точки зрения классификации С. Е. Малова. Фактически почти во всех тюркских языках среднеязычные согласные представлены одной фонемой /j/. С. Е. Малов разделил все тюркские языки на старые и новые по встречаемости этой среднеязычной фонемы. В алтайском, хакасском, киргизском, узбекском, каракалпакском, казахском, казанском татарском, туркменском, и уйгурском языках /j/ является единственной среднеязычной фонемой. В тех же тюркских языках, в которых она не является единственной среднеязычной, основная частотная нагрузка падает на нее, напр., в азербайджанском она покрывает 72,33% частоты всех среднеязычных фонем, а в турецком еще больше (99,27%). Несколько меньше веса среди среднеязычных фонем /j/ имеет в якутском (47,11%) и в барабинском татарском (34,54%). По утверждению С. Е. Малова фонема /j/ встречается только в новых тюркских языках, в то время как в старых тюркских языках ее место занимала какая-нибудь другая фонема. Исходя из теории С. Е. Малова, было бы логично предположить, что фонема /j/ встречается в новых тюркских языках с большей частотностью. По мнению С. Е. Малова новыми тюркскими языками являются азербайджанский, алтайский, башкирский, гагаузский, казахский, караимский, каракалпакский, киргизский, кумандинский, кумыкский, кечакский, ногайский, ойротский, печенежский, половецкий, саларский, татарский, турецкий, туркменский, узбекский, уйгурский и чулымский.

К старым тюркским языкам С. Е. Малов относит чувашский, болгарский, хазарский, рунический огузский, старо-уйгурский, тофоларский, старо-чатагайский, хакасский, шорский, и язык желтых уйгуров (М а л о в 1952,139). В связи с этой теорией хакасский должен относиться к языкам с низкой встречаемости фонемы /j/, а все остальные языки, которые мы исследовали не должны оказаться ниже хакасского. Так фактически и происходит и только уйгурский выпадает из этой теории.

Анализ распределения заднеязычных согласных в тюркских языках показывает, что меньше всего они используются в азербайджанском (6,84%) и больше всего в уйгурском (13,47%). Размах R = 6,63; средний размах MR = 0,47.

Введение в эту группу монгольского языка (11,32%) оставляет размах прежним, но уменьшает средний размах до 0,44. Это говорит о том, что данная новая группа становится более компактной. В уральских языках заднеязычные согласные функционируют в пределах 7,44%–17,48%;  $R = 10,04$ ;  $MR = 0,50$ . В объединенной монгольской и тунгусо-маньчжурской группе интервал составил 12,81%–8,50%;  $R = 4,31$ ;  $MR = 0,62$ ; в палео-азиатской: 23,21%–13,66%;  $R = 9,55$ ;  $MR = 1,36$ ; в славянской: 6,87% – 5,63%;  $R = 1,24$ ;  $MR = 0,15$ . Это говорит о том, что интервал функционирования тюркских заднеязычных не самый узкий. Диффузность тюркской группы при введении в него монгольского языка уменьшается по КВ с 14,49% до 13,96%, а по Т с 0,137 до 0,129. Последний параметр далек от единицы, что говорит о значительной однородности данной группы. Аналогичные параметры для финно-угорских языков ( $КВ = 18,67\%$ ,  $T = 0,15$ ) несколько больше, что показывает меньшую компактность этих языков по сравнению с тюркскими.

Введя монгольский в группу тюркских языков, мы получаем следующий упорядоченный ряд по частоте встречаемости заднеязычных согласных: уйгурский (13,47%) – якутский (13,27%) – хакасский (12,59) – кумандинский (11,92%) – алтайский (11,71%) – казахский (11,55) – узбекский (11,48%) – барабинский татарский (11,35%) – монгольский (11,32%) – казанский татарский (10,92%) – турецкий (10,91%) – туркменский (10,90%) – каракалпакский (10,25%) – киргизский (9,86%) – азербайджанский (6,84%). Монгольский попадает между двумя диалектами татарского языка.

Монгольский язык показывает дистрибуцию заднеязычных согласных, которая похожа дистрибуцию турецкого и барабинского татарского. Схожесть последнего может быть объяснена территориальной близостью.

Сонорные согласные в тюркских языках изменяются в пределах 30,68%–18,65%;  $R = 12,03$ ;  $MR = 0,86$ . Введение в эту группу монгольского языка (23,56%) делает ее компактнее, снижая средний разброс до 0,80. Рассмотрим, в каких пределах изменяется частотность сонорных в других группах языков. В уральских языках интервал составил 34,76%–17,45%;  $R = 17,31$ ;  $MR = 0,87$ ; в объединенной группе монгольских и тунгусо-маньчжурских языков: 28,32%–20,05%;  $R = 8,27$ ;  $MR = 1,18$ ; в палеоазиатских языках: 32,63%–19,11%;  $R = 13,52$ ;  $MR = 1,93$ ; в славянских языках, которые можно выбрать за стандарт, мы имеем: 25,06%–20,25%;  $R = 4,81$ ;  $MR = 0,60$ .

Упорядоченный ряд по сонорным имеет следующий вид: каракалпакский (30,68%) – туркменский (29,40%) – азербайджанский (26,66%) – казанский татарский (26,54%) – барабинский татарский (25,73%) – узбекский (25,24%) – турецкий (24,89%) – якутский (24,29%) – киргизский (24,31%) – хакасский

(23,96%) – алтайский (23,63%) – монгольский (23,56%) – казахский (23,20%) – уйгурский (22,93%) – кумандинский (18,65%).

Таким образом, звуковые цепочки монгольского, казахского и алтайского языков сходны в плане функционирования в них сонорных согласных. При этом интересно, что по разным параметрам монгольский близок то одному, то другому тюркскому языку. Это же можно заметить и в отношении других тюркских языков. Так, например, в данном случае якутский язык близок турецкому и киргизскому. Кумандинский опять значительно отдален от других тюркских языков. Введение монгольского языка в тюркскую группу языков понижает ее диффузность по КВ с 11,59% до 11,31%, и понижает ее разнородность по Т с 0,195 до 0,188. Заметим, что по распределению сонорных согласных тюркские языки компактнее финно-угорских, у которых КВ = 18,08%; Т = 0,48. В отличие от кумандинского языка, монгольский по дистрибуции сонорных согласных также ближе к тюркским языкам.

Смычные согласные в тюркских языках функционируют в пределах 37,26%–18,94%; R = 18,32%; MR = 1,31. Введение монгольского (23,56%) улучшает компактность этой группы до MR = 1,22%. В уральских языках этот интервал от 32,50% до 16,35%; R = 16,15; MR = 0,81. В объединенной группе монгольских и тунгусо-маньчжурских языков: 25,46%–21,72%; R = 3,74; MR = 0,53; в палеоазиатской: 28,21%–21,21%; R = 7,00; MR = 1,00; в славянской: 21,91%–14,96%; R = 6,95; MR = 0,87. Таким образом мы видим, что в среднем зона функционирования смычных согласных в объединенной монгольско-тюркской группе шире, чем в объединенной монгольской и тунгусо-маньчжурской группе, палеоазиатской, или славянской группах.

Упорядоченный ряд по частоте встречаемости смычных согласных: кумандинский (37,26%) – алтайский (25,76%) – киргизский (25,76%) – уйгурский (23,62%) – барабинский татарский (23,54%) – казанский татарский (23,17%) – узбекский (22,16%) – казахский (23,12%) – турецкий (23,22%) – монгольский (22,66%) – якутский (22,14%) – каракалпакский (20,97%) – хакасский (20,93%) – туркменский (20,50%) – азербайджанский (18,94%).

Опять монгольский язык входит в группу тюркских языков. В данном случае, монгольский близок турецкому и якутскому. По признаку частотности смычных тюркские языки не очень компактны КВ = 19,83%, Т = 0,115. Введение монгольского языка повышает ее компактность, снижая диффузность с 19,83% до 19,14%, а Т с 0,551 до 0,520. Следовательно, по этому признаку монгольский принадлежит группе тюркских языков. По этому параметру финно-угорские языки более компактны КВ = 13,56%; Т = 0,23), чем тюркские языки.

Щелевые согласные функционируют в интервале 13,65%–5,73%; R = 7,92%; MR = 0,566%. Данный интервал означает, что в тюркских языках нет языка, у

которого щелевые согласные по частоте превышают 13,65%, но в тоже время нет языка, у которого меньше 5,73% щелевых по отношению ко всем фонемам звуковой цепочки. Введение монгольского языка понижает диффузность этой группы по КВ с 22,12% до 21,34%, и по Т с 0,273 до 0,261. Уральские языки используют щелевые в интервале 16,91%–4,58%;  $R = 12,33$ ;  $MR = 0,62$ ; смешанная монгольско-тунгусо-маньчжурская группа: 18,76%–2,94%;  $R = 15,82$ ;  $MR = 2,26$ ; палеоазиатская группа: 17,64%–3,44%;  $R = 14,20$ ;  $MR = 2,03$ ; славянская группа: 17,43%–9,44%;  $R = 7,99$ ;  $MR = 1,00$ . Отсюда следует, что тюркская группа с включенным в нее монгольским языком является наиболее компактной.

Упорядоченный ряд по частоте встречаемости щелевых согласных имеет следующий вид: хакасский (13,65%) – кумандинский (12,02%) – азербайджанский (11,48%) – якутский (10,92%) – казахский (10,84%) – монгольский (10,67%) – уйгурский (9,84%) – турецкий (9,52%) – узбекский (10,68%) – каракалпакский (8,46%) – туркменский (8,10%) – казанский татарский (7,95%) – барабинский татарский (7,85%) – киргизский (7,56%) – алтайский (5,73%).

Как видно из этого упорядоченного ряда по распределению щелевых согласных, монгольский язык ближе всего к казахскому и уйгурскому, в то время как якутский близок азербайджанскому и казахскому.

Звонкие шумные согласные являются нашим восьмым признаком. Как мы увидим далее, это не очень стабильный признак, поэтому компактность по нему всегда меньше. Звонкие шумные согласные распределены по тюркским языкам в интервале 15,00%–7,22%;  $R = 7,78$ ;  $MR = 0,560$ . В уральских языках они распределены в интервале 13,00%–0,00%;  $R = 13,00$ ;  $MR = 0,65$ ; в объединенной монгольско-тунгусо-маньчжурской группе: 19,99%–9,91%;  $R = 10,08$ ;  $MR = 0,76$ ; в палео-азиатской: 16,37%–0,00%;  $R = 16,37$ ;  $MR = 2,34$ ; в славянской группе: 26,48%–19,23%;  $R = 7,25$ ;  $MR = 0,91$ .

Упорядоченный ряд по частоте звонких шумных согласных: туркменский (15,00%) – азербайджанский (14,54%) – киргизский (14,28%) – узбекский (14,20%) – алтайский (11,94%) – казахский (13,48%) – казанский татарский (11,93%) – турецкий (11,60%) – хакасский (10,26%) – уйгурский (9,90%) – кумандинский (9,23%) – якутский (9,17%) – каракалпакский (8,89%) – татарский (7,22%). Якутский ближе кумандинскому и каракалпакскому. Диффузность составила по КВ = 21,58% и Т = 0,313. Это единственный признак, по которому кумандинский язык инкорпорирован в группу тюркских языков. Это также единственный признак, по которому монгольский язык (18,57%) не принадлежит группе тюркских, так как введение монгольского языка понижает компактность, и наоборот, увеличивает диффузность по КВ с 21,58% до 25,04%,

и по  $T$  с 0,313 до 0,445. Тем не менее, другие языковые семьи по этому признаку также диффузны, например, финно-угорская семья  $KB = 46,21\%$ ;  $T = 1,11$ .

У нас имеется 8 признаков, по каждому из которых определяется величина  $KB$  и «хи-квадрат»  $XK$ . Мы можем воспользоваться тем, что они обладают свойствами аддитивности, т.е. сложимости. Из этого следует, что  $KB$  можно складывать по всем 8 признакам и затем оперировать с их суммой, которая будет показывать общую диффузность (или рыхлость). Аналогичным образом можно поступать и с величиной «хи-квадрат» ( $XK$ ), которую мы будем применять в виде коэффициента  $T$ , т.е. сопоставив эмпирическую и теоретическую величину  $XK$ . Теоретическая величина берется из таблицы процентных точек распределения «хи-квадрат», которую легко можно найти в любом учебнике по математической статистике. Полученную нами эмпирическую величину «хи-квадрата» ( $XK$ ) мы делим на ее теоретическое значение с учетом степеней свободы. Подробнее об этом мы рассуждаем при описании метода «хи-квадрат» ( $XK$ ) в той части работы, которая описывает методы исследования.

Интересно отметить, что коэффициент вариации ( $KB$ ) и коэффициент  $T$  достаточно тонко реагируют как на объединение нескольких групп в одну, так и на введение в группу нового элемента, т.е. какого-то нового языка. Мы будем вводить в монгольскую и тюркскую семьи, и отмечать при этом повышение или понижение диффузии, которая равна величинам  $KB$  и  $T$ . Компактность будет обратна этим величинам, как мы уже указывали выше.

С помощью компактности попробуем установить, насколько правомерно включение в монгольскую семью таких генетически изолированных языков как кетский, юкагирский, нивхский, японский и корейский. Заметим, что все эти языки не могут быть признаны типологически похожими на монгольские языки, так как все они повышают диффузность монгольской семьи ( $KB = 86,24\%$ ,  $T = 0,66$ ) в той или иной степени: корейский ( $KB = 149,22\%$ ,  $T = 1,55$ ), японский ( $KB = 152,23\%$ ,  $T = 1,60$ ), кетский ( $KB = 192,89\%$ ,  $T = 1,90$ ), нивхский ( $KB = 269,33\%$ ,  $T = 4,75$ ), юкагирский ( $KB = 272,26\%$ ,  $T = 4,38$ ).

Несколько по-другому реагируют на введение этих языков тюркские языки (диффузность которых без кумандинского составила:  $KB = 150,16\%$ ,  $T = 1,66$ ): корейский ( $KB = 149,95\%$ ,  $T = 1,70$ ), японский ( $KB = 153,39\%$ ,  $T = 1,79$ ), кетский ( $KB = 161,63\%$ ,  $T = 1,98$ ), юкагирский ( $188,16\%$ ,  $T = 2,69$ ), нивхский ( $KB = 199,64\%$ ,  $T = 3,05$ ). Только корейский язык следует считать типологически похожим на тюркские языки. Достаточно похож на тюркские языки и японский, чего нельзя сказать об остальных изолированных языках. Интересно, что в некоторой степени похожи на тюркские языки бурятский ( $KB = 155,40\%$ ,  $T = 1,81$ ) и китайский ( $KB = 165,52\%$ ,  $T = 2,00$ ) языки.

Совершенно непохож на остальные тюркские языки кумандинский, который резко увеличивает диффузность тюркских языков (ср. КВ = 150,16%, Т = 1,66 и КВ = 204,32%, Т = 3,40). Странно и достаточно неожиданно, что один из тюркских языков (т.е. кумандинский), увеличивает диффузность тюркской семьи даже больше, чем такие финно-угорские языки как мари-луговой (КВ = 156,69, Т = 1,72), мари-горный (163,53%, Т = 2,06) или мансийский (КВ = 174,51%, Т = 2,35). Понятно, почему мансийский язык более отличен от тюркских языков, нежели горно-марийский или лугово-марийский языков. По всей вероятности долгие контакты последних с тюркскими языками не прошли даром.

Еще больше поражает тот факт, что венгерский язык повышает компактность тюркских языков, при этом диффузность понижается с КВ = 150,16%, Т = 1,66 до 147,36%, Т = 18,42. Это может говорить как о контактах древних венгров с древними тюрками в урало-сибирском регионе, так и о тысячелетнем стоянии предков венгров на Волге. На наш взгляд, это позволяет понять невозможность коммуникации между венграми и обскими уграми. По крайней мере распределение согласных в их звуковых цепочках различно.

Продолжая анализ финно-угорских языков, нужно отметить, что угорские и прибалтийско-финские языки показывают различные тенденции употребления групп согласных.

Величина КВ по губным согласным в группе угорских языков составила 12,28, а в группе прибалтийско-финских языков – 14,38. Посчитаем величину КВ всей финно-угорской семьи. Она составила 14,42. Если смешать угорские и пермские языки, то их общий КВ уменьшится (11,92), что может говорить о большей схожести функционирования в них губных согласных, нежели в одной только угорской группе. Можно сказать, что они дополняют друг друга, заполняя хиатусы (пробелы). Если смешать угорские и волжские языки, то губные согласные дают резкое повышение значения КВ. Это может говорить о большей типологической отдалённости угорских и волжских языков. Достаточно далеко друг от друга по этому параметру расположены пермские и волжские языки, так как их КВ = 13,86. Введение волжских языков в группу прибалтийско-финских также немного увеличивает значение КВ (14,39), но увеличение это столь незначительное (ср. 14,38–14,39), что можно говорить о близости волжских и прибалтийско-финских языков по распределению губных согласных в их звуковых цепочках. Неожиданно то, что смешение прибалтийско-финской и пермской групп даёт резкое снижение значения КВ (12,46). Это говорит о повышении компактности этой объединённой группы, хотя и не такое сильное, как при смешении угорских и пермских языков (ср. 11,92 и 12,46). Всё это указывает на то, что пермские языки близки как угорским, так и прибалтийско-финским, хотя и ближе к пер-

вым. В то же время волжские языки достаточно далеки от угорских, но близки к прибалтийско-финским. Напомним, что этот вывод мы можем сделать пока только лишь на основании распределения по языкам губных согласных.

Для того, чтобы понять общую тенденцию, нужно сложить величины КВ по семи остальным параметрам. Сложение величин коэффициента вариации возможно в силу принципа соизмеримости. Суммированное значение КВ показывает общую диффузность языковых подгрупп, групп и семей. Из суммированных значений коэффициента вариации следует, что подгруппа угорских языков более диффузна (менее компактна), чем подгруппа прибалтийско-финских (ср. КВ = 221,27%, Т = 3,77 и КВ = 185,90%, Т = 2,79) языков. Наиболее же компактна подгруппа волжских языков (КВ = 143,19%, Т = 1,02). Введение коми-зырянского языка (пермская подгруппа) в подгруппу угорских языков понижает ее диффузность до КВ = 211,67%, Т = 3,72. Это может говорить о том, что коми-зырянский язык имеет распределения согласных, которое похоже на аналогичное распределение согласных в угорских языках.

Важно понять, более ли похоже распределены согласные в волжских и угорских языках, или же в волжских и прибалтийско-финских языках. Объединение волжских и угорских языков дает менее компактную группу (ср. КВ = 210,78%, Т = 3,63), нежели объединение волжских и прибалтийско-финских языков (ср. КВ = 185,73%, Т = 2,82). Это означает, что волжские и прибалтийско-финские языки имеют больше общих тенденций в распределении согласных.

Введение изолированных азиатских языков в финно-угорскую семью (КВ = 193,13%, Т = 3,77) показывает увеличение ее диффузности: кетский (КВ = 198,04%, Т = 3,94), юкагирский (КВ = 199,17%), корейский (КВ = 199,24%, Т = 3,88), японский (КВ = 200,51%, Т = 3,91), нивхский (КВ = 206,48%). Китайский язык показывает с ними большую схожесть (КВ = 190,01%, Т = 3,65), чем указанные выше изолированные азиатские языки. Наименьшую схожесть с финно-угорскими языками показывает язык американских индейцев свит кри (КВ = 212,98%, Т = 4,47).

Самодийская языковая семья менее компактна (более диффузна), чем финно-угорская (ср. КВ = 216,14% и КВ = 193,13%), если учитывать распределение звонких согласных. Проблема заключается в том, что в ненецком и селькупском нет звонких согласных. В связи с этим, принцип соизмеримости диктует нам необходимость анализировать самодийские языки без этого параметра (признака). Таким образом, нам придется производить сравнения только по семи параметрам (признакам), вместо восьми (КВ = 128,04%, Т = 1,10). Введение изолированных азиатских языков в самодийскую семью дает следующие результаты: японский (КВ = 133,60%, Т = 1,23), кетский (КВ = 140,22%, Т = 1,25),

корейский (КВ = 150,99%, Т = 1,33), юкагирский (КВ = 191,82%, Т = 2,65), нивхский (КВ = 212,65%, Т = 3,40). Как мы видим, все они понижают компактность самодийской семьи, следовательно, типологически ей не принадлежат.

Отметим, что уральская языковая общность (КВ = 226,49%, Т = 4,58) менее компактна, чем самодийская (КВ = 128,04%, Т = 1,10) или финно-угорская (КВ = 193,13%, Т = 3,77) языковые семьи. Проанализируем изменение компактности уральской языковой общности (КВ = 226,49%, Т = 4,58) при внесении в нее изолированных азиатских и других языков: кетский (КВ = 226,50%, Т = 5,43), корейский (КВ = 228,85%, Т = 4,83) японский (КВ = 228,99%, Т = 5,01), юкагирский (КВ = 228,99%, Т = 5,18), нивхский (КВ = 233,20%, Т = 5,46), суит грасс кри (КВ = 238,19%, Т = 5,48).

Тюркская языковая семья намного компактнее финно-угорской (ср. величины их диффузности: КВ = 193,13%, Т = 3,77 и КВ = 150,16%, Т = 1,66). Наши данные подтверждают близость тюркских языков, что подчеркивается многими исследователями тюркских языков (И с а е в 1978, 109; М у с а е в 1984, 16). Вероятно, это происходит потому, что в добавок к своей генетической близости тюркские языки сохраняли друг с другом интенсивные контакты в течение длительного времени, постоянно смешиваясь между собой (К л и м о в 1971, 11; М у с а е в 1984, 33; Щ е р б а к 1970). Сравнивая совокупность финно-угорских и тюркских языков, языковеды отмечают тот факт, что прошло намного меньше времени от эпохи совместного компактного проживания тюркских племен в Центральной Азии (Б а с к а к о в 1981, 60–69; М у с а е в 1984, 31–33; Н а д ж и п 1970, 87–89; С а м о й л о в и ч 1922, 6–14; Щ е р б а к 1970, 20). В связи с этим, на наш взгляд, следует ожидать меньшую диффузность тюркской семьи, чем других семей, например, финно-угорской индоевропейской, языки которых разошлись намного ранее. Вне всякого сомнения стоит тот факт, что находясь в контакте языки уподобляются, и наоборот, потеряв контакт, они теряют былое сходство, т.е. типологически расходятся.

Анализируя по 8 указанным признакам 6 языков тунгусо-маньчжурской языковой семьи (эвенкийский, негидальский, эвенский, нанайский, ульчский и ороцкий), заметим, что ее компактность довольно высокая. Так, ее диффузность составила КВ = 148,79%, Т = 1,73. Она почти такая же как у тюркской (153,39%, Т = 1,79), и намного (почти в два раза) ниже диффузности палеоазиатской (КВ = 271,14%, Т = 3,71) семьи.

Проверим, как будет реагировать компактность на введение в тунгусо-маньчжурскую семью с одной стороны некоторых изолированных языков, а с другой – языка, который к ней принадлежит. В данном случае орокского языка. Введение всех языков, кроме орокского, повышает диффузность тунгусо-маньчжур-

ской семьи: юкагирский (КВ = 148,87%, Т = 1,93), кетский (150,00%, Т = 1,84), корейский (159,22%, Т = 2,01), японский (КВ = 169,23%, Т = 2,25), суит грасс кри (КВ = 171,69%, Т = 2,67), нивхский (КВ = 183,03%, Т = 2,99). Введение орокского языка резко понижает диффузность до КВ = 139,26%, Т = 1,59. Это говорит о явной типологической схожести орокского языка с указанными выше шестью тунгусо-маньчжурскими языками. Интересно, что введение в эту группу китайского языка тоже понижает диффузность, хотя и не так сильно (КВ = 144,36, Т = 1,72). Это говорит о типологической схожести тунгусо-маньчжурских языков и китайского по употреблению групп согласных в звуковой цепочке. Это не удивительно, так как китайский язык находится с тунгусо-маньчжурскими языками во многолетнем языковом контакте.

Продолжим процедуру введения изолированных азиатских языков в другие языковые семьи, на этот раз, в палеоазиатскую. Еще раз напомним, что ее компактность измерялась по 5 языкам (чукотский, корякский, ительменский, науканский и имаглинский диалекты эскимосского языка). Ее диффузность составила КВ = 271,14%, Т = 3,71. Получаем следующий упорядоченный ряд: японский (263,67%, Т = 4,21), корейский (275,58%, Т = 4,21), китайский (304,55%, Т = 5,24), суит грасс кри (305,38%, Т = 4,99), кетский (319,83%, Т = 8,22) юкагирский (342,61%, Т = 7,46), нивхский (359,74%, Т = 9,28). Японский и корейский языки выказывают несомненную типологическую близость к указанным пяти палеоазиатским.

Перед языковедом часто встает задача, включать ли в список языков какой-либо семьи мертвые языки, которые ранее в нее предположительно входили. Или же эти языки были настолько отличными, что мало напоминают современные языки, которые от них произошли. Перед нами вставала проблема следует или не следует в тюркскую семью включать древнетюркский. Такая же проблема стоит в отношении древнеанглийского и языков германской группы или латинского языка и романской группы индоевропейской семьи. В принципе, можно проверить насколько древнеанглийский (или латинский) язык отличается от других германских (или романских) языков по дистрибуции согласных только на основе величины компактности, но лучше сделать это на базе двух предложенных методов: метода компактности и метода расстояний. В данной главе мы измерим компактность германской группы с древнеанглийским языком и без него. Измерить конкретные фоностатистические расстояния от древнеанглийского до каждого из германских языков предполагается в следующей главе.

Мы включили древнеанглийский язык в группу германских языков, состоящую из современного английского нидерландского (голландского), немецкого, датского, норвежского и шведского языков. По встречаемости губных

согласных древнеанглийский типологически близок в большей степени датскому языку, что видно из следующего упорядоченного ряда: современный английский (13,05%) – нидерландский (12,03%) – датский (11,95%) – древнеанглийский (11,81%) – норвежский (10,60%) – шведский (11,00%) – немецкий (9,88%). В упорядоченном ряду переднеязычных согласных древнеанглийский язык занимает место рядом с норвежским: немецкий (42,60%), шведский (41,20%) – современный английский (40,79%) – нидерландский (39,98%) древнеанглийский (33,88) – норвежский (32,70%) – датский (29,40%). Таким же образом можно рассмотреть все остальные упорядоченные ряды. Суммарная величина коэффициента вариации по 8 признакам без древнеанглийского равна 196,09%, т.е. 24,51% в среднем по каждому признаку. После введения в группу германских языков древнеанглийского языка суммарная величина КВ уменьшилась до 188,48%, т.е. до 23,56% в среднем по признаку. Это говорит об уменьшении вариации по признакам, что, в свою очередь, говорит о принадлежности древнеанглийского языка к группе германских языков. Это же подтверждает уменьшение коэффициента Т (ср. 2,32 и 2,30).

Проверим, насколько близки германским языкам такие изолированные индоевропейские языки как армянский, греческий и албанский. Поочередно помещая их в германскую группу, получаем следующий упорядоченный ряд по величине диффузности: греческий (196,99%, Т = 2,64), албанский (КВ = 202,30%, Т = 2,69), армянский (КВ = 206,31, Т = 2,75), армянский (206,31%, Т = 2,75). Все данные языки увеличивают диффузность германской группы и, следовательно, с ней типологически не схожи.

Известно, что романские языки произошли от латинского языка. Проверим по методу компактности, принадлежит ли типологически латинский язык к группе романских языков. Введение латинского языка в группу романских языков повышает ее компактность по КВ: ее диффузность понижается с КВ = 222,49%, Т = 2,84 до 209,97%, Т = 2,66. Из этого следует, что латинский язык типологически схож с романскими языками и может быть включен в эту группу как ее член.

Интересно, что искусственный язык «Эсперанто» очень сильно похож по звучанию на романские языки. Введение эсперанто в эту группу понижает ее диффузность с 222,49%, Т = 2,84 до 204,48%, Т = 2,54.

Ближе всего по звучанию к романским языкам греческий (КВ = 214,78%), затем идет албанский (КВ = 217,08%), затем – армянский (КВ = 218,56%).

По индоарийской группе индоевропейской языковой семьи нам удалось проанализировать данные по пяти языкам: хинди, бенгальскому, гуджарати, маратхи и цыганскому (КВ = 163,20%, Т = 1,83). Будем поочередно помещать в индоарийскую группу как изолированные индоевропейские языки, так и

древние индийские языки. В результате получаем следующий упорядоченный ряд: албанский (156,59%,  $T = 1,97$ ), греческий (167,13%,  $T = 2,14$ ), махараштри (168,33%,  $T = 2,25$ ), санскрит (168,77%,  $T = 2,44$ ), армянский (169,08%,  $T = 2,68$ ), ведийский (193,11%,  $T = 3,10$ ). Из этих значений видно, что албанский язык типологически входит в группу индоарийских языков. Греческий язык имеет более похожие звуковые цепочки, нежели санскрит или среднеиндийский пракрит махараштри. Выше мы видели, что древнеанглийский и латинский типологически входили в соответствующие группы. Вероятно, современные индоарийские языки сильнее отклонились от их первоначальной типологии распределения групп согласных в звуковой цепочке.

Наша иранская группа индоевропейских языков включает такие языки как иранский, таджикский, осетинский и курдский ( $KB = 105,64\%$ ,  $T = 0,68$ ). Как мы видим она очень компактна. Может быть поэтому, все изолированные языки повышают ее диффузность: албанский ( $KB = 125,60\%$ ,  $T = 1,22$ ) армянский ( $KB = 138,18\%$ ,  $T = 1,49$ ), греческий ( $KB = 139,74\%$ ,  $T = 1,58$ ). Следовательно, типологически они в нее не входят.

Наша славянская группа включила все основные славянские языки: русский, украинский, белорусский, серболужицкий, сербскохорватский, болгарский, польский, словацкий и чешский ( $KB = 121,71\%$ ,  $T = 1,34$ ). Многие языковеды считают, что когда-то существовало балто-славянское языковое единство. В этой связи было бы интересно рассмотреть типологическую близость балтийских языков к славянским. Включение их в единую группу дало следующие результаты: литовский ( $KB = 121,55\%$ ,  $T = 1,35$ ), латышский ( $KB = 125,77\%$ ,  $T = 1,43$ ). Из этого следует, что, строго говоря, они не принадлежат типологически одной группе. При этом они типологически ближе к славянским языкам, чем изолированные языки: греческий ( $KB = 134,37\%$ ,  $T = 1,67$ ), албанский ( $KB = 143,63\%$ ,  $T = 1,77$ ) и армянский ( $KB = 155,59\%$ ,  $T = 2,09$ ).

Известно, что древнетюркский язык имеет в себе уйгуро – огузо – кыпчакские характеристики, поэтому он родственен всем современным тюркским языкам.

Его уйгуро – огузо – кыпчакские характеристики ясно видны из древнетюркских текстов (К о н о н о в 1982, 307–318; М а л о в 1951, 7–32; М у с а е в 1984, 61–81). Многие тюркологи отмечают, что современные литературные тюркские языки отличаются друг от друга сильнее, чем каждый из них отличается от языков древнетюркских памятников (И с х а к о в а 1973; М у с а е в 1984, 61–81; Т е н и ш е в 1973, 59–60; 1976, 170). Мы ввели данные по величине консонантного коэффициента древнетюркского в группу аналогичных данных по тюркским языкам. Компактность этой объединенной группы увеличилась (Т а м б о в ц е в 1986). Так произошло, по-видимому, в связи с тем, что

древнетюркский язык имеет форму наддиалектного койне (Тенишев 1973, 59–61; 1976, 170–171). Именно поэтому его введение в семью современных языков нисколько не ухудшает обобщенную картину компактности. К нашему большому сожалению, мы не смогли выработать достаточно надежный алгоритм перевода древнетюркских «букв» в четко им соответствующие звуки, поэтому мы отказались от подсчетов частоты встречаемости фонем в древнетюркских текстах. Нами были взяты подсчеты по древнеанглийскому, латинскому и древнегреческому языкам, близость которых к другим языкам мы обсудим в следующей главе. В отношении же тюркских языков заметим, что в таблицах мы даем их в алфавитном порядке, в отличие от финно-угорских или индоевропейских языков, а не в порядке развертки их генетического дерева. Как считают многие тюркологи, строгое построение генетического дерева тюркских языков в принципе невозможно вследствие их слишком большой близости: их многие черты слишком переплетаются. Кроме того, следует учесть скептическое отношение со стороны лингвистов к теории родословного дерева, которое в последние несколько десятков лет находит все больше и больше сторонников (Жирмунский 1964, 104; Чемоданов 1982, 94–98). Многие другие лингвисты считают, что отказ от шлейхеровского генеалогического дерева не приводит к отказу от генетического родства языков (Шарадзе 1982, 71).

Следует отметить, что величина компактности группы является достаточно тонким показателем схожести языков даже по одному параметру. Возьмем, к примеру, величину консонантного коэффициента, т.е. отношение частоты встречаемости согласных к гласным в звуковой цепочке. Проанализируем по консонантному коэффициенту урало-алтайскую языковую общность, куда, как мы уже отмечали, лингвисты относят финно-угорские, самодийские, тюркские, тунгусо-маньчжурские и монгольские языки (Баранникова 1973, 299–306; Головин 1966, 264; СИИЯРС 155). Рассмотрим компактность условной тюркско-финно-угорской группы. В связи с тем, что мы анализируем языки прежде всего на синхронном уровне, древнетюркский язык из этой условной группы будет исключен. Компактность этой условной группы составила 12,39. Отметим, что компактность тюркской группы 32,16, а финно-угорской 8,32, то есть почти в четыре раза ниже. Следовательно, данные финно-угорской группы сильно понижают компактность объединенной группы. Вероятно, это может в какой-то степени говорить о том, что тюркские языки более схожи типологически по распределению гласных и согласных.

Продолжим анализ семей с учетом всех 8 признаков по аналогичному алгоритму. Объединение в одну супер-семью тюркских и монгольских языков уменьшает компактность (ср. 154,55 и 170,90). Объединение самодийских и

финно-угорских языков в уральскую супер-семью также понижает компактность (ср. 203,90 и 231,06). Величина компактности тунгусо-маньчжурских языков находится примерно на уровне тюркских (ср. 154,72 и 154,55). Алтайская языковая общность имеет компактность меньшую (207,76), чем компактность каждой их входящих в неё семей, что не говорит в пользу «алтайской гипотезы». Урало-алтайская языковая супер-общность ещё менее компактна (247,82). Величина диффузности индоевропейской языковой семьи меньше (223,79) урало-алтайской (247,82), но больше алтайской (207,76), что говорит в пользу «алтайской гипотезы». Действительно, если большинство лингвистов поддерживают существование индоевропейской семьи, то нельзя отвергать существование алтайской языковой общности, чья компактность выше.

Наш метод схожести языков на фонетическом уровне можно применить и к группе языков, которые называются «балканским языковым союзом». В него входят языки разных групп индоевропейской семьи: славянской (болгарский и сербскохорватский), романской (румынский), и изолированные языки (греческий и албанский). Некоторые лингвисты считают необходимым включать в балканский языковой союз и армянский язык. Проверим нашим методом как будет изменяться компактность – диффузность при введении в него таких языков как армянский и турецкий. На наш взгляд, турецкий язык на протяжении нескольких веков сильно влиял на языки балканского языкового союза, что не может не проявиться на фонетическом уровне. Величина диффузности балканского языкового союза по величине коэффициента вариации составила 237,89%,  $T = 2,90$ . Введение армянского языка несколько понижает его диффузность ( $KB = 234,23\%$ ,  $T = 3,13$ ), что может говорить о правомерности включения армянского языка в этот союз языков. Введение в него турецкого языка еще более понижает его диффузность ( $KB = 224,92\%$ ,  $T = 2,89$ ), что говорит о еще большей схожести языков балканского союза с турецким по дистрибуции групп согласных в их звуковых цепочках.

Из языков австронезийской языковой семьи были привлечены следующие: тагалог, индонезийский, даякский, гавайский и самоанский ( $KB = 369,65\%$ ,  $T = 7,22$ ). Как мы видим, эта семья достаточно диффузна. Может быть поэтому японский язык повышает ее компактность ( $KB = 340,33\%$ ,  $T = 6,49$ ).

Весьма компактной по 7 признакам (без звонких) оказывается языковая семья австралийских языков ( $KB = 95,04\%$ ,  $T = 1,29$ ). Даже с привлечением данных по звонким она достаточно компактна ( $KB = 150,59\%$ ,  $T = 3,83$ ). Она намного компактнее, чем семья языков американских индейцев ( $KB = 346,95\%$ ,  $T = 8,59$ ).

Более детально величины компактности можно увидеть из многочисленных таблиц, которые приводятся отдельно.

Из всего вышесказанного можно заключить:

1) Переднеязычные и смычные согласные распределены по языковым семьям наиболее компактно.

2) Звонкие шумные согласные – это наиболее изменяемый признак, который может отсутствовать в некоторых языках.

3) Самой компактной по 7 признакам (без частоты звонких шумных), по суммарному коэффициенту вариации и коэффициенту T, является монгольская языковая семья. Далее идут самодийская, тюркская, тунгусо-маньчжурская и финно-угорская. Палео-азиатская языковая семья настолько диффузна, что скорее напоминает языковую общность, а не семью.

4) Обычно языковая подгруппы компактнее группы, а языковая группа компактнее языковой семьи. В свою очередь языковая семья компактнее языковой общности или супер-общности. Наименее компактной является языковая супер-общность, которая объединяет все языки мира.

5) Объединение двух групп или двух семей обычно дает более диффузное образование, чем каждая из них.

## LINGUISTIC TAXONOMY: DENSITY OF LANGUAGE SUBGROUPS, GROUPS AND FAMILIES

### *Summary*

Language subgroups, groups, families and unities were investigated from the point of view of their dispersion in the way it was first proposed in (Тамбовцев 1986). We consider how compact this or that language taxon (i.e. subgroup, group, family or unity) on the basis of distribution of certain consonantal groups in the speech sound chain. Therefore, one can speak about a compact or disperse language family. If language taxon is compact, then its internal connections are shorter than its outer connections. Actually, the same notion of compact object is accepted in pattern recognition. The more compact the family, the more correctly its languages are chosen. If we put in the family a language, which does not belong to the family, then the dispersion of the family rises, thus it becomes less compact. If the language has a similar sound chain, then the dispersion of the group remains the same or becomes less. It means that the family became more compact. In this case we speak about the typological properties of the families. We measure the dispersion of a family by the sum of dispersions of 8 phonostatistical features: frequency of occurrence of labial, front, palatal, velar, sonorant, occlusive, fricative and voiced consonants. It is important that the features do not intersect.

The values of the coefficient of variance (V) and T coefficient show the degree of dispersion. The principle is the greater the dispersion, the less compact the family.

We have chosen the coefficient of variance and T coefficient since they both keep to the law of commensurability. Comparing different languages of different language families and different morphological structures was possible since all of them have the same 8 phonetic features mentioned above. We have considered Indo-European, Turkic, Mongolian, Tungus-Manchurian, Samoyedic, Finno-

Ugric, Paleo-Asiatic, Austronesian, Australian and American Indian language families. The most compact is Mongolian ( $V = 10.78\%$ ;  $T = 0.08$ ), the least compact is Austronesian ( $V = 46.21\%$ ;  $T = 0.90$ ). Tungus-Manchurian ( $V = 17.41\%$ ;  $T = 0.20$ ) is more compact than Samoyedic ( $18.29\%$ ;  $T = 0.16$ ), Turkic ( $18.77\%$ ;  $T = 0.21$ ), Indo-European ( $V = 28.00\%$ ;  $T = 0.61$ ) or Paleo-Asiatic ( $V = 33.89\%$ ;  $T = 0.46$ ). Language groups are more compact than language families. Iranian group ( $V = 13.21\%$ ;  $T = 0.09$ ) of Indo-European language family is the most compact, the least compact is Romanian ( $V = 26.25\%$ ;  $T = 0.33$ ). Slavonic group ( $V = 15.21\%$ ;  $T = 0.17$ ) is more compact than Indo-Arien ( $V = 20.40\%$ ;  $T = 0.23$ ) or Germanic ( $V = 24.51\%$ ;  $T = 0.29$ ). Volga group of the Finno-Ugric language family ( $V = 17.90\%$ ;  $T = 0.13$ ) is more compact than Ugric ( $V = 27.66\%$ ;  $T = 0.47$ ) or Finnic ( $V = 29.24\%$ ;  $T = 0.35$ ) group. Altaic super-family is rather compact ( $V = 25.97\%$ ;  $T = 0.45$ ). It is more compact than Indo-European ( $V = 28.00\%$ ;  $T = 0.61$ ) or Paleo-Asiatic family ( $V = 33.89\%$ ;  $T = 0.46$ ). This fact supports those linguists who consider Altaic as a family, not a super family. Uralic super family is less compact than that ( $V = 28.31\%$ ;  $T = 0.47$ ), though it is more compact than Balkan ( $V = 29.74\%$ ;  $T = 0.36$ ) language unity (Sprachbund). Ural-Altaic language unity ( $V = 30.98\%$ ;  $T = 0.88$ ) is more (much more) compact than American Indian ( $43.37\%$  1.07) language unity. Measuring the typological density of language taxons (subgroups, groups, families, unities) may help to understand how natural (or correct) these taxons are.

## ЛИТЕРАТУРА

Апресян Ю. Д., 1966, Алгоритм построения классов по матрице расстояний, – Машинный перевод и прикладная лингвистика, 9, Москва, 72–79.

Аркадьев А. Г., Э. М. Браверман, 1964, Обучение машины распознаванию образов, Москва.

Аркадьев А. Г., Э. М. Браверман, 1971, Обучение машины классификации объектов, Москва.

Баранникова Л. И., 1973, Введение в языкознание, Саратов.

Баскаков Н. А., 1981, К историко-типологической фонологии тюркских языков, – Вопросы языкознания, 1, 60–69.

Бонгард М. М., 1967, Проблемы узнавания, Москва.

Васильев В. И., 1969, Распознающие системы, Киев.

Вентцель Е. С., 1964, Теория вероятностей, Москва.

Гамкрелидзе Т. В., Вяч. Вс. Иванов, 1984, Индоевропейский язык и индоевропейцы, I–II, Тбилиси.

Головин Б. Н., 1966, Введение в языкознание, Москва.

Дидэ Э., 1985, Методы анализа данных: Подход, основанный на методе динамических сгущений, Москва.

Друянов Л. А., 1982, Законы природы и их познание, Москва.

Елисеева И. И., В. О. Рукавишников, 1977, Группировка, корреляция, распознавание образов (Статистические методы, классификация и измерение связей), Москва.

Елисеева И. И., В. О. Рукавишников, 1982, Логика прикладного статистического анализа, Москва.

Жирмунский В. М., 1964, Введение в сравнительно-историческое изучение германских языков, Москва–Ленинград.

- Загоруйко Н. Г., 1972, Методы распознавания и их применение, Москва.
- Загоруйко Н. Г., 1978, Таксономия в анизотропном пространстве, – Эмпирическое предсказание и распознавание образов (Вычислительные системы 76), Новосибирск, 26–33.
- Загоруйко Н. Г., 1981, Методы обнаружения закономерностей, Москва.
- Исаев М. И., 1978, О языках народов СССР, Москва.
- Исхакова С. М., 1973, Древнетюркские элементы в народноразговорном языке западносибирских татар, – Советская тюркология, VI, 36–39.
- Климов Г. А., 1971, Вопросы методики сравнительно-генетических исследований, Ленинград.
- Кондаков Н. И., 1975, Логический словарь-справочник, Москва.
- Кононов А. Н., 1982, История изучения тюркских языков в России, Ленинград.
- Кордонский С. Г., 1983, Таксоны и аналитические объекты как системы, – Системный метод и современная наука, Новосибирск, 141–150.
- Кульбак С., 1967, Теория, информация и статистика, Москва.
- Малов С. Е., 1951, Памятники древнетюркской письменности: Тексты и исследования, Москва–Ленинград.
- Малов С. Е., 1952, Древние и новые тюркские языки, – Изв. АН СССР, Отд. лит. и языка, II, 137–142.
- Мартынов В. В., 1983, Язык в пространстве и времени: К проблеме глоттогенеза славян, Москва.
- Мельников Г. П., 1978, Системология и языковые аспекты кибернетики, Москва.
- Миркин Б. Г., 1976, Анализ качественных признаков (математические методы и модели), Москва.
- Миркин Б. Г., 1980, Анализ качественных признаков и структур, Москва.
- Мусаев К. М., 1984, Лексикология тюркских языков, Москва.
- Наджиб Э. Н., 1970, О средневековых литературных традициях и смешанных письменных языках, – Советская тюркология, 1970, I, 87–89.
- Никифоров А. М., 1984, Метризация некоторых описаний при анализе разнотипных данных, – Обнаружение эмпирических закономерностей с помощью ЭВМ, Вычислительные системы, 102, Новосибирск, 92–93.
- Ожегов С. И., 1970, Словарь русского языка, Москва.
- Павловский З., 1967, Введение в математическую статистику, Москва.
- Петров В. В., 1974, Критический анализ проблемы несоизмеримости в современной «философии науки», Новосибирск.
- Петров В. В., 1988, Язык и логическая теория: В поисках новой парадигмы, – Вопросы языкознания, IV, 39–48.
- Пфанцагель И., 1976, Теория измерений, Москва.
- Рассадин В. И., 1988, Роль контактов в образовании тюркоцимонгольской языковой общности, – Вопросы языкознания, IV, 103–108.
- Романовский В. И., 1947, Применения математической статистики в опытном деле, Москва–Ленинград.
- Самойлович А., 1922, Некоторые дополнения к классификации турецких языков, Петроград.
- Селютин И. Я., 1983, Кумандинский консонантизм: Экспериментально-фонетическое исследование, Новосибирск.
- СИИЯРС – Сравнительно-историческое изучение языков разных семей: (Задачи и перспективы), Москва, 1982.
- СИС – Словарь иностранных слов, Москва, 1964.

- Тамбовцев Ю. А., 1982, Эмпирическое распределение частотности фонем в казымском диалекте хантыйского языка, – Уч. зап. Тартусского ГУ, 628, 122–134.
- Тамбовцев Ю. А., 1983, Эмпирическое распределение частотности фонем в орочском языке, – Уч. зап. Тартусского ГУ, 658: Квантитативная лингвистика и стилистика, Тарту, 124–135.
- Тамбовцев Ю. А., 1985, Устойчивость консонантного коэффициента в зависимости от величины выборки в некоторых финно-угорских языках, – Исследование языков народов СССР, Новосибирск, 110–114.
- Тамбовцев Ю. А., 1986, Консонантный коэффициент в языках разных семей, Новосибирск.
- Тамбовцев Ю. А., 1988, Некоторые фоностатистические характеристики языка барабинских татар, – Фонетика и грамматика языков Сибири, Новосибирск, 135–139.
- Тамбовцев Ю. А., 1990, Компактность финно-угорской языковой семьи по данным консонантного коэффициента, – Лингвистика Урала, XXVI (1), 13–20.
- Тенишев Э. Р., 1973, Тюркская историческая диалектология и Махмуд Кашгарский, – Советская тюркология, 1973, VI, 49–71.
- Тенишев Э. Р., 1976, О наддиалектном характере языка тюркских рунических памятников, – Turcologica (К 70-летию акад. А. Н. Кононова), Ленинград, 163–179.
- Третьяков П. Н., 1970, У истоков древнерусской народности, Ленинград.
- Третьяков П. Н., 1982, По следам древних славянских племен, Ленинград.
- Турбович и др., 1971 – И. Т. Турбович, В. Г. Гитис, В. К. Маслов, Оpozнание образов, Москва.
- Убрятова Е. И., 1966, Якутский язык, – ЯНСТЮ, Москва, 261–324.
- Хургин Я. И., 1979, Как объять необъятное, Москва.
- Чемоданов Н. С., 1982, Германские языки, – Сравнительно- историческое изучение языков разных семей, Москва, 91–115.
- Шараденидзе Т. С., 1982, Родство языков, процессы дивергенции., конвергенции и соответствующие им классификации языков мира, – Проблемы родства, Москва, 63–107.
- Щербак А. М., 1970, Сравнительная фонетика тюркских языков, Ленинград.
- ЭК – Энциклопедия кибернетики, Киев, 1975.