

ФОНОТИПОЛОГИЧЕСКАЯ БЛИЗОСТЬ ЛИНГВИСТИЧЕСКИХ ОБЪЕКТОВ ПО КРИТЕРИЮ «ХИ-КВАДРАТ» *

В статье предлагается метод исследования близости языков, основанный на применении информационных технологий и прикладных математических методов в языкознании. После обработки большого массива протранскрибированных текстов на компьютере была вычислена схожесть лингвистических объектов с помощью методов математической статистики и распознавания образов. Анализируется типология строения звуковых цепочек в языках и их диалектах в качестве образа. На основании доминантных фонетических признаков звуковых картин построены фонотипологические расстояния между лингвистическими объектами.

Ключевые слова: типология, звуковая цепочка, язык, диалект, звуковые расстояния, языковой таксон.

Введение

Языкознание в настоящее время находится на таком этапе развития, когда без применения информационных технологий продвижение вперед невозможно. С помощью компьютера нами были обработаны большие массивы протранскрибированных текстов, что позволило получить надежные статистические результаты в численном виде по звуковым картинам различных языков. На основе применения информационной технологии была получена возможность решения лингвистической задачи «язык-диалект», которая не могла быть решена в рамках классического языкознания.

Язык и диалект в данном случае трактуются как фонетические объекты. Таким образом, язык (как и любой другой объект) можно интерпретировать в терминах распознавания образов, где объекты понимаются как отдельные представители образов в пространстве признаков [1. С. 7]. Поскольку выбранные нами фонетические признаки имеют количественное выражение, мы можем сказать, что к изучению языков в данном случае применена самая сильная, т. е. абсолютная, шкала. Точная нумерическая информация позволяет измерять точные расстояния как между отдельными объектами, так и между таксонами [2. С. 61]. Под таксонами мы будем понимать подгруппы, группы, семьи или общности (супер-семьи) языков. В то же время таксоном может считаться один язык, если он имеет несколько диалектов. В том случае, когда мы имеем какой-либо лингвистический объект (т. е. язык или диалект), который мы должны поместить в какой-либо лингвистический таксон (т. е. подгруппу, группу, семью и т. д.), мы можем это сделать на основе типологического расстояния, измеренного по выбранным признакам. По величине расстояния между лингвистическими объектами можно определить величину функции принадлежности объекта к образу [2. С. 167–169]. Если расстояние от лингвистического объекта до некоторого образа (таксона)

* Автор выражает благодарность за консультацию по диалектам алтайского языка и применению методов математической статистики профессору А. А. Кочееву (Новосибирский государственный университет). Мы также благодарим профессора Н. Г. Загоруйко (Институт математики СО РАН) за внимательное прочтение статьи и его замечания по применению статистических критериев и методов распознавания образов.

меньше, чем до других образов, то можно считать, что объект принадлежит к данному образу. Велико или невелико расстояние, можно судить по величине коэффициента ТМВ [3]. Вычислить этот коэффициент достаточно просто. Фактически это величина отношения полученного значения «хи-квадрата» к его теоретическому значению, представленному в таблицах с учетом уровня значимости и числа степеней свободы.

В данной работе осуществлен типологический подход, в основе которого лежит выделение каких-либо признаков систем лингвистических объектов и их дальнейшая группировка с помощью обобщенной модели на основе этих выделенных признаков. Таким образом, лингвистический объект представлен уже не целостно, а в виде некоторого количества признаков, которые имеют нумерические значения. Этот подход обычно применяется в распознавании образов, но мы также используем его и в языкознании. Мы можем назвать этот подход численно-типологическим. Следует отметить, что применение типологического подхода на основе определенных выбранных признаков, но без их нумерического значения, осуществлялось в языкознании и ранее. Наиболее последовательным представителем типологического подхода в языкознании является В. Д. Аракин. Таким образом, типологический подход позволяет сопоставлять не только генетически родственные языки (например, русский, украинский и белорусский), но и достаточно далекие языки, например, русский и английский [4. С. 5–8]. Концепция типологического подхода в языкознании позволяет сравнивать любые языки мира с разными структурами не только на фонетическом, но и на других уровнях языка. В то же время важность сопоставления различных языков в типологическом плане на фонетическом уровне подчеркивается в современном языкознании (см., например, [5. С. 455–456]). Метод фундаментального типологического сопоставления позволил В. Д. Аракину сравнивать между собой такие разные языки, как славянские, германские, тюркские, малайско-полинезийские и другие языки с разным генетическим происхождением и разных систем и структур [4].

Обсуждая проблему языка и диалекта, Д. И. Эдельман отмечает, что в условиях отсутствия письменности проблема квалификации степени самостоятельности локальных языковых разновидностей относительно друг друга и окружающих языков представляет большие трудности в современной лингвистике. Вопрос о том, являются ли языковые образования того или иного ареала а) самостоятельными близкородственными языками; б) диалектами какого-либо единого языка; в) диалектами, составляющими «переходную зону», или «зону вибрации», между двумя (или более) родственными языками; г) рядом близкородственных языков с относящимися к ним диалектами, не получает однозначного решения [6]. Мы полностью согласны с Д. И. Эдельман в том, что до сих пор не ясны лингвистические критерии, на основании которых два лингвистических объекта следует считать самостоятельными языками или диалектами одного языка. Мы попытались выработать такой критерий.

Еще раз отметим, что в общем виде диалект и язык можно определять как лингвистический объект. При определении степени самостоятельности лингвистических объектов, т. е. языков и их диалектов, мы используем фоностатистический критерий близости звуковых картин [7; 8]. Под звуковой картиной языка мы подразумеваем совокупность звуковых цепочек анализируемых языковых объектов. Исследование проводилось на материале финно-угорских, монгольских и тюркских языков. С целью сопоставления привлекались данные, полученные ранее при изучении языков индоевропейской семьи.

Проблему языка и диалекта поставили еще древнегреческие философы Сократ и Аристотель, которые пытались решить ее в общем, т. е. как проблему целого и частного. Известно, что частное включает в себя характеристики целого. Действительно, можно ли диалект считать частью определенного языка или же этот языковой объект следует считать самостоятельным языком? Иными словами, служат ли отличия, которые характеризуют диалект, подтверждением общих тенденций функционирования языка и его диалекта? В связи с тем, что любой язык достаточно вариативен, это является трудной задачей. Диалект от языка может отделять зона неопределенности, которая будет тем шире, чем больше их схожесть.

Таблица 1

Частота встречаемости согласных и гласных звуков,
% ко всем фонемам

№	Звуки	Группы фонем									
		ЯКТ	ДЛГ	АКЖ	АЧЛ	ТКЗ	ТБР	КРТ	ТЧЛ	ЭРЗ	МКШ
1	Губные	6,10	8,43	5,98	7,87	8,03	9,04	9,79	11,03	10,21	10,96
2	Переднеязычные	32,77	32,45	38,25	36,86	35,52	34,60	34,30	35,18	41,17	40,09
3	Среднеязычные	5,22	3,53	2,54	2,31	3,19	2,13	3,11	1,87	2,12	1,81
4	Заднеязычные	13,27	12,72	11,71	10,70	10,92	11,35	10,46	10,85	6,86	7,79
5	Сонорные	24,29	23,03	23,63	20,57	26,54	25,73	26,29	24,17	22,01	20,35
6	Смычные шумные	22,14	28,48	29,12	27,63	23,17	23,54	21,55	25,35	21,84	22,77
7	Щелевые шумные	10,92	5,62	5,73	9,54	7,95	7,85	9,82	9,41	16,51	17,53
8	Звонкие шумные	9,17	11,98	11,94	10,76	11,93	7,22	12,74	10,91	11,70	11,31
9	Гласные	42,64	42,87	41,52	42,26	42,34	42,88	42,34	41,07	39,64	39,35

Материал исследования

Материалом исследования послужили тексты на различных языках, которые были представлены в фонетической транскрипции (табл. 1). Были взяты выборки по различным языкам (тюркским, финно-угорским и др.). Выборки фонем по лингвистическим объектам имеют следующий объем: долганский (ДЛГ) – 5 407; якутский (ЯКТ) – 23 6245; алтай-кижи (АКЖ) – 190 525; алтай-чалкан (АЧЛ) – 5 147; татарский казанский (ТКЗ) – 31 000; татарский барабинский (ТБР) – 67 569; крымско-татарский (КРТ) – 10 362; татарский чулымский (ТЧЛ) – 5 602; эрзя-мордовский (ЭРЗ) – 148 164; мокша-мордовский (МКШ) – 143 720.

Методика исследования

Расстояние между лингвистическими объектами измеряется при помощи величины коэффициента ТМВ, вычисленного на основе критерия хи-квадрат. Известно, что критерий хи-квадрат используется в статистических исследованиях для сравнения распределений объектов двух совокупностей на основе измерений в числовом виде признаков в двух независимых выборках. Данный статистический критерий широко известен. В рамках статьи невозможно, да и нет необходимости вдаваться в тонкости математических подсчетов с помощью критерия хи-квадрат. Его описание можно найти в любых учебниках по математической статистике или математической лингвистике (например, [9]). Кроме того, они даются в компьютерных приложениях в виде таких программ, как «Статистика», «Математика», «XL» и др. Критерий хи-квадрат применялся нами ранее для определения типологической близости языков различных семей [10; 11].

Важность применения критерия хи-квадрат для лингвистических исследований заключается в том, что он позволяет исключить субъективизм суждений путем точного сопоставления любых лингвистических объектов, у которых признаки выражены в числовых величинах. Например, сравнение числовых значений признаков звуковых картин венгерского, хантыйского и мансийского языков подтвердило правомерность классификации венгерского языка в качестве члена угорской группы финно-угорской семьи. В то же время оно не позволило от-

нести его к подгруппе обско-угорских языков, куда традиционно относятся хантыйский и мансийский языки. Наш метод позволил выделить его в отдельную, венгерскую, подгруппу, которая до сих пор еще не выделялась внутри угорской группы [7. С. 123].

При трактовке полученных данных следует учитывать наличие обратной зависимости между близостью языков и фонетическими расстояниями между ними: схожесть языков тем больше, чем меньше расстояния, и наоборот [12]. Таким образом, величина хи-квадрат показывает несхожесть языков и диалектов. Мы воспользовались методом сравнения двух выборок, которые могут иметь разный объем [13].

В качестве основополагающего критерия мы предложили ввести понятие «лингвистическая компактность». Его сущность заключается в том, что два типологически близких друг другу лингвистических объекта дают один лингвистический таксон, если они близки друг другу. Это покажет типологическое расстояние [14]. Чем меньше расстояние, тем более похожи эти объекты. В то же время чем меньше сумма расстояний в каком-либо таксоне, тем более похожи лингвистические объекты в этом таксоне. Под таксоном, как мы уже указывали выше, мы имеем в виду любую группу языков, даже такую, которая состоит только из двух членов, т. е. любая пара языков или диалектов [15. С. 131]. Так, можно говорить о таксоне мансийского языка, представленного в данном случае сосвинским и кондинским диалектами. Карельский язык состоит из трех диалектов, татарский язык представлен также тремя диалектами, алтайский язык имеет пять диалектов и т. д.

Выбор признаков

В типологических исследованиях важно определить базовые признаки [10], на которых строится наше сравнение. В данном случае было выбрано 9 доминантных фонетических признаков, к которым мы относим показатели частоты встречаемости восьми групп согласных в звуковых цепочках языка: губных, переднеязычных, среднеязычных, заднеязычных, сонорных, шумных смычных, шумных щелевых, шумных звонких; а также гласных [11]. Фактически данные фонетические признаки являются необходимыми и достаточными для распознавания звуковых картин различных языков. Обычно в распознавании образов признаки должны быть проверены с точки зрения их необходимости и достаточности [1. С. 55]. Нам нет необходимости делать такую проверку, так как мы взяли все базовые артикуляционные признаки согласных, которые выделяются в фонетике. Гласные берутся без деления на группы потому, что данные 9 признаков хорошо различают различные лингвистические объекты на фонетическом уровне. Мы согласны с Н. Г. Загоруйко, что нужно по возможности минимизировать количество признаков, поскольку это уменьшает размерность пространства, что, в свою очередь, уменьшает затраты усилий при распознавании объектов [1. С. 54–57; 2. С. 102–108].

Под звуковой цепочкой имеется в виду последовательность звуков, т. е. слова, которые в совокупности составляют текст. В языкознании отмечается, что в диалектах звуковые цепочки похожи. Более того, они похожи и в генетически родственных языках. Метод поиска генетической близости языков основан на анализе материального сходства, т. е. сходства звуковых цепочек на уровне слов. Эти материальные сходства приводятся в любом учебнике по введению в языкознание (см., например, [16. С. 276–279]). Действительно, в родственных языках слова имеют сходные звуковые цепочки: русский – «мать»; украинский – «мати»; белорусский – «маці»; болгарский – «майка»; сербский – «мати»; чешский – «матер»; польский – «мац». В то же время, в тюркских языках слово «мать» имеет уже другую звуковую цепочку: азербайджанский – «ана»; башкирский – «ана»; казахский «ана»; киргизский – «энэ»; кумандинский – «анам»; татарский – «ана»; тувинский – «ава»; туркменский – «ана»; узбекский – «она».

Современное состояние проблемы «язык-диалект»

В качестве примера нерешенности вопроса о статусе языкового объекта можно привести алтайский язык, пять диалектов которого сильно отличаются друг от друга. Традиционно южно-алтайский (алтай-кижи) и северно-алтайские языки (кумандинский, телеутский и др.) считались одним языком – алтайским. Однако в последнее время специалисты по алтайскому языку все чаще высказывают мнение, что алтай-кижи, кумандинский, телеутский, челканский и тубаларский являются разными языками. Ситуацию усложняет то, что до сих пор нет достаточных оснований считать их ни отдельными языками, ни диалектами одного языка. Наши исследования, в которых применяются информационные технологии, помогут принять решение об их статусе.

До конца не решен вопрос о статусе языка долган, который до сих пор квалифицируется в тюркологической литературе то как отдельный язык, то как диалект якутского языка. Против того, чтобы считать долганский язык диалектом якутского языка, резко выступали известные тюркологи В. М. Наделяев и Е. И. Убрятова. Они неоднократно подчеркивали, что совокупность объективных лингвистических, функциональных и социальных факторов позволяет

Таблица 2

Расстояния между языками и диалектами
по коэффициенту ТМВ

№	Языки и диалекты	Коэффициент расстояния
1	Мордовский: Эрзя – Мокша	1,59
2	Татарский (Казанский) – Татарский (Крымский)	3,16
3	Русский – Украинский	3,90
4	Эстонский – Водский	4,87
5	Русский – Белорусский	5,49
6	Монгольский – Бурятский	5,75
7	Чешский – Словацкий	6,00
8	Алтайский (Телеутский) – Алтайский (Челканский)	6,71
9	Алтайский (Шорский) – Алтайский (Челканский)	7,39
10	Татарский (Казанский) – Татарский (Чулымский)	7,58
11	Алтайский (Алтай-кижи) – Киргизский	8,79
12	Татарский (Казанский) – Татарский (Барабинский)	9,48
13	Татарский (Казанский) – Туркменский	9,48
14	Алтайский (Алтай-кижи) – Алтайский (Чалканский)	10,38
15	Алтайский (Алтай-кижи) – Татарский (Казанский)	11,07
16	Алтайский (Алтай-кижи) – Алтайский (Телеутский)	12,66
17	Татарский (Казанский) – Башкирский	14,18
18	Алтайский (Кумандинский) – Алтайский (Телеутский)	16,29
19	Марийский: Горный – Луговой	17,27
20	Алтайский (Алтай-кижи) – Древнетюркский	18,02
21	Карельский: Тихвинский – Ливвиковский	18,24
22	Алтайский (Алтай-кижи) – Алтайский (Шорский)	18,69
23	Мансийский: Северный – Кондинский	19,58
24	Татарский (Казанский) – Хакасский	22,24
25	Якутский – Долганский	23,33
26	Карельский: Тихвинский – Людиковский	58,77
27	Хантыйский: Казымский – Восточный	98,44

считать язык долган самостоятельным языком, отличным от якутского. Результаты, полученные нами с помощью методов фонологической статистики, подтверждают правомерность данного утверждения (табл. 2). Требуется решения и проблема языкового статуса лингвистических образований, входящих в состав мансийского языка. В данной статье мы также проверяем, насколько близки друг другу северный и кондинский диалекты мансийского языка по типологии встречаемости базовых фонетических признаков.

В научной литературе нет единства взглядов и в отношении трактовки степени самостоятельности мордовских языков эрзя и мокша, имеющих ряд общих фонетических и грамматических признаков. А. С. Чикобава в 1953 г. рассматривал их как диалекты единого мордовского языка [17]. Но спустя несколько лет А. П. Феокистов говорит уже об отдельных эрзя-мордовском и мокша-мордовском языках, которые близки между собой, как русский, украинский и белорусский [18; 19].

Обсуждение результатов

Получив числовые характеристики звуковой картины исследуемых языков и диалектов, мы сможем измерить расстояния между ними на фонетическом уровне [20]. Расстояние измеряется при помощи величины критерия хи-квадрат.

Для построения типологической модели звуковых картин в языках по названным фонетическим критериям необходимо протранскрибировать тексты с учетом инвентарей фонем и особенностей их реализации в речи. Так, в мокшанском и эрзянском языках транскрибирование проводилось в соответствии с общепринятой в этих мордовских языках системой фонем, описанной А. П. Феокистовым [18; 19], в мансийском языке – по выявленной нами фонологической системе с учетом фонетических исследований предыдущих авторов [21].

Следует отметить, что объективное типологическое расстояние между финно-угорскими языками не показывает их генетическое родство. В то же время генетически близкие языки обычно близки типологически [22]. Например, генетически родственные славянские языки – русский, украинский и белорусский – имеют значительную типологическую близость, что подтверждается фоностатистическими данными. Мы исходим из того, что близость русского, украинского и белорусского языка не вызывает сомнения. В связи с этим расстояние, которое получено между этими языками, послужит эталоном их близости (см. табл. 2).

Наши предыдущие исследования вносят в решение проблемы языкового статуса некоторые объективные основания. Результаты анализа фонетических систем трех диалектов карельского языка с точки зрения распределения фонем в их звуковых цепочках выявили существенные различия между тихвинским, ливвиковским и людиковским диалектами [22]. Расстояния были получены на отдельных фонемах и по другой формуле, но результаты в принципе показали большое различие. Фактически на основании полученных нами типологических расстояний их следует считать не диалектами, а отдельными языками.

Мы будем также сравнивать расстояние не только между диалектами, но и между языками, которые имеют близкие номенклатуры фонем. В частности, А. П. Феокистов отметил почти полную идентичность фонемных номенклатур и сходство артикуляционно-акустических баз русского и эрзя-мордовского языков: по его мнению, русские и эрзянские звуки произносятся почти одинаково [19]. Мы проверили, насколько похожи звуковые картины этих языков по статистическим критериям. Расстояние между русским и эрзя-мордовским языками составило 11,89 % (от всего количества фонем). Это говорит о том, что даже похожие фонемные номенклатуры могут производить несхожие звуковые цепочки [23].

Впервые нами были исследованы звуковые картины ряда финно-угорских языков [21], что позволило определить степень близости этих языков к тюркским, монгольским и некоторым другим языкам. Показана частота встречаемости звуков – согласных и гласных, выявленная

по результатам анализа фольклорных и художественных текстов, в процентном отношении ко всем фонемам выборки по каждому из языков и диалектов (см. табл. 1).

Отметим при этом, что объем обработанного материала был достаточным, и это снимает проблему рассмотрения вида распределения и других статистических ограничений [23]. Статистическая обработка звучащей речи, имеющая целью выявление типологии распределения частоты встречаемости некоторых групп согласных и гласных в звуковых цепочках, позволила получить звуковые картины по разным языкам.

На основе данных, представленных в табл. 1, методами математической лингвистики с применением критерия хи-квадрат были определены межъязыковые расстояния (табл. 2). При подсчете использовался введенный нами коэффициент Тамбовцева (ТМВ), означающий величину отношения вычисленного (эмпирического) значения критерия хи-квадрат к его теоретическим критическим значениям с учетом степеней свободы и уровня значимости. Коэффициент ТМВ облегчает работу лингвиста тем, что позволяет оценить статистическую идентичность или несхожесть двух языковых таксонов без расчета степеней свободы и анализа уровня значимости. Заметим, что для квалифицирования лингвистических образований в качестве диалектов одного языка величина коэффициента ТМВ не должна превышать единицу. Получение коэффициента ТМВ в деталях описано в нашей монографии, где подробно рассматриваются некоторые статистические критерии, применяемые в лингвистике [11].

При трактовке полученных данных мы исходили из постулата о том, что если русский и украинский признаются самостоятельными языками (ТМВ = 3,90), то лингвистические объекты, расстояние между которыми по коэффициенту ТМВ превышает этот пороговый показатель, безусловно, должны квалифицироваться как имеющие статус отдельных языков.

Сравнение звуковых картин эрзя-мордовского и мокша-мордовского языков показало наличие расхождений их звуковых цепей по критерию хи-квадрат: величина коэффициента ТМВ больше единицы (ТМВ = 1,59). Из этого можно сделать вывод, что звуковые картины этих языков различны. В то же время сравнение с эталонным в данном контексте ТМВ = 3,90 (расстояние между русским и украинским языками) не дает оснований для трактовки эрзя-мордовского и мокша-мордовского лингвистических образований в качестве самостоятельных диалектов. Звуковые картины мокша-мордовского и эрзя-мордовского языков намного более сходны с типологической точки зрения, чем звуковые картины восточно-славянских языков, которые нами взяты в качестве эталона близости. Из этого можно сделать вывод, что эрзя-мордовский и мокша-мордовский так же, как казанско-татарский и крымско-татарский (ТМВ = 3,16), следует рассматривать в статусе диалектов, а не отдельных языков.

Выводы

1. Важнейшим элементом всякой классификации является выбор характеристик, по значениям которых будет делаться эта классификация. Язык – это объект, который описывается большим количеством характеристик разного иерархического уровня – фонетического, лексического, синтаксического, семантического и прагматического. Классификации языковых объектов на базе характеристик каждого уровня могут не совпадать друг с другом. Общая классификация языков должна строиться по совокупности языковых характеристик всех уровней и отражать обоснованный компромисс их несовпадающих влияний. В данной работе сделана попытка построить классификацию языков и диалектов на базе характеристик только одного иерархического уровня – фонетического. По этой причине полученная классификация является лишь некоторым приближением к полной классификации языков и диалектов. Цель данной работы состояла в отработке методики классифицирования языковых объектов по их звуковым картинам, которая может оказаться полезной и для получения более полной классификации языков. Кроме того, отдельные результаты полученной фонетической классификации могут восприниматься как очевидные, другие – как спорные

или неприемлемые. Это может повысить интерес сообщества лингвистов и специалистов по информатике к проблеме построения общей классификации языков и диалектов.

2. Предлагаемый нами метод исследования типологической близости языков является новым в языкознании. Он основан на использовании информационных технологий и показывает свою пригодность для решения лингвистических проблем, которые невозможно решить традиционными методами языкознания. Он позволяет делать достоверные выводы о типологическом расстоянии между языками на основе их звуковых цепочек, совокупность которых составляет звуковую картину языка. Если русский, украинский, белорусский, эстонский, водский, монгольский, бурятский, чешский и словацкий признаются лингвистами самостоятельными языками, то есть все основания для внесения поправок в классификацию некоторых тюркских и финно-угорских диалектов. Тихвинский, ливвиковский и людиковский диалекты карельского языка следует квалифицировать в статусе отдельных – хотя и близкородственных – языков, но никак не диалектов. Литературный алтайский язык (Алтай-кижи) ближе к киргизскому, чем к своим диалектам. Из этого следует, что диалекты алтайского языка нужно считать отдельными языками. Якутский и долганский, несомненно, являются самостоятельными языками. К сожалению, до сих пор такие разные лингвистические объекты, как северный и кондинский диалекты мансийского языка в лингвистической литературе фиксируются в качестве диалектов, что совершенно нелогично: точнее было бы северный и кондинский диалекты мансийского, а также казымский и восточный диалекты хантыйского также признать самостоятельными языками. Это же касается барабинского и чулымского лингвистических объектов, до сих пор рассматриваемых некоторыми языковедами в статусе диалектов татарского языка.

Список литературы

1. *Загоруйко Н. Г.* Методы распознавания и их применение. М., 1972.
2. *Загоруйко Н. Г.* Прикладные методы анализа данных и знаний. Новосибирск, 1999.
3. *Tambovtsev Yu.* How Can Typological Distances between Latin and Some Indo-European Language Taxa Improve Its Classification? // *The Prague Bulletin of Mathematical Linguistics.* 2007. № 88. P. 73–90.
4. *Аракин В. Д.* Сравнительная типология английского и русского языков. М., 2000. 256 с.
5. *Реформатский А. А.* Введение в языковедение. М., 2005. 536 с.
6. *Эдельман Д. И.* К проблеме «язык или диалект» в условиях отсутствия письменности // *Теоретические основы классификации языков мира.* М., 1980. С. 127–147.
7. *Tambovtsev Yu.* How Should the Hungarian Language Be Correctly Classified within the Ugric Language Taxon? // *Asian and African Studies.* 2009. Vol. 18, № 1. P. 103–137.
8. *Tambovtsev Yu.* The Imitation of Style: Frequency of Occurrence of Linguistic Elements // *California Linguistic Notes.* 2009. Vol. 34, № 2. P. 3–11.
9. *Пиотровский Р. Г., Бектаев К. Б., Пиотровская А. А.* Математическая лингвистика. М., 1977.
10. *Тамбовцев Ю. А.* Измерение фоностатистических расстояний между уральскими языками // *Fenno-Ugristica.* 2003а. № 25. P. 120–168.
11. *Тамбовцев Ю. А.* Типология функционирования фонем в звуковой цепочке индоевропейских, палеоазиатских, урало-алтайских и других языков мира: компактность подгрупп, групп, семей и других языковых таксонов. Новосибирск, 2003б. 143 с.
12. *Tambovtsev Yu.* The Phono-Typological Distances between Ainu and other World Languages as a Clue for Closeness of Languages // *Asian and African Studies.* 2008. Vol. 17, № 1. P. 40–62.
13. *Крамер Г.* Математические методы статистики. М., 1975. 485 с.

14. *Tambovtsev Yu.* Degrees of Dispersion of Various Language Taxa as the Index of Their Typological Similarity // *Favete Linguis. Studies in Honour of Viktor Krupa.* Bratislava: Institute of Oriental Studies of Slovak Academy of Sciences, 2006. P. 219–260.
15. *Тамбовцев Ю. А.* Лингвистическая таксономия: компактность языковых подгрупп, групп и семей // *Baltistica.* 2002. Vol. 32 (1). P. 131–161.
16. *Баранников Л. И.* Введение в языкознание. Саратов: Изд-во Сарат. ун-та, 1973.
17. *Чикобава А. С.* Введение в языкознание: Учеб. пособие. М., 1953.
18. *Феоктистов А. П.* Мордовские языки // *Языки народов СССР.* М., 1966. Т. 3: Финно-угорские и самодийские языки. С. 172–176.
19. *Феоктистов А. П.* Мордовские языки // *Языки мира: уральские языки.* М., 1993.
20. *Tambovtsev Yu. A.* The Value of the Confidence Interval of the Consonant-Vowel Ratio as an Indicator of the Type of Linguistic Material // *Literary and Linguistic Computing.* 1987. Vol. 2, № 2. P. 120–124.
21. *Тамбовцев Ю. А.* Некоторые характеристики распределения фонем мансийского языка // *Советское финно-угроведение.* 1977. Т. 13, № 3. С. 195–198.
22. *Tambovtsev Yu. A.* Phoneme Frequency and Closeness Quotient: Establishing Relationship Degrees by Phonostatistics // *Ural-Altai Yearbook.* 1984. № 56. P. 103–119.
23. *Tambovtsev Yu. A.* The Linguistic Distances among some Languages of Asia // *The Study of Sounds.* 1988. Vol. 22. P. 17–34.

Материал поступил в редколлегию 10.03.2010

Yu. A. Tambovtsev

**PHONO-TYPOLOGICAL CLOSENESS OF LINGUISTIC OBJECTS
BY THE «CHI-SQUARE» CRITERION**

The article investigates the phono-typological closeness of linguistic objects by the Chi-square criterion. Any dialect or language is considered to be a linguistic object. The Chi-square distance is based on the typology of the structure of the speech sound chains in languages and dialects. The total of the speech sound chains comprise a phonetic picture of a language. The typological distances on the phonetic level between the linguistic objects have been built on the basis of threshold value of the TMB coefficient which is the ratio of the computed Chi-square value in a text to the theoretical value of the Chi-square criterion with respect to the degrees of freedom and the level of statistical significance. It allows us to classify the linguistic object either as a dialect or as a separate language from the point of view of their phonetic pictures.

Keywords: phono-typological closeness, linguistic objects, Chi-square criterion, dialect, language, distance, typology.