

**Ю.А. Тамбовцев**

*Новосибирский государственный педагогический университет*

**Распределения групп фонем  
в звуковых цепочках языков и диалектов**

*Аннотация:* В статье исследуется типология строения звуковых цепочек в языках и их диалектах. На основании доминантных фонетических признаков звуковых картин построены типологические расстояния между лингвистическими объектами, т.е. языками и диалектами. С учетом порогового значения коэффициента Тамбовцева (ТМВ) лингвистические объекты, которые взяты для анализа, классифицируются на диалекты и отдельные языки на основании показателей типологических расстояний.

The article investigates the typology of the structure of the speech sound chains in languages and dialects. The typological distances between the linguistic objects have been built on the basis of threshold value of the TMB coefficient. It allows us to classify the linguistic object either as a dialect or as a separate language.

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

Typology, speech sound chains, language, dialect, distances, language taxon.

УДК: 801.5.3.83.11.87.3.1.

Контактная информация: Новосибирск, ул. Виллюйская, 28. НГПУ, ИФМиП. Тел. (383) 2441072. E-mail: yutamb@mail.ru.

В данной работе осуществлен типологический подход, в основе которого лежит расчленение систем лингвистических объектов и их дальнейшая группировка с помощью обобщенной модели на основе выделенных признаков. Этот подход обычно применяется в языкознании. Наиболее последовательным представителем этого подхода 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 типа (например, русский и английский) [Аракин, 2000, 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 на фонетическом уровне всегда подчеркивается при описании общей типологии [см., например: Реформатский, 2005, с. 455–456]. Метод фундаментального типологического сопоставления позволил В.Д. Аракину сравнивать между собой такие разные языки, как славянские, германские, тюркские, малайско-полинезийские и др. с разным генетическим происхождением, разных систем и структур [Аракин, 2000].

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 [Эдельман, 1980]. 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.

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. 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 нами 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 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 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. Чем меньше расстояние, тем более похожи эти объекты. В то же время, чем меньше сумма расстояний, тем более похожи лингвистические объекты в таксоне. Под таксоном мы имеем в виду любую группу языков, даже такую, которая состоит только из двух членов, т.е. любая пара языков или диалектов [Тамбовцев, 2002, с. 131].

В качестве примера нерешенности вопроса о статусе языкового объекта можно привести язык долган, иногда квалифицируемый в тюркологической литературе как диалект якутского языка, против чего выступали известные тюркологи В.М. Надеяев и Е.И. Убрятова. Они неоднократно подчеркивали, что совокупность объективных лингвистических, функциональных и социальных факторов позволяет считать язык долган самостоятельным языком, отличным от якутского. Результаты, полученные нами с помощью методов фонологической статистики, подтверждают правомерность данного положения. 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 [Чикобава, 1953]. Но через несколько лет после этого А.П. Феоктистов говорит уже об отдельных эрзя-мордовском и мокша-мордовском языках, которые близки между собой, как русский, украинский и белорусский [Феоктистов, 1966; 1993].

Получив числовые характеристики звуковой картины исследуемых языков и диалектов, мы сможем измерить расстояния между ними на фонетическом уровне. Расстояние измеряется при помощи величины критерия хи-квадрат, который 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, [Пиотровский и др., 1977]). Кроме того, они даются в компьютерных приложениях в виде таких программ, как «Статистика», «Математика», «XL» и другие. Хи-квадрат критерий уже применялся нами ранее для определения типологической близости языков различных семей [Тамбовцев, 2003а; 2003б].

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 отнести его к подгруппе обско-угорских языков, куда традиционно относятся хантыйский и мансийский язык. Наш метод дает объективные основания для выделения его в отдельную – венгерскую – подгруппу, которая до сих пор не выделялась внутри угорской группы [Тамбовцев, 2009, 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 [Тамбовцев, 2008]. Таким образом, величина «хи-квадрат» показывает несхожесть языков и диалектов. Мы воспользовались методом сравнения двух выборок, которые могут иметь разный объем [Крамер, 1975].

В типологических исследованиях важно определить базовые признаки [Тамбовцев, 2003а], 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 [Тамбовцев, 2003б].

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. Феоктистовым [Феоктистов, 1966; 1993], в мансийском языке – по выявленной нами фонологической системе с учетом фонетических исследований предыдущих авторов [Тамбовцев, 1977].

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 [Тамбовцев, 1984]. Например, генетически родственные славянские языки – русский, украинский и белорусский имеют значительную типологическую близость, что подтверждается фоностатистическими данными.

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 [Тамбовцев, 1984]. 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. Феоктистов 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 [Феоктистов, 1993]. 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 [Tambovtsev, 1988].

Впервые нами были исследованы звуковые картины мансийского и венгерского языков [Тамбовцев, 1977], что позволило рассчитать фоно-типологическое расстояние, т.е. определить степень близости этих языков к некоторым другим финно-угорским языкам.

В таблице 1 показаны величины выборок, а также частота встречаемости звуков – согласных и гласных, выявленная по результатам анализа фольклорных и художественных текстов в % ко всем фонемам выборки по каждому из языков и диалектов.

Выборки фонем по лингвистическим объектам: долганский (ДЛГ) – 35407; якутский (ЯКТ) – 236245; алтай-кижи (АКЖ) – 190525; алтай-чалкан (АЧЛ) – 5147; татарский-казанский (ТКЗ) – 31000; татарский-барабинский (ТБР) – 67569; крымско-татарский (КРТ) – 10362; татарский-чулымский (ТЧЛ) – 5602; эрзя-мордовский (ЭРЗ) – 148164; мокша-мордовский (МКШ) – 143720.

Отметим при этом, что объем обработанного материала был значительным, и это снимает проблему рассмотрения вида распределения и других статистических ограничений [Tambovtsev, 1998]. 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.

Таблица 1

**Частота встречаемости групп фонем в некоторых языках и диалектах**

№	Группы фонем	ЯКТ	ДЛГ	АКЖ	АЧЛ	ТКЗ	ТБР	КРТ	ТЧЛ	ЭРЗ	МКШ
		Частота встречаемости звуков, % ко всем фонемам									
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

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 квалификации 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 [Тамбовцев, 2003б].

Пример расчета близости двух лингвистических объектов – казанского татарского и барабинского татарского языков – по девяти фонетическим признакам приведен в таблице 2.

Таблица 2

№	Группы фонем	Губной	Переднеязычный	Среднеязычный	Заднеязычный	Сонорный	Смычный	Щелевой	Звонкий	Гласный
1.	Казанский татарский	8,03	35,52	3,19	10,92	26,54	23,17	7,95	11,93	42,34
2.	Барабинский татарск.	9,04	34,60	2,13	11,35	25,73	23,54	7,85	7,22	42,88
3.	Хи-квадрат	5,98	1,21	21,12	0,83	1,26	0,29	0,06	115,84	0,34

Сумма «хи-квадрата» – 146,93. Табличное значение «хи-квадрата» при уровне значимости 0,05 и 8 степенях свободы составляет 15,507. ТМВ=9,48.

Таблица 3

### Расстояния между языками и диалектами по коэффициенту ТМВ

№	Языки и диалекты	Расстояние
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.	Татарский (Казанский) – Татарский (Чулымский)	7,58
9.	Татарский (Казанский) – Татарский (Барабинский)	9,48
10.	Татарский (Казанский) – Туркменский	9,48

11.	Алтайский (Алтай-кижи) – Алтайский (Чалканский)	10,35
12.	Татарский (Казанский) – Башкирский	14,18
13.	Марийский: Горный – Луговой	17,27
14.	Карельский: Тихвинский – Ливвиковский	18,24
15.	Мансийский: Северный – Кондинский	19,58
16.	Татарский (Казанский) – Хакасский	22,24
17.	Якутский – Долганский	23,33
18.	Карельский: Тихвинский – Людиковский	58,77
19.	Хантыйский: Казымский – Восточный	98,44

При трактовке полученных данных мы исходили из постулата о том, что, если русский и украинский признаются самостоятельными языками ( $TMB=3,90$ ), то лингвистические объекты, расстояние между которыми по коэффициенту  $TMB$  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 их коэффициента  $TMB$  больше единицы ( $TMB=1,59$ ). Из этого можно сделать вывод, что звуковые картины этих языков различны. В то же время, сравнение с эталонным в данном контексте  $TMB=3,90$  (расстояние между русским и украинским языками), не дает оснований для трактовки эрзя-мордовского и мокша-мордовского лингвистических образований в качестве самостоятельных языков. Звуковые картины мокша-мордовского и эрзя-мордовского языков намного более сходны с типологической точки зрения, чем звуковые картины восточнославянских 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 ( $TMB=3,16$ ), следует рассматривать в статусе диалектов.

*Выводы.* 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 финно-угорских и тюркских диалектов. Тихвинский, ливвиковский и людиковский диалекты карельского языка следует квалифицировать в статусе отдельных – хотя и близкородственных – языков, но никак не диалектов. 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.

## Литература

- Аракин В.Д. Сравнительная типология английского и русского языков. М., 2000.
- Крамер Г. Математические методы статистики. М., 1975.
- Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А. Математическая лингвистика. М., 1977.
- Реформатский А.А. Введение в языковедение. М., 2005.
- Тамбовцев Ю.А. Некоторые характеристики распределения фонем мансийского языка // Советское финно-угроведение XIII. 1977. № 3. С. 195–198.
- Тамбовцев Ю.А. Лингвистическая таксономия: компактность языковых подгрупп, групп и семей // Baltistica. XXXII (1). 2002. P. 131–161.
- Тамбовцев Ю.А. Измерение фоно-статистических расстояний между уральскими языками // Fenno-Ugristica. № 25. (Tartu). 2003a. P. 120–168.
- 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.
- Феоктистов А.П. Мордовские языки // Языки народов СССР. М., 1966. Т. 3: Финно-угорские и самодийские языки. С. 172–176.
- Феоктистов А.П. Мордовские языки // Языки мира: уральские языки. М., 1993.
- Чикобава А.С. Введение в языкознание: Учеб. пособие. М., 1953.
- Эдельман Д.И. К проблеме «язык или диалект» в условиях отсутствия письменности // Теоретические основы классификации языков мира. М., 1980. С. 127–147.
- Tambovtsev Yu.A. Phoneme Frequency and Closeness Quotient: Establishing Relationship Degrees by Phonostatistics // Ural-Altai Yearbook. 1984. N 56. P. 103–119.
- Tambovtsev Yu.A. The Linguistic Distances among some Languages of Asia // The Study of Sounds. Vol. 22. December, 1988. P. 17–34.
- Tambovtsev Yu.A. Some phonological features as measure of closeness of dialects // Philologia Fenno-Ugrica (Freiburg). 1998. N 4. P. 1–19.
- Tambovtsev Yu. The phono-typological distances between Ainu and other world languages as a clue for closeness of languages // Asian and African Studies. Vol. 17. 2008. N 1. P. 40–62.
- Tambovtsev Yu. How should the Hungarian language be correctly classified within the Ugric language taxon? // Asian and African Studies. Vol. 18. 2009. N 1. P. 103–137.