

## ПРОБЛЕМА «ЯЗЫК ИЛИ ДИАЛЕКТ» И ПОПЫТКА ЛЕКСИКОСТАТИСТИЧЕСКОГО ПОДХОДА\*

© 2017

Юрий Борисович Коряков

Институт языкознания РАН, Москва, 125009, Российская Федерация; ybkoryakov@gmail.com

Разграничение языков и диалектов — один из «вечных» вопросов лингвистики. В статье дается краткий обзор двух типов подходов к этой проблеме: социо-этно-политических и структурных. Далее предлагается методика разграничения разных языков и диалектов одного языка, основанная на проценте совпадений по 100-словному списку Сводеша. Предложенная методика рассматривается на большом количестве примеров.

**Ключевые слова:** классификация, лексикостатистика, социолингвистика, таксономия, языки мира

### LANGUAGE VS. DIALECT: A LEXICOSTATISTIC APPROACH

Yuri B. Koryakov

Institute of Linguistics, Russian Academy of Sciences, Moscow,  
125009, Russian Federation; ybkoryakov@gmail.com

Linguists have always seen the notion of “language” as inherently problematic, and the question of whether a particular form of speech should be classified as a separate language or a dialect of a language cannot be answered easily indeed. In this paper, criteria such as shared standard norm, common ethnic identity, and mutual intelligibility are shortly reviewed. Then another method of disentangling languages and dialects based on the number of shared cognates in the 100-word Swadesh list is proposed. The method is tested on various examples.

**Keywords:** classification, languages of the world, lexicostatistics, sociolinguistics, taxonomy

### 1. Введение

Вопрос, является ли некоторая языковая разновидность (идиом) языком или диалектом, относится к одной из сложных проблем лингвистики, причем последствия такого решения могут выходить и далеко за ее пределы.

Лингвисты сами постоянно оперируют понятием языка, но старательно избегают попыток определить, чем он отличается от диалекта. С одной стороны, это оправдано в тех случаях, когда речь идет о «точке» в языковом мире — языке одного человека, одного села, одного народа, одного государства. Действительно в таких случаях не так важно, какой именно термин выбрать, ср. названия работ: «язык Пушкина» [Виноградов 2000], «даргинский говор селения Танты» [Сумбатова, Ландер 2014], «язык бесермян» [Кузнецова 2012].

Однако же, когда мы рассматриваем некоторое цельное множество идиомов, все большее значение приобретает их относительная иерархия как внутри множества, так и по сравнению с другими множествами. Будь то говоры нескольких сел, языки внутри одной генетической группы или языки одного государства.

Так, умение отличать язык от диалекта важно при инвентаризации лингвистических объектов, например, чтобы иметь представление о том, сколько языков в стране, регионе,

\* Исследование выполнено при поддержке гранта РФФ № 17-18-01649.

семье / группе, мире или же сколько языков описано в той или иной работе; для составления справочников [ЯиДМ 1982: 3—4] и баз данных (Ethnologue, Glottolog, Linguasphere [Dalby 2000], Catalogue of Endangered Languages [Okura 2015]), так как обычно базовыми входами в них являются отдельные языки; для присвоения ISO-кодов; при создании языковых карт (так как многие языковые карты, в отличие от текста, недискретны, а иерархии обозначений на карте должна соответствовать иерархия языковых объектов). Также это умение может быть важно для таких прикладных аспектов, как подсчет индексов языковой плотности, например LDI (Linguistic Diversity Index) и ILD (Index of Linguistic Diversity), которые рассчитываются исходя из числа языков; разработка программ по сохранению языков, которые, как правило, ориентированы на сохранение отдельных языков, а не диалектов<sup>1</sup>; выработка литературных норм; создание стройной терминологии, не допускающей ситуаций вида «диалект X диалекта Y диалектной группы Z» или таких названий статей в справочных работах, как «Сави язык / диалект» [Романова и др. 2014].

Обобщая, можно сказать, что понимание того, что является языком, а что диалектом, важно для создания некоторой непротиворечивой системы или реестра всех языковых разновидностей, или языковой систематики<sup>2</sup>. Кроме того, оно важно для специалистов в смежных областях, например, демографов — при проведении переписей, этнографов, — для описания этнических процессов.

## 2. Обзор проблематики

Проблема выработки обоснованной и удобной таксономии уже привлекала внимание лингвистов, см. [Коряков, Майсак 2001]. Некоторые исследователи предлагали отказаться от жесткой дихотомии язык — диалект, заменив ее на трихотомию «внешний язык / внутренний язык / диалект»<sup>3</sup> [Dalby 2000], расширить ее дополнительными единицами [Kaufman 1994] (см. подробнее ниже) или использовать параллельно термины из несколько иной плоскости: «*linguoid*», «*doculect*», «*glossonym*» [Cysouw, Good 2013].

Таким образом, способность отличать язык от диалекта представляется в достаточной степени востребованной. Тут следует оговорить, что в нашем случае мы используем понятия «язык» и «диалект» как находящиеся в иерархическом отношении: диалект должен входить в язык, а не находиться с ним на том же уровне иерархии [Касаткин 1990: 132; Durtell 1990: 499]. Условно говоря, есть совокупность диалектов<sup>4</sup>, и наша задача — разбить их на перепесекающиеся множества, которые мы и называем «языками». Следовательно, мы будем рассматривать язык как множество диалектов, а не как некоторую языковую разновидность, противопоставленную другим языковым разновидностям, называемым диалектами (как, например, в [Hudson 1996: 30—34]). Иными словами, любой диалект входит в какой-нибудь язык, и каждый язык состоит из диалектов (в частном случае из одного диалекта).

Значит, нужно найти некоторые критерии, на основе которых можно провести это разделение. Прежде этого необходимо определить, каким **требованиям** должны отвечать такие критерии, чтобы с их помощью можно было решать поставленные выше практические

<sup>1</sup> Хотя выбор объектов для таких программ часто ориентируется на экстралингвистические обстоятельства (в том числе диктуемые политикой), мнение лингвистов также довольно часто учитывается.

<sup>2</sup> Как будет показано ниже, использование далеко не всех критериев дает непротиворечивый результат.

<sup>3</sup> **Внешний язык** соответствует одному или группе нескольких близкородственных языков в традиционной классификации, которые демонстрируют не менее 71—85 % совпадений в базовом словаре. **Внутреннему языку** соответствует наречие, группа диалектов или отдельный язык с более 86 % совпадений в базовом словаре. **Диалект** соответствует определенной территориальной, социальной или письменной разновидности внутреннего языка [Dalby 2000].

<sup>4</sup> В данном случае под диалектом мы понимаем любую языковую разновидность (или идиом), входящую в состав языка, будь то наречие, диалект (в том числе социальный), говор или даже идиолект.

задачи. К таким требованиям можно отнести универсальность, постоянство во времени и непротиворечивость (однозначность).

Под **универсальностью** имеется в виду то, что рассматриваемый критерий должен быть применим к любой паре идиомов. Это можно считать наиболее важным требованием: если к каким-то идиомам критерий неприменим, то мы не можем ответить на поставленный вопрос, и нам нужен другой критерий.

**Постоянство во времени** подразумевает, что факт отнесения идиома к тому или иному языку не должен меняться слишком быстро. Все живые языки постоянно меняются, и, скажем, два диалекта через 500—1000 лет могут стать разными языками, но если, согласно принятому критерию, в любой момент могут возникнуть условия, заставляющие нас пересмотреть его таксономический статус, то это непрактично.

Под **непротиворечивостью** мы понимаем то, что один идиом может относиться только к одному языку. Не должно быть так, что при использовании одного критерия один и тот же диалект (или близкие диалекты) входит в два разных языка. В противном случае мы, например, не сможем однозначно проводить языковые границы или точно посчитать количество языков на заданной территории.

Теперь рассмотрим **критерии**, используемые для решения вопроса о том, куда относится тот или иной идиом. Все существующие критерии можно разделить на две большие группы: внешние (социо-этно-политические) критерии и внутренние (структурные).

### 3. Социо-этно-политические критерии

Внешние, или социо-этно-политические, критерии апеллируют не к самому языку или его структурным свойствам, а к его восприятию людьми. В отечественной лингвистике внешние критерии, как правило, считаются более важными по сравнению с чисто языковыми [Аванесов 1962: 26; Пиотровский 1973: 38; Мечковская 1996; Алпатов 2000: 14; Мурсорин 2001]. В меньшей степени такая точка зрения распространена в западных работах [Hudson 1996].

Можно выделить следующие внешние критерии:

- 1) общая наддиалектная норма,
- 2) этническая ориентация,
- 3) мнение носителей,
- 4) официальный статус.

Среди других критериев также предлагались: достаточно большое количество носителей и/или большая территория распространения; высокий уровень престижа; использование в качестве средства обучения; (широкое) использование в СМИ, в интернете; разработанность письменной нормы [Hudson 1996: 34—35; Hussain Sandhu, Maldonado 2015: 213—214]; автономность языковой системы [Montes Giraldo 1980: 255]. Однако данные критерии не могут быть использованы уже на основании того, что язык и диалект могут оказаться единицами одного уровня, а не входить один в другой, что противоречит нашей изначальной установке.

Рассмотрим подробнее каждый из перечисленных выше четырех критериев.

Согласно **первому** критерию, два идиома считаются диалектами одного языка, если у них есть общий письменный стандарт (литературная норма, нем. *Dachsprache* ‘язык-крыша’), независимый от других языков [Hudson 1996; Brown 1998]; либо, для бесписьменных языков, общее наддиалектное койне. Подробно разные ситуации взаимоотношения письменных и бесписьменных форм языка были разобраны С. Е. Яхонтовым [1980: 154—157].

Уточнение о независимости от других языков необходимо в силу того, что существует немало так называемых литературных диалектов, которые, как правило, не позиционируются как отдельные языки и носят вспомогательный характер по отношению к основному литературному языку, тем не менее обладают письменным стандартом отличным

от общезыкового, ср. восточнословацкий литературный язык и банатско-болгарский книжный язык [Дуличенко 2005а: 608—609] или региональные нормы английского, испанского или французского языков.

Хотя данный критерий можно теоретически применять к бесписьменным языкам, которые пользуются литературной нормой соседнего языка, он практически неприменим к регионам, где письменности совсем отсутствуют или отсутствовали до недавнего времени. Поскольку в таких регионах (вся Америка, Австралия, Новая Гвинея и Океания, большая часть Африки) еще недавно было распространено большинство языков мира, данный критерий нельзя назвать универсальным.

Другой сложностью является то, что общая письменная традиция часто объединяет довольно далеко разошедшиеся идиомы (например, итальянские, немецкие, арабские, китайские «диалекты») или даже слабо родственные языки (например, латынь в Средние века для романских, германских, кельтских языков; классический арабский для многих не-арабских языков Ближнего Востока). Кроме того, у довольно близких диалектов могут оказаться разные литературные нормы. Например, западнополесские говоры Белоруссии ориентируются на белорусский литературный язык, а Украины — на украинский, хотя структурно представляют собой один и тот же тип [Жовтобрюх, Молдован 2005: 543; Бирилло и др. 2005: 588; Коряков 2005: 649]; для носителей нижнесаксонских диалектов в Нидерландах литературным языком является нидерландский, а для носителей тех же диалектов в Германии — немецкий литературный язык или региональные письменные варианты [Lenz et al. 2009]. Таким образом, не соблюдаются принцип непротиворечивости.

Наконец, используя этот критерий, мы так или иначе ориентируемся на степень близости между идиомами. Скажем, едва ли на основании того, что баски, бретонцы или эльзасцы пользуются (или пользовались) литературным французским, кто-нибудь будет называть их языки диалектами французского. Или, например, официально все население Греции кроме официально признанных национальных меньшинств (мусульмане (турки, помаки и часть цыган), армяне, евреи) и недавних иммигрантов считается греками, а их языки, соответственно, диалектами греческого, и литературный язык у всех общий — греческий [Сlogg 2002]. Таким образом, славяне Северной Греции считаются греками-славофонами, а их речь официально называется *σλαβικά διάλεκτος* ‘славянский диалект’ [Усикова 2005: 103]. При использовании этого критерия необходимо сперва определить, где проходит граница между языком и диалектом.

Таким образом, для данного критерия, очевидно, не соблюдается требование постоянства во времени: как показывает практика, создание письменного стандарта — дело нескольких лет, максимум десятилетий. Например, в СССР в 1920—1930-е гг. было создано около 70 письменных норм для бесписьменных до того языков [Алпатов 2000: 74]. И тогда, следуя данному критерию, все эти 70 языков буквально за одно десятилетие должны были изменить свой статус с диалектов на языки.

**Второй** критерий ориентирован на этническую принадлежность носителей: если носители двух идиом относятся к одному этносу, значит это один язык, если к разным — разные [Аврорин 1975: 55; Brown 1998]. Однако языковые и этнические границы слишком часто не совпадают, чтобы определять первые на основе вторых. Например, в Швейцарии четыре традиционных языка (немецкий, французский, итальянский и романшский), хотя в этническом плане их носители считают себя прежде всего швейцарцами [Duttell 1990: 497]; население итальянской провинции Валле-д’Аоста с точки зрения этнических и культурных связей составляют преимущественно итальянцы, хотя традиционные диалекты данного региона относятся к франко-провансальскому (арпитанскому) языку, при том что региональным языком признан литературный французский [Viatta 1982: 302]. Вообще взаимно однозначное соответствие между этносом и языком встречается довольно редко, оно характерно либо для родоплеменного строя (племя ↔ язык, скажем, в Амазонии или Новой Гвинее), либо для европейского национального размежевания и советского национального строительства 1920-х гг., когда для большинства языков Советского Союза были определены соответствующие им народы и наоборот [Алпатов 2000: 47—48]. В некоторых

случаях для этого пришлось либо объединить несколько этнических групп (например, хакасы [G. Anderson 1998: 1], узбеки [Кармышева 1969]), либо объединить под одной «шапкой» несколько идиомов (например, «даргинский язык» [Коряков 2013: 193]). Тем не менее даже на постсоветском пространстве остаются случаи несоответствия как в одну сторону (носители мокшанского и эрзянского языков обычно считают себя мордвой; носители 13 андо-цезских народностей, как правило, считают себя аварцами), так и в другую (на карачаево-балкарском говорят карачаевцы и балкарцы; на кабардино-черкесском — кабардинцы и черкесы) [Comrie 1981: 4—5]. Наконец, в некоторых регионах понятие этнической принадлежности не имеет прямой связи с языком, основываясь большей частью либо на расовых категориях (Латинская Америка<sup>5</sup>), либо на социально-конфессиональной принадлежности (Ближний Восток [Огувал 1990]), иногда с учетом кастового положения (Южная Азия<sup>6</sup>).

Хотя в целом этническое самосознание достаточно устойчиво, при резком изменении внешних факторов или целенаправленном воздействии на сознание людей оно может меняться достаточно быстро. Ф. Д. Климчук приводит пример двукратной смены этнической аффилиации в Закарпатье в течение нескольких десятилетий: «Вообще-то мы русские [в данном случае, видимо, один из вариантов самоназвания «русин». — Ю. К.], теперь мы украинцы (с 1945 г. Закарпатье в составе УССР), а до войны мы были чехами (область находилась в составе Чехословакии)» [Климчук 1990: 97], цит. по [Беликов, Крысин 2001: 49]. Другим примером является распространение этнонима «черногорцы» в переписях Черногории 2003 и 2011 гг., см. подробнее ниже.

Встречаются также случаи сложной этноязыковой идентификации: так, кубанские греки-понтийцы признают, что их язык слишком далек от греческого, чтобы утверждать, что они говорят по-гречески, но назвать свою речь отдельным языком они также не решаются [Перехвальская 1997: 7—8]. Цалкинские туркоязычные греки, говорящие фактически на одном из восточных турецких диалектов, отказываются соотносить свой диалект с турецким языком, именуя его *bizim dil* ‘наш язык’, при этом часто называя своим родным греческий, которым они не владеют [Елоева 1995: 4—5].

Все эти примеры говорят о том, что определение народа само является достаточно комплексным и не универсальным, и, скорее, одним из параметров для выделения этнической группы является язык, а не наоборот. С точки зрения соответствия выдвинутым нами требованиям, этот критерий далеко не универсален, результаты его применения вступают в противоречие друг с другом, а относительное постоянство во времени зависит от многих факторов.

Согласно **третьему** критерию основополагающим является мнение носителей: «если коллектив говорящих считает родную речь отдельным языком, отличным от языков всех соседей, значит то, на чем данный коллектив говорит, — это отдельный самостоятельный язык» [Мечковская 1996: 94—95] (ср. также [Пиотровский 1973: 38; Касаткин 1990: 132—133]).

Однако знание о «собственно языке» и его границах с другими языками на самом деле не является частью наивной картины мира. Для «наивной» картины мира, как правило, достаточно противопоставления «наш» язык — «чужой», что до сих пор можно увидеть в самоназваниях многих языков, которые переводятся как «наш / обычный / простой язык». Например, *tutэйишш / no-просту* — западнополесские говоры (польско-белорусско-украинское Полесье), *našinski* — диалект горан (торлакское наречие, Косово), *naš jezik* — диалект моллизских хорват (Италия), *qatō:k* ‘наш язык’ — язык мочо (Мексика, майяские языки), *lishán didán* ‘тж.’ — урмийский еврейско-арамейский язык (Иран), *ombeayiiüts* ‘тж.’ — язык уаве (Мексика, изолят), *nou ling* ‘тж.’ — индо-португальский креол (Индия), *nyeshan* ‘тж.’ — язык

<sup>5</sup> Ср. списки этнических групп по данным переписей в странах Южной Америки в «Demographic Yearbook (Population Censuses’ Datasets (1995 — Present))», выпускаемом ООН: <http://data.un.org/Data.aspx?d=POP&f=tableCode:26>.

<sup>6</sup> Ср., например, список языков и список этнических групп в непальской переписи 2011 г.: <https://unstats.un.org/unsd/demographic/sources/census/wphc/Nepal/Nepal-Census-2011-Vol1.pdf> (National Population and Housing Census 2011).

мананг (Непал, тибето-бирманские языки), *meänkieli* ‘тж.’ — финский диалект меанкиели (Швеция) и т. д. [Беликов, Крысин 2001: 49].

В большинстве же случаев носители повторяют то, что выучили в школе, прочитали в книге, слышали по радио и т. д. Вообще мнение о собственном языке легко формируется государством и СМИ. Например, до 1991 г. все население Молдавии употребляло название «молдавский язык» (*лимба молдовеняскэ*), затем, параллельно с переходом на латиницу, активно стал вводиться лингвоним «румынский язык», однако он так до конца не прижился, более того, позднее начался частичный обратный процесс. По результатам переписи 2004 г. 78,4 % молдаван назвали своим родным языком молдавский, и лишь 18,8 % — румынский [Гульский 2006]. В качестве другого примера можно привести черногорский язык, само название которого появилось лишь в 1995 г. (до этого языком большинства населения был сербохорватский, а еще ранее — сербский) [Кречмер, Невекловский 2005: 145]. Благодаря развернутой пропаганде, по данным переписи 2003 г., 22 % населения Черногории назвало своим родным языком черногорский, а согласно переписи 2011 г. — уже 37%. 63 % и 43 %, соответственно, по-прежнему считали, что говорят по-сербски. Процент считающих себя черногорцами при этом каждый раз отличается в большую сторону от процента назвавших свой язык черногорским: 43 % и 45 % населения назвали себя черногорцами в 2003 и 2011 гг. соответственно<sup>7</sup>.

Кроме того, мнение одних носителей языка может противоречить мнению других (нарушая тем самым требование непротиворечивости). Например, луры (Иран) утверждают, что они говорят по-персидски, персы же отрицают это и говорят, что они не понимают языка луров [С. Voegelin, F. Voegelin 1977: 4]. Наконец, далеко не всегда ясно, что считать одним и тем же названием, то есть насколько велики должны быть различия, чтобы считать (само)названия разными. Так, словенцы называют свой язык — *slovenski jezik*, *slovenščina*, словаки — *slovenčina*, *slovenský jazyk*, словинцы (близкие кашубам) — *slowińskô mòwa*, а самоназвание старославянского — *словѣньскъ ѡязыкъ*. Названия практически идентичны, и, согласно критерию общего самоназвания, все эти идиомы следовало бы считать одним языком. В то же время носители разных диалектов, например словенского языка, могут называть свой язык весьма отличными лингвонимами: резьянцы — *rozajanski langač*, прекмурцы — *prekmörščina*.

Таким образом, как пишет Л. Э. Калнынь [1976: 38], «дифференция “язык — диалект” лингвистически не искушенному индивидууму не дана». Или, как говорят некоторые информанты: «Ты же лингвист, ты и скажи нам, диалект это или язык». С точки зрения соответствия нашим требованиям этот критерий относительно универсален, однако результаты его применения вступают в противоречие друг с другом и могут быстро меняться во времени.

Согласно **четвертому** критерию языком считается тот идиом, статус которого прописан в соответствующем официальном документе, например в конституции, законе о языке или других официальных актах некоторого государства. Политики не так часто выносят решения относительно языков или диалектов, однако в отдельных случаях политический фактор может существенно влиять на мнение как самих носителей, так и лингвистов.

Так, законы могут влиять на названия языков (обычно официальных или государственных) и тем самым на самоидентификацию носителей. Среди примеров можно назвать отдельные языковые варианты в рамках сербохорватского языкового континуума (хорватский, боснийский, сербский, черногорский) [Кречмер, Невекловский 2005], молдавский / румынский язык (см. выше), малайский язык (в Индонезии официально именуется «индонезийским», тогда как в Малайзии используются два названия: «малайский» и «малайзийский») [Алиева 2011: 625]).

Кроме того, в некоторых странах принимаются особые законы, касающиеся малых языков, например Европейская хартия региональных языков или языков меньшинств<sup>8</sup>,

<sup>7</sup> Данные переписи 2003 г. доступны по адресу: <http://www.monstat.org/userfiles/file/popis03/saopstenje60.pdf>; о переписи 2011 г.: [http://www.monstat.org/userfiles/file/popis2011/saopstenje/saopstenje\(1\).pdf](http://www.monstat.org/userfiles/file/popis2011/saopstenje/saopstenje(1).pdf).

<sup>8</sup> Текст Хартии доступен по адресу: <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168007bf44>.

ратифицированная уже в 25 государствах Европы. Согласно этой хартии, защите подлежат все языки, заметно отличающиеся от основного языка, традиционно распространенные на соответствующей территории и не имеющие статуса государственного. В то же время языковая политика некоторых стран Европы далека от такого языкового плюрализма: так, например, Франция и Грузия отказываются ее ратифицировать, а Греция, Литва, Латвия и Эстония — даже подписывать<sup>9</sup>. При этом иногда случаются некоторые несостыковки, например, русинский признан отдельным языком наряду с украинским в Польше, Румынии и Словакии, но лишен такого статуса на самой Украине<sup>10</sup>.

Языковая политика в области признания языков может быть также опосредована, например, через задание списка языков при проведении переписи или выбор языков, которые преподаются в школе или на которых ведется радиовещание.

В этой связи уместно упомянуть известное выражение «**Язык — это диалект с армией и флотом**» — шутивную фразу, популяризованную лингвистом Максом Вайнрайхом [Weinreich 1944]. В буквальном понимании она имеет мало смысла. Есть много государств с довольно сильными армиями (США, Бразилия, арабские страны), диалекты которых от этого ничуть не становятся языками, и множество малых языков, не могущих похвастаться ни вооруженными силами, ни флотом. Единственное реальное применение она получает при возникновении нового государства, одной из программных линий которого может быть стремление к языковому отделению. Ярким примером такой политики являются государства, возникшие на месте бывшей Югославии, каждое из которых провозгласило свой собственный язык вместо бывшего общего сербохорватского ([Багдасаров 1994: 208]; кроме Словении, язык которой рассматривался отдельно еще со Средних веков [Дуличенко 2005б: 200—202], и Македонии, язык которой был провозглашен отдельным от сербского в 1945 г. [Усикова 2005: 104]). Что интересно, единый сербохорватский язык был похожим образом провозглашен по идеологическим мотивам в 1850 г. на волне панславянских идей [Кречмер, Невекловский 2005]. В то же время возможен и обратный сценарий: после провозглашения независимости отдельный до того язык официально объявляется идентичным языку соседнего государства (румынский / молдавский язык в Молдавии) [Лухт, Нарумов 2001: 576].

Следовательно, критерий официального статуса не универсален, не постоянен во времени и относительно противоречив.

Ниже в таблице 1 отмечено, насколько предложенные выше требования выполняются каждым из внешних критериев.

Таблица 1

## Выполнение внешними критериями предложенных требований

Требования \ Критерии	Литературная норма	Этническая общность	Мнение носителей	Официальный статус
Универсальность	–	–	+	–
Постоянство во времени	–	–	–	–
Непротиворечивость	–	–	–	–

По таблице 1 видно, что ни для одного из внешних критериев не выполняются все три требования. Можно отметить, что на практике используется, скорее, некоторое мозаичное

<sup>9</sup> Ср. таблицу подписей и ратификации хартии: <http://www.coe.int/ru/web/conventions/full-list/-/conventions/treaty/148/signatures>; о положении дел во Франции и Грузии см. также: <http://www.conseil-etat.fr/Decisions-Avis-Publications/Avis/Selection-des-avis-faisant-l-objet-d-une-communication-particuliere/Ratification-de-la-Charte-europeenne-des-langues-regionales-ou-minoritaires> и [http://www.ecmcaucasus.org/upload/publications/working\\_paper\\_42\\_rus.pdf](http://www.ecmcaucasus.org/upload/publications/working_paper_42_rus.pdf).

<sup>10</sup> См. подробнее: <http://www.coe.int/ru/web/conventions/search-on-treaties/-/conventions/treaty/148/declarations>.

переплетение разных критериев. В одних случаях лингвисты ориентируются на один, в иных на другой. Однако иногда среди лингвистов имеется консенсус относительно того, что некоторый идиом является отдельным языком, даже если ни один из внешних критериев не работает.

Таблица 2

#### Примеры применения внешних критериев

Язык \ Критерии	Литературная норма	Этническая общность	Мнение носителей	Официальный статус	Мнение лингвистов
Мегрельский	–	–	–	–	+ <sup>11</sup>
Андийские	–	–	–	–	+ <sup>12</sup>
Рюкюские	–	–	–	–	+ <sup>13</sup>

Во всех этих случаях согласно всем внешним критериям данные идиомы должны считаться диалектами, соответственно, грузинского, аварского и японского языков, однако, по мнению большинства лингвистов, перед нами отдельные языки (мегрельский) или даже группы языков (андийские и рюкюские языки). Очевидно, что роль здесь играют только лингвистические различия.

## 4. Структурные критерии

Перейдем теперь к лингвистическим (структурным) или внутренним критериям. Такие критерии должны опираться на фактические различия между идиомами и быть независимы от мнения людей и других внешних факторов.

При выборе и использовании структурных критериев требуется определить признаки сравнения, а затем пороговый уровень различий, необходимых для выделения отдельного языка. Собственно, отсутствие такого универсально принятого порога является одним из главных возражений против использования лингвистических критериев [Глисон 1959: 436—440].

Можно выделить следующие подходы к структурному определению степени близости идиомов:

- 1) измерение взаимопонятности;
- 2) измерение структурных различий.

### 4.1. Взаимопонятность

Пожалуй, наиболее популярный лингвистический критерий — это степень **взаимопонятности** (mutual intelligibility) [Petyt 1980: 11—36; Comrie 1981: 6; Мусорин 2001; Беликов, Крысин 2001: 48—53; Kurpaska 2010; Mané 2012], иногда называемой общепонятностью [Глисон 1959: 436]. Согласно ему, если носители разных идиомов достаточно хорошо понимают друг друга (говоря каждый на своем идиоме), они говорят на диалектах одного языка, если же нет — на разных языках. На практике применение этого метода осложняется некоторыми другими факторами: предварительное знакомство с языком собеседника, тема общения, языковые способности информанта, само желание или нежелание понимать собеседника [Sallabank 2011: 496; Schüppert, Gooskens 2010], поэтому этот структурный

<sup>11</sup> Ср., например, [Hewitt 2004: 19—21; Ростовцев-Попель 2012; Hammarström et al. 2017].

<sup>12</sup> Ср., например, [Hewitt 2004: 24—26; Майсак 2005; Hammarström et al. 2017].

<sup>13</sup> Ср., например, [Shimoji 2010; Hammarström et al. 2017]; традиция описывать рюкюские как диалекты японского языка сохранялась до недавнего времени в отечественной лингвистике, ср. [Алпатов 1997: 325].

критерий становится социально обусловленным [Эдельман 1980: 129—130]. Кроме того, нередки ситуации «полукоммуникации», когда каждый из собеседников говорит на родном языке, но при этом неплохо (хотя и не свободно) воспринимает язык собеседника [Haugen 1966: 926]. Хотя нередко утверждается, что взаимопонятность не обязательно является симметричной (то есть носитель идиома А может понимать идиом Б, но не наоборот), как пишет Х. Хаммарстрём, практически все известные примеры такого рода обусловлены разным уровнем мотивации [S. Anderson 2004] или первоначальным знакомством носителя идиома А с идиомом Б [Hammarström 2008: 35]. При экспериментальных измерениях степени взаимопонятности идиомов А и Б называется среднее степеней взаимопонятности идиома А носителем идиома Б и идиома Б носителем идиома А [Tang, van Heuven 2009].

Другое возражение против использования этого критерия — ситуации диалектных континуумов, когда в цепочке идиомов степень взаимопонятности уменьшается с расстоянием между идиомами и может отсутствовать между крайними точками [Hudson 1996: 35]. Довольно часто в качестве примера приводится нидерландско-немецкое языковое пространство, двигаясь в рамках которого из Амстердама в Вену можно не встретить ни одной резкой языковой границы [S. Anderson 2004; Comrie 2009: 2—3]. Одно из возможных решений этой проблемы рассматривается Х. Хаммарстрёмом, который предлагает считать верным утверждение «если две разновидности относятся к одному языку, то они взаимно понятны», но вместо обратного утверждения «если две разновидности взаимно понятны, они обязательно относятся к одному языку» предлагает требование о том, что число языков не должно увеличиваться без необходимости. При таком решении диалектный континуум делится на некоторое минимальное число языков, внутри которых все разновидности взаимопонятны, и также взаимопонятны некоторые разновидности, отнесенные к разным языкам [Hammarström 2008].

Однако основная сложность в применении этого критерия заключается в том, что взаимопонятность — это не бинарный, а континуальный признак, для использования которого необходимо некоторое пороговое значение [Hudson 1996; Yule 2010]. Частичная попытка установить такие пороговые значения была предпринята С. Е. Яхонтовым [1980: 151—153], однако он определял уровень взаимопонятности описательно, не прибегая к числовым показателям. Хотя чаще всего взаимопонятность между разными идиомами определяется «на глазок», разрабатывается все больше экспериментальных способов ее измерения [Hickerson et al. 1952; Pierce 1952; Casad 1974; Cheng 1997; Brye, Brye 2002; H. Anderson 2005], которые требуют значительных усилий и вместе с тем, с одной стороны, не исключают до конца субъективности суждений (прогресс в этом направлении связан с использованием так называемого «функционального тестирования», которое стремится к измерению именно того, как испытуемый понимает другой идиом, а не его субъективного мнения [C. Voegelin, Harris 1951; van Bezooijen, van den Berg 1999; Tang, van Heuven 2009; Swarte, Hilton 2013; Castro, Xingwen 2015: 157—179], однако этот способ еще более ресурсоемок), а с другой, их результаты плохо сравнимы между собой. Наконец, пока отсутствуют обоснованные формальные способы нахождения универсального порога взаимопонятности для различения языка и диалекта, хотя для одного из типов тестов (Recorded Text Test, RTT) были предложены следующие ориентиры: идиомы, взаимопонятность которых выше 85 %, разумно считать принадлежащими к одному языку и развивать для них единую литературную норму; при уровне взаимопонятности между 70 % и 84 % следует учитывать другие лингвистические и экстралингвистические факторы для более точного определения их принадлежности [Casad 1974; Starr, Stalder 1990].

Таким образом, как суммирует Ш. Госкенс [Gooskens 2007], взаимопонятность зависит от трех факторов:

- 1) отношения слушателя к языку собеседника,
- 2) предварительные контакты или знакомство с языком собеседника,
- 3) лингвистическое расстояние до языка собеседника.

Если более подробно рассматривать только третий фактор, то на взаимопонимание между близкородственными языками и диалектами наибольшее влияние оказывают фонетические различия и лексика [Delsing, Lundin-Åkesson 2005; Gooskens 2007].

Таким образом, взаимопонятность может быть хорошим критерием, если научиться измерять ее независимо от отношения носителей к другому идиому (убрать фактор 1), хотя влияние фактора 2 при этом все равно сохранится. Другим подходом может быть использование некоторого структурного параметра, хорошо коррелирующего с объективной степенью взаимопонятности и при этом не отягощенного фактором 2.

Помимо технического удобства использование степени взаимопонятности для различения языков обеспечивает большую эффективность при языковом планировании, так как успешная коммуникация жизненно важна при оказании медицинской помощи, в образовании и даже в локальной политике [Roche, Suzuki (in print): 9].

## 4.2. Измерение структурных различий

Можно предположить, что степени взаимопонятности соответствует степень структурных различий между идиомами. Однако возникает вопрос: как измерить эти различия? Языки различаются по очень многим параметрам и для разных языков эти параметры разные, поэтому если попытаться сравнить пошагово всю грамматическую систему, начиная от фонологии и заканчивая синтаксисом и лексикой, то результаты сравнения между разными парами идиомов, особенно достаточно далеких друг от друга, будут сами по себе плохо сравнимы. Чтобы получить объективные и сравнимые результаты, необходимо ограничиться некоторым конкретным подмножеством языковых различий. В разных работах для измерения лингвистического расстояния (*linguistic distance*, *linguistic difference*) предлагались разные явления или характеристики для сравнения: фонетический облик слов, звуковая структура словаря, и предложений [Gooskens, Heeringa 2004; Heggarty et al. 2005; Ellison, Kirby 2006]; синтаксические структуры [Homola, Kuboň 2006]; определенные грамматические явления [Wichmann, Holman 2009; Hammarström, O'Connor 2013; McGregor 2013]; даже различия в графике [Cysouw, Good 2013]. Однако ни один из предложенных способов не предлагалось использовать для разграничения языков и диалектов и ни один из них не является универсальным, то есть едва ли возможно найти такое явление или характеристику, применимую к любому языку, с помощью которой можно осмысленно считать лингвистическое расстояние. Кроме того, недавно было отмечено, что многие грамматические явления склонны к распространению через языковые границы (см. [Aikhenvald 2007] и мн. др.) и, следовательно, часто могут свидетельствовать, скорее, о типологической близости или об ареальном сближении языков. В таком случае более универсальным материалом для сравнения является базисная лексика, но сравнение целых словарей трудозатратно и также не универсально. Поэтому в разных работах делались попытки ограничить набор сравниваемых лексических данных: сравнивались эквивалентные переводы, автоматически извлеченные из параллельных корпусов [Kondrak 2013; Wälchli, von Waldenfels 2013] или из двуязычных словарей [Sejane, Eger 2013]; набор диалектной лексики для выявления места региолекта в языковой системе [Nerbonne et al. 2013].

Видимо, наиболее популярным материалом для сравнения является некоторое подмножество (список) базовой лексики [Borin 2013: 7], см. работы [Глисон 1959: 437; Carling et al. 2013; Snoek 2013; Heeringa et al. 2013], особенно список Сводеша или его модификации [Schepens et al. 2013; Saxena, Borin 2013; Blust 2013: 34]. Как известно, М. Сводеш предложил список слов, точнее понятий, предположительно базовых для любой культуры и потому пригодных для универсального сравнения, что крайне важно для наших целей [Swadesh 1952; 1955]. Кроме того, использование базовой лексики в противовес культурной лексике, фонетике или синтаксису дает возможность оценить исконное родство языковых разновидностей, а не внешнее, вызванное поздними языковыми контактами.

## 5. Использование лексикостатистики

Хотя списки базовых слов, особенно список Сводеша, широко используются при сравнении языков для различных целей, в первую очередь для построения языковой классификации и вычисления лингвистического расстояния между языками (собственно лексикостатистика), можно ли использовать такие списки для сравнения близкородственных идиомов с целью различения языков и диалектов? Следует сказать, что идея использования лексикостатистики для этой цели не нова. Еще И. Дайен предложил использовать в качестве порога (он называет его «Provisional language limit») 70 % для 200-слового списка Сводеша или 77 % для 100-слового списка [Dyen 1965: 18; Dyen, Aberle 1974: 11]. Для папуасских языков С. Вурм использовал интервал 78—81 % по слегка измененному 200-словному списку Сводеша как пороговое значение для различения языков и диалектов [Wurm 1977: 152]. В пределах 3-процентного «зазора» (а в исключительных случаях — и вне его) исследователь при выборе в пользу языка или диалекта должен был основываться на том, какие именно лексические единицы оказываются «родственными». Т. Кауфман в своей классификации мезоамериканских языков [Kaufman 1994] предлагает более сложную схему: выше 84 % (по 100-словному списку) — однозначно диалекты одного языка, 76—84 % — диалекты или так называемые зарождающиеся языки (emergent languages), менее 76 % — однозначно отдельные языки (хотя иногда без четких границ между ними, в таком случае они называются virtual languages). Однако ни один из упомянутых авторов не приводит сколько-нибудь серьезного обоснования своим цифрам, довольствуясь лишь общими соображениями и некоторым количеством примеров.

Попробуем точнее оценить, насколько этот метод можно использовать для определения границы между языком и диалектом. Предположим, что есть такой **порог**, что если процент совпадений (ПС) по списку базовой лексики между двумя идиомами **ниже** его, то их можно считать двумя разными языками, если **выше** — диалектами одного языка. Для обнаружения и верификации такого порога необходимо собрать большое количество соответствующих списков и посчитать проценты совпадений между достаточно близкими языками, а затем сравнить полученные результаты с экспертной оценкой взаимопонятности между оцениваемыми идиомами и с тем, как обычно трактуются данные идиомы в лингвистической литературе.

В нашем исследовании будет использоваться 100-словный список Сводеша — Старостина [С. Старостин 2007: 784], так как смысловые концепты этого списка более универсальны, чем у 200-слового [Swadesh 1955], и лишь для него обоснованы и разработаны строгие семантические спецификации [G. Starostin 2010; Kassian et al. 2010], позволяющие получить достаточно точные сравниваемые данные для разных языков. Согласно предложенной С. А. Старостиным процедуре, из подсчета удаляются заимствования, дается экспертная оценка родственности слов и затем подсчитывается процент совпадений между списками двух идиомов. Что касается синонимов, то сравнение результатов обработки списков, составленных по более строгим критериям (обычно не более пяти синонимов на весь список, ср. [G. Starostin 2010: 100]), и списков с большим количеством синонимов говорит в пользу минимизации количества синонимов (М. Сводеш вовсе не допускал синонимов), для чего могут быть эффективно использованы уточняющие контексты [Kassian et al. 2010]. Хотя неучитывание заимствований вполне обосновано для целей общей компаративистики и построения генеалогических классификаций [G. Starostin 2013: 133—136], вопрос их учета при синхронном сравнении близкородственных языков и диалектов остается открытым и требует дальнейшего изучения.

Как уже отмечалось, при лексикостатистических исследованиях главную роль играют не алгоритмы, а качество исходных данных [Касьян 2014: 78]. Поэтому для проверки нашей гипотезы особое внимание было обращено на качество сбора и обработки лексических списков. Хотя для получения списков слов и процентов совпадения между ними использовались несколько источников, главным и наиболее надежным была Глобальная лексикостатистическая база данных [GLD]. На март 2017 г. она включала списки по 483 языкам и диалектам из 100

групп (это число продолжает увеличиваться). Помимо этой базы данных имеются достаточно полные источники по славянским [Kushniarevich et al. 2015], даргинским [Koryakov 2013], уральским [Zhivlov 2014], манде [Vydin 2009] и другим языкам. Хотя, как в данной базе, так и в других источниках, существует много списков по языкам всех частей света, мы старались ограничиться языками наиболее изученных регионов (Европа, Северная Евразия, Ближний Восток), относительно которых у лингвистов (во всяком случае российских) есть некоторые представления о взаимопонятности и о том, что можно назвать отдельным языком с лингвистической точки зрения, а что нет. В общей сложности было взято около 220 пар близкородственных идиомов (с ПС от 80 до 100), в которых участвовало около 340 языков и диалектов.

Таблица 3

Распределение использованных языков по семьям / регионам<sup>14</sup>

Регионы и семьи	Языки	Пары
Евразия	264	181
индоевропейские	110	89
кавказские	68	34
уральские	57	33
Африка	47	23
Америка	28	16
<b>Всего</b>	<b>339</b>	<b>220</b>

Распределим выбранные пары по пятибалльной шкале взаимопонятности. Описания каждого из баллов даны во втором столбце таблицы 4. В третьем столбце даны определения соответствующих степеней родства по шкале, предложенной Яхонтовым [1980: 154—157].

Таблица 4

Пятибалльная шкала взаимопонятности

Баллы	Описание	Яхонтов
1	Полностью взаимопонятны; могут отмечаться особенности каждого диалекта, обычно в произношении (акцент) и употреблении некоторых слов.	1. «Наименьшие языковые различия, заметные для говорящих: носители разных идиом свободно общаются друг с другом, но по особенностям произношения и отчасти лексики могут приблизительно определить, откуда каждый из них родом».
2	Взаимопонимаемы, но не полностью; встречается довольно много особенностей у каждого идиома; могут возникнуть недопонимания или ошибки в понимании.	2. «Носители разных идиом без особого труда общаются между собой, хотя возможны отдельные случаи непонимания».
3	Частично взаимопонимаемы или данных нет.	—
4	Взаимопонимание сильно затруднено, но заметно улучшается после некоторого опыта общения.	3. «Носители разных идиом не могут свободно общаться, но постоянно слышат в речи друг друга знакомые слова и даже короткие фразы. Говорящий на одном языке может научиться понимать другой, “постепенно привыкая” к нему, без учебника или переводчика».
5	Практически невзаимопонимаемы.	4. «Общение невозможно, но при систематическом изучении языков мы обнаруживаем множество общих слов и правил грамматики; это может заметить даже и не языковед».

<sup>14</sup> Подробный список использованных языков и пар с процентами совпадений и присвоенными статусами, а также ссылками доступен по адресу: [http://lingvarium.org/koryakov/Lg\\_vs\\_dt.shtml](http://lingvarium.org/koryakov/Lg_vs_dt.shtml).

Суждения о степени взаимопонятности взяты в каждом случае из языковых описаний, именно поэтому так велико количество попавших в третью зону, одно из значений которой — «данных нет» и которая, соответственно, отсутствует у Яхонтова.

Предварительный анализ показывает, что для пар с ПС > 93 % всегда отмечается хорошее взаимопонимание и входящие в такие пары идиомы в подавляющем большинстве считаются диалектами одного языка. Напротив, между идиомами в парах с ПС < 87 % наблюдается отсутствие взаимопонимания, и такие идиомы как правило считаются отдельными языками. Поэтому попробуем подробнее рассмотреть среднюю часть интервала, с ПС от 87 до 93 %. В этот интервал попадают 118 пар.

Таблица 5

## Распределение пар языков по процентам совпадений и уровню взаимопонятности

ПС	1	2	3	4	5	Всего пар	%, 3 <sup>15</sup>	Среднее
93	4	10	6	1		21	29	2,2
92	2	8	1			11	9	1,9
91	3	6	10	1		20	50	2,5
90		1	8	13	1	23	35	3,6
89			6	10		16	38	3,6
88			3	11	2	16	19	3,9
87			3	7	1	11	27	3,8
<b>Всего</b>	9	25	37	43	4	<b>118</b>	31	3,1

Итак, мы видим, что, если не брать во внимание 3-й столбец, практически все пары с ПС, равным 91 % и больше, попадают в 1 и 2 столбцы, а с ПС, равным 90 % и меньше, — в 4 и 5 столбцы.

Теперь рассмотрим, насколько лексикостатистические проценты совпадений коррелируют с традиционным разграничением языков и диалектов. Распределим пары с ПС от 87 до 93 % по таксономическому статусу, то есть считаются ли они обычно в современных лингвистических работах разными языками или диалектами одного языка. Также используем 5-балльную шкалу:

- 1 — всегда диалекты,
- 2 — чаще диалекты,
- 3 — то языки, то диалекты,
- 4 — чаще языки,
- 5 — всегда языки.

Таблица 6

## Распределение пар языков по процентам совпадений и таксономическому статусу

ПС	1	2	3	4	5	Всего	%, 3	Среднее
93	5	7	2	5		19	11	2,4
92	5	4	1	2		12	8	2,0
91	7	0	5	7		19	26	2,6
90	2	3	5	4	9	23	22	3,7
89		4	2	5	5	16	13	3,7
88		1	2	10	4	17	12	4,0
87			1	9	2	12	8	4,1
<b>Всего</b>	19	19	18	42	20	<b>118</b>	15	3,2

<sup>15</sup> В столбце «%, 3» указаны для каждого процентного уровня доля пар, попавших в неопределенную зону (взаимопонятность = 3), а в столбце «среднее» — среднее значение взаимопонятности с учетом 3-й зоны.

По таблице 6 видно, что диалектами одного языка обычно считаются идиомы, ПС между которыми  $\geq 92\%$ , а разными языками — с ПС  $\leq 89\%$ . Однако в данном случае распределение в целом не такое однозначное, как в предыдущей таблице. Это можно объяснить разными традициями в описании языков разных языковых групп (когда даже взаимопонятные идиомы считаются разными языками) и значительным влиянием исторических и этно-социальных факторов. Например, ПС = 93 % демонстрируют такие пары, как стандартный немецкий — люксембургский, средненемецкие диалекты — идиш, исландский — букмол, фриульский — северно-венетский, сусу — дьялонка (манде); 92 % наблюдается в таких парах, как чешский — словацкий, теннет — мурле (сурмийские), мегленитский — арумьинский.

Таким образом, в качестве рабочей гипотезы можно достаточно уверенно считать, что при ПС  $\geq 91\%$  идиомы являются взаимопонимаемыми и могут считаться диалектами одного языка, а при ПС  $\leq 90\%$  — не являются взаимопонимаемыми и могут считаться разными языками. Диапазон между 90 и 91 % является переходной зоной и при установлении таксономического статуса в таких случаях следует собирать дополнительные данные и учитывать другие факторы.

То, что мы не проводим абсолютно точную границу, а оставляем узкую «зону сомнения», говорит не о недостатках метода, а скорее о том, что он отражает некоторую объективную реальность, так как вполне естественно, что любой диалектный пучок в своем развитии в какой-то момент проходит через такую переходную стадию, когда нельзя с определенностью сказать, по-прежнему ли это один язык или уже несколько.

## 6. Терминология

Можно сказать, что существует (как минимум) две сущности под названием «язык» (ср. [Haugen 1966: 926; Durtell 1990: 499; Salminen 2007: 213—216]):

- «собственно язык» (СЯ; «структурное понимание» по Хаугену) — совокупность лингвистических характеристик, ментальная грамматика носителя; определяется прежде всего разговорными проявлениями языка и не зависит от престижа, социального статуса или наличия литературной традиции;
- «социальный язык» (ФЯ; «функциональное понимание» по Хаугену) — это проекция собственно языка, на которую наложены другие факторы типа функционального статуса, этнического сознания, литературной обработки, политики и т. д.

Кроме того, можно говорить также о «традиционном языке» (ТЯ) — это любой идиом, имеющий статус языка как в лингвистических работах, так и вне их.

Рассмотрим возможные терминологические решения для ситуаций, когда традиционный язык совпадает или не совпадает с собственно языком. Тут возможны три ситуации:

1. Один традиционный язык соответствует одному собственно языку (СЯ = ТЯ ниже в таблице). Идеальная ситуация, которая встречается не так уже редко, но и не слишком часто. Примеры: каталанский, латышский, валлийский, корейский языки.
2. Один традиционный язык соответствует нескольким собственно языкам (ТЯ > СЯ). В последнее время для обозначения такой группы родственных языков стало употребляться название «макроязык», призванное показать некоторое этно-социальное единство рассматриваемых языков<sup>16</sup> [Lewis et al. 2017]. К примерам можно отнести арабский макроязык, китайский макроязык, сапотекский макроязык, даргинский макроязык.
3. Несколько традиционных языков соответствует наречиям одного собственно языка (ТЯ < СЯ). Для данной ситуации было предложено следующее решение [Коряков

<sup>16</sup> Хотя стоит упомянуть, что термин «макроязык» может употребляться также для случая 3, когда наречия одного языка (= макроязыка-2) на основе этно-социальных критериев считаются отдельными языками.

2006: 9—11]: в случае, если собственно язык включает в себя идиомы (наречия или диалекты), с этно-социально-политической точки зрения считающиеся отдельными языками, он может называться «языковым кластером», с пониманием того, что входящие в него «языки» соответствуют наречиям<sup>17</sup> в лингвистическом смысле слова. Примерами таких кластеров являются чехословацкий (ПС = 92 %), сербохорватский, болгаро-македонский (90,5 %), континентально-скандинавский (95 %), адыгский, португало-галисийский (96 %), коми (включающий коми-зырянский и коми-пермяцкий языки и коми-язьвинское наречие; 96 %).

Более наглядно это можно представить в таблице 7. (Значения, отделяющие последний уровень от предпоследнего, даны приблизительно, так как разделение на наречия, диалекты и говоры не является предметом данного исследования. Ср. также [Бурлак, С. Старостин 2005: 19], где считается, что «родственные диалекты обычно имеют более 95 % совпадений в базисной лексике».)

Таблица 7

## Употребление таксономических рангов

ПС	Баллы	Таксономические ранги			Примеры (в скобках ПС)		
		СЯ = ТЯ	ТЯ > СЯ	ТЯ < СЯ	СЯ = ТЯ	ТЯ > СЯ	ТЯ < СЯ
< 91 %	5	группа языков	макро-язык	группа языков	западнороманские (77)	«даргинский макроязык» (67)	западнославянские и восточнославянские языки (80)
	3—4	язык	язык	языковой кластер	окситанский / каталанский (85)	севернодаргинский / муиринский (85)	чехословацкий кластер и польский язык (86)
91—96 %	2—3	наречие	наречие	«язык»	собственно каталанское / валенсийское / балеарское (92—95)	акушинское / мюрего-губденское и т. д. (91)	чешский и словацкий (92), польский и кашубский (91)
> 96—97 %	1	диалект / говор	диалект / говор	диалект / говор	кастельонский / центральноваленсийский и др. (97—99)	мюрегинский / губденский / мекегинский (96)	собственно чешские и ганакские диалекты, мазовецкий и великопольский диалекты

## 7. Примеры использования метода

В заключение хотелось бы рассмотреть некоторые приложения описанного выше метода.

Характерным примером несовпадения этно-социо-политических и лингвистических признаков разграничения языков и диалектов являются **скандинавские** языки. Существует пять наций, пять литературных языков (причем норвежский еще и в двух вариантах), пять отдельных политических образований (хотя Фарёвские острова и входят в состав Королевства Дания). Но как минимум между континентальными разновидностями (норвежский, шведский и датский) высока степень взаимопонятности [Delsing, Lundin-Åkesson

<sup>17</sup> Под «наречиями» в данном случае понимается крупный диалект или группа диалектов.

2005; Gooskens 2006] и структурной близости. Лексикостатистически выделяются в лучшем случае два языка: островной (исландский и фарерский, ПС = 94 %) и континентальный (ПС = 95 %), степень близости между которыми (ПС = 90,6 %) попадает в переходную зону [G. Starostin 2016].

**Романские языки Италии** представляют наиболее сложную в классификационном плане часть романского ареала. Традиционно выделяются итальянский язык с большим количеством гетерогенных диалектов, которые можно объединить в три или четыре группы [Чельшева 2001: 92—93], сардинский язык (или языки) и условная группа ретороманских языков (ранее называвшихся одним языком): романшский, ладинский и фриульский.

Данные лексикостатистики [Saenko 2016] показывают, что, ожидаемо, дальше всех отстоит сардинская группа, противопоставленная всем западнороманским языкам (ПС = 74 %). Далее отдельно стоит романшский язык, который сам, возможно, делится на два языка (рейнский и энгадинский, ПС = 90 %). Далее четко выделяются ладинский язык, североитальянский ареал, тосканский язык и южно-итальянский язык. Вопрос о количестве языков внутри североитальянского ареала, даже несмотря на большое количество собранных данных (22 разновидности), остается не до конца ясным, так как, во-первых, этот ареал представляет собой сложный диалектный континуум, а во-вторых, проценты совпадений между многими попадают как раз в переходную зону 90—91 %. Условно можно говорить о четырех-шести языках, причем не исключено, что фриульский образует один язык с венецким. Напротив, южно-итальянский ареал представляет собой довольно тесное единство с ПС = 95 между неаполитанскими и сицилийскими говорами.

**Обско-угорские языки** традиционно включают два языка: хантыйский и мансийский. В первом выделяется три наречия и большое количество диалектов. Тем не менее взаимопонимание между отдельными наречиями отсутствует [Comrie 1981: 7; Janhunen, Salminen 2000; Кошкарева 2013: 50], а попытки создать единую литературную норму не увенчались успехом — отдельные письменности развиваются для четырех диалектов. Таким образом, есть все основания усомниться в единстве хантыйского языка, что лишь подтверждается лексикостатистическими данными [Zhivlov 2014]. Согласно последним надежно выделяется три хантыйских языка: собственно хантыйский (северный), хандэйский (южный) и кантыкский (восточный). Более того, единство кантыкского также находится под вопросом: его диалекты образуют диалектный континуум, между крайними членами которого ПС опускается до 86—88 %.

Мансийский, который традиционно делится на четыре наречия, оказывается более однородным. С точки зрения лексикостатистики в нем можно выделить максимум два языка: северномансийский и вымерший южно-средне-мансийский (соответствующий западному, восточному и южному наречиям), ПС между которыми равен 87 % [Zhivlov 2014].

## 8. Выводы

Утверждая, что некий идиом является именно языком или именно диалектом, необходимо оговаривать, какой смысл мы вкладываем в понятие «язык» (см. выше о различении структурного и функционального понимания) и на основании каких критериев делается этот вывод. Причем для разных целей может быть востребовано то или иное понимание языка. Для целей, описанных в начале работы, важно, скорее, структурное понимание.

Как показано выше, предлагаемый критерий имеет достаточно высокую корреляцию со степенью взаимопонятности (которая, в свою очередь, является основополагающим критерием для разграничения двух языков) и в то же время свободен от субъективных недостатков последней (отношение носителей к другому идиому и предварительное знакомство с языком собеседника). Более того, хотя получение качественного стословного списка требует определенных усилий, в целом оно гораздо проще проведения качественных тестов на взаимопонятность и далеко не всегда требует непосредственного контакта с носителями. Таким образом, предлагаемый метод может быть применен к гораздо большему

числу языков, потенциально — ко всем, по которым имеются достаточно подробные лексические данные или доступ к носителям.

Большинство же цитируемых сложных случаев связано либо с путаницей между лингвистическими и социо-этно-политическими критериями, либо со смешиванием диахронии и синхронии [Salminen 2007: 214]. Немногие по-настоящему затруднительные случаи связаны с действительно переходными ситуациями и по своему количеству сильно уступают противоречивым ситуациям, возникающим при применении социо-этно-политических критериев.

## СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Аванесов 1962 — Аванесов Р. И. (ред.). Вопросы теории лингвистической географии. М.: Изд-во Академии наук СССР, 1962. [Avanesov R. I. (ed.). *Voprosy teorii lingvisticheskoi geografii* [Problems of the theory of linguistic geography]. Moscow: Academy of Sciences of the USSR Publ., 1962.]
- Аврорин 1975 — Аврорин В. А. Проблемы изучения функциональной стороны языка (к вопросу о предмете социолингвистики). Л.: Наука, 1975. [Avrorin V. A. *Problemy izucheniya funktsional'noi storony yazyka (k voprosu o predmete sotsiolingvistiki)*. [Problems of study of the functional aspect of language (on the subject matter of sociolinguistics)]. Leningrad: Nauka, 1975.]
- Алиева 2011 — Алиева Н. Р. Малайский язык // Большая российская энциклопедия. Т. 18. М.: Большая российская энциклопедия, 2011. С. 625—626. [Alieva N. R. Malay. *Bol'shaya Rossiiskaya entsiklopediya*. Vol. 18. Moscow: Bol'shaya Rossiiskaya Entsiklopediya, 2011. Pp. 625—626.]
- Алпатов 1997 — Алпатов В. М. Японский язык // Алпатов В. М., Кормушин И. В., Пюрбеев Г. Ц. и др. (ред.). Языки мира. Монгольские языки. Тунгусо-маньчжурские языки. Японский язык. Корейский язык. М.: Academia, 1997. С. 324—360. [Alpatov V. M. Japanese. *Yazyki mira. Mongol'skie yazyki. Tunguso-man'chzhurskie yazyki. Yaponskii yazyk. Koreiskii yazyk*. Alpatov V. M., Kormushin I. V., Pyurbeev G. Ts. et al. (eds.). Moscow: Academia, 1997. Pp. 324—360.]
- Алпатов 2000 — Алпатов В. М. 150 языков и политика. 1917—2000. Социолингвистические проблемы СССР и постсоветского пространства. Москва: Крафт+, ИВ РАН, 2000. [Alpatov V. M. *150 yazykov i politika. 1917—2000. Sotsiolingvisticheskie problemy SSSR i postsovetского prostranstva* [150 languages and politics. 1917—2000. Sociolinguistic problems of the USSR and the former Soviet Union]. Moscow: Kraft+, Institute of Oriental Studies, Russian Academy of Sciences, 2000.]
- Багдасаров 1994 — Багдасаров А. Р. Социолингвистический аспект языковых отношений в Хорватии и Сербии во второй половине XX в. // Арутюнов С. А., Багдасаров А. Р., Белоусов В. Н. и др. (ред.). Язык. Культура. Этнос. М.: Наука, 1994. С. 198—210. [Bagdasarov A. R. A sociolinguistic aspect of language relations in Croatia and Serbia in the 2<sup>nd</sup> half of the 20<sup>th</sup> century. *Yazyk. Kul'tura. Etnos*. Moscow: Nauka, 1994. Pp. 198—210.]
- Беликов, Крысин 2001 — Беликов В. И., Крысин Л. П. Социолингвистика: Учебник для вузов. М.: Изд-во РГГУ, 2001. [Belikov V. I., Krysin L. P. *Sotsiolingvistika: Uchebnik dlya vuzov* [Sociolinguistics. A college textbook]. Moscow: Russian State Univ. for the Humanities Publ., 2001.]
- Бирилло и др. 2005 — Бирилло Н. В., Мацкевич Ю. Ф., Михневич А. Е., Рогова Н. В. Белорусский язык // Молдован А. М., Скорвид С. С., Кибрик А. А. и др. (ред.). Языки мира. Славянские языки. М.: Academia, 2005. С. 548—594. [Birillo N. V., Matskevich Yu. F., Mikhnevich A. E., Rogova N. V. Belarusian. *Yazyki mira: Slavyansknye yazyki*. Moldovan A. M., Skorvid S. S., Kibrik A. A. et al. (eds.). Moscow: Academia, 2005. Pp. 548—594.]
- Бурлак, Старостин 2005 — Бурлак С. А., Старостин С. А. Сравнительно-историческое языкознание: Учебник для студентов высших учебных заведений. М.: Издательский центр «Академия», 2005. [Burlak S. A., Starostin S. A. *Sravnitel'no-istoricheskoe yazykoznanie: Uchebnik dlya studentov vysshikh uchebnykh zavedenii* [Comparative linguistics: A college textbook]. Moscow: Academia Publ., 2005.]
- Виноградов 2000 — Виноградов В. В. (ред.). Словарь языка Пушкина. Т. 1. Москва: Азбуковник, 2000. [Slovar' yazyka Pushkina [A dictionary of Pushkin's language]. Vol. 1. Moscow: Azbukovnik, 2000.]
- Глисон 1959 — Глисон Г. Введение в дескриптивную лингвистику. М.: Изд-во иностранной литературы, 1959. [Gleason H. *Vvedenie v deskriptivnyuyu lingvistiku* [An introduction to descriptive linguistics]. Moscow: Izdatel'stvo Inostrannoi Literatury, 1959.]
- Дуличенко 2005a — Дуличенко А. Д. Малые славянские литературные языки // Молдован А. М., Скорвид С. С., Кибрик А. А. и др. (ред.). Языки мира: Славянские языки. М.: Academia, 2005.

- C. 595—615. [Dulichenko A. D. Minor Slavic literary languages. *Yazyki mira: Slavyanskije yazyki*. Moldovan A. M., Skorvid S. S., Kibrik A. A. et al. (eds.). Moscow: Academia, 2005. Pp. 595—615.]
- Дуличенко 2005б — Дуличенко А. Д. Словенский язык // Молдован А. М., Скорвид С. С., Кибрик А. А. и др. (ред.). Языки мира: Славянские языки. М.: Academia, 2005. С. 198—233. [Dulichenko A. D. Slovenian. *Yazyki mira: Slavyanskije yazyki*. Moldovan A. M., Skorvid S. S., Kibrik A. A. et al. (eds.). Moscow: Academia, 2005. Pp. 198—233.]
- Елоева 1995 — Елоева Ф. А. Тюркоязычные православные греки Восточной Грузии (Цалкинский и Тетрицкароисский районы). СПб: Языковой центр филол. фак-та СПбГУ, 1995. [Eloeva F. A. *Turkoyazychnye pravoslavnye greki Vostochnoi Gruzii (Tsalkinskii i Tetrtskaroisskii raiony)* [Turkic-speaking Orthodox Greeks of East Georgia (Tsalkin and Tetrtsqaro areas)]. St. Petersburg: Language Center of the Philology Department, St. Petersburg State Univ., 1995.]
- Жовтобрюх, Молдован 2005 — Жовтобрюх М. А., Молдован А. М. Украинский язык // Молдован А. М., Скорвид С. С., Кибрик А. А. и др. (ред.). Языки мира: Славянские языки. М.: Academia, 2005. С. 513—548. [Zhovtobryukh M. A., Moldovan A. M. Ukrainian. *Yazyki mira: Slavyanskije yazyki*. Moldovan A. M., Skorvid S. S., Kibrik A. A. et al. (eds.). Moscow: Academia, 2005. Pp. 513—548.]
- Калнынь 1976 — Калнынь Л. Э. Диалектологический аспект проблемы «язык и диалект» // Известия АН СССР. Серия литературы и языка. 1976. Т. 35. № 1. С. 34—45. [Kalnyn' L. E. Dialectological aspect of the problem "language and dialect". *Izvestiya AN SSSR. Ser. literatury i yazyka*. 1976. Vol. 35. No. 1. P. 34—45.]
- Кармышева 1969 — Кармышева Б. Х. Сарты // Советская историческая энциклопедия. Т. 12: Репарации — Славяне. М.: Советская энциклопедия, 1969. [Karmysheva B. Kh. Sarts. *Sovetskaya istoricheskaya entsiklopediya. T. 12: Reparatsii — Slavyane*. Moscow: Sovetskaya Entsiklopediya, 1969.]
- Касаткин 1990 — Касаткин Л. Л. Диалект // Лингвистический энциклопедический словарь. М.: Советская энциклопедия, 1990. [Kasatkin L. L. Dialect. *Lingvisticheskii entsiklopedicheskii slovar'*. Moscow: Sovetskaya Entsiklopediya, 1990.]
- Касьян 2014 — Касьян А. С. К формальной генеалогической классификации лезгинских языков (Северный Кавказ) // Вопросы языкового родства. 2014. № 11. С. 63—80. [Kassian A. S. Towards a formal genealogical classification of Lezgian languages of the North Caucasus. *Journal of Language Relationship*. 2014. No. 11. Pp. 63—80.]
- Климчук 1990 — Климчук Ф. Д. Этнос и перепись: парадоксы статистики // Гусейнов Г., Драгунский Д. (ред.). Ожог родного очага. М.: Прогресс, 1990. С. 92—106. [Klimchuk F. D. Ethnos and census: Paradoxes of statistics. *Ozhog rodnogo ochaga*. Guseinov G., Dragunskii D. (eds.). Moscow: Progress, 1990. Pp. 92—106.]
- Коряков 2005 — Коряков Ю. Б. Карты славянских языков. 10. Украинский язык // Молдован А. М., Скорвид С. С., Кибрик А. А. и др. (ред.). Языки мира: Славянские языки. М.: Academia, 2005. С. 649. [Koryakov Yu. B. Maps of the Slavic languages. 10. Ukrainian. *Yazyki mira: Slavyanskije yazyki*. Moldovan A. M., Skorvid S. S., Kibrik A. A. et al. (eds.). Moscow: Academia, 2005. P. 649.]
- Коряков 2006 — Коряков Ю. Б. Атлас кавказских языков. М.: Институт языкознания РАН, 2006. [Koryakov Yu. B. *Atlas kavkazskikh yazykov* [Atlas of Causasian languages]. Moscow: Institute of Linguistics, Russian Academy of Sciences, 2006.]
- Коряков, Майсак 2001 — Коряков Ю. Б., Майсак Т. А. Систематика языков мира и базы данных в интернете // Труды Международного семинара «Диалог '2001» по компьютерной лингвистике и ее приложениям. Т. 2. М.: Аксаково, 2001. С. 215—229. [Koryakov Yu. B., Maisak T. A. Systematics of the world languages and databases in the Internet. *Trudy Mezhdunarodnogo seminar "Dialog 2001" po komp'yuternoi lingvistike i ee prilozheniyam*. Vol. 2. Moscow: Aksakovo, 2001. Pp. 215—229.]
- Кошкарёва 2013 — Кошкарёва Н. Б. Актуальные вопросы совершенствования хантыйской графики и орфографии // Вестник угроведения. 2013. № 3(14). С. 52—72. [Koshkareva N. B. Current issues of the development of Khanty graphics and orthography. *Vestnik ugrovedeniya*. 2013. No. 3(14). Pp. 52—72.]
- Кречмер, Невекловский 2005 — Кречмер А. Г., Невекловский Г. Сербохорватский язык (сербский, хорватский, боснийский языки) // Молдован А. М., Скорвид С. С., Кибрик А. А. и др. (ред.). Языки мира: Славянские языки. М.: Academia, 2005. С. 139—198. [Krechmer A. G., Nevekovskii G. Serbo-Croatian (Serbian, Croatian, Bosnian). *Yazyki mira: Slavyanskije yazyki*. Moldovan A. M., Skorvid S. S., Kibrik A. A. et al. (eds.). Moscow: Academia, 2005. Pp. 139—198.]
- Кузнецова 2012 — Кузнецова А. И. Грамматические и лексические раритеты в языке бесермян в сопоставлении с другими уральскими языками // Финно-угорские языки: фрагменты грамматического описания. Формальный и функциональный подходы. М.: Языки славянских культур, 2012.

- С. 41—59. [Kuznetsova A. I. Grammatical and lexical rarities in the language of Besermians in comparison with the other Uralic languages. *Finno-ugorskie yazyki: fragmenty grammaticheskogo opisaniya. Formal'nyi i funktsional'nyi podkhody*. Moscow: Yazyki Slavyanskikh Kul'tur, 2012. Pp. 41—59.]
- Лухт, Нарумов 2001 — Лухт Л. И., Нарумов Б. П. Румынский язык // Чельшева И. И., Нарумов Б. П., Романова О. И. (ред.). Языки мира: Романские языки. М., Academia, 2001. С. 574—636. [Lukht L. I., Narumov B. P. Romanian. *Yazyki mira: Romanskije yazyki*. Chelysheva I. I., Narumov B. P., Romanova O. I. (eds.). Moscow, Academia, 2001. Pp. 574—636.]
- Майсак 2005 — Майсак Т. А. Андийские языки // Большая российская энциклопедия. Т. 1. М.: Большая российская энциклопедия, 2005. С. 718—719. [Maisak T. A. Andic languages. *Bol'shaya rossiiskaya entsiklopediya*. Vol. 1. Moscow: Bol'shaya Rossiiskaya Entsiklopediya, 2005. Pp. 718—719.]
- Мечковская 1996 — Мечковская Н. Б. Социальная лингвистика. М.: Аспект-Пресс, 1996. [Mechkovskaya N. B. *Sotsial'naya lingvistika* [Social linguistics]. Moscow: Aspekt-Press, 1996.]
- Мусорин 2001 — Мусорин А. Ю. Что такое отдельный язык? // Сибирский лингвистический семинар. 2001. № 1. С. 12—16. [Musorin A. Yu. What is a separate language? *Sibirskii lingvisticheskii seminar*. 2001. No. 1. Pp. 12—16.]
- Перехвальская 1997 — Перехвальская Е. В. Статус идиома и этническое самосознание // Кузнецова А. И., Раевская О. В., Скорвид С. С. (ред.). Малые языки Евразии: Социолингвистический аспект. М.: МГУ имени М. В. Ломоносова, 1997. С. 5—11. [Perekhval'skaya E. V. Status of an idiom and ethnic identity. *Malye yazyki Evrazii: Sotsiolingvisticheskii aspekt*. Kuznetsova A. I., Raevskaya O. V., Skorvid S. S. (eds.). Moscow: Lomonosov Moscow State Univ., 1997. Pp. 5—11.]
- Пиотровский 1973 — Пиотровский Р. Г. Лингвистические оценки расхождения близкородственных языков // Вопросы языкознания. 1973. № 5. С. 36—42. [Piotrovskij R. G. Linguistic estimates of differences in closely related languages. *Voprosy jazykoznanija*. 1973. No. 5. Pp. 36—42.]
- Романова и др. 2014 — Романова О. И., Кибрик А. А., Коряков Ю. Б. Презентация новых книг лингвистической энциклопедии «Языки мира» (Москва, 31 марта 2014 г.) // Вестник Российского гуманитарного научного фонда. 2014. № 2. С. 166—172. [Romanova O. I., Kibrik A. A., Koryakov Yu. B. Presentation of new volumes of the linguistic encyclopedia “Languages of the world” (Moscow, March 31, 2014). *Vestnik Rossiiskogo gumanitarnogo nauchnogo fonda*. 2014. No. 2. Pp. 166—172.]
- Ростовцев-Попель 2012 — Ростовцев-Попель А. А. Мегрельский язык // Большая российская энциклопедия. Т. 19. М.: Большая российская энциклопедия, 2012. С. 455—456. [Rostovtsev-Popel A. A. Mingrelian. *Bol'shaya rossiiskaya entsiklopediya*. Vol. 19. Moscow: Bol'shaya Rossiiskaya Entsiklopediya, 2012. Pp. 455—456.]
- С. Старостин 2007 — Старостин С. А. О доказательстве языкового родства // Старостин С. А. Труды по языкознанию. М.: Языки славянских культур, 2007. С. 779—793. [Starostin S. A. On the validation of linguistic affinity. Starostin S. A. *Trudy po jazykoznaniju*. Moscow: Yazyki Slavyanskikh Kul'tur, 2007. Pp. 779—793.]
- Сумбатова, Ландер 2014 — Сумбатова Н. Р., Ландер Ю. А. Даргинский говор селения Танты: грамматический очерк, вопросы синтаксиса. М.: Языки славянской культуры, 2014. [Sumbatova N. R., Lander Yu. A. *Dargin'skii govor seleniya Tanty: grammaticheskii ocherk, voprosy sintaksisa* [Dargin dialect of Tanty village: A grammatical sketch, problems of syntax]. Moscow: Yazyki Slavyanskoi Kul'tury, 2014.]
- Тулский 2006 — Тульский М. Некоторые итоги переписи 2004 года в Молдавии // Демоскоп Weekly. 2006. № 249—250. [Tul'skii M. Some results of 2004 census in Moldavia. *Demoskop Weekly*. 2006. No. 249—250.]
- Усикова 2005 — Усикова Р. П. Македонский язык // Молдован А. М., Скорвид С. С., Кибрик А. А. и др. (ред.). Языки мира: Славянские языки. М.: Academia, 2005. С. 102—139. [Usikova R. P. Macedonian. *Yazyki mira: Slavyanskije yazyki*. Moldovan A. M., Skorvid S. S., Kibrik A. A. et al. (eds.). Moscow: Academia, 2005. Pp. 102—139.]
- Чельшева 2001 — Чельшева И. И. Диалекты Италии // Чельшева И. И., Нарумов Б. П., Романова О. И. (ред.). Языки мира: Романские языки. М.: Academia, 2001. С. 90—146. [Chelysheva I. I. Dialects of Italy. *Yazyki mira: Romanskije yazyki*. Chelysheva I. I., Narumov B. P., Romanova O. I. (eds.). Moscow: Academia, 2001. Pp. 90—146.]
- Эдельман 1980 — Эдельман Д. И. К проблеме «язык или диалект» в условиях отсутствия письменности // Ярцева В. Н. (ред.). Теоретические основы классификации языков мира. М.: Наука, 1980. С. 127—147. [Edel'man D. I. Towards the problem of language vs. dialect in the absence of a writing system. *Teoreticheskie osnovy klassifikatsii yazykov mira*. Yartseva V. N. (ed.). Moscow: Nauka, 1980. Pp. 127—147.]

- ЯнДМ 1982 — Языки и диалекты мира. Проспект и словарь. М.: Наука, 1982. [*Yazyki i dialekty mira. Prospekt i slovník* [Languages and dialects of the world. Prospect and vocabulary]. Moscow: Nauka, 1982.]
- Яхонтов 1980 — Яхонтов С. Е. Оценка степени близости родственных языков // Ярцева В. Н. (ред.). Теоретические основы классификации языков мира. М.: Наука, 1980. С. 148—157. [Yakhontov S. E. Estimate of the degree of affinity of related languages. *Teoreticheskie osnovy klassifikatsii yazykov mira*. Yartseva V. N. (ed.). Moscow: Nauka, 1980. Pp. 148—157.]
- Aikhenvald 2007 — Aikhenvald A. Y. Grammars in contact: A cross-linguistic perspective. *Grammars in contact: A cross-linguistic typology*. Aikhenvald A. Y., Dixon R. M. W. (eds.). Oxford: Oxford Univ. Press, 2007. Pp. 1—66. (Explorations in Linguistic Typology, 4.)
- G. Anderson 1998 — Anderson G. D. S. *Xakas*. München: Lincom Europa, 1998.
- S. Anderson 2004 — Anderson S. R. *How many languages are there in the world?* Washington (DC): Linguistic Society of America, 2004. Available at: <https://www.linguisticsociety.org/sites/default/files/how-many-languages.pdf>.
- H. Anderson 2005 — Anderson H. *Intelligibility testing (RTT) between Mendankwe and Nkwen*. SIL Electronic Survey Reports 2005—002. SIL International, 2005. Available at: <http://www.sil.org/resources/publications/entry/9217>.
- Van Bezooijen, van den Berg 1999 — van Bezooijen R., van den Berg R. Word intelligibility of language varieties in the Netherlands and Flanders under minimal conditions. *Linguistics in the Netherlands 1999*. Van Bezooijen R., Kager R. (eds.). John Benjamins: Amsterdam, 1999. Pp. 1—12.
- Blust 2013 — Blust R. *The Austronesian languages*. Canberra: Australian National Univ., Asia Pacific Linguistics, 2013.
- Borin 2013 — Borin L. The why and how of measuring linguistic differences. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 3—28.
- Brown 1998 — Brown R. On criteria for identifying language groups and language clusters. *Notes on Sociolinguistics*. 1998. Vol. 3. No. 1. Pp. 3—42.
- Brye, Brye 2002 — Brye E., Brye E. Rapid appraisal and intelligibility testing surveys of the Eastern Boid group of languages. SIL Electronic Survey Reports, 2002. Available at: <http://www-01.sil.org/silesr/2002/019/SILESR2002-019.htm>.
- Carling et al. 2013 — Carling G., Eriksen L., Holmer A., van de Weijer J. Contrasting linguistics and archaeology in the matrix model: GIS and cluster analysis of the Arawakan languages. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 29—56.
- Casad 1974 — Casad E. H. *Dialect intelligibility testing*. Norman (OK): Summer Institute of Linguistics, Univ. of Oklahoma, 1974. (Summer Institute of Linguistics Publications in Linguistics and Related Fields, 38.)
- Castro, Xingwen 2015 — Castro A., Xingwen P. (eds). *Sui dialect research*. Dallas: SIL International, 2015.
- Cheng 1997 — Cheng C.-C. Measuring relationship among dialects: DOC and related resources. *Computational Linguistics and Chinese Language Processing*. 1997. Vol. 2. No. 1. Pp. 41—72.
- Clogg 2002 — Clogg R. (ed.). *Minorities in Greece: Aspects of a plural society*. London: Hurst & Company, 2002.
- Comrie 1981 — Comrie B. *The languages of the Soviet Union*. Cambridge: Cambridge Univ. Press, 1981.
- Comrie 2009 — Comrie B. (ed.). *The world's major languages*. 2<sup>nd</sup> ed. London: Croom Helm, 2009.
- Cysouw, Good 2013 — Cysouw M., Good J. Languoid, doculect, glossonym: Formalizing the notion “language”. *Language Documentation and Conservation*. 2013. Vol. 7. Pp. 331—359.
- Dalby 2000 — Dalby D. *Linguasphere register of the world's languages and speech communities*. Hebron: Linguasphere Observatory, 2000.
- Delsing, Lundin-Åkesson 2005 — Delsing L.-O., Lundin-Åkesson K. *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska*. Köpenhamn: Nordiska ministerrådet, 2005.
- Durrell 1990 — Durrell M. Language as geography. *An encyclopaedia of language*. Collinge N. E. (ed.). London: Routledge, 1990.
- Dyen 1965 — Dyen I. *A lexicostatistical classification of the Austronesian languages*. Baltimore: Waverly Press, 1965.
- Dyen, Aberle 1974 — Dyen I., Aberle D. F. *Lexical reconstruction. The case of the Proto-Athapaskan kinship system*. London: Cambridge Univ. Press, 1974.
- Ellison, Kirby 2006 — Ellison T. M., Kirby S. Measuring language divergence by intra-lexical comparison. *Proceedings of the 21<sup>st</sup> International Conference on Computational Linguistics and 44<sup>th</sup> Annual Meeting of the ACL*. Burwood: ACL, 2006. Pp. 273—280.

- GLD — The Global Lexicostatistical Database. Available at: <http://starling.rinet.ru/new100>.
- Gooskens 2006 — Gooskens Ch. Linguistic and extra-linguistic predictors of Inter-Scandinavian intelligibility. *Linguistics in the Netherlands 23*. Van de Weijer J., Los B. (eds.). Amsterdam: John Benjamins. Pp. 101—113.
- Gooskens 2007 — Gooskens Ch. The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development*. 2007. Vol. 28. No. 6. Pp. 445—467.
- Gooskens, Heeringa 2004 — Gooskens Ch., Heeringa W. The position of Frisian in the Germanic language area. *On the boundaries of phonology and phonetics*. Gilbers D., Schreuder M., Knevel N. (eds.). Groningen: Univ. of Groningen, 2004. Pp. 61—87.
- Hammarström 2008 — Hammarström H. Counting languages in dialect continua using the criterion of mutual intelligibility. *Journal of Quantitative Linguistics*. 2008. Vol. 15. No. 1. Pp. 34—45.
- Hammarström et al. 2017 — Hammarström H., Forkel R., Haspelmath M., Bank S. (eds.). Glottolog 2.7. Jena: Max Planck Institute for the Science of Human History, 2017. Available at: <http://glottolog.org>.
- Hammarström, O'Connor 2013 — Hammarström H., O'Connor L. Dependency-sensitive typological distance. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 329—352.
- Haugen 1966 — Haugen E. Dialect, language, nation. *American Anthropologist*. 1966. Vol. 68. No. 4. Pp. 922—935.
- Heeringa et al. 2013 — Heeringa W., Golubovic J., Gooskens Ch., Schüppert A., Swarte F., Voigt S. Lexical and orthographic distances between Germanic, Romance and Slavic languages and their relationship to geographic distance. *Phonetics in Europe: Perception and production*. Gooskens Ch., van Bezooijen R. (eds.). Frankfurt am Main: Peter Lang, 2013. Pp. 99—137.
- Heggarty et al. 2005 — Heggarty P., McMahon A., McMahon R. From phonetic similarity to dialect classification: a principled approach. *Perspectives on variation: Sociolinguistic, historical, comparative*. Delbecque N., Geeraerts D., van der Auwera J. (eds.). Amsterdam: De Gruyter Mouton, 2005. Pp. 43—91.
- Hewitt 2004 — Hewitt B. G. *Introduction to the study of the languages of the Caucasus*. München: Lincom Europa, 2004.
- Hickerson et al. 1952 — Hickerson H., Turner G. D., Hickerson N. P. Testing procedures for estimating transfer of information amongst Iroquois languages and dialects. *International Journal of American Linguistics*. 1952. Vol. 18. Pp. 1—8.
- Homola, Kuboň 2006 — Homola P., Kuboň V. A structural similarity measure. *Linguistic distances*. Nerbonne J., Hinrichs E. (eds.). Sydney: ACL, 2006. Pp. 91—99.
- Hudson 1996 — Hudson R. A. *Sociolinguistics*. New York: Cambridge Univ. Press, 1996.
- Hussain Sandhu, Maldonado 2015 — Hussain Sandhu A., Maldonado G. M. I. Language and dialect: Criteria and historical evidence. *Grassroots*. 2015. Vol. 49. No. 1. Pp. 203—217.
- Janhunen, Salminen 2000 — Janhunen J., Salminen T. UNESCO red book on endangered languages: North-east Asia. Available at: [http://www.helsinki.fi/~tasalmin/nasia\\_index.html](http://www.helsinki.fi/~tasalmin/nasia_index.html).
- Kassian et al. 2010 — Kassian A., Starostin G., Dybo A., Chernov V. The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship*. 2010. No. 4. Pp. 46—89.
- Kaufman 1994 — Kaufman T. The native languages of Latin America: General remarks. *Atlas of the world's languages*. Moseley Ch., Asher R. E. (eds.). London: Routledge, 1994. Pp. 31—33.
- Kondrak 2013 — Kondrak G. Word similarity, cognation, and translational equivalence. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 375—386.
- Koryakov 2013 — Koryakov Yu. B. Convergence and divergence in the classification of Dargwa languages. *46<sup>th</sup> Annual Meeting of the Societas Linguistica Europaea (SLE 2013). Book of abstracts*. Part 1. Split: Univ. of Split, 2013. Pp. 193—194.
- Kurpaska 2010 — Kurpaska M. *Chinese language(s)*. Berlin: De Gruyter Mouton, 2010.
- Kushniarevich et al. 2015 — Kushniarevich A., Utevska O., Kassian A., Dybo A., Villems R., Balanovsky O. et al. Genetic heritage of the Balto-Slavic speaking populations: A synthesis of autosomal, mitochondrial and Y-chromosomal data. *PLoS ONE*. 2015. Vol. 10. No. 9. Pp. 1—19.
- Lenz et al. 2009 — Lenz A. N., Gooskens Ch., Reker S. *Low Saxon across borders — Niedersächsische Dialekte über Grenzen hinweg*. Stuttgart: Franz Steiner Verlag, 2009.
- Lewis et al. 2017 — The problem of language identification. *Ethnologue: Languages of the world*. 20<sup>th</sup> ed. Lewis M. P., Simons G. F., Fennig Ch. D. (eds.). Dallas (TX): SIL International, 2017. Available at: <https://www.ethnologue.com/about/problem-language-identification>.
- Mané 2012 — Mané D. Concepts of language and dialect and sociolinguistic prejudice. *Via Litterae*. 2012. Vol. 4. No. 1. Pp. 39—51.

- McGregor 2013 — McGregor W. B. Comparing linguistic systems of categorisation. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 387—428.
- Montes Giraldo 1980 — Montes Giraldo J. J. Lengua, dialecto y norma. *Thesaurus*. 1980. XXXV. Vol. 2. No. 23. Pp. 237—257.
- Nerbonne et al. 2013 — Nerbonne J., van Ommen S., Gooskens Ch., Wieling M. Measuring socially motivated pronunciation differences. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 107—140.
- Okura 2015 — Okura E. Language vs. dialect in language cataloguing: The vexed case of Otomanguean dialect continua. *University of Hawai'i at Mānoa Working Papers in Linguistics*. 2015. Vol. 46. No. 5. Pp. 1—19.
- Orywal 1990 — Orywal E. A. VIII 13 Middle East. Ethnic groups — the Emic view [Karte]. *Tübinger Atlas des Vorderen Orients (TAVO)*. Tübingen: Dr. Ludwig Verlag, 1990.
- Petyt 1980 — Petyt K. M. *The study of dialect: An introduction to dialectology*. London: André Deutsch Ltd., 1980.
- Pierce 1952 — Pierce J. E. Dialect distance testing in Algonquian. *International Journal of American Linguistics*. 1952. Vol. 18. Pp. 203—210.
- Roche, Suzuki (in print) — Roche G., Suzuki H. Tibet's minority languages: Diversity and endangerment. To appear in *Modern Asian Studies*.
- Saenko 2016 — Saenko M. Annotated Swadesh wordlists for the Romance group (Indo-European family). The Global Lexicostatistical Database, 2016. Available at: <http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\ier\rom&limit=-1>.
- Sallabank 2011 — Sallabank J. Language endangerment. *The SAGE handbook of sociolinguistics*. Wodak R., Johnstone B., Kerswill P. E. (eds.). London: SAGE, 2011. Pp. 496—512.
- Salminen 2007 — Salminen T. Europe and North Asia. *Encyclopedia of the world's endangered languages*. Moseley Ch. (ed.). London: Routledge, 2007. Pp. 211—282.
- Saxena, Borin 2013 — Saxena A., Borin L. Carving Tibeto-Kanauri by its joints: Using basic vocabulary lists for genetic grouping of languages. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 175—198.
- Schepens et al. 2013 — Schepens J., van der Slik F., van Hout R. The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 199—230.
- Schüppert, Gooskens 2010 — Schüppert A., Gooskens Ch. The influence of extra-linguistic factors on mutual intelligibility: Some preliminary results from Danish and Swedish pre-schoolers. *Proceedings of Methods in dialectology*. Heselwood B., Upton C. (eds.). Frankfurt am Main: Peter Lang, 2010. Pp. 194—203.
- Sejane, Eger 2013 — Sejane I., Eger S. Semantic typologies by means of network analysis of bilingual dictionaries. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 447—474.
- Shimoji 2010 — Shimoji M. Ryukyuan languages: An introduction. *An introduction to Ryukyuan languages*. Shimoji M., Pellard T. (eds.). Tokyo: ILCAA, 2010.
- Snoek 2013 — Snoek C. Using semantically restricted word-lists to investigate relationships among Athapaskan languages. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 231—248.
- G. Starostin 2010 — Starostin G. Preliminary lexicostatistics as a basis for language classification: A new approach. *Journal of Language Relationship*. 2010. No. 3. Pp. 79—116.
- G. Starostin 2013 — Starostin G. Lexicostatistics as a basis for language classification: Increasing the pros, reducing the cons. *Classification and evolution in biology, linguistics and the history of science: Concepts — Methods — Visualization*. Fangerau H., Geisler H., Halling Th., Martin W. (eds.). Stuttgart: Franz Steiner Verlag, 2013. Pp. 125—146.
- G. Starostin 2016 — Starostin G. Annotated Swadesh wordlists for the Germanic group (Indo-European family). The Global Lexicostatistical Database, 2016. Available at: <http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\ier\grm&limit=-1>.
- Starr, Stalder 1990 — Starr A., Stalder J. *Principles of sociolinguistic surveys: Actualities of Cameroon*. Yaoundé: Summer Institute of Linguistics, 1990.
- Swadesh 1952 — Swadesh M. Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society*. 1952. Vol. 96. No. 4. Pp. 452—463.
- Swadesh 1955 — Swadesh M. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*. 1955. Vol. 21. No. 2. Pp. 121—137.

- Swarte, Hilton 2013 — Swarte F., Hilton N. H. Mutual intelligibility between speakers of North and West Frisian. *Phonetics in Europe: Perception and production*. Gooskens Ch., van Bezooijen R. (eds.). Frankfurt am Main: Peter Lang, 2013. Pp. 281—302.
- Tang, van Heuven 2009 — Tang Ch., van Heuven V. J. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*. 2009. Vol. 119. No. 5. Pp. 709—732.
- Viatta 1982 — Viatte A. French outside France. *Trends in Romance linguistics and philology*. Green J. N., Posner R. (eds.). The Hague: Mouton, 1982. Pp. 299—317.
- C. Voegelin, Harris 1951 — Voegelin C. F., Harris Z. S. Methods for determining intelligibility among dialects of natural languages. *Proceedings of the American Philosophical Society*. 1951. Vol. 95. Pp. 322—329.
- C. Voegelin, F. Voegelin 1977 — Voegelin C. F., Voegelin F. M. *Classification and index of the world's languages*. New York: Elsevier, 1977.
- Vydrin 2009 — Vydrin V. On the problem of the Proto-Mande homeland. *Journal of Language Relationship*. 2009. No. 1. Pp. 107—142.
- Wälchli, von Waldenfels 2013 — Wälchli B., von Waldenfels R. Measuring morphosemantic language distance in parallel texts. *Approaches to measuring linguistic differences*. Borin L., Saxena A. (eds.). Berlin: De Gruyter Mouton, 2013. Pp. 475—506.
- Weinreich 1944 — Weinreich M. The place of the Yivo in Jewish life. *YIVO Bleter. Journal of the Yiddish Scientific Institute*. 1944. Vol. 23. No. 3. Pp. 4—16. Available at: <http://download.hebrewbooks.org/downloadhandler.ashx?req=43629>.
- Wichmann, Holman 2009 — Wichmann S., Holman E. W. *Temporal stability of linguistic typological features*. Munich: LINCOM Europa, 2009.
- Wurm 1977 — Wurm S. A. (ed.). *New Guinea area languages and language study. Vol. 1: Papuan languages and the New Guinea linguistic scene*. Canberra: Research School of Pacific and Asian Studies, Australian National University, 1977.
- Yule 2010 — Yule G. *The study of language*. 4<sup>th</sup> ed. Cambridge: Cambridge Univ. Press, 2010.
- Zhivlov 2014 — Zhivlov M. Annotated Swadesh wordlists for the Ob-Ugric group (Uralic family). The Global Lexicostatistical Database, 2014. Available at: <http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\ura\oug&limit=-1>.

Получено / received 01.06.2016

Принято / accepted 27.06.2017