

УДК 81.286

ЯВЛЯЕТСЯ ЛИ ОПРЕДЕЛЕНИЕ ЯЗЫКОВОГО СОЮЗА НЕЧЕТКИМ?

В.Д. Соловьев

Аннотация

В статье анализируются определения понятия «языковой союз». Предложен количественный подход для оценки степени конвергенции языков в языковых союзах. Для этих целей используется типологическая база данных «Языки мира». На ее основе определяется типологическое расстояние между языками и вводится мера близости языков в языковом союзе. Приведены данные расчетов для балканского и поволжского языковых союзов. Показано, что введенная мера близости может быть использована в спорных случаях, таких как принадлежность турецкого языка к балканскому языковому союзу. Обосновано утверждение, что эта мера близости меняет статус определения языкового союза, придавая ему более строгий характер.

Ключевые слова: языковой союз, количественная лингвистика, типологические базы данных, балканские языки, контактная зона.

Введение

Хотя интуитивно понятие языкового союза (в немецкой литературе *Sprachbund*, в англоязычной литературе используется также термин *linguistic area* ('лингвистическая область')) кажется весьма простым и ясным, при его ближайшем рассмотрении обнаруживаются серьезные проблемы. Возьмем определение из [1]. Оно использует нечеткий термин «общность» и потому само является расплывчатым, а его уточнение оказывается затруднительным. Во многих работах определения языкового союза (далее ЯС) критиковались, предлагались новые, но и они не получали общего признания. На нечеткость определений было обращено внимание в работах [2–4]. Критика суммирована в работе [5], в которой в итоге вообще высказаны сомнения в возможности хорошего определения этого понятия и его плодотворности. Основные вопросы, поставленные в [5], следующие.

1. Какое минимальное число языков должно быть для признания языкового союза? Достаточно ли двух языков? Трех?

2. Сколько черт должно разделяться языками? Достаточно ли одного свойства?

3. Какова должна быть степень неродственности языков – должны ли они относиться к разным семьям?

Отметим, что вопросы 1 и 2 носят количественный характер. В последнее время на смену понятию ЯС пришел термин «контактная зона» [6], который можно трактовать как более общий по сравнению с ЯС. Однако сформулированные выше вопросы остаются актуальными и для этого понятия.

На наш взгляд, общим недостатком всех существующих определений является то, что в них не учитываются возможности и результаты применения методов квантитативной лингвистики. Цель статьи – предложить квантитативный подход, используя типологическую базу данных «Языки мира» [7]. Созданная в Институте языкознания РАН база данных «Языки мира» содержит описание 315 языков Евразии по 3815 грамматическим признакам. Это делает ее одной из крупнейших в мире типологических баз данных. В статье [8] было проведено сравнение ее с другой крупной базой данных WALS и установлено, что при статистических расчетах «Языки мира» точнее отражают типологическую близость языков.

В первом разделе статьи дается обзор существующих определений ЯС. Во втором разделе описан предлагаемый подход по отношению к количественной мере близости языков. Приводятся результаты применения этого подхода к балканскому, поволжскому и балтийскому ЯС. В заключении подводятся итоги и обсуждается вопрос, в каком смысле понятие ЯС является нечетким.

1. Обзор подходов к определению языкового союза

При анализе определений основной акцент будет сделан на втором вопросе о числе разделяемых черт. Начнем с одного из ранних определений ЯС.

Согласно [9], “This term ‘linguistic area’ may be defined as meaning an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families.” («Лингвистическая область» может быть определена как область, которая включает языки, принадлежащие более чем к одной семье, но обладающие общими признаками, не характерными для других языков (по крайней мере) одной из этих семей.»¹)

Как видим, в этом определении уточняется вопрос о степени неродственности языков, но остается неясным, сколько общих признаков должно у них быть. Кроме того, если следовать этому определению, то Балканы нельзя будет признать ЯС, так как в классическом виде в него включаются только индоевропейские языки [1].

Согласно [10], “A *linguistic area* is defined here as an area in which *several* linguistic traits are shared by the languages of the area and furthermore, there is evidence... that contact between the speakers of the languages contributed to the spread and/or retention of these traits...”. («Лингвистическая область определяется как область, языки которой обладают несколькими общими лингвистическими признаками, и, более того, существуют примеры... когда контакты между носителями этих языков способствовали распространению и/или сохранению этих черт.»)

Здесь опять остается неясным, чему равно это *several* (*несколько*). Кроме того, остается не вполне понятным следующий момент. Должны ли общие признаки разделяться **всеми** языками ЯС или только частью из них? Данное определение вроде бы предполагает, что общие признаки должны разделяться всеми языками ЯС. Однако в некоторых случаях столь жесткое ограничение может

¹ Здесь и далее перевод наш.

быть излишне обременительным. Рассмотрим классический пример ЯС – Юго-Восточную Азию. В [11] описаны свойства (всего 21), характерные для языков этого региона (относящихся к 5 языковым семьям). При этом оказалось, что ни один язык не обладает всеми свойствами. Рекордсменом является тайский язык, обладающий 19 свойствами, вьетнамский и кхмерский обладают 18 свойствами и т. д.

Если отказаться от требования, чтобы признаки разделялись непременно всеми языками, то как интерпретировать ситуации такого типа? Пусть три языка А, В, С из трех разных семей разделяют признаки (не принадлежащие другим членам по крайней мере одной из соответствующих семей) следующим образом. Языки А и В имеют n общих признаков, языки А и С – m общих признаков (вообще говоря, других), языки В и С – p общих признаков (тоже, в общем случае, других). Считать ли это ЯС, особенно, если ни один из этих признаков не разделяется всеми тремя языками?

Приведенный пример не является искусственным. Как справедливо замечает С.Г. Томасон [12], “...Changes... spread widely through region – but always from language to languages in a series of events, not in some single mystical area-wide process that affects many languages at once.” («...Изменения... широко распространяются по всему региону – но всегда от языка к языку вследствие ряда событий, а не какого-то одного мистического процесса, происходящего на данной территории, который влияет сразу на множество языков.»)

Аналогичная точка зрения высказывается также в [13, 14].

В [12] предусматривается выделение в ЯС центра и периферии по числу разделяемых признаков. Томасон пишет: “In the majority of cases, the boundaries are fussy; often there is a central group of languages that share a large proportion of the characteristic features, and scattered peripheral languages or groups that share a considerably smaller of the features. Frequently, too, there are shared features that are found only in a small subset of the area’s languages.” («В большинстве случаев границы нечетки. Часто существует центральная группа языков, которым свойственна большая доля общих характерных признаков, а также отдельные периферийные языки или группы языков, обладающие значительно меньшим количеством общих признаков. Зачастую также встречаются такие общие признаки, которые свойственны лишь немногим языкам данной области.»)

И снова термины ‘large proportion’ («большая доля»), ‘considerably smaller’ («значительно меньше»), ‘small subset’ («небольшое подмножество») остаются нечеткими.

Совершенно строгое определение предложено в [15]. В нем группа языков трактуется как языковой союз, если выполняются следующие условия: а) языки располагаются в непрерывной географической области; б) группа включает по крайней мере 2 языка; в) языками разделяется по крайней мере одно общее свойство.

Однако большинство исследователей, видимо, считает, что разделяемых признаков должно быть больше, минимум два [16]. В [12] предлагается рассматривать языковые союзы как более сильные или более слабые в зависимости от числа разделяемых черт. Характеризация ЯС как сильных или слабых

обсуждается также в [5]. ЯС с единственным разделяемым признаком является, безусловно, слабейшим из возможных [17].

Наиболее современными (предложенными уже в XXI веке) и тщательно проработанными являются следующие определения. Согласно [12], “A linguistic area is a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor.” («Лингвистическая область представляет собой географический регион, включающий группу из трех или более языков, обладающих рядом общих структурных особенностей, возникающих главным образом в результате языковых контактов, нежели вследствие случайности или наличия общего языкового предка.») Как видим, в этом определении уточняется число языков, но по-прежнему не уточняется число разделяемых признаков.

Согласно [18], “a linguistic area (or *Sprachbund*) is generally taken to be a geographically delimited area including languages from two or more language families, sharing significant traits (which are not found in languages from these families spoken outside the area). There must be a fair number of common traits and they should be reasonably distinctive.” («Под лингвистической областью обычно понимается географически ограниченная область, включающая языки из двух или более языковых семей, обладающие значимыми общими чертами (которые не встречаются в языках этих семей за пределами данной области). Вероятно, количество этих общих черт достаточно велико, и они должны быть весьма характерными.»)

Здесь уточняется число языковых семей, но также не оговаривается число разделяемых признаков. Более того, в этом определении требуется, чтобы число разделяемых признаков было ‘достаточным’ (fair) и признаки были существенными, что делает определение еще более нечетким. Не ставится также вопроса, должны ли признаки разделяться всеми языками ЯС.

Таким образом, большинство определений ЯС не уточняют, сколько признаков должно быть общими и должны ли они разделяться всеми языками. В ряде работ предлагается идея сравнения ЯС по числу разделяемых признаков.

2. Количественная мера степени конвергенции ЯС

Типологическая (грамматическая, структурная) близость языков может быть строго определена численно как расстояние между языками в пространстве признаков. Расстояние может определяться различными способами, наиболее простой и чаще используемой является метрика Хэмминга [19] – число признаков, по которым объекты (языки) различаются. Или в такой форме, или в слегка модифицированном виде она широко используется в квантитативной лингвистике [7, 11]. В настоящей работе для расчета расстояний между языками использовалась программа Similarity [7], разработанная В.Н. Поляковым.

Естественно, что за счет заимствованных друг у друга свойств в ЯС расстояние между языками должно быть меньше, чем у языков тех же языковых групп, которые не контактируют между собой. Это позволяет следующим образом определить степень близости языков в ЯС.

Допустим, что в исследуемом регионе присутствуют две группы языков G1 и G2 из двух семей (или ветвей семьи) S1 и S2 соответственно. Подсчитаем

среднее расстояние $\rho(G1, G2)$ между языками из групп $G1$ и $G2$ и среднее расстояние $\rho(S1, S2)$ между языками из $S1$ и $S2$. Разность $R = \rho(S1, S2) - \rho(G1, G2)$ назовем мерой схождения языков в регионе. Данное определение содержит некую формализацию идеи, заложенной в вышеприведенном определении из [9].

При $R = 0$ (и $R < 0$) нет оснований постулировать наличие в регионе ЯС. Чем величина R больше, тем ЯС сильнее. Аналогичные расчеты можно сделать и в случае языков из нескольких семей. Сразу отметим ограничение применения этого метода. Если в регионе присутствуют все языки соответствующих семей (то есть $G1 = S1$ и $G2 = S2$), то применение метода невозможно. Такая ситуация имеет место, например, на Северном Кавказе, где полностью находятся абхазо-адыгская и нахско-дагестанская семьи.

Вначале в порядке тестирования применим предложенный подход к классическому ЯС – балканскому и также к поволжскому (волго-камскому). В балканский ЯС традиционно [1] включают южнославянские и балкано-романские языки, а также албанский и новогреческий. К сожалению, в базе данных «Языки мира» отсутствуют албанский и новогреческий. Поэтому применим предложенное определение к двум оставшимся группам, принадлежащим славянской и романской ветвям индоевропейской семьи. К балкано-романским относятся арумунский, истророманский, истрорумынский, мегленорумынский, молдавский, румынский. К южнославянским относятся македонский, болгарский, сербохорватский.

Расчеты показывают, что среднее расстояние между балкано-романскими и южнославянскими языками равно 237, в то время как между романскими и славянскими в целом – 243. Как и следовало ожидать, расстояние между всеми романскими и всеми славянскими оказалось больше. Хотя разница $243 - 237 = 6$ кажется не столь и большой, все же она указывает на большую типологическую близость между балкано-романскими и южнославянскими языками, которая не может быть объяснена их родством, так как их генеалогическая близость точно такая же, как и у романских и славянских языков в целом.

Перейдем к Поволжью. В этом регионе проживают носители тюркских языков (алтайская семья: башкирский, татарский, чувашский) и финно-угорских языков (уральская семья: марийский, горномарийский литературный, эрзянский, мокшанский). В этом регионе присутствует также удмуртский язык, но его описания нет в нашей базе данных.

Среднее расстояние между языками этих двух групп языков Поволжья – 210, среднее же расстояние между тюркскими и финно-угорскими языками – 215. Здесь разница равна 5, что весьма точно соответствует величине, полученной выше для балканского ЯС. Таким образом, на основе применения предложенных формул к двум классическим ЯС можно предположить, что в ЯС языки типологически ближе друг к другу, чем языки из их генеалогических групп, в среднем на величину, примерно равную 5–6.

Теперь применим метод к спорному случаю. В [17, 20] к балканскому ЯС отнесен также турецкий язык, что признается не всеми. Среднее расстояние от турецкого до балкано-романских и южнославянских языков равно 275, а расстояние от турецкого до всех славянских и романских – 279. Таким образом,

разность составляет 4, что несколько меньше, чем в случае балканского ЯС без турецкого языка.

Подсчитаем отдельно близость турецкого языка к обеим основным группам балканского региона. Расстояние от турецкого до южнославянских языков равно 268, до славянских – 283. Разность – 15, что достаточно естественно, учитывая длительную оккупацию Турцией региона проживания южных славян. Хорошо известно, что болгарский и македонский языки содержат много тюркизмов. В то же время расстояние от турецкого до балкано-романских языков 278, а до всех романских – 277. (При применении наших формул получается отрицательная величина –1.) В этом случае числа практически совпадают. Хотя балкано-романские языки территориально находятся много ближе к Турции, чем другие подветви романской ветви, однако это не привело к их большей типологической близости. Возможно, это является следствием длительных (с XI века) культурных контактов предшественников современных турков – тюрков-османов и тюрков-сельджуков – с византийцами и итальянцами. Получается, что турецкий язык входит в балканский ЯС как бы наполовину.

3. Обсуждение и заключение

Во многих статьях отмечалось, что определение языкового союза является нечетким. В статье [5], в которой наиболее детально проанализированы различные подходы к определению, в итоге фактически сделан вывод о невозможности строгого удовлетворительного определения ЯС.

Хотелось бы обратить внимание вот на какой аспект этой проблемы. Когда авторы статей называют определение нечетким, то это является отрицательной характеристикой определения. Однако не следует забывать, что мы систематически пользуемся огромным количеством нечетких понятий, и вполне успешно. Например, нечеткими, очевидно, являются понятия ‘молодой’, ‘старый’. Любая попытка дать строгое определение этих терминов будет искусственной.

В математике уже несколько десятков лет успешно разрабатывается нечеткая логика и нечеткая теория множеств [21]. Причем в рамках этих теорий само понятие нечеткости является четким! Таким образом, при правильном подходе нет ничего плохого в нечеткости. Одним из важных свойств нечетких понятий является то, что многие существенные операции с ними являются уже строгими. Вероятно, важнейшей из таких операций следует признать сравнение. Так, хотя мы и не имеем четкой границы старости, ни у кого не вызовет сомнения, что 80-летний человек более старый, чем 70-летний.

В настоящей работе предложена количественная мера близости языков в ЯС. Она основана на мере типологической близости между языками, подсчитанной с использованием базы данных «Языки мира». Эта мера близости определяется абсолютно строго и позволяет выполнять операцию сравнения, устанавливая, какое языковое сообщество является в большей степени ЯС, а какое в меньшей степени.

Применение данного подхода показало, что балканский и поволжский регионы являются практически в равной степени ЯС. Кроме того, подтвержден пограничный статус турецкого языка по отношению к балканскому союзу. Он достаточно близок только к половине (славянской) языков балканского союза.

Таким образом, хотя при применении предлагаемого количественного подхода определение ЯС по-прежнему остается нечетким, но сама эта нечеткость теперь становится строгой, в частности позволяет проводить строго определенную операцию сравнения двух ЯС по степени сближения языков.

Итак, продемонстрировано еще одно возможное применение типологических баз данных в ареальной лингвистике.

Работа выполнена при поддержке РФФИ (проект № 10-06-00087-а).

Summary

V.D. Solovev. Language Union: Is Its Definition Fuzzy?

The paper analyses definitions of the “language union” concept. A quantitative approach is suggested for the assessment of linguistic convergence in language unions. The “World Languages” typological database is used for this purpose. Based on this database, typological distance between languages is defined, and a measure of the proximity of languages in a language union is introduced. The calculation results for the Balkan and the Volga-Region language unions are presented. It is shown that the suggested measure of proximity can be used in controversial matters such as the Turkish language belonging to the Balkan language union. It is proved that this measure of proximity changes the status of the language union definition, making it stricter.

Key words: language union, quantitative linguistics, typological databases, Balkan languages, contact area.

Литература

1. Лингвистический энциклопедический словарь / Ред. В.Н. Ярцева. – М.: Сов. энцикл., 1990. – 685 с.
2. *Thomason S.G., Kaufman T.* Language Contact, Creolization, and Genetic Linguistics. – Berkeley: Univ. California Press, 1992. – 411 p.
3. *Tosco M.* Is There An Ethiopian Language Area? // *Anthropol. Ling.* – 2000. – V. 42, No 3. – P. 329–365.
4. *Brend H., Kuteva T.* Convergence and Divergence in the Development of African Languages // *Aikhenvald A.Y., Dixon R.M.W.* (eds.) *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics.* – Oxford: Oxford Univ. Press, 2001. – P. 393–411.
5. *Campbell L.* Areal linguistics: a closer scrutiny // *Matras Y., McMahon A., Vincent N.* (eds.) *Linguistic Areas: Convergence in Historical and Typological Perspective.* – Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2006. – P. 1–31.
6. *Wälchli B.* The Circum-Baltic languages // *The Languages and Linguistics of Europe / Ed. by B. Kortmann, J. van der Auwera.* – Berlin; Boston: Mouton de Gruyter, 2011. – P. 325–340.
7. *Поляков В.Н., Соловьев В.Д.* Компьютерные модели в типологии и компаративистике. – Казань: Казан. гос. ун-т, 2006. – 208 с.
8. *Polyakov V., Solovyev V., Wichmann S., Belyaev O.* Using WALS and Jazyki Mira // *Linguist. Typol.* – 2009. – V. 13, No 1. – P. 135–165.
9. *Emeneau M.B.* India as a Linguistic Area // *Language.* – 1956. – V. 32, No 1. – P. 3–16.
10. *Sherzer J.* Areal linguistics in North America // *Curr. Trends Linguist.* – 1973. – V. 10. – P. 749–795.

11. *Comrie B.* Areal typology of mainland Southeast Asia: what we learn from the WALS maps // Kullavanijaya P. (ed.) Trends in Thai Linguistics (=Manusya, Special Issue 13). – Bangkok: Chulalongkorn Univ., 2007. – P. 18–47.
12. *Thomason S.G.* Language Contact: An Introduction. – Edinburgh: Edinburgh Univ. Press, 2001. – 324 p.
13. *Dahl O.* Principles of Areal Typology // Haspelmath M., König E., Oesterreicher W., Raible W. (eds.) Language Typology and Language Universals: An International Handbook, V. 2. – Berlin: Mouton de Gruyter, 2001. – P. 1456–1470.
14. *Kortjevskaja-Tamm M.* The Circum-Baltic Languages – a Contact-Superposition Zone in the European Periphery // Ramat P., Stolz T. (eds.) Mediterranean Languages: Papers from the MEDTYP Workshop, Tirrenia, June 2000. – Bochum: Brockmeyer, 2002. – P. 209–222.
15. *Katz H.* Generative Phonologie und phonologische Sprachbünde des Ostjakischen und Samojedischen. – München: Fink, 1975. – 159 S.
16. *Schaller H.W.* Die Balkansprachen: Eine Einführung in die Balkanphilologie. – Heidelberg: Winter, 1975. – 312 S.
17. *Masica C.* Defining a Linguistic Area: South Asia. – Chicago: Univ. Chicago Press, 1976. – 389 p.
18. *Aikhenvald A., Dixon R.M.W.* Introduction // Aikhenvald A., Dixon R.M.W. (eds.) Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics. – Oxford: Oxford Univ. Press, 2001. – P. 1–29.
19. *Хэмминг P.B.* Теория кодирования и теория информации. – М.: Радио и связь, 1983. – 176 с.
20. Атлас языков мира / Ред. Б. Комри, С. Мэттьюс, М. Полински. – Лик Пресс, 1998. – 224 с.
21. *Заде Л.* Понятие лингвистической переменной и его применение к принятию приближенных решений. – М.: Мир, 1976. – 165 с.

Поступила в редакцию
06.04.11

Соловьев Валерий Дмитриевич – доктор физико-математических наук, профессор кафедры теоретической кибернетики Казанского (Приволжского) федерального университета.

E-mail: maki.solovyev@mail.ru