



УДК 81

А. А. Шумков, И. М. Теплыгина

## ОБ ИЗОМОРФНОСТИ ГЕНЕТИЧЕСКОГО И ЛИНГВИСТИЧЕСКОГО КОДОВ

*Проводится обзор современных научных исследований в сфере изоморфности генетического и лингвистического кодов. Особый интерес представляет рассмотрение этой проблемы средствами как сопоставительной лингвистики, так и молекулярной биологии. В рамках идеи двухчастности и с учетом традиционных взглядов на устройство физического мира выдвигается возможное объяснение сущности изоморфизма генетического и лингвистического кодов.*

**Лингвистический код, генетический код, язык, лингвистика, генетика, диалекты, география**

Язык является постоянно изменяющимся и развивающимся живым организмом. Взгляд на язык как на живой организм, подчиняющийся естественным законам природы, восходит к глубокой древности. Эту идею можно встретить в диалоге Платона «Кратил», в сочинениях Гераклита, а также в трактате Марка Теренция Варрона «О латинском языке» (кн. II и VII). Также она отразилась в знаменитой поэме Тита Лукреция Кара (I в. до н. э.) «О природе вещей»: «По побуждению природы язык стал различные звуки произносить, при нужде выражая названия предметов» [1].

Многие ученые высказывают мысль о подобии лингвистического и генетического кодов. В настоящей статье мы намерены осветить современные исследования в данной сфере и работы лингвистов, ставшие классическими.

Генетика, как более молодая наука, заимствовала большую часть терминологии и способы классификации у лингвистики. В работах известного итальянского ученого Кавалли-Сфорца показана явная конгруэнтность между фенограммой (диаграммой), отображающей изменения частот аллелей (генов) у различных народов, и древом языковых семей. Кавалли-Сфорца и его коллеги предполагают, что эта конгруэнтность показывает «значительный параллелизм между развитием языка и гена». Последующие исследования других ученых также выявили различную степень близости у лингвистических и генетических классификаций (Barbujani and Sokal 1990; Excoffier 1991; Barbujani and Pilastro 1993; Sajantila 1995; Poloni 1997; Dupanloup 2000; Rosser 2000) в зависимости от исследуе-

мого региона. А некоторые из этих лингвистов даже утверждают, что подобный параллелизм лингвистических изменений и изменений аллель-частот не случайность, а общая тенденция (Barbuĵani 1997). Также отмечается внутренняя взаимосвязь между генетикой и языком (Chen 1995) [2].

Некоторые исследователи обсуждают важность проявления корреляции между генетическими и лингвистическими классификациями. Из нескольких возможных объяснений самым весомым является следующее. Подобие генетического и лингвистического кодов отражает одни и те же события в истории народа. Например, переселение первооткрывателей, которые размножались биологически и лингвистически в новом месте, а затем сохраняли свою биологическую и языковую самоидентичность до настоящего времени. Такая сильная связь предполагается, например, в исследованиях освоения и заселения Америки. Эти исследования используют генетические сходства как «дополнительную поддержку... для первичного взаимодействия... основанного на лингвистических данных». Модели, показывающие демографическое расселение в эпоху неолита во взаимосвязи с распространением языков, основаны на той же предполагаемой связи лингвистического и генетического кодов [2].

Менее убедительно объясняет изоморфизм лингвистических и генетического кодов их развитие отдельно друг от друга, но под влиянием одних и тех же факторов. Наиболее очевидным из таких факторов является география. И генетический, и лингвистический коды передаются непосредственно через прямой контакт. И в прошлом такое взаимодействие зависело от географической близости и доступности. Географическая удаленность служит наибольшей предпосылкой генетической удаленности людей [2]. В языковых семьях наблюдается та же тенденция. Наиболее близкие языки чаще всего расположены близко к друг другу. Причем их родство ослабевает при увеличении физического расстояния между ними. Таким образом, изменения в генетическом и лингвистическом кодах подвергнуты влиянию реального расстояния.

Возникает вопрос: являются ли эти проявления подобия генетического и лингвистического кодов параллельными, но не зависящими друг от друга процессами, или же это – явный показатель их совместного взаимодействия и развития?

Для выяснения данного вопроса британскими специалистами по биологии человека Даниелью Нетл и Луизой Харрис было проведено исследование, в котором рассматривались жители 5 регионов (Европа, Западная Азия, Юго-Восточная Азия, Восточная и Центральная Азия, Западная Африка). Выбор регионов был основан на наличии нескольких хорошо изученных языковых семей и на четкой идентифицируемости географического месторасположения и языка каждого народа. Результаты исследования показали взаимозависимость языкового и генетического родства. Существенная связь прослеживается только в Европе и Восточной и Центральной Азии. Языковое родство связано с географическим расстоянием во всех регионах кроме Западной Африки, но данный результат для этого региона объясняется малым количеством этнических групп. Для нас важно само наличие такой связи [2].

В работе «Теория Дарвина в применении к науке о языке» (1863) есть такие утверждения: «Законы, установленные Дарвином для видов животных и растений, применимы, по крайней мере в главных чертах своих, и к организмам языка... Что же касается уста-

новленной Дарвином изменчивости видов, которая, если только она не однородна и не равномерна у всех особей, содействует возникновению из одной формы многих новых, то в отношении к организмам языка эта способность уже давно доказана» [1]. Два французских ученых, специалисты по молекулярной биологии Жан-Жак Купек и Пьер Саниго, предлагают применить теорию Дарвина к их дисциплине. В книге «Ни Бога, ни гена», опубликованной недавно, они объясняют, что сообщество клеток организма определяется не генетической программой, а конкуренцией, которая происходит между различными составляющими живого организма в целях получения внешних ресурсов, без которых он не может жить. Молекулы возникают произвольным образом и принцип «естественного отбора» поощряет жизнеспособные сочетания [3], [11]. Здесь можно провести параллели с развитием и изменением языка, например через заимствование, а также появление имен для названия новых реалий. И закрепление в языке тех из них, которые обладают наибольшей популярностью среди носителей языка.

Результаты исследования показывают, что для Европы и, в меньшей степени, для Восточной и Центральной Азии существует очевидная связь между языковой принадлежностью и генетическим родством. При этом учитывается географическая близость. Четкость модели в Европе подтверждает заявления предыдущих исследований этого региона (Sokal 1989; Barbujani and Sokal 1990; Sajantila 1995). С давних пор идет обсуждение истории заселения Европы. Что заселение происходило путем миграции с Востока последние 10 тыс. лет, связанной с распространением сельского хозяйства и, согласно гипотезе, индоевропейской группы языков вместе с ним (Ammerman and Cavalli-Sforza 1984; Renfrew 1987; Renfrew 1991; Barbujani and Bertorelle 2001). Результаты настоящего исследования подтверждают этот сценарий развития, а также предполагают, что последующие местные примеси были недостаточны, чтобы полностью стереть следы этого процесса [2, с. 341].

В данной связи будет интересно упомянуть об одной этнической группе, которая столетиями мигрировала на большие расстояния и взаимодействовала с большим количеством других этнических групп.

До недавнего времени большинство открытий, связанных с происхождением цыган и их историей, были основаны на лингвистических исследованиях. Индийское происхождение цыганского языка было доказано несколько веков назад, модели миграции были воссозданы впоследствии на лингвистических данных. Исторические и археологические подтверждения практически отсутствовали. Относительно недавно генетика человека предоставила дополнительные доказательства этим историческим открытиям, особенно результаты исследований молекулы ДНК. Результаты лингвистических и генетических исследований оказываются сходными. Генетическое исследование доказывает индийское происхождение, последующий раскол и основание цыганских общин немногочисленными предками, зачастую они отчасти ассимилировались с другими этническими группами [4, с. 91].

И лингвистика, и генетика занимаются изменениями. Изменения, которые произошли в языковом сообществе, передаются последующему поколению. Таким образом, если изменения происходят в сообществе А, но не в сообществе Б, а сообщество А в свою очередь распадается на В и Г (как в случае с цыганами, например), то исследователи могут

говорить о том, что эти 3 сообщества (А, В, Г) имеют общую недавнюю историю. Другие сообщества, возможно, не подвергались тем же изменениям. Поэтому изучение изменений языка у нескольких сообществ одной этнической группы позволяет выявить их демографические изменения, передвижения и контакты [4, с. 94].

Генетика может воссоздать историю отдельных людей, семей, сообществ и этнических групп путем сравнения молекул ДНК, определить изменения и, возможно, датировать их. Генетические изменения могут быть унаследованы только прямыми потомками, тем самым позволяя исследователям проследить передвижения этнической группы и взаимодействие с представителями окружающих других этнических групп.

При исследовании эволюции этнической группы лингвистика и генетика имеют общие основания, мотивы, причины. Обе области исследования проливают свет на исторические взаимоотношения между языковыми сообществами: чужие генетические особенности и языковые особенности передаются при взаимодействии с окружающими языками, доказывая историческую взаимосвязь наследственности и непосредственного контакта. Генетические и языковые изменения учитывают относительную и абсолютную датировки. Относительная датировка – это возможность определить, используя определенные методы [4, с. 97], в каком порядке происходили изменения.

Взаимодействие генетики и лингвистики может пролить свет на проблемы диалектологии и лингвистики. С традиционной точки зрения, диалектам присуща дивергенция, т. е. влияние диалектов друг на друга после разделения может быть сведено к минимуму. Такого взгляда на диалекты цыганского языка придерживаются ученые Борецкий и Игла (2004). Матрас (2002, 2005) придерживается иной, более ареальной точки зрения на классификацию диалектов. Виктор Элсик (2006) критикует точку зрения Борецкого и Иглы, указывая на явное ареальное распространение диалектов в некоторых случаях. Он также говорит, что распространение диалектов во многом зависит от географического фактора. Различия между цыганскими диалектами часто являются результатом влияния соседствующих этнических групп. Языковые особенности могут передаваться от одного языкового сообщества другому только при непосредственном контакте их представителей. Возможно, они контактируют с целью торговли или имеют более тесные связи, например заключение брака между представителями двух языковых сообществ. Возможно, и мужчины, и женщины вступают в брак с представителями другого сообщества, а возможно, только женщины из одного сообщества выходят замуж за мужчин из другого. И именно генетика может ответить на вопрос, по какой причине одно языковое сообщество оказало влияние на другое, будь то деловые контакты или супружеские узы [4, с. 98].

Наследственность и изменчивость – два изначальных свойства жизни не только органического мира, но и языка [1]. Именно благодаря этому ученые могут проследить историю происхождения, развития, территориального перемещения языка и его носителей. И зачастую именно генетика дает понять, почему те или иные процессы происходили, и помогает составить более полную картину истории развития языкового сообщества. В связи с этим можно говорить о таком направлении в языкознании, как «генетическая лингвистика».

Говоря об историческом развитии языка, следует упомянуть работы Н. Я. Марра и его последователей в области «палеонтологии языка», которые основывались на выявлении так называемых первоэлементов. Н. Я. Марр сводит исторически возникшее многообразие языков именно к четырем исходным элементам, состоящим, как ни странно, из своеобразных звуковых «троек» – бессмысленных последовательностей – сал, бер, йон, рош. Любой текст произвольной длины на любом языке мира есть, в конечном счете, результат фонетического преобразования только этих исходных четырех, самих по себе ничего не значащих элементов, скомбинированных в определенной линейной последовательности. Этим, по мнению Н. Я. Марра, и определяется единство глоттогонического процесса. И хотя эта теория противоречит логике современной теоретической лингвистики, она представляет собой своеобразную структурную модель языка, весьма близкую к генетическому коду [1].

Проблема изоморфизма между генетическим и лингвистическим кодами долгое время была предметом пристального внимания Р. О. Якобсона. Он писал: «Можно утверждать, что среди всех систем, несущих информацию, генетический код и языковой код являются единственными семиотическими сущностями, основанными на использовании дискретных компонентов, которые сами по себе лишены какого-либо значения и служат для конституирования минимальной значимой единицы, т. е. единицы, наделенной своим собственным внутренним значением в пределах данного кода». Р. Якобсон подчеркивал, что одно и то же свойство может быть закодировано различными средствами, а различные свойства могут быть закодированы одним и тем же способом [5, с. 438].

Изоморфизм лингвистического и генетического кодов, на наш взгляд, действительно может заключаться в дискретности компонентов значимой единицы. Эта дискретность вызвана изначальной квантованностью микроуровня, приводящей, среди всего прочего, к принципиальной неопределенности пространственной и временной координат при точном определении импульса и энергии соответственно. Эта ситуация выражается математической записью принципа неопределенности Гейзенберга, в которой пространственная координата и импульс, а равно и временная координата и энергия, представлены произведениями их приращений. Важно отметить, что в уравнении Шредингера приращение пространственной координаты умножается, в конечном итоге, не на приращение импульса, а на приращение массы, что соответствует ньютоновскому представлению о кинетической энергии. На классическом уровне половина массы, умноженной на квадрат пространственной координаты, уравнивается с энергией, умноженной на квадрат временной координаты.

Рассмотренные физические соотношения, относящиеся как к микро-, так и к макроуровню природы, обнаруживают удивительное сходство с формально-логическими соотношениями, выявленными в ходе применения идеи двухчастности (ЛЭТИ, 1993) к моделированию синтаксических явлений естественных языков. Так, массу, умноженную на пространство, можно сопоставить с субстантивным семифинитивом, а энергию, умноженную на время, – с глагольным семифинитивом. На микроуровне, разумеется, эти произведения являются произведениями приращений, и вследствие дискретности своих компонентов семифинитивная единица приобретает лексическое значение. Вероятно, мозг человека воспринимает семифинитив и на квантовом, и на классическом уровне – микроуровень ответ-

ственен за генерацию значений, а макроуровень – за упорядочение их в смысловое целое. В таком рассмотрении совокупность глагольного и субстантивного семифинитивов предстает перед нами удивительной репликой какой-нибудь элементарной частицы, имеющей энергию и массу, например электрона. Уже на уровне элементарных частиц вследствие квантово-механической неопределенности можно ожидать генерации значения. При образовании атомов и далее – молекул, в том числе молекул ДНК, элементарные значения упорядочиваются в череду смыслов. Не исключено, что процесс упорядочения значений имеет отношение к процессу декогеренции открытой квантово-механической системы с преобразованием ее в макросистему. Не исключено также, что результатом декогеренции является не один акт редукции волновой функции, а несколько таких актов, отчего и возникают цепочки все более усложняющихся единиц, биологических и лингвистических: элементарная частица – атом – молекула – клетка – живой организм и член предложения – предложение – абзац – текст – язык. Может быть, в этой параллели язык действительно можно считать своеобразным живым организмом, умеющим воспринимать, обрабатывать, хранить и производить информацию? Ведь в случае сложных органических молекул мы действительно имеем дело с сегментами текстов, содержащими генетическую информацию, а в текстах на естественных языках ощущается некоторая «молекулярность».

Видный специалист по молекулярной биологии П. П. Гаряев сопоставляет особенности генетического и лингвистического кодов для решения проблемы «прочтения» РНК. Одна из фундаментальных характеристик генетического кода – его омонимия, утверждает Гаряев. Это второй вектор вырожденности триплетного кода. Синонимия (первый вектор вырожденности), т. е. кодовая однозначность набора различающихся кодонов, была обнаружена сразу. Она понятна и хорошо изучена в функциях изоакцепторных транспортных РНК. Омнимия – это кодовая неоднозначность первых двух одинаковых нуклеотидов в кодонах. Третьи нуклеотиды «воблируют» (могут быть любыми из четырех) и поэтому не участвуют в кодировании аминокислот. Иными словами, рибосома читает информационную РНК по правилу «два из трех». Это было отмечено в 1978 г. шведским исследователем У. Лагерквистом, но оставлено без внимания научной общественностью. Совершенно очевидно, что при использовании рибосомой такого правила возникает ситуация неоднозначности прочтения омонимичных кодонов. По этой причине существует опасность выбора неправильной аминокислоты или стоп-кодона. Это может привести к синтезу неправильных белков и гибели организма. Однако синтез белков – очень точный процесс. Почему рибосомы не ошибаются? Эксперименты свидетельствуют, что аппарат клетки, синтезирующий белок, использует лингвистический механизм контекстной ориентации при точном прочтении рибосомой омонимичных кодонов. Тогда возникает следующий вопрос: термин «чтение» рибосомой (в комплексе с транспортной РНК) информационной РНК – это метафора (как принято в генетике) или осмысленный процесс, реальное чтение и понимание? Придать омониму точный смысл можно, только поняв смысл полного текста (контекста). Значит, рибосомный аппарат клетки читает и понимает РНК не в метафорическом смысле? Однозначного ответа на этот вопрос на сегодняшний день не существует. Принять идею квазисознания, квазиразумности генома для биологов непросто. Гаряев детально анализирует этот теоретический (в сфере биологии), а по существу философ-

ский, тупик и определяет генетическую и биологическую роль омонимической вырожденности триплетного кода. Омонимия кода, по Гаряеву, есть фактор выведения работы рибосомного аппарата и клетки в целом на уровень квазиразумности и, следовательно, в иные, многомерные смысловые ареалы. Чтобы выйти из омонимического тупика, необходима простая, но ключевая идея: вновь обратиться к лингвистике и почерпнуть оттуда понятие контекста, которое снимает эту проблему. Омоним утрачивает неоднозначность только в контексте, т. е. роль части становится ясной, когда ее рассматривают в составе целого. В этом смысле понятие контекста (целостного текста) иРНК – отнюдь не метафорично. Фактически молекулярные биологи и генетики признают это, используя идею «второго генетического кода». Как ведет себя квазимыслящая рибосомная система при встрече с омонимичными ситуациями на иРНК? Можно пояснить это на простом примере. Скажем, в предложении надо выбрать одно из двух слов. Эти слова – аналоги омонимичных нуклеотидных дублетов в неопределенно кодирующих триплетах иРНК (иРНК выступает как аналог фразы или предложения). Эти слова –  $коX^1$  и  $коX^2$  с воблирующей третьей буквой  $X$ .  $X^1$ ,  $X^2$  могут произвольно принимать значения букв  $L$  и  $D$ . Составим предложение: «Не надо думать, что триплетный  $коX$  является некой мыслительной конструкцией-вампиром, которую надо уничтожить, вбивая ей в сердце осиновый  $коX$  волновой генетики». Ясно, что выбор из двух букв  $L$  и  $D$  и наделение омонимичного дублета  $ко$  точной семантикой слов  $код$  или  $кол$  зависит от целого предложения, от контекста, который выступает как смысловой фон, позволяющий выделить сигнал из шума неопределенности, т. е. выделить необходимое слово. Если не учитывать контекст – возможны смысловые нарушения. Но в действительности, даже если эти буквы будут поставлены неверно, вопреки контексту, то контекст (фон) дает такой избыток информации, что читающему ясно, как на самом деле трактовать омонимичное  $ко$  в целом предложении [3].

В лингвистике с такой проблемой часто сталкиваются при машинном переводе, поскольку, в отличие от человека, компьютеру нужно гораздо больше вспомогательных инструментов, чтобы различить значения при переводе многозначного слова. Например, значения слова «can». Помимо основного значения модального вспомогательного глагола у слова «can» имеется несколько официальных и жаргонных значений в качестве существительного: «банка», «отхожее место», «тюрьма». Кроме этого, существует архаичное значение этого слова – «знать» или «понимать». Для того чтобы правильно выбрать вариант перевода, необходимо знать, в каком контексте оно употребляется. Определенные успехи были достигнуты в сфере разработки программ перевода, различающих смысл с упором на контекст (тематические и семантические словари). Тем не менее, полностью автоматизированный машинный перевод текстов с обширной тематикой все еще является невыполнимой задачей. Все чаще высказывается необходимость организации систем перевода не на пословном принципе, а на принципе сопоставления словосочетаний [6]. На этом фоне идея двухчастности, оперирующая в качестве основных семантических единиц глагольным и субстантивным семифинитивами, выражающими в большинстве случаев словосочетания, представляется весьма интересной. Если основная часть семантического рельефа действительно генерируется в семифинитиве, по этим «средневзвешенным» состояниям можно будет производить более точный поиск смысловых соответствий. Тогда

появляется возможность охватить языки самого разного строя, от инкорпорирующих до флективных. Отождествление языка с флорой и фауной, берущее свое начало в натуралистическом направлении (А. Шлейхер), в череде эволюционно усложняющихся систем может иметь под собой определенные основания. Но это относится к макроуровню, рождающемуся в процессе декогеренции из микроуровня, где происходит генерация основных значений. Именно микроуровень ответственен в конечном итоге как за генетический, так и за лингвистический смыслы; без значения отдельного элемента система элементов не имела бы никакого смысла.

Таким образом, исследование природы, сущности, характера и причин изоморфизма генетических и лингвистических закономерностей является для языкознания нашего времени одной из наиболее важных проблем. К сожалению, эта проблема, поставленная Якобсоном в рамках семиотических категорий, до сих пор не исследовалась в более широком плане – в плане выявления сущности и механизма процессов, организующих языковую систему. Возможно, становление и языковых, и биологических систем начинается на квантово-механическом уровне, если именно этот уровень является исходным уровнем мироздания.

#### СПИСОК ЛИТЕРАТУРЫ

1. Маковский М. М. Лингвистическая генетика. Проблема онтогенеза слова в индоевропейских языках. М.: Изд-во ЛКИ, 2007.
2. Nettle D., Harriss L. Genetic and linguistic affinities between human populations in Eurasia and West Africa. Detroit, Michigan: Wayne State University Press, 2003.
3. Гаряев П. П. Лингвистико-волновой геном: теория и практика / Ин-т квантовой генетики. Киев, 2009.
4. Bakker P. Romani genetic linguistics and genetics: results, prospects and problems / Aarhus University. Aarhus, Denmark, 2012.
5. Jacobson R. Linguistics Main trends of research in the social and human sciences I. Paris: The Hague, 1970.
6. Белоногов Г. Г., Зеленков Ю. Г. Интерактивная система русско-английского и англо-русского машинного перевода / ВИНТИ. М., 1993.

A. A. Shumkov, I. M. Teplygina

#### GENETIC AND LINGUISTIC AFFINITIES

*This article reviews the modern studies of genetic and linguistic affinities. The issue is examined by means of both comparative linguistics and molecular Biology. We propose the possible interpretation of genetic and linguistic affinities considering the concept of twofoldness and traditional view on the physical world.*

**Linguistic code, genetic code, language, linguistics, genetics, dialects, geography**