

УДК 800 : 002

ББК 81.1

Т 49

В.Ш. Тлюстен

О некоторых параллелях процессов формирования и организации внутренней структуры искусственных и естественных языков

(Рецензирована)

Аннотация:

Статья посвящена анализу и обсуждению некоторых общих аспектов процесса формирования внутренней структуры и смыслового наполнения искусственных и естественных языков. В ней выдвигается *гипотеза историко-лингвистической рекреации* (HLR-гипотеза), предполагающая возможность наличия связи между формальной структурой естественных языков и истоками исторического процесса их зарождения.

Ключевые слова:

Язык, смысл, лингвистический объект, язык программирования, естественный язык, искусственный язык, сравнение языков, языковые примитивы, смысловые парадигмы, лингвистическое кодирование, носитель языка, исторический процесс.

В работе [1], посвящённой технологическим аспектам построения программных систем компьютерной обработки данных, вводится понятие *типичной программной ситуации*, основанное на следующем представлении о процессе создания нового языка программирования (ЯП).

Выбрав предметную область, т.е. планируемую сферу применения нового языка, разработчик, на основе детального анализа внутренней структуры этой предметной области и с учётом известного ему опыта программирования, определяет возникающие при создании программ данной сферы применения *типичные ситуации*, для решения которых необходимо предусмотреть введение особых, явно обозначаемых в создаваемом языке лингвистических средств.

Обобщая это представление, позволим себе предположить, что не только вновь создаваемый искусственный язык (в частности, ЯП), но и **любой естественный язык (ЕЯ), в начале его зарождения и в процессе формирования, в максимальной степени вбирает в себя структуру и особенности той предметной среды, на которую он (язык) изначально ориентирован.** В частности, в естественном языке, в момент его зарождения, прежде всего, возникают и лингвистически *наиболее экономным способом* оформляются смысловые схемы, отражающие исторические, социальные, психологические и иные особенности быта *наиболее первоосновные и характерные* для народа-носителя этого языка.

Указанная предметная ориентация как искусственных, так и естественных языков внешне проявляется в наличии в языке специальных лингвистических объектов, которые, являясь встроенными *языковыми примитивами*, непосредственно отражают типичные ситуации или, как мы будем говорить, *базовые смысловые парадигмы* (БСП), характерные для выбранной предметной области.

Например, языковыми примитивами для ЯП можно считать закреплённые во встроенных типах данных ато-

марные операции, а для естественных языков – имеющие самостоятельное значение и выражающие целостные понятия лексические основы, аффиксы, предлоги и т.п.

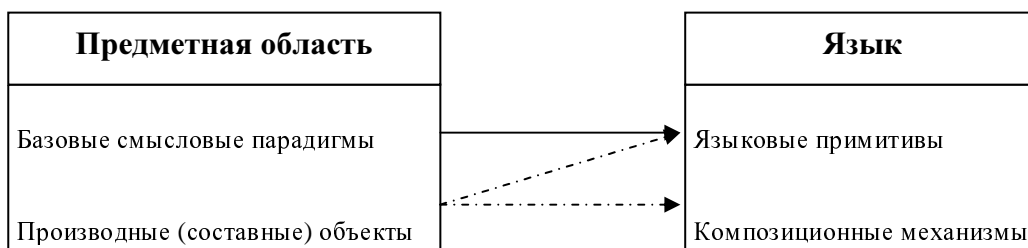
Помимо *базовых смысловых парадигм*, множество которых в структуре предметной области всегда конечно, в неё, как правило, оказывается включённым так же бесконечный класс *производных смысловых парадигм* (ПСП), получаемых за счёт комбинирования БСП.

Языковое выражение производных смысловых парадигм, осуществляется за счёт, собственно, «программирования» – привлечения *композиционных языковых механизмов*, позволяющих передать соответствующие указанным парадигмам смыслы не напрямую, а косвенно, выбором подходящих *сочетаний* базовых лингвистических примитивов.

Кроме того, если конструируемый, или исторически возникший язык является *универсальным* (таковы, кстати сказать, все естественные языки и большая часть применяемых на практике ЯП), теоретически должно быть возможным его использование в любой предметной области, а не только в той, для которой он изначально создан. Такая возможность так же реализуется за счёт композиционных языковых механизмов. В ЕЯ производным смысловым парадигмам соответствуют отдельные фразы и законченные предложения.

Композиционные механизмы, таким образом, необходимые в двух выше указанных случаях, расширяют сферу применимости языка и наряду с лингвистическими примитивами, должны быть включены во вновь создаваемый (или исторически возникающий) язык, претендующий хотя бы на какую-то степень универсальности.

Описанная выше зависимость структуры языка от структуры той предметной среды, которую данный язык представляет, может быть иллюстрирована следующей схемой:



Краткий вывод из выше изложенного состоит в следующем: структура предметной среды определяет структуру отражающего эту среду в человеческом сознании языка (безразлично, искусственного или естественного).

Но, по-видимому, до некоторой степени справедливо и обратное – формальная структуры языка может быть использована для воссоздания особенностей той предметной среды, которую данный язык отражает.

Последнее утверждение, применительно к естественным языкам, представляет особый интерес и может быть сформулировано в виде следующей *гипотезы историко-лингвистической рекреации* (HLR- гипотезы): **формальная структура любого естественного языка содержит неявную информацию об историческом контексте зарождения, традиционном укладе жизни, национальных особенностях и характере народа – носителя этого языка.**

Возможно, нелишне подчеркнуть, что в HLR- гипотезе речь идёт не о смысловой информации, явно содержащейся в печатных и рукописных текстах на том или ином языке (эта информация воспроизводится отдельными людьми и, преднамеренно или невольно, может быть искажена), а об информации скрытой, которая заключена в исторически сложившейся и потому, не имеющей прямого авторства формальной структуре (грамматическом строе) языка.

Как отмечено выше, универсальные языки, за счёт наличия в них композиционных механизмов, пригодны для выражения бесконечного множества смыслов, в том

числе, и тех, на которые эти языки не были изначально ориентированы. Именно этот факт, кстати сказать, является основой общепринятой практики использования ЕЯ в качестве метаязыков для изложения основ самых различных отраслей знаний.

То, насколько экономно, в информационном отношении, языки справляются с задачей лингвистического кодирования не только «своих», но и «чужих» смысловых схем, может быть использовано для построения шкал сравнительных оценок выразительной силы этих языков. Причём указанные оценки следует рассматривать не в абсолютном их значении, а по отношению к той или иной отдельно взятой смысловой парадигме, тематически обособленной совокупности таких парадигм, или целой предметной области.

Степень «чужеродности» для некоторого языка той или иной рассматриваемой смысловой парадигмы выделенной предметной области может быть выражена количественно, в зависимости от минимально требуемой длины реализующего эту парадигму в данном языке лингвистического кода.

В работе автора [2] приводится подробный анализ того, как это может быть сделано в случае рассмотрения языков программирования.

В качестве простейшего примера лингвистического выражения одной и той же смысловой парадигмы в четырёх различных ЯП, ниже приведена взятая из указанной работы и несколько видоизменённая таблица.

Таблица 1. Реализация смысловой парадигмы “Замена значения целочисленной переменной V на остаток от её деления на целочисленное значение выражения E” в языках программирования C++, PASCAL, ALGOL-68 и ALGOL-60.

№ п.п.	Язык программирования	Реализация (лингвистический код) ситуационной парадигмы	Число операций в реализации	Степень реализуемости
1	C++	$V \% = E$	1	1,00
2	ALGOL-68	$V \text{ modab } E$	1	1,00
3	PASCAL	$V := V \text{ mod } E$	2	0,98
4	ALGOL-60	$V := V - V \text{ div } E * E$	4	0,84

Таблица иллюстрирует то, каким образом с одной и той же смысловой парадигмой справляются различные, представленные в этой таблице ЯП.

Лингвистический код, наиболее экономным образом выражающий указанную смысловую парадигму в каждом из рассматриваемых ЯП, как видно из третьей колонки таблицы, заметно различается для этих языков по форме, по структуре и по длине.

Степень реализуемости данной смысловой парадигмы в том или ином ЯП, согласно нечёткому подходу

[3], принятому в выше указанной работе автора, зависит от минимально необходимого в реализации числа операций и определяется как значение специальным образом подобранной функции принадлежности, выбираемое из отрезка [0;1]. При этом значение 0 отражает факт полной невозможности реализации, а 1 – факт наличия реализации в языке в качестве встроенного лингвистического примитива.

Анализируя последнюю колонку таблицы, мы видим, что хуже всех (грубо говоря, – на 84%), с рассматри-

ваемой задачей справляется «прародитель» многих современных ЯП – язык ALGOL-60. Соответствующий лингвистический код обладает явной избыточностью (в частности, содержит повторяющиеся элементы) и передаёт необходимый смысл только за счёт определённого сочетания четырёх базовых операций.

Напротив, языки ALGOL-68 и C++ выражают рассматриваемую смысловую парадигму в форме БСП, т.е. непосредственно, за счёт использования специально предусмотренной для неё операции (“**modab**” или “%=", соответственно).

Язык PASCAL, как видно из той же таблицы, имея оценку 0.98, занимает в этом ряду промежуточное положение.

Более внимательное рассмотрение указанного примера показывает, что на самом деле, в нём представлено не четыре, а пять языков. Действительно, пятым является русский язык. Именно в этом *естественном языке* формулируется смысл обсуждаемой парадигмы: “замена значения целочисленной переменной V на остаток от её деления на целочисленное значение выражения E ”.

С функциональной точки зрения, текст выше приведённой формулировки играет роль *обозначения* (идентификатора) рассматриваемой парадигмы. Но, в то же время, указанную формулировку можно, очевидно, трактовать как лингвистический код (реализацию) соответствующей смысловой парадигмы *в русском языке*.

Подобная трактовка немедленно ставит вопрос о *степени реализуемости* рассматриваемой смысловой парадигмы в русском языке, т.е. о том месте, которое мог бы занять русский язык в выше приведённой таблице. Очевидная громоздкость выражающего обсуждаемую парадигму русскоязычного текста, позволяет сделать предположение о том, что указанная степень не может быть высокой. И действительно, не вдаваясь здесь в методику расчёта принятую в упомянутой работе [2], мы можем констатировать, что значение этой степени приблизительно равно величине 0.37.

Таким образом, русский язык в выше приведённой таблице с большим отрывом от языка PASCAL занимает самую последнюю позицию. Это является отражением бесспорного факта – естественные языки не удовлетворительны в качестве средства реализации базовых сущностей сферы автоматизированной обработки данных и в этом качестве существенно проигрывают специально оптимизированным для таких применений языкам программирования.

Но подойдём к вопросу о степени реализуемости (или, попросту говоря, о *степени экономности*) языкового выражения базовых сущностей предметной области с другой стороны. Что, если в качестве сравниваемых лингвистических объектов мы, вместо ЯП, будем теперь рассматривать естественные языки, а в качестве предметной области – некоторую *естественную* сферу их применения? Можем ли мы и в этом случае строить таблицы, подобные той, которая рассмотрена выше?

Интересуясь, прежде всего, наиболее универсальными аспектами общечеловеческой языковой практики, в качестве предметной области выберем, например, бытовые применения ЕЯ. Зафиксируем два естественных языка L_1 и L_2 , которые мы хотели бы в этой сфере сравнить

и зададимся некоторой, принадлежащей этой сфере смысловой парадигмой s .

Теперь, если $L_1(s)$ и $L_2(s)$ – реализации s в выбранных языках, а $\mu(L_1(s))$ и

$\mu(L_2(s))$ – соответствующие значения степени реализуемости, то, сравнивая эти значения, можно выяснить:

- какой из сравниваемых естественных языков лучше (с меньшей информационной избыточностью) выражает данную парадигму, и, если принять HLR – гипотезу,
- в какой мере парадигма s отвечает духу, историческим традициям, образу жизни и характеру народа – носителя того, или иного из двух сравниваемых языков.

Возьмём, например, в качестве сравниваемых языков L_1 и L_2 , русский и адыгейский языки, соответственно. Выберем в качестве s_0 смысл сформулированной в русском языке фразы $L_1(s_0) =$ “По моему заказу, мне принесли немного мёда”. Выражение смысла s_0 на адыгейском языке будет тогда иметь вид: $L_2(s_0) =$ “шъоу тIэкIу къыз-фязгъэхьыгъ”.

Оценку степени компактности выражения некоторого смысла в разных языках естественно увязывать с длиной записей соответствующих фраз. Но при этом, может оказаться необходимой нейтрализация фактора избыточности кодирования, который проявляется при имеющем место несоответствии используемого в языке алфавита фонетическому строю этого языка. Такое несоответствие наблюдается, в частности, при использовании кириллицы в адыгейском языке.

Проблему можно обойти, если длину текста выражать не числом составляющих его знаков, а числом кодируемых этим и знакам фонетических единиц.

В рассматриваемом нами примере, длина $L_1(s_0)$ при таком подходе равна 35, а *приведённая* (т.е. подсчитанная с учётом избыточности кодирования в кириллическом алфавите) длина $L_2(s_0) = 17$.

Из этого следует, что смысловая парадигма s_0 в адыгейском языке кодируется в два раза компактней, чем в русском!

Конечно, могут существовать и обратные примеры. Но, тем не менее, этот интересный и несколько неожиданный факт порождает целый ряд требующих дальнейшего рассмотрения вопросов. Вот некоторые из них:

1. Насколько наблюдаемое преимущество Адыгейского языка над русским в компактности выражения смысловой парадигмы s_0 носит случайный характер и определяется только лишь выбором этой парадигмы?

2. За счёт каких лингвистических различий адыгейского и русского языков достигается это преимущество?

3. Не следует ли на основе HLR-гипотезы и с учётом выше упомянутых различий считать, что парадигма s_0 или отдельные её элементы характеризуют какие-то исторически обусловленные различия славянского и адыгского этносов?

4. Можно ли на основе некоторого статистически репрезентативного набора смысловых парадигм отражающих, например, наиболее универсальные общечело-

веческие категории, ранжировать ЕЯ по их «уровням», подобно тому, как по уровням ранжируются ЯП?

5. Как идентифицировать вышеуказанный набор в некотором смысле эталонных для выбранной предметной области парадигм?

Исследования в направлении поиска ответов на эти и другие подобные вопросы поможет, как нам кажется, не только лучше понять глубинную структуру ЕЯ, но и выявить неясные связи этой структуры с истоками зарождения языковой практики народов-носителей этих ЕЯ.

Примечания:

1. А.Л. Фуксман. Технологические аспекты создания программных систем / М. «Статистика», 1979 г., 184 с.
2. В.Ш. Глюстен. Функционально ориентированная система аттестации языков программирования и её использование для решения прикладных задач / Киев, Препринт 6-8, Институт кибернетики АН УССР, 1987 г.
3. Л. Заде. Понятие лингвистической переменной и его применение к принятию приближённых решений / М., «Мир», 1976 г.