

Оригинальная статья
УДК 81–116
<http://doi.org/10.32603/2412-8562-2024-10-1-100-113>

Об одной лингвистической проблеме

Олег Маратович Поляков

*Санкт-Петербургский государственный университет
аэрокосмического приборостроения, Санкт-Петербург, Россия,
road.dust.spb@gmail.com, <https://orcid.org/0000-0002-8520-3936>*

Введение. Статья продолжает серию публикаций по R-лингвистике и посвящена тернарной гипотезе и связанной с ней проблеме моделирования мира. Суть тернарной гипотезы состоит в предположении, что любые естественные языки не используют глаголы арности (валентности) выше трех. Эта гипотеза порождает гносеологическую проблему, суть которой состоит в несоответствии ограничений размерности мышления и размерности некоторых явлений и процессов окружающего мира. Указанная проблема имеет прямое отношение к языку, поскольку язык отражает мышление и структуру модели мира человека.

Методология и источники. В качестве инструментов исследования используются результаты, полученные в предыдущих частях серии. Для разработки необходимых математических представлений для глаголов с высокой валентностью используются аксиомы категоризации, сформулированные ранее в рамках серии.

Результаты и обсуждение. Показано, что случаи множественности косвенных дополнений связаны с зависимостями данных в предложении. Разработаны методы категоризации для тернарных отношений и отношений с более высокой арностью. Показано, что и в этом случае все условные категоризации и условные глаголы могут быть сведены к одной категоризации.

Заключение. Сформулирована тернарная гипотеза и тернарная проблема. На основании исследования зависимостей данных показано, что, скорее всего, тернарная гипотеза выполняется для всех естественных языков. В ходе исследования методов категоризации отношений с арностью выше трех, показано, что и в этом случае лингвистические пространства проявляют стабильность независимо от места расположения категорий в предложении, и это создает возможность описания высокоразмерных взаимодействий с помощью придаточных предложений. Таким образом, появление придаточных предложений во всех языках мира связано с необходимостью описания высокоразмерных взаимодействий в модели мира человека.

Ключевые слова: R-лингвистика, категоризация, тернарные глаголы, фиктивная арность

Для цитирования: Поляков О. М. Об одной лингвистической проблеме // ДИСКУРС. 2024. Т. 10, № 1. С. 100–113. DOI: 10.32603/2412-8562-2024-10-1-100-113.

© Поляков О. М., 2024

Контент доступен по лицензии Creative Commons Attribution 4.0 License.

This work is licensed under a Creative Commons Attribution 4.0 License.



About One Linguistic Problem

Oleg M. Polyakov

*Saint Petersburg State University of Aerospace Instrumentation, St Petersburg, Russia,
road.dust.spb@gmail.com, <https://orcid.org/0000-0002-8520-3936>*

Introduction. The article continues a series of publications on R-linguistics and is dedicated to the ternary hypothesis and the problem of modeling the world associated with it. The essence of the ternary hypothesis is based on the assumption that any natural language does not use verbs with arity (valency) above three. This hypothesis creates an epistemological problem, the essence of which lies in the mismatch between the constraints of the dimensionality of thinking and the dimensionality of some phenomena and processes of the surrounding world. The indicated problem is directly related to language, since language reflects human thinking and the structure of the world model.

Methodology and sources. The results obtained in the previous parts of the series are used as research tools. To develop the necessary mathematical representations for verbs with high valence, the categorization axioms formulated earlier in the series are used.

Results and discussion. It has been shown that cases of multiple indirect objects are related to data dependencies in the sentence. Methods of categorization have been developed for ternary relations and relations with higher arity. It has been demonstrated that even in this case, all conditional categorizations and conditional verbs can be reduced to a single categorization.

Conclusion. The ternary hypothesis and the ternary problem are formulated. Based on the study of data dependencies, it is shown that the ternary hypothesis is most likely true for all natural languages. Based on the study of methods of categorizing relations with arity higher than three, it is demonstrated that in this case linguistic spaces show stability regardless of the location of categories in the sentence, which makes it possible to describe high-dimensional interactions using subordinate clauses. Thus, the appearance of subordinate clauses in all languages of the world is associated with the need to describe high-dimensional interactions in the world model human.

Keywords: R-linguistics, categorization, ternary verbs, fictitious arity

For citation: Polyakov, O.M. (2024), "About One Linguistic Problem", *DISCOURSE*, vol. 10, no. 1, pp. 100–113. DOI: 10.32603/2412-8562-2024-10-1-100-113 (Russia).

Введение. Цель статьи – обратить внимание лингвистов, аналитических философов, специалистов в области когнитивных наук и искусственного интеллекта (ИИ) на достаточно очевидный факт, имеющий важное значение для дальнейшего развития этих областей знания. Будем называть этот факт лингвистической проблемой, поскольку в наиболее очевидном виде он проявляется в лингвистике. Язык проявляет (отражает) работу сознания и структуру модели мира у человека, поэтому обсуждаемая здесь проблема автоматически становится проблемой когнитивных наук. Кроме того, современные текстовые системы искусственного интеллекта имеют модели, основанные на статистике применения языка, так что обсуждаемая проблема естественным образом перетекает и в эту область.

Методология и источники. В качестве методологии в данной статье используется методология и результаты R-лингвистики, подробно изложенные в рамках серии в этом журнале. Кроме того, здесь применен инструмент теории зависимостей реляционных баз данных.

Результаты и обсуждение. Сознание моделирует мир, наблюдаемый в виде различных отношений между значениями параметров. Мы не будем останавливаться на этом факте, поскольку он подробно исследован в аналитической философии (см., например, [1]). Сознание перерабатывает полученные данные в отношения между категориями. В языке это проявляется прежде всего в виде валентностей (арностей) у глаголов [2], которые, как правило, являются именами этих отношений. Язык описывает различные отношения, наблюдаемые в мире. Принципы преобразования различных наблюдаемых бинарных отношений в категории и глаголы подробно описана в [3]. Например, типовая форма предложения: подлежащее (категория 1) – сказуемое (имя отношения) – прямое дополнение (категория 2), характерна для бинарных глаголов, соединяющих две категории. Появление косвенного дополнения далеко не всегда говорит о повышении арности до трех. Тернарных глаголов в языке не так уж много. В этой статье арность глагола означает, что в его стандартном использовании любое уменьшение числа связанных с глаголом объектов вызовет вопрос о пропавшей категории, потому что семантика этого глагола (возможность понять его действие) требует некоторого минимального числа объектов. Скажем, глагол *дарить* является тернарным глаголом (кто дарит, кому дарит, что дарит). Если вы услышите предложение «*Я подарил Лене*», то неизбежно зададите вопрос: «*Что подарил?*». Вместе с тем в предложении «*Я подарил Лене конфеты на Новый год*» объект *Новый год* не является обязательным. Другими словами, есть обязательные дополнения, которые связаны с арностью глаголов, и необязательные (фиктивные), которые создают фиктивную арность. Читатель можете обратиться к искусственному интеллекту GPT4 или GPT3,5, чтобы убедиться, какая неразбериха творится в этом вопросе. Именно поэтому в дальнейшем мы будем использовать только термин «арность». Кстати, обращение к этим версиям ИИ очень продуктивно для лингвистических исследований, поскольку они владеют большинством языков планеты и учитывают гигантский объем текстов, написанных до 2021 г. (возможно, уже и более поздних). В дальнейшем обсудим фиктивное повышение арности, хотя иногда имеет место кажущееся уменьшение арности. Например, если вы увидите у Лены шикарные конфеты и на вопрос: «*Откуда у тебя такие конфеты?*», получите ответ «*Подарили*», то в этом случае кажущаяся арность глагола *дарить* будет равна нулю. Но в этом примере глагол обозначает процесс [4], а не действие объектов.

Суть проблемы. Суть проблемы заключается в следующем. Мир дает нам для познания отношения различной арности. В этих отношениях зачастую действуют некоторые зависимости данных (функциональные, многозначные, взаимные, зависимости соединения). Эти зависимости позволяют декомпозировать данные и уменьшать арность отношений без потери исходной информации. Однако зависимости данных в отношениях существуют далеко не всегда. Другими словами, зачастую мы не можем понизить размерность поступающей информации, не создав искажений в данных. Например, отношение *быть футбольной командой* является 11-арным и совсем не обязательно содержит какие-то зависимости данных. Если мы, несмотря на отсутствие зависимостей, все-таки попробуем разделить данные на части меньшей размерности, то это неизбежно приведет к возникновению ложных данных или «кентавров» [5]. Другими словами, мы перестанем адекватно отражать и моделировать мир.

Обычно мы полагаем верной гипотезу об адекватном отображении модели мира человека в языке. В противном случае язык вряд ли мог служить средством коммуникации и

обучения. Далее, в языке мы наблюдаем глаголы с максимальной арностью три (тернарные глаголы), которые отражают моделирование сознанием отношений арности до трех включительно. Возникает вопрос: если максимальная арность глаголов в языке не превышает трех, то что происходит, когда человек сталкивается с данными более высокой арности? Здесь возможно несколько вариантов.

Первый вариант. Человек не может адекватно познавать события мира, имеющие арность более трех. Другими словами, мы принципиально ограничены в познании мира, поскольку не способны сформулировать и передать информацию о таких взаимодействиях в силу отсутствия соответствующих глаголов в языке.

Второй вариант. Человек искусственно (без достаточных оснований) снижает размерность и, значит, наше сознание в целом химерично, т. е. в наших моделях мира постоянно присутствуют химеры (кентавры) и мы принципиально неадекватны.

Третий вариант. При превышении размерности исходных данных более трех мы переходим к какому-то другому способу моделирования высоко размерных событий, который отображается в языке иначе (без введения многоарных глаголов).

Все эти варианты имеет смысл обсуждать при условии, что справедлива следующая тернарная гипотеза: ни в одном языке мира не используются глаголы с арностью более трех. Такая формулировка этой гипотезы переносит акцент на модель мира в сознании человека, поскольку речь идет о любых языках.

Перейдем к рассмотрению тернарной гипотезы. Для начала обратимся к автору идеи о валентностях – С. Д. Кацнельсону. Он пишет: «Из сказанного вытекает, что в принципе валентность не превышает двух-трех мест и что эти места определяются фундаментальным различием событийных предикатов, непереходных и переходных» [2, с. 25]. Итак, автор теории валентности в целом был склонен принять тернарную гипотезу. Обратимся теперь к ИИ с его знанием сотен языков и гигантского количества текстов. Это достаточно корректное обращение, поскольку мы обращаемся к нему только как к гигантской поисковой системе, и не более. Предварительно надо заметить, что в целом арность и валентность глаголов он понимает одинаково, хотя иногда наблюдаются незначительные отличия. Кроме того, порой наблюдаются небольшие различия в ответах GPT3,5 и GPT4, по-видимому, связанные с объемом обучающих данных. Вот некоторые примеры из опыта общения по данному вопросу (текст и пунктуация ИИ).

Человек. *Существуют ли глаголы арности более трех?*

ИИ. *Да, в некоторых языках существуют глаголы арности более трех. Например, в языке навахо (один из языков индейцев Навахо в США) есть глаголы, которые могут выражать шесть или более аргументов, и русские глаголы иногда могут выражать четыре или пять аргументов, например, «он сказал мне, чтобы я сделал это для него». Однако в большинстве языков, в том числе в русском, глаголы обычно имеют один, два или три аргумента.*

Человек. *Приведи пример глагола языка племени навахо, у которого арность больше трех.*

ИИ. *Язык племени навахо не имеет глаголов с арностью больше трех.*

В этом диалоге пример высокой арности глаголов русского языка некорректный. Глаголы *сказать* и *сделать* в примере имеют арность два, поскольку семантическое осмысление предложения «Он сказал, чтобы я сделал это» не вызывает с необходимостью допол-

нительных вопросов. В отношении глагола с высокой арностью в языке племени Навахо ИИ, как мы видим, ушел в отказ. Теперь спросим в терминах валентности.

Человек. *Существуют ли глаголы валентности более трех?*

ИИ. *Да, существуют. Некоторые глаголы имеют валентность более трех, что означает, что они могут образовывать предложения с более чем тремя аргументами (субъект, дополнение и пр.). Например, глагол «передать» может иметь четыре аргумента: «Я передал книгу другу через Интернет». Глагол «передавать» может иметь еще более аргументов в зависимости от контекста.*

Здесь мы имеем дело с типичным примером ошибочной трактовки понятия валентности, которое будет рассмотрено далее. Глагол *передать* имеет валентность три (кто передает, кому передает, что передает). Скажем, предложение «*Студенты передают администрации*» неизбежно вызовет вопрос: «*Что передают?*». Без ответа на этот вопрос интерпретация предложения не может быть завершена. Вместе с тем мы без дополнительных вопросов закончим интерпретацию предложения «*Я передал книгу через Интернет*», если из контекста будет понятно значение отсутствующей необходимой составляющей – «*другу*». А вот предложение «*Я передал книгу другу*» будет понятно и без контекста.

Человек. *Приведи пример четырехарного глагола.*

ИИ. «Создавать» – глагол, имеющий четыре позиции и обозначающий процесс создания чего-либо. Другие примеры четырехвалентных глаголов: «получать», «принимать», «использовать», «воспроизводить».

Предложение «*Я создаю шедевр*» не вызывает дополнительных вопросов и имеет законченную семантику, так что его валентность не превышает двух. Анализ других глаголов оставляем читателю.

Эту переписку можно продолжить, но фактически ИИ так и не смог привести реальные примеры глаголов с арностью выше трех. Читатель и сам может продолжить подобные опыты. Разумеется, ответы этой самой большой и интеллектуальной в истории человечества поисковой системы не могут гарантировать справедливость тернарной гипотезы, но они повышают ее достоверность.

О фиктивных арностях. Прежде, чем перейти к обсуждению тернарной проблемы, рассмотрим причины появления повышенной арности глаголов, которая явно прослеживается в ошибочных примерах ИИ. В принципе безопасное введение в предложение новых дополнений должно означать, что их появление не приводит к изменению объема других дополнений. В противном случае их появление может сказаться на арности глагола. Например, в предложении «*Х дарит у (цветы) z*» конкретное значение *у* может существенно влиять на *z*. Скажем, если в качестве *у* выступает девушка *Лиза* (*у = Лиза*), то перечень возможных подарков может существенно отличаться от *у = Даша*, потому что *Лиза*, например, терпеть не может лилии и гвоздики, но зато обожает кремовые розы, в отличие от *Даши*. Аналогично значение *z* зависит от дарителя *х*, хотя бы в силу его финансовых возможностей. Все это будет означать, что при использовании глагола *дарить* семантически законченное восприятие предложения с этим глаголом потребует указания всех трех переменных, поскольку их значения взаимно определяют объемы друг друга.

Итак, появление в предложении нового дополнения, разумеется, затронет объем подлежащего (он уменьшится из-за появления нового ограничения), но не должно затрагивать

объемы уже присутствующих дополнений. В этом случае арность глагола не изменится, поскольку семантическая интерпретация предложения не будет зависеть от нового дополнения (исключая подлежащее). Другими словами, если мы удалим новое дополнение, то объем оставшихся дополнений не изменится, хотя увеличится объем подлежащего, который, впрочем, вполне будет определяться оставшимися дополнениями, что не приведет к нарушению семантической интерпретации. Это и будет означать неизменность арности глагола. Описанный случай показывает, что объем вводимого дополнения не зависит от объема существующего дополнения. Этот случай для бинарных отношений (бинарных глаголов) математически соответствует многозначной зависимости (МЗ) $X \rightarrow Y$, которая означает, что в рамках данного отношения (глагола) значения Y (дополнения) зависят только от значений X (подлежащего) и не зависят от значения других составляющих. Рассмотрение бинарного случая позволит разобраться с вопросом о валентностях, но бинарный выбор совсем не означает, что в тернарном случае (взаимная зависимость – раздел «Кентавры» в [5]) что-либо меняется, кроме усложнения доказательств.

Итак, нам необходимо разобраться, в каких случаях можно безболезненно для интерпретации вводить и удалять дополнения. Напомним содержание понятия многозначной зависимости [6]. Формально определение многозначной зависимости выглядит следующим образом. Пусть задано отношение R на атрибутах (параметрах) X, Y, Z . Здесь X, Y и Z могут представлять собой некоторые множества атрибутов. Полагаем, что $Y \cap Z = \emptyset$. Многозначная зависимость $X \rightarrow Y$ в отношении $R(XYZ)$ выполняется тогда и только тогда, когда R может быть в точности восстановлено при соединении проекций отношения R по общим атрибутам X : $R = R[XY] \bowtie R[XZ]$. Здесь $R[XY]$ и $R[XZ]$ – проекции отношения R на множества атрибутов XY и XZ соответственно, а \bowtie – операция соединения проекций. Это определение можно сформулировать в другом виде. В отношении $R(XYZ)$ выполняется многозначная зависимость $X \rightarrow Y$ тогда и только тогда, когда из $(x, y, z) \in R$, $(x, y', z') \in R$ следует $(x, y, z') \in R$. Во втором определении обращает на себя внимание независимое сочетание значений атрибутов Y и Z при выполнении многозначной зависимости. Кроме того, из определения видно, что поведение значений атрибутов Y и Z симметрично. Таким образом, если $X \rightarrow Y$, то $X \rightarrow Z$ и наоборот (правило дополнения МЗ).

Под состоянием отношения R обычно понимают некоторое конкретное заполнение его кортежами. В одном состоянии R может выполняться МЗ $X \rightarrow Y$, а в другом – нет. Поэтому говорят, что в R выполняется МЗ $X \rightarrow Y$, если она выполняется в любом состоянии отношения R . Это можно сформулировать иначе. В отношении выполняется многозначная зависимость, если это определяется не конкретным составом накопленных данных, а природой вещей.

В качестве примера рассмотрим авторов и их книги. На любой конкретный момент времени каждая книга определяет написавших ее авторов, список которых зависит только от названия книги. Этот список не зависит ни от наличия этих книг в различных библиотеках, ни от числа экземпляров, ни от списков их читателей. Книги многозначно определяют своих авторов. Это связано с природой вещей, т. е. с тем, как происходит написание книг в окружающем мире. Короче говоря, многозначная зависимость – это факт существования в мире определенных устойчивых связей, которые не зависят от других факторов и позволяют выделять и рассматривать эти связи отдельно (позволяют декомпозировать) от других факторов.

По аналогии с МЗ определим взаимную зависимость (ВЗ) [7] как частный случай зависимости соединения [8], позволяющий без потерь информации декомпозировать отношение на три части. Взаимная зависимость $X \sim Y/Z$ в отношении $R(XYZ)$ выполняется тогда и только тогда, когда R может быть в точности восстановлено при соединении проекций отношения R по общим атрибутам X : $R = R[XY] \bowtie R[XZ] \bowtie R[YZ]$. Это определение можно сформулировать в другом виде. В отношении $R(XYZ)$ выполняется взаимная зависимость $X \sim Y/Z$ тогда и только тогда, когда из $(x, y, z) \in R$, $(x, y', z') \in R$ и $(x', y, z') \in R$ следует $(x, y, z') \in R$.

Итак, пусть в некотором отношении $R(XYZT)$ выполняется МЗ $X \rightarrow Y$. Другими словами, МЗ $X \rightarrow Y$ выполняется в обрамлении дополнительных атрибутов ZT . Что произойдет с МЗ, если удалить часть атрибутов из обрамления?

Предложение 1. Пусть в отношении $R(XYZT)$ выполняется МЗ $X \rightarrow Y$. И пусть $R' = R[XYZ]$ – проекция отношения R на атрибуты XYZ . Тогда в R' справедлива МЗ $X \rightarrow Y$.

Доказательство. Итак, пусть $(x, y, z) \in R'$ и $(x, y', z') \in R'$, где x, y, z – значения атрибутов X, Y, Z соответственно. Поскольку R' – проекция отношения R на XYZ , найдутся такие значения атрибута T t и t' , что $(x, y, z, t) \in R$ и $(x, y', z', t') \in R$. Поскольку в R выполняется МЗ $X \rightarrow Y$, то $(x, y, z', t') \in R$, а значит $(x, y, z') \in R'$ и в R' также выполняется МЗ $X \rightarrow Y$.

Из предложения 1 следует, что удаление из отношений некоторых атрибутов, не участвующих в МЗ, не приводит к нарушению этой зависимости. В нашем примере с книгами и авторами это означает следующее. Если в некоторой базе хранятся данные об авторах, книгах, читателях, наличии книг в библиотеках и прочем, то удаление части этих данных, не затрагивающих зависимость книг и авторов, не приводит к нарушениям этой зависимости. Если речь идет о языке и между подлежащим и прямым дополнением действует МЗ, то удаление любого косвенного дополнения в таком предложении не повредит эту связь.

Займемся теперь обратным процессом. Что будет происходить с МЗ, если наращивать дополнительные атрибуты в отношении? Здесь ситуация выглядит несколько сложнее. Произвольное добавление атрибутов может разрушить МЗ, создав зависимость Y от добавленных атрибутов.

В табл. 1 представлено отношение, в котором выполняется МЗ $A \rightarrow B$. В табл. 2 к этому отношению добавлены значения нового атрибута D , которые разрушили МЗ $A \rightarrow B$. В самом деле, при выполнении МЗ $A \rightarrow B$ в новом отношении согласно определению, при выборе любого значения атрибута A значения атрибутов B и CD должны произвольно сочетаться (в силу их независимости друг от друга). Возьмем, например, значение a . При этом значении атрибуты B и CD сочетаются произвольно. Хотя, например, b и b' одинаково сочетаются с cd , но b' сочетается с $c'd$, а b – нет. Итак, необходимо определить условия корректного добавления атрибутов.

Табл. 1. Тернарное отношение с МЗ
Table 1. Ternary relation with the MV

A	B	C
a	b	c
a	b'	c
a	b	c'
a	b'	c'
a'	b'	c'

Табл. 2. Расширенное тернарное отношение
Table 2. Extended ternary relation

A	B	C	D
a	b	c	d
a	b'	c	d
a	b	c'	d'
a	b'	c'	d
a'	b'	c'	d'

Предложение 2. Пусть в отношении R' на атрибутах XYZ выполняется $M3\ X \rightarrow Y$. И пусть к отношению R' добавляется атрибут T , и кортежи отношения R' дополняются значениями атрибута T для получения отношения R . Если это добавление осуществляется так, что в R выполняется $M3\ X \rightarrow T$, то в R также выполняется и $M3\ X \rightarrow Y$.

Доказательство. Если $C = \emptyset$, то справедливость предложения 2 следует из правила дополнения. Пусть $(x, y, z, t) \in R$ и $(x, y', z', t') \in R$. Для того чтобы в R выполнялась $M3\ X \rightarrow Y$, надо показать, что $(x, y, z', t') \in R$. Поскольку $(x, y, z) \in R'$, $(x, y', z') \in R'$ и в R' выполняется $M3\ X \rightarrow Y$, то $(x, y, z') \in R'$, а значит существует t'' такое, что $(x, y, z', t'') \in R$. Поскольку $(x, y, z', t'') \in R$, $(x, y', z', t') \in R$ и в R выполняется $M3\ X \rightarrow T$, то $(x, y, z', t') \in R$, а значит в R выполняется $M3\ X \rightarrow Y$.

Предложение 2 описывает условия, когда в предложение с бинарным глаголом можно безопасно добавлять новые дополнения. Именно если новое дополнение связано зависимостью только с подлежащим и его значения не зависят от существующих дополнений (как, впрочем, и наоборот), то такое дополнение может быть добавлено в предложение без изменения уже существующей информации.

В тех случаях, когда между объектами действует взаимная зависимость, ее действие можно свести к рассмотренному выше случаю действия $M3$, если иметь в виду следующее. Взаимная зависимость подобна многозначной, только в ней есть конъюнктивно добавленное условие. Если $(x, y, z) \in R$, $(x, y', z') \in R$, то $(x, y, z') \in R$, если $(y, z') \in R[YZ]$. Условие $(y, z') \in R[YZ]$ фактически является защитой от появления химер, вызванных слишком смелым предположением о независимости атрибутов Y и Z . Это последнее защитное условие может быть выражено в предложении любым приемлемым способом: *мужчина подарил женщине цветы, которые ей когда-то уже дарил отец*. На этом примере мы и ограничим рассмотрение зависимостей соединения.

В качестве примера для $M3$ рассмотрим тернарное отношение между авторами (A), книгами (K) и читателями ($Ч$). Как уже отмечалось, атрибут *книга* многозначно определяет своих авторов и читателей. Связь книги с автором не зависит от читателей, так что в отношении на атрибутах $АКЧ$ действует $M3\ K \rightarrow A$ и по правилу дополнения также $M3\ K \rightarrow Ч$. Глагол *писать* бинарный, так что предложение «*Книга N написана Иваном Васильевичем*» не вызывает проблем с интерпретацией. Точно также для второй $M3$ предложение «*Книга N написана для детей*» тоже не вызывает проблем с интерпретацией. Если же мы захотим сказать сразу обо всех трех объектах, то в предложении «*Книга N написана Иваном Васильевичем для детей*» арность глагола останется бинарной, но дополнение «для детей» будет вписано по правилу предложения 2 на том основании, что состав авторов книги не зависит от списка ее читателей. Если мы используем возвратную форму предложения «*Книга N написана Иваном Васильевичем*», то получим: «*Иван Васильевич написал книгу N*». К этому семантически тождественному предложению по-прежнему можно добавить дополнение и получить: «*Иван Васильевич написал книгу N для детей*». Другими словами, источник зависимостей в предложении со многими дополнениями необязательно должен занимать место подлежащего, он вполне может быть и дополнением.

В качестве второго примера рассмотрим предложение с бинарным глаголом *быть*: «*Девочка была мальчика*». Допустим, мы хотим добавить в это предложение косвенное дополнение,

связанное с орудием избиения. Что это может быть? Да, что угодно. Это может быть ладонь, кулак, веник и т. д. Все зависит от девочки, о которой идет речь в предложении: одни девочки дерутся кулаками, а другие могут использовать подручные средства. Итак, мы получаем, например, новое предложение: «*Девочка билa мальчика палкой*». Появление палки сузило область значений для категории *девочка*, потому что далеко не все девочки используют в драке палку, но не изменило область значений жертвы. Уменьшенный объем категории *девочка* по-прежнему входит в эту категорию, так что новое дополнение не изменяет семантической корректности текста. Продолжая по той же схеме, мы получаем предложение: «*Девочка билa мальчика палкой по голове*» или даже «*На улице девочка билa мальчика палкой по голове*». И опять каждый раз область значений категории *девочка* сужается, поскольку редко встретишь столь кровавную девочку, которая прилюдно лупит ребенка палкой по голове. Точно также в примере с ИИ «*Я передал книгу другу через Интернет*» к тернарному глаголу *передавать* добавлено косвенное дополнение *через Интернет*. Здесь вполне могло стоять *почтой* или *через деканат*, или еще как-нибудь. Это зависит от решения и предпочтений передающего.

Итак, мнение о том, что появление множества косвенных дополнений связано с возможностями глаголов управлять дополнительными переменными не совсем верно, хотя, разумеется, важно, что такие глаголы должны иметь возможность сочетаться с такими дополнениями. Появление косвенных дополнений (кроме первого косвенного дополнения «истинно» тернарного глагола) связано с их взаимодействием с уже существующими дополнениями в предложении. Поэтому их количество определяется лингвистической искушенностью говорящего и может быть в значительной степени произвольным и непредсказуемым. В этой ситуации решение относить число таких дополнений к базовой характеристике глагола кажется неразумным. Принимая во внимание сказанное, необходимо признать, что тернарная гипотеза представляется весьма вероятной, если под арностью (валентностью) понимать минимальное число дополнений, необходимых для интерпретации (формирования семантики) предложения.

О высоких арностях. На рис. 1 изображено тернарное отношение *R дарить*.

Кто (1)	Кому (2)	Что (3)
x ₁	y ₁	z ₁
x ₂	y ₁	z ₁
x ₁	y ₂	z ₁
x ₂	y ₂	z ₁
x ₁	y ₁	z ₂
x ₃	y ₁	z ₂

Рис. 1. Тернарное отношение

Fig. 1. Ternary relation

В качестве доменов (универсумов) для каждого атрибута отношения выберем проекцию $R[i]$ ($i = 1, 2, 3$) отношения на соответствующий атрибут. Отношение на рис. 1 можно рассматривать как бинарное, например, на универсумах $R[1] \times R[2,3]$. Пространство и ко-пространство для этого отношения изображено на рис. 2, а. Их построение и свойства описаны в [3]. Особенность заключается только в том, что в ко-пространстве категории состоят из бинарных отношений (на рисунке кортежи отношений перечислены через запятую). Это в свою очередь позволяет для каждой ко-категории построить свое пространство и ко-пространство. Результат этих построений изображен на рис. 2, б.

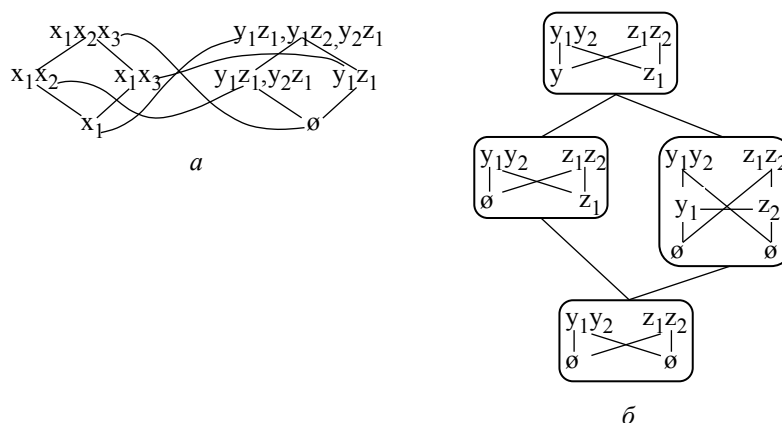


Рис. 2. Пространства для тернарного отношения
Fig. 2. Spaces for ternary relation

Подобную вложенную систему лингвистических пространств удобно представить в виде дерева (рис. 3). Здесь корень дерева соответствует пространству 1, первый уровень – пространству 2 и конечные вершины – пространству 3. Ветви дерева одного уровня слева направо соответствуют категориям соответствующего пространства, выбираемым слева направо и сверху вниз.

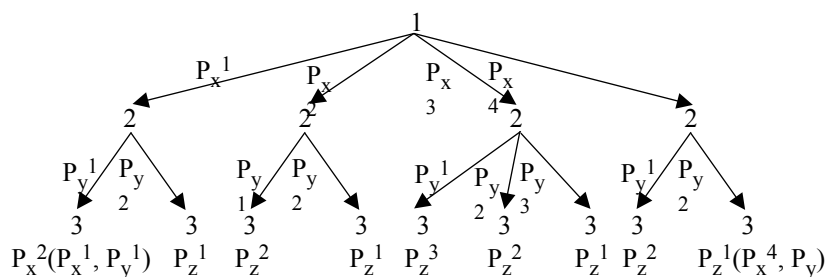


Рис. 3. Иерархическое представление отношения
Fig. 3. Hierarchical representation of the relation

На рис. 3 дуги дерева отмечены обозначениями категорий соответствующего пространства. Для примера: самый левый лист дерева кроме отметки номера категории имеет отметку пути от корня к листу. Например, запись $P_z^2(P_x^1, P_y^1)$ означает, что тернарный глагол соединяет три категории: *КТО дарит* – $P_x^1 = \{x_1, x_2, x_3\}$; *КОМУ дарит* – $P_y^1 = \{y_1, y_2\}$; *ЧТО дарит* – $P_z^2 = \emptyset$. Итак, каждый путь от корня дерева к концевой вершине соответствует некоторой фразе с глаголом *дарить* и тремя категориями, например, P_x^1, P_y^1 и $P_z^2 = \emptyset$. Поскольку последняя категория пуста, то эта фраза в целом имеет пустую семантику.

Определение. Дерево на рис. 3 будем называть иерархическим представлением (ИП) глагола (отношения S).

Легко видеть, что для отношения любой арности можно построить такое дерево, причем арность определит его высоту (число уровней). Кроме того, надо заметить, что в нашем примере первое пространство находилось в привилегированном положении. А именно, атрибут 1 был выбран первым в n -арном отношении, и ему было поставлено в соответствие $n-1$ -арное отношение. Впрочем, в качестве первого атрибута можно было бы выбрать любой другой, изменив порядок слов в соответствующем предложении.

Определение. Пространство атрибута, выбранного в качестве первого, назовем **безусловным** пространством этого атрибута. Обычно в языке ему соответствует подлежащее. Все остальные пространства называются условными. Условность каждого из условных пространств определяется категориями (условиями) в ИП, лежащими на пути от корня дерева до атрибута условного пространства.

Лемма 3. По ИП можно восстановить исходное n -арное отношение.

Доказательство. Как следует из [3], утверждение справедливо для $n = 2$. Пусть лемма справедлива для $n = k$ и пусть отношение S имеет арность $k + 1$. Рассмотрим некоторое ИП отношения S , у которого корень помечен номером i . Каждая дочерняя вершина корня ИП является корнем поддерева ИП k -арного отношения $S(P_i^j)$ для некоторой категории P_i^j пространства i . По нашему предположению, $S(P_i^j)$ совпадает с объединением произведений категорий, помечающих ветви поддерева от корня поддерева до листьев поддерева. В соответствии с [3] получаем:

$$\bigcup_{P_i^j} P_i^j \times S(P_i^j) = S.$$

Замечание. В ИП отношения S на дугах могут встречаться пустые категории соответствующих пространств. В этом случае удаление из ИП ветвей, содержащих такие категории, с учетом доказательства леммы 3 не влияет на восстанавливающую способность ИП. В связи в этом будем говорить, что два пространства почти равны, если они отличаются не более, чем на пустую категорию. Аналогично одно пространство почти принадлежит другому, если все непустые категории одного пространства входят в другое пространство.

Хотя лемма 3 дает механизм работы с n -арными отношениями без искажения информации, представляя их в виде последовательностей из n категорий (путь от корня ИП к листу), в то же время это приводит к возникновению проблемы. Появление локальных пространств означает зависимость пространств и их категорий от положения в ИП. Другими словами, изменение порядка слов изменяет категории и сами пространства. Решение этой проблемы дает теорема 4 (см. далее). Ее доказательство дано для тернарного случая по чисто техническим причинам: доказательства для произвольного n получаются на основе тернарного случая с помощью индукции. Это существенно усложняет и без того непростую систему обозначений и соответственно восприятие и объем такого доказательства, хотя реально помогает только заменить цифру 3 на букву n .

Теорема 4. Пусть S – тернарное отношение. Для каждого x первого атрибута определим бинарное отношение $S^x \subseteq S[2] \times [3]$ следующим образом: $(y, z) \in S^x$ тогда и только тогда, когда $(x, y, z) \in S$. Обозначим через \sqcup операцию сведения пространств, а сами пространства обозначим через \mathbb{P} с соответствующими индексами. Тогда $\mathbb{P}_3 \cong \bigsqcup_{x \in S[1]} \mathbb{P}_3^x$, где знак \cong обозначает отношение почти равенства, а \mathbb{P}_3^x – пространство по третьему атрибуту отношения S^x . Аналогично для атрибута 2.

Доказательство. Выберем некоторые x, y . Тогда $(x, y)^\Delta$ – категория безусловного пространства \mathbb{P}_3 . Но тогда для S^x имеем $y^\Delta = (x, y)^\Delta$ и $(x, y)^\Delta \in \mathbb{P}_3^x$, а значит

$(x, y)^\Delta \in \prod_{x \in S[1]} \mathbb{P}_3^x$. Таким образом, $\mathbb{P}_3 \subseteq \prod_{x \in S[1]} \mathbb{P}_3^x$. Обратно, пусть $Z \neq \emptyset$ и $Z \in \prod_{x \in S[1]} \mathbb{P}_3^x$. Тогда найдутся такие $x_i S[1]$, ($i = 1, \dots, k$), что для каждого i и отношения S^{x_i} в условном пространстве на $S[2]$ найдется категория Y^{x_i} , которой по условному глаголу соответствует ко-категория Z^{x_i} , такая что $Z = \bigcap_{x_i} Z^{x_i}$. Но $Z^{x_i} = \bigcap_{y \in Y_{x_i}} y^{\Delta x_i}$ по соответствующему условному глаголу. То есть $Z^{x_i} = \bigcap_{y \in Y_{x_i}} (x_i, y)^\Delta$, а значит $Z \in \mathbb{P}_3$.

Итак, сведение условных пространств с точностью до пустой категории равно безусловному пространству. Другими словами, непустые категории в условных пространствах остаются такими же, как и в безусловном пространстве. Например, для отношения на рис. 1 безусловное пространство для второго атрибута (*КОМУ*) показано на рис. 4, а. На рис. 4, б изображены условные пространства для этого атрибута. На рис. 4, в показано сведенное пространство для всех условных пространств, которое с точностью до пустой категории совпадает с безусловным пространством. Заметим, что после фиксации категории первого атрибута условные пространства второго и третьего атрибутов находятся в равном положении независимо от их перестановки, так что сказанное в полной мере относится и к третьему атрибуту (*ЧТО*).

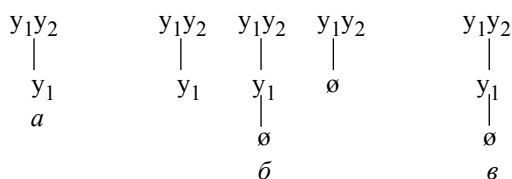


Рис. 4. Условные и безусловные пространства
Fig. 4. Conditional and unconditional spaces

Разумеется, если в качестве начального взять, допустим, третий атрибут, то условные пространства по второму атрибуту изменятся, но их сведение по-прежнему с точностью до пустой категории даст безусловное пространство. Итак, изменение порядка слов в предложениях, описывающих n -арные взаимодействия, не меняет объемы категорий, входящих в эти предложения. Без этого свойства отражение в языке таких взаимодействий было бы невозможно. Эти цепочки категорий передаются в языке с помощью придаточных предложений с глаголами, соответствующими проекциям исходного отношения, которые описываются данным придаточным предложением.

На описанном свойстве базируется присутствие в языке возвратной(!) формы тернарных глаголов [2]. Например, предложение: «Директор одолжил сотруднику деньги» (тернарный глагол *одалживать*) имеет возвратную форму: «Сотрудник одолжил у директора деньги», о которой сигнализируют изменения падежей. Понятно, что существование такой возвратной формы было бы невозможным без справедливости теоремы 4.

Закключение. Итак, наша модель мира должна адекватно отражать различные наблюдаемые взаимодействия разной размерности. Большой класс взаимодействий описывается соотношениями не более трех категорий, которым соответствуют глаголы арности, не превышающей трех. Это положение в статье получило название тернарной гипотезы. Много-

численные случаи присутствия в предложении одного и более косвенных дополнений не противоречат тернарной гипотезе, а связаны наличием различных зависимостей данных.

Отражение в языке неразложимых взаимодействий более трех, скорее всего, реализуется с помощью придаточных предложений (и некоторых других форм), которые позволяют разбить общую цепочку взаимосвязанных категорий на части. Это возможно лишь в том случае, если сами категории в цепочках являются стабильными, не зависящими от их положения в цепочке. Как было показано, это утверждение действительно справедливо, что на первый взгляд неочевидно. Если все это так, и наша модель мира, несмотря на тернарную гипотезу, действительно адекватно отражает многомерные взаимодействия, то наличие в любом языке придаточных предложений является совершенно необходимым. Другими словами, не может существовать язык, состоящий только из простых предложений. В заключение снова обратимся за помощью к ИИ GPT-4.

Человек. *Почему не существует языков без придаточных предложений?*

ИИ. *Не существует языков без придаточных предложений, потому что придаточные предложения необходимы для того, чтобы передать сложные мысли и отношения между идеями в рамках предложений. Некоторые языки могут иметь меньшее количество придаточных предложений, чем другие, но даже в таких языках они все же используются. Например, в языке китайских иероглифов можно представить сложные идеи через использование комплексных иероглифов и комбинаций слов, но все равно существует необходимость использования придаточных предложений при передаче более сложных связей и отношений между идеями в предложениях.*

СПИСОК ЛИТЕРАТУРЫ

1. Витгенштейн Л. Логико-философский трактат / пер. с нем. И. С. Добронравова, Д. Г. Лахути. М.: Канон+, 2017.
2. Кацнельсон С. Д. К понятию типов валентности // Вопросы языкознания. 1987. № 3. С. 20–32.
3. Поляков О. М. Лингвистическая модель данных для естественных языков и искусственного интеллекта. Часть 1. Категоризация // ДИСКУРС. 2019. Т. 5, № 4. С. 102–114. DOI: 10.32603/2412-8562-2019-5-4-102-114.
4. Поляков О. М. R-лингвистика. Принцип двойственности // ДИСКУРС. 2023. Т. 9, № 2. С. 103–110. DOI: 10.32603/2412-8562-2023-9-2-103-110.
5. Поляков О. М. Лингвистическая модель данных для естественных языков и искусственного интеллекта. Часть 2. Идентификация // ДИСКУРС. 2019. Т. 5, № 5. С. 98–111. DOI: 10.32603/2412-8562-2019-5-5-98-111.
6. Fagin R. Multivalued Dependencies and a New Normal Form for Relational Databases // ACM Transactions on Database Systems. 1977. Vol. 2, no. 3. P. 262–278.
7. Nicolas J.-M. First Order Logic Formalization for Functional, Multivalued and Mutual Dependencies // Proc. of the 1978 ACM SIGMOD Intern. Conf. on Management of Data, Austin, Texas, May 31 – June 2, 1978. P. 40–46.
8. Rissanen J. Independent Components of Relations // ACM Transactions on Database Systems. 1977. Vol. 2, no. 4. P. 317–325.

Информация об авторе.

Поляков Олег Маратович – кандидат технических наук (1982), доцент кафедры информационных технологий предпринимательства Санкт-Петербургского государственного университета аэрокосмического приборостроения, ул. Большая Морская, д. 67, литера А,

Санкт-Петербург, 190000, Россия. Автор более 35 научных публикаций. Сфера научных интересов: лингвистика, искусственный интеллект, математика, теория проектирования баз данных, философия.

О конфликте интересов, связанном с данной публикацией, не сообщалось.
Поступила 24.03.2023; принята после рецензирования 04.05.2023; опубликована онлайн 21.02.2024.

REFERENCES

1. Wittgenstein, L. (2017), *Logisch-philosophische Abhandlung*, Transl. by Dobronravov, I.S. and Lakhuti, D., Kanon+, Moscow, RUS.
2. Katsnel'son, S.D. (1987), "To the concept of valence types", *Voprosy Jazykoznanija*, no. 3, pp. 20–32.
3. Polyakov, O.M. (2019), "Linguistic Data Model for Natural Languages and Artificial Intelligence. Part 1. Categorization", *DISCOURSE*, vol. 5, no. 4, pp. 102–114. DOI: 10.32603/2412-8562-2019-5-4-102-114.
4. Polyakov, O.M. (2023), "R-linguistics. The Principle of Linguistic Duality", *DISCOURSE*, vol. 9, no. 2, pp. 103–110. DOI: 10.32603/2412-8562-2023-9-2-103-110.
5. Polyakov, O.M. (2019), "Linguistic Data Model for Natural Languages and Artificial Intelligence. Part 2. Identification", *DISCOURSE*, vol. 5, no. 5, pp. 98–111. DOI: 10.32603/2412-8562-2019-5-5-98-111.
6. Fagin, R. (1977), "Multivalued Dependencies and a New Normal Form for Relational Databases", *ACM Transactions on Database Systems*, vol. 2, no. 3, pp. 262–278.
7. Nicolas, J.-M. (1978), "First Order Logic Formalization for Functional, Multivalued and Mutual Dependencies", *Proceedings of the 1978 ACM SIGMOD International Conference on Management of Data*, May 31 – June 2, Austin, Texas, USA, pp. 40–46.
8. Rissanen, J. (1977), "Independent Components of Relations", *ACM Transactions on Database Systems*, vol. 2, no. 4, pp. 317–325.

Information about the author.

Oleg M. Polyakov – Can. Sci. (Engineering, 1982), Associate Professor at the Department of Information Technology of Entrepreneurship, Saint Petersburg State University of Aerospace Instrumentation, 67 Bol'shaya Morskaya str., St Petersburg 190000, Russia. The author of over 35 scientific publications. Area of expertise: linguistics, artificial intelligence, mathematics, database design theory, philosophy.

No conflicts of interest related to this publication were reported.
Received 24.03.2023; adopted after review 04.05.2023; published online 21.02.2024.