Конструктивное доказательство теорем Тарского о невыразимости истины и Гёделя о неполноте

Алексей Торгашин

Date: 18 мая 2026 г.

Аннотация

В статье дается принципиально новое конструктивное доказательство теорем Тарского о невыразимости истины и Гёделя о неполноте, основанное на невыразимости экзистенциального содержания внутри формализма («дефицит остенсии»), без которого сам формализм теряет смысл. Показано, почему классическая теорема Гёделя возникает по той же причине.

Содержание

1 Введение

В работе «Конечное о бесконечном» дается теорема с использованием «табличной», «денотативной» предметной области, имеющая те же последствия, что и теорема Гёделя для «достаточно богатых» теорий первого порядка [Torgashin, 2025, pp.16-27]. Она имеет следующие недостатки:

1. это просто неполная теория, которая приводит к нужному результату с помощью рискованного шага потери полноты из-за утраты остенсии исчислением предикатов первого порядка (заведомо полная теория) с последующим её восстановлением в виде актуального указания;

2. это не теорема Гёделя в буквальном смысле.

В этой статье предлагается конструктивное доказательство теоремы Тарского о невыразимости истины и теоремы Гёделя о неполноте для арифметики Робинсона. В основе доказательства следующий механизм:

1. Теория T может быть записана разными способами, имеющими не взаимозаменяемые методики прочтения.

2. По данным теории невозможно определить, какая из записей (они все семантически и синтаксически корректные) - «правильная», с точки зрения методики прочтения.

3. Разные способы чтения записей теории могут давать противоречащие другу другу интерпретации.

4. Невыразимость «правильной» записи чисто формальными средствами (без внешнего указания) приводит к неполноте самой T.

2 Формализация

T - произвольная теория первого порядка с одной обязательной метааксиомой:

Аксиома 2.1. *Любые символы алфавита T могут быть заменены любыми другими символами, если это не приводит к коллизиям.*

Это элементарное требование, запрет на которое равносилен абсурдной онтологизации алфавита. Почему метааксиома, а не просто аксиома? Замене подлежат, вообще говоря, и логические символы.

Определение 2.1. *Назовем записью S_i ($i \in \mathbb{N}$) теории T любую последовательность символов, достатоточную для ее изложения.*

Определение 2.2. *Назовем канонической записью S_0 теории T одну из ее записей.*

Каноническая запись просто фиксирует одну произвольную из них.

Определение 2.3. *Назовем чтением R_i записи S_i ($i \in \mathbb{N}$) теории T алгоритм, преобразующий S_i в S_0.*

Очевидно $R_0(S_0) = id(S_0)$.

Определение 2.4. *Пусть R - некоторый алгоритм. Назовем запись S_j распознанной R, если она либо переведена в S_0, либо признана не T (дисквалифицирована).*

Определение 2.5. *Назовем субграфемой символа алфавита T графические составляющие этого символа, для которых существует методика разделения.*

Примерами субграфем могут служить оба символа русской Ы, литера без диакритики и сама диакритика, ключи иероглифа и т.п.

Предложение 2.1. *T - произвольная теория первого порядка, допускающая замену символов, с канонической записью S_0. Тогда для T может быть составлено неограниченное число записей S_i, $i \in \mathbb{N}$.*

Доказательство. Достаточно конвенционально оговоренного кратного увеличения всех субграфем символов и изменения локации *символов* (не субграфем!) в записи. Неограниченность кратности дает неограниченное число символов и, как следствие, записей. Для устранения однообразия *символы* могут менять значения. В разных записях кратность субграфем разных символов может быть произвольной.

□

Символ, как минимум, состоит из одной субграфемы: всегда возможно преобразование КОТ $\Longrightarrow$ ККООТТ. Примером разной локации могут служить общепринятый и польский стиль написания операций: $3 + 4$ - это и $3\ 4\ +$, и $+\ 3\ 4$. Однако никто не мешает последовательность символов считывать из каких угодно мест от начала записи.

Лемма 2.1. *T - произвольная теория, допускающая замену символов, с записями S_i, $i \in \mathbb{N}$. Тогда для любого i существует $j \neq i$ такой, что $S_i \neq S_j$ и $R_i(S_i) \neq R_j(S_i)$, и $R_i(S_j) \neq R_j(S_j)$.*

Доказательство. Пусть существует единый конечный алгоритм R_g, который переводит все S_j в S_0 или признает, что последовательность символов - не запись T (дисквалифицирует).

Используем метааксиому 2.1. Достаточно ввести нотацию для некоторого j, которая не учитывается R_g (предложение 2.1), например, введением новых неизвестных идентификаторов. Если R_g умеет распознавать все эти идентификаторы, то он перестает быть конечным.

Есть ровно два исхода:

1. R_g - конечен, поэтому введение новых идентификаторов, приводит к его поломке, когда он не распознает запись семантически пригодную для классификации её в качестве T.

   Проблема останова «бесконечно» выполняющегося алгоритма решается диагностикой многократного повторения одних и тех же операций, которое неизбежно из-за обоюдной конечности алгоритма и списка данных.

2. R_g - бесконечен.

   Для начального сегмента R_g длины n обозначим существование некоторой нераспознанной S_j как $C_{R_g}(n) = 0$ и $C_{R_g}(n) = 1$ как несуществование такой S_j ($1 < n \in \mathbb{N}$). Чтобы избежать потери значения $C_{R_g}(n)$ для некоторых n по причине обрыва алгоритма, будем считать значением последовательности длины n значение максимального по длине начального сегмента R_g, поддающегося интерпретации, недостроенную для осмысленной интерпретации часть R_g считаем идемпотентной, но участвующей в подсчете n.

   Тогда, очевидно, для для любого n: $C_{R_g}(n) = 0$, и по свойствам стационарной последовательности

   $$\lim_{n\to\infty} C_{R_g}(n) = 0,$$

   то есть алгоритм ломается и при неограниченном введении в него новых данных.

$\square$

Стоит уточнить на примере способ интерпретации R_g. Легче всего провести аналогию с булевыми выражениями. Например, если это неправильная формула $(a \vee b) \wedge c \wedge (t \vee$, интерпретации поддается $(a \vee b) \wedge c$, а $\wedge(t \vee$ - недостроенная часть формулы, интерпретируемая как идемпотентная. Это позволяет избежать потери значения некоторыми номерами $C_{R_g}(n)$.

Обратим внимание, что S_j - это не теория, и даже не предикат, а просто последовательность символов, которая представляет собой строку графических объектов. То же самое относится к R_g, который представляет собой запись инструкций.

Требуется разъяснить также, что дисквалификация записи T и сбой алгоритма - не одно и то же. Дисквалификация - это, когда алгоритм точно установил, что запись не способна выражать теории, например по принципу Дирихле о клетках и кроликах. Поломка алгоритма, когда алгоритм завершился, не переведя S_j в S_0.

Может возникнуть сомнение: а существует ли настолько сложная теория, которая гарантированно не будет распознаваться при изменении алфавита.

Для обеспечения сложности существует разный порядок чтения, который не обязательно слева направо, и даже не обязательно с первого символа, технически представим и такой, когда первым читается какой-нибудь 10243 символ, потом 452 и далее по схеме. Речь идет не о рациональности, а о технической возможности такой записи. Фрагменты могут заполняться шумом из художественной литературы и браться из разных источников.

Однако всё перечисленное - излишне, и больше для дидактической убедительности в гуманитарном контексте. Достаточно такого контрпримера: пусть теория T_π содержит всего одно предложение $t = \pi$, чтобы она не рапознавалась, достаточно взять неиспользованные идентификаторы для t и для π, потому что на вход R_g может быть подана и теория T_π, и «чепуха» из одной формулы. Пригодность ее в качестве T_π отдается на откуп семантике.

Пронумеруем разные алгоритмы чтения T в зависимости от записи S_l, $l \in \mathbb{N}$.

Теорема 2.1. *Не существует общей валидационной формулы* $VT(l)$

теории T, которая бы определяла «правильный» выбор параметра l, определяющего алгоритм чтения записи теории T.

Доказательство. Докажем утверждение от противного. Предположим, что такая формула $VT(l)$ для T существует.

Достаточно изменить способ записи, согласно лемме 2.1, и получить противоречие, то есть ситуацию, когда семантически корректная запись T ставится в соотвествие «правильному» алгоритму, который ее не может читать.

$\square$

Следует обратить внимание, что $VT(l)$ теоремы 2.1 определяет «правильный» алгоритм для теории T, а не для ее записи. Для доказательства нам не нужно представлять, как $VT(l)$ могла бы это делать, достаточно самой невозможности функции на чистом сопоставлении Dom и Im.

Возникает сомнение в границах применимости теоремы 2.1: нельзя ли методику чтения определить заранее, тем более она «естественная» в виде «с одного края к другому», и остальное угадать в рамках какой-нибудь эвристики? На это есть свой контрпример: даже если существует глобальная каноническая запись теории, невозможно заранее записать и упорядочить привязку идентификатора к предмету материального мира. Знакомство со словом «кошка» или «апостроф» не имеет порядка и у каждого своё, а есть еще «потокосцепление» в электромагнетизме и синоним слова «творог» среди фермионов.

Также возникает сомнение «А вдруг теория T сразу ориентирована на каноническую запись S_0?». Для этого достаточно отдать валидационной формуле номера без номера S_0, она будет вынуждена вернуть или пустое множество, или номер S_l и $l \neq 0$. И то и другое опровергает предположение «природной ассоциации» T и S_0.

Определение 2.6. *Назовем параметр l, однозначно определяющий интерпретацию теории, остенсивным параметром этой теории.*

Теорема 2.2 (Тарского)**.** *Истина формул T невыразима в записи теории T.*

Доказательство. Достаточно взять некоторую формулу *P(x)* с одной свободной переменной, принадлежащую теории T.

Пусть $P(a) = True$ и $P(b) = False$ в некоторой интерпретации, и понимание $P(a)$ как истины определено в записи T, но это противоречит теореме 2.1. Например, достаточно переобозначить a как b, «правильность» которого невозможно определить в записи.

$\square$

Рассмотрим арифметику Q Робинсона [Булос-Джеффри, 1994, с.214].

Язык теории Q содержит константу 0, одноместный функциональный символ S (следование), двуместные функциональные символы $+$ и $\cdot$, а также двуместный предикатный символ $<$. Аксиомами Q являются следующие формулы:

A1. $\forall x \forall y\, (Sx = Sy \to x = y)$

A2. $\forall x\, (0 \neq Sx)$

A3. $\forall x\, (x \neq 0 \to \exists y\, (x = Sy))$

A4. $\forall x\, (x + 0 = x)$

A5. $\forall x \forall y\, (x + Sy = S(x + y))$

A6. $\forall x\, (x \cdot 0 = 0)$

A7. $\forall x \forall y\, (x \cdot Sy = (x \cdot y) + x)$

A8. $\forall x \forall y\, (x < y \leftrightarrow \exists z\, (Sz + x = y))$

Предложение 2.2. *Пусть $M = \{0, 2, 4, 6, \dots\}$ — множество чётных неотрицательных чисел. Определим на M структуру $\mathfrak{M}$:*
$0^{\mathfrak{M}}$,
$S^{\mathfrak{M}}(2n) = 2n + 2$,
$(2a) +^{\mathfrak{M}} (2b) = 2(a + b)$,
$(2a) \cdot^{\mathfrak{M}} (2b) = 2(ab)$,
$<^{\mathfrak{M}}$ *интерпретируется как* $<$.

Тогда отображение $\varphi : \mathbb{N} \to M$, $\varphi(n) = 2n$, является изоморфизмом между $\mathfrak{N}$ и $\mathfrak{M}$. В частности, $\mathfrak{M} \models Q$.

Доказательство. φ биективно. Проверим сохранение операций:

- $\varphi(0_{\mathfrak{N}}) = 0 = 0_{\mathfrak{M}}$.

- $\varphi(S_{\mathfrak{N}}(n)) = \varphi(n+1) = 2n+2 = S_{\mathfrak{M}}(2n) = S_{\mathfrak{M}}(\varphi(n))$.
- $\varphi(a +_{\mathfrak{N}} b) = 2(a+b) = 2a +^{\mathfrak{M}} 2b = \varphi(a) +^{\mathfrak{M}} \varphi(b)$ по определению $+^{\mathfrak{M}}$.
- $\varphi(a \cdot_{\mathfrak{N}} b) = 2(ab) = (2a) \cdot^{\mathfrak{M}} (2b) = \varphi(a) \cdot^{\mathfrak{M}} \varphi(b)$ по определению $\cdot^{\mathfrak{M}}$.
- Сохранение порядка очевидно.

Следовательно, $\mathfrak{M}$ изоморфна $\mathfrak{N}$, а значит, удовлетворяет тем же предложениям, что и $\mathfrak{N}$. Так как $\mathfrak{N} \models Q$, то $\mathfrak{M} \models Q$. □

Теорема 2.3 (Гёделя). *Формальная запись арифметики Q или противоречива, или неполна.*

Доказательство. В силу теоремы 2.2 (Тарского) в записи Q невозможно определить, какая модель используется $\mathfrak{N}$ или $\mathfrak{M}$, поэтому строго формальное определение Q оказывается неполным.

Попытка ввести элементы определения и $\mathfrak{N}$, и $\mathfrak{M}$ *на одной и той же сигнатуре* фатально приводят к противоречию.

□

Возникает сомнение, а нельзя ли использовать пары

$\{0^{\mathfrak{M}}, 0^{\mathfrak{N}}\}$, $\{S^{\mathfrak{M}}, S^{\mathfrak{N}}\}$, $\{+^{\mathfrak{M}}, +^{\mathfrak{N}}\}$, $\{\cdot^{\mathfrak{M}}, \cdot^{\mathfrak{N}}\}$, $\{<^{\mathfrak{M}}, <^{\mathfrak{N}}\}$ вместе? Но это не решает вопрос полноты, потому что для модели $\mathfrak{M}$ существует свой аналог $\mathfrak{M}$ уже с её «чётными числами». Попытка ввести элементы определения двух моделей, чтобы они «видели» друг друга и «называли» друг друга, не уступая друг другу общности интерпретации Q, неминуемо приведет к противоречию, потому что «одно и то же число» для первой окажется 64, а для второй 128.

Проверим полноту, как в суде присяжных:

Истинно ли $4 \cdot 16 = 64$ (четыре раз по шестнадцать яблок)?

- Да, истинно.

- Поддается ли это чисто формальному определению?

- Нет, не поддается.

3 Остенсивная природа классической теоремы

Геделева нумерация - это как раз способ чисто синтаксического определения. Ей «всё равно», она нумерует без оглядки, противоречит это какому-то «пониманию» или не противоречит. Формальный способ определения «сломался», когда в него попытались вложить валидационную формулу, обязанную иметь в то же время семантическое значение.

Механизм доказательства у Геделя следующий: формула, выражающая невыводимость, вынуждена обладать истинностью, при этом истинность из-за требования полноты предполагает выводимость. Само по себе это не страшно, но Гедель нашел такую формулу, которая выражает невыводимость той же формулы [Булос-Джеффри, 1993, с.233].

«Внутри семантики» и «внутри синтаксиса» такое противоречие невыразимо. Геделевы номера $gn(\neg G)$ и $gn(G)$, с синтаксической точки зрения, - *просто номера, множества, в которые они включены, - просто множества.* Противоречие возникает только при переходе к семантике.

«Внутри семантики» формулы $\neg G$ и G одновременно могут появиться в качестве истинных разве что в результате технической ошибки.

Формула *Prov(x)*, определяющая номера теорем T, - это попытка выразить обе взаимоисключающие остенсии и обойтись одним синтаксисом. Гёдель обязал теорию определиться синтаксически, не дав возможности «заглянуть за край» по Тарскому, и гарантированно нашел такую частную формулу, которая выражает *невозможный смысл* на эту тему.

Откуда тогда вообще берутся полные теории? Для полных теорий не требуется приведения к чисто синтаксическому виду, оно всегда отчасти семантическое и интуитивно понятное, которое подавляет другие определения в качестве *маргинальных.* Арифметика оказалась не настолько прозрачной, её определение пошло по строго формальному пути создания процедуры, «полностью» избавленной от семантического компонента. Этого не хватило для выполнения программы формализации, но хватило с избытком для потери остенсии. Именно такие теории называются «достаточно богатыми» для неполноты. Исчисление предикатов первого порядка и остальные полные теории в данном аспекте *принципиально* не отличаются от арифметики, кроме того, что остенсия там интуитивно понятна на уровне: «вот предметная область!». Именно на этом строился первначальный вариант теоремы о неполноте, на который указывалось во введении. Правила игры «Крестики и нолики» можно считать полной теорией (для ограниченного игрового поля мы знаем все комбинации расстановки обоих символов сигнатуры), но она станет неполной, если мы забудем, что такое «крестики» и что такое «нолики». Тоже самое произойдет при поиптке ее строго формального определения.

Коротко результат этой статьи можно выразить так: о формализации теории с остенсией говорить не имеет смысла, это *не общий вид*, а теория без остенсии - заведомо не полна.

4 Итоги

Трудно переоценить глубину теоремы Геделя и теоремы Тарского. Теперь мы понимаем, что их «физический смысл» в игре того, что выражает, и того, что может выразить. Это основной результат теории формализации как таковой. Понимание его природы и связанных с нею коллизий дает возможность наиболее точных и безопасных способов определения. Это особенно актуально в эпоху искусственного интеллекта, когда автоматизации подлежат даже творческие паттерны, которые невозможно проверить актуальным сопоставлением с предметной областью.

Список литературы

[1] Булос Дж., Джеффри Р. Вычислимость и логика - М., Мир, 1994 - 396 с.

[2] Torgashin A. Finite about Infinite [Электронный ресурс] / A. Torgashin. – Режим доступа: https://zenodo.org/records/18100849 (дата обращения: 18.05.2026).