Cezary Cieśliński

Dlaczego prawda jest (nie)definiowalna

Filozofia Nauki 13/1, 15-23

2005

Filozofia Nauki
Rok XIII, 2005, Nr 1(49)

Cezary Cieśliński

Dlaczego prawda jest (nie)definiowalna

WSTĘP

W 1933 roku Alfred Tarski przedstawił dowód słynnego twierdzenia o niedefiniowalności prawdy.[1] W swobodnym sformułowaniu twierdzenie to mówi, że predykat prawdy dla zdań danego języka nie należy do tego języka. Wynik ten posiada ciekawe konsekwencje zarówno w matematyce, jak i w filozofii. Z filozoficznego punktu widzenia wspomniane twierdzenie dostarczyło cząstkowej lecz istotnej odpowiedzi na pytanie o granice wyrażalności naszego języka. Stanowi również ważny punkt odniesienia w filozoficznej debacie o naturę prawdy arytmetycznej, którego żaden teoretyk (niezależnie od zajmowanego stanowiska) nie może zignorować.

W pierwszej części artykułu przedstawimy znane wyniki o pozytywnym charakterze, dzięki którym dowiedzieliśmy się, ile można powiedzieć w danym języku o semantyce tego języka. Podstawowy temat niniejszej pracy zostanie rozwinięty w części drugiej. Będziemy tam rozważać pytanie o przyczyny niedefiniowalności prawdy — dlaczego niektóre zbiory zdań są „za duże", abyśmy mogli dysponować dla nich predykatem prawdy. W tym kontekście zaprezentujemy także pewne rozszerzenie klasycznych rezultatów przedstawionych w części pierwszej; skonstruujemy też nieco inny od klasycznego dowód twierdzenia o hierarchii arytmetycznej. Na koniec przedstawimy otwarte zagadnienia, wymagające dopiero rozwiązania.

---

[1] A. Tarski, *Pojęcie prawdy w językach nauk dedukcyjnych*, „Prace Tow. Naukowego Warszawskiego", t. III nr 34, 1933.

I

O predykatach prawdy można myśleć na dwa sposoby. Z jednej strony, możemy przyjąć perspektywę naszej ulubionej teorii; z drugiej, możemy spojrzeć na nie z punktu widzenia modelu. Tym dwóm ujęciom odpowiadają dwa podpunkty poniższej definicji.

DEFINICJA 1

(a) Formuła $Tr_K(x)$ jest predykatem prawdy dla zbioru zdań K w modelu M wtedy i tylko wtedy, gdy dla dowolnego $\psi \in K$, $M \models Tr_K(\psi) \equiv \psi$.

(b) Formuła $Tr_K(x)$ jest predykatem prawdy dla zbioru zdań K w teorii T wtedy i tylko wtedy, gdy dla dowolnego $\psi \in K$, $T \vdash Tr_K(\psi) \equiv \psi$

Oczywiście istnienie predykatu prawdy dla K w pewnej niesprzecznej teorii T (czyli w sensie warunku (b)) pociąga za sobą istnienie takiego predykatu w pewnym modelu M (mianowicie w modelu dla T). Warto jednak zauważyć, że jeśli poza niesprzecznością będziemy wymagać od T również aksjomatyzowalności (a jest to często stawiany wymóg), to warunki (a) i (b) nie będą równoważne: dany zbiór K może mieć predykat prawdy w danym modelu, choć w żadnej aksjomatyzowalnej teorii T nie istnieje dla niego predykat prawdy. Przykład takiego zbioru K podamy w części drugiej. Oczywiście jeśli nie wymagamy aksjomatyzowalności, to z istnienia predykatu prawdy dla K w M wynika istnienie predykatu prawdy dla K w pewnej teorii T (wystarczy wziąć za T zbiór wszystkich zdań prawdziwych w M).

Posługując się stylistyką definicji 1, możemy w następujący sposób wyrazić twierdzenie Tarskiego, biorąc za K zbiór *Sent* (wszystkich zdań).

TWIERDZENIE TARSKIEGO

Nie istnieje $M \models PA$ oraz formuła $Tr(x)$ taka że $Tr(x)$ jest predykatem prawdy dla *Sent* w M.

Z tego naturalnie wynika, że nie istnieje niesprzeczna teoria T (aksjomatyzowalna lub nie) zawierająca PA, mająca predykat prawdy dla zbioru *Sent*.

Twierdzenie Tarskiego pokazało, jak doniosłą rolę odgrywa rozróżnienie „język przedmiotowy — metajęzyk". Nie powinniśmy tu jednak posuwać się zbyt daleko. Nie ma sensu np. upierać się, że zdanie „Kot ma cztery nogi" należy do języka przedmiotowego, a wyrażenie „Zdanie „Kot ma cztery nogi" składa się z czterech słów" należy do metajęzyka, skoro wspomina się w nim o zdaniu języka przedmiotowego. Wynik Tarskiego nie uzasadnia takiego rozróżnienia. Jeśli poruszamy się w obrębie teorii składni (a w powyższym przykładzie poza nią nie wykraczamy), to nie potrzebujemy apelować do żadnego bogatszego języka, przynajmniej jeśli nasz język przedmiotowy jest wystarczająco bogaty, np. jeśli możemy w nim wyrazić elementarną arytmetykę dodawania i mnożenia. Jak pokazał Gödel,[2] składniowy opis języka arytmetyki nie wymaga wykroczenia poza ten język: w języku arytmetyki można np. zdefiniować pojęcie termu i formuły tego języka.

Przejście do metajęzyka staje się konieczne dopiero przy wprowadzeniu semantycznego predykatu prawdy. I tu jednak trzeba zachować ostrożność. Pytanie polega na tym, dla jakiej klasy K zdań chcemy uzyskać predykat prawdy. Twierdzenie Tarskiego mówi, że musimy wejść na poziom metajęzyka w przypadku, gdy K = zbiór wszystkich zdań. Co się stanie, jeśli będziemy rozważać mniejsze zbiory? Nie ma jednej odpowiedzi na to pytanie — wszystko zależy od tego, jaki zbiór rozważamy. Zacznijmy od przykładów skrajnych (i trywialnych). Po pierwsze, jeśli K jest zbiorem skończonym, to pewien predykat prawdy $Tr_K(x)$ dla K należy do języka przedmiotowego. Po prostu, dla $K = \{\psi_1 \dots \psi_n\}$, bierzemy formułę:

$$Tr_K(x) := (x = \ulcorner\psi_1\urcorner \wedge \psi_1) \vee \dots \vee (x = \ulcorner\psi_n\urcorner \wedge \psi_n)$$

i odpowiednia równoważność dla K jest trywialnie spełniona. Po drugie, jeśli K jest zbiorem ko-skończonym, to żaden predykat prawdy dla K nie należy do języka przedmiotowego. To wynika łatwo z twierdzenia Tarskiego — po prostu gdyby taki predykat $Tr_K(x)$ istniał, zaś $Sent - K = \{\psi_1 \dots \psi_n\}$, to następująca formuła języka przedmiotowego byłaby predykatem prawdy dla *Sent*:

$$(x \neq \ulcorner\psi_1\urcorner \wedge \dots\ x \neq \ulcorner\psi_n\urcorner \wedge Tr_K(x)) \vee (x = \ulcorner\psi_1\urcorner \wedge \psi_1) \vee \dots \vee (x = \ulcorner\psi_n\urcorner \wedge \psi_n)$$

Otrzymalibyśmy zatem jawną sprzeczność z twierdzeniem Tarskiego.

Pomiędzy tymi trywialnymi ekstremami sytuują się mniej banalne przypadki. Znany i często wykorzystywany rezultat dotyczy klas zdań w hierarchii arytmetycznej. Przypomnijmy podstawowe definicje i fakty.

DEFINICJA 2

(a) Kwantyfikatorem ograniczonym nazywamy kwantyfikator o postaci $\forall x < y$ oraz $\exists x < y$.

(b) Dana formuła jest klasy Δ_0 ($=\Pi_0$, $=\Sigma_0$) gdy wszystkie występujące w niej kwantyfikatory są ograniczone.

(c) Dana formuła jest klasy Σ_{n+1}, gdy ma ona postać $\exists \mathbf{x}\varphi$, gdzie $\mathbf{x}$ jest ciągiem zmiennych, zaś $\varphi \in \Pi_n$.

(d) Dana formuła jest klasy Π_{n+1}, gdy ma ona postać $\forall \mathbf{x}\varphi$, gdzie $\mathbf{x}$ jest ciągiem zmiennych, zaś $\varphi \in \Sigma_n$.

[2] K. Gödel, *Über formal unentscheidbare Sätze der Prinzipia Mathematica und Verwandter Systeme I*, „Monatschefte Math. Phys.", t. 38, 1931, s. 173-198.

Okazuje się, że jeśli chcemy posługiwać się predykatami prawdy dla klas Σ_n oraz Π_n, nie potrzebujemy wznosić się na poziom metajęzyka — wystarczy nam zwykły język arytmetyki. Uzyskano tu następujący rezultat.

TWIERDZENIE O DEFINIOWALNOŚCI PRAWDY DLA OGRANICZONYCH KLAS FORMUŁ

(a) Istnieją formuły φ, ψ takie że $\varphi \in \Pi_1$ i $\psi \in \Sigma_1$ i zarówno φ jak ψ są predykatami prawdy dla Δ_0 w *PA*, a przy tym $PA \vdash \varphi \equiv \psi$.

(b) Dla dowolnego $n > 0$, istnieje formuła $Tr_{\Sigma n}(x)$ klasy Σ_n, będąca predykatem prawdy dla klasy Σ_n w *PA*.

(c) Dla dowolnego $n > 0$, istnieje formuła $Tr_{\Pi n}(x)$ klasy Π_n, będąca predykatem prawdy dla klasy Π_n w *PA*.

Dowód tego twierdzenia można odnaleźć w licznych podręcznikach i monografiach.[3]

Jak więc widzimy, dla rozmaitych klas zdań uzyskuje się na interesujące nas pytanie pozytywne lub negatywne odpowiedzi. W tej sytuacji pojawia się kolejne pytanie: czy za tymi wynikami kryje się jakaś głębsza, ogólna prawidłowość? Mówi się czasem, że niektóre klasy zdań są „za duże", by można było skonstruować dla nich predykat prawdy. Cóż to jednak znaczy „za duże"? (Przyjmujemy tu, że nie zadowala nas nieciekawe wyjaśnienie: takie, które nie dopuszczają predykatu prawdy.) Przekonaliśmy się już, że „za duża" to nie to samo, co „nieskończona" — klasa Σ_n ma przecież predykat prawdy, choć skończona nie jest. Czy da się o tym powiedzieć cokolwiek pozytywnego?

Tę właśnie kwestię będziemy rozważać w drugiej części niniejszego artykułu.

II

Niech *T* będzie aksjomatyzowalnym, niesprzecznym rozszerzeniem *PA*. Określimy obecnie klasę zdań, które są Σ_n (ewentualnie Π_n) względem teorii *T*:

DEFINICJA 3

$$\Sigma_n(T) = \{\varphi\colon \exists\psi \in \Sigma_n\ T \vdash \varphi \equiv \psi\}$$

$$\Pi_n(T) = \{\varphi\colon \exists\psi \in \Pi_n\ T \vdash \varphi \equiv \psi\}$$

$\Sigma_n(T)$ jest zatem klasą zdań, o których *T* dowodzi, że są one równoważne Σ_n zdaniom; analogiczny komentarz dotyczy klasy $\Pi_n(T)$.

---

[3] Zob. np. P. Hajek, P. Pudlak, *Metamathematics of First Order Arithmetic*, Berlin, Nowy Jork, Paryż, Springer Verlag 1933 albo R. Kaye, *Models of Peano Arithmetic*, Oxford Logic Guides, Oxford University Press 1991.

Obecnie pokażemy, że klasy $\Sigma_n(T)$ oraz $\Pi_n(T)$ mają predykaty prawdy w teorii T. Wspomniane predykaty są przy tym odpowiednio klasy Σ_n oraz Π_n.

TWIERDZENIE 1

Dla dowolnej liczby naturalnej n i dla dowolnej teorii T będącej aksjomatyzowalnym, niesprzecznym rozszerzeniem PA, istnieje formuła $Tr_{\Sigma n(T)}(x)$ klasy Σ_n, będąca predykatem prawdy dla klasy $\Sigma_n(T)$ w T.

DOWÓD:

Określamy $Tr_{\Sigma n(T)}(x)$ jako następującą formułę:

$$\exists\beta d[Prov_T(d, \ulcorner\beta \equiv x\urcorner) \wedge \beta \in \Sigma_n \wedge Tr_{\Sigma n}(\beta) \wedge \forall\alpha, d_1 < d(Prov_T(d_1, \alpha \equiv x) \Rightarrow \alpha \notin \Sigma_n)]$$

gdzie:

- „$Prov_T(x, y)$" to formuła, którą odczytujemy: „x jest dowodem y na gruncie teorii T" (dla aksjomatyzowalnych rozszerzeń PA, taka formuła będzie klasy Δ_0)
- „$x \in \Sigma_n$" to Δ_0 formuła, która znaczy „x jest formułą klasy Σ_n".
- „$Tr_{\Sigma n}(x)$" to Σ_n formuła, będąca predykatem prawdy dla klasy Σ_n w PA.

Przy intuicyjnej interpretacji, nasza formuła $\mathrm{Tr}_{\Sigma n(T)}(x)$ mówi zatem:

1. d jest dowodem (na gruncie T) równoważności x oraz pewnego prawdziwego Σ_n zdania β
2. poniżej d nie ma dowodu równoważności x z jakąś formułą klasy Σ_n.

Obecnie twierdzimy, że nasza formuła jest predykatem prawdy dla $\Sigma_n(T)$ w T, czyli:

dla dowolnej ψ należącej do $\Sigma_n(T)$, $T \vdash \mathrm{Tr}_{\Sigma n(T)}(\psi) \equiv \psi$

Rozważymy kolejno obie implikacje.

($\Leftarrow$)

Niech $M \models T$, a przy tym $M \models \psi$. Pokazujemy, że $M \models \mathrm{Tr}_{\Sigma n(T)}(\psi)$. Skoro ψ należy do $\Sigma_n(T)$, to wybierzmy β oraz d spełniające warunki:

(1) d jest dowodem równoważności $\ulcorner\beta \equiv \psi\urcorner$ na gruncie teorii T

(2) $\beta \in \Sigma_n$

(3) dla dowolnego α, $d_1 < d$, jeśli d_1 jest dowodem równoważności $\ulcorner\alpha \equiv \psi\urcorner$ na gruncie teorii T, to $\alpha \notin \Sigma_n$.

Czyli d jest najmniejszym dowodem (na gruncie T) równoważności ψ oraz pewnego zdania β klasy Σ_n. Należy zauważyć, że wówczas zarówno d jak β są standardowe.

Warunek (3) uzyskujemy w następujący sposób. Załóżmy, że $\psi \in \Sigma_n(T)$, $M \models T$ i $M \models \psi$. Istnieje zatem d takie że $\exists\beta[\beta \in \Sigma_n(T) \wedge d$ jest dowodem $\ulcorner\beta \equiv \psi\urcorner$ na gruncie teorii $T]$. Istnieje zatem najmniejsze d o tej własności. Zatem dla dowolnego $d_1 < d$ i dla dowolnego α, jeśli d_1 jest dowodem $\ulcorner\beta \equiv \psi\urcorner$ na gruncie teorii T, to $\alpha \notin \Sigma_n$.

To zaś jest równoważne wersji, w której kwantyfikator wiążący α zostaje ograniczony do d.

Otrzymujemy:

(a) $M \models Prov_T(d, \ulcorner\beta \equiv \psi\urcorner)$

(b) $M \vDash \beta \in \Sigma_n$

(c) $M \vDash \forall\alpha, d_1 < d\,(Prov_T(d_1, \ulcorner\alpha \equiv \psi\urcorner) \Rightarrow \alpha \notin \Sigma_n)$.

Dowód polega na tym, że zarówno d jak β są standardowe, a przy tym powyższe zdania są klasy Δ_0, jeśli więc są prawdziwe w standardowym modelu, to również i w M.

Co więcej, mamy także: (d) $M \vDash Tr_{\Sigma_n}(\beta)$, bo gdyby $M \vDash \neg Tr_{\Sigma_n}(\beta)$, to $M \vDash \neg\beta$ (formuła β jest bowiem Σ_n, a formuła $Tr_{\Sigma_n}(x)$ jest predykatem prawdy dla Σ_n formuł w PA, a zatem i w M), więc $M \vDash \neg\psi$, gdyż $T \vdash \beta \equiv \psi$. Wniosek ten przeczy zaś naszemu początkowemu założeniu, zgodnie z którym $M \vDash \psi$.

Ale skoro (a), (b), (c) i (d), to $M \vDash \mathrm{Tr}_{\Sigma n(T)}(\psi)$.

($\Rightarrow$)

Załóżmy, że $M \vDash \mathrm{Tr}_{\Sigma n(T)}(\psi)$. Wybierzmy β oraz d, takie że:

$M \vDash Prov_T(d, \ulcorner\beta \equiv \psi\urcorner) \wedge \beta \in \Sigma_n \wedge Tr_{\Sigma_n}(\beta)$

$M \vDash \forall\alpha, d_1 < d\,(Prov_T(d_1, \ulcorner\alpha \equiv \psi\urcorner) \Rightarrow \alpha \notin \Sigma_n)$.

Skoro $\psi \in \Sigma_n(T)$, to podobnie jak w dowodzie implikacji odwrotnej, wybierzmy c oraz γ, takie że

(1) c jest dowodem równoważności $\ulcorner\gamma \equiv \psi\urcorner$ na gruncie teorii T

(2) $\gamma \in \Sigma_n$

(3) dla dowolnego α, $d_1 < c$, jeśli d_1 jest dowodem równoważności $\ulcorner\alpha \equiv \psi\urcorner$ na gruncie teorii T, to $\alpha \notin T$. Należy zauważyć, że wówczas zarówno c jak γ są standardowe.

Zatem z tych samych powodów co w dowodzie odwrotnej implikacji:

$M \vDash Prov_T(c, \ulcorner\gamma \equiv \psi\urcorner)$

$M \vDash \gamma \in \Sigma_n$

$M \vDash \forall\alpha, d_1 < c\,(Prov_T(d_1, \ulcorner\alpha \equiv \psi\urcorner) \Rightarrow \alpha \notin \Sigma_n)$.

Zauważamy, że wówczas $d = c$. Załóżmy przeciwnie, przyjmując bez utraty ogólności, że $d < c$. Wtedy:

$M \vDash \forall\alpha < c[Prov_T(d_1, \ulcorner\alpha \equiv \psi\urcorner) \Rightarrow \alpha \notin \Sigma_n]$.

Ale $M \vDash Prov_T(d, \ulcorner\beta \equiv \psi\urcorner)$, a przy tym $\beta < c$, bo β jest podformułą ostatniego wyrazu ciągu d, zaś $d < c$. Więc $\beta \notin \Sigma_n$. Sprzeczność, bo $\beta \in \Sigma_n$.

Zatem $\beta = \gamma$.

Więc $M \vDash \mathrm{Tr}_{\Sigma n}(\gamma)$, zatem $M \vDash \gamma$ (pamiętajmy, że γ jest standardowa), stąd: $M \vDash \psi$.

Możemy uzyskać również analogiczny wynik dotyczący klasy $\Pi_n(T)$.

TWIERDZENIE 2

Dla dowolnej liczby naturalnej n i dla dowolnej teorii T będącej aksjomatyzowalnym, niesprzecznym rozszerzeniem PA, istnieje formuła $Tr_{\Pi n(T)}(x)$ klasy Π_n, będąca predykatem prawdy dla klasy $\Pi_n(T)$ w T.

SZKIC DOWODU

Określamy $Tr_{\Pi n(T)}(x)$ jako następującą formułę:

$$\forall\beta d\,[(Prov_T(d, \ulcorner\beta \equiv x\urcorner) \wedge \beta \in \Pi_n \wedge \forall\alpha, d_1 < d(Prov_T(d_1, \alpha \equiv x) \Rightarrow \alpha \notin \Sigma_n)) \Rightarrow Tr_{\Pi_n}(\beta)]$$

gdzie:
„$x \in \Pi_n$" to Δ_0 formuła, która mówi „x jest formułą klasy Π_n"
„$Tr_{\Pi n}(x)$" to Π_n formuła, będąca predykatem prawdy dla klasy Π_n w *PA*.
I dalej postępujemy podobnie jak w dowodzie twierdzenia 1.

Przedstawione powyżej wyniki możemy następnie wykorzystać, aby uzyskać nieco inny od klasycznego dowód twierdzenia o hierarchii arytmetycznej.

TWIERDZENIE O HIERARCHII ARYTMETYCZNEJ

Niech T będzie aksjomatyzowalnym, niesprzecznym rozszerzeniem *PA*. Wtedy dla dowolnej liczby naturalnej n:

(a) Istnieje zdanie ψ należące do $\Pi_n(T)$, takie że $\forall\varphi \in \Sigma_n(T)\ T \nvdash \psi \equiv \varphi$.

(b) Istnieje zdanie ψ należące do $\Sigma_n(T)$, takie że $\forall\varphi \in \Pi_n(T)\ T \nvdash \psi \equiv \varphi$.

DOWÓD:

Z lematu przekątniowego, weźmy zdanie ψ takie że $T \vdash \psi \equiv \neg\, Tr_{\Sigma n(T)}(\psi)$. Skoro formuła $Tr_{\Sigma n(T)}(x)$ należy do klasy Σ_n, wolno nam przyjąć, że ψ należy do Π_n. Załóżmy nie wprost, że dla pewnego φ należącego do $\Sigma_n(T)$, $T \vdash \psi \equiv \varphi$. Zatem ψ należy do $\Sigma_n(T)$, więc skoro $Tr_{\Sigma n(T)}(x)$ jest predykatem prawdy dla klasy $\Sigma_n(T)$ w T, to otrzymujemy: $T \vdash \psi \equiv Tr_{\Sigma n(T)}(\psi)$. Ostatecznie okazuje się więc, że T jest sprzeczna, wbrew początkowemu założeniu, co kończy dowód części (a). Dowód części (b) jest podobny.

Z twierdzenia o hierarchii wynika, że istnienie predykatu prawdy w danym modelu nie przekłada się na istnienie odpowiedniego predykatu w teorii. Sformułujemy to w postaci następującego wniosku.

WNIOSEK 1

Istnieje zbiór zdań Z oraz model M taki że:
1. W modelu M istnieje predykat prawdy dla Z.
2. W żadnej aksjomatyzowalnej, niesprzecznej teorii T nie istnieje predykat prawdy dla Z.

DOWÓD:

Wystarczy tu rozważyć przykład dowolnego niesprzecznego, zupełnego zbioru zdań Z. Niech $M \models Z$. Wtedy formuła „$x = x$" jest w M predykatem prawdy dla Z, warunek (1) jest więc spełniony. Gdyby jednak istniała aksjomatyzowalna, niesprzeczna teoria T z predykatem prawdy dla Z, otrzymalibyśmy sprzeczność z twierdzeniem o hierarchii: taki predykat musiałby być Σ_n dla pewnego n, a zatem ze względu na zupełność Z, każde zdanie lub jego negacja należałoby do klasy $\Sigma_n(T)$.

WNIOSEK 2

Zbiór zdań Z ma predykat prawdy w teorii T wtedy i tylko wtedy, gdy dla pewnej liczby naturalnej n, $Z \subseteq \Sigma_n(T)$.

DOWÓD:

Jeśli $Z \subseteq \Sigma_n(T)$, to predykat prawdy dla $\Sigma_n(T)$ jest zarazem predykatem prawdy dla Z. Z kolei implikacja w drugą stronę jest oczywista, jeśli bowiem „$Tr_Z(x)$" jest predykatem prawdy dla Z, to możemy założyć, że dla pewnego n, należy on do klasy Σ_n. Zatem $Z \subseteq \Sigma_n(T)$, bo każde zdanie ψ należące do Z jest równoważne (dowodliwie w T) pewnemu zdaniu klasy Σ_n — mianowicie zdaniu $Tr_Z(\psi)$.

W ten sposób uzyskujemy odpowiedź na wyjściowe pytanie — pytanie o przyczyny niedefiniowalności prawdy w teorii T dla rozmaitych zbiorów zdań. Nie potrafimy zdefiniować prawdy dla tych i tylko dla tych zbiorów zdań, które z punktu widzenia T są „za duże", tzn. nie mieszczą się w żadnej klasie, którą z perspektywy T postrzegamy jako klasę skończonej złożoności kwantyfikatorowej. Określenie „za duża klasa zdań" uzyskuje zatem w tym kontekście precyzyjny sens.

Analogiczne pytanie można postawić w odniesieniu do definiowalności w modelu. Tu również nie dysponujemy predykatem prawdy dla „zbyt dużych" zbiorów. Cóż to jednak znaczy „zbyt dużych"? Jak się przekonany, tu sytuacja nie przedstawia się tak prosto jak w poprzednim przypadku.

Zauważmy przede wszystkim, że nasze wcześniejsze rozwiązanie tu nie działa. Oczywiście moglibyśmy określić klasę $\Sigma_n(M)$ w analogicznym stylu jak zbiór $\Sigma_n(T)$ — należałyby do niej te zdania, które w M są równoważne Σ_n formułom. Jednakże w tym przypadku nie powstałaby żadna hierarchia: każde zdanie należałoby po prostu do klasy $\Sigma_0(M)$, byłoby bowiem równoważne Σ_0 — zdaniu „$0 = 0$" (czyli prawdziwe) lub jego negacji (czyli fałszywe).

Inne możliwe podejście polegałoby na bezpośrednim odwołaniu się do twierdzenia Tarskiego. Wiemy, że w M nie ma predykatu prawdy dla zbioru *Sent*. A może na tym właśnie polega przyczyna niedefiniowalności prawdy dla dowolnych zbiorów zdań? Zgodnie z tym przypuszczeniem klasa zdań byłaby „zbyt duża", gdy rozszerzenie naszego języka o predykat prawdy dla tej klasy pozwoliłoby od razu zdefiniować prawdę dla całego zbioru *Sent*. W ten sposób dochodzimy do następującej hipotezy (sformułowanej poniżej w wersji dla modelu standardowego).

HIPOTEZA

Nie istnieje zbiór zdań K, taki że:

(1) nie ma w standardowym modelu predykatu prawdy dla K

(2) rozszerzenie języka o predykat prawdy dla K nie pozwala zdefiniować w modelu standardowym zbioru wszystkich prawdziwych zdań języka arytmetyki.

Prawdziwość tej hipotezy oznaczałaby, że argument za niedefiniowalnością prawdy, podany przez Tarskiego w dowodzie jego twierdzenia, ma ogólny charakter i możemy się nim posłużyć w dowolnej sytuacji, w której mamy do czynienia z tym zjawiskiem. Czy jednak hipoteza jest prawdziwa? Pewne racje każą w to wątpić. Badania nad zjawiskiem definiowalności pokazały, że zbiory liczb naturalnych mogą okazać się niedefiniowalne z rozmaitych powodów. W szczególności okazuje się, że istnieją niedefiniowalne zbiory, w przypadku których rozszerzenie języka arytmetyki o odpowiadający im predykat nie pozwala nam bynajmniej na zdefiniowanie zbioru wszystkich prawd arytmetycznych.[4] W świetle tego wyniku, spodziewalibyśmy się raczej istnienia kontrprzykładu, obalającego powyższą hipotezę. Taki kontrprzykład należałoby jednak dopiero skonstruować.

---

[4] Formalnie: istnieje zbiór Z taki że Z nie jest definiowalny w N, a przy tym $Th(N)$ (czyli zbiór wszystkich prawdziwych zdań arytmetyki) nie jest definiowalna w (N, Z). Własność tę posiadają niektóre zbiory ω-generyczne. Zob. P. G. Odifreddi, *Clasical Recursion Theory*, t. II, Amsterdam, Elsevier 1999, s. 737-780.