Klasyczna definicja prawdy Alfreda Tarskiego i prawda w modelach skończonych

Marek Czarnecki

15 stycznia 2009

1 Twierdzenie Tarskiego o niedefiniowalności prawdy arytmetycznej

1.1 Pojęcie spełniania

Słownikiem nazywamy czwórkę $(\mathbb{P}, \mathbb{F}, \mathbb{C}, \mathrm{ar})$, gdzie $\mathbb{P}$ jest zbiorem predykatów, $\mathbb{F}$ zbiorem symboli funkcyjnych, $\mathbb{C}$ zbiorem stałych, a $ar : \mathbb{P} \cup \mathbb{F} \to \mathbb{N}$ jest funkcją arności. Zakładamy, że zbiory $\mathbb{P}, \mathbb{F}, \mathbb{C}$ są parami rozłączne.

Każdy słownik wyznacza zbiór **termów** oraz **formuł**.
Niech $\sigma = (\{P_i\}_{i \leq r}, \{f_i\}_{i \leq s}, \{c_i\}_{i \leq t}, \mathrm{ar})$ będzie słownikiem. Przez **Var** oznaczać będziemy przeliczalny zbiór zmiennych indywiduowych. Zbiór Trm_σ **termów** nad słownikiem σ definiujemy indukcyjnie, jako najmniejszy zbiór spełniający warunki:

- Każda zmienna jest termem,
- Każda stała c_i, dla $i \leq t$ jest termem,
- Dla $i \leq s$, jeżeli $t_1, ..., t_{ar(f_i)}$ są termami, to $f(t_1, ..., t_{ar(f_i)})$ również jest termem.

Formułą atomową nad σ nazywamy dowolne wyrażenie postaci: $t_1 = t_2$ oraz $P_i(t_1, ..., t_{ar(P_i)})$, gdzie $i \leq r$, a $t_1, t_2, ..., t_{ar(P_i)} \in \mathrm{Trm}_\sigma$. Zbiór $\mathcal{F}_\sigma$ **formuł** słownika σ definiujemy jako najmniejszy zbiór spełniający warunki:

- Każda formuła atomowa jest formułą,
- Dla $\varphi, \psi \in \mathcal{F}_\sigma$ zachodzi $\neg\varphi \in \mathcal{F}_\sigma$ i $\varphi \vee \psi \in \mathcal{F}_\sigma$,
- Dla $v \in \mathbf{Var}$ oraz $\varphi \in \mathcal{F}_\sigma$ również $\forall v \varphi \in \mathcal{F}_\sigma$.

Modelem słownika σ nazywamy czwórkę: $\mathcal{A} = (A, \{R_i\}_{i \leq r}, \{F_i\}_{i \leq s}, \{a_i\}_{i \leq t})$, gdzie A jest niepustym zbiorem zwanym uniwersum modelu i oznaczanym przez $|\mathcal{A}|$, $R_i \subseteq A^{ar(P_i)}$ są relacjami na A dla $i \leq r$ interpretującymi odpowiednie predykaty, $F_i : A^{ar(f_i)} \to A$ są funkcjami dla $i \leq s$ interpretującymi odpowiednie symbole funkcyjne oraz $a_i \in A$ są elementam modelu interpretującymi odpowiednie stałe dla $i \leq t$. Klasę modeli nad słownikiem σ oznaczamy przez Mod_σ.

Wartościowaniem w modelu $\mathcal{A}$ nazywamy dowolną funkcję $v : \mathbf{Var} \to |\mathcal{A}|$. Przez $v(x_i/a)$ oznaczamy wartościowanie v' takie, że dla $j \neq i$ zachodzi $v(x_j) = v'(x_j)$ oraz $v'(x_i) = a$. W naturalny sposób rozszerzamy pojęcie wartościowania na termy:

- $v(c_i) = a_i$ dla $i \leq t$,
- $v(f_i(t_1, ..., t_{ar(f_i)})) = F_i(v(t_1), ..., v(t_{ar(f_i)}))$ dla $i \leq s$.

Relacja spełniania $\models$ jest relacją pomiędzy modelami, formułami i wartościowaniami. Dla danych $\mathcal{A} \in \mathrm{Mod}_\sigma$, $\varphi(x_{i_1}, ..., x_{i_k}) \in \mathcal{F}_\sigma$ oraz wartościowania v przez indukcję definiujemy pojęcie $\mathcal{A} \models \varphi(x_{i_1}, ..., x_{i_k})[v]$ spełniania formuły $\varphi(x_{i_1}, ..., x_{i_k})$ w modelu $\mathcal{A}$ przez wartościowanie v:

- $\mathcal{A} \models (t = t')[v]$, gdy $v(t) = v(t')$,
- $\mathcal{A} \models P_i(t_1, ..., t_{ar(P_i)})[v]$, gdy $(v(t_1), ..., v(t_{ar(P_i)})) \in R_i$,
- $\mathcal{A} \models \neg\varphi[v]$, gdy $\mathcal{A} \not\models \varphi[v]$,
- $\mathcal{A} \models \varphi \vee \psi[v]$, gdy $\mathcal{A} \models \varphi[v]$ lub $\mathcal{A} \models \psi[v]$,
- $\mathcal{A} \models \forall x\varphi[v]$, gdy dla dowolnego $b \in |\mathcal{A}|$ $\mathcal{A} \models \varphi[v(x/b)]$.

1.2 Klasyczna definicja prawdy i twierdzenie Tarskiego

W 1933 roku Alfred Tarski w swojej pracy *Pojęcie prawdy w językach nauk dedukcyjnych* pokazał, jak można w naukowy sposób uprawiać semantykę. Był to wielki przełom w sposobie patrzenia na język nauki, zdominowany wówczas propagowanym przez Rudolfa Carnapa podejściem, polegającym na badaniu wyłącznie składni języka. Wynikało to w dużej mierze z powszechnego przekonania, że badania nad semantyką w istotny sposób wykraczają poza nasze możliwości poznawcze, i że jedynym aspektem języka dającym się sensownie opisywać jest składnia, ponieważ podczas rozważań na poziomie semantycznym wikłamy się w antynomie — takie jak na przykład słynne antynomialne zdanie Eubulidesa: *Ja teraz kłamię*. Praca Tarskiego doprowadziła do rewizji sposobu patrzenia na badania nad językiem, kładąc fundament dla badań z dziedziny semantyki, których pierwszym ważnym wynikiem jest właśnie twierdzenie Tarskiego o niedefiniowalności prawdy.

W 1931 roku Kurt Gödel podał przykład prawdziwego zdania arytmetycznego, które nie może być dowiedzione na gruncie systemu z Principia Mathematica, oraz którego negacja również jest w nim niedowodliwa. Oczywiście twierdzenie o niezupełności nie odnosi się jedynie do tego jednego systemu formalnego. Analizując klasyczny dowód twierdzenia o niezupełności arytmetyki można wyabstrahować warunki wystarczające, by móc przeprowadzić ten dowód. Jednym z ważniejszych lematów, z których Gödel korzysta jest — wyodrębniony po raz pierwszy przez Carnapa — lemat przekątniowy, pozwalający konstruować zdania samoodnoszące, dla dowolnej formuły z co najmniej jedną zmienną wolną. To właśnie możliwość wykorzystania samoodniesienia jest jednym z kluczowych składników potrzebnych do dowodu obu omawianych tutaj twierdzeń. Gödel pokazał, że w standardowej arytmetyce można udowodnić lemat przekątniowy. Podamy teraz pewną wersję tego lematu.

Twierdzenie 1. *(Lemat przekątniowy) Niech $\mathcal{K}$ będzie wystarczająco silną klasą formuł arytmetycznych oraz niech $\varphi(x) \in \mathcal{K}$ będzie formułą z jedną zmienną wolną x. Wtedy istnieje zdanie $\psi \in \mathcal{K}$ takie, że $\mathcal{N} \models \psi \equiv \varphi(\ulcorner\psi\urcorner)$.*

Powyższe twierdzenie niewiele mówi, dopóki nie wyjaśnimy co znaczy *wystarczająco silna klasa formuł arytmetycznych*, oraz w jaki sposób można nazywać zdania wewnątrz języka (klasy formuł), co oznaczaliśmy powyżej za pomocą symboli $\ulcorner\,\urcorner$. Gödel wymyślił sposób w jaki można zarytmetyzować język, czyli kodować (nazywać) formuły za pomocą liczb oraz pokazał, że gdy możemy skonstruować w języku formuły $\mathbf{Name}(x)$ oraz $\mathbf{Subst}_0(x, y)$ takie, że $\mathbf{Name}(x) = y$ wyraża fakt, że y jest nazwą liczby x oraz $\mathbf{Subst}_0(x, y) = z$ gdy z jest nazwą formuły powstającej z formuły o nazwie x przez podstawienie w miejsce pierwszej zmiennej termu o nazwie y, wtedy też możemy udowodnić lemat przekątniowy. Idea zawarta w dowodzie polega właśnie na próbie rekonstrukcji samoodnoszenia wyrażeń języka — tak jak w przypadku, gdy rozważamy zdanie: *Ja teraz kłamię* — wypowiadając to zdanie twierdzimy coś o naszej aktualnej wypowiedzi.

Drugą istotną własnością języka wykorzystywaną w dowodzie twierdzenia Tarskiego jest domkniętość na negację. Zauważmy, że łatwo możemy skonstruować zdania samoodnoszące, które nie są antynomialne, jak na przykład: *To zdanie składa się z siedmiu słów.* Zdanie to jest po prostu prawdziwe, a jego wartość logiczna nie budzi naszych wątpliwości, mimo iż odnosi się ono samo do siebie. Konieczność domknięcia na negację rozważanego języka jest istotna — antynomia Eubulidesa opiera się na zdaniu, które orzeka o sobie, że **nie** jest prawdzie, zdanie Gödla orzeka o sobie, że **nie** jest dowodliwe w danym systemie formalnym. Przedstawimy teraz twierdzenie Tarskiego o niedefiniowalności prawdy.

Twierdzenie 2. *(Tarskiego o niedefiniowalności prawdy) Dla dowolnej klasy formuł arytmetycznych $\mathcal{K}$ zawierającej Δ^0_0 oraz domkniętej na konstrukcje pierwszego rzędu — w tym na negację — nie istnieje formuła $\varphi(x) \in \mathcal{K}$ z jedną zmienną wolną x taka, że dla dowolnego zdania $\psi \in \mathcal{K}$ zachodzi $\mathcal{N} \models \psi \equiv \varphi(\ulcorner\psi\urcorner)$.*

Zauważmy, że to co nazywaliśmy wystarczającą siłą wyrazu teorii zostało wyeksplikowane jako $\Delta^0_0 \subseteq \mathcal{K}$ — faktycznie zarówno formułę $\mathbf{Name}(x)$ jak i $\mathbf{Subst}_0(x, y)$ możemy zdefiniować jako Δ^0_0–formuły. Uwzględniona została też konieczność domknięcia na negację klasy $\mathcal{K}$ — zauważmy, że klasy formuł Σ^0_n i Π^0_n zawierają Δ^0_0 dla dowolnego $n \geq 1$, jednak mieszczą w sobie własne definicje prawdy — nie są zatem domknięte na negację.

Twierdzenie Tarskiego dostarcza nam również bardzo użytecznego narzędzia do klasyfikacji języków pod względem ich siły semantycznej. Pokazuje ono, że gdy staramy się określić semantykę, dla pewnego języka bazowego $\mathcal{L}_0$ musimy posługiwać się pewnym silniejszym językiem $\mathcal{L}_1$, oraz że $\mathcal{L}_0$ nie może wystarczać do tego celu. Z drugiej strony, gdy rozważając siłę semantyczną dwóch różnych języków $\mathcal{L}_0$ i $\mathcal{L}_1$ skonstruujemy w $\mathcal{L}_1$ definicję prawdy dla zdań z $\mathcal{L}_0$, natychmiastowo możemy wywnioskować, że siła semantyczna $\mathcal{L}_1$ musi znacząco przewyższać siłę semantyczną $\mathcal{L}_0$.

1.3 Dowód twierdzenia Tarskiego o niedefiniowalności prawdy arytmetycznej

Na początku udowodnimy odpowiednią wersję lematu przekątniowego.

Twierdzenie 3. *Niech $\mathcal{K}$ będzie klasą wszystkich formuł arytmetycznych pierwszego rzędu oraz niech $\varphi(x) \in \mathcal{K}$ będzie formułą z jedną zmienną wolną x. Wtedy istnieje zdanie $\psi \in \mathcal{K}$ takie, że $\mathcal{N} \models \psi \equiv \varphi(\ulcorner\psi\urcorner)$, gdzie $\mathcal{N}$ jest standardowym modelem arytmetyki.*

Dowód. Ustalmy $\varphi(x)$ formułę aytmetyczną z jedną zmienną wolną x. W klasie formuł $\mathcal{K}$ mamy funkcje $\mathbf{Name}(x)$ oraz $\mathbf{Subst}_0(x, y)$ o następujących własnościach:

- $\mathcal{N} \models \mathbf{Name}(n) = \ulcorner\underbrace{s(...s(}_{n}0)...)\urcorner$,
- $\mathcal{N} \models \mathbf{Subst}_0(\ulcorner\varphi(x)\urcorner, \ulcorner t\urcorner) = \ulcorner\varphi(t)\urcorner$,

gdzie $\ulcorner\ \urcorner$ jest funkcją zwracającą kod dowolnego ciągu symboli arytmetycznych, s jest funkcją następnika, t jest dowolnym termem arytmetycznym. Wprowadzimy teraz kilka pomocniczych definicji:

- $\zeta(x) \equiv_{df} \varphi(\mathbf{Subst}_0(x, \mathbf{Name}(x)))$,
- $m = \ulcorner\zeta(x)\urcorner$,
- $\psi \equiv_{df} \zeta(m)$.

Następujące wyrażenia są równoważne w $\mathcal{N}$:

- ψ,
- $\zeta(m)$,
- $\zeta(\ulcorner\zeta(x)\urcorner)$,
- $\varphi(\mathbf{Subst}_0(\ulcorner\zeta(x)\urcorner, \mathbf{Name}(m)))$,
- $\varphi(\ulcorner\zeta(m)\urcorner)$
- $\varphi(\ulcorner\psi\urcorner)$.

□

Posiadając lemat przekątniowy z łatwością dowodzimy twierdzenia Tarskiego o niedefiniowalności prawdy arytmetycznej.

Twierdzenie 4. *Nie istnieje formuła arytmetyczna $\varphi(x)$ z jedną zmienną wolną x taka, że dla dowolnego zdania arytmetycznego ψ zachodzi $\mathcal{N} \models \varphi(\ulcorner\psi\urcorner) \equiv \psi$.*

Dowód. Załóżmy przeciwnie, że istnieje formuła $\varphi(x)$, o własnościach jak w twierdzeniu. Zastosujemy lemat przekątniowy do formuły $\neg\varphi(x)$ - niech zdanie ψ_0 będzie zdaniem przekątniowym, czyli $\mathcal{N} \models \psi_0 \equiv \neg\varphi(\ulcorner\psi_0\urcorner)$. Z drugiej jednak strony, z założenia wiadomo, że $\varphi(x)$ jest definicją prawdy dla zdań arytmetycznych, mamy więc w szczególności dla zdania ψ_0 $\mathcal{N} \models \psi_0 \equiv \varphi(\ulcorner\psi_0\urcorner)$. Otrzymujemy zatem $\mathcal{N} \models \psi_0 \equiv \neg\psi_0$ – co jest niemożliwe. $\square$

2 Definicje prawdy

2.1 Definicja prawdy dla arytmetyki pierwszego rzędu

Pokażemy teraz w jaki sposób w arytmetyce drugiego rzędu zdefiniować prawdę dla wszystkich zdań arytmetycznych pierwszego rzędu.

Ustalmy pewne kodowanie składni arytmetycznej. Formuły rozumiemy jako (zbudowane w odpowiedni sposób) ciągi nad alfabetem złożonym z symboli: $s, +, \times, (,), \neg, \Rightarrow, \forall, v, ', =, 0$. Dla dowolnego symbolu α niech $\overline{\alpha}$ będzie kodem α (pewną liczbą naturalną). Skończone ciągi nazywamy w nastepujący sposób: $\ulcorner\alpha_0, ..., \alpha_n\urcorner = p_0^{\overline{\alpha_0}+1} \dots p_n^{\overline{\alpha_n}+1}$. Kodowanie, które otrzymujemy jest różnowartościowe.

Niech $\mathbf{value}(x)$ będzie funkcją częściową, określoną jedynie dla nazw termów zamkniętych, zwracającą ich wartość. Niech $\mathbf{Form}(x)$ będzie predykatem prawdziwym o nazwach poprawnie zbudowanych formuł arytetycznych i niech $\mathbf{Form}_0(x)$ będzie predykatem prawdziwym o nazwach formuł zamkniętych. Niech $\mathbf{Term}(x)$ będzie predykatem prawdziwym o nazwach termów oraz niech $\mathbf{Var}(x)$ będzie predykatem prawdziwym o nazwach zmiennych. Zbiory definiowane przez powyższe predykaty są pierwotnie rekurencyjne, zatem wyrażalne w arytmetyce pierwszego rzędu. Trójargumentowa funkcja $\mathbf{Subst}(x, y, z)$ działa podobnie do rozważanej wcześniej funkcji $\mathbf{Subst}_0(x, z)$, z taką różnicą, że $\mathbf{Subst}(x, y, z)$ zwraca kod formuły będącej podstawieniem o nazwie z do formuły o nazwie x za zmienną o nazwie y. Symbolem $*$ oznaczamy kontatenację ciągów.

Opiszemy teraz formułę $\Phi(P)$ z dodatkowym predykatem P. Naszą intencją jest wyrażenie faktu, że P jest zbiorem (kodów) zdań prawdziwych pierwszego rzędu.

$$\begin{aligned}
&\forall x(P(x) \Rightarrow \mathbf{Form}_0(x))\wedge\\
&\quad \forall x\forall y((\mathbf{Term}(x) \wedge \mathbf{Term}(y) \wedge \mathbf{Form}_0(x * \overline{\equiv} * y)) \Rightarrow\\
&\quad (P(x * \overline{\equiv} * y) \equiv \mathbf{value}(x) = \mathbf{value}(y)))\wedge\\
&\quad \forall x(\mathbf{Form}_0(x) \Rightarrow (P(\overline{\neg} * x) \equiv \neg P(x)))\wedge\\
&\quad \forall x\forall y((\mathbf{Form}_0(x) \wedge \mathbf{Form}_0(y)) \Rightarrow\\
&\quad (P(\overline{(} * x * \overline{\Rightarrow} * y * \overline{)}) \equiv (P(x) \Rightarrow P(y))))\wedge\\
&\quad \forall x\forall y((\mathbf{Form}(x) \wedge \mathbf{Form}_0(\overline{\forall} * y * x) \wedge \mathbf{Var}(y)) \Rightarrow\\
&\quad (P(\overline{\forall} * y * x) \equiv \forall z P(\mathbf{Subst}(x, y, \mathbf{Name}(z))))).
\end{aligned}$$

Przez indukcję ze względu na długość ciągów definiujących formuy możemy pokazać, że:

$$\mathcal{N} \models \forall P \forall P'((\Phi(P) \wedge \Phi(P')) \Rightarrow \forall x(P(x) \equiv P'(x))).$$

Zatem istnieje co najwyżej jeden zbiór $\mathbf{P} \subset \mathbb{N}$ taki, że gdy zinterpretujemy P jako nazwę $\mathbf{P}$, to formuła $\Phi(P)$ będzie prawdziwa w $\mathcal{N}$. Istnienie takiego zbioru $\mathbf{P}$ pokazujemy przez przykład:

$$\mathbf{P} = \{\ulcorner \varphi \urcorner : \varphi \, \text{jest zdaniem pierwszego rzędu i } \mathcal{N} \models \varphi\}.$$

Możemy zatem zdefiniować na dwa sposoby predykat arytmetyczny $\mathrm{Tr}(x)$ prawdziwości dla zdań arytmetycznych pierwszego rzędu.

- $\forall x(\mathrm{Tr}(x) \equiv \exists P(\Phi(P) \wedge P(x)))$,
- $\forall x(\mathrm{Tr}(x) \equiv \forall P(\Phi(P) \Rightarrow P(x)))$.

Przez indukcję ze względu na długość ciągu definiującego zdanie φ możemy udowodnić następujące twierdzenie:

Twierdzenie 5. *Dla dowolnego zdania arytmetycznego pierwszego rzędu φ zachodzi $\mathcal{N} \models \varphi \equiv \mathrm{Tr}(\ulcorner \varphi \urcorner)$.*

Zatem formuła $\mathrm{Tr}(x)$ jest definicją prawdy dla zdań arytmetycznych pierwszego rzędu.

2.2 Klasy formuł zawierające własne definicje prawdy

Wprowadzimy najpierw kilka oznaczeń. Przez Δ_0^0 (również Σ_0^0 i Π_0^0) oznaczymy klasę formuł arytmetyki dodawania i mnożenia, w których wszystkie kwantyfikatory są ograniczone. Formuły Σ_n^0 są następującej postaci: $\exists x_1 \forall x_2 ... Q x_n \, \varphi(x_1, ..., x_n, y_1, ..., y_k)$, gdzie Q jest $\forall$ dla n parzystych, a $\exists$ dla n nieparzystych oraz $\varphi(x_1, ..., x_n, y_1, ..., y_k)$ jest formułą Δ_0^0. Podobnie definiujemy klasę Π_n^0, z tą różnicą, że pierwszy kwantyfikator jest ogólny.

Na mocy twierdzenia o preneksowej postaci normalnej oraz możliwości wyrażania rzutowań dla par uporządkowanych za pomocą formuł Δ_0^0 wiemy, że każda formuła arytmetyczna równoważna jest w modelu standardowym pewnej formule Σ_n^0.

Podobnych oznaczeń używamy dla klasyfikowania relacji arytmetycznie definiowalnych (definiowalnych za pomocą formuł pierwszego rzędu z dodawaniem i mnożeniem). Mówimy, że relacja jest Σ_n^0 (Π_n^0), dla $n \geq 1$, jeżeli jest definiowalna w modelu standardowym przez pewną formułę Σ_n^0 (Π_n^0). Dla $n \geq 1$ relacja jest Δ_n^0 dokładnie wtedy, gdy jest zarazem Σ_n^0 oraz Π_n^0. W szczególności wiadomo, że Σ_1^0 są to dokładnie relacje rekurencyjnie przeliczalne. Relacje Δ_1^0 są to dokładnie relacje rekurencyjne. Relacje Δ_2^0 określa się też jako relacje rekurencyjne w granicy. W szczególności wiadomo, że zbiór relacji Δ_1^0 zawarty jest w zbiorze relacji Σ_1^0 (każda relacja rekurencyjna jest rekurencyjnie przeliczalna), wszystkie relacje Σ_1^0 lub Π_1^0 są również Δ_2^0.

Najpierw pokażemy w jaki sposób możemy wyrazić prawdę dla formuł Δ_0^0 za pomocą formuły Σ_1^0 (a również Π_1^0), czyli pokażemy, że w klasach Σ_1^0 oraz Π_1^0 są definicje prawdy dla Δ_0^0.

Z dowolną formułą $\psi \in \Delta_0^0$ możemy związać jej drzewo składniowe. Drzewo to jest skończone i możemy badać jego własności w języku arytmetyki, czyli: jednoznacznie kodować drzewa, definiować wierzchołki, liście, korzeń, bycie potomkiem oraz bycie przodkiem, itp. Wszystkie powyższe pojęcia związane z drzewami możemy wyrażać za pomocą formuł Δ_0^0. Na podstawie drzewa składniowego budujemy indukcyjnie względem budowy formuły kolejne drzewo. Modyfikujemy drzewo syntaktyczne ψ w następujący sposób:

- Dla spójników logicznych pozostawiamy drzewo bez zmian,
- Dla kwantyfikatorów organiczonych $\forall_{x<n}\varphi(x,\overline{y})$ oraz $\exists_{x<n}\varphi(x,\overline{y})$ kładziemy odpowiedni kwantyfikator jako etykietę wierzchołka oraz tworzymy n potomków – $\varphi(0,\overline{y}),\ldots,\varphi(n-1,\overline{y})$.

Liście uzyskanego w ten sposób drzewa to zamknięte formuły atomowe. Tworzymy kopię uzyskanego drzewa zmieniając jedynie etykiety. Etykiety liści są zerami i jedynkami w zależności od prawdziwości zdań, które są etykietami odpowiadających im liści pierwszego drzewa. Prawdziwość całego zdania zależy od etykiety korzenia drugiego drzewa, którą poznamy etykietując w odpowiedni sposób całe drugie drzewo zerami i jedynkami. Predykat $\mathrm{Tr}^{\Sigma_1^0}_{\Delta_0^0}$, czyli Σ_1^0–definicją prawdy dla zdań z Δ_0^0 definiujemy następująco.
$\mathrm{Tr}^{\Sigma_1^0}_{\Delta_0^0}(x)$ dokładnie wtedy, gdy x jest Δ_0^0 zdaniem i istnieją drzewa y oraz z, które spełniają odpowiednie warunki. Warunki te są czysto algorytmiczne – dla pierwszego drzewa wynikają z jego indukcyjnej konstrukcji, dla drugiego drzewa natomiast wynikają z warunków Tarskiego w definicji spełniania.

Analogicznie defniujemy predykat $\mathrm{Tr}^{\Pi_1^0}_{\Delta_0^0}$. $\mathrm{Tr}^{\Pi_1^0}_{\Delta_0^0}(x)$ dokładnie wtedy, gdy x jest Δ_0^0 zdaniem i dla dowolnych drzew y oraz z, jeżeli spełniają one odpowiednie warunki, to korzeń drugiego drzewa ma etykietę 1.

Ponieważ wszystkie pojęcia potrzebne do ścisłego zapisu powyższych definicji można wyrazić za pomocą Δ_0^0 formuł otrzymujemy odpowiednio (po sprowadzeniu do postaci normanej) $\mathrm{Tr}^{\Sigma_1^0}_{\Delta_0^0}(x) \equiv_{df} \exists y\exists z\varphi_1(x,y,z)$ oraz $\mathrm{Tr}^{\Pi_1^0}_{\Delta_0^0}(x) \equiv_{df} \forall y\forall z\varphi_2(x,y,z)$, gdzie $\varphi_1,\varphi_2 \in \Delta_0^0$.

Pokażemy teraz, że klasy Σ_1^0 oraz Π_1^0 zawierają definicje własne prawdy.

Wiadomo, że można efektywnie sprowadzać formuły do postaci normalnej. Niech $f : \Sigma_1^0 \to \Sigma_1^0$ będzie funkcją przyporządkowującą formułom Σ_1^0 ich postaci normalne (aby f była funkcją przyjmujemy, że f zwraca formułę o najmniejszym kodzie Gödla) takie, że $f(\theta) = \exists v_0\varphi$, gdzie $\varphi \in \Delta_0^0$. Ponieważ f jest rekurencyjna (Δ_1^0) istnieje Σ_1^0 formuła $\psi(x,y)$ reprezentująca f. Niech $\psi(x,y)$, wtedy przez $\hat{y}$ oznaczymy Δ_0^0–formułę taką, że $y = \exists v_0\hat{y}$. Podobnie dla funkcji $f' : \Pi_1^0 \to \Pi_1^0$ sprowadzającej do postaci normalnej formuły Π_1^0 istnieje Σ_1^0 formuła $\psi'(x,y)$ reprezentująca f'. Tu również przez $\hat{y}$ oznaczymy kod formuły powstającej z formuły o kodzie y przez pominięcie kwantyfikatora.
Definiujemy $\mathrm{Tr}_{\Sigma_1^0}(x) = \exists y(\psi(x,y) \wedge \exists z\mathrm{Tr}^{\Sigma_1^0}_{\Delta_0^0}(\mathbf{Subst}_0(\hat{y},\mathbf{Name}(z)))$.
Analogicznie $\mathrm{Tr}_{\Pi_1^0}(x) = \forall y(\psi'(x,y) \Rightarrow \forall z\mathrm{Tr}^{\Pi_1^0}_{\Delta_0^0}(\mathbf{Subst}_0(\hat{y},\mathbf{Name}(z)))$.
Po sprowadzeniu do postaci normalnych łatwo zauważyć, że $\mathrm{Tr}_{\Sigma_1^0}(x) \in \Sigma_1^0$ oraz $\mathrm{Tr}_{\Pi_1^0}(x) \in \Pi_1^0$

Za pomocą indukcji możemy pokazać, że dla $n > 1$ Σ_n^0 oraz Π_n^0 zawierają własne definicje prawdy. Z twierdzenia o postaci normalnej wiemy, że każdą

formułę Σ^0_{n+1} (Π^0_{n+1}) można (rekurencyjnie) sprowadzić do postaci $\exists x\varphi_1$ ($\forall x\varphi_2$), gdzie $\varphi_1 \in \Pi^0_n$ ($\varphi_2 \in \Sigma^0_n$), funkcję tłumaczącą do postaci normalnej reprezentuje formuła ψ_n (ψ'_n). Założenie idukcyjne to: istnieje Σ^0_n definicja prawdy $\mathrm{Tr}_{\Sigma^0_n}$ dla zdań z Σ^0_n oraz istnieje Π^0_n definicja prawdy $\mathrm{Tr}_{\Pi^0_n}$ dla zdań z Π^0_n. Niech $\hat{y}$ będzie kodem formuły powstającej z formuły o kodzie y przez usunięcie pierwszego kwantyfikatora. Mamy następujące definicje prawdy:
$\mathrm{Tr}_{\Sigma^0_{n+1}}(x) = \exists y(\psi_n(x,y) \wedge \exists z \mathrm{Tr}_{\Pi^0_n}(\mathbf{Subst}_0(\hat{y}, \mathbf{Name}(z))),$
$\mathrm{Tr}_{\Pi^0_{n+1}}(x) = \forall y(\psi'(x,y) \Rightarrow \forall z \mathrm{Tr}_{\Sigma^0_n}(\mathbf{Subst}_0(\hat{y}, \mathbf{Name}(z))).$

Inne przykłady klas, w których ograniczona została możliwość wyrażania negacji to klasy Σ^1_1, Π^1_1, czy EFPL (Existential Fixed-Point Logic). Σ_1 to klasa formuł drugiego równoważnych logicznie formułom postaci $\exists f_1 ... \exists f_n \varphi$, gdzie φ jest pierwszego rzędu, a $f_1, ..., f_n$ są zmiennymi drugiego rzędu (Π^1_1 definiujemy analogicznie zamieniając kwantyfikatory na ogólne). EFPL jest rozszerzeniem EL (Existential Logic), w której negację można stosować jedynie do formuł atomowych, o operator najmniejszego punktu stałego. Wszystkie powyższe klasy zawierają własne definicje prawdy.

3 Prawda w modelach skończonych

Definicja 1. *Niech $\mathcal{K}$ będzie pewną klasą formuł ustalonego słownika σ. Powiemy, że σ–formuła $\varphi(x)$ jest FM–definicją prawdy dla $\mathcal{K}$, gdy dla dowolnego zdania $\psi \in \mathcal{K}$ i dla prawie wszystkich (wszystkich poza skończoną liczbą) skończonych σ–modeli M zachodzi:*

$$M \models \psi \equiv \varphi(\ulcorner \psi \urcorner)$$

Niech $R \subseteq \omega^r$ będzie relacją arytmetyczną, wtedy $R^{(n)} = R \cap \{0, ..., n-1\}^n$.
W naszych rozważaniach ograniczamy się do skończonych modeli (o słownikach relacyjnych), których uniwersa są segmentami początkowymi liczb naturalnych. Dla dowolnego modelu $\mathcal{A}$ ustalonego słownika relacyjnego $\sigma = \{R_1, ..., R_k\}$ definiujemy FM–dziedzinę $\mathrm{FM}(\mathcal{A}) = \{\mathcal{A}_n : n = 0, 1, 2, ...\}$, gdzie $\mathcal{A}_n = (\{0, ..., n-1\}, R^{(n)}_1, ..., R^{(n)}_k)$.

Wprowadzimy następujące oznaczenie:
$\mathrm{FM}(\mathcal{A}) \models_{sl} \psi$ wtedy i tylko wtedy, gdy $\exists k \forall n > k\, \mathcal{A}_n \models \psi$.
O zdaniu ψ powiemy w takim wypadku, że jest prawdziwe w FM–dziednie $\mathrm{FM}(\mathcal{A})$ (prawdziwe w dostatecznie dużych modelach z $\mathrm{FM}(\mathcal{A})$).

Definicja 2. *Niech $R \subseteq \omega^k$ będzie relacją arytmetyczną. Powiemy, że R jest FM–reprezentowana przez formułę $\varphi(x_1, ..., x_k)$ w FM–dziedzinie* $\mathrm{FM}(\mathcal{A})$ *wtedy i tylko wtedy, gdy dla dowolnych $a_1, ..., a_k \in \omega$ spełnione są następujące warunki:*

1. $\mathrm{FM}(\mathcal{A}) \models_{sl} \varphi[a_1, ..., a_k]$ *wtedy i tylko wtedy, gdy* $R(a_1, ..., a_k)$,

2. $\mathrm{FM}(\mathcal{A}) \models_{sl} \neg\varphi[a_1, ..., a_k]$ *wtedy i tylko wtedy, gdy* $\neg R(a_1, ..., a_k)$.

Powiemy, że R jest FM–reprezentowalna w $\mathrm{FM}(\mathcal{A})$, jeżeli istnieje formuła φ, która FM–reprezentuje R w $\mathrm{FM}(\mathcal{A})$.

Twierdzenie 6. *(Twierdzenie o FM–reprezentowalności dla $\mathrm{FM}(\mathcal{N})$) Niech R będzie relacją arytmetyczną. R jest FM–reprezentowalna w $\mathrm{FM}(\mathcal{N})$ wtedy i tylko wtedy, gdy $R \in \Delta^0_2$.*

Powyższe twierdzenie mówi nam, o jakich relacjach arytmetycznych można sensownie mówić w modelach skończonych. Z naszego punktu widzenia Δ^0_2 jest dużym zbiorem – równoważnie zawiera on relacje stopnia $\leq 0'$, czy rekurencyjne z rekurencyjnie przeliczalną wyrocznią. Jak wspominaliśmy funkcje $\mathbf{Name}(x)$ oraz $\mathbf{Subst}_0(x, y)$ są definiowalne przy pomocy Δ^0_0 formuł – zatem również należą do klasy Δ^0_2.

Twierdzenie 7. *Nie istnieje formuła arytmetyczna $\varphi(x)$ z jedną zmienną wolną taka, że dla dowolnego zdania arytmetycznego ψ zachodzi $\mathrm{FM}(\mathcal{N}) \models_{sl} \psi \equiv \varphi(\ulcorner\psi\urcorner)$.*

Powyższe twierdzenie to FM–wersja twierdzenia Tarskiego o niedefiniowalności prawdy arytmetycznej. Jego dowód jest analogiczny do klasycznego: najpierw dowodzimy lematu przekątniowego w wersji skończonomodelowej, czyli że dla dowolnej formuły $\varphi(x)$ z jedną zmienną wolną x istnieje zdanie ψ takie, że $\mathrm{FM}(\mathcal{N}) \models_{sl} \psi \equiv \varphi(\ulcorner\psi\urcorner)$. Ponieważ $\mathbf{Name}(x)$ i $\mathbf{Subst}(x, y)$ są FM–reprezentowalne, od pewnego miejsca (dla dostatecznie dużych modeli) ich wartość logiczna się ustala (i jest zgodna z ich wartością logiczną w modelu standardowym) – stąd dostajemy tezę. Przy pomocy skończonej wersji lematu przekątniowego łatwo dowieść skończonej wersji Twierdzenia Tarskiego.

Chcemy pokazać, że w przeciwieństwie do modelu standardowego, już względna pierwszość wystarczy, by wyrażać własności semantyczne – i wystarczy, by udowodnić odpowiednią wersję twierdzenia o niedefiniowalności prawdy. W modelu standardowym arytmetyka mnożenia jest rozstrzygalna. To pociąga rozstrzygalność arytmetyk podzielności i względnej pierwszości. W teoriach tych nie da się również zdefiniować predykatu prawdy – ich języki są zbyt ubogie. Czego brakuje zatem, by otrzymać pełną moc arytmetyki (bądź przynajmniej wystarczającą, by mówić o semantyce)? Odpowiedź brzmi – porządku. Przy pomocy mnożenia, w modelu standardowym, nie możemy zdefinować porządku. Co zatem dają nam modele skończone? Rozważmy następującą definicję: $a \prec b \equiv \exists c\, a \times b \wedge \neg\exists c\, c \times b$. Wykorzystujemy fakt, że pewne elementy nie mieszczą się w modelu ustalonej wielkości, by odtworzyć porządek na pewnym fragmencie początkowym tego modelu. Fragment ten rośnie nieograniczenie, gdy zwiększamy rozpatrywane modele.

Trochę trudniejszy trick da nam interpretację FM–dziedziny pełnej arytmetyki $\mathrm{FM}(\mathcal{N})$ w $\mathrm{FM}((\omega, \perp))$. Interpretację buduje się na indekach liczb pierwszych. Co prawda 2 nie jest definiowalne w języku względnej pierwszości (bo nie sposób odróżnić 2 od 4, czy na przykład 8), jednak następujące dwa lematy pokazują o czym możemy powiedzieć w tym języku.

Lemat 1. *Następujące formuły są definiowalne w języku względnej pierwszości:*

- $P(x) \equiv_{df} \forall y \forall z\, (\neg z \perp x \wedge \neg y \perp x \Rightarrow \neg z \perp y)$ *(x jest potęgą liczby pierwszej),*

- $\{p, q\} \equiv_{df} \forall z\, (z \perp a \equiv (z \perp p \wedge z \perp q))$ *funkcja zdefiniowana dla par liczb p i q, zwracająca jako wartość klasę elementu* $a \approx pq$.

Następujący lemat pokazuje wyrażalność w języku z samym predykatem $\perp$ następujących predykatów, pozwalających na operowanie na klasach równoważności relacji $\approx$.

Lemat 2. *Istnieją formuły* $\varphi_{\cup}(x, y, z)$, $\varphi_{\cap}(x, y, z)$ *oraz* $\varphi_{-}(x, y, z)$ *spełniające w dowolnym modelu względnej pierwszości M następujące warunki:*

- $M \models \varphi_{\cup}[a, b, c]$ *wtedy i tylko wtedy, gdy* $Supp(a) \cup Supp(b) = Supp(c)$,

- $M \models \varphi_{-}[a, b, c]$ *wtedy i tylko wtedy, gdy* $Supp(a) - Supp(b) = Supp(c)$,

- $M \models \varphi_{\cap}[a, b, c]$ *wtedy i tylko wtedy, gdy* $Supp(a) \cap Supp(b) = Supp(c)$,

dla dowolnych $a, b, c \in |M|$.

Możemy teraz zdefiniować bardzo ważną dla naszych potrzeb relację $\prec$, którą wyrażać będzie następująca formuła:

$$\varphi_{\prec}(x, y) \equiv_{df} \exists z\, (P(z) \wedge x \perp z \wedge y \perp z \wedge \exists w\, \varphi_{\cup}(x, z, w) \wedge \neg \exists w\, \varphi_{\cup}(y, z, w))$$

Jest to szukany przez nas porządek na klasach indeksów liczb pierwszych. Porządek ten pozwala zdefiniować następnik. Następnie dowodzi się twierdzenia, że formuły FM–reprezentowalne w $\mathrm{FM}((\omega, \perp))$, gdy do języka dołożymy pseudoporządek $\prec$ domknięte są na rekursję prostą. Otrzymujemy dodawanie i mnożenie na klasach indeksów liczb pierwszych. Dodatkowo pokazujemy, że jeżeli R jest relacją arytmetyczną to jej tłumaczenie na indeksy liczb pierwszych R^* jest FM–reprezentowalna w $\mathrm{FM}((\omega, \perp))$, dokładnie wtedy, gdy R była FM–reprezentowalna w $\mathrm{FM}(\mathcal{N})$.

Biorąc teraz dowolne kodowanie Gödla $\ulcorner\ \urcorner$ formuł arytmetycznych (dodatkowo z $\perp$) i definiując $\mathrm{GN}(x) = p_{\ulcorner x \urcorner}$ otrzymujemy FM–wersję twierdzenia Tarskiego o niedefiniowalności prawdy dla $\mathrm{FM}((\omega, \perp))$.

Twierdzenie 8. *(FM–wersja lematu przekątniowego dla* $\mathrm{FM}((\omega, \perp))$*) Dla dowolnej formuły w języku względnej pierwszości* $\varphi(x)$ *z jedną zmienną wolną* x *istnieje zdanie* ψ *spełniające następujący warunek:*

$$\mathrm{FM}((\omega, \perp)) \models_{sl} \psi \equiv \varphi(\mathrm{GN}(\psi)).$$

Dowód. Pokazaliśmy, że w $\mathrm{FM}(\mathcal{N})$ FM–reprezentowalne są funkcje $\mathbf{Name}(x)$ oraz $\mathbf{Subst}(x, z)$. Wiemy również, że dzięki istnieniu interpretacji skończonych modeli arytmetyki dodawania i mnożenia w skończonych modelach względnej pierwszości, w $\mathrm{FM}((\omega, \perp))$ FM–reprezentowalne są $\mathbf{Name}^*(x)$ oraz $\mathbf{Subst}^*(x, z)$ — tłumaczenia tych funkcji na klasy równoważności relacji $\approx$ liczb pierwszych.
Jeżeli $\mathrm{GN}(\psi)$ jest kodem formuły ψ, takim jaki zdefiniowaliśmy w jednym z wcześniejszych punktów niniejszej pracy: $\mathrm{GN}(\psi) = p_{\ulcorner \psi \urcorner}$, to $\mathbf{Name}^*(x)$ oraz $\mathbf{Subst}^*(x, z)$ spełniają następujące warunki:

- $\mathrm{FM}((\omega,\perp)) \models_{sl} \mathbf{Subst}^*(\mathrm{GN}(\varphi(x)), \mathrm{GN}(t)) = \mathrm{GN}(\varphi(t))$
- $\mathrm{FM}((\omega,\perp)) \models_{sl} \mathbf{Name}^*(x) = \mathrm{GN}(\underbrace{s(s...s}_{x}(0)))$

Przedstawimy teraz następujące pomocnicze definicje:
Niech $\zeta(x) \equiv_{df} \varphi(\mathbf{Subst}^*(x, \mathbf{Name}^*(x)))$,
Niech $m =_{df} \mathrm{GN}(\zeta(x))$,
Niech $\psi \equiv_{df} \zeta(m)$.
Chcemy pokazać, że:

$$\mathrm{FM}((\omega,\perp)) \models_{sl} (\psi \equiv \varphi(\mathrm{GN}(\psi))).$$

W dostatecznie dużych modelach z $\mathrm{FM}((\omega,\perp))$ równoważne są następujące formuły:

- ψ
- $\zeta(m)$
- $\zeta(\mathrm{GN}(\zeta(x)))$
- $\varphi(\mathbf{Subst}^*(\mathrm{GN}(\zeta(x)), \mathbf{Name}^*(\mathrm{GN}(\zeta(x)))))$
- $\varphi(\mathrm{GN}(\zeta(\mathbf{Name}^*(\mathrm{GN}(\zeta(x))))))$
- $\varphi(\mathrm{GN}(\zeta(m)))$
- $\varphi(\mathrm{GN}(\psi))$

□

Twierdzenie 9. *(FM–wersja twierdzenia Tarskiego dla* $\mathrm{FM}((\omega,\perp))$*) Nie istnieje formuła* $\varphi(x)$ *z jedną zmienną wolną* x *w języku względnej pierwszości taki, że dla dowolnego zdania* ψ *w języku względnej pierwszości zachodzi:*

$$\mathrm{FM}((\omega,\perp)) \models_{sl} (\varphi(\mathrm{GN}(\psi)) \equiv \psi)$$

Dowód. Załóżmy nie wprost, że $\varphi(x)$ istnieje. Stosujemy FM–wersję lematu przekątniowego dla $\mathrm{FM}((\omega,\perp))$ i formuły $\neg\varphi(x)$ i otrzymujemy istnienie takiego zdania ψ, że:

$$\mathrm{FM}((\omega,\perp)) \models_{sl} \psi \equiv \neg\varphi(\mathrm{GN}(\psi))$$

Ponieważ φ jest FM–definicją prawdy, więc zachodzi:

$$\mathrm{FM}((\omega,\perp)) \models_{sl} \psi \equiv \varphi(\mathrm{GN}(\psi))$$

Zatem:

$$\mathrm{FM}((\omega,\perp)) \models_{sl} \psi \equiv \neg\psi$$

Sprzeczność — $\varphi(x)$, będące FM–definicją prawdy dla $\mathrm{FM}((\omega,\perp))$ nie może istnieć.

□