Studia Semiotyczne, t. XXX, nr 2 (2016), s. 185–206
ISSN 0137-6608
DOI: 10.26333/sts.xxx2.07

Rozprawa

Mateusz Łełyk, Bartosz Wcisło*

SILNE I SŁABE WŁASNOŚCI POJĘCIA PRAWDY**

Streszczenie: Niniejsza praca stanowi przegląd niedawnych wyników, zarówno opublikowanych, jak i jeszcze czekających na publikację, dotyczących różnych pojęć słabości i siły pojęcia prawdy, a także próbę ich systematyzacji i ukazania na tle szerszego nurtu badań. Omawiamy pojęcie granicy Tarskiego oddzielającej słabe i silne teorie prawdy. Omawiamy znane twierdzenia dotyczące niekonserwatywnych rozszerzeń podstawowej kompozycyjnej teorii prawdy oraz opisujemy pewną naturalną silną teorię prawdy, którą można scharakteryzować wieloma pozornie ze sobą niezwiązanymi układami aksjomatów. Na koniec przytaczamy inne możliwe eksplikacje pojęcia „siły" aksjomatycznych teorii prawdy.

Słowa kluczowe: aksjomatyczne teorie prawdy, arytmetyka Peana, konserwatywność, granica Tarskiego

1. WSTĘP

1.1. AKSJOMATYCZNE TEORIE PRAWDY

Formalne teorie prawdy stanowią część filozofii, na gruncie której pojęcie prawdy bada się przy zastosowaniu metod logiki matematycznej. Jedną z często przyjmowanych metod formalizacji jest badanie

---

* Uniwersytet Warszawski, Wydział Filozofii i Socjologii. E-mail: bar.wcislo@gmail.com, mlelyk@student.uw.edu.pl

** Badania, w ramach których powstał niniejszy artykuł, były prowadzone przy wsparciu Narodowego Centrum Nauki (NCN), nr grantu 2014/13/B/HS1/02892.

aksjomatycznych teorii prawdy skonstruowanych w następujący sposób:

– Ustalamy pewną teorię bazową B, która modeluje ogół naszej wiedzy niedotyczącej pojęć semantycznych, takich jak znaczenie lub prawda.

– Do języka tej teorii dodajemy nowy predykat jednoargumentowy $T(x)$, który interpretujemy jako „x jest zdaniem prawdziwym", natomiast do samej teorii aksjomaty rządzące tym nowym predykatem.

Następnie badamy, jak własności otrzymanej teorii zależą od wyboru aksjomatów rządzących predykatem prawdy.

W praktyce teorię bazową B często stanowi arytmetyka Peana PA. Za wyborem tym stoi przeświadczenie, że większość wyników dotyczących zależności między teorią prawdy a odpowiadającą jej teorią bazową nie zależy w istotny sposób od konkretnego wyboru tej drugiej. Jedyne, czego naprawdę od niej żądamy, to żeby była w stanie wyrazić i udowodnić podstawowe fakty dotyczące składni, jak na przykład: każde poprawnie zbudowane zdanie ma tyle samo prawych nawiasów, ile lewych. PA całkowicie wystarczy do tego celu.

Warto zaznaczyć, że wielu logików nie myśli o PA jako o teorii liczb, ale jako o teorii skończonych obiektów matematycznych, takich jak zbiory dziedzicznie skończone, grafy skończone, skończone ciągi znaków nad skończonym alfabetem. Teoria ta wystarcza do udowodnienia zaskakująco wielu faktów dotyczących tego rodzaju przedmiotów[1].

Ponieważ zdania, formuły lub dowody wyrażone w językach formalnych również można traktować jako skończone obiekty matematyczne, mianowicie ciągi znaków o łatwej do opisania budowie, arytmetyka Peana pozwala mówić o nich w zupełnie swobodny sposób. Przyznajemy jednak, że wybór PA jako teorii bazowej jest nieco arbitralny. Sformułowanie w abstrakcyjny sposób wymagań dotyczących teorii bazowej – gwarantujących, że teoria ta jest wystarczająco silna z punktu widzenia teorii prawdy, i wystarczających, aby zachodziły wszystkie wyniki, które jesteśmy w stanie otrzymać dla PA – wydaje się jednak zadaniem dość trudnym. Na początkowym etapie badań strategia „od szczegółu do ogółu" wydaje się lepszym pomysłem.

W niniejszym artykule skupiamy się na teoriach opisujących predykat prawdy dla języka teorii bazowej. Zaznaczmy jednak, że wła-

[1] Dostępnych jest wiele źródeł dotyczących formalizacji składni i uściślających nasze powyższe uwagi. Polecamy szczególnie (Franzen 2003) oraz (Kaye 1991).

sności samoodnośnych predykatów prawdy formalizujących pojęcie prawdziwości dla wszystkich zdań języka, do którego należą, stanowią przedmiot bardzo intensywnych badań (dobry opis wyników dotyczących takich teorii znajdzie Czytelnik w (Halbach 2011)).

Aksjomaty nakładane na predykat prawdy mają rzecz jasna uchwytywać różne intuicje związane z tym pojęciem. W ramach badań nad formalnymi teoriami prawdy staramy się więc wytłumaczyć, jakie są związki między tymi intuicyjnymi własnościami i jakie są konsekwencje faktu, że predykat prawdy posiada te własności. Jeden z najprostszych warunków, które możemy narzucić na predykat prawdy, to układ aksjomatów następującej postaci:

$$T(\ulcorner \varphi \urcorner) \equiv \varphi$$

dla wszystkich zdań φ z języka teorii bazowej. Aksjomaty te wyrażają, że predykat prawdy spełnia równoważności Tarskiego dla języka teorii bazowej. Teorię rozszerzającą PA, w której jedynymi aksjomatami rządzącymi predykatem prawdy są wyżej opisane równoważności, nazywamy TB^{-}[2].

Inną własnością, której możemy oczekiwać od predykatu prawdy, jest kompozycyjność. Wyrażamy ją aksjomatami formalizującymi zasady takie, jak:

(Dla dowolnych zdań φ i ϑ) Koniunkcja zdań φ oraz ϑ z języka teorii bazowej jest prawdziwa dokładnie wtedy, gdy prawdziwe są oba jej człony.

Lub:

(Dla dowolnej zmiennej v i dowolnej formuły $\varphi(v)$ z co najwyżej jedną zmienną wolną v) Zdanie uniwersalne $\ulcorner \forall v\varphi(v) \urcorner$ z języka teorii bazowej jest prawdziwe dokładnie wtedy, gdy dla dowolnego liczebnika $\underline{x}$ prawdziwe jest zdanie $\ulcorner \varphi(\underline{x}) \urcorner$.

Dodajmy, że liczebnikiem $\underline{x}$ nazywamy kanoniczny term oznaczający liczbę x np. $(\ldots((\underline{0} + \underline{1}) + \underline{1}) \ldots + \underline{1})$, gdzie symbol dodawania pojawia się x-krotnie (natomiast $\underline{0}$ i $\underline{1}$ to pewne ustalone symbole reprezentujące odpowiednio 0 i 1).

Zauważmy, że już teoria TB^{-} dla każdych dwóch konkretnych zdań φ, ϑ języka arytmetyki udowodni, że ich koniunkcja[3] jest prawdziwa

[2] W literaturze częściej przyjmuje się oznaczenie $\mathrm{TB}\restriction$.

[3] Oczywiście mamy tu na myśli kod Gödla reprezentujący koniunkcję tych formuł. Dla zwiększenia czytelności pozwolimy sobie jednak na mniej precyzyjne sformułowania.

dokładnie wtedy gdy φ i ϑ są prawdziwe. Nie udowodni ona jednak ogólnego faktu o wszystkich zdaniach arytmetyki, który wyraża właśnie pierwszy z cytowanych aksjomatów. Teorię, której aksjomaty głoszą, że predykat prawdy jest kompozycyjny nazywamy CT^-[4]. Precyzyjną definicję tej teorii znajdzie Czytelnik w (Kaye 1991). Inną zasadą, którą można postulować, jest zasada ekstensjonalności predykatu prawdy:

Dla dowolnych zdań $\varphi(t)$, $\varphi(s)$ z języka teorii bazowej, jeśli wartości termów t, s są równe, to zdanie $\varphi(t)$ jest prawdziwe wtedy i tylko wtedy, gdy zdanie $\varphi(s)$ jest prawdziwe.

Kolejny często wyodrębniany postulat głosi, że zdania zawierające predykat prawdy spełniają własność indukcji lub – równoważnie – zasadę minimum:

Każdy niepusty zbiór liczb naturalnych zdefiniowany formułą zawierającą predykat prawdy ma element najmniejszy.

W języku pierwszego rzędu, a w takim sformułowane są rozważane przez nas teorie, powyższą zasadę wyraża układ nieskończenie wielu aksjomatów, tzw. schemat indukcji dla formuł języka rozszerzonego o predykat prawdy. Powyższa zasada ma bardziej techniczny charakter niż te opisano poprzednio. Można ją jednak odczytywać w następujący sposób: własności zdefiniowane przy użyciu predykatu prawdy są „dobrze określone". Nie zachodzi dla nich zjawisko nieostrości.

Dodajmy, że teorie CT^- oraz TB^- z dodanym schematem indukcji dla zdań zawierających predykat prawdy nazywamy odpowiednio CT i TB.

Mamy nadzieję, że Czytelnik dostrzega szeroki wachlarz naturalnych własności, których oczekujemy od predykatu prawdy. Możliwości przybywa, gdy zaczynamy rozważać samoodnośne predykaty prawdy, to znaczy, jeśli zaczynamy uwzględniać w aksjomatach zachowanie predykatu prawdy, gdy odnosi się do zdań, w których występuje tenże właśnie predykat.

1.2. SŁABE I SILNE TEORIE PRAWDY

Jeden z podstawowych wyników teorii prawdy głosi, że teoria CT dowodzi pewnych zdań arytmetycznych, których nie dowodzi sama

[4] Ponownie, częściej przyjmowanym oznaczeniem jest $\mathrm{CT}\restriction$. Ogólnie, teorie oznaczane przez nas Th^-, często oznacza się $\mathrm{Th}\restriction$.

PA. Mianowicie na mocy drugiego twierdzenia Gödla wiemy, że jeśli PA jest niesprzeczna, to nie dowodzi zdania Con_{PA}, które formalizuje tezę o niesprzeczności PA. Tymczasem zachodzi następujący fakt:

Twierdzenie 1 (Tarski). CT dowodzi zdania Con_{PA}.

Przedstawimy nieformalny szkic dowodu tego twierdzenia (pełny dowód można znaleźć m.in. w (Halbach 2011): Najpierw pokazujemy, że CT dowodzi zdania „Wszystkie aksjomaty PA są prawdziwe". Ponieważ PA ma nieskończenie wiele aksjomatów, nie jest to zupełnie banalne. Nie wystarczy udowodnić, że każdy aksjomat z osobna jest prawdziwy (co potrafi zrobić już TB^{-}), lecz potrzeba dowodu zdania ogólnego. Intuicja stojąca za dowodem nie jest bardzo skomplikowana, jednak zawiera dość techniczne szczegóły. Poprzestaniemy na jego naszkicowaniu.

Pracując w CT^{-}, ustalmy dowolną formułę $\varphi(x)$. Zdanie

$$(T(\varphi(\underline{0})) \wedge \forall x(T(\varphi(\underline{x})) \rightarrow T(\varphi(\underline{x+1})))) \longrightarrow \forall x T(\varphi(\underline{x}))$$

jest (prawdziwym) podstawieniem schematu indukcji (z parametrem φ) dla pewnej formuły z predykatem T, jest zatem dostępne w CT jako aksjomat. Korzystając z aksjomatów kompozycyjnych w CT^{-}, dostajemy

$$T((\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1))) \longrightarrow \forall x \varphi(x))$$

Zdanie powyższe stwierdza zaś, że podstawienie schematu indukcji dla formuły φ jest prawdziwe. Z dowolności φ dostajemy zdanie ogólne[5]. Poza schematem indukcji PA ma jedynie skończenie wiele aksjomatów, więc korzystając skończenie wiele razy z kompozycyjności predykatu prawdy, jesteśmy w stanie pokazać, że one wszystkie są prawdziwe.

Udowodniwszy, że aksjomaty PA są prawdziwe, dowodzimy następnie przez indukcję po liczbie kroków w dowodzie, że każde zdanie, które można wywieść z aksjomatów PA, jest prawdziwe. Jesteśmy tymczasem w stanie pokazać, korzystając ponownie z aksjomatów kompozycyjnych, że żadne zdanie postaci $\varphi \wedge \neg\varphi$ nie jest prawdziwe. Zatem

---

[5] Celowo prześlizgujemy się tutaj nad pewnymi szczegółami: między innymi nad tym, że w ten sposób pokazujemy prawdziwość jedynie tzw. schematu indukcji bez parametrów. Jak już jednak wspomnieliśmy kompletny dowód wymaga przezwyciężenia koncepcyjnie nietrudnych, ale dość technicznych trudności.

żadne zdanie tej postaci nie jest dowodliwe w PA, co kończy szkic dowodu Twierdzenia 1.

Powyższe twierdzenie można uznać za filozoficznie istotne: okazuje się, że dołączenie do PA predykatu prawdy spełniającego bardzo naturalne warunki daje w wyniku teorię silniejszą niż PA. Fakt ten stał się podstawą argumentu wymierzonego przeciwko deflacyjnej teorii prawdy[6].

W sytuacji, gdy dane są dwie teorie Th_1, Th_2 takie, że $Th_1 \subseteq Th_2$ oraz istnieje takie zdanie w języku teorii Th_1, którego dowodzi Th_2, ale nie Th_1, mówimy, że Th_2 jest niekonserwatywna nad Th_1. W przeciwnym przypadku mówimy, że Th_2 jest konserwatywna nad Th_1. Z Twierdzenia 1 (i Twierdzenia Gödla) wynika zatem, że CT jest niekonserwatywna nad PA. Niezależnie od filozoficznej doniosłości tego konkretnego faktu ciekawe wydaje się ogólne pytanie: jakie własności predykatu prawdy sprawiają, że teoria prawdy Th staje się niekonserwatywna nad swoją teorią bazową *B*? W niniejszej pracy opisujemy pewne wyniki związane z tym pytaniem. Innymi słowy, staramy się zrozumieć, jakie własności pojęcia prawdy sprawiają, że teoria prawdy jest „silniejsza" od teorii bazowej.

2. ZNANE WYNIKI DOTYCZĄCE KONSERWATYWNOŚCI

W świetle wyniku Tarskiego, przedstawionego w poprzedniej sekcji, głoszącego, że kompozycyjna teoria prawdy spełniająca wszystkie aksjomaty indukcji nie jest konserwatywna nad PA, naturalne jest pytanie, czy teoria prawdy CT^-, w której o predykacie prawdy zakłada się jedynie tyle, że jest kompozycyjny, jest konserwatywna nad arytmetyką.

Na pytanie to odpowiada następujące twierdzenie:

Twierdzenie 2 (Kotlarski–Krajewski–Lachlan, Enayat–Visser, Leigh). CT^- jest konserwatywna nad PA.

Zanim przejdziemy do omówienia tego twierdzenia, skomentujmy kwestię autorstwa. Kotlarski, Krajewski i Lachlan (1981) udowodnili twierdzenie teoriomodelowe, które pociągało za sobą konserwatyw-

[6] Por. (Ketland 1999), (Shapiro 1998).

ność pewnej teorii bardzo zbliżonej do CT^-. W czasie, gdy powstawał ich artykuł, aksjomatyczne teorie prawdy nie wyodrębniły się jeszcze jako osobny przedmiot badań i nie sformułowano ich standardowych definicji. W związku z tym teoria, której konserwatywność można wyprowadzić jako wniosek z twierdzenia Kotlarskiego–Krajewskiego–Lachlana, różni się od CT^- (aksjomatyzuje predykat spełniania, nie predykat prawdy) i nie jest jasne, jak należałoby zmodyfikować dowód tego twierdzenia, tak aby pokazywał konserwatywność CT^-. Dowód konserwatywności kompozycyjnej teorii prawdy, znacznie prostszy od argumentu z (Kotlarski, Krajewski, Lachlan 1981), sformułowali dopiero Enayat i Visser (2015). Teoria, którą badali, różniła się jednak wciąż od CT^-, choć była to już nie tak znaczna różnica. Dowód konserwatywności CT^- podał dopiero Leigh (2015), korzystając z jeszcze innych technik niż wyżej wymienieni autorzy.

2.1. ZASADY DOMKNIĘCIA I POPRAWNOŚCI

Twierdzenie 2 mówi, że czysto kompozycyjna teoria prawdy nie dowodzi więcej faktów arytmetycznych niż sama PA. Dopiero schemat indukcji dla zdań zawierających predykat prawdy pozwala udowodnić np., że PA jest niesprzeczna.

Widzimy więc dwie bardzo naturalne teorie, CT^- i CT, z których jedna jest konserwatywna nad PA, a druga nie. Indukcja dla predykatu prawdy dostępna w drugiej teorii pozwala udowodnić wiele faktów o jego strukturze. Zasadnicza własność tego predykatu, kompozycyjność, nie wystarczy, żeby udowodnić nowe twierdzenia arytmetyczne. Nasze pytanie o naturalną granicę między teoriami konserwatywnymi a niekonserwatywnymi nad PA możemy zawęzić więc do następującego problemu: jakie naturalne aksjomaty charakteryzujące predykat prawdy dodane do CT^- sprawiają, że teoria ta przestaje być konserwatywna nad PA?

Ali Enayat zaproponował określenie granicy oddzielającej konserwatywne i niekonserwatywne teorie prawdy zawarte pomiędzy CT^- a CT *granicą Tarskiego*[7]. Zatem nasze pytanie można wyrazić

[7] Uzasadnijmy pokrótce wybór tej nazwy: Tarski, zdaje się jako pierwszy, zwrócił uwagę na „słabość" niektórych arytmetycznych teorii prawdy (m.in. TB^-), a poza tym CT^- jest modelowana na wzór jego indukcyjnych warunków definiujących relację spełniania. Zarówno Ali Enayat, jak i autorzy używali tego sformu-

w następujący zwarty sposób: jak przebiega granica Tarskiego? Spróbujmy teraz wymienić kilka naturalnych aksjomatów, które wykraczają poza CT^-, ale są dowodliwe w CT.

Bardzo naturalną grupą zasad rozszerzających CT^-, a dowodliwych w CT są zasady domknięcia i zasady poprawności. Zasady domknięcia głoszą, że zdania prawdziwe są domknięte na rozumowania w danym systemie wnioskowania. Zasady poprawności głoszą, że pewien zbiór zdań zawiera wyłącznie zdania prawdziwe. Za chwilę podamy typowe przykłady takich zasad.

Z dowodu twierdzenia o niekonserwatywności CT możemy wyodrębnić jedną prostą zasadę poprawności, która z pewnością nie jest konserwatywna, zasadę poprawności arytmetyki Peana:

Każde twierdzenie arytmetyki Peana jest prawdziwe.

Powyższą zasadę nazywa się też zasadą globalnej refleksji nad PA. Zauważmy, że w dowodzie niekonserwatywności CT skorzystaliśmy dokładnie z faktu, że CT dowodzi poprawności PA. Możemy wyodrębnić dwie kolejne naturalne zasady dowodliwe w CT, które łącznie implikują zasadę poprawności PA. Pierwsza z nich to zasada domknięcia na logikę pierwszego rzędu:

Każde zdanie dowodliwe w logice pierwszego rzędu z prawdziwych przesłanek jest prawdziwe.

Można powiedzieć, że zasada ta ma bardziej fundamentalny charakter niż zasada poprawności PA. Mówi bowiem jedynie o związku między prawdą a logiką pierwszego rzędu i nie zależy w jawny sposób od naszego zaufania do prawdziwości aksjomatów arytmetyki. O statusie aksjomatów arytmetyki Peana mówi wprost zasada poprawności aksjomatycznej PA:

Każdy aksjomat PA jest prawdziwy.

Standardowymi technikami teoriodowodowymi można pokazać, że konsekwencje arytmetyczne CT ściśle zawierają konsekwencje arytmetyczne CT^- z dodanymi zasadami poprawności aksjomatycznej PA i do-

łowania kilkakrotnie podczas wystąpień konferencyjnych, jednak, zgodnie z naszą najlepszą wiedzą, niniejszy artykuł jest pierwszym dłuższym studium, w którym to pojęcie zostało wprowadzone w druku.

mknięcia na logikę pierwszego rzędu. Wyodrębniliśmy więc naturalną teorię prawdy, która jest ściśle słabsza niż CT, ale wciąż niekonserwatywna nad PA. Mogłoby się przy tym wydawać, że pozostawienie któregokolwiek z dwóch aksjomatów tej teorii dodanych do CT^- również daje teorię niekonserwatywną. Okazuje się jednak, że sama w sobie zasada poprawności aksjomatycznej PA zalicza się do zasad słabych, co zostało ogłoszone już w (Kotlarski, Krajewski, Lachlan 1981). Wynik ten został ogłoszony również w (Enayat, Visser 2015) i (wraz z dowodem) w (Leigh 2015). Każde z powyższych źródeł dostarcza innych technik udowodnienia tego twierdzenia.

Twierdzenie 3 (Kotlarski–Krajewski–Lachlan, Enayat–Visser, Leigh). CT^- z dodaną zasadą poprawności aksjomatycznej PA jest konserwatywna nad PA.

Okazuje się więc, że jesteśmy w stanie dość precyzyjnie zawęzić pole poszukiwań granicy między słabymi i silnymi kompozycyjnymi teoriami prawdy. Dodanie do CT^- zasady głoszącej, że aksjomaty PA są prawdziwe, nie pozwala uzyskać nowych konsekwencji arytmetycznych. Natomiast dalsze wzbogacenie teorii prawdy o zasadę głoszącą, że prawdziwe są zdania dowodliwe w PA, daje już teorię niekonserwatywną.

2.2. OGRANICZONY SCHEMAT INDUKCJI DLA PREDYKATU PRAWDY

Inną perspektywą, która pozwala szukać naturalnych teorii niekonserwatywnych nad arytmetyką Peana, ale zarazem istotnie słabszych od CT jest ograniczenie schematu indukcji do pewnych wybranych klas formuł.

Powiemy, że formuła jest klasy Δ_0, jeśli wszystkie występujące w niej kwantyfikatory $\forall$, $\exists$ są kwantyfikatorami ograniczonymi, czyli można je przedstawić w postaci $\forall x < t$, $\exists y < s$ dla pewnych termów t, s. Prawdziwość zdań klasy Δ_0 zależy więc jedynie od własności obiektów pewnej z góry ustalonej wielkości. Można więc o nich myśleć jako o pewnym szczególnym przypadku zdań, które można efektywnie rozstrzygać. Ta klasa formuł odgrywa znaczną rolę w badaniach dotyczących metamatematycznych własności arytmetyki.

Inną ważną klasą formuł są formuły klasy Π_1. Są to formuły postaci

$$\forall x_1,\ldots,\forall x_n \varphi,$$

gdzie φ jest formułą klasy Δ_0. Można o nich myśleć jako o formułach czysto uniwersalnych. Takich, które stwierdzają, że pewne proste, efektywnie rozstrzygalne fakty zachodzą dla wszystkich obiektów.

Istotną klasą podteorii arytmetyki Peana są jej fragmenty powstałe przez ograniczenie formuł występujących w schemacie indukcji do formuł klasy $\prod_1$ lub Δ_0. Podążymy tą drogą również w przypadku teorii prawdy. Przez CT_1 będziemy rozumieli teorię CT^-, do której dodajemy te podstawienia schematu indukcji:

$$(\varphi(0) \wedge \forall x\ (\varphi(x) \rightarrow \varphi(x+1))) \longrightarrow \forall x \varphi(x),$$

w których formuła φ jest dowolną formułą klasy $\prod_1$.

Zauważmy, że ograniczenie do formuł klasy $\prod_1$ dotyczy tylko formuł zawierających predykat prawdy, gdyż już CT^-, jako rozszerzenie PA, zawiera wszystkie aksjomaty indukcji dla formuł arytmetycznych.

Teoria CT_1 jest dość naturalna w kontekście naszych rozważań, gdyż analiza dowodu niekonserwatywności CT pozwala stwierdzić, że tak naprawdę korzystaliśmy tylko z aksjomatów dostępnych już w CT_1. Otrzymujemy zatem następujący wniosek:

Twierdzenie 4. CT_1 jest niekonserwatywna nad PA.

Istotnie, łatwo można pokazać, że zasada poprawności PA, a co za tym idzie również zasada poprawności aksjomatycznej PA, są dowodliwe w CT_1, podobnie jak zasada domknięcia na logikę pierwszego rzędu. Zatem uzyskaliśmy jeszcze inną perspektywę pozwalającą zawęzić nasze poszukiwania naturalnych zasad, które sprawiają, że teoria prawdy staje się istotnie silniejsza od swojej teorii bazowej. W szczególności naturalne staje się pytanie o konserwatywność teorii CT_0, która powstaje przez ograniczenie schematu indukcji dla formuł zawierających predykat prawdy do formuł klasy Δ_0.

Dodajmy jeszcze, że TB (czyli TB^- z pełną indukcją) jest konserwatywna nad arytmetyką Peana, co stanowi zresztą znacznie prostszy wynik niż Twierdzenie 3.Wynika stąd, że teoria prawdy z naturalną aksjomatyką obejmującą pełen schemat indukcji nie musi automatycznie dowodzić nowych faktów arytmetycznych. Co więcej, można pokazać przykłady teorii istotnie rozszerzających TB, opartych na pewnych wariantach schematu równoważności Tarskiego i zawierających pełną indukcję, które wciąż są konserwatywne. Zatem fakt, że rozważamy teorie kompozycyjne, jak najbardziej ma znaczenie dla wyników o konserwatywności.

3. WYZNACZANIE GRANICY TARSKIEGO

W niniejszym rozdziale przedstawiamy główne znane fakty dotyczące przebiegu granicy Tarskiego – większość twierdzeń, które przytaczamy jest jeszcze nieopublikowana, dlatego treść tego rozdziału należy traktować jako komunikat o stanie badań.

Zacznijmy od drobnego uporządkowania naszej wiedzy: powyżej widzieliśmy, że (nad CT^-) zasada domknięcia na logikę pierwszego rzędu w połączeniu z zasadą poprawności aksjomatycznej PA dowodzi zasady poprawności PA. Okazuje się, że już ta pierwsza zasada wystarczy do tego celu: pracując w CT^- z dodaną zasadą domknięcia na logikę pierwszego rzędu, udowodnimy, że wszystkie aksjomaty PA są prawdziwe. Dowód tego faktu jest raczej standardowy i intuicyjny: skończenie wiele aksjomatów określających własności dodawania, mnożenia i porządku jest prawdziwych już w CT^-. Pozostają aksjomaty indukcji: PA (nawet zinterpretowana w modelu niestandardowym) „myśli, że" obiekty, o których mówi, są uporządkowane tak jak liczby naturalne, tzn. od 0 dzieli każdą liczbę skończenie wiele kroków (oczywiście w modelu niestandardowym to „skończenie" może oznaczać liczbę niestandardową). Pracując w CT^- z zasadą domknięcia na logikę pierwszego rzędu i zakładając, że $\varphi(\underline{0})$ i $\forall x(\varphi(x) \rightarrow \varphi(x + 1))$ są prawdziwe dla ustalonej formuły arytmetycznej $\varphi(x)$, dla dowolnego a zbudujemy w czystej logice pierwszego rzędu dowód zdania $\varphi(\underline{a})$ (a razy stosujemy reguły *dictum de omni* i *modus ponens*). Dowód ten wygląda tak samo jak klasyczne uzasadnienie, że w modelu standardowym arytmetyki prawdziwa jest zasada indukcji, z tą tylko różnicą, że w modelu niestandardowym wykorzystujemy aksjomat indukcji dla formuły:

$$\theta(x) := \mathrm{Pr}_{\mathrm{PA}}(\phi(\underline{x})).$$

Skoro zbiór zdań prawdziwych jest domknięty na rozumowania w logice pierwszego rzędu, możemy wnioskować, że $\varphi(\underline{a})$ jest prawdziwe. Z dowolności a wnioskujemy, że dla dowolnego a $\varphi(\underline{a})$ jest prawdziwe, a zatem (na mocy aksjomatów kompozycyjnych), że prawdziwe jest zdanie $\forall x\varphi(x)$. Nasz dowód jest zakończony.

Rozumując analogicznie, można pokazać (po drodze omijając jedną dodatkową komplikację – zainteresowanych odsyłamy do artykułu (Cieśliński 2010), w którym zostało to po raz pierwszy udowodnione), że zasada poprawności PA równoważna jest następującej, znacznie ograniczonej, zasadzie poprawności:

Wszystkie twierdzenia logiki pierwszego rzędu są prawdziwe.

Powyższe zasady dadzą się dość naturalnie podzielić na te stwierdzające, że zbiór zdań prawdziwych jest domknięty na pewne reguły wnioskowania (np. logikę pierwszego rzędu), i te stwierdzające, że wszystkie zdania z pewnego zbioru są prawdziwe (np. twierdzenia PA, twierdzenia logiki pierwszego rzędu). Intuicyjnie zasady pierwszego typu mówią coś więcej niż ich odpowiedniki drugiego typu. W zeszłym roku jednak Cezary Cieśliński przedstawił pomysłowy dowód, że nad CT^- te zasady są równoważne. Wyodrębnijmy to jako osobne

Twierdzenie 5 (Cieśliński). CT^- z dodanym aksjomatem „Wszystkie twierdzenia logiki są prawdziwe" dowodzi z a s a d y d o m k n i ę c i a n a l o g i k ę p i e r w s z e g o r z ę d u.

W poszukiwaniu słabszych zasad dowodliwych w CT, a właściwie rozszerzających CT, wyodrębnijmy z zasady domknięcia na logikę pierwszego rzędu następującą z a s a d ę d o m k n i ę c i a n a k l a s y c z n y r a c h u n e k z d a ń:

Każde zdanie dowodliwe w klasycznym rachunku zdań z prawdziwych przesłanek jest prawdziwe.

Oczywiście nad CT^- powyższe zdanie jest dowodliwe z zasady domknięcia na logikę pierwszego rzędu. Jak pokazał Cieśliński (2010), rozszerzenie CT^- o zasadę domknięcia na klasyczny rachunek zdań jest równoważne kompozycyjnej teorii prawdy opartej na indukcji ograniczonej, czyli CT_0.

Twierdzenie 6 (Cieśliński). Zasada domknięcia na klasyczny rachunek zdań jest dowodliwa w CT_0. Na odwrót: każdy aksjomat CT_0 jest dowodliwy w CT^- z dodanym aksjomatem domknięcia na klasyczny rachunek zdań.

Na długo przed artykułem Cieślińskiego, Henryk Kotlarski podał dowód, że CT_0 dowodzi zasady poprawności PA (Kotlarski 1986). Argument, choć skrótowy, wydawał się poprawny i był na tyle przekonujący, że cytowany był jeszcze w (Cieśliński 2010) i (Halbach 2011). Jednakże w 2008 roku Albert Visser i Richard Heck dostrzegli lukę w dowodzie Kotlarskiego: problem dotyczył wykazania, że CT_0 dowodzi zdania:

Każdy aksjomat logiki pierwszego rzędu jest prawdziwy.

Jeśli wzbogacić CT_0 o powyższy aksjomat, dowód Kotlarskiego przechodzi bez dalszych problemów. Jest dość oczywiste, że powyższy aksjomat można udowodnić, wykorzystując indukcję dla formuł klasy $\prod_1$. Problem odtworzenia tego dowodu przy użyciu indukcji tylko dla formuł ograniczonych wydawał się na tyle nie do pokonania, że wielu logików (w tym i autorzy tego artykułu) zaczęło szukać dowodu, że CT_0 leży po konserwatywnej stronie Granicy Tarskiego.

W końcu pokazano (dowód ukaże się w (Łełyk, Wcisło 2017a, w druku)), że CT_0 dowodzi tych samych zdań arytmetycznych co CT_0 z dodaną zasadą poprawności PA. Tak naprawdę w tym dowodzie wykorzystano jedynie rozszerzenie CT^- o dwie naturalne zasady dla predykatu prawdy (dowodliwe w CT_0): jedną z nich była już wprowadzona zasada poprawności aksjomatycznej PA, a drugą następująca uogólniona zasada przemienności z alternatywą, zwana także zasadą poprawności dysjunktywnej[8]:

Dla dowolnego x i dowolnych x zdań $\varphi_0, \ldots, \varphi_x$, ich alternatywa jest prawdziwa wtedy i tylko wtedy, gdy któreś z $\varphi_0, \ldots, \varphi_x$ jest prawdziwe.

Doprecyzujmy: dla dowolnej liczby naturalnej n, CT^-, dzięki aksjomatom kompozycyjnym, będzie w stanie udowodnić, że alternatywa n zdań jest prawdziwa dokładnie wtedy, gdy któreś z tych zdań jest prawdziwe. Nie będzie jednak w stanie udowodnić powyższego zdania ogólnego[9]. Zaskakujące jest to, że już tak proste uogólnienia podstawowych aksjomatów kompozycyjnych, w połączeniu z (samodzielnie konserwatywną) zasadą poprawności aksjomatycznej PA, dają teorię,

[8] Używamy tego nieco sztucznego terminu, gdyż „alternatywna poprawność" brzmi trochę jak „druga świeżość" albo „uczciwy inaczej".

[9] Ten wynik po raz pierwszy uzyskali Kotlarski, Krajewski i Lachlan (1981). Inny dowód można podać, opierając się na metodach wykorzystanych przez Enayata-Vissera.

która dowodzi tych samych twierdzeń arytmetycznych co CT^- z zasadą poprawności PA. Podsumujmy to w następującym twierdzeniu:

Twierdzenie 7 (W.). CT^- rozszerzona o zasady poprawności dysjunktywnej i poprawności aksjomatycznej PA dowodzi tych samych twierdzeń arytmetycznych co CT^- rozszerzona o zasadę poprawności PA.

Warto już teraz podkreślić nieoczywistość powyższych odkryć: jak na razie okazuje się, że wszystkie zasady, o których wiemy, że leżą po niekonserwatywnej stronie granicy Tarskiego dowodzą (przynajmniej) konsekwencji arytmetycznych zasady poprawności PA (dla uproszczenia napisów oznaczmy tę zasadę przez TPA). Zauważmy teraz, że ten zbiór zawiera dużo więcej niedowodliwych (w PA) zdań niż po prostu zdanie stwierdzające niesprzeczność PA (ozn. Con_{PA}). Widzieliśmy już, że to zdanie jest dowodliwe w CT^- + TPA. Jest to ponadto zdanie arytmetyczne, zatem możemy pokazać, że jest prawdziwe, wykorzystując skończenie wiele razy aksjomaty kompozycyjne.

Ponieważ ta teoria dowodzi zasady domknięcia na logikę pierwszego rzędu i wiemy, że wszystkie aksjomaty teorii PA + Con_{PA} sa prawdziwe, to do konsekwencji tej teorii nie należy żadne zdanie fałszywe (w szczególności 0 = 1). Udowodniliśmy zatem niesprzeczność teorii PA + Con_{PA}, czyli zdanie

$$\mathrm{Con}_{PA+\mathrm{Con}_{PA}}$$

Nic nam nie przeszkadza w iterowaniu tego procesu. Możemy w ten sposób udowodnić coraz silniejsze stwierdzenia

$$\mathrm{Con}_{PA+\mathrm{Con}_{PA+\mathrm{Con}_{PA}}}$$
$$\mathrm{Con}_{PA+\mathrm{Con}_{PA+\mathrm{Con}_{PA+\mathrm{Con}_{PA}}}}$$
$$\mathrm{Con}_{PA+\mathrm{Con}_{PA+\mathrm{Con}_{PA+\mathrm{Con}_{PA+\mathrm{Con}_{PA}}}}}$$

i tak w nieskończoność.

Tak naprawdę CT^- + TPA jest dużo silniejsza arytmetycznie: nietrudno jest się przekonać, że dowodzi ona wszystkich zdań postaci:

$$\forall x \ (\mathrm{Pr}_{PA}(\varphi(\underline{x})) \rightarrow \varphi(x)) \qquad (*)$$

dla dowolnej formuły $\varphi(x)$ języka arytmetyki. Zbiór wszystkich zdań tej postaci nazywa się zasadą jednorodnej refleksji nad PA[10]. Mały frag-

[10] Widać teraz dlaczego zasadę prawdziwości twierdzeń PA nazwaliśmy „globalną" refleksją – w obecności predykatu prawdy możemy „wyrazić" powyższą zasadę jako jedno zdanie.

ment tego zbioru zdań (dla formuł klasy Π_1) wystarczy do udowodnienia wszystkich powyższych iteracji zdań stwierdzających niesprzeczność coraz silniejszych teorii. I to nie koniec: zbiór (kodów Gödla) wszystkich zdań powyższej postaci jest silnie reprezentowany w PA (jest pierwotnie rekurencyjny), zatem w arytmetyce można zdefiniować teorię

$$\mathrm{PA}^1 := \mathrm{PA} \cup \{\forall x \ (\mathrm{Pr}_{\mathrm{PA}}(\varphi(\underline{x})) \rightarrow \varphi(x)) \mid \varphi(x) \in \mathsf{L}_{\mathrm{PA}}\} \qquad (**)$$

dla której standardowy predykat dowodliwości będzie spełniał warunki Gödla–Loba. W CT^- + TPA udowodnimy wszystkie zdania postaci (*) dla PA^1, tj. wszystkie zdania

$$\forall x(\mathrm{Pr}_{\mathrm{PA}^1}(\varphi(\underline{x})) \rightarrow \varphi(x))$$

gdzie $\varphi(x)$ jest dowolną formułą arytmetyczną z co najwyżej jedną zmienną wolną. Następnie możemy zdefiniować teorię PA^2, w (**) zastępując PA przez PA^1. Iterując ten proces w nieskończoność, w kroku granicznym biorąc teorię

$$\mathrm{PA}^{\omega} := \bigcup_{n \in \omega} \mathrm{PA}^n$$

dopiero otrzymamy arytmetyczną aksjomatyzację konsekwencji arytmetycznych CT^- + TPA. Bardzo elegancki dowód, że PA^{ω} to faktycznie aksjomatyzacja wszystkich konsekwencji arytmetycznych tej teorii, przedstawił Kotlarski (1986).

Sytuacja zaczyna wyglądać zatem tak, jak gdyby każda „naturalna" teoria prawdy dowodząca niesprzeczności arytmetyki dowodziła od razu wszystkich zdań powyższej postaci. Oczywiście można podać sztuczne kontrprzykłady dla tego „twierdzenia": na przykład teoria CT^- z dołączonym aksjomatem „$\mathrm{Con}_{\mathrm{PA}}$ jest prawdziwe" jest niekonserwatywna nad PA i słabsza od rozważanych teorii (nie dowodzi np. zdania $\mathrm{Con}_{\mathrm{PA}+\mathrm{Con}_{\mathrm{PA}}}$.) Oczywiście „naturalna" nie jest terminem formalnym, lecz wyraża tylko pewną heurystykę: pozwala tymczasowo zablokować kontrprzykłady wymyślane *ad hoc*. W tej chwili staramy się znaleźć „naturalny" kontrprzykład, być może przekonując się po drodze, że taki kontrprzykład nie może istnieć. Wtedy pewnie zrozumiemy też, co znaczy „naturalna".

Podkreślmy, że powyżej nie powiedzieliśmy, że CT_0 dowodzi zasady poprawności PA. Dowód pokazujący niekonserwatywność tej teorii polega na konstrukcji takiej formuły $T'(x)$, która dowodliwie w CT_0 zachowuje się tak jak predykat prawdy CT_0 spełniający dodatkowo zasadę poprawności PA. Używając definicji, której wprowadzenie odkładamy na później (Definicja 13), pokazano zatem, że CT_0 rozszerzona o zasadę poprawności PA jest interpretowalna w sensie Fujimoto w CT_0. Nie udało się jednak pokazać, że skonstruowana formuła $T'(x)$ dowodliwie w CT_0 będzie miała tę samą ekstensję co wyjściowy predykat prawdy. Mówiąc niezbyt precyzyjnie, mogłoby być tak, że CT_0 jest w stanie „poprawić" swój własny predykat prawdy, ale nie może udowodnić że jej własny predykat prawdy jest tak „dobry" (tzn. spełnia zasadę poprawności PA). W międzyczasie Ali Enayat[11] pokazał, że koncentrując się na rozszerzeniu CT^- o zasady poprawności aksjomatycznej PA i poprawności dysjunktywnej, nie osłabiliśmy naszych założeń w żaden sposób. Zachodzi bowiem następujące

Twierdzenie 8 (Enayat). CT^- rozszerzona o zasadę poprawności aksjomatycznej PA i aksjomat poprawności dysjunktywnej dowodzi CT_0.

Okazało się zatem, że z dokładnością do domknięcia na konsekwencje logiczne i ograniczając się do teorii, których niekonserwatywność jesteśmy w stanie dowieść, istnieją dwie minimalne niekonserwatywne teorie: CT_0 i $CT^- + TPA$, które dodatkowo mają te same konsekwencje arytmetyczne. Następnie jednak udało się pokazać, że obrazek ten jest jeszcze prostszy: mianowicie można bezpośrednio załatać lukę w dowodzie Kotlarskiego. Podsumujmy nasze dotychczasowe rozważania w następującym twierdzeniu:

Twierdzenie 9 (Cieśliński, Enayat, Kotlarski, Ł.). Następujące teorie są równoważne:
1. CT_0
2. CT^- rozszerzone o zasadę poprawności PA.
3. CT^- rozszerzone o zasadę domknięcia na logikę pierwszego rzędu.
4. CT^- rozszerzone o zasadę domknięcia na klasyczny rachunek zdań.
5. CT^- rozszerzone o zasadę prawdziwości twierdzeń logiki pierwszego rzędu.
6. CT^- rozszerzone o zasadę poprawności dysjunktywnej i zasadę poprawności aksjomatycznej PA.

Jak na razie sytuacja na granicy Tarskiego wygląda zatem tak, jakby po niekonserwatywnej stronie granicy istniała najmniejsza (czyt.: naj-

[11] Wynik ten nie jest jeszcze nigdzie opublikowany. O jego uzyskaniu poinformował nas bezpośrednio prof. Enayat.

mniejsza „naturalna”) teoria, którą można zaksjomatyzować na wiele różnych sposobów. Oczywiście, jak uważny Czytelnik zapewne zauważył, kilka pytań w dalszym ciągu pozostawiliśmy bez odpowiedzi. Nie był to przypadek: w tej chwili nie wiemy np., czy rozszerzenie CT^- o samą tylko zasadę poprawności dysjunktywnej jest konserwatywne. Intuicje osób pracujących nad tym problemem są takie, że znajduje się ona po „słabej” stronie granicy Tarskiego. Jednakże te intuicje już nie raz nas zawiodły…

Warto, przynajmniej pokrótce, wspomnieć również o możliwej roli powyższego twierdzenia w kontekście debaty nad deflacjonizmem[12]. Jeśli bowiem deflacjonista powinien przedstawić teorię dowodzącą ogólnych praw rządzących predykatem prawdy i konserwatywną nad PA, to w świetle wspomnianego wyniku ma bardzo ograniczone pole manewru. Nie może np. w jednej teorii połączyć własności „klasycznej kompozycyjności” i domknięcia na dowodliwość w rachunku zdań. Co być może bardziej zastanawiające – nie może połączyć uogólnionej kompozycyjności (implikującej warunek poprawności dysjunktywnej) i zasady poprawności aksjomatycznej PA. Wydaje się, że jest to sytuacja bez wyjścia. Wniosek ten otrzymamy jednak przy założeniu, że jego teoria ma dowodzić pewnych ogólnych praw rządzących predykatem prawdy i być konserwatywna. Powyższe wymaganie opiera się na innym znów założeniu: musimy przyjąć, że dowodliwość w teorii jest w naszym przypadku wystarczająco dobrą eksplikacją pojęcia uzasadnienia lub wyjaśnienia (w zależności od tego, jak dokładnie brzmi teza deflacjonizmu). Pogląd ten został ostatnio poddany wyczerpującej krytyce w (Cieśliński 2017, w przygotowaniu) i w tej chwili musimy przyznać, że kwestia wykorzystania powyższego twierdzenia w debacie nad deflacjonizmem pozostaje dla nas nierozstrzygnięta.

3.1. GRANICA TARSKIEGO A INNE TEORIE PRAWDY

Zauważmy, że o przebieg granicy Tarskiego możemy też pytać, używając jako punktu wyjścia innych teorii prawdy. Możemy na przykład zacząć od naszej (jak do tej pory) najmniejszej „naturalnej” niekonserwatywnej teorii czyli CT_0 i osłabić aksjomaty kompozycyjne, modelując je nie za pomocą logiki klasycznej, lecz np. silnej logiki

---

[12] Dziękujemy anonimowemu recenzentowi za sugestię umieszczenia tej uwagi w artykule.

Kleene'ego. W takiej teorii, znanej jako PT_0 [13], nie ma aksjomatu dla negacji w stylu CT^{-}, to znaczy

(Dla dowolnego zdania φ) Negacja φ jest prawdziwa wtedy i tylko wtedy, gdy φ nie jest prawdziwe.

Zamiast niego dla każdego spójnika (w tym i dla negacji) mówimy, co to znaczy, że prawdziwa jest negacja zdania złożonego z danym głównym spójnikiem. A zatem aksjomatem PT_0 jest na przykład zdanie

(Dla dowolnych dwóch arytmetycznych zdań φ, ψ) Negacja koniunkcji zdań φ i ψ jest prawdziwa wtedy i tylko wtedy, gdy negacja φ jest prawdziwa lub negacja ψ jest prawdziwa.

Możemy teraz zapytać: czy przebieg granicy Tarskiego zależy od tego jaką logikę wybierzemy dla kompozycyjnego predykatu prawdy spełniającego Δ_0 indukcję? Jest to jeden z obszarów naszych obecnych badań.

4. INNE POJĘCIA KONSERWATYWNOŚCI

Pytanie o konserwatywność jest tylko pierwszym krokiem w rozróżnianiu aksjomatycznych teorii prawdy. Służy ono do pierwszego przybliżenia, które teorie są silne, a które słabe. W ogólności możemy pytać o to, które są silniejsze od innych (tzn. rozróżniać także niekonserwatywne teorie). Służy do tego następujące, oczywiste w gruncie rzeczy, uogólnienie pojęcia konserwatywności:

Definicja 10. Teoria Th_1 jest s y n t a k t y c z n i e s i l n i e j s z a niż Th_2 wtedy i tylko wtedy, gdy konsekwencje arytmetyczne Th_2 stanowią właściwy podzbiór konsekwencji arytmetycznych Th_1.

I tak można pokazać, że CT_1 jest silniejsza arytmetycznie od CT_0, a CT od CT_1. Teorie niestratyfikowane, kompozycyjne i dowodzące wszystkich aksjomatów indukcji dla rozszerzonego języka są zazwyczaj jeszcze silniejsze, np. FS jest silniejsza niż CT, KF jest silniejsza niż FS, a VF jest silniejsza od KF[14].

[13] Tak naprawdę w literaturze znana jest teoria PT^{-} (albo $PT \upharpoonright$), a PT_0 to po prostu ta teoria z aksjomatami indukcji dla formuł klasy Δ_0 rozszerzonego języka.

[14] Nazwy tych teorii są już standardowe. Definicje KF i FS można znaleźć np. w (Halbach 2011), po definicję VF trzeba sięgnąć do artykułu Cantiniego (1990).

Zauważmy jednak, że pytanie jedynie o konsekwencje arytmetyczne nie pozwala nam rozróżniać wielu teorii, których aksjomaty mają zupełnie różny charakter. Np. zarówno CT^-, jak i TB są konserwatywne nad PA, a zatem są nieporównywalne za pomocą samych konsekwencji arytmetycznych. Można rozważać inną miarę siły pozwalającą rozróżniać teorie tak samo silne pod względem syntaktycznym. Miara ta jest oparta na, znanym wcześniej, pojęciu konserwatywności semantycznej:

Definicja 11. Teoria prawdy Th jest semantycznie konserwatywna nad PA wtedy i tylko wtedy, gdy każdy model PA możemy rozszerzyć (z zachowaniem uniwersum i funkcji arytmetycznych) do modelu dla Th.

Intuicja filozoficzna stojąca za tym pojęciem jest następująca: myślimy o modelach danej teorii jak o „światach możliwych" (z punktu widzenia tej teorii). Jeśli jakiegoś modelu PA nie można rozszerzyć do modelu dla teorii Th, to znaczy, że pewna możliwość dopuszczana przez PA jest eliminowana przez Th. Warto zauważyć, że konserwatywność semantycznie implikuje syntaktyczną (na mocy twierdzenia o pełności), ale nie zachodzi implikacja w drugą stronę: ani CT^-, ani TB nie są teoriami semantycznie konserwatywnymi. Pojęcie to można uogólnić w następujący sposób:

Definicja 12. Teoria prawdy Th_1 jest semantycznie silniejsza niż Th_2 wtedy i tylko wtedy, gdy klasa modeli PA, które można rozszerzyć do modeli teorii Th_1 jest właściwą podklasą klasy modeli PA, które można rozszerzyć do modeli teorii Th_2.

Odwołując się do wprowadzonej wyżej intuicji, możemy powiedzieć, że Th_1 jest semantycznie silniejsza od Th_2, jeśli Th_1 eliminuje więcej „światów możliwych" niż Th_2. Stosując to rozróżnienie, można pokazać że TB jest semantycznie słabsza niż CT^-, co odpowiada intuicji, że aksjomaty kompozycyjne „mówią więcej" o pojęciu prawdy niż same równoważności Tarskiego (nawet z pełną zasadą indukcji).

Najbardziej drobnoziarnista relacja pozwalająca rozróżnić teorie prawdy została wprowadzona przez Kentaro Fujimoto (2010) i znana jest jako relatywna definiowalność prawdy (*relative truth definability*).

Będziemy ją nazywać relacją *interpretowalności w sensie Fujimoto*:

Definicja 13. Niech Th_1, Th_2 będą teoriami prawdy. Powiemy, że Th_1 jest *interpretowalna w sensie Fujimoto* w Th_2 wtedy i tylko wtedy, gdy istnieje formuła $\varphi(x)$ taka, że Th_2 dowodzi wszystkich zdań powstających z aksjomatów Th_1 przez podstawienie $\varphi(x)$ w miejsce predykatu prawdy teorii Th_1.

Mówiąc najprościej: Th_1 jest interpretowalna w sensie Fujimoto w Th_2, jeśli Th_2 jest w stanie zdefiniować predykat prawdy spełniający aksjomaty Th_1. Powiemy, że Th_2 jest *silniejsza w sensie Fujimoto* niż Th_1 wtedy i tylko wtedy, gdy Th_1 jest interpretowalna w sensie Fujimoto w Th_2, ale Th_2 nie jest interpretowalna w sensie Fujimoto w Th_1. Dowód Twierdzenia 7 pokazuje właśnie, że CT_0 z zasadą poprawności PA jest interpretowalna w sensie Fujimoto w CT^- z zasadami poprawności dysjunktywnej i poprawności aksjomatycznej PA. Znamy przykłady teorii, które są rozróżniane dopiero przez powyższą relację np. TB^- i UTB^-[15].

5. PODSUMOWANIE I PYTANIA OTWARTE

Zaczęliśmy od wprowadzenia podstawowej miary siły aksjomatycznych teorii prawdy. Zgodnie z nią teoria silna to teoria, która dowodzi arytmetycznych zdań niedowodliwych w PA. Granicę między silnymi a słabymi rozszerzeniami podstawowej kompozycyjnej teorii prawdy CT^- nazwaliśmy granicą Tarskiego. Kluczowy wynik, o którym pisaliśmy, pokazuje, że jak do tej pory każda „naturalna" teoria prawdy, o której niekonserwatywności wiemy, dowodzi CT^- z zasadą poprawności PA. Ponadto ta ostatnia teoria ma wiele równoważnych aksjomatyzacji, a jedną z nich jest CT^- rozszerzona o schemat indukcji dla formuł ograniczonych z predykatem prawdy (teorię tę nazwaliśmy CT_0). Na koniec pokazaliśmy, że istnieją ciekawe „wzmocnienia" zaproponowanej przez nas relacji, pozwalające rozróżnić teorie, dla których podstawowa miara była zbyt „gruboziarnista". Warto podkreślić, że dalszym ciągu wiele naturalnych pytań dotyczących „siły" (różnorako

[15] Zdefiniowane w (Halbach 2011) jako $TB\restriction$ i $UTB\restriction$.

rozumianej) aksjomatycznych teorii prawdy pozostaje bez odpowiedzi. Oto kilka z nich:

1. Czy teoria CT^- rozszerzona o aksjomat dysjunktywnej poprawności jest konserwatywna nad PA?

2. Czy teoria CT^- rozszerzona o aksjomat „Wszystkie podstawienia tautologii klasycznego rachunku zdań są prawdziwe" jest konserwatywna nad PA? Zauważmy, że powyższy aksjomat to zasada poprawności odpowiadająca zasadzie domknięcia na klasyczny rachunek zdań. Ta ostatnia zasada jest zaś (nad CT^-) równoważna m.in. zasadzie globalnej refleksji, a zatem jest bardzo silna.

3. Czy teoria CT^- jest przynajmniej tak silna jak UTB (tzn. każdy model, który można rozszerzyć do modelu CT^-, można rozszerzyć do modelu dla teorii UTB; twierdzenie to zostało udowodnione w (Łełyk, Wcisło 2017b, w druku).

4. Czy teoria CT^- jest silniejsza w sensie Fujimoto niż teoria UTB?

BIBLIOGRAFIA

Cantini, A. (1990), *A Theory of Formal Truth Arithmetically Equivalent to ID_1*, „Journal of Symbolic Logic" 55(1), s. 244–259.

Cieśliński, C. (2017, w przygotowaniu), *The Epistemic Lightness of Truth. Deflationism and Its Logic*, Cambridge: Cambridge University Press.

Cieśliński, C. (2010), *Deflationary Truth and Pathologies*, „Journal of Philosophical Logic" 39(3), s. 325–337.

Cieśliński, C. (2010), *Truth, Conservativeness, and Provability*, „Mind" 119(474), s. 409–422.

Enayat, A., Visser, A. (2015), *New Constructions of Satisfaction Classes*, w: T. Achourioti, H. Galinon, J.M. Fernández (eds.), *Unifying the Philosophy of Truth*, s. 321–335, Dordrecht: Springer Netherlands.

Franzen, T. (2003), *Inexhaustibility. A Non-Exhaustive Treatment*, Association for Symbolic Logic.

Fujimoto, K. (2010), *Relative Truth Definability of Axiomatic Truth Theories*, „Bulletin of Symbolic Logic" 16(3), s. 305–344.

Halbach, V. (2011), *Axiomatic Theories of Truth*, Cambridge: Cambridge University Press.

Kaye, R. (1991), *Models of Peano Arithmetic*, New York: Clarendon Press.

Kaufmann, M., Schmerl, J. (1987), *Remarks on Weak Notions of Saturation in Models of Peano Arithmetic*, „Journal of Symbolic Logic" 52(1), s. 129–148.

Ketland, J. (1999), *Deflationism and Tarski's Paradise*, „Mind" 108(429), s. 69–94.

Kotlarski, H. (1986), *Bounded Induction and Satisfaction Classes*, „Zeitschrift für mathematische Logik und Grundlagen der Mathematik" 32(31–34), s. 531–544.

Kotlarski, H., Krajewski, S., Lachlan, A. (1981), *Construction of Satisfaction Classes for Nonstandard Models*, „Canadian Mathematical Bulletin" 24, s. 283–293.

Leigh, G. (2015), *Conservativity for Theories of Compositional Truth via Cut Elimination*, „Journal of Symbolic Logic" 80(3), s. 845–865.

Łełyk, M., Wcisło, B. (2017a, w druku), *Notes on Bounded Induction for the Compositional Truth Predicate*, „The Review of Symbolic Logic".

Łełyk, M., Wcisło, B. (2017b, w druku), *Models of Weak Theories of Truth*, „Archive for Mathematical Logic".

Shapiro, S. (1998), *Proof and Truth: Through Thick and Thin*, „Journal of Philosophy" 95(10), s. 493–521.

STRONG AND WEAK TRUTH PRINCIPLES

SUMMARY: This paper is an exposition of some recent results concerning various notions of strength and weakness of the concept of truth, both published and not. We try to systematically present these notions and their relationship to the current research on truth. We discuss the concept of Tarski's boundary between weak and strong theories of truth and we give an overview of nonconservativity results for the extensions of the basic compositional truth theory. Additionally, we present a natural strong theory of truth, which admits a number of apparently unrelated axiomatisations. Finally, we discuss other possible explications for the notion of 'strength' in axiomatic theories of truth.

KEYWORDS: axiomatic truth theories, Peano Arithmetic, conservativity, Tarski boundary