Universitatea din București

Facultatea de Filosofie

LUCRARE DE LICENȚĂ

Argumente de logică formală în dezbaterea despre referință: Quine și Putnam

Coordonator științific:
Prof. univ. dr. Mircea Dumitru

Autor:
Andrei-Alexandru Oltean

București, 2022

Cuprins

1 Introducere

În lucrarea de față îmi propun să analizez argumentele lui W. V. Quine și Hilary Putnam în legătură cu teza nedeterminării referinței. Mă voi concentra cu precădere asupra argumentelor formale prin care cei doi autori încearcă să își justifice poziția. Mai precis, îmi propun să ofer construcția logică de care Quine are nevoie pentru a-și formula așa-numitul *argument prin funcții proxy*; și, similar, să introduc teoremele folosite de Putnam în cadrul *argumentului său de teoria modelelor*. Deși cele două argumente sunt similare din punctul de vedere al concluziei pentru care militează, voi încerca să arăt că argumentul lui Quine este mai parcimonios în legătură cu premisele de care are nevoie, motiv pentru care voi consideră că este mai imun la critică. Îmi voi propune, totodată, să indic punctele slabe ale argumentului de teoria modelelor al lui Hilary Putnam, identificând în sfera de aplicare a metodei lui formale o posibilă sursă de eroare.

Din punct de vedere tehnic, destinația la care vor ajunge definițiile și demonstrațiile pe care le voi oferi este teorema Löwenheim-Skolem. În termenii cei mai generali cu putință, am putea spune că ea privește relația dintre limbaje formale și referința lor. Descrisă astfel, ea ar părea să adreseze de la bun început o problemă eminamente filosofică: problema referirii. Însă limbajul și lumea de care se leagă teorema Löwenheim-Skolem sunt construcții formale ale logicii matematice și nu e obligatoriu ca ele să coincidă, fără o discuție suplimentară, cu conceptele numite identic pe care le întrebuințează un filosof. Orice astfel de discuție ar trebui să clarifice mai întâi sensul tehnic matematic al teoremei, înainte de a arăta cum îi poate fi realizată puntea către filosofie. În primele două secțiuni îmi propun să pun bazele acestei discuții tehnice, urmând apoi să arăt cum se împletește ea cu probleme din filosofia matematicii și filosofia limbajului. Cu excepția cazului în care voi menționa altceva, noțiunile de teoria modelelor sunt adaptate din Kirby 2019.

2 Limbajul

> *„On mastering a finite vocabulary and a finitely stated set of rules, we are prepared to produce and to understand any of a potential infinitude of sentences."*
>
> — Davidson 1967

Să începem prin a introduce limbajul logicii de ordinul I, înăuntrul căruia atât Quine cât și Putnam vor presupune că poate fi formalizată întreaga știință. Vom introduce, pe rând, vocabularul unui limbaj formal (la care ne vom referi drept *semnătura* lui), apoi termenii care pot fi definiți pe baza acestui vocabular și, în fine, formulele de ordinul I.

Definiția 2.1 (Semnături)**.** O *semnătură* L e un tuplu $\langle R, F, C, V, ar \rangle$, unde:

i. R este o mulțime de simboluri pentru relații;

ii. F este o mulțime de simboluri pentru funcții;

iii. C este o mulțime de simboluri pentru constante;

iv. V este o mulțime de simboluri pentru variabile;

v. *ar* este o funcție care atribuie fiecărui simbol un număr natural pozitiv numit *aritate*.

Definiția 2.2 (Termeni)**.** Mulțimea termenilor lui L se definește recursiv astfel:

i. orice simbol de variabilă e un termen;

ii. orice simbol de constantă e un termen;

iii. dacă f e un simbol de funcție cu aritate n și $t_1,...,t_n$ sunt termeni, atunci $f(t_1,...,t_n)$ e un termen;

iv. doar ceva construit prin clauzele i. – iii. în număr finit de pași e un termen.

Definiția 2.3 (Formule)**.** Mulțimea formulelor lui L se definește recursive astfel:

i. dacă t_1 și t_2 sunt termeni, atunci $(t_1 = t_2)$ e o formulă;

ii. dacă $t_1,...,t_n$ sunt termeni și R e un simbol de relație de aritate n, atunci $R(t_1,...,t_n)$ e o formulă

Formulele construite doar prin clauzele i. – ii. se numesc *formule atomare*, pentru că ele servesc drept unități de bază care pot fi compuse prin conectori logici:

iii. dacă φ e o formulă, atunci $\neg\varphi$ e o formulă;

iv. dacă φ și ψ sunt formule, atunci $\varphi \wedge \psi$ e o formulă;

Formulele construite doar plin clauzele i. - iv. se numesc *formule necuantificate*.

v. dacă φ e o formulă și x e o variabilă, atunci $\exists x\ \varphi$ e o formulă;

O variabilă care nu e legată de un cuantificator se numește variabilă liberă. O formulă lipsită de variabile libere se numește enunț.

vi. doar ceva construit prin clauzele i. - v. în număr finit de pași e o formulă.

În plus, vom stipula următoarele echivalențe între formule:

a. $\varphi \vee \psi$ e o prescurtare pentru $\neg(\neg\varphi \wedge \neg\psi)$

b. $\forall x\ \varphi$ e o prescurtare pentru $\neg(\exists x \neg \varphi)$

Iată un fapt care se va dovedi crucial în demonstrația teoremei Löwenheim-Skolem: regulile de construcție a formulelor oferite mai sus produc o mulțime *infinit numărabilă* de formule, și nu mai mult de atât. Demonstrația acestui fapt se poate găsi în anexă la secțiunea 9.2.

Am oferit până acum o serie de reguli prin care se pot construi formule logice cu un anumit vocabular. Avansul pe care l-am făcut nu este nicidecum neînsemnat. Am putea, deja, suplini regulile de construcție ale formulelor cu o mulțime de reguli de deducție, adică reguli de *rescriere*, și am putea astfel demonstra enunțuri noi. În punctul în care ne aflăm, însă, termenii logici și formulele care sunt compuse din ei nu posedă încă o *semnificație*; nu sunt încă nimic mai mult decât simple înlănțuiri de simboluri. Dacă nu am avea de a face cu cuantificatori în limbaj, am putea asocia o semnificație formulelor într-o manieră foarte simplă: pur și simplu am asocia valori de adevăr formulelor atomare și am face demonstrații semantice pe baza tabelelor de adevăr. Dar cuantificarea complică acest calcul. Cuantificarea, în interpretarea ei clasică, necesită mai mult decât valori de adevăr – necesită *lucruri* peste care să se cuantifice. Pentru un logician, asta înseamnă că e nevoie de un domeniu pentru a da sens formulelor cuantificate. Iar pentru un filosof aplecat spre logica formală, acel domeniu reprezintă o *ontologie*. Vom vedea în continuare cum se pot introduce precis aceste noțiuni, sub denumirea tehnică de „structuri”, pe care o vom tot folosi de-a lungul acestei lucrări.

3 Interpretarea limbajului

Definiția 3.1 (Structuri)**.** O L-structură $\mathscr{A}$ constă într-o mulțime A numită *domeniu* și o mulțime de *interpretări*:

i. pentru orice simbol relațional R de aritate n, o submulțime $R^{\mathscr{A}}$ a lui A^n;

ii. pentru orice simbol de funcție f, o funcție $f^{\mathscr{A}} : A^n \to A$;

iii. pentru orice simbol de constantă c, un element $c^{\mathscr{A}}$ din A.

Definiția 3.2 (Interpretarea termenilor)**.** Interpretarea unui termen $t(\overline{x})$, unde $\overline{x}$ este o listă $x_1, ..., x_n$ de variabile libere prezente în termenul t, este o funcție $t^{\mathscr{A}} : A^n \to A$, definită astfel:

i. dacă t e constanta c, atunci $t^{\mathscr{A}}$ este funcția constantă care ia valoarea $c^{\mathscr{A}}$;

ii. dacă t e a i-a variabilă din $\overline{x}$, x_i, atunci $t^{\mathscr{A}}(\overline{a}) = a_i$;

iii. dacă t e funcția $f(t_1, ..., t_n)$, atunci $t^{\mathscr{A}}(\overline{a}) = f^{\mathscr{A}}(t_1^{\mathscr{A}}(\overline{a}), ..., t_n^{\mathscr{A}}(\overline{a}))$.

Definiția 3.3 (Interpretarea formulelor)**.** Fie $\varphi(\overline{x})$ o L-formulă cu variabile libere, $\mathscr{A}$ o L-structură și $\overline{a}$ o listă oarecare de elemente ale lui A. Atunci, spunem că $\mathscr{A}$ *satisface* $\varphi(\overline{x})$ cu parametrii $\overline{a}$, definind recursiv relația $\mathscr{A} \models \varphi(\overline{a})$:

i. $\mathscr{A} \models t_1(\overline{a}) = t_2(\overline{a})$ dacă și numai dacă $t_1^{\mathscr{A}}(\overline{a}) = t_2^{\mathscr{A}}(\overline{a})$

ii. $\mathscr{A} \models R(t_1(\overline{a}), \ldots, t_n(\overline{a}))$ dacă și numai dacă $(t_1^{\mathscr{A}}(\overline{a}), \ldots, t_2^{\mathscr{A}}(\overline{a})) \in R^{\mathscr{A}}$

iii. $\mathscr{A} \models \neg\varphi(\overline{a})$ dacă și numai dacă $\mathscr{A} \not\models \varphi(\overline{a})$

iv. $\mathscr{A} \models (\varphi \wedge \psi)(\overline{a})$ dacă și numai dacă $\mathscr{A} \models \varphi(\overline{a})$ și $\mathscr{A} \models \psi(\overline{a})$

v. $\mathscr{A} \models \exists\varphi(x, \overline{a})$ dacă și numai dacă există un $b \in A$ așa încât $\mathscr{A} \models \varphi(b, \overline{a})$

Dacă φ e un enunț și $\mathscr{A} \models \varphi$, spunem că $\mathscr{A}$ **este model pentru** φ. Dacă Γ e o mulțime de enunțuri și $\mathscr{A} \models \varphi$ pentru orice φ din Γ, vom scrie $\mathscr{A} \models \Gamma$ și vom spune că $\mathscr{A}$ **este model pentru** Γ.

Acum, dacă cel care citește va avea o aplecare filosofică, ar putea protesta: cum ar putea definiția 3.3 să reprezinte o atribuire de sens pentru formulele din limbajul L, din moment ce acele propoziții la care se reduce adevărul formulelor lui L rămân neinterpretate? Cu alte cuvinte, la ce bun să instituim o relație de felul „dacă și numai dacă” între două propoziții, dacă sensul propoziției din dreapta acestei relații este la fel de obscur ca al celei din stânga? Cum ne-ar putea clarifica înțelegerea, de pildă, să afirmăm că interpretarea unui simbol f din L este *o funcție*, din moment ce nu am explicitat în prealabil cum interpretăm cuvântul „funcție”? Această observație ar surprinde diferența dintre ceea ce a fost numit, grație lui Tarski, *limbaj obiect* și *metalimbaj*. În cazul nostru, L reprezintă limbajul obiect, iar limbajul (natural) în care am definit relația de satisfacere a formulelor lui L este metalimbajul. Înțelegerea metalimbajului este *presupusă* atunci când el este folosit pentru a defini interpretarea limbajului obiect. Mai departe, metalimbajul poate fi și el supus interpretării unui metametalimbaj, ocupând astfel locul unui limbaj obiect; însă regresia la infinit e prevenită prin faptul că, într-un punct anume din ierarhia metalimbajelor, adevărul unei propoziții în limbajul aflat pe acea poziție a ierarhiei *nu va mai fi controversabil*. În general, acel limbaj necontroversabil va fi considerat a fi chiar limbajul natural – parafrazându-l pe Quine, trebuie să ne mulțumim să luăm de bun limbajul natural. Vom reveni la această remarcă.

4 Un exemplu de structură. Teorema lui Tarski.

Avem acum la dispoziție aparatul formal necesar pentru a construi modele pentru mulțimi de propoziții formale. Ce înseamnă asta, în concret, pentru cineva preocupat de probleme filosofice? Iată un răspuns care merge în aceeași direcția în care va trata Hilary Putnam relația dintre filosofie și teoria modelelor. Fiecare disciplină științifică are propriul ei vocabular și se sprijină pe propriul ei nucleu de adevăruri incontestabile. Am putea, atunci, asocia vocabularului științei o semnătură (să îi spunem L), și am putea încerca să formulăm nucleul dur al disciplinei ca o mulțime de L-formule de ordinul I (să le spunem axiome). Sigur, disciplinele formale precum matematica sunt cele care se pretează cel mai bine la un asemenea procedeu de *înregimentare* logică. Însă, a priori, nu este evident niciun motiv pentru care științele empirice nu ar fi și ele formalizabile în logica de ordinul I. Să luam, așadar, exemplul teoriei mulțimilor. În mod minimal, vocabularul ei constă într-un singur simbol relațional, $\in$. Avem, deci, $L = \langle \in \rangle$ – dar va fi în general mult mai comod să folosim și simboluri precum „$\varnothing$”, „$\bigcup$”, „$\aleph$”, ca prescurtări pentru formule mult mai lungi care conțin doar $\in$. Axiomatizarea cea mai comună constă în nouă formule, axiomele ZFC. Accentul filosofic al acestei lucrări nu ne va permite să intrăm în prea multe detalii în legătură cu axiomele teoriei mulțimilor, deși ne vom folosi de ele pentru a oferi demonstrațiile metateoretice de care vom avea nevoie; ele sunt enumerate în anexă la secțiunea 9.1. Aceste axiome, luate împreună, permit definirea conceptelor matematice comune, precum perechi ordonate, relații, funcții, bijectivitate, cardinali. Avem, așadar, o semnătură și o mulțime de formule[1] construite în limbajul respectiv. Mai avem nevoie de un lucru – să facem acele formule să fie *adevărate*. Pentru asta, avem nevoie de o semantică, deci de un domeniu și de o funcție de interpretare.

Ne vom întreba, deci: în ce domeniu, și sub ce interpretare a simbolului $\in$, sunt adevărate axiomele teoriei mulțimilor? „În domeniul tuturor mulțimilor, firește! Și ce altă interpretare, alta decât interpretarea lui *uzuală* ca apartenență?” Acest răspuns e perfect corect; mai târziu, ne vom referi la acest model drept *modelul standard* al teoriei mulțimilor. Am putea cu ușurință nota domeniul tuturor mulțimilor cu litera V și i-am putea specifica extensiunea drept „mulțimea tuturor mulțimilor identice cu ele însele”; adică, într-o notație oarecum mai formală, $V = \{x \mid x = x\}$. Însă trebuie să fim atenți în ce fel de teorie ne plasăm *noi*, în metalimbaj, atunci când oferim un astfel de răspuns. În general, mulțimea V definită astfel s-ar putea nici măcar să nu fie un obiect despre care putem trata. Acesta ar fi cazul, de exemplu, dacă metateoria noastră ar fi tot teoria mulțimilor, pentru că „clasa tuturor mulțimilor” nu e o mulțime conform axiomelor ZFC.[2] Iar dacă metaaxiomele cu care lucrăm sunt doar axiomele teoriei mulțimilor, în mod cert nu vom putea să definim interpretarea modelului $\langle V, \in \rangle$, ca urmare a unui rezultat limitativ demonstrat de Tarski, pe care îl vom detalia mai jos.

Ce începem să intuim este că atunci când definim un model pentru teoria mulțimilor, trebuie să ne aflăm deja într-un limbaj *mai expresiv* decât cel al teoriei mulțimilor. Acest fapt nu vizează doar

[1] Pe parcursul acestei lucrări, voi folosi interschimbabil termenii „mulțime de formule” și „teorie”

[2] Pentru a vedea de ce, să presupunem că V ar fi o mulțime și să definim mulțimea $R = \{x \in V \mid x \notin x\}$. Datorită axiomei specificației, R ar trebui să desemneze o mulțime, deci un obiect al ZFC. Însă, desigur, această mulțime ar cauza paradoxul lui Russell, deci V nu e o mulțime conform ZFC. În general, o teorie nu își poate vedea clasa tuturor obiectelor ei tot ca un obiect al teoriei.

modelul $\langle V, \in \rangle$ pe care l-am oferit ca primă sugestie de structură. El se înscrie într-un principiu mai general: *nicio* interpretare a limbajului teoriei mulțimilor nu poate fi oferită tot în limbajul teoriei mulțimilor. E nevoie de un limbaj mai expresiv pentru a putea defini predicatul $\models$ pe care l-am introdus în 3.3, și, în general, e nevoie de un limbaj mai expresiv pentru a defini *orice* predicat care exprimă adevărul formulelor dintr-un limbaj. Acest lucru este cunoscut drept *teorema lui Tarski*, pe care o voi enunța și demonstra pe scurt în cele ce urmează.

Să desemnăm predicatul de adevăr prin simbolul „T" și să îl considerăm ca relație unară pe formule. Că „exprimă adevărul formulelor din limbaj" înseamnă că, pentru orice formulă φ, ar fi adevărat $\varphi \iff T(\varphi)$. Însă o astfel de notație nu ar fi tocmai exactă. T nu poate fi propriu-zis o relație unară *pe formule*, și deci nu putem lua φ ca parametru al lui T, pentru că o formulă nu e o mulțime, iar noi încercăm să definim T ca predicat al teoriei mulțimilor. Însă am văzut în 9.2 că există o asociere unu la unu a formulelor cu numerele naturale – or numerele naturale *sunt* reductibile la mulțimi. Gödel definește în mod explicit o astfel de funcție; fără a intra în detaliile construcției ei, să o notăm cu $g(\varphi)$. T ar putea fi, atunci, o relație pe numere naturale, iar echivalența s-ar putea rescrie ca $\varphi \iff T(g(\varphi))$.

Teorema 4.1 (Tarski)**.** Predicatul T nu poate fi definit în teoria mulțimilor. [Jech 2006]

Demonstrație. Vom folosi un argument diagonal. Folosind funcția lui Gödel, putem să ordonăm formulele din limbaj după numărul natural care le este asociat. Fie formula $\psi(x)$ definită ca $\neg T(g(\varphi_x(x)))$, unde φ_x indică formula căreia îi este asociat numărul x. Formulei ψ îi corespunde propriul număr natural, să îi zicem k. Atunci, $\psi(k) \iff \neg T(g(\varphi_k(k)))$ *(din definiție)* $\iff \neg T(g(\psi(k)))$ *(din alegerea lui k)* . Dar, din presupunerea că T e predicat de adevăr, $\psi(k) \iff T(g(\psi(k)))$. Am obținut o contradicție. □

Pentru a ajunge la acest rezultat, ne-am folosit în mod esențial de capacitatea autoreferențială a teoriei mulțimilor, aplicându-i lui $\psi(x)$ propria ei gödelizare ca parametru. Nu avem, însă, niciun motiv să presupunem că *orice* limbaj ar avea capacitate autoreferențială. Așadar, deși în teoria mulțimilor putem demonstra că nu există un predicat de adevăr, în limbaje mai puțin expresive acest stil de demonstrație nu ne este disponibil în mod necesar. Cu toate acestea, *concluzia* demonstrației este mult mai larg aplicabilă decât doar în legătură cu teoria mulțimilor. Într-adevăr, *niciun* limbaj nu își poate defini propriul predicat T. Semantica unui limbaj formal nu poate fi oferită decât de pe pozițiile unei teorii de fundal mai puternice; pentru că, așa cum ilustrează definiția 3.3, semantica *constă* în traducerea limbajului obiect în metalimbaj.

5 Nedeterminarea referinței și argumentul prin funcții proxy

Acum că am introdus mijloacele de bază prin care se construiește un limbaj formal și prin care i se atribuie o semantică, putem formula întrebarea centrală aflată la intersecția dintre filosofie și teoria modelelor. Și anume: care este relația dintre o teorie și modelele sale posibile? Ce fel de constrângeri impun enunțurile unei teorii asupra domeniului care trebuie să le satisfacă? Spre exemplu, sunt oare domeniile unic determinate pentru fiecare teorie? Ori, dacă nu *orice* teorie își determină în mod unic modelul, am putem spera măcar ca o teorie cu suficiente constrângeri să își caracterizeze în întregime obiectele despre care tratează?

Că asemenea întrebări atrag interes filosofic nu e greu de explicat. Nu e absurd să spunem că atunci când vorbim despre *lume*, în limbajul comun sau în limbajul tehnic al unei științe, ne-o închipuim ca fiind o mulțime de obiecte și de relații în care intră acele obiecte. Iar acele relații obiective sunt cele în virtutea cărora părerile noastre sunt adevărate sau false. Privită astfel, lumea ia aparența unui model, iar intersecția ontologiei cu teoria modelelor încetează să mai pară o alăturare arbitrară de domenii. Iată atunci în ce constă pentru un filosof însemnătatea întrebărilor din paragraful anterior. Știm că teoriile noastre curente au succes. Dar despre ce lume tratează aceste teorii? Succesul lor ne garantează totodată accesul imediat la realitate, așa cum e ea în sine? Sau este posibil ca teoriile noastre să fie tot atât de bine satisfăcute de *o mulțime de ontologii diferite*, dintre care să ne fie imposibil să alegem una singură corectă, *reală*? În mod dezamăgitor pentru cei care păstrează aspirația accesului imediat la realitate, vom vedea că mijloacele formale pentru a varia modelele oricărei teorii *există*. Teorema Löwenheim-Skolem, în diferitele ei forme, este o astfel de afirmație existențială. Însă rămâne deschis dezbaterii filosofice dacă aceste mijloace formale sunt aplicabile în contextul general al tuturor științelor; și dacă, chiar aplicabile fiind, există criterii de a distinge *realitatea* de celelalte ontologii posibile.

Situația în care o teorie admite mai multe modele poate fi exemplificată fără a face uz de aparatură formală suplimentară. Numerele naturale, de pildă, pot fi identificate cu mulțimi fie în maniera propusă de Frege (numărul n este clasa tuturor mulțimilor cu n elemente), fie în cea a lui Zermelo (n este reuniunea tuturor numerelor mai mici decât n), fie în cea a lui von Neumann (n este mulțimea care are ca unic element pe $n-1$). Oricare din aceste variante este suficientă pentru a satisface axiomele lui Peano. Care dintre ele, atunci, surprinde esența reală a numerelor naturale? În *What Numbers Could Not Be*, Paul Benacerraf argumentează în favoarea unei concepții structurale a numerelor naturale. Un număr natural oarecare, potrivit lui Benacerraf, nu este *identic* cu niciuna din mulțimile propuse de Frege, Zermelo sau von Neumann; pentru că un număr natural nu e un *obiect*, ci este o poziție în progresia tuturor numerelor naturale. Mai general, ce identifică un număr natural sunt relațiile în care el intră față de toate celelalte numere. Chestiunea ontologiei numerelor naturale este astfel retrogradată ca neesențială: *orice* obiect poate sta drept un număr natural, dacă intră în relațiile corespunzătoare cu celelalte obiecte.[3] Cu alte cuvinte, poziția lui Benacerraf este că *orice model al aritmeticii este acceptabil*; fără ca numerele naturale să fie identice cu elementele vreunui model în particular, ci cu poziții din structura aritmetică pe care toate modelele trebuie să o redea.

Dar a privi ontologia drept neesențială nu înseamnă a o dizolva ca problemă. Ar fi evident absurd să susținem că ne putem descotorosi de obiecte în favoarea relațiilor, pentru că relațiile sunt aplicabile doar între obiecte. Poziția lui Benacerraf are, în schimb, consecința de *a relativiza* ce obiecte pot sta *drept* numere naturale. Câtă vreme conservă structura relațională de care are nevoie aritmetica, orice domeniu de obiecte poate sta drept referința termenilor aritmetici. Am început, deci, prin a ne întreba dacă teoriile noastre în ansamblu își pot determina în mod unic obiectele care le satisfac și vedem acum, în cazul particular al aritmeticii, că răspunsul este *negativ*. Vedem, cu alte cuvinte, că aritmetica imprimă relațiile în care obiectele ei trebuie să intre, însă nu identifică și obiectele în sine. Obiectele sunt lăsate la latitudinea interpretării pe care o acordăm limbajului aritmetic: interpretarea lui Frege, ori a lui Zermelo, ori a lui von Neumann. Este important să remarcăm că interpretarea e un procedeu lingvistic. În secțiunea 3.3, am văzut că ea constă într-o mulțime de propoziții construite conform schemei „φ e adevărată dacă și numai dacă Φ”, unde φ e o formulă din limbajul obiect și Φ e o expresie din metalimbaj. Din punct de vedere filosofic, urmându-l pe Davidson, putem echivala o astfel de schemă cu una de tipul φ *înseamnă* Φ, identificând condițiile de adevăr ale unei propoziții cu înțelesul ei. Așadar, rolul pe care îl joacă o interpretare este cel al unui manual de traducere: specifică în metalimbaj înțelesul formulelor dintr-un limbaj obiect. Iar specificându-le înțelesul, le specifică și referința. Întorcându-te la Benacerraf, ne putem declara convinși că suntem liberi să variem manualul de traducere a limbajului aritmetic, câtă vreme nu îi afectăm structura relațională. Poziția structuralistă a lui Benacerraf poate fi, deci, sintetizată astfel: obiectele aritmetice sunt relative la felul în care *limbajul aritmeticii e tradus în limbajul teoriei mulțimilor*. Iar a nu afecta structura relațională ce presupune, în fond? Doar ca enunțurile care *ne așteptăm* să fie adevărate (ie.: cele care pot fi demonstrate plecând de la axiome) să fie traduse în expresii *care sunt* adevărate în metalimbaj. Așadar, orice traducere care face teoria aritmetică să fie adevărată e acceptabilă ca determinant al referinței ei. Părem să fi descoperit o teorie a referinței (în miniatură) pentru limbajul aritmetic. Rezultatele ar fi însă catastrofale dacă am aplica-o asupra limbajului în ansamblul lui – din motive care țin de o dezbatere inițiată de Hilary Putnam, despre care urmează să vorbim.

Dar de ce ne-am gândi să o aplicăm asupra întregului limbaj? La mijloc s-ar putea afla temerea că referința *tuturor* termenilor, nu doar a celor aritmetici, poate fi specificată în moduri

[3] Benacerraf 1965, în special pp. 69-73. Spre exemplu, ce îl identifică pe 3 este că 1 și 2 sunt mai mici decât el, iar el este mai mic decât oricare alt număr diferit de 3 însuși. Orice obiect poate ocupa locul lui 3, dacă domeniul per ansamblu satisface condiția lui de identificare sub o interpretare anume a relației „mai mic decât”.

multiple. Când Quine vorbește despre „o teorie relațională a ceea ce sunt obiectele teoriilor”,[4] argumentele lui merg în această direcție. Teza lui este că traducerea dispozițiilor lingvistice ale oricărei alte persoane în limbajul persoanei proprii se poate face în moduri distincte, dar imposibil de diferențiat empiric. Sub umbrela acestei teze intră o clasă enormă de traduceri, căci ea vizează în egală măsură traducerea unei limbi extraterestre, interpretarea cuvintelor vecinului tău (care vorbește aceeași limbă ca tine, însă care ar putea să *înțeleagă* altceva prin aceleași cuvinte) sau reducerea unei teorii științifice la o alta (în sensul în care fenomenele chimice, de pildă, sunt reductibile la fenomene fizice). La baza acestei convingeri stă respingerea semnificațiilor ca entități și a numelor ca etichete puse pe respectivele entități.[5] Concepția quineană a semnificației este, în schimb, în întregime empiristă și behavioristă: semnificațiile sunt corelații între stimuli empirici și dispoziții lingvistice. Semnificația unui enunț precum „Pisica stă pe masă” constă în acei stimuli empirici care m-ar determina să aprob că pisica stă pe masă. O astfel de naturalizare a semnificațiilor produce ecouri puternice în filosofia limbajului. Spre exemplu: nu toate enunțurile vor avea o semnificație în mod direct, în sensul că nu vor fi în general condiționate de niciun stimul empiric. Acesta e cazul enunțurilor teoretice. Ele sunt semnificative doar prin faptul că adevărul lor condiționează adevărul *unor alte enunțuri*, abia acestea din urmă fiind corelate direct cu stimuli. O consecință și mai importantă din perspectiva discuției actuale este cea pe care am formulat-o mai devreme: traducerile între aproape orice pereche de limbaje nu sunt unic determinate. Pentru că o traducere este o funcție între limbaje care *conservă semnificația* rostirilor; drept urmare o traducere trebuie să transpună enunțuri pe care vorbitorul unei limbi ar fi dispus să le aprobe sub o mulțime de stimuli, în enunțuri pe care vorbitorul celeilalte limbi ar fi dispus să le aprobe sub aceeași mulțime de stimuli. Cu alte cuvinte, o traducere trebuie să trimită adevăruri în adevăruri. *Dar există numeroase funcții de traducere între două limbaje oarecare care trimit adevăruri în adevăruri.* În consecință, fiecare din aceste traduceri *va atribui o referință diferită* termenilor din limbajul-sursă. Așadar, referința termenilor din limbajul-sursă *este relativă* la interpretarea în limbajul nostru pe care le-o oferim. Dacă tot acest mănunchi de teze filosofice e adevărat, atunci consecințele lor se revarsă puternic asupra ceea ce presupune *înțelegerea* unei limbi. Căci actul de traducere nu ar fi doar o preocupare strâmtă a lexicografului: orice interacțiune lingvistică cu altcineva ar presupune o traducere. Pentru a-i înțelege pe ceilalți, conform lui Quine, suntem întotdeauna obligați să interpretăm discursul lor în discursul nostru propriu. Asta în timp ce avem la dispoziție o mulțime de interpretări distincte, în mod egal de bune!

Vom vedea în scurt timp cum se poate justifica existența mai multor manuale de traducere. Însă înainte de asta trebuie făcute câteva precizări, în lipsa cărora am risca să cădem pradă unor conclu-

[4] Quine 1968, p. 201: „[...] a relational theory of what the objects of theories are”.

[5] Quine descrie metaforic poziția pe care o critică drept „limba ca un muzeu”: exponatele sunt semnificații, numele sunt plăcuțele indicative din dreptul fiecărui exponat.

zii absurde. Calificam mai devreme drept „enormă” clasa traducerilor vizate de Quine – cam cât de mare e „enorm”, totuși? Ori, pentru a pune problema altfel – ce fel de traduceri sunt excluse din această clasă? Suntem liberi să interpretăm vorbirea unui extraterestru și pe a vecinului nostru – dar suntem liberi să ne reinterpretăm și *propriii termeni*? Să vedem, oare ce ar presupune asta? Ar trebui să ne angajăm la un manual de traducere care ar avea reguli de tipul „«X» se referă la Y”, în condițiile în care *și X, și Y sunt cuvinte din propriul vocabular*. De pildă: „«Iepure» se referă la complemente cosmice ale iepurilor”. Dar asta ar fi contradictoriu! Pentru că formularea însăși a manualului de traducere nu se poate face decât înăuntrul unui vocabular, presupunând deja cunoscut înțelesul cuvintelor folosite în regulile de traducere. Atunci când *menționăm* termenul Y într-o regulă de traducere, îi folosim semnificația; dar dacă ne îndoim de semnificația propriilor noastre cuvinte, trebuie să admitem că nici semnificația lui Y nu e fixată, deci că manualul de traducere nu specifică până la capăt referința lui X. „Dacă întrebăm: «Se referă oare ’iepure’ la iepuri?» cineva poate contra cu următoarea întrebare: «Să se refere la iepuri în ce sens al lui ’iepuri’?», avântându-se astfel într-un regres; avem nevoie de un limbaj de fundal în care să regresăm”.[6] Limbajul de fundal în care vor fi formulate toate manualele de traducere va fi *limbajul propriu*, pe care, în cuvintele lui Quine, trebuie *să acceptăm să îl luăm de bun*. „În practică oprim regresul la limbi de fundal acceptând limba proprie și luându-i cuvintele așa cum sunt.”[7] Deci referința cuvintelor unui străin este relativă de două ori: întâi este relativă la un manual de traducere, apoi este relativă la limbajul de fundal în care respectivul manual de traducere a fost formulat. Acel limbaj este limba maternă a celui care concepe traducerea; iar de sensul propriei lui vorbiri nu se mai poate îndoi.

Așadar, pentru noi înșine *întotdeauna* va fi cazul că „iepure” se referă la iepuri. Dar acest lucru nu ar trebui să ne consoleze prea mult: limba nu poate fi niciodată privată, ci e o activitate socială. Să revenim, deci, la teza centrală pe care ne-am propus să o demonstrăm: că întotdeauna vor exista interpretări distincte pentru unul și același limbaj. Argumentul cel mai direct pe care Quine îl aduce în favoarea acestei idei este unul de natură formală. Pentru a accepta relativitatea referențială a discursului aritmetic, am avut nevoie să găsim modele diferite pentru aritmetică; pentru a accepta relativitatea referinței *în general*, am avea nevoie de o procedură prin care să obținem un model nou plecând de la *orice* model al unei teorii oarecare. Quine demonstrează că cerința pentru o astfel de procedură poate fi îndeplinită cu ușurință. Dacă am putea specifica o funcție care să asocieze fiecărui obiect dintr-un univers M câte un reprezentant dintr-un univers M', în așa fel încât toate relațiile care țineau pentru obiecte din M să fie reinterpretate ca ținând pentru reprezentanții lor din M', am fi capabili să oferim modele distincte pentru orice teorie. Astfel de

[6] „When we ask, "Does ’rabbit’ really refer to rabbits?" someone can counter with the question: "Refer to rabbits in what sense of ’rabbits’?" thus launching into a regress; and we need the background language to regress into.” Quine 1968, p. 200.

[7] „[I]n practice we end the regress of background languages, in discussions of reference, by acquiescing in our mother tongue and taking its words at face value.” ibid., p. 201.

funcții, Quine le numește *funcții proxy*; ele ne vor servi ca manual de traducere, ori funcții de reinterpretare. Funcții proxy pentru diferite clase de teorii nu sunt greu de găsit. Am menționat deja câteva exemple, de altfel. Pe una am folosit-o în demonstrația teoremei 4.1: funcția lui Gödel. Cu ajutorul ei am reinterpretat obiectele sintaxei, formulele, ca numere naturale; pe care, la rândul lor, le-am tratat implicit ca fiind mulțimi. În felul acesta am înglobat discursul despre formule în discursul despre mulțimi. Putem da exemple de funcții proxy și pentru teorii mai puțin abstracte decât sintaxa. Ne întrebam mai devreme dacă am putea reinterpreta „iepure" ca referindu-se la complemente cosmice ale iepurilor. Câtă vreme „iepure" e rostit de altcineva, am putea. Căci orice teorie fizicalistă, adică orice teorie ale cărei obiecte sunt regiuni din spațiu, admite funcția de *complement cosmic* ca funcție proxy;[8] adică funcția care trimite o regiune x din spațiu în *întregul spațiu, mai puțin x*. Nu e greu de văzut cum poate fi folosită această funcție la construcția unui nou model. Fiecare regiune din spațiu din universul teoriei o înlocuim cu complementul ei cosmic; tot așa, toate relațiile care țineau în vechiul univers între regiuni din spațiu vor ține în noul univers între complementele cosmice ale acelor regiuni din spațiu. Dacă reinterpretăm apoi enunțurile teoriei ca fiind adevărate nu despre obiectele și relațiile lor uzuale, ci despre cele cu care le-am înlocuit în noul univers, vom vedea că niciun enunț nu-și va schimba valoarea de adevăr. Teoria ține la fel de bine în universul clasic, pe cât ține în universul complementat.[9] Cu alte cuvinte, teoriei fizicaliste îi este indiferentă referința precisă a termenilor săi: orice termen X se poate referi și la complementul cosmic al referinței uzuale a lui X, fără ca adevărul teoriei per ansamblu să fie afectat.

Acestea sunt, însă, doar câteva exemple particulare. Am putea spera, oare, să găsim o funcție proxy pentru orice model al unei teorii? Există un truc formal care ne confirmă că răspunsul e pozitiv. Și anume: orice permutare a obiectelor dintr-un domeniu, reflectată în mod corespunzător asupra relațiilor și asupra interpretării termenilor, conservă adevărul tuturor enunțurilor. Vom vedea în continuare că această afirmație reprezintă un adevăr metalogic demonstrabil cu ușurință odată ce conceptele necesare sunt introduse. Să le introducem, așadar.

În afara contextului filosofic, funcțiile proxy ale lui Quine sunt cunoscute ca *scufundări elementare*: funcții care trimit domeniul unui model în domeniul altuia, conservând adevărul tuturor formulelor. Definiția care urmează nu este nimic altceva decât formalizarea a ceea ce am prezentat deja în limbaj natural.

Definiția 5.1 (Scufundări)**.** O scufundare a unei L-structuri $\mathscr{A}$ într-o L-structură $\mathscr{B}$ este o funcție

[8] Quine [1994] 2008, pp. 457-8.

[9] Inspirându-ne dintr-un exemplu al lui Putnam, să luăm „pisică" și „preș" ca fiind termenii folosiți în sensul nostru uzual, și „pisică*" și „preș*" ca fiind termenii (omofoni) folosiți de un locuitor al universului complementelor cosmice. Prin felul în care am construit acest univers – mai precis, datorită condiției ca toate relațiile noastre uzuale între obiecte să poarte și asupra reprezentanților obiectelor din noul model – „Pisica* stă pe preș*" e adevărat dacă și numai dacă pisica stă pe preș. Așadar, atât universul nostru cât și universul remodelat sunt în acord asupra valorii de adevăr a acestui enunț (și asupra tuturor celorlalte). Însă termenii au altă referință în cele două universuri. Vezi Putnam 1981, cap. 2. Vom reveni asupra comentariului lui Putnam.

$\pi: A \to B$ care satisface următoarele condiții:

i. π este injectivă ($\pi(a_1)$ e diferit de $\pi(a_2)$, pentru orice a_1, $a_2 \in A$)

ii. pentru orice simbol de constantă c din L, $c^{\mathscr{B}} = \pi(c^{\mathscr{A}})$

iii. pentru orice simbol relațional R din L și orice $a_1,...,a_n \in A$, $(a_1,...,a_n) \in R^{\mathscr{A}}$ dacă și numai dacă $(\pi(a_1),...,\pi(a_n)) \in R^{\mathscr{B}}$

iv. pentru orice simbol funcțional F din L și orice $a_1,...,a_{n+1} \in A$, $F^{\mathscr{A}}(a_1,...,a_n) = a_{n+1}$ dacă și numai dacă $F^{\mathscr{B}}(\pi(a_1),...,\pi(a_n)) = \pi(a_{n+1})$.

În particular, orice permutare a unui domeniu este o scufundare; adică exact la ce ne-am aștepta după ce am parcurs argumentele lui Quine. Să definim mai precis ce înțelegem prin „permutare”:

Remarcă. Dacă π este chiar *bijectivă*, adică pentru orice $b \in B$ există un unic $a \in A$ cu $\pi(a) = b$, spunem că π este un *izomorfism*.

Remarcă. Permutările sunt cazuri particulare de izomorfisme. Mai precis, o permutare este o scufundare bijectivă a lui $\mathscr{A}$ în $\mathscr{A}$, care nu e funcția identitate $\pi(x) = x$.

Dacă există o scufundare a lui $\mathscr{A}$ în $\mathscr{B}$, înseamnă asta că cele două structuri se acordă în adevărul tuturor formulelor? Nu neapărat. Scufundările asigură că valoarea de adevăr a formulelor *necuantificate* rămâne neschimbată. Acest fapt se poate demonstra prin inducție pe formule, însă demonstrația în sine este oarecum mecanică și mai puțin interesantă pentru o lucrare al cărei scop principal este unul filosofic; ea poate fi consultată în anexă, în secțiunea 9.4. Însă, deși scufundările nu afectează satisfacerea formulelor necuantificate, nu același lucru se întâmplă în mod obligatoriu în legătură cu valoarea de adevăr a *oricărei* formule, inclusiv din cele cu cuantificatori.[10]

Putem, atunci, rafina noțiunea de scufundare în așa fel încât să ne asigurăm că îndeplinește cerința de acord absolut al adevărurilor. Am avea, atunci:

Definiția 5.2 (Scufundări elementare)**.** O scufundare π a unei L-structuri $\mathscr{A}$ într-o L-structură $\mathscr{B}$ se numește elementară dacă, pentru orice L-formulă $\varphi(\bar{x})$ (unde $\bar{x} = x_1,...,x_n$ este o listă de L-variabile care include toate variabilele libere ale lui φ) și orice listă de parametri $a_1,...,a_n \in A$, e adevărat că:

$$\mathscr{A} \models \varphi(a_1,...,a_n) \text{ ddacă } \mathscr{B} \models \varphi(\pi(a_1),...,\pi(a_n))$$

În particular, se poate demonstra că:

[10]Spre exemplu, numerele întregi cu relația de ordine $\leq$ formează o structură $(\mathbb{Z},0,1,\leq)$ care poate fi scufundată în structura $(\mathbb{R},0,1,\leq)$ prin funcția identitate $\pi(x) = x$. Însă $(\mathbb{Z},0,1,\leq) \models \forall x\,(x \leq 0) \vee (x = 1) \vee \neg(x \leq 1)$, adică în structura numerelor întregi nu există niciun număr aflat între 0 și 1. Evident, structura numerelor reale nu satisface aceeași formulă.

Teorema 5.1. Orice izomorfism e o scufundare elementară.

Demonstraţie. Odată stabilită teorema 9.4, e suficient să arătăm că pentru orice $\varphi(\bar{x},y)$ și orice $a_1,\ldots,a_n \in A$:

$$\mathscr{A} \models \exists y\ \varphi(\bar{a},y) \text{ ddacă } \mathscr{B} \models \exists y\ \varphi(\pi(\bar{a}),y)$$

Conform semanticii tarskiene, avem:

$$\begin{aligned}\mathscr{A} \models \exists y\ \varphi(\bar{a},y) &\text{ ddacă există un } d \in A \text{ așa încât } \mathscr{A} \models \varphi(\bar{a},d)\\ &\text{ ddacă există } \pi(d) \in B \text{ așa încât } \mathscr{B} \models \varphi(\pi(\bar{a}),\pi(d))\\ &\text{ ddacă } \mathscr{B} \models \exists y\ \varphi(\pi(\bar{a}),y)\end{aligned}$$

□

Corolar. Orice permutare e o scufundare elementară; pentru că orice permutare e un izomorfism.

Această ultimă teoremă și corolarul ei reies, așadar, aproape imediat din definiții. Iată, în cuvinte simple, ce ne spune teorema: dacă o teorie de ordinul I e adevărată într-un model, atunci e adevărată în orice permutare a elementelor din respectivul model. Reformulat într-o notă mai filosofică, nicio teorie nu este suficient de puternică încât să își determine în mod unic interpretarea termenilor; pentru că întotdeauna vor mai exista și alte interpretări izomorfe (în particular, permutări) sub care exact aceleași propoziții vor fi adevărate.

Or, argumentul lui Quine se prezintă astfel. Presupunând că am putea formaliza orice teorie din limbaj natural în limbajul logicii de ordinul I, teorema 5.1 și corolarul ei ar trebui să fie aplicabile asupra oricărei teorii. Cu ajutorul ei putem transpune o parte a științei în diverse alte părți, putem transpune cuvintele cuiva care vorbește aceeași limbă cu noi în diverse interpretări neomofone, și putem transpune în diverse moduri o limbă extraterestră în limba noastră. Iar acest lucru, susține Quine, ar trebui să ne convingă să abandonăm ideea că există o referință absolută a termenilor și o ontologie a lucrurilor în sine de care acele referințe aparțin; ci doar referință relativizată la interpretare și la limbajul de fundal. Pentru că singurul mod în care referința poate fi specificată este prin traducerea respectivilor termeni în limba noastră comună, ale cărei cuvinte trebuie să le acceptăm de bun sensul. Izomorfismele demonstrează multiplicitatea acestor traduceri – deci multiplicitatea referinței. Ce este cel mai important, în mod absolut ne putem pronunța doar despre structura relațională pe care o teorie o imprimă asupra obiectelor sale, și pe care orice izomorfism o conservă prin definiție. Concluzia aceasta este sintetizată de Quine într-o singură propoziție: „Structura e ceea ce contează la o teorie, nu felul în care îi sunt alese obiectele.”[11] Aici găsim

[11] „Structure is what matters to a theory, and not the choice of its objects.” Quine 1981, p. 20.

esența argumentului quinean pentru nedeterminarea referinței.

Remarcabil este că raționamentul lui Quine are nevoie de foarte puține presupoziții ca să funcționeze. Existența izomorfismelor e un adevăr logic; mai departe, trebuie doar să acceptăm 1. că ontologia este o funcție a limbajului, și nu viceversa („a fi înseamnă a fi valoarea unei variabile"), și în consecință 2. că semnificațiile nu sunt entități în sine, mai presus de felul în care e folosit limbajul. E adevărat, cineva care preferă în filosofia matematicii să adopte o poziție platonistă ar refuza poziția 1., susținând că limbajul aritmetic e construit *cu intenția* de a se referi la *ceva* independent de el. Acel ceva poate fi luat în varii moduri: ar putea fi entități *bona fide* din universul platonic, sau ar putea fi conținuturi ale intuiției. Aceasta e o problemă care va deveni relevantă odată ce vom trece la discuția teoremei Löwenheim-Skolem. Deocamdată, însă, dintr-un punct de vedere pur empirist, e complet plauzibil să acceptăm premisele de care are nevoie Quine și, deci, să îi acceptăm concluzia. Dar trebuie totuși să tratăm cu delicatețe concluzia lui Quine și să nu ne avântăm în presupuneri care depășesc ce a fost demonstrat. Căci concluzia nu e una sceptică, nici antirealistă: *există* o lume exterioară, iar cuvintele *proprii* ale oricărei persoane se referă la obiecte externe din realitate în mod complet determinat. Doar în ochii *altuia* obiectele din realitate devin locuri într-o structură. Chiar fără a fi sceptică, aceasta e o afirmație absolut remarcabilă: Benacerraf înlocuia ontologia aritmeticii cu o structură relațională; Quine înlocuiește ontologia oricărei teorii cu o structură relațională.

Următorul pas din mișcarea de *dizolvare a ontologiei* e al lui Hilary Putnam. Poziția lui este de departe cea mai radicală dintre toate. Putnam încearcă să ne convingă că, ducând premisele până la consecințele lor ultime, referința nu e fixată *nici măcar pentru vorbitor*. Ca urmare, *aproape* orice teorie poate fi adevărată, în sensul de a corespunde unor stări de fapt din lume, căci suntem liberi să stabilim orice corespondență a termenilor ei cu lumea dorim. Putnam recunoaște absurditatea acestei concluzii și identifică problema în premisele asumate. Anume, în premisa că există o realitate exterioară, independentă de mintea umană și de capacitățile noastre de a o cunoaște. Dacă argumentul lui are să fie unul de succes, Putnam ar vrea să înlocuim realismul semantic – teza conform căreia adevărul unui enunț depinde de referința termenilor lui din realitatea obiectivă – cu o poziție *antirealistă*. Adică ar vrea să îl urmăm în a crede că adevărul unui enunț e determinat de existența unei proceduri de investigație prin care am putea ajunge *noi* să aprobăm respectivul enunț, fără a ține cont de vreo presupusă corespondență a lui cu realitatea. Putnam încearcă, așadar, să facă pasul de a ceda *în întregime* locul referinței în filosofia limbajului în favoarea normelor de comportament social. Nu e deloc evident, la o primă vedere, cum ar putea fi dus acest țel la bună îndeplinire. Modul în care încearcă Putnam să își construiască argumentul este tot unul cvasiformal, așa cum este și argumentul lui Quine prin funcții proxy. Putnam invocă teorema Löwenheim-Skolem, folosindu-se astfel de o formă mai puternică de echivalență lingvistică decât pot oferi izomorfismele. Căci teorema Löwenheim-Skolem garantează că anumite teorii pot fi satisfăcute de modele *de cardinalități diferite*, pe câtă vreme izomorfismele nu pot fi instituite decât între modele de aceeași cardinalitate. Teorema Löwenheim-Skolem este ea însăși încărcată filosofic; nu pentru că demonstrația ei ar avea nevoie de vreo presupunere filosofică, ci pentru că încă de când a apărut a provocat dezbateri în legătură cu interpretarea corectă a axiomelor teoriilor formale. Este orice model care satisface axiomele teoriei mulțimilor, de pildă, o referință acceptabilă pentru termenii teoriei mulțimilor? Are sens să vorbim despre vreo intenție referențială care să fi fost atribuită limbajului teoriei mulțimilor atunci când teoria mulțimilor a fost axiomatizată? Sau suntem liberi să alegem orice referință dorim, cu condiția ca ea să facă teoria adevărată? Acesta este tipul de întrebări filosofice care înconjoară teorema Löwenheim-Skolem. Vom vedea că și autori care altminteri s-ar pronunța pentru o formă de structuralism vor susține că unele modele sunt pur și simplu inacceptabile. Din discuția aceasta care ține de filosofia matematicii se va inspira Putnam pentru a-și apăra antirealismul generalizat, luând partea celor care susțin că nu există nicio intenție de a te referi la un anumit model în formalizarea teoriilor. Să facem, așadar, un ocol până în anii '20 ai secolului trecut, când Thoralf Skolem propune o demonstrație nouă pentru o teoremă a lui Leopold Löwenheim, și remarcă un aparent paradox pe care această teoremă ar părea că îl generează în teoria axiomatizată a mulțimilor.

6 Teorema Löwenheim-Skolem și paradoxul lui Skolem

Să începem prin a defini noțiunile cu care teorema Löwenheim-Skolem va lucra. Primul lucru remarcabil pe care ea îl demonstrează este că două structuri pot fi puse în acord în legătură cu adevărul tuturor propozițiilor, fără a fi însă izomorfe. Numele tehnic pentru o astfel de structură este acela de *substructură elementară*. Să vedem, pentru început, ce anume este o substructură:

Definiția 6.1 (Substructuri)**.** O structură $\mathscr{A}$ se numește substructură a lui $\mathscr{B}$ dacă $A \subseteq B$ și funcția identitate $\pi\colon A \to B$ $\pi(x) = x$ este o scufundare a lui $\mathscr{A}$ în $\mathscr{B}$.

Următoarea teoremă e o consecință imediată a incluziunii lui A în B:

Teorema 6.1. Fie $\mathscr{A}$ o L-substructură a lui $\mathscr{B}$. Fie $\varphi(\bar{x}, y)$ o L-formulă necuantificată și $\bar{a} \in A$. Atunci:

$$\text{Dacă } \mathscr{B} \models \forall y\ \varphi(\bar{a}, y), \text{ atunci } \mathscr{A} \models \forall y\ \varphi(\bar{a}, y).$$

$$\text{Dacă } \mathscr{A} \models \exists y\ \varphi(\bar{a}, y), \text{ atunci } \mathscr{B} \models \exists y\ \varphi(\bar{a}, y).$$

Demonstrație. Demonstrația poate fi oferită cel mai precis prin explicitarea condițiilor de adevăr pentru cele două formule, însă esența demonstrației poate fi surprinsă în câteva cuvinte astfel.

Întrucât $A \subseteq B$ și funcția de scufundare a lui A în B este identitatea, este evident că dacă toate elementele lui B satisfac o anumită proprietate, atunci și elementele lui A o vor satisface; și dacă există un element în A care satisface o anumită proprietate, atunci acel element va fi și în B. □

Acum, la fel ca în cazul scufundărilor oarecare, putem defini o versiune mai tare de substructură; și anume, una care conservă valoarea de adevăr a *tuturor* formulelor structurii mai mari.

Definiția 6.2 (Substructuri elementare)**.** O L-substructură $\mathscr{A}$ a lui $\mathscr{B}$ este substructură elementară, $\mathscr{A} \preceq \mathscr{B}$, dacă scufundarea lui A în B, $\pi\colon A \to B$ $\pi(x) = x$, este o scufundare elementară.

Cu alte cuvinte, $\mathscr{A} \preceq \mathscr{B}$ dacă, pentru orice L-formulă $\varphi(\bar{x})$ (unde $\bar{x}$ este o listă de L-variabile care include toate variabilele libere ale lui φ) și orice listă de parametri $a_1,...,a_n \in A$, e adevărat că:

$$\mathscr{A} \models \varphi(a_1,...,a_n) \text{ ddacă } \mathscr{B} \models \varphi(a_1,...,a_n)$$

Intuitiv vorbind, ce are în plus o substructură elementară față de una oarecare este că menține și adevărul formulelor cuantificate existențial. Cu alte cuvinte, dacă există o instanță a unei formule existențiale în domeniul structurii mai mari, acea instanță trebuie să aibă un corespondent care să satisfacă aceeași formulă existențială și în domeniul substructurii (e posibil ca instanța însăși să își fie propriul corespondent). Aceasta este intuiția care stă la baza testului Tarski-Vaught pentru verificarea elementarității unei substructuri.

Teorema 6.2 (Testul Tarski-Vaught)**.** *Fie $\mathscr{A}$ o L-substructură a lui $\mathscr{B}$. Fie, de asemenea, o L-formulă $\varphi(\bar{x},y)$ și parametrii $\bar{a} \in A$, $b \in B$ cu proprietatea că $\mathscr{B} \models \varphi(\bar{a},b)$. $\mathscr{A}$ e o substructură elementară a lui $\mathscr{B}$ dacă și numai dacă există un $d \in A$ așa încât $\mathscr{B} \models \varphi(\bar{a},d)$.*

Demonstrație. Prin teoremele 9.4 și 6.1, știm deja că $\mathscr{A}$ conservă toate formulele lui $\mathscr{B}$, cu excepția celor existențiale. Trebuie, așadar, doar să arătăm că $\mathscr{A}$ conservă adevărul formulelor existențiale dacă și numai dacă proprietatea din enunțul teoremei e îndeplinită. Să demonstrăm atunci ambele laturi ale acestei echivalențe.

Fie $\varphi(\bar{x},y)$, $\bar{a} \in A$, $b \in B$, cu $\mathscr{B} \models \varphi(\bar{a},b)$. Așadar, $\mathscr{B} \models \exists y\ \varphi(\bar{a},y)$. Fie $\mathscr{A}$ o substructură a lui $\mathscr{B}$.

Presupunem că există $d \in A$ așa încât $\mathscr{B} \models \varphi(\bar{a},d)$. Cum $\mathscr{A}$ e o substructură a lui $\mathscr{B}$ și $\bar{a}$, d se regăsesc în A, e adevărat că $\mathscr{A} \models \varphi(\bar{a},d)$. Deci $\mathscr{A} \models \exists y\ \varphi(\bar{a},y)$, de unde $\mathscr{A} \preceq \mathscr{B}$.

Să presupunem acum că $\mathscr{A} \preceq \mathscr{B}$. Conform definiției 6.2, dacă $\mathscr{B} \models \exists y\ \varphi(\bar{a},y)$, atunci $\mathscr{A} \models \exists y\ \varphi(\bar{a},y)$. Să reformulăm acest condițional în lumina condițiilor de adevăr ale formulelor cuantificate existențial introduse în definiția 3.3: dacă există un $b \in B$ așa încât $\mathscr{B} \models \varphi(\bar{a},b)$, atunci există un $d \in A$ așa încât $\mathscr{A} \models \varphi(\bar{a},d)$. Antecedentul e satisfăcut prin ipoteză, deci consecventul e adevărat. Iar cum $A \subseteq B$, știm că acel $d \in B$, deci d este o instanță pentru formula existențială φ și în structura $\mathscr{B}$. Formal, există un $d \in A$ așa încât $\mathscr{B} \models \varphi(\bar{a}, d)$. □

E momentul acum să demonstrăm prima variantă a teoremei Löwenheim-Skolem. În forma ei modernă, teorema ne garantează că de la orice submulțime a domeniului unei structuri putem obține domeniul unei substructuri elementare, dacă adăugăm cel mult o infinitate numărabilă de elemente noi la respectiva submulțime. Cum poate un fapt atât de tehnic să aibă orice fel de importanță filosofică vom arăta după ce vom oferi demonstrația. Înainte de asta, însă, ce este mai suprinzător e că enunțul teoremei pe care a demonstrat-o Skolem pare să aibă foarte puțin de-a face cu enunțul pe care tocmai l-am formulat. Iată formularea lui Skolem din 1920 pentru teoremă:

Teoremă (Skolem 1920)**.** Orice propoziție în forma normală fie este o contradicție, fie este satisfiabilă într-un domeniu finit sau infinit numărabil.[12]

Ce vom demonstra în continuare e un caz mai general decât cazul tratat de Skolem. Acest lucru va deveni evident odată ce vom apropia formularea din 1920 de formularea modernă, făcând câteva observații. Prima dintre ele, *i)* despre contradicții. Sensul în care Skolem folosește acest termen nu e sintactic, ci semantic. Așadar ce are în vedere când vorbește despre „o propoziție care nu e contradictorie” este o propoziție satisfiabilă, deci o propoziție care are un model oarecare. O a doua mențiune care trebuie făcută este că Skolem arată în același articol că teorema se poate generaliza pentru *conjuncții infinite de propoziții* în forma normală. Conform convențiilor noastre de limbaj (2.3), o formulă nu poate avea decât lungime finită; însă putem trata o conjuncție infinită de formule drept o mulțime infinită de formule finite, adică drept o *teorie* infinită. Apoi, *ii)* despre forme normale. Skolem, în același articol, arată că orice formulă e echivalentă semantic cu o formulă ai cărei cuantificatori se află toți la început; aceasta fiind forma ei normală. Deși *formele normale Skolem* încă mai sunt folosite (spre exemplu, pentru demonstrarea automată a formulelor), noi nu vom avea nevoie de o astfel de rescriere, fără a pierde în felul acesta din gradul de generalitate al demonstrației. Nu în ultimul rând, *iii)* despre maniera în care Skolem construiește demonstrația. Esența demonstrației constă în construirea unui domeniu nou, în care se adaugă rând pe rând câte un singur element *din domeniul în care propoziția era deja satisfiabilă*, pentru fiecare cuantificator existențial al propoziției. Domeniul care rezultă trebuie, deci, să fie inclus în domeniul original. Devine atunci evident că teorema vorbește despre structuri și substructuri care satisfac o anumită teorie:

Teoremă. Orice teorie satisfiabilă într-o structură oarecare este satisfiabilă și într-o substructură cu domeniu finit sau infinit numărabil.

[12] Skolem [1920] 1981a, p. 256.

Această din urmă variantă a teoremei va reieși drept corolar al teoremei moderne. Varianta modernă mai garantează două lucruri în plus. Primul este că substructura va fi *elementară* – adică nu doar enunțurile, ci și propozițiile deschise cu parametri din substructură vor fi satisfăcute în ambele structuri. A doua este mai importantă pentru cititorul cu aplecare filosofică: că *oricare* elemente ale structurii originale pot fi prezente în substructură, dacă cel care aplică teorema își dorește dinadins să rețină acele elemente. Atunci, evident, dacă mulțimea elementelor care trebuie păstrate cu orice preț este infinit *nenumărabilă*, substructura nu va mai avea cum să aibă domeniu finit sau infinit numărabil. Însă un lucru va fi în continuare garantat: *cel mult* o mulțime finită sau infinit numărabilă de elemente va mai trebui adăugată *pe lângă* elementele pe care substructura e forțată să le păstreze. Intuiția este că vom „înveli” orice submulțime a unei structuri cu elemente noi din structură până când vom obține o substructură elementară. Nu întâmplător procedeul e cunoscut ca *învelitoarea Skolem*. Pe faptul că putem înveli orice submulțime a unei structuri se reazemă Putnam în așa-numitul argument al constructivizării.

Putem acum să formulăm și să demonstrăm cea mai generală formă a teoremei Löwenheim-Skolem:

Teorema 6.3 (Teorema Löwenheim-Skolem (în jos))**.** Pentru orice L-structură $\mathscr{B}$ și orice mulțime $S \subseteq B$, există o L-structură $\mathscr{A}$, cu $S \subseteq A \subseteq B$, în așa fel încât $\mathscr{A} \preceq \mathscr{B}$ și $|A| \leq max(|S|, \aleph_0)$.

Demonstrație. Conform testului Tarski-Vaught (6.2), dacă $A \subseteq B$, și orice formulă existențială satisfăcută de $\mathscr{B}$ cu parametri din A (ie. $\mathscr{B} \models \exists y\, \varphi(\bar{a}, y)$, cu $\bar{a} \in A$) poate fi instanțiată de un element din A (ie. $\mathscr{B} \models \varphi(\bar{a}, d)$, cu $\bar{a}$ și $d \in A$), atunci $\mathscr{A} \preceq B$.

Fie $S \subseteq B$. Vom construi mulțimea A așa încât $S \subseteq A$ și condiția cerută de testul Tarski-Vaught să fie respectată. Pentru a face asta, vom avea nevoie să construim un șir de mulțimi intermediare S_0, S_1, S_2, ..., cu $S_i \subseteq S_{i+1}$. A va fi reuniunea tuturor acestor mulțimi.

Începem cu $S_0 = S$. Vrem ca S_{i+1} să conțină toate elementele de care e nevoie pentru ca formulele existențiale satisfăcute de $\mathscr{B}$ cu parametri din S_i să aibă o instanță în S_{i+1}. Pentru fiecare L-formulă φ cu n variabile libere, atunci, să definim următoarea funcție $g_\varphi : B^n \to B$:

$$g_\varphi(\bar{a}) = \begin{cases} \text{un } d \in B \text{ pentru care } \mathscr{B} \models \varphi(\bar{a}, d) & \text{dacă } \mathscr{B} \models \exists y\ \varphi(\bar{a}, y) \\ \text{orice } d \in B & \text{altfel} \end{cases} \tag{1}$$

Funcția $g_\varphi(\bar{a})$ alege o singură instanță din B care garantează adevărul formulei $\exists y \varphi(\bar{a}, y)$, dacă ea e adevărată deja în structura $\mathscr{B}$. Această funcție e cunoscută drept *funcția Skolem*. Avem nevoie de axioma alegerii în metateorie ca să ne asigurăm că g_φ e bine definită (vezi 9.1); în practica modernă e însă complet neproblematic să acceptăm axioma alegerii când facem metalogică fără justificări suplimentare.

Echipați cu o familie de funcții Skolem (câte o funcție pentru fiecare formulă), putem acum să definim S_{i+1}:

$$S_{i+1} = S_i \cup \{g_\varphi(\bar{a}) \mid \varphi \in Form(L) \text{ și } \bar{a} \in S_i^n\} \tag{2}$$

În fine, A va fi reuniunea tuturor mulțimilor intermediare:

$$A = \bigcup_{i \in \mathbb{N}} S_i \tag{3}$$

Evident $S \subseteq A$. Similar, că $A \subseteq B$ este evident din faptul că $S \subseteq B$ și codomeniul funcțiilor Skolem este B. Să arătăm acum că $\mathscr{A}$ satisface condiția cerută de testul Tarski-Vaught. Să presupunem că $\mathscr{B} \models \exists y\ \varphi(\bar{a}, y)$, pentru $\bar{a} \in A$. Dacă $\bar{a} \in A$, înseamnă că există o mulțime intermediară S_i în care toate elementele $a_1, a_2, \dots, a_n$ au apărut laolaltă pentru prima oară în șirul de mulțimi intermediare (adică cel puțin unul dintre $a_1, \dots, a_n \notin S_{i-1}$, dar toate $a_1, \dots, a_n \in S_i$). Atunci, prin definiția funcției g_φ și prin construcția șirului $(S_i)_{i \in \mathbb{N}}$, există o instanță $d = g_\varphi(\bar{a})$ în S_{i+1}. Cum $S_{i+1} \subseteq A$, înseamnă că $d \in A$. Atunci, conform Tarski-Vaught, $\mathscr{A} \preceq \mathscr{B}$.

Mai avem un singur pas: să stabilim mărimea lui A. Vom face un mic abuz de limbaj și vom vorbi despre „aritatea” unei formule deschise (funcția $ar\colon Form(L) \to \mathbb{N}$), înțelegând prin asta numărul ei de variabile libere. Atunci, conform ecuației 2:

$$|S_{i+1}| \leq |S_i| + |\bigcup_{\varphi \in Form(L)} S_i^{ar(\varphi)}|, \tag{4}$$

pentru că S_{i+1} adaugă la S_i *cel mult* un element pentru fiecare formulă φ și fiecare listă de $ar(\varphi)$ parametri. Or $Form(L)$ este numărabilă (vezi 9.2 în anexă). Atunci, dacă S_i este finită sau numărabilă, atunci $\bigcup_{\varphi \in Form(L)} S_i^{ar(\varphi)}$ este numărabilă (conform 9.2 din anexă), deci $|S_{i+1}| \leq \aleph_0$. Iar dacă S_i este infinit nenumărabilă, reuniunea după $Form(L)$ va avea cardinalitatea lui S_i, deci $|S_{i+1}| = |S_i|$. Ambele cazuri pot fi exprimate în mod comprimat prin inegalitatea $|S_{i+1}| \leq max(|S_i|, \aleph_0)$. Deci $|S_i| \leq max(|S|, \aleph_0)$. Conform ecuației 3, atunci, $|A| \leq max(|S|, \aleph_0)$ □

Corolar. Dacă $\mathscr{B}$ este o structură cu domeniu de cardinalitate infinit nenumărabilă și alegem $S = \varnothing$, există o substructură $\mathscr{A}$ a lui $\mathscr{B}$ cu $|A| \leq \aleph_0$. Dacă știm, în plus, că teoria lui $\mathscr{B}$ nu poate fi modelată într-un domeniu finit, atunci $|A| = \aleph_0$.

Într-un articol din 1922, după ce arată că teorema se poate demonstra și fără a face apel la axioma alegerii, Skolem discută implicațiile pe care ea le are asupra metodei axiomatice în matematică. Să mergem pe urmele argumentului lui, remarcând pentru început că axiomele ZFC sunt o teorie de tipul celei menționate în corolar, adică una care nu poate fi modelată într-un domeniu finit.[13] Cu acest lucru în minte, Skolem susține că modelul lor uzual $(V, \in)$ poate fi trecut prin teorema Löwenheim-Skolem, rezultatul fiind o structură numărabilă pentru teoria mulțimilor. Trebuie, totuși, să fim ceva mai atenți decât Skolem la presupunerile metalingvistice cu care lucrăm. Până acum demonstrațiile din secțiunea aceasta nu au necesitat o metateorie mai expresivă decât teoria mulțimilor. Dar dacă am lua tot teoria mulțimilor ca metateorie și am aplica teorema Löwenheim-Skolem asupra universului V, ne-am afla în plină contradicție cu teorema lui Tarski (4.1) – căci obținând o structură obținem un predicat de adevăr. Asta nu putea fi ceva care să-l preocupe pe Skolem în 1922, fiind nevoie de încă un deceniu până când Tarski să își publice rezultatele. Există o cale de rezolvare – deși e îndoielnic că Skolem ar fi acceptat-o – și anume să acceptăm în metateorie existența unui număr cardinal foarte mare κ numit *cardinal inaccesibil*. Sub această presupunere se poate arăta că $(V_\kappa, \in)$ este un model pentru axiomele ZFC, și putem aplica fără grijă teorema Löwenheim-Skolem asupra domeniului V_κ. Sigur, pentru Skolem, existența oricărei cardinalități infinite era problematică. Dar dacă vom vrea să ducem argumentul prin substructuri numărabile până la concluziile pe care Skolem i le intenționează, nu avem de ales decât să respingem acele premise care împiedică formularea lui coerentă.

Obținem, deci, o substructură elementară numărabilă a modelului uzual pentru axiomele ZFC. Care sunt acele concluzii pe care Skolem le intenționează? Să numim domeniul acestei substructuri B. Iată atunci un aparent paradox remarcat de Skolem: „Cu ajutorul axiomelor putem să demonstrăm existența cardinalităților superioare, a claselor de numere transfinite, și așa mai departe. Cum se poate, atunci, ca întregul domeniu B să fie deja enumerabil prin intermediul numerelor naturale?"[14] Reformulând: din axiomele teoriei mulțimilor se deduce ca teoremă că unele mulțimi nu pot fi puse în corespondență unu la unu cu numerele naturale, adică nu sunt numărabile: aceasta este teorema lui Cantor. Un exemplu de astfel de mulțime ar fi mulțimea numerelor reale $\mathbb{R}$. Noi dispunem acum de o structură cu domeniu B pentru axiomele teoriei mulțimilor; ceea ce înseamnă că structura noastră trebuie să satisfacă propoziția conform căreia $\mathbb{R}$ nu e o mulțime numărabilă. Însă B este e o mulțime numărabilă ea însăși, pentru că am obținut-o prin teorema Löwenheim-Skolem. *Deci un domeniu numărabil spune despre unul din elementele sale că este nenumărabil!*

Dar este ceva cu adevărat surprinzător aici? Nimic nu împiedică o mulțime numărabilă, sau chiar finită, să aibă ca element o mulțime nenumărabilă. Fie, așadar, o mulțime M din B, despre care structura lui B spune că este nenumărabilă. M poate să conțină o mulțime nenumărabilă de

[13]Din axioma infinitului și axioma specificației putem demonstra existența mulțimii vide $\varnothing$. Axioma perechii garantează că dacă există o mulțime x, atunci există o mulțime $\{x\}$, diferită de x. Putem deduce din axiome, deci, existența unui număr infinit de mulțimi: $\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\{\{\varnothing\}\}\}\dots$.

[14]„By virtue of the axioms we can prove the existence of higher cardinalities, of higher number classes, and so forth. How can it be, then, that the entire domain B can already be enumerated by means of the finite positive integers?" Skolem [1922] 1981b, p. 295.

elemente, fără ca fiecare element al lui M să fie și element al lui B. Deci M poate aparține de B și poate fi nenumărabilă, în timp ce B însăși e numărabilă. Tocmai de aceea, Paul Benacerraf concluzionează că Skolem mai adăuga în mod tacit o constrângere legată de natura lui B.[15] Mai precis, Skolem ar fi presupus că dacă M e în B, atunci orice x care e în M trebuie să fie și în B. Termenul tehnic pentru o mulțime B care are proprietatea presupusă de Skolem este acela de *mulțime tranzitivă*. Dar simpla aplicare a teoremei Löwenheim-Skolem nu ne garantează existența unui model cu domeniu *tranzitiv*. Din fericire, există o lemă datorată lui Mostowski care arată că orice model al teoriei mulțimilor este izomorf cu un model cu domeniu tranzitiv (vezi 9.1 în anexă pentru enunțul formal și demonstrație). Deci putem să aplicăm lema lui Mostowski asupra substructurii numărabile cu domeniu B obținute prin teorema Löwenheim-Skolem, obținând o structură nouă cu domeniu B' numărabil *și tranzitiv*, izomorfă cu structura inițială. Prin izomorfism, teorema lui Cantor e satisfăcută și în noul domeniu B'. Iar acum ne apropiem mai mult de ceva care are aerul unei contradicții: B' spune despre unul din elementele sale M' că este nenumărabil, în timp ce B' e tranzitivă. Or, dacă M' este nenumărabilă, atunci prin tranzitivitate B' ar trebui să fie și ea nenumărabilă. *Dar știm că B' e numărabilă!* mai precis Versiunea aceasta tranzitivă pare un mod mai precis mod de a caracteriza ceea ce a rămas cunoscut drept paradoxul al lui Skolem; în mod cert, pare să coincidă cu ce a intenționat Skolem să evidențieze în 1922.[16] Acum, dacă paradoxul lui Skolem reprezintă cu adevărat un paradox, atunci în orice caz nu e un paradox *matematic*. Pentru că, matematic, aparenta contradicție se disipă imediat ce ne dăm seama că felul în care vorbim despre numărabilitate în modelul uzual $(V, \in)$, sau chiar în metateorie, diferă de felul în care B' interpretează numărabilitatea. Că o mulțime e numărabilă e o propoziție care exprimă *existența* unei anumite funcții; o funcție bijectivă între respectiva mulțime și mulțimea numerelor naturale. Deci a spune despre o mulțime că nu e numărabilă înseamnă a spune că *nu există* o bijecție cu proprietatea cerută. Până în punctul acesta, V și B' ar trebui să fie de acord. Însă ele nu mai sunt de acord asupra interpretării cuantificatorului existențial. Cuantificatorul interpretat la B' are acces *doar la domeniul B'*, deci a spune în B' că M' e nenumărabilă înseamnă a spune că *nu există în B'* respectiva bijecție. Pe când, atunci când vorbim despre existență în V, avem acces la întregul univers al mulțimilor: privind M' dinafara submodelului, avem la dispoziție o ontologie suficient de largă încât să oferim bijecția necesară. Însă dinăuntru, bijecția poate pur și simplu să nu ne fie

[15] Benacerraf 1985, p. 102.

[16] Trebuie menționat totuși că modelul obținut prin aplicarea lemei lui Mostowski nu va mai fi obligatoriu un submodel al lui V_κ, deși $B' \subset V_\kappa$, pentru că funcția de scufundare a lui B' în V_κ poate să nu fie identitatea. Dar ce e garantat e că scufundarea *este* o scufundare elementară. Button 2011 demonstrează existența unui submodel tranzitiv numărabil al lui V_κ – dar în cazul acela scufundarea nu va mai fi elementară, tocmai pentru că va oferi altă valoare de adevăr formulelor care atestă nenumărabilitatea unor mulțimi. Pentru că nicio demonstrație nu poate asigura și numărabilitatea domeniului, și tranzitivitatea lui, și elementaritatea scufundării; și pentru că argumentul lui Button are nevoie de mai multă teorie decât ar fi adecvat scopului acestei lucrări; ne vom mulțumi cu ce am demonstrat până acum.

accesibilă – ceea ce transformă M', din perspectiva internă a lui B', într-o mulțime nenumărabilă.

Nu e nimic controversabil din punct de vedere formal la această rezolvare, pe care o oferă chiar Skolem. Teorema Löwenheim-Skolem nu scoate în evidență vreo contradicție ascunsă a teoriei mulțimilor, așa cum au făcut-o paradoxuri autentice precum cel al lui Russell. Dar începând cu acest moment, în care contrastăm interpretarea internă a unui model cu interpretarea externă a metateoriei, începe să se contureze o dilemă reală *în filosofia matematicii*. Cum ar trebui să ne raportăm la perspectiva internă a lui B'? Suntem la fel de îndreptățiți să îi oferim lui B' statutul de univers care interpretează teoria mulțimilor, pe cât suntem să îi oferim și universului clasic von Neumann același statut? Să ne amintim despre teoria relațională a obiectelor, pe care o apără Quine. E suficient ca un domeniu să asigure structura relațională de care are nevoie o teorie, pentru a ne convinge că el reprezintă o ontologie posibilă pentru respectiva teorie?[17] Considerațiile din secțiunile anterioare ne-ar putea tenta să răspundem pozitiv la această întrebare. Dar iată ce consecință ar avea răspunsul pozitiv. În cuvintele lui Skolem, „axiomatizarea teoriei mulțimilor duce la o relativitate a noțiunilor ei, iar această relativitate este legată în mod inseparabil de orice axiomatizare completă”.[18] Ce dovedește paradoxul lui Skolem este că axiomele nu sunt suficient de puternice încât să prevină noțiuni precum numărabilitatea să fie relative la model. Iar relativitatea acestor noțiuni face ca una și aceeași mulțime să fie privită diferit, în funcție de modelul de care ea aparține – M' e numărabilă în V, dar nenumărabilă în B'. Asta în vreme ce modelele sunt echivalente elementar!

Implicațiile acestui fapt sunt în mare parte dependente de pozițiile filosofice pe care te plasezi de la bun început. Așa cum comentează și Benacerraf, „ce pentru unul e modus tollens, pentru altul e modus ponens”.[19] Dacă ținem la premisa că axiomele sunt singurele în măsura de a impune constrângeri asupra modelelor, atunci vom lua relativitatea despre care vorbește Skolem ca o consecință inevitabilă a felului în care funcționează matematica și științele. Vom fi atunci înclinați să spunem că nicio mulțime nu e nenumărabilă decât într-un sens relativ – căci oricând o metateorie mai largă ar putea să enumere chiar și elementele mulțimilor pe care noi le considerăm nenumărabile. Singura cardinalitate infinită absolută între toate modelele unei teorii ar fi cea numărabilă; pentru că dacă bijecția între o mulțime și $\mathbb{N}$ există într-un model, atunci ea există în toate supermodelele lui. Iar atunci am putea să ne mulțumim prin a reduce *orice* ontologie la o ontolo-

[17] Quine personal nu acorda teoremei Löwenheim-Skolem prea multă semnificație ontologică, pentru că nu admitea decât ca transferul unei ontologii în alta să fie realizat printr-o funcție proxy, iar funcțiile proxy să fie definibile explicit. Or funcțiile Skolem de care e nevoie în demonstrația teoremei L-S nu specifică explicit care element al structurii mari e transpus în care element al substructurii; demonstrația ne spune doar că *se poate* face o asemenea transpunere. Vezi Quine 1964. Dar putem, cel puțin de dragul argumentului, lăsa deoparte aceste scrupule formale și putem admite că teorema chiar transpune un domeniu mai mare într-un domeniu mai mic; pentru a vedea, atunci, ce consecințe filosofice ar avea acest lucru.

[18] Skolem [1922] 1981b, p. 296.

[19] „One person’s modus tollens is another’s modus ponens." Benacerraf 1965.

gie numărabilă, din moment ce despre nicio cardinalitate nenumărabilă nu putem ști ceva în mod absolut. Concluziile la care ajungem astfel sunt în general văzute drept concluzii *sceptice*; pentru că ne condamnă la un punct de vedere relativ, deși știm că întotdeauna vor exista puncte de vedere mai largi (eventual inaccesibile) în care unele din credințele noastre (imposibil de specificat care) vor fi contestate. Deci a duce poziția structuralistă până la concluziile ei ultime pare să anunțe vești proaste pentru însăși capacitatea noastră de a dobândi credințe adevărate despre lume.

Dar consecințele acestea sunt greu de acceptat. Cu atât mai mult cu cât paradoxul lui Skolem, despărțit de premisa lui structuralistă radicală, pare să sugereze cu totul altceva decât un scepticism generalizat. Dimpotrivă, ar putea să fie luat ca demonstrația faptului că *nu doar axiomele* determină structura ontologică a unei teorii, tocmai *pentru că* unele concepte rămân relative în orice axiomatizare. Deci paradoxul ar dovedi că mai este *încă ceva*, pe lângă axiome, cu ajutorul căruia putem deosebi interpretarea intenționată a formulelor de toate celelalte posibile. Chiar și pentru Benacerraf, care am văzut deja că susține că numerele naturale sunt doar poziții în structura aritmetică, acea structură e determinată de mai mult decât doar de axiome. Numai prin intuiție, sugerează Benacerraf, putem să ne dăm seama care a fost *intenția* cu care a fost formulat un anumit sistem axiomatic. „Mulțimea teoriilor axiomatizată își are rădăcinile în matematica informală, și pentru a-și păstra sensul în urma axiomatizării și formalizării e obligată să păstreze aceste puncte de legătură [*cu disciplina informală*]”.[20] Privit astfel, Skolem arată doar că există modele care nu corespund intuiției, modele *neintenționate*. Pe de-o parte, deci, teorema Löwenheim-Skolem dovedește că din matematica axiomatizată decurge o anumită relativitate; pe de altă parte, însăși acea relativitate funcționează ca un *reductio* împotriva identificării teoriilor matematice cu axiomele lor.

Mai există o versiune a teoremei Löwenheim-Skolem, care arată că dacă o teorie are un model infinit $\mathscr{A}$, atunci are un model de orice cardinalitate infinită mai mare decât a lui $\mathscr{A}$. Așadar, nu suntem obligați doar să restrângem domeniul unui model „în jos” după dimensiunea lui; putem și să îl extindem până la orice cardinalitate infinită. În mod nesurprinzător, și versiunea aceasta poate fi interpretată în moduri radical diferite în funcție de preferințele filosofice ale celui care o comentează. Interpretarea sceptică am prefigurat-o deja; ei văd în versiunea „în sus” a teoremei dovada faptului că o structură cu domeniu mai larg ar putea să enumere și elementele mulțimilor pe care noi le vedem a fi nenumărabile. Latura cealaltă se folosește de noua teoremă pentru a-i dovedi scepticului că a pune pe un piedestal infinitatea *numărabilă* este nejustificat– tot atât de bine și-ar putea fixa atenția și pe $\aleph_{17}$ drept unic cardinal absolut, și ar putea susține că oricare altă cardinalitate e relativă.

Să demonstrăm, deci, această variantă a teoremei Löwenheim-Skolem. Vom avea nevoie, pen-

[20]„Axiomatized set theory took its roots in informal mathematics and in order to retain its sense through axiomatization and formalization it must retain these points of contact.” Benacerraf 1985, p. 109.

tru început, să introducem câteva noțiuni formale noi. Primele se vor referi la teorema de compacitate, care reprezintă unul din cele mai importante rezultate din teoria modelelor și pe care doar o vom enunța.

Definiția 6.3 (Satisfiabilitate finită)**.** Spunem despre o teorie T că este finit satisfiabilă dacă orice submulțime finită a lui T e satisfiabilă.

Teorema 6.4 (Teorema de compacitate)**.** O teorie este satisfiabilă dacă și numai dacă este finit satisfiabilă.

Va fi util să discutăm despre două moduri în care putem modifica limbajul unei structuri (semnătură ei), fără a-i afecta în vreun fel domeniul sau interpretările limbajului rămas nemodificat. Mai precis, putem fie elimina, fie adăuga simboluri în semnătură; prima variantă se numește o restrângere a semnăturii, cealaltă o extindere:

Definiția 6.4 (Restrângeri și extinderi)**.** Fie L și L' două semnături, L fiind obținută prin eliminarea unor simboluri din L'. Fie $\mathscr{A}$ o $L'-$structură.

a. *Restrângerea* lui $\mathscr{A}$ la L, $\mathscr{A}|L$, este L-structura cu domeniu A care atribuie aceleași interpretări simbolurilor din L ca L'-structura $\mathscr{A}$.

b. Dacă $\mathscr{B}$ este restrângerea lui $\mathscr{A}$ la L, $\mathscr{A}$ se numește o *extindere* a lui $\mathscr{B}$ la L'.

O proprietate utilă a extinderilor, și motivul pentru care avem nevoie de ele ca să demonstrăm teorema Löwenheim-Skolem, este că dacă adăugăm câte un simbol nou pentru fiecare obiect din domeniul unei structuri, pe care îl interpretăm ca obiectul însuși, atunci mulțimea *enunțurilor* adevărate în structura cu limbajul extins coincide într-un anumit sens cu mulțimea formulelor adevărate din structura inițială. Acest lucru îl vom arăta în următoarea definiție și teoremă.

Definiția 6.5 (Diagonala elementară)**.** Fie $\mathscr{A}$ o L-structură, L' o semnătură care conține toate simbolurile lui L și câte o nouă constantă c_a pentru fiecare $a \in A$. Fie $\mathscr{A}'$ extinderea lui $\mathscr{A}$ la L', cu $c_a^{\mathscr{A}'} = a$ pentru orice $a \in A$.

Diagonala elementară a lui $\mathscr{A}$ este mulțimea $eldig(\mathscr{A}) = \{\varphi \mid \varphi$ e un L'-enunț și $\mathscr{A}' \models \varphi\})$.

Remarcă. O L'-formulă $\varphi(\bar{c_a})$ este în $eldiag(\mathscr{A})$ dacă și numai dacă există o L-formulă $\varphi(\bar{x})$ în așa fel încât $\mathscr{A} \models \varphi(\bar{a})$.

Teorema 6.5. Dacă $\mathscr{A}$ e o $L-$structură, $\mathscr{B}$ e o $L'-$structură ($L \subseteq L'$) și $\mathscr{B} \models eldiag(\mathscr{A})$, atunci există o scufundare elementară a lui $\mathscr{A}$ în $\mathscr{B}|L$.

Demonstrație. Dacă $\mathscr{B} \models \varphi(\bar{c_a})$, pentru $\varphi(\bar{c_a}) \in L'$, atunci $\mathscr{B}|L \models \varphi(\bar{c_a^{\mathscr{B}}})$ pentru $\varphi(\bar{x}) \in L$. Conform remarcii de la definiția anterioară, $\mathscr{A} \models \varphi(\bar{a})$. Așadar, funcția $\pi \colon A \to B$ $\pi(a) = c_a^{\mathscr{B}}$ este o scufundare elementară a lui $\mathscr{A}$ în $\mathscr{B}|L$. $\square$

Acum avem la dispoziție mijloacele de a demonstra varianta promisă a teoremei Löwenheim-Skolem.

Teorema 6.6 (Teorema Löwenheim-Skolem (în sus))**.** Pentru orice L-structură infinită $\mathscr{A}$ de cardinalitate κ și pentru orice $\lambda > \kappa$, există o L-structură $\mathscr{B}$ cu $|B| = \lambda$ și $\mathscr{A} \preceq \mathscr{B}$.

Demonstrație. Fie $\mathscr{A}$ o L-structură, cu $|A| = \kappa$ și λ un număr cardinal mai mare decât κ. Vom construi un model de cardinalitate λ pentru $eldiag(\mathscr{A})$.

Vom defini un nou limbaj L' care, pe lângă toate simbolurile lui L, va include două noi mulțimi de constante C_1 și C_2 ($|C_1| = \kappa$, $|C_2| = \lambda$), în așa fel încât fiecărui element $a \in A$ să îi corespundă o constantă $c_a \in C_1$. Să considerăm acum următoarea mulțime de L'-formule:

$$\Gamma = eldiag(\mathscr{A}) \cup \{c_1 \neq c_2 \mid c_1,\, c_2 \in C_2 \text{ și } c_1 \neq c_2\}$$

Vom arăta că Γ este finit satisfiabilă.

Fie, deci, Δ o submulțime finită a lui Γ. Cum Δ e finită, înseamnă că doar un număr finit de formule de tipul $\{c_1 \neq c_2 \mid c_1,\, c_2 \in C_2 \text{ și } c_1 \neq c_2\}$ sunt prezente în Δ, deci doar un număr de n constante din C_2 sunt folosite. Respectivele constante pot fi enumerate ca $c_1, \ldots, c_n$. Fie $a_1, \ldots, a_n$ o enumerare oarecare de n elemente distincte din A. Luăm $\mathscr{A}'$ ca $L'-$extinderea lui $\mathscr{A}$ în care $c_a^{\mathscr{A}'} = a$, pentru $c_a \in C_1$; și $c_i^{\mathscr{A}'} = a_i$, pentru i de la 1 la n. Din interpretarea simbolurilor din C_1 reiese că $\mathscr{A}' \models eldiag(\mathscr{A})$. Apoi, cum $a_1, \ldots, a_n$ sunt toate distincte, $\mathscr{A}' \models c_i \neq c_j$ pentru $1 \leq i \neq j \leq n$. Deci $\mathscr{A}' \models \Delta$.

Prin teorema de compacitate, atunci, Γ are un model. Să îl numim $\mathscr{B}'$. Mai știm ceva despre $\mathscr{B}'$ în afară de faptul că e model pentru Γ? Da: știm că $|B'| \geq \lambda$, pentru că $\mathscr{B}' \models c_1 \neq c_2$ oricare ar fi c_1, c_2 distincte din C_2, și $|C_2| = \lambda$.

Acum, conform 6.5, există o scufundare elementară π a lui $\mathscr{A}$ în $\mathscr{B}'|L$. Însă $\mathscr{A}$ nu e o substructură a lui $\mathscr{B}'|L$, și B' nu are cardinalitatea exact λ. Putem remedia însă cu ușurință ambele probleme. Să găsim întâi o substructură a lui $\mathscr{A}$ care este izomorfă cu $\mathscr{B}'|L$. Putem să transpunem toate elementele din B' care sunt în imaginea lui π în corespondentele lor din domeniul lui π; iar toate celelalte elemente din B' pot fi transpuse în cate un element dintr-o mulțime oarecare disjunctă de A. E ușor de văzut că o astfel de funcție de transpunere e un izomorfism de structuri, și că A este inclusă în imaginea ei. Fie $\mathscr{B}$ structura care rezultă prin aplicarea izomorfismului asupra lui $\mathscr{B}'$, urmată de Löwenheim-Skolem în jos pentru a o aduce la cardinalitate λ. Atunci, $\mathscr{A} \preceq \mathscr{B}$ și $|B| = \lambda$. $\square$

Iată acum ce arată împreună cele două versiuni ale teoremei Löwenheim-Skolem: dacă o teorie este satisfiabilă într-un model infinit, atunci e satisfiabilă într-un model de *orice* cardinalitate infinită. Observația aceasta stă la baza argumentului lui Putnam împotriva realismului. Acum e timpul să vedem în ce constă el.

7 Argumentul de teoria modelelor

„Nu susțin că paradoxul lui Skolem e o antinomie în logica formală. Dar voi argumenta că *este* o antinomie, sau ceva asemănător, în filosofia limbajului”,[21] spune Putnam în deschiderea articolului său din 1980, *Models and Reality*. De ce paradoxul lui Skolem nu e o antinomie în logica formală am văzut deja: aparența de paradox dispare odată ce ne dăm seama că modele diferite interpretează bijectivitatea diferit. Ce rămâne însă este dezbaterea despre modele intenționate: din toate modelele posibile pentru teoria mulțimilor, avem vreun motiv să le preferăm pe cele cu domeniu nenumărabil în locul celor cu domeniu numărabil? În termenii filosofiei limbajului, întrebarea aceasta se poate reformula astfel: dispunem de mijloace de *a fixa referința* vocabularului teoriei mulțimilor? A răspunde „da” la această întrebare te-ar plasa într-o poziție realistă în legătură cu teoria mulțimilor. Dar există *diferite feluri* în care poți răspunde „da”, și diferite nuanțe ale realismului. Căci ai putea fi exponentul unui realism dur, platonic: putem fixa referința termenilor pentru că avem acces imediat la o realitate abstractă a mulțimilor în sine, care nu se intersectează cu realitatea sensibilă. Asta e ce Putnam ar numi o „teorie magică a referinței”. Dar poți oferi un răspuns pozitiv și fără să te angajezi la o teorie magică a referinței, explicând felul în care referința e posibilă în termeni compatibili cu știința empirică. Benacerraf are, probabil, o astfel de variantă în minte când recurge la intuiție pentru a argumenta care e cardinalitatea intenționată a modelelor teoriei mulțimilor; căci intuiția poate fi un obiect de studiu al psihologiei, și psihologia ar putea explica ce fel de intuiții sunt la mijloc atunci când facem raționamente matematice. Acesta ar fi un exemplu de realism *moderat*, care se angajează la existența independentă a referinței termenilor săi fără a depăși limitele unei orientări filosofice naturaliste; adică fără a postula capacități mintale supranaturale de a intra în legătură cu realitatea platonică. Ce își propune Putnam să demonstreze, însă, este că „din păcate, poziția realistă moderată e cea pusă în primejdie de teorema Löwenheim-Skolem”.[22] Iar cum realismul *dur* iese din discuție de la bun început, Putnam va alege să ocolească în întregime întrebarea despre fixarea referinței și să recurgă la o formă de antirealism: enunțurile noastre nu sunt adevărate în virtutea corespondenței lor cu o realitate anume, ci sunt adevărate *pentru că spunem că sunt adevărate*. Soluția nu ar trebui să ne pară artificială; tot ce încearcă Putnam să facă este să echivaleze între *adevărul epistemic* – cel în virtutea căruia ajungem *noi* să spunem că propozițiile sunt adevărate, în urma aplicării unei proceduri empirice de investigare a

[21] „It is not my claim that the Lowenheim-Skolem paradox is an antinomy in formal logic. But I shall argue that it *is* an antinomy, or something close to it, in philosophy of language.” Putnam 1980, p. 464.

[22] „I shall argue that it is, unfortunately, the moderate realist position which is put into deep trouble by the Löwenheim-Skolem Theorem.” ibid., p. 464.

naturii – și *adevărul simpliciter, absolut*. Dacă Putnam are succes cu argumentul lui, ar trebui să ne convingă că nu există nicio instanță mai presus de propria rațiune care să determine ce e adevărat și ce e fals. Referința nu mai e doar nedeterminată, ea pur și simplu încetează să mai fie o problemă. Observați că nu mai vorbim acum doar despre teoria mulțimilor sau aritmetică – vorbim despre întreaga teorie asupra lumii, în ansamblul ei.

Cum poate Putnam să justifice așa o teză? Cum are de gând să respingă realismul moderat nu doar în ceea ce privește teoria mulțimilor, ci în ceea ce privește *absolut tot*? Strategia lui este să reproducă un argument similar cu al lui Skolem, pentru a reduce la absurd poziția realistă moderată. Acest argument depinde, printre altele, de următoarea premisă: realismul, în orice formă a lui, susține că *adevărul e neepistemic*. Cu alte cuvinte, că până și cea mai bună teorie la care putem spera să ajungem folosindu-ne de capacitățile noastre raționale – teoria *ideală*, care respectă orice constrângeri teoretice și operaționale am vrea să îi impunem – poate fi falsă *în realitate*. „Falsă”, în sensul că *nu corespunde* stării de fapt din realitate. Dar, protestează Putnam folosindu-se de teorema Löwenheim-Skolem, un realist e obligat să accepte că *orice teorie ideală poate fi pusă în corespondență cu realitatea*. Deci orice teorie ideală e adevărată *de fapt*, nu doar epistemic – iar realismul devine autocontradictoriu. Iată argumentul așa cum l-a formulat Putnam:

> Fie T_I o teorie ideală, din punctul nostru de vedere. Ridicând restricțiile asupra puterilor noastre finite, ne putem imagina că T_I are toate proprietățile pe care ni le dorim, cu excepția *adevărului obiectiv* – care rămâne nestabilit. De exemplu, T_I poate fi imaginată ca fiind completă, consistentă, ca prezicând corect toate propozițiile de observație [...], ca îndeplinind orice „constrângeri operaționale” există, ca fiind „frumoasă”, „simplă”, „plauzibilă” etc.. Presupunerea [*realismului*] este că T_I poate fi toate acestea, dar *tot poate fi* (în realitate) *falsă*.[23]

Putnam va exploata *consistența* teoriei T_I. Teorema de completitudine a logicii de ordinul I garantează că orice teorie consistentă, adică necontradictorie, are un model. Cum T_I este o teorie *ideală*, ea trebuie să fie și necontradictorie, deci să aibă un model. Mai departe, dacă T_I are un model *infinit*, atunci din cele două teoreme Löwenheim-Skolem reiese că T_I are modele *de orice cardinalitate infinită*. Această proprietate a lui T_I este inima argumentului lui Putnam. Căci continuă:

> Presupun că *LUMEA* are (sau poate fi descompusă într-o) o infinitate de elemente. Mai presupun și că T_I atestă existența unei infinități de obiecte [...]. Acum, T_I e consistentă (prin ipoteză) și are modele infinite. Prin [teorema Löwenheim-Skolem], T_I are un model de orice cardinalitate infinită. Alege un model M de aceeași cardinalitate cu a *LUMII*. Asociază fiecărui element al lui M câte un singur element al *LUMII*,

[23] Putnam 1977, p. 485.

> și folosește funcția rezultată pentru a defini relațiile lui M direct în *LUME*. Rezultatul este o relație de satisfacție SAT – o „corespondență” între termenii teoriei și elementele *LUMII* – în virtutea căreia teoria T_I devine *adevărată*, cu condiția să interpretăm „adevărat” ca adevărat-conform-SAT. Deci ce rămâne din presupunerea că până și teoria ideală T_I ar putea de fapt fi falsă?[24]

După consistența lui T_I, următoarea cea mai importantă parte a argumentului lui Putnam se regăsește în aceste cuvinte: „cu condiția să interpretăm «adevărat» ca adevărat-conform-SAT”. De ce am face așa ceva? Poate că natura tehnică a expunerii face mai puțin evident cât de bizară e această condiție. Voi prelua un exemplu al lui Catherine Elgin ca să arăt ce ne invită Putnam, de fapt, să facem. Să luăm, de pildă, teoria filosofică a lui Thales: „Totul e apă”.[25] Evident, teoria aceasta e consistentă, pentru că ea constă într-o singură formulă; și, evident, e satisfiabilă în modele infinite. Deci putem interpreta termenul „apă” în așa fel încât *toate elementele lumii să facă parte din extensiunea lui*. Interpretarea aceasta ar stabili o corespondență a teoriei lui Thales cu *lumea*, și ar fi totodată o relație de satisfacție, căci teoria lui Thales ar fi adevărată așa înțelegând cuvântul „apă”. *Deci teoria lui Thales e adevărată* – cu condiția să interpretăm „adevărat” ca adevărat-sub-SAT! Repet acum întrebarea: *de ce am face așa ceva?* Nu ar trebui să fie deloc surprinzător că dacă suntem liberi să atribuim *orice referință dorim* cuvintelor, atunci putem transforma orice teorie într-o teorie adevărată. Dar nu e deloc clar de ce ne-am asuma libertatea aceasta – nici noi, nici Thales nu folosim termenul „apă” pentru a ne referi la *tot*, ci îl folosim cu intenția de a ne referi *exact la apă*. Putnam ar vrea să forțeze asupra noastră o poziție precum cea a scepticului skolemit: dacă dispunem de mai multe interpretări pentru aceeași teorie și toate reușesc să o satisfacă, suntem îndreptățiți în egală măsură să o alegem pe oricare dintre ele. Dar atunci când dăm la schimb dezbaterea despre modele neintenționate ale teoriei mulțimilor cu o discuție despre interpretări nonstandard ale *termenilor comuni*, argumentele relativiste devin mult mai greu de acceptat. Dacă Putnam are dreptate, atunci ar trebui să acceptăm că referința este *absolut nedeterminată*, nu doar în interpretarea celorlalți, ci și în interpretarea propriei noastre vorbiri. Cum ar putea Putnam să aibă dreptate?

Aici își face intrarea cea mai controversată parte a argumentului său. Dacă crezi că referința e determinată în vreun fel anume, ar spune Putnam, atunci *doar accepți mai multă teorie* („just more theory”). Deci dacă te-ai împotrivi argumentului de mai devreme susținând că „apă” se referă la *apă*, sau poate adoptând o teorie mai cuprinzătoare a referinței precum teoria cauzală, tot ce ai reuși să faci ar fi să adaugi enunțuri noi la teoria ideală T_I. E imposibil să specifici la ce te referi altfel decât prin cuvinte, până la urmă; iar cuvintele pot fi formalizate și tratate ca teorie de ordinul I. Orice încercare de a fixa referința ar determina, atunci, o nouă teorie, să îi spunem T_R. Iar dacă

[24] Putnam 1977, p. 485.
[25] Elgin 1995, p. 290.

T_R nu determină nicio contradicție cu T_I, atunci $T_I \cup T_R$ e consistentă, deci are modele de orice cardinalitate infinită, deci poate fi satisfăcută printr-o corespondență directă cu *lumea*! Pe scurt, Putnam repetă argumentul de teoria modelelor *împotriva realistului* care susține că termenii au o referință determinată, în încercarea de a-l face să accepte că termenii pe care el însuși îi folosește *pentu a explica mecanismul de fixare a referinței* nu au o referință determinată. Dacă realistul ar susține, de pildă, că numele se referă la lucruri datorită unui act inițial de botez, propagat apoi de la vorbitor la vorbitor printr-un lanț cauzal, Putnam l-ar întâmpina astfel: ce înțelegi prin „se referă”?, ce înțelegi prin „botez”?, ce înțelegi prin „cauză”?... Ar fi inutil ca realistul să încerce să îi răspundă: prin „cauză” mă refer la X. Pentru că Putnam ar aplica *din nou* argumentul de teoria modelelor și ar reveni cu întrebarea: ce înțelegi prin „X”? Deci Putnam are, într-un fel, dreptate. Raționamentul său nu poate fi falsificat de niciun răspuns posibil. Însă pentru un filosof care se pretinde naturalist, tocmai acest lucru ar trebui să producă nu încredere sporită în propriul argument, ci îngrijorare. Căci știm, grație lui Popper, că o ipoteză care nu poate fi falsificată în niciun mod imaginabil ține de *dogmă*, nu de știință. O secundă – Popper era realist? Atunci poate ar trebui chestionat ce înțelege prin „dogmă”, pentru că există un model de aceeași cardinalitate cu lumea...

7.1 De ce argumentul eșuează

În continuare, propun să aruncăm o privire asupra răspunsurilor cu care a fost întâmpinat argumentul lui Putnam în literatură. Pentru că, în ciuda a ceea ce ar spera el, argumentul său este departe de a fi imun la critică. Într-adevăr, este imun la critica *frontală*, adică cea care încearcă să îi răspundă lui Putnam acceptându-i toate premisele și acceptându-i maniera de raționare. Însă premisele înseși, și procedura deductivă pe care se reazemă argumentul prin teoria modelelor sunt discutabile. Trebuie să fie clar de la început: cred că argumentul lui Putnam nu reușește, în ultimă instanță, să reprezinte un atac serios la adresa realismului. Asta nu înseamnă că teza în sine a lui Putnam – faptul că ar trebui să adoptăm un antirealism generalizat – nu poate fi susținută în niciun fel. Dar înseamnă, cred, că o astfel de teză nu se poate justifica prin argumente de felul celor oferite de Putnam, adică nu cu ajutorul teoriei modelelor. Dacă vom vrea să folosim tehnici din metalogică pentru a apăra vreo teză filosofică în legătură cu natura referinței, cred că nu putem depăși teza relativistă a lui Quine. Cred, adică, că nu e cu putință să punem la îndioală prin aceste mijloace capacitatea referențială până și a *propriei* vorbiri.

Să începem prin a considera *forma* argumentului oferit de Putnam. Nu e deloc clar în ce fel se presupune că ar conduce el la concluzia că trebuie să identificăm adevărul cu satisfiabilitatea. Pentru că el nu are o formă deductivă, în care să se aplice un număr finit de raționamente asupra premiselor până când se ajunge la concluzie. Nu – Putnam se bazează pe faptul că poate skolemiza răspunsul adversarului său realist *la nesfârșit*. Problema nu ar fi neapărat că „explicația trebuie să se oprească undeva",[26] așa cum îi reproșează Devitt. Problema este, în schimb, că tot la fel de bine poate și realistul să își repete răspunsul la infinit, ori de câte ori Putnam îi cere să explice la ce se referă prin cuvintele sale. Privind dintr-o poziție neutră, niciuna din cele două părți ale dezbaterii nu are vreun avantaj. Deci, cu excepția cazului în care realistul se plictisește să mai încerce să interacționeze cu Putnam (ceea ce e foarte posibil, dar, în orice caz, nu obligatoriu), argumentul nu demonstrează niciuneia dintre părți *nimic* din ce n-ar fi fost dispusă să accepte de la bun început. Privindu-i doar forma, deci, raționamentul lui Putnam are aparența clasică a unui *petitio principii* – ceea ce e, fără discuție, bizar pentru un argument produs de un autor serios ca Putnam. Ar fi posibil, deci, să punem la îndoială corectitudinea înțelegerii *noastre* a demonstrației pe care încearcă Putnam să facă. Vom explora și varianta aceasta un pic mai târziu. Dar, așa cum se prezintă acum, argumentul are o evidentă hibă logică. Și chiar dacă nu ar suferi de circularitate, toți comentatorii ar fi încă îndreptățiți dacă ar face observația (și au făcut-o) că *o constrângere asupra interpretărilor acceptabile ale unei teorii* pur și simplu *nu este* „doar mai multă teorie". Pentru că a impune restricții asupra funcțiilor de interpretare pe care le putem alege *nu e totuna* cu a interpreta restricțiile! Cu alte cuvinte, atunci când Putnam îi oferă realistului o interpretare SAT a $T_I \cup T_R$ și îi arată că astfel $T_I \cup T_R$ e adevărată-conform-SAT, realistul ar contra că teoria sa T_R nu e adevărată dacă e satisfăcută alături de T_I, ci e adevărată dacă însăși relația de satisfacere SAT a lui T_I se conformează restricțiilor impuse de T_R. Acest argument nu e foarte diferit de a spune că o *constrângere* nu e o propoziție cu același fel de conținut semantic ca cel al unei propoziții *descriptive*. Or, cu o remarcabilă consistență, Putnam pare să echivaleze între cele două.

Mai există vreo speranță pentru argumentul lui Putnam? Există vreun mod în care i-am putea reconstrui demonstrația, în așa fel încât să fie ferită de majoritatea criticilor care i-au fost aduse? Igor Douven crede că da, și oferă o reconstrucție a argumentului lui Putnam în manieră formală.[27] În cele ce urmează voi sintetiza în limbaj natural reconstrucția lui Douven. Amintesc că Putnam își intenționează raționamentul ca un *reductio* la adresa realismului moderat, încercând să demonstreze că o implicație particulară a realismului duce la rezultate contradictorii. Acea implicație a realismului este teza că până și o teorie ideală din punct de vedere epistemic poate fi falsă *în realitate*. Douven o numește teza *failibilismului metodologic*. Acum, argumentul de teoria modelelor al lui Putnam, dacă îl despărțim de subargumentul „just more theory", arată că pentru orice teorie consistentă există o corespondență SAT a termenilor ei cu *lumea* în așa fel încât teorie este adevărată-conform-SAT. Mai departe, Putnam ar vrea să convingă realistul că adevărul-conform-SAT e identic cu adevărul *pur și simplu*. Douven arată că poate obține acest rezultat *fără să recurgă* la „just more theory", adică la partea cea mai șubredă a demonstrației sale. Căci chiar dacă realistul

[26] Devitt 1983, p. 299.
[27] Douven 1999, în special pp. 488-9.

susține că dispune de o teorie a referinței care îi permite să distingă între o corespondență oarecare cu realitatea și corespondența *intenționată* a termenilor (să presupunem chiar că e *cea mai bună* teorie a referinței cu putință, din punct de vedere epistemic); câtă vreme se angajează și la teza failibilismului metodologic, realistul e forțat să admită că acea teorie a referinței *poate fi falsă*. Or, în posibilitatea în care ea e falsă, nu ar avea cum să mai distingă între interpretarea intenționată a termenilor și interpretările nonstandard; deci ar fi obligat să admită, odată cu Putnam, că orice corespondență a unei teorii cu realitatea e acceptabilă dacă satisface teoria. Dar argumentul de teoria modelelor arată că orice teorie consistentă admite o relație de satisfacție asupra realității. În particular, teoria ideală a referinței la care ține realistul e consistentă (chiar dacă, sub posibilitatea luată în considerare, ea e falsă). Deci teoria ideală a referinței admite o corespondență SAT cu realitatea. Cum nu poate face distincția între corespondențe *intenționate* și corespondențe *nonstandard*, realistul trebuie să accepte că SAT garantează adevărul *simpliciter* al teoriei referinței. Dar tocmai îmbrățișase presupunerea că ea e falsă! Așadar, în lumea posibilă în care teoria ideală a referinței e falsă ajungem la o contradicție. Pe care presupunere cade responsabilitatea acestei contradicții? Pe teza failibilismului metodologic – căci de pe urma ei a fost dispus realistul să accepte că teoria ideală a referinței *poate fi falsă în realitate*. Dar dacă am redus la absurd failibilismul metodologic, atunci am arătat absurditatea realismului în ansamblul lui, pentru ca failibilismul metodologic este o consecință directă a realismului.

Reconstrucția lui Douven pare să fie cea mai promițătoare variantă de a resuscita argumentul de teoria modelelor. Problema lui cea mai acută, faptul că subargumentul „just more theory" se rezumă la un cerc vicios, dispare – pentru că „just more theory" nu mai este invocat *deloc* în noua demonstrație. Ceea ce rămâne în noua demonstrație este în mod evident corect, presupunând corectitudinea argumentului prin care Putnam determină o corespondență a oricărei teorii cu realitatea. Dar ce susțin eu este că exact în acel argument se află eroarea. Iar, de data aceasta, ea nu mai poate fi corectată printr-o reconstrucție a întregii demonstrații, pentru că ea este inerentă formalismului cu care Putnam alege să își susțină antirealismul. Mai precis, Putnam scapă din vedere teorema lui Tarski (4.1).

Un contraargument similar celui pe care îl voi oferi în continuare e formulat de Timothy Bays împotriva unei demonstrații mai restrânse pe care încearcă Putnam să o facă, și anume împotriva *argumentului constructivizării*.[28] Fără a mai zăbovi asupra detaliilor, în esență ce își propunea Putnam prin respectivul argument era să arate că adevărul unui anumit enunț dovedit *independent* de axiomele ZFC ($V = L$) este compatibil cu *orice* domeniu de referință am vrea să atribuim modelului în care axiomele sunt satisfăcute. Dar ce evidențiază Bays este că Putnam aplică greșit teorema Löwenheim-Skolem. Că, asemenea lui Skolem în 1922, omite faptul că unele clase de obiecte sunt *prea mari* ca să poată fi tratate ca mulțimi – în cazul lui Skolem, era vorba de universul

[28] Bays 2001, în special secțiunea 2.

von Neumann V, în cazul lui Putnam e vorba de universul constructibil al lui Gödel L – și nu își poate corecta greșeala decât dacă adoptă în metateorie existența unui cardinal inaccesibil. În lipsa acestei axiome adiționale, și Skolem și Putnam ar fi puși în situația absurdă de a defini un predicat de adevăr pentru *însăși teoria în care se află*. Dar pentru Putnam, adoptarea de axiome noi e fatală: căci se presupune că teoria pe care o modelează e *deja* teoria ideală, care conține toate constrângerile pe care am putea dori să i le impunem. Acesta este, pe scurt, contraargumentul lui Bays.

Acum, aceste observații cred că ar trebui să ne facă mai atenți la detaliile formale prin care Putnam încearcă sa își justifice și argumentul principal, cel în legătură cu relația oricărei teorii ideale cu lumea. Căci, în primul rând, este discutabil dacă un realist ar accepta să trateze „lumea" ca obiect al discuției sale, așa cum o face Putnam – și nu doar *elementele* lumii. Iată la ce mă refer. Când Putnam ne invită să alegem un model „de aceeași cardinalitate cu a lumii", ne putem întreba ce înțeles are în minte pentru termenul de „cardinalitate". Dacă se referă la ea în sensul clasic din teoria mulțimilor, atunci ar trebui să se gândească la ceva de felul *celui mai mic număr cardinal care poate fi pus într-o bijecție cu lumea*. Dar acea bijecție ar trebui să existe *ea însăși* în lume, iar domeniul și codomeniul ei ar trebui să fie obiecte ale lumii. Dar asta ar face lumea un obiect al ei înseși. Nu e nimic contrar poziției realiste să susții că poți vorbi despre *elementele* lumii, dar nu poți vorbi despre lumea însăși; în același fel în care înăuntrul teoriei mulțimilor poți vorbi despre *mulțimi*, dar nu poți vorbi despre *clasa tuturor mulțimilor*. Or, doar acest fapt ar fi suficient pentru a respinge argumentul lui Putnam de pe o poziție realistă. Dar să presupunem că suntem realiști *și acceptăm* că lumea își este sieși element. Argumentul lui Putnam este în continuare profund eronat. Pentru că teoria ideală ar trebui să fie cea mai cuprinzătoare teorie cu putință; ea ar trebui să fie și metateoria în care orice argument formal ar fi exprimat. Dar ce ar face Putnam ar fi să ofere un model pentru teoria ideală, deci, practic, să definească un predicat de adevăr. Dar asta în timp ce s-ar afla exact în cadrul aceleiași teorii al cărei model îl determină! Or asta e imposibil, pentru că ar contrazice teorema lui Tarski, exact ca în situația evidențiată de Bays în legătură cu argumentul constructivizării. Mai mult, un realist nu s-ar face neapărat vinovat de a încălca teorema lui Tarski. Pentru că, așa cum arată Lewis, teoria cauzală a referinței nu vizează felul în care *orice termen* al vocabularului nostru își dobândește referința.[29] Dimpotrivă, doar referința unei clase foarte restrânse de termeni poate fi explicată astfel (doar termenii singulari și unele clase naturale) – și, în consecință doar adevărul unei clase foarte restrânse de enunțuri decurge din această teorie a referinței. Deci un realist moderat *nu* definește un predicat de adevăr pentru întreaga teorie în care se află, pe câtă vreme Putnam, dacă i-am accepta argumentul, ar face exact acest lucru. Iar din acest motiv, până și argumentul de teoria modelelor *lipsit* de subargumentul „just more theory" cred că trebuie respins. Așadar, nici reconstrucția lui Douven, în ciuda atractivității ei, nu poate sta

[29] Lewis 1984, p. 235.

în picioare, pentru că presupune corectitudinea procedeului formal urmat de Putnam.

8 Concluzii

Cum ar trebui să înțelegem argumentul lui Putnam? Ce ar trebui să extragem din el, odată ce am hotărât că trebuie respins? Din ce motiv este el mai greu de acceptat decât argumentul similar oferit de Quine?

Motivul trebuie căutat, cred, în următoarea direcție. Putnam probabil consideră argumentele lui Quine și ale lui Benacerraf ca un fel de *precedent* în legătură cu ce e justificată teoria modelelor să demonstreze din punct de vedere ontologic. Însă nici Benacerraf și nici Quine nu își propun să dinamiteze sensul limbajului în care propriile lor argumentele formale sunt exprimate. Pentru că, în măsura în care Benacerraf poate fi luat ca promotorul unui fel de relativism referențial, argumentele lui vizează totuși *doar* limbajul aritmetic. Mai departe, Quine face un salt care este fără îndoială remarcabil: referința atribuită vorbirii *oricărei alte persoane* este nedeterminată, pentru că poate fi transpusă în feluri diferite în propriii noștri termeni. Și cum *limba e o artă socială*, ontologia devine pentru Quine o problemă secundară adevărului și teoretizării științifice. Dar Putnam ar vrea să obțină mult. Putnam ar vrea să folosească un fel foarte similar de argumentare formală ca al lui Quine pentru a reduce la absurd întreaga noțiune de *lume exterioară* – nu doar pentru că vorbirea *celorlalți* poate fi pusă în corespondențe diferite cu lumea, dar și pentru că *propriul limbaj* admite orice corespondență cu exteriorul am prefera să îi atribuim. Dar astfel, Putnam plonjează direct în absurditate, căci împinge o metodă logică de argumentare mai departe decât mai poate fi ea folosită cu sens. În momentul în care până și propria vorbire poate fi *dezinterpretată* și apoi *reinterpretată*, ar putea părea că suntem liberi pe cale formală să definim adevărul exact așa cum dorim. *Dar formalismul însuși nu mai poate funcționa* atunci, pentru că el are sens numai *relativ* la teoria în care îl exprimăm. Numai relativ, cu alte cuvinte, la o înțelegere a propriei limbi în care acceptăm că sensurile cuvintelor trebuie „luate de bune”, după cum sugerează Quine. Niciun rezultat util din punct de vedere filosofic nu poate fi produs dacă încercăm să ne ridicăm deasupra propriului limbaj.

Să recapitulăm, așadar, cum am ajuns la aceste concluzii. După ce am definit noțiunile de bază legate de sintaxa logicii de ordinul I și construcția modelelor pentru mulțimi de formule, am demonstrat un rezultat limitativ fundamental din metalogică: teorema lui Tarski. Am aruncat apoi o privire asupra conexiunii dintre teoria modelelor și argumentele lui Paul Benacerraf și W. V. Quine în legătură cu natura *structurală* a obiectelor la care se angajează diferite teorii. Am început apoi să discutăm teza relativistă mult mai dură a lui Hilary Putnam. Pentru a face asta, a fost nevoie să demonstrăm două versiuni ale teoremei Löwenheim-Skolem și să discutăm implicațiile ei din filosofia matematicii. Am văzut, apoi, că încercarea lui Putnam de a extinde dezbaterea lansată de

Skolem asupra referinței *în general* nu poate fi susținută. Căci ea fie se sprijină pe un argument circular, fie intră în contradicție cu teorema lui Tarski. Per ansamblu, acest periplu filosofic a fost util pentru a ne arăta tipul de dezbateri prin care ne poate purta metoda analizei logice a limbajului, dar și limitările ei.

9 Anexă

9.1 Axiomele ZFC

1. **Axioma extensionalității**: $\forall u(u \in X \iff u \in Y) \implies X = Y$
   Care asigură un principiu al identitații indiscernabililor pentru mulțimi: dacă două mulțimi au aceleași elemente (sunt indiscernabile după relația $\in$), ele sunt identice.

2. **Schema axiomatică a specificației**: $\forall p \forall X \exists Y \forall u(u \in Y \iff u \in X \wedge \varphi(u, p))$
   Care garantează că pentru orice mulțime X, există $Y = \{y \,|\, y \in X \wedge \varphi(y, u)\}$, adică putem forma o submulțime a lui X specificându doar elementele care satisfac o anumită formulă φ. Cum φ poate fi oricare dintr-o mulțime infinită de formule (după cum vom arăta în 9.2), axioma aceasta este de fapt o schemă axiomatică.

3. **Axioma perechilor:** $\forall a \forall b \exists X \forall u(u \in X \iff u = a \vee u = b)$
   Care asigură că se pot forma mulțimi care conțin exact două elemente: pentru orice a și b, există $X = \{a, b\}$. Printre altele, acest lucru e util pentru că ne permite să definim perechi ordonate $\langle a, b\rangle = \{a, \{a, b\}\}$, care se pot extinde la n-tupluri $\langle v_1, ..., v_n\rangle$.

4. **Axioma reuniunii:** $\forall X \exists Y \forall u(u \in Y \iff \exists x(x \in X \wedge u \in z))$
   Care garantează că există o mulțime $Y = \bigcup X = \{x \in A \,|\, A \in X\}$, care conține toate elementele elementelor lui X.

5. **Axioma mulțimii părților:** $\forall X \exists Y \forall u(u \in Y \iff \forall v(v \in u \implies v \in X))$
   Care garantează că există o mulțime $Y = \mathscr{P}(X) = \{x \,|\, x \text{ e o submulțime a lui } X\}$, prin care putem introduce produse carteziene, grafice și funcții.

6. **Schema axiomatică a înlocuirii:** $\forall p \forall X \exists Y \forall u(u \in Y \iff \exists x(x \in X \wedge u = f(x, p)))$
   Introdusă independent de Fraenkel și Skolem, garantează că imaginea oricărei funcții e o mulțime.

7. **Axioma infinitului:** $\exists S(\exists y\, y \in S \wedge \forall x(x \in S \implies x \cup \{x\} \in S))$
   Care garantează că există o mulțime infinită, luată ca mulțime *inductivă*.

8. **Axioma alegerii:** $\forall X \exists f : X \to \bigcup X\, (\forall A(A \in X \implies f(A) \in A))$
   Am avea nevoie de mult mai mult decât un scurt paragraf pentru a prezenta cum se cuvinte

axioma alegerii. Ne vom limita să o descriem, însă, foarte succint: pentru orice familie de mulțimi (posibil infinită), există o funcție care alege exact un element din fiecare mulțime. În demonstrația teoremei Löwenheim-Skolem, am folosit-o pentru a alege câte o instanță a fiecărei propoziții existențiale adevărate.

9. **Axioma regularității:** $\forall X \exists u (u \in X \wedge u \cap X = \varnothing)$

   Orice mulțime X are un element minimal relativ la relația de apartenență $\in$. (Un element u e $\in$-minimal în X dacă nu există niciun v în X așa încât $u \in v$. Sau, echivalent: $u \cap X = \varnothing$, ceea ce duce la formularea din axiomă.)

Pe această ultimă axiomă ne vom concentra ceva mai mult. Iată importanța ei, în termeni mai potriviți unei lucrări de filosofie: înlătură posibilitatea regresiei la infinit atunci când enumerăm toate elementele unei mulțimi, și toate elementele acelor elemente, și așa mai departe. Lanțul apartenențelor *trebuie* să aibă o lungime finită, indiferent de mulțimea de la care plecăm. În particular, o mulțime T care își înlănțuie propriile elemente – dacă $u \in T$, atunci toate elementele lui u sunt și ele în T, deci $u \subset T$ – se numește o mulțime tranzitivă.

O relație E pe o clasă C care are proprietatea cerută de axioma regularității – anume, orice subclasă a lui C are un element E-minimal – se numește o relație *bine fondată*. În mod uzual când interpretăm axioma regularității, relația la care considerăm că face referire este relația uzuală de apartenență $\in$, iar clasa pe care o luăm ca domeniu al relației este clasa uzuală a tuturor mulțimilor V. Dar axioma regularității impune constrângerea de bună fondare asupra *oricărui model al ei*, nu doar asupra perechii $(V, \in)$: orice pereche (C, E), dacă prin intermediul ei se interpretează relația de apartenență din teoria mulțimilor, desemnează o relație bine fondată. Deci axioma regularității garantează că *orice model* al teoriei mulțimilor este bine fondat – orice înlănțuire de elemente ale lui C care intră în relația E *trebuie* să fie finită ($c_1 \; E \; c_2 \; E \; c_3 \; E \ldots E \; c_n$).

Iată acum motivul pentru care axioma regularității ne interesează în mod special. Consecința faptului că orice model al axiomelor ZFC este bine fondat este lema lui Mostowski:

Teorema 9.1 (Lema lui Mostowski)**.** Pentru orice model $\mathscr{A} = (A, E)$ al axiomelor ZFC, există un alt model $\mathscr{T} = (T, \in)$ izomorf cu $\mathscr{A}$, al cărui domeniu T este tranzitiv și care atribuie relației de apartenență interpretarea ei uzuală. [Jech 2006]

Demonstrație (schiță). E suficient să găsim un izomorfism între $\mathscr{A}$ și $\mathscr{T}$. Fie următoarea funcție pe A:

$$\pi(a) = \{\pi(v) \mid v \; E \; a\}$$

π e o funcție recursivă; însă buna fondare a relației (A, E) garantează că recursia are un caz de bază care împiedică regresul la infinit, pentru că nicio mulțime a nu determină un lanț infinit de elemente ... $E\ v_2\ E\ v_1\ E\ a$.

Remarcăm, în primul rând, că $\pi(a) = \pi(b)$ dacă și numai dacă a și b intră în relația E cu exact aceleași obiecte din domeniu. Cum $\mathscr{A}$ e un model al ZFC, $\mathscr{A}$ satisface axioma extensionalității: două obiecte indiscernabile după relația E sunt de fapt unul și același obiect. Deci $\pi(a) = \pi(b)$ dacă și numai dacă $a = b$. Așadar π e o bijecție între A și $\mathrm{Im}(\pi)$.

Mai apoi:

$$\text{Dacă } u\ E\ v, \text{ atunci } \pi(u) \in \pi(v),$$

adică relația E pe domeniul A e transpusă în apartenența uzuală $\in$ pe domeniul format de $\mathrm{Im}(\pi)$. Aceste două proprietăți luate împreună arată că π e un izomorfism între (A, E) și $(\mathrm{Im}(\pi), \in)$.

Acum, cel mai important, imaginea lui π e tranzitivă:

$$\text{Dacă } u \in Im(\pi) \text{ și } v \in u, \text{ atunci } v \in Im(f).$$

Fie atunci $T = Im(\pi)$. $\mathscr{A}$ e izomorf cu $\mathscr{T}$ și T e tranzitiv. □

Teorema 9.2. Orice reuniune numărabilă de mulțimi finite sau numărabile este numărabilă.

9.2 Rezultate ajutătoare pentru Löwenheim-Skolem

Teorema 9.3 (Cardinalitatea lui *Form(L)*)**.** Fie $L = \langle R, F, C, V, ar \rangle$ o semnătură. Atunci, $|Form(L)| = \aleph_0$.

Demonstrație. Fie $X = R \cup F \cup C \cup V \cup \{\exists, \forall, \implies, \land, \lor\}$ mulțimea tuturor simbolurilor care pot apărea într-o formulă a lui L. Cum orice formulă este o concatenare finită de simboluri din X, avem $Form(L) \subseteq \bigcup_{n \in \mathbb{N}} X^n$, iar potrivit 9.2, $\bigcup_{n \in \mathbb{N}} X^n$ e numărabilă. Așadar, $|Form(L)| \leq \aleph_0$.

Fie R o relație de aritate $ar(R)$ și v o variabilă. Luăm $f : \mathbb{N} \to Form(L)$, $f(n) = \exists v \bigwedge_n R(\bar{v})$, unde $\bar{v}$ reprezintă o listă $v, \ldots, v$ în care variabile v este repetată de $ar(R)$ ori. Cum pentru orice număr natural n există o unică formulă $f(n)$, f este injectivă, deci $|Form(L)| \geq \aleph_0$.

Așadar, $|Form(L)| = \aleph_0$.

□

Teorema 9.4. Orice scufundare conservă adevărul formulelor fără cuantificatori.

Demonstrație. Fie o scufundare π de la o L-structură $\mathscr{A}$ la o L-structură $\mathscr{B}$. Pentru orice L-formulă necuantificată $\varphi(\bar{x})$ și orice listă de parametri $\bar{a} \in A$, vrem să arătam că:

$$\mathscr{A} \models \varphi(\bar{a}) \text{ ddacă } \mathscr{B} \models \varphi(\pi(\bar{a}))$$

Vom proceda prin inducție pe formule. O formulă necuantificată e formată dintr-una sau mai multe formule atomare, legate prin conectori logici. La rândul ei, o formulă atomară poate avea una din două forme: $t_1 = t_2$, sau $R(t_1,...,t_n)$, pentru t_i termeni. Să începem, deci, prin a vedea ce efect are scufundarea π asupra interpretării termenilor.

Fie $t(x_1,...,x_n)$ un termen din L, $x_1,...,x_n$ fiind o listă de variabile ale lui L incluzând toate variabilele care apar în t. Fie $t^{\mathscr{A}}$ funcția de interpretare a lui t în $\mathscr{A}$, și $a_1,...,a_n \in A$ parametrii pentru care se evaluează variabilele lui t. Atunci:

$$\begin{aligned}
\pi(t^{\mathscr{A}}(a_1,...,a_n)) &= \begin{cases} \pi(c^{\mathscr{A}}) & \text{dacă t e constanta c,} \\ \pi(a_i) & \text{dacă t e a i-a variabilă,} \\ \pi(F^{\mathscr{A}}(a_1,...,a_n)) & \text{dacă t e funcția F cu n variabile libere} \end{cases} \\
&= \begin{cases} c^{\mathscr{B}} & \text{dacă t e constanta c,} \\ \pi(a_i) & \text{dacă t e a i-a variabilă,} \\ F^{\mathscr{B}}(\pi(a_1),...,\pi(a_n)) & \text{dacă t e funcția F cu n variabile libere} \end{cases} \\
&= t^{\mathscr{B}}(\pi(a_1),...,\pi(a_n))
\end{aligned}$$

Folosindu-ne de acest lucru, vom arăta că orice scufundare conservă adevărul formulelor atomare.

$$\begin{aligned}
\mathscr{A} \models t_1(a_1,...,a_n) = t_2(a_1,...,a_n) \text{ ddacă } & t_1^{\mathscr{A}}(a_1,...,a_n) = t_2^{\mathscr{A}}(a_1,...,a_n) \\
\text{ddacă } & f(t_1^{\mathscr{A}}(a_1,...,a_n)) = f(t_2^{\mathscr{A}}(a_1,...,a_n)) \\
\text{ddacă } & t_1^{\mathscr{B}}(\pi(a_1),...,\pi(a_n)) = t_2^{\mathscr{B}}(\pi(a_1,...,a_n) \\
\text{ddacă } & \mathscr{B} \models t_1(\pi(a_1),...,\pi(a_n)) = t_2(\pi(a_1),...,\pi(a_n))
\end{aligned}$$

$$\begin{aligned}
\mathscr{A} \models R(t_1(\bar{a}),...,t_m(\bar{a})) \text{ ddacă } & (t_1^{\mathscr{A}}(\bar{a}),...,t_m^{\mathscr{A}}(\bar{a})) \in R^{\mathscr{A}} \\
\text{ddacă } & (\pi(t_1^{\mathscr{A}}(\bar{a})),...,\pi(t_m^{\mathscr{A}}(\bar{a}))) \in R^{\mathscr{B}} \\
\text{ddacă } & (t_1^{\mathscr{B}}(\pi(\bar{a})),...,t_m^{\mathscr{B}}(\pi(\bar{a}))) \in R^{\mathscr{B}} \\
\text{ddacă } & \mathscr{B} \models R(t_1(\pi(\bar{a})),...,t_m(\pi(\bar{a})))
\end{aligned}$$

Așadar, π conservă adevărul tuturor formulelor atomare.

O formulă necuantificată e formată din formule atomare legate prin conectori logici. Cum conectorii logici sunt interpretați la fel în orice structură, π conservă adevărul și al oricărei formule necuantificate. $\square$

Bibliografie

Bays, Timothy (2001), „On Putnam and His Models”, în *Journal of Philosophy* 98.7, pp. 331–350.

Benacerraf, Paul (1965), „What Numbers Could Not Be”, în *The Philosophical Review* 74.1, pp. 47–73.

— (1985), „Skolem and the Skeptic”, în *Proceedings of the Aristotelian Society, Supplementary Volumes* 59, pp. 85–115.

Button, Tim (2011), „The Metamathematics of Putnam's Model-Theoretic Arguments”, în *Erkenntnis (1975-)* 74.3, pp. 321–349.

Davidson, Donald (1967), „Truth and Meaning”, în *Synthese* 17.3, pp. 304–323.

Devitt, Michael (1983), „Realism and the Renegade Putnam: A Critical Study of Meaning and the Moral Sciences”, în *Noûs* 17.2, pp. 291–301.

Douven, Igor (1999), „Putnam's Model-Theoretic Argument Reconstructed”, în *Journal of Philosophy* 96.9, pp. 479–490.

Elgin, Catherine Z. (1995), „Unnatural Science”, în *Journal of Philosophy* 92.6, pp. 289–302.

Jech, Thomas (2006), *Set Theory*, The Third Millennium Edition, revised and expanded, Pure and applied mathematics, a series of monographs and textbooks 79, Academic Press.

Kirby, Jonathan (2019), *An Invitation to Model Theory*, Cambridge University Press.

Lewis, David (1984), „Putnam's Paradox”, în *Australasian Journal of Philosophy* 62.3, pp. 221–236.

Putnam, Hilary (1977), „Realism and Reason”, în *Proceedings and Addresses of the American Philosophical Association* 50.6, pp. 483–498.

— (1980), „Models and Reality”, în *Journal of Symbolic Logic* 45.3, pp. 464–482.

— (1981), *Reason, Truth and History*, Cambridge University Press.

Quine, Willard van Orman (1964), „Ontological Reduction and the World of Numbers”, în *The Journal of Philosophy* 61.7, pp. 209–216.

— (1968), „Ontological Relativity”, în *The Journal of Philosophy* 65.7, pp. 185–212.

— (1981), „Things and Their Place in Theories”, în *Theories and Things*, Londra, Marea Britanie: Harvard University Press, pp. 1–23.

Quine, Willard van Orman [1994] (2008), „Assuming Objects”, în *Confessions of a Confirmed Extensionalist and Other Essays*, ed. de Dagfinn Føllesdal și Douglas Quine, Cambridge (Mass.): Harvard University Press, pp. 449–460.

Skolem, Thoralf [1920] (1981a), „Logico-combinatorial investigations in the satisfiability or provability of mathematical propositions: A simplified proof of a theorem by L. Löwenheim and generalizations of the theorem”, în *From Frege to Goedel: a source book in mathematical logic, 1879 - 1931*, ed. de Jean Van Heijenoort, Cambridge (Mass.): Harvard University Press, pp. 252–263.

— [1922] (1981b), „Some remarks on axiomatized set theory”, în *From Frege to Goedel: a source book in mathematical logic, 1879 - 1931*, ed. de Jean Van Heijenoort, Cambridge (Mass.): Harvard University Press, pp. 290–301.