Modeliranje pojma istine pomoću najveće intrinsične fiksne točke jake Kleeneove trovaljane semantike

Boris Čulina

Zagreb, 2004.

SADRŽAJ

Uvod

Kad o nečem govorimo prirodno se nameće i razmišljanje o istinitosti i lažnosti rečenica kojima to govorimo. Tako se pojam istine javlja jednim od temeljnih pojmova. U svakodnevnoj upotrebi čini se da imamo jasnu intuiciju o tom pojmu. Jasan nam je smisao rečenica kojima tvrdimo da su neke rečenice istinite ili lažne. Uzmimo za primjer sljedeće rečenice:

(1) Prva rečenica Uvoda je istinita.

(2) Sedma rečenica Uvoda je lažna.

Čini se da znamo što ove rečenice govore. Štoviše, na osnovu poznavanja njihova smisla određujemo i njihovu istinitost. Da bismo utvrdili je li istinita rečenica (1) moramo pogledati što ona govori. A ona govori o istinitosti prve rečenice Uvoda. Tako se problem njene istinitosti svodi na problem istinitosti prve rečenice Uvoda:

Rečenica (1) je istinita ako i samo ako prva rečenica Uvoda je istinita.

Kako je, po meni, prva rečenica Uvoda istinita to zaključujemo i da je rečenica (1) istinita. Na isti način, razumijevajući smisao rečenice (2), pitanje njene istinitosti svodimo na pitanje istinitosti sedme rečenice Uvoda:

Rečenica (2) je istinita ako i samo ako sedma rečenica Uvoda je lažna.

Lako je prebrajanjem utvrditi da je sedma rečenica Uvoda upravo rečenica (2). Dakle, iz prethodnog kriterija dobivamo:

Rečenica (2) je istinita ako i samo ako rečenica (2) je lažna.

A to je kontradikcija!

Prethodno razmatranje ilustrira opće pravilo: *intuitivno poimanje istine koje nas obično vodi odgovoru na pitanje je li neka rečenica istinita ili ne, u nekim situacijama vodi kontradikciji ili, općenito govoreći, zbunjujućim uvjetima, tzv.* **paradoksima istine**. Oni nisu samo "zavrzlame za razbibrigu" već i "simptomi bolesti" — pokazuju da nešto nije u redu s našim temeljnim poimanjem značenja jezika. Pošto svako logičko zasnivanje znanja teži iznalaženju što univerzalnijeg jezika u kojem se u idealnom slučaju može modelirati sve razmišljanje, pa i ono temeljno o istinitosti rečenica, to je analiza paradoksa istine nezaobilazni dio tog zasnivanja. *Paradoksi ukazuju na postojanje problema i test su za svako ponuđeno rješenje.* Kako je pojam istine ujedno i jedna od temeljnih filozofskih preokupacija koju možemo sažeto opisati kao traženje odgovora na vječno pitanje "Što je istina?", to je u prvom poglavlju točnije opisano kojim aspektom tog pojma se ovaj rad bavi. Taj aspekt je ovdje nazvan **logički aspekt pojma istine**. Također je opisano nekoliko paradoksa istine od kojih svaki nosi određeni aspekt problema s kojima se svako rješenje treba suočiti.

Mada paradoksi istine imaju tisućljetnu i na momente dramatičnu povijest[1], odlučujući korak u logičkoj analizi pojma istine napravio je Alfred Tarski u članku "Pojęcie prawdy w językach nauk dedukcyjnych", 1933 g [Tar33]. *U tom članku Tarski se bavi problemom definicije pojma istine i pokazuje pod kojim uvjetima i kako se ona može formulirati. Negativan rezultat analize je da za bogatiji jezik (npr. jezik koji sadrži jezik aritmetike) pojam istine ne može biti definiran u samom jeziku (to je sadržaj Tarskijevog teorema o nedefinabilnosti istine) već u bitno bogatijem metajeziku. Pozitivan rezultat je ustanovljavanje definicije pojma istine jezika u odgovarajućem metajeziku (to je poznata Tarskijeva rekurzivna definicija istine.* Time je pokazao da se precizno može opisati ne samo forma nego i značenje naučnih jezika, koje su logičari do tada koristili neformalno, smatrajući semantičke pojmove suviše nejasnim (dijelom i pod utjecajem paradoksa istine) za logičku analizu (vidjeti [Wol02, Sin01]). Uvođenjem semantike u metamatematiku (matematičko proučavanje matematičih teorija), koja se do tada sastojala od formalnih istraživanja jezika i teorija u duhu Hilbertovog finitizma, bitno je obogaćena matematička logika. Semantika ima svoj konačan izraz u teoriji modela, danas možda i najplodnijem dijelu logike. Teorija modela koristi jezik teorije skupova kao jezik u kojem se može opisati semantika, specijalno i pojam istinite rečenice, gotovo svih jezika osim, naravno, samog jezika teorije skupova. Ovakvo rješenje je sasvim zadovoljavajuće za matematičku praksu. Čak i po pitanju zasnivanja znanja ima određenu uvjerljivost. Naime, razmišljanje o pojmu istine nekog jezika je refleksija (pogled izvana) nad tim jezikom i prirodno se odvija u odgovarajućem metajeziku. Isto tako se i razmišljanje o metajeziku

[1] Zapisi o Aleksandrijskom pjesniku Philetasu govore da je na njegovom nadgrobnom spomeniku pisalo da su ga ubile besane noći u kojima je razmišljao o paradoksima istine.

prirodno odvija u njegovom metajeziku, itd. Tako je stalno pribjegavanje metajeziku imanentno refleksivnosti razmišljanja. Tarskijeva analiza pojma istine sadržaj je drugog poglavlja ovo rada.

Što se tiče paradoksa istine Tarskijeva analiza ih blokira na način da se oni uopće ne mogu izraziti, jer definicija istine jezika u njegovom metajeziku ne dozvoljava da rečenice govore, direktno ili indirektno, o vlastitoj istinitosti. Problem je što se takvim rješenjem blokiraju i mnoge druge sasvim uobičajene i bezazlene rečenice. Nadalje, nesumnjiva je privlačnost jezika koji bi bio zatvoren na refleksivnost razmišljanja. Naš prirodni jezik je takav univerzalan jezik, no to je jezik neprecizne forme i značenja. Idealan jezik bi trebao biti njegova sinaktički i semantički pročišćena varijanta. Takav precizan univerzalni jezik je ideal logičkog zasnivanja znanja. U umjetnoj inteligenciji to je idealan jezik za inteligentne sustave. U lingvistici to bi bio traženi model za razumijevanje prirodnog jezika. Konstrukcija jezika koji sadrži vlastiti pojam istine sigurno bi bio značajan korak u tom smjeru. Sve su to razlozi zašto je usprkos Tarskijevom teoremu o nedefinabilnosti istine i uspješnosti njegovog rješenja za matematiku i nauku općenito, ipak nastavljeno razmišljanje o mogućnosti takvog jezika. Odlučujući korak u tom smjeru u filozofskom i matematičkom smislu uradio je Saul Kripke u članku "Outline of a Theory of Truth", 1975.godine[Kri75]. *Kripke je dao jednu drugačiju logičku analizu pojma istine, koja je mnogo bliža korištenju tog pojma u prirodnom jeziku nego Tarskijeva analiza, i koja daje jezike koji sadrže vlastiti predikat istine.* Ti jezici imaju trovaljanu semantiku (istina, laž i neodređeno) i zbog toga ne potpadaju pod Tarskijev teorem o nedefinabilnosti istine koji vrijedi za klasične jezike s dvovaljanom semantikom (istina, laž). Kripkeova analiza je filozofski argumentirana i daje preciznu matematičku konstrukciju takvih jezika. Za primjer uzmimo ponovo rečenicu

(2) Sedma rečenica Uvoda je lažna.

Ono zbog čega pojam istine u slučaju ove rečenice vodi kontradikciji je *empirička činjenica* da je upravo ta rečenica sedma rečenica Uvoda. Da je neka druga rečenica Uvoda bila sedma rečenica vjerojatno bismo imali normalnu upotrebu pojma istine. No, *ovo razlikovanje normalnog i paradoksalnog ne možemo izolirati na sintaktičkom niti na unutrašnjem semantičkom nivou, jer to ovisi o realnosti, a ne o načinu kako koristimo jezik. Zato je po Kripkeu nužno ovaj rizik uključiti u teoriju istine. Rečenice koje govore o istinitosti drugih rečenica, mada sintaktički ispravne i smislene, pod nekim uvjetima koji ovise o realnosti na koju se jezik odnosi mogu ne dati određeno tvrđenje o toj realnosti, tj. ne dati klasičnu istinitosnu vrijednost, istinu ili* *laž. Njima tada pridružujemo treću vrijednost, neodređeno .* Smisao treće vrijednosti je naprosto da rečenica nema klasičnu istinitosnu vrijednost. Takva analiza vodi proučavanju jezika s trovaljanom semantikom. Kripke je pokazao određenom matematičkom konstrukcijom kako za cijelu klasu trovaljanih semantika, tzv. monotonih semantika, postoje jezici koji sadrže vlastiti pojam istine, tzv. **fiksne točke date semantike**. Kripkeova analiza pojma istine sadržaj je trećeg poglavlja ovog rada.

Pošto postoji više trovaljanih monotonih semantika i za svaku od njih općenito više pripadnih fiksnih točaka nad datim jezikom, nameće se pitanje koja semantika i koja fiksna točka date semantike odgovaraju intuitivnom poimanju istine. Kripke je ostao neutralan po tom pitanju. No iz članka je jasno da je preferirao tzv. **najmanju fiksnu točku jake Kleeneove trovaljane semantike**. Za nju je dao i jednu intuitivnu motivaciju koja odgovara ideji refleksivnosti razmišljanja. Kripke je u svom radu istaknuo i tzv. **najveću intrinsičnu fiksnu točku** zbog njenog posebnog položaja u strukturi fiksnih točaka. No nije dao nikakvo intuitivno poimanje istine koje bi odgovaralo ovoj fiksnoj točki. Upravo je to glavni predmet ovog rada. *Dana je jedna logička analiza pojma istine koja vodi najvećoj intrinsičnoj fiksnoj točki jake Kleeneove trovaljane semantike.* U njenoj osnovi je sljedeća intuicija. Da bismo utvrdili je li data rečenica istinita ili lažna moramo ispitati što ona govori. Ako govori o vanjezičnoj realnosti kao npr. prva rečenica Uvoda tada treba ispitati tu realnost i utvrditi je li istinita ili lažna. Ako govori o istinitosti druge rečenice, kao npr rečenica (1), tada treba ispitati istinitost te druge rečenice. Ako je pak rečenica složena od jednostavnijih rečenica kao npr. rečenica koja je disjunkcija rečenica (1) i (2)

Prva rečenica Uvoda je istinita ili sedma rečenica Uvoda je lažna.

tada se ispitivanje njene istinitosti svodi na ispitivanje istinitosti rečenica (1) i (2). Tako, koristeći klasične uvjete istinitosti rečenica, da bismo odredili istinitost date rečenice moramo ispitati istinitost svih rečenica o kojima njena istinitost ovisi, zatim eventualno, iz istih razloga, i istinitost rečenica o kojima istinitost tih rečenica ovisi, itd. Svaki takav put po rečenicama vodi prema jednostavnim rečenicama tipa subjekt-predikat. U običnim situacijama jezik ne govori o istinitosnim vrijednostima vlastitih rečenica, tako da istinitost jednostavnih rečenica ne ovisi o istinitosti drugih rečenica. Za ispitati njihovu istinitost moramo istražiti vanjsku realnost o kojoj one govore. Utvrđivanjem njihove istinitosti proces determiniranja istinitosti početne rečenice je završen. Dobivamo definitivan odgovor je li ona istinita ili lažna. No ova klasična situacija može biti narušena (i jeste) ako jednostavne rečenice govore o istinitosti drugih rečenica. Tada se proces ispitivanja istinitosti nastavlja. Putevi vode ponovo prema složenijim rečenicama. Zbog mogućih kruženja ništa više ne garantira uspjeh procedure. Paradoksi istine upravo su svjedoci takvih situacija. *Paradoksi istine proizlaze iz situacija kada klasična procedura određenja istine ne daje klasično pretpostavljeni odgovor. Oni pokazuju da je klasična pretpostavka o uspješnom završetku procedure neopravdano poopćena s uobičajenih situacija na sve situacije. Klasičnu proceduru određenja istine možemo sačuvati, kao i internu semantičku strukturu jezika, ali moramo odbaciti univerzalnost pretpostavke njenog uspjeha. Moramo prihvatiti da neke sintaktički i semantički ispravne rečenice nemaju klasičnu istinitosnu vrijednost. Njima možemo pridružiti treću vrijednost, neodređeno, kao oznaku definitivnog neuspjeha klasične procedure.* Analiza propagiranja tog neuspjeha u strukturi rečenica daje upravu jaku Kleeneovu trovaljanu semantiku, ali ne kao investigativnu proceduru, kakvom se javlja kod Kripkea, već kao klasičnu proceduru određenja istinitosti nadopunjenu propagacijom vlastitog neuspjeha. Analiza cirkularnosti u postupku određivanja klasične istinitosne vrijednosti rečenice daje kriterij kad će postupak uspjeti a kad ne, kada će rečenice imati klasičnu istinitosnu vrijednost a kad ne. *Pokaže se da je tako dobijen skup istinitih i lažnih rečenica upravo najveća intrinsična fiksna točka jake Kleeneove trovaljane semantike. Time je dana argumentacija za izbor između svih fiksnih točaka svih monotonih semantika upravo te točke za model logičkog pojma istine, a ujedno je dan i njen neposredni matematički opis.* No ovaj jezik je trovaljan, dakle i manje udoban od dvovaljanog jezika na koji smo naviknuti svakodnevnim razmišljanjem. Također neodređene rečenice su područje šutnje jezika, dok u metajeziku sasvim jednostavno možemo zaključivati o njima. Npr. rečenica (2) je neodređena u tom jeziku. Iako to jednostavno možemo izreći u metajeziku (što je upravo urađeno), ne možemo izreći u samom jeziku. Ne samo da je ovo područje šutnje jezika nezadovoljavajuće (vodi trovaljanoj logici i slabi izražajnu moć jezika), već ono može biti prirodno prekinuto dodatnom valuacijom takvih rečenica koja proizlazi upravo iz prepoznavanja neuspjeha klasične procedure na njima. Ta dodatna valuacija s jedne strane nadopunjuje primarnu semantiku a s druge strane upravo nju ima za predmet - ona je njen metajezik. Pogledajmo npr. rečenicu (2). Klasična procedura određenja istinitosti na njoj propada. Dakle rečenica (2) je neodređena. To ne možemo izreći u primarnoj semantici (jeziku) ali možemo izreći u nadopunjenoj dvovaljanoj semantici (metajeziku). Štoviše u toj novoj semantici iz toga možemo zaključiti da nije lažna u trovaljanoj semantici, suprotno od onoga što govori. Dakle ona je lažna u nadopunjenoj dvovaljanoj semantici. *Ovim se ne obnavlja paradoks* *jer je napravljen semantički pomak od istinitosti u trovaljanoj semantici o kojoj (2) govori do istinitosti u dvovaljanoj semantici kojom je (2) na kraju vrednovana. Tako refleksivnost razmišljanja daje prirodnu nadopunu trovaljane semantike do dvovaljane klasične semantike koja s jedne strane govori o trovaljanoj semantici a s druge je nadopunjuje. Taj jezik,* **klasični zatvarač najveće intrinsične fiksne točke Jake Kleeneove trovaljane semantike***, krajnji je produkt ove logičke analize pojma istine.* Ona je razrađena u četvrtom poglavlju .

Zahvaljujem se voditelju rada prof. dr. Zvonimiru Šikiću, te prof. dr. Deanu Rosenzwiegu i docentu dr. Mladenu Vukoviću, kao i ostalim članovima Seminara za matematičku logiku i osnove matematike u okviru kojeg je ovaj rad dobio svoj konačan oblik.

POGLAVLJE 1

Logička analiza pojma istine i paradoksi

1.1 Logička analiza pojma istine

Pojam istine ima razne aspekte i čest je predmet filozofskih rasprava. Cilj je ovog odjeljka objasniti kojim aspektom tog pojma se ovaj rad bavi. Ovaj rad na žalost ne daje nikakvu ontološku teoriju koja bi nam rekla što je istina a što nije. On neće dati odgovor na pitanje je li bilo ičeg prije Big Banga ili jesu li se Tuđman i Milošević dogovorili u Karađorđevu o podjeli Bosne i Hercegovine. On također ne daje nikakvu epistemološku teoriju na koji način je istinitosna vrijednost rečenica povezana sa svijetom. Tim pitanjem se više manje neuspješno bave razne filozofske teorije, npr. korespondencijska teorija istine, koherencijska teorija istine, instrumentalistička teorija istine i pragmatistička teorija istine. To su primjeri tzv. inflacijskih teorija istine koje nastoje opisati vezu istine i realnosti. Nasuprot njima su tzv. deflacijske teorije istine koje minimiziraju ulogu pojma istine smatrajući da je u pitanju više ili manje suvišan način izražavanja. To su npr. teorija redundantosti istine i minimalistička teorija istine. Zainteresirani čitatelj može pogledati npr. [Kir92] da stekne uvid o kakvim je diskusijama tu riječ. *Ono što je zajedničko svim tim filozofskim teorijama istine je da one žele analizirati pojam istine unutar šireg konteksta - postoji li uopće veza istine sa svijetom i ako postoji kakva je.* U tom kontekstu se često spominju Tarskijeva i Kripkeova analiza pojma istine. Pogotovo je Tarski često kritiziran, ili sa stajališta da njegova teorija ne rješava filozofski problem istine ili, suprotno tome, da ga rješava na filozofski jednostran način [1]. Mislim da su te kritike u osnovi promašene jer brkaju dva razna aspekta pojma istine.

Aspekt pojma istine kojim se bavi ovaj rad, a po mom mišljenju i Tarskijeva i Kripkeova analiza, nije odnos istine i realnosti nego međusobna povezanost istinitosnih vrijednosti rečenica koje mogu i same sadržavati pojam istine. Svaki jezik pored svoje formalne (gramatičke) strukture i unutarnje značenjske strukture ima također i svoju vanjsku značenjsku strukturu, vezu između jezičnih formi i vanjskog svijeta. Ova vanjska značenjska struktura počiva na određenim vanjskim pretpostavkama o upotrebi jezika. Za klasični jezik, recimo za jezik logike prvog reda, to su pretpostavke da postoje predmeti o kojima jezik govori, da je svako ime ime nekog predmeta, da je svakom funkcijskom ili relacijskom simbolu pridružena odgovarajuća funkcija odnosno relacija među objektima, te da je svaka atomarna rečenica istinita ili lažna. Ova analiza ne problematizira vanjsku značenjsku strukturu jezika. Ne bavi se pitanjem ispunjivosti pretpostavki na kojima je ta struktura zasnovana. Ne analizira na koji način se atomarnim rečenicama koje govore o vanjskom svijetu pridružuje istinitosna vrijednost niti kakva je veza te vrijednosti s realnošću. *Za razliku od navedenih filozofskih teorija koje se upravo time bave, ova analiza polazi od vanjske značenjske strukture jezika kao unaprijed zadane i bavi se unutrašnjom značenjskom strukturom jezika, međusobnom povezanošću rečenica, prije svega pitanjem povezanosti rečenica koje govore o istinitosti drugih rečenica i rečenica o čijoj istinitosti govore.* Dakle ona je *logičke prirode*, i zato će biti i nazvana **logičkom analizom problema istine**. Tako ona ima vrlo malo zajedničkog s navedenim filozofskim teorijama.

Analiza će se provoditi na nivou jezika, u smislu da će se baviti *istinitošću rečenica* a ne istinitošću propozicija, koje su po nekim zamislima pravi subjekti mišljenja dok rečenice tek izražavaju propozicije. Ovakvo određenje ima nesumnjivu tehničku prednost jer su predmet proučavanja konkretne jezične forme, a ne apstraktni objekti nejasne prirode. Ono je ujedno odraz mog dubokog uvjerenja da jezik nije tek sredstvo zapisivanja i komuniciranja misli (propozicija) već bitan dio razmišljanja, te da razmišljanje u svojoj apstraktnoj formi nije ništa drugo do tvorba i upotreba jezika.

Radi određenosti, formalni dio analize će biti proveden nad interpretiranim jezicima prvog reda. Pod **interpretiranim jezikom prvog reda** podrazumijeva se formalni jezik logike prvog reda zajedno sa svojim klasičnim

[1] Tarski se u polemičkom dijelu članka [Tar44] ironično izvinjava što ne može odgovoriti na tu kritiku jer mu još nitko nije objasnio šta je to filozofski problem istine.

modelom. Takvim jezicima će biti dodan predikat istine kojeg ćemo obično bilježiti "T" i čije će značenje biti posebno dodano jeziku. Pretpostavljeno znanje logike prvog reda može se naći u svakom elementarnom udžbeniku matematičke logike.

1.2 Paradoks Lašca

Najstariji poznati paradoks istine je tzv. **Lažac**. On se pripisuje sofistu Eubulidu iz Mileta (4. stoljeće prije Krista) koji je rekao: "Neki čovjek kaže da laže. Je li to što on govori istina ili laž?". U nešto direktnijoj formi Lažac glasi:

Ova rečenica je lažna.

Ako pretpostavimo da je ova rečenica istinita tada je istina šta ona govori, da je lažna. Ako pak pretpostavimo da je lažna tada je laž šta govori — da je lažna. Dakle istinita je. Tako je ona kontradiktorna samoj sebi.

Problem nije u upotrebi kontekstno ovisne pokazne zamjenice "ova" jer samoreferiranje možemo postići i na drugi način tzv **baptizmom**. Naprosto se dogovorimo da oznaka "L" imenuje rečenicu "Rečenica L je lažna."[2]:

L: Rečenica L je lažna.

Samoreferiranje možemo postići i na sintaktičkom nivou, pomoću sintaktičke operacije **dijagonalizacije**. To je operacija koja predikatu $P(x)$ pridružuje njegovu dijagonalizaciju, rečenicu koja nastaje primjenom predikata na svoje vlastito ime "$P(x)$":

$P($"$P(x)$"$)$.

Npr dijagonalizacija predikata

Kaspar čita x

je rečenica

Kaspar čita "Kaspar čita x",

koja govori da Kaspar čita prethodnu rečenicu. No rečenica koja je dijagonalizacija predikata

Kaspar čita dijagonalizaciju od x

je rečenica

[2] Kao što Visser reče [Vis89]: "Koji Semantički Bog bi nas mogao u tome zaustaviti?"

Kaspar čita dijagonalizaciju od “ Kaspar čita dijagonalizaciju od x”

a koja govori da Kaspar čita upravo nju. [3]

Dijagonalizacija daje uniforman način samoreferiranja kojim možemo za svako svojstvo P dobiti rečenicu koja govori da za nju vrijedi P. Specijalno možemo dobiti Lašca dijagonalizacijom predikata:

Dijagonalizacija predikata x je lažna rečenica

Tako Lažac glasi

Dijagonalizacija predikata “Dijagonalizacija predikata x je lažna rečenica” je lažna rečenica.

Postupak dijagonalizacije možemo intuitivnije razumjeti na sljedeći način. Neka smo npr. leksikografskim redoslijedom numerirali sve predikate jezika tako da oni tvore niz $P_i(x), i \in N$. Tada rečenice koje nastaju supstitucijom imena predikata na mjesto x možemo rasporediti u sljedeću beskonačnu tablicu:

	$P_1(x)$	$P_2(x)$	$P_3(x)$	
$P_1(x)$	$P_1("P_1(x)")$	$P_1("P_2(x)")$	$P_1("P_3(x)")$	.....
$P_2(x)$	$P_2("P_1(x)")$	$P_2("P_2(x)")$	$P_2("P_3(x)")$	.....
$P_3(x)$	$P_3("P_1(x)")$	$P_3("P_2(x)")$	$P_3("P_3(x)")$	.....
..	....	....	....	.....
..				

Na dijagonali se nalaze upravo dijagonalizacije predikata. Nazovimo $d(x)$ dijagonalizaciju predikata x. Tada za svaki predikat $A(x)$ možemo gledati njegovu ispunjivost na dijagonali. Time dobijemo novi predikat $A(d(x))$. No i on se nalazi među numeriranim predikatima, tj. postoji neki δ takav da je $A(d(x)) = P_\delta(x)$ Pogledamo li rečenice u tom retku, vidimo da svaka od njih govori da za rečenicu na dijagonali povrh nje vrijedi A:

[3] Ovaj sintaktički mehanizam samoreferiranja putem dijagonalizacije dugujemo Gödelu.

	$P_1(x)$	$P_2(x)$	$P_3(x)$	
$P_1(x)$	$P_1("P_1(x)")$	$P_1("P_2(x)")$	$P_1("P_3(x)")$	.....
$P_2(x)$	$P_2("P_1(x)")$	$P_2("P_2(x)")$	$P_2("P_3(x)")$	.....
$P_3(x)$	$P_3("P_1(x)")$	$P_3("P_2(x)")$	$P_3("P_3(x)")$	.....
.....	....	....	....	.....
$P_\delta(x)$	$P_\delta("P_1(x)")$	$P_\delta("P_2(x)")$	$P_\delta("P_3(x)")$	$P_\delta("P_\delta(x)")$

Tako je rečenica koja se nalazi u tom retku na dijagonali samoreferirajuća: govori da ima svojstvo A:

$$P_\delta(\text{“}P_\delta(x)\text{”}) \;\leftrightarrow\; A(\text{“}P_\delta(\text{“}P_\delta(x)\text{””})$$

No Lažac se može postići i na **empirički** način — pozivanjem na realnost opiše se rečenica za koju se tvrdi da je lažna i onda se ispitivanjem te realnosti utvrdi da je to upravo rečenica kojom se to tvrdi. Takva je konstrukcija iz Uvoda:

(2) Sedma rečenica Uvoda je lažna.

Empiričko ispitivanje Uvoda je pokazalo da je to upravo ta ista rečenica. Dakla ona o sebi tvrdi da je lažna. Kripke u [Kri75] takvu konstrukciju opisuje kao konstrukciju rečenice $\forall x\ (P(x) \rightarrow Q(x))$, pri čemu je $P(x)$ empirički predikat za kojeg se pokaže da je jedino ispunjiv na samoj toj rečenici. Tako je ta rečenica samoreferirajuća — o sebi kaže da za nju vrijedi $Q(x)$. Već je u Uvodu naglašena važnost takvog Lašca za Kripkeovu analizu pojma istine. Za razliku od baptističkog i dijagonalnog Lašca koji su uvijek paradoksalni, empirički konstruiran Lažac ne mora biti paradoksalan, već to ovisi o realnosti. Da je kojim slučajem, npr. prisustvom buga u editoru teksta, iz Uvoda izbrisana rečenica (1) tada rečenica (2) ne bi bila paradoksalna. Ne bi se više odnosila na sebe već na neku drugu rečenicu. Zato se takav tip Lašca zove **Kontingentni Lažac**.

Zanimljivo je da je jedna od najpoznatijih starinskih verzija Lašca upravo tog tipa. U Bibliji, u Evanđelju po Pavlu, u poslanici Titu piše:

> Reče jedan od njih, njihov vlastiti prorok: “Krećani su uvijek lažljivi, zle živine, besposleni trbusi”. Ovo svjedočanstvo je istinito.

No to svjedočanstvo ne može biti istinito, jer bi to značilo da svi Krećani lažu pa i Krećanin koji je ovo izrekao. Dakle to svjedočanstvo je lažno. Za sada ovo i nije neki paradoks osim za one koji vjeruju da sveci uvijek govore istinu, odnosno da je sve što je rečeno u Bibliji istina. Iz toga da je ovo svjedočanstvo lažno slijedi da barem neki Krećani ponekad govore istinu ili pak nisu zle živine ili besposleni trbusi. I to je vjerojatno istina koju više ne možemo provjeriti. No moguće je zamisliti situaciju koja proroku vraća vjerodostojnost — da su svi Krećani zaista bili zle živine i besposleni trbusi i da su sve njihove rečenice koje su bile izrečene prije prethodno citirane bile lažne. Međutim, kad bi i ovo svjedočanstvo bilo lažno, to bi po navedenim uvjetima značilo da bi moralo biti istinito. Ono je dakle Kontingentni Lažac koji jako dobro ilustrira Kripkeova razmatranja. Ovaj paradoks je nazvan **Epimenidov paradoks** jer je grčki filozof Epimenid, po nekim autorima, bio Krećanin koji se u paradoksu spominje.

Već je u Uvodu spomenuto moguće rješenje paradoksa Lašca — da ta rečenica nije ni istinita ni lažna već neodređena. Varijanta paradoksa koja naizgled negira tu mogućnost je **Jaki Lažac**:

Ova rečenica nije istinita.

Naime, dok iz Lašca slijedi da je on istinit ako i samo ako je lažan, pa se kontradikcija može izbjeći uvođenjem vrijednosti neodređeno, iz ovog Lašca slijedi da je on istinit ako i samo ako nije istinit, pa se čini da se ne može pribjeći trećoj vrijednosti, jer to spada pod slučaj da nije istinita. Burge [Bur79] daje i sljedeću verziju koja će ovdje biti nazvana **Trovaljani Lažac**

Ova rečenica je lažna ili neodređena.

Sljedeća verzija se suprotstavlja rješenju po kojem je Lažac besmislen. Ona će ovdje biti nazvana **Besmisleni Lažac**:

Ova rečenica je lažna ili besmislena.

Kad bi ova rečenica bila lažna dobili bismo da je istinita i smislena. Kad bi ova rečenica bila istinita to bi značilo da je lažna ili besmislena. Ali po pretpostavci nije lažna, a ne može biti ni besmislena jer kako bi tada bila istinita. Kad bi ova rečenica bila besmislena tada je istina šta ona govori pa nije besmislena. Tako sve mogućnosti vode kontradikciji.

Za rješenje paradoksa Lašca predloženo je i rješenje u okviru zamisli da rečenice samo izražavaju propozicije i da su propozicije istinite ili lažne. Po tome je rješenje da je Lažac u stvari krnja rečenica koja treba glasiti

Propozicija izražena ovom rečenicom je lažna.

Pretpostavka istinitosti propozicije odmah daje kontradikciju. Ali pretpostavka da je lažna ne daje kontradikciju već zaključak da ta rečenica ne izražava propoziciju. No sljedeća verzija, tzv. **Propozicijski Lažac** ([GB93]) suprotstavlja se takvom rješenju:

Ne postoji istinita propozicija koju ova rečenica izražava

Razmišljanje analogno predhodnom vodi zaključku da ne postoji istinita propozicija koju ova rečenica izražava. Pa ispada da je ona istinita. Ali kako može biti istinita kad ne izražava propoziciju?

Sljedeća verzija, **Metalažac**, ([Gai92]) suprotstavlja se rješenju Lašca po kojem problem treba prenijeti u metajezik:

1. Rečenica na liniji 1 nije istinita.
2. Rečenica na liniji 1 nije istinita.

Rečenica na liniji 1 je Lažac pa je neodređena. Ako drugu rečenicu shvatimo kao refleksiju nad prvom rečenicom za koju smo utvrdili da je neodređena, tada je ona istinita. Tako ispada da je jedna te ista rečenica i neodređena i istinita.

Sljedeća verzija, slična prethodnoj, tzv. **Intenzionalni Lažac** ([Sky84]) ukazuje na eventualni intenzionalni karakter Lašca. Naime ako u rečenici

(1) Rečenica (1) nije istinita.

zamijenimo znak (1) detaljnim opisom rečenice koju taj znak imenuje dobijemo rečenicu

Rečenica "Rečenica (1) nije istinita" nije istinita.

Dok je rečenica (1) neodređena, čini se da nam je ova bezazlena zamjena dala rečenicu koja nije neodređena nego istinita, jer rečenica o kojoj ona govori je neodređena, dakle nije istinita.

1.3 Ostali paradoksi istine

Paradoks Lašca, u raznim varijantama, je tek jedan tip paradoksa istine, mada i najpoznatiji. Sasvim drugi tip paradoksa istine predstavlja sljedeća rečenica, **Istinoljubac**:

Ova rečenica je istinita.

Za razliku od Lašca kojem ne možemo pridružiti ni istinu ni laž, ovoj rečenici na jednako (ne)uvjerljiv način možemo pridružiti i istinu i laž. Visser [Vis89] ovakve rečenice zove **bikonzistentnim**. Nema nekih dodatnih određenja koja bi uradili izbor između dviju mogućnosti, ne zbog našeg neznanja nego "u principu". Tako ni ovoj rečenici ne možemo odrediti istinitost. Mada ovdje nema kontradikcije, prisutna je ista paradoksalnost — nemogućnost da rečenici odredimo istinitosnu vrijednost.

Sljedeći paradoks, **Curryev paradoks** pokazuje da paradoksalnost nije vezana uz negaciju. On se obično formulira u dramatičnoj varijanti kao dokaz postojanja Boga:

Ako je ova rečenica istinita tada Bog postoji.

Kad bi ova rečenica bila lažna njen antecedent bi bio lažan pa bi ona po tablici istinitosti kondicionala bila istinita. Dakle ona mora biti istinita. Ali tada joj je istinit antecedent, pa mora biti istinit i konzekvent, tj istina je da Bog postoji. Naravno, problem je u tome da bismo na isti način mogli dokazati i da Bog ne postoji. Kontradiktornost ovakve rečenice se najbolje može vidjeti ako se za konzekvent stavi logička laž , npr. da Bog postoji i Bog ne postoji. Tada ponavljanjem prethodnog razmišljanja direktno dobijemo kontradikciju, da je logička laž istinita.

Slični tipovi paradoksa javljaju se i kod ispitivanja istinitosti skupova rečenica. Prvi primjer reproducira Lašca indirektnim referiranjem:

(1) Rečenica (2) je istinita.

(2) Rečenica (1) je lažna.

Lako je utvrditi da pretpostavka lažnosti kao i pretpostavka istinitosti bilo koje od ovih rečenica vodi kontradikciji. Drugi primjer je **Buridanov dokaz** postojanja Boga:

(1) Bog postoji.

(2) Rečenice (1) i (2) su lažne.

Rečenica (2) ne može biti istinita jer bi to značilo, po onom što govori, da je lažna. Dakle ona je lažna. Po onom što govori to znači da je jedna od rečenica (1) i (2) istinita. Kako je (2) lažna to mora (1) biti istinita. Dakle, Bog postoji. Naravno problem je isti kao i kod Curryeva paradoksa — na isti način bismo mogli dokazati i da Bog ne postoji.

Sljedeći primjer, **Yabloov paradoks** ([Yab93]), pokazuje da paradoks istine možemo dobiti i bez samoreferiranja, direktnog ili indirektnog. Promotrimo sljedeći beskonačan skup rečenica $(i), i \in N$:

(i) Za svaki $k > i$ (k) nije istinita rečenica.

Kad bi neka rečenica (i) bila istinita, tada sve sljedeće rečenice ne bi bile istinite. No to bi značilo s jedne strane i da (i+1) nije istinita a s druge strane, pošto i sve nakon nje nisu istinite, da je (i+1) istinita. Dakle sve navedene rečenice nisu istinite. Ali, ako pogledamo šta one tvrde, to za sobom povlači da su sve one istinite.

Kao što je u Uvodu rečeno *ovi paradoksi istine nisu "zavrzlame za razbibrigu" već su "simptomi bolesti". Oni ukazuju na postojanje problema i na njima će biti testirana sva rješenja opisana u ovom radu.* Pri tome treba razlikovati *normativni* od *analitičkog* aspekta rješenja. Prvi nastoji osigurati da se paradoksi neće pojaviti, dok ih drugi nastoji objasniti. Naravno važan je i *praktični* aspekt rješenja koji nastoji osigurati što bolji okvir za logičko zasnivanje znanja, umjetnu inteligenciju kao i analizu prirodnog jezika.

POGLAVLJE 2

Tarskijeva analiza pojma istine

2.1 Tarskijeva semantička definicija pojma istine

U članku "Pojecie prawdy w jezykach dedukcyjnych" [Tar33] Alfred Tarski je izložio svoju analizu pojma istine.[1] Članak se odlikuje izuzetnom preciznošću. Stoga, usprkos značaju teme, čudi enormno mnoštvo rasprava u kojima se tumači šta je Tarski mislio i šta je uradio u tom članku.

Tarski započinje članak jasnim opisom cilja:

> ... konstruirati — za zadani jezik — *sadržajno adekvatnu i formalno korektnu definiciju termina "istinita rečenica"*.

U [Tar44] Tarski objašnjava zašto je cilj definicija pojma istine — definicija objašnjava značenje tog pojma pomoću pojmova kojim ga definiramo i osigurava da pojam neće voditi kontradikciji, ako pojmovi koji ga definiraju ne vode kontradikciji. Što je formalno korektna definicija istine u logici nije sporno. Pored pravila definiranja ona podrazumijeva i jezik u kojem će definicija biti izražena. Što se sadržajne adekvatnosti tiče Tarskijev cilj je **semantička definicija istine**, tj. definicija koja pitanje istinitosti rečenice svodi na pitanje istinitosti onoga što ona govori. To je upravo ona svakodnevna intuicija koja je korištena i u Uvodu. Da bismo utvrdili je li prva

[1] Članak je prezentirao J. Lukasiewicz 21.3.1931 pred Varšavskim znanstvenim društvom, a publiciran je 1933. godine. Ista analiza je na popularan način izložena u [Tar44], gdje je popraćena i polemičkim dijelom u kojem Tarski odgovara na kritike.

rečenica Uvoda istinita, morali smo vidjeti što ona govori, tj. koje je njeno značenje. Da bismo utvrdili je li istinita rečenica "Pada snijeg" moramo ispitati pada li snijeg. Tako za svaku rečenicu imamo odgovarajuću parcijalnu definiciju istine. Npr. za rečenicu "Pada snijeg" imamo sljedeću parcijalnu definiciju istine:

"Pada snijeg" je istinita rečenica ako i samo ako pada snijeg.

Sadržajno adekvatna definicija pojma istine je po Tarskom definicija iz koje za svaku rečenicu jezika za koji definiramo pojam istine slijedi odgovarajuća parcijalna definicija. Preciznije, neka u jeziku ML u kojem definiramo pojam istine jezika L, za svaku rečenicu S jezika L postoji njeno ime $\overline{S}$ i prijevod u metajezik S^*, koji može biti i sama rečenica S . Tada iz definicije istine jezika L u metajeziku ML moraju slijediti sve tvrdnje oblika

$\overline{S}$ je istinita rečenica ako i samo ako S^*

To je tzv **shema T**, a navedeni uvjet sadržajne adekvatnosti semantičke definicije pojma istine Tarski je nazvao **konvencija T**.

Ovaj uvjet sadržajne adekvatnosti definicije pojma istine čest je argument tvrđenju da je Tarskijeva semantička koncepcija istine određeni oblik teorije korespondencije koja istinitost smatra podudaranje (u nekom smislu) s realnošću. Sam Tarski kaže da je shema T formulacija intuicije o istini koja korijen vuče još iz Aristotelove koncepcije istine koju Aristotel opisuje u svojoj Metafizici sljedećim riječima:

> Reći za ono što jeste da nije, ili za ono što nije da jeste, je laž, dok reći za ono što jeste da jeste, ili za ono što nije da nije, je istina.

No Tarski u polemičkom dijelu članka [Tar44], osvrćući se na takve kritike eksplicitno kaže da semantička definicija istine ništa ne implicira o realnosti, tj. je li snijeg pada ili ne, već samo implicira da su uvjeti istinitosti rečenice

"Pada snijeg" je istinita rečenica

jednaki uvjetima istinitosti rečenice

Pada snijeg

Dakle, *Tarskijeva analiza pojma istine se ne bavi odnosom istine i realnosti nego međusobnim odnosom istinitosnih vrijednosti rečenica. Ona je logičke prirode.*

U članku [Tar33] Tarski se prvo bavi mogućnošću definiranja pojma istine kolokvijalnog jezika u samom tom jeziku. Bez obzira na nepreciznu formu kolokvijalnog jezika i mogući oblik definicije, ona mora povlačiti shemu T. No, u prirodnom jeziku je lako na empirički način konstruirati Lašca, rečenicu L takvu da je

$L =$"L nije istinita rečenica" (1)

Shema T primijenjena na ovu rečenicu daje

"L nije istinita rečenica" je istinita rečenica ako i samo ako L nije istinita rečenica

Obična supstitucija na osnovi jednakosti (1) daje kontradikciju

L je istinita rečenica ako i samo ako L nije istinita rečenica

Tarski se u članku bavi i drugim problemima koje definicija istine kolokvijalnog jezika donosi kao i mogućim rješenjima. No prethodni paradoks Lašca mu je glavni argument za sljedeći zaključak:

> *...sama mogućnost konsistentne upotrebe izraza "istinita rečenica" koja je u harmoniji sa zakonima logike i duhom svakodnevnog jezika čini se veoma upitnom, a samim tim iste sumnje se odnose i na mogućnost konstrukcije korektne definicije ovog izraza.*

Nakon toga Tarski prelazi na mogućnost definicije pojma istine za jezike deduktivnih nauka. To su kod Tarskoga interpretirani jezici s preciznom formom. Za takav jezik L Tarski prvo formulira odgovarajući metajezik ML. Za formulaciju sheme T jezik ML pored standardnog logičkog vokabulara mora imati vokabular koji mu omogućuje sintaktički opis rečenica jezika L (da bi za svaku rečenicu S jezika L imao njeno ime $\overline{S}$), kao i vokabular koji ima isto značenje kao i vokabular jezika L (da bi za svaku rečenicu S jezika L postojao njen prijevod S^* u jezik ML). Tarskijeva analiza mogućnosti definicije istine urađena je za jezike logika višeg reda. U tom okviru je pokazano kada je definicija moguća a kada ne. Mogućnost je pokazana danas već klasičnom *rekurzivnom definicijom pojma ispunjivosti formule na zadanim objektima, te definicijom istinitosti rečenice pomoću tako definiranog pojma ispunjivosti*. Suvremeni oblik te definicije je definicija istinitosti rečenice jezika L u datom modelu a koja je moguća u jeziku teorije skupova uvijek kada je domena modela skup.[2] On za te definicije pokazuje da su formalno korektne i da su semantičke definicije, tj. ispunjavaju konvenciju T. Nemogućnost takve definicije Tarski pokazuje rekonstrukcijom paradoksa Lašca. Naime postojanje definicije istine jezika L u jeziku ML značilo bi

[2] ta definicija je prvi put dana u [TA56])

da je u ML istinit svaki primjerak sheme T. Koristeći Gödelov dijagonalni metod Tarski pokazuje kako za neke L i ML možemo opisati rečenicu jezika L koja o sebi govori da nije istinita. To je sadržaj njegovog teorema nedefinabilnosti istine koji će detaljno biti razmotren u sljedećem odjeljku.

Tako je Tarski precizno opisao šta je to semantička definicija istine. Izolirao je shemu T kao kriterij sadržajne adekvatnosti takve definicije. Pokazao je da za dati jezik ona mora biti provedena u odgovarajućem metajeziku. Konstrukcijom definicije istine za neke jezike u odgovarajućem metajeziku pokazao je mogućnost takve definicije. U suvremenom obliku najznačajnija je definicija istine za jezik s datim modelom kojem je domena skup, a koja je provediva u jeziku teorije skupova. *Teoremom nedefinabilnosti pokazao je kada ta definicija nije moguća.* Najznačajnija moderna varijanta je nemogućnost definicije istine jezika aritmetike ili pak jezika teorije skupova (općenito svakog "dovoljno bogatog" jezika) u samom tom jeziku.

2.2 Nedefinabilnost istine

U ovom odjeljku je formuliran problem definabilnosti istine jezika logike prvog reda L u jeziku logike prvog reda ML i pokazano je da je u slučaju $ML = L$ za dovoljno bogate jezike odgovor negativan — pojam istinite rečenice jezika L nije definabilan u samom jeziku L. Izlaganje se uglavnom zasniva na [GB93] i [Smu92].

Tarski je postavio sljedeće uvjete na metajezik ML:

1. ML mora sadržavati ime $\overline{S}$ za svaku rečenicu S jezika L
2. za svaku rečenicu S jezika L mora postojati njen prijevod S^* u jezik ML

To mu je bilo nužno da bi se u ML mogli izreći svi primjerci sheme T koji tvore uvjet sadržajne adekvatnosti tražene definicije istine:

$\overline{S}$ je istinita rečenica ako i samo ako S^*

No ti uvjeti se mogu oslabiti. Drugi uvjet se može posve izostaviti jer će shema T biti izrečena u metametajeziku, neformalnom matematičkom jeziku na kojem se sve ovo opisuje. Što se tiče prvog uvjeta dovoljno će biti pretpostaviti da ML može na neki način pričati o rečenicama jezika L. Zato ćemo pretpostaviti da njegova domena sadrži kodove rečenica iz L.

Definicija 21. *Neka je* $<>$ *kodiranje (injektivna funkcija) rečenica jezika* L *u domenu jezika* ML*. Kažemo da predikat* $T(x)$ *(formula s jednom slobodnom varijablom* x*)* **definira pojam istinite rečenice jezika** L *preko kodiranja* $<>$ *tj. da je* **predikat istine** *jezika* L *ako i samo ako za svaku valuaciju* v *varijabli jezika* ML *vrijedi:*

$ML \models T(x)[v]$ *ako i samo ako postoji rečenica* $S \in L$ *takva da je* $v(x) =< S >$ *i* $L \models S$

Tada kažemo da je **pojam istinite rečenice u** L **definabilan u** ML*.*

Uvjet iz definicije je upravo Tarskijeva shema T, sada izrečena kao uvjet u neformalnom jeziku. Taj uvjet pitanje ispunjivosti predikata $T(x)$ u jeziku ML i valuaciji v svodi na pitanje istinitosti rečenice $v(x)$ u jeziku L. Samo kodiranje može biti i identitet, tj. može biti i da je

$< S >= S$

Tada jezik ML može direktno govoriti o rečenicama jezika L. Ako pak ML ima za svaku rečenicu S ime $\overline{S}$ za njen kod $< S >$ tada shemu T možemo jednostavnije izreći:

$ML \models T(\overline{S})$ ako i samo ako $L \models S$, za svaku rečenicu $S \in L$

(i lažan je na objektima koji nisu kodovi rečenica)

Ako postoji prevođenje * rečenica iz L u rečenice iz ML, tj. injektivna funkcija $^* : L \longrightarrow ML$ takva da je

$L \models S$ ako i samo ako $ML \models S^*$

tada shemu T možemo izreći i u samom metajeziku ML:

$ML \models T(\overline{S}) \leftrightarrow S^*$, za svaku rečenicu $S \in L$

Ako je $L \subseteq ML$ tada možemo izbjeći prevođenje:

$ML \models T(\overline{S}) \leftrightarrow S$, za svaku rečenicu $S \in L$

No bez obzira o kojem obliku sheme T *je riječ njena uloga je uvijek ista i odgovara Tarskijevoj semantičkoj koncepciji — ona problem istinitosti rečenice* $T(\overline{S})$ *u jednom jeziku svodi na problem istinitosti rečenice* S *u drugom jeziku.* U ovom radu je najzanimljiviji slučaj kada je $ML = L$, tj. problem definabilnosti istine jezika u samom jeziku :

$L \models T(\overline{S}) \leftrightarrow S$, za svaku rečenicu $S \in L$

Strategija dokaza nedefinabilnosti istine jezika u samom jeziku svodi se na konstrukciju Lašca u tom jeziku na osnovu pretpostavke definabilnosti istine. Ako se može konstruirati Lažac [3] tada se može rekonstuirati paradoks Lašca, tj. dobiti kontradikcija iz koje slijedi nemogućnost definabilnosti istine. Rekonstrukcija kontingentnog Lašca zahtjeva jezik s *empiričkim* predikatima. Naprosto bismo dio prirodnog jezika u kojem je bilo provedeno razmatranje neke verzije kontingentnog Lašca trebali rekonstruirati kao jezik prvog reda s odgovarajućim empiričkim predikatima. No najjednostavnija konstrukcija Lašca je putem *baptizma* — odgovarajućom interpretacijom konstanti jezika.

Lema 21. *Neka domena jezika L sadrži za svaku rečenicu S jezika njen kod $< S >$ i neka je $A(x)$ formula s jednom slobodnom varijablom x. Ako postoji zatvoreni term t jezika takav da imenuje kod rečenice $\neg A(t)$, tj.*

$t^L =< \neg A(t) >,$

tada $A(x)$ nije predikat istine jezika L.

Dokaz : Kada bi $A(x)$ bio predikat istine tada bi rečenica $\neg A(t)$ bila Lažac u jeziku L i vodila bi kontradikciji. Naime po definiciji predikata istine je

$L \models A(t)$ ako i samo ako je t ime koda istinite rečenice

a po interpretaciji terma t

t je ime koda istinite rečenice ako i samo ako je $L \models \neg A(t)$

iz čega slijedi kontradikcija

$L \models A(t) \leftrightarrow \neg A(t)$

Dakle $A(x)$ nije predikat istine jezika L. ■

Nedefinabilnost istine sada je lako dobiti — samo trebamo za svaku formulu $A(x)$ imati konstantu c_A koja će imenovati $< \neg A(c_A) >$. To je **baptistička verzija teorema o nedefinabilnosti istine**:

Teorem 21. *Neka domena jezika L sadrži za svaku rečenicu S jezika njen kod $< S >$ i neka za svaku formulu $A(x)$ s jednom slobodnom varijablom x sadrži konstantu c_A takvu da je*

$c_A^L =< \neg A(c_A) >$

Tada u jeziku L nije definabilan pojam istinite rečenice jezika L.

[3] Često korištena fraza "dovoljno bogat jezik" ne znači ništa drugo nego da jezik ima sredstva za takvu konstrukciju.

Dokaz : Pretpostavke teorema osiguravaju primjenjivost leme na svaku formulu $A(x)$ jezika L. ■

Jezici koji mogu izraziti sintaktičku operaciju dijagonalizacije svojih formula imaju uniforman način za tvorbu traženih terma, pa tako i za sintaktičku konstrukciju lašca. Ta konstrukcija je neformalno opisana u prethodnom poglavlju. Sada slijedi formalni opis. Pretpostavit ćemo da domena jezika L sadrži za svaku formulu $A(x)$ jezika L s najviše jednom slobodnom varijablom x njen kod $< A(x) >$ a sam jezik ime $\overline{A(x)}$ tog koda. Za takve jezike možemo definirati **dijagonalizaciju**, funkciju d koja svakoj formuli $A(x)$ jezika L s jednom slobodnom varijablom x pridružuje rečenicu nastalu supstitucijom na mjesto x imena koda same te formule:

$$d(A(x)) = A(\overline{A(x)}))$$

Pri navedenom kodiranju dijagonalizacija d prelazi u odgovarajuću funkciju $< d >$ nad kodovima:

$$< d > (< A(x) >) =< A(\overline{A(x)})) >$$

Ova funkcija nad ostalim objektima domene može biti zadana proizvoljno . Nadalje, pretpostavit ćemo da jezik L, ili eventualno njegovo definicijsko proširenje, sadrži funkcijski simbol D koji imenuje ovu funkciju $< d >$:

$$D^L =< d >$$

Dakle u jeziku L za svaku formulu $A(x)$ će vrijediti

$$L \models D(\overline{A(x)}) = \overline{A(\overline{A(x)})}$$

Takav funkcijski simbol ćemo zvati **funkcijski simbol dijagonalizacije**. Njegovo prisustvo u jeziku omogućuje **dijagonalizacijsku verziju teorema o nedefinabilnosti istine**:

Teorem 22. *Neka domena jezika L sadrži za svaku formulu $A(x)$ jezika L s najviše jednom slobodnom varijablom x njen kod $< A(x) >$ a sam jezik ime $\overline{A(x)}$ tog koda. Nadalje neka L sadrži funkcijski simbol dijagonalizacije D. Tada u tom jeziku nije definabilan pojam istinite rečenice tog jezika.*

Dokaz : Pretpostavimo da jezik ima predikat istine $T(x)$. Izrazivost dijagonalizacije omogućuje konstrukciju Lašca u datom jeziku. To je dijagonalizacija predikata

$\neg T(D(x))$, tj. rečenica

$$\neg T(D(\overline{\neg T(D(x))}))$$

Lako je vidjeti iz smisla simbola D da term $D(\overline{\neg T(D(x))})$ imenuje upravo tu rečenicu. Po lemi to znači da T nije predikat istine. ■

2.3 Posljedice Tarskijeve analize pojma istine i kritika

Baveći se problemom definicije pojma istinite rečenice datog jezika Tarski je uveo shemu T *kao kriterij sadržajne adekvatnosti te definicije i tako izolirao jedno osnovno svojstvo tog pojma kojim se koristimo u svakodnevnom govoru:*

$T(\overline{S}) \leftrightarrow S^*$

gdje je $\overline{S}$ ime rečenice S jezika L o kojem govorimo u jeziku ML kojim govorimo, a S^* je prijevod te rečenice u metajezik ML.

Mogućnost definicije pojma istine odnosno njene suvremene verzije, definicije istinitosti rečenica datog jezika u datom modelu, a koja se izriče u jeziku teorije skupova, dovela je do uvođenja semantike u matematičku logiku i time bitno pridonijela njenom budućem razvoju. Ta definicija osigurava za svaki model M jezika L, kojem je domena skup, postojanje predikata T_M u jeziku LS teorije skupova za kojeg vrijedi

$LS \models T_M(\overline{S}) \leftrightarrow S^*$ za svaku rečenicu S jezika L

Na takvim definicijama počivaju dokazi relativne konzistentnosti matematičkih teorija — za datu teoriju Th pronađe se u okviru neke teorije skupova (obično ZFC) model M u kojem su istiniti svi aksiomi teorije Th. Iz T - sheme, uz pretpostavku da * čuva kontradikciju, slijedi da konzistentnost teorije skupova povlači i konzistentnost teorije Th. Pošto je to urađeno za sve matematičke teorije osim teorije skupova, time je problem konzistentnosti matematike sveden na problem konzistentnosti teorije skupova. A ovo se opet svodi na analizu značenja jezika teorije skupova. Ova redukcija nije samo formalna već ima i značenjski karakter. Naime, matematičke teorije možemo nalaženjem ovakvih modela razumjeti kao teorije o objektima jezika skupova, dakle kao dio teorije skupova. *Tako je definicija istine doprinijela zasnivanju cijele matematike na jeziku skupova, odnosno na teoriji skupova* ZFC *kojom nastojimo izraziti značenje tog jezika.*

Nemogućnost definicije pojma istine izražena je Tarskijevim teoremima nedefinabilnosti — u dovoljno bogatom jeziku nije definabilan pojam istinite rečenice tog jezika. Preciznije, za takav jezik L ne postoji predikat istine T, tj. predikat za kojeg će vrijediti

$L \models T(\overline{S}) \leftrightarrow S$ za svaku rečenicu S jezika L.

Teoremi nedefinabilnosti istine pokazuju da jezik koji ima dovoljno bogatu domenu (beskonačnu) možemo jednostavno upotpuniti dodavanjem odgovarajućih konstanti ili pak funkcijskog simbola dijagonalizacije do jezika koji nema svoj predikat istine. *Tako možemo reći da je nedefinabilnost vlastitog predikata istine uobičajeno svojstvo bogatijih jezika logike prvog reda. Ovo naravno predstavlja ograničenje za jezik koji pretendira da bude osnovni jezik matematike, kao što je to danas jezik teorije skupova.* Ma kako precizirali njegovo značenje postoje neki relevantni pojmovi, kao pojam istinite rečenice tog jezika, koji ne mogu biti izraženi u tom jeziku. Mada se u njemu jednostavno može definirati predikat istine bilo kojeg interpretiranog jezika kojem je model skup, on nema predikat istine za samog sebe. Naravno, da bismo ovo precizno pokazali morali bismo opisati kodiranje njegovih rečenica u skupove i definirati funkciju dijagonalizacije. U osnovi, to bi bila precizna formulacija u jeziku skupova prethodnih poluformalnih razmatranja o jeziku. Bogatstvo jezika teorije skupova sugerira da je to moguće uraditi.

Teorem nedefinabilnosti ne govori da ne postoji parcijalni predikat istine jezika, tj. predikat istine za dio jezika. Npr. u jeziku LS teorije skupova je definabilan $T_n(x)$, predikat istinitosti Σ_n rečenica (nazovimo taj dio jezika $LS\Sigma_n$) :

$LS \models T_n(\overline{S}) \leftrightarrow S$ za svaku rečenicu S jezika $LS\Sigma_n$

Nasuprot nekim interpretacijama Tarskijevih rezultata po kojima oni zabranjuju samoreferiranje, ne samo da je samoreferiranje moguće, nego je moguće i samoreferiranje koje obuhvaća istinitost rečenica. Npr. pokaže se da je predikat $T_n(x)$ definabilan u jeziku $LS\Sigma_n$ tako da se može konstruirati i Σ_n rečenica koja kaže da je ona istinita Σ_n rečenica. Može se konstruirati i rečenica koja kaže da ona nije istinita Σ_n rečenica. No ona nije Lažac jer ona nije Σ_n rečenica. Štoviše, pošto nije Σ_n rečenica to nije ni istinita Σ_n rečenica, pa je to istinita rečenica. Nadalje uz neke male pretpostavke o skupovima može se pokazati da je svaka rečenica ekvivalentna nekoj Σ_n rečenici. Tako se jezik LS raslojava na beskonačan niz jezika od kojih je svaki u nizu proširenje svih prethodnih jezika i ima svoj predikat istine, koji je tako i parcijalni predikat istine cijelog jezika. No ukupan predikat istine tog jezika nije u njemu definabilan.

S druge pak strane refleksijom nad istinitošću jezika LS možemo ga proširiti do jezika LS_1 tako da jeziku LS dodamo predikat istine $T(x)$. Tako u jeziku LS_1 možemo govoriti o istinitosti njegovog podjezika:

$LS_1 \models T(\overline{S}) \leftrightarrow S$ za svaku rečenicu S jezika LS

Na isti način možemo proširiti i jezik LS_1 do jezika LS_2 koji sadrži njega i njegov predikat istine $T_1(x)$:

$LS_2 \models T_1(\overline{S}) \leftrightarrow S$ za svaku rečenicu S jezika LS_1

Tako dobivamo cijelu hijerarhiju sve bogatijih jezika. Mada nemamo jedan univerzalni jezik dobivamo niz sve boljih jezika koji po pitanju zasnivanja znanja ima određenu uvjerljivost. Naime, razmišljanje o pojmu istine nekog jezika je uvijek refleksija (pogled izvana) nad tim jezikom i prirodno se odvija u odgovarajućem metajeziku. Isto tako se i razmišljanje o metajeziku prirodno odvija u njegovom metajeziku, itd. *Tako je stalno pribjegavanje metajeziku imanentno refleksivnosti razmišljanja, pa hijerarhija jezika može biti zadovoljavajuće rješenje za zasnivanje znanja.*

Što se tiče primjene ove analize na prirodni jezik ona je upitna. Sam Tarski je izrazio rezervu po tom pitanju smatrajući prirodan jezik suviše nepreciznim za takvu analizu. No svaki se model koji bi htio precizirati prirodni jezik, logički ga pročistiti, mora suočiti s Tarskijevim rezultatima. Ako ijedan jezik ima pretenziju na univerzalnost to je onda prirodan jezik. No njegova univerzalnost je nespojiva s Tarskijevim rezultatima, jer univerzalnost zahtijeva i da ima vlastiti predikat istine, a što je u okvirima klasične semantike nemoguće. Alternativa je zamisliti prirodan jezik raslojen u nivoe, takve da svaki nivo ima predikat istine za rečenice prethodnih nivoa. To bi bila neka vrsta prihvaćanja principa refleksije za prirodan jezik. Naprosto bi nivo upotrebe predikata istine ovisio o kontekstu u kojem je upotrebljen. Taj kontekst bi odlučio o kojem je predikatu istine riječ. No Kripke je u svojoj analizi pojma istine pokazao da je postojanje takvih nivoa krajnje problematično. Kripkeova kritika će biti izložena u sljedećem poglavlju.

Što se tiče rješenja paradoksa, *Tarskijeva teorija je normativnog karaktera (osigurava da se oni neće pojaviti) ali ne i analitičkog karaktera (ne objašnjava njihov uzrok).* Tarski naprosto koristi paradokse, odnosno njihove formalne verzije, da bi pokazao da bogati jezici nemaju svoj predikat istine. Ne daje nikakvo objašnjenje zašto je to tako već samo konstataciju da jeste tako. Predikat istine takvog jezika L mora se opisati u bogatijem jeziku ML. Najčešće je to proširenje jezika L, pa je taj predikat istine tek parcijalni predikat istine jezika ML. Tako je Tarskijevo rješenje u stvari ograničenje sheme T:

$ML \models T(\overline{S}) \leftrightarrow S$ za svaku rečenicu S jezika $L \subset ML$

Paradoksi su izbjegnuti time da se naprosto rečenice tipa Lašca ne mogu konstruirati u jeziku ML jer takve rečenice ne pripadaju jeziku L za koje se može definirati djelomični predikat istine T. Za bogati jezik L razlog tome je da sam predikat $T(x)$ ne pripada jeziku L već striktno njegovom metajeziku ML. Tako i rečenica koja bi trebala biti Lažac glasi

“ Ja nisam istinita rečenica jezika L”.

No pošto ona uopće nije rečenica jezika L to je ona na trivijalan način istinita rečenica jezika ML. Ako imamo hijerahiju jezika od kojih svaki sadrži predikat istine za jezike koji su ispod njega, paradoksi se analogno izbjegavaju zahvaljujući tome da predikat istine datog jezika ne pripada tom jeziku pa se Lažac ne može ni sintaktički ni semantički konstruirati. Ograničavajuća sposobnost jezika rješava se pribjegavanjem metajeziku. No, takvo **hijerarhijsko rješenje** definitivno postavlja ograničenje na ono što možemo izreći u takvoj hijerarhiji. Drugim riječima, *Tarskijevo rješenje paradoksa je normativno — paradoks Lašca je izbjegnut time da se Lažac uopće ne može izreći. Isto normativno rješenje “liječi” i ostale paradokse istine. Ono se također javlja i previše restriktivnim jer se u njemu ne može izreći nijedna situacija u kojoj postoji cirkularno referiranje jednih rečenica na istinitost drugih rečenica, ma koliko takva situacija bila uobičajena i bezazlena.*

Tarskijeva analiza je pokazala da su za dovoljno bogat jezik nespojiva dva intuitivno posve prihvatljiva principa:

1. *Shema T*: $T(\overline{S}) \leftrightarrow S$ za svaku rečenicu S
2. *klasična dvovaljana semantika jezika*: svaka rečenica je ili istinita ili lažna

Tarskijevo rješenje se sastoji u zadržavanju dvovaljane semantike i restrikciji sheme T. Kripke je izabrao drugi put: zadržavanje sheme T i restrikciju dvovaljane semantike jezika.

POGLAVLJE 3

Kripkeova analiza pojma istine

3.1 Kripkeova teorija istine

Kripkeov članak "Outline of a Theory of Truth" [Kri75] je nakon Tarskijevog rada sljedeći značajan korak u logičkoj analizi pojma istine koji je istovremeno i filozofski argumentiran i matematički precizan. Mada je i prije njega bilo prijedloga da rješenje treba tražiti van klasične dvovaljane semantike ([Mar84]) on je u tom članku točno pokazao kako se to može uraditi.

U prvom dijelu članka Kripke se bavi utvrđivanjem što je glavni problem kod paradoksa istine. Za njegovu analizu je ključan Kontingentni Lažac. Za primjer uzima jednu sasvim uobičajenu situaciju, rečenice koje su npr. mogli izreći Jones i Nixon:

Jones: "Većina Nixonovih tvrdnji o Watergateu je lažna."

Nixon: "Sve Jonesove tvrdnje o Watergateu su istinite."

Ne samo da su te rečenice gramatički i značenjski uobičajene, već je za očekivati da se kroz jedan popis tvrdnji koje su Nixon i Jones izrekli o Watergateu može ustanoviti njihova istinitost. No moguća je (mada malo vjerojatna) situacija da je među svim ostalim Nixonovim tvrdnjama o Watergateu upravo pola lažno a pola istinito, te da su sve ostale Jonesove tvrdnje o Watergateu istinite. Pod tim uvjetima su te rečenice kontradiktorne. Naime, pod navedenim uvjetima, kad bi Jonesova tvrdnja bila istinita tada bi Nixonova tvrdnja bila lažna, a što bi značilo da je Jonesova tvrdnja lažna. No kad bi Jonesova tvrdnja bila lažna tad bi Nixonova tvrdnja bila istinita što bi značilo da je Jonesova tvrdnja istinita. Tako dobijemo kontradikciju. Kripke iz ovog po njemu paradigmatskog primjera izvlači sljedeću pouku:

> ... adekvatna teorija mora dopustiti da su rečenice koje sadrže pojam istine *riskantne*: one nose rizik da budu paradoksalne ako su empiričke činjenice ekstremno (i neočekivano) nepovoljne. Nema sintaktičkog niti semantičkog "sita" koje bi propustilo "loše slučajeve" a zadržalo "dobre".

Kripke u daljnjem proširuje razmatranje i na druge rečenice koje sadrže pojam istine i koje nisu kontradiktorne ali ipak jesu intuitivno problematične kao npr. Istinoljubac:

Istinoljubac: Istinoljubac je istinita rečenica

Kripke na primjeru Jonesove tvrdnje ukazuje na uzrok problematičnosti. Da bi se ispitala istinitost Jonesove tvrdnje mora se pogledati istinitost svih rečenica na koje ona referira. Ako te rečenice imaju utvrđenu istinitosnu vrijednost, tada ćemo i Jonesovoj tvrdnji odrediti istinitosnu vrijednost. Međutim ako neke od njih također sadrže pojam istine tada moramo ispitati istinitost rečenica na koje one referiraju itd. Ako ovaj proces završava na rečenicama koje ne sadrže pojam istine tada ćemo moći utvrditi istinitost Jonesove rečenice. Takve rečenice Kripke zove **utemeljenim** a one druge **neutemeljenim**. Naravno, opet je značajno da se to ne može odrediti na sintaktičkom niti na unutrašnjem značenjskom nivou jezika, već se mora pogledati vanjska značenjska struktura jezika — to ovisi o, kako Kripke kaže, "empiričkim činjenicama". Jednom od osnovnih vrijednosti svoje teorije smatra to što ona daje formalnu definiciju pojma utemeljenosti koji po njemu formulira intuiciju o neproblematičnim rečenicama.

U drugom dijelu članka Kripke se bavi analizom prethodnih pristupa problemu istine. Ističe Tarskijev pristup kao jedini razrađeni pristup. Tarskijevo rješenje unutar klasične logike (Kripke takav pristup zove ortodoksnim) vodi hijerarhiji jezika, tj. nizu jezika od kojih svaki ima predikat istine za rečenice jezika koji su prije njega u hijerarhiji. Tako predikat $istinita_1$ opisuje istinitost rečenica koje nemaju predikat istine, predikat $istinita_2$ opisuje istinitost rečenica koje najviše sadrže predikat istine $istinita_1$ itd. No Kripke pokazuje da ovakvo raslojavanje intuitivnog pojma istine na hijerarhijski niz predikata istine ima ozbiljne probleme u prirodnom jeziku, kao i u bogatijem formaliziranom jeziku. Prije svega, indeksi takvih predikata istine ne mogu se eksplicitno ili implicitno odrediti na sintaktičkom nivou ili pak unutarnjem semantičkom nivou već moraju ovisiti o "empiričkim činjenicama". Npr. kako odrediti indeks tog predikata u rečenici koju je izrekao Dean o Watergateu:

Dean: Sve Nixonove tvrdnje o Watergateu su lažne.

Indeks pojma istine u ovoj rečenici ovisi o indeksima pojma istine koji se javljaju u rečenicama koje je Nixon izrekao o Watergateu i to se može jedino *empirički* utvrditi. Još gore, empiričke činjenice mogu biti takve da je nemoguće takav indeks pridružiti datoj rečenici. Ako se npr. među Nixonovim tvrdnjama nalazi sljedeća rečenica

Nixon: Sve Deanove tvrdnje o Watergateu su lažne.

Zamisao o hijerarhiji indeksa sad vodi kontradikciji: indeks pojma istine u Deanovoj rečenici mora biti viši od indeksa pojma istine u Nixonovoj rečenici, ali i indeks pojma istine u Nixonovoj rečenici mora biti viši od indeksa pojma istine u Deanovoj rečenici. Nasuprot tome na intuitivnom nivou može biti sasvim jednostavno odrediti istinitost ovih rečenica. Npr. ako je Dean izrekao barem jednu istinitu tvrdnju o Watergateu pored navedene, tada je Nixonova tvrdnja lažna. Ako su i sve ostale Nixonove tvrdnje o Watergateu lažne tada je Deanova tvrdnja istinita, u protivnom je lažna. Na osnovu ovakvih razmatranja Kripke zaključuje da je teško intuiciju o pojmu istine prilagoditi zahtjevima hijerarhijskog pristupa.

Što se tiče postojeće literature koja se bavi traženjem alternative ortodoksnom pristupu Kripke smatra da se ona u jednom slaže: mora biti jedan predikat istine, primjenjiv i na rečenice koje sadrže taj predikat istine, a paradoksi se mogu izbjeći dopuštanjem da neke rečenice, kao one koje vode paradoksima, mogu ne imati istinitosnu vrijednost. Kripkeova glavna zamjerka svim tim zamislima je da su one, za razliku od Tarskijeve analize, samo sugestije a nikako i razrađene teorije. No Kripke eksplicitno kaže da postoji jasan utjecaj tih ideja na njegovu analizu.

U trećem dijelu članka Kripke izlaže svoju analizu pojma istine. *U osnovi te analize je prihvatanje T - sheme (jer ne postoje kriteriji koji bi razlikovali dobre od loših slučajeva i koji bi tako dali prirodnu restrikciju na tu shemu) i prijelaz na trovaljanu semantiku (jer teorija istine treba prihvatiti rizik da neke rečenice, mada smislene nisu ni istinite ni lažne). Kripke se odmah ograđuje da njegova analiza neće dati točno određenu interpretaciju pojma istine niti točno određeno rješenje paradoksa istine*, jer:

> ...u ovom momentu još nisam osmislio pažljivo filozofsko opravdanje za određeni prijedlog, niti sam siguran koja su područja i granice primjenjivosti takvog prijedloga.

No on se nada da njegova analiza daje dobar matematički okvir za takvo rješenje:

> Nadam se da model koji je ovdje dan ima dvije vrline: prvo, da daje područje bogato formalnom strukturom i matematičkim svojstvima; drugo, da u razumnoj mjeri ova svojstva obuhvaćaju važne intuicije. Model tako treba biti testiran s obzirom na tehničku plodnost. On ne mora obuhvatati svaku intuiciju, ali za nadati se je da će obuhvatati mnoge.

Kripke predlaže ispitivanje jezika kojem ne moraju sve rečenice imati istinitosne vrijednosti. On rečenice smatra pokušajima tvrđenja ili izražavanja propozicija. Sve rečenice, pa i one koje vode paradoksima smatra smislenim, jer one daju upute za tvorbu određenog tvrđenja no one ne moraju i tvoriti tvrđenje, tj. imati istinitosnu vrijednost. Nepostojanje istinitosne vrijednosti zahtijeva određenu shemu valuacije složenih rečenica koja će odrediti kako nepostojanje istinitosnih vrijednosti jednih rečenica utječe na postojanje istinitosnih vrijednosti drugih rečenica koje su od njih složene. Kripke smatra tzv. **jaku Kleeneovu trovaljanu semantiku** jednom takvom shemom koja odgovara zamisli da su rečenice smislene bez obzira imaju li istinitosnu vrijednost ili ne. Po toj shemi $\neg S$ je istinita (lažna) ako je S lažna (istinita), a neodređena ako je S neodređena. Disjunkcija je istinita ako je barem jedan disjunkt istinit, lažna je ako su oba disjunkta lažna, a u preostalim slučajevima je neodređena. Ostali veznici se na standardni način definiraju pomoću ovih i time je određeno i njihovo ponašanje s obzirom na neodređenu vrijednost. Egzistencijalna rečenica se tretira kao beskonačna disjunkcija, a univerzalna rečenica kao beskonačna konjukcija. Precizna matematička formulacija Kripkeove konstrukcije dana je u sljedećem odjeljku.

Konstrukcija jezika koji ima predikat istine i kojem sve rečenice ne moraju imati istinitosnu vrijednost treba da obuhvati sljedeću intuiciju. Kripke razmatra na koji način bismo nekome tko ne zna za pojam istine objasnili koje su rečenice istinite a koje lažne. Neka je u pitanju kompetentan govornik nekog jezika u kojem ne postoji predikat istine. Takav subjekt je sposoban svaku rečenicu tog jezika ili prihvatiti ili odbaciti. On npr. prihvaća rečenicu

(1) Snijeg je bijel.

ali ne zna šta uraditi s rečenicama koje sadrže pojam istine kao npr.

(2) “Snijeg je bijel” je istinita rečenica.

(3) ““Snijeg je bijel” je istinita rečenica” je lažna rečenica.

(4) Snijeg je bijel ili ““Snijeg je bijel” je istinita rečenica” je lažna rečenica.

(5) Snijeg je bijel i ""Snijeg je bijel" je istinita rečenica" je lažna rečenica.

No dat ćemo mu *pravilo da prihvati (odbaci) rečenicu "$\overline{S}$ je istinita rečenica" uvijek kada prihvati (odbaci) rečenicu S.* Tako će po tom pravilu on sada prihvatiti i rečenicu (2). Za složene rečenice, kao što su (4) i (5) dat ćemo mu neku *trovaljanu semantiku po kojoj će moći vrednovati složene rečenice i kada neke njihove komponente nije ni prihvatio ni odbacio.* Tako npr. po Jakoj Kleeneovoj trovaljanoj semantici prihvatit će rečenicu (4) bez obzira što nije vrednovao drugi član disjunkcije, jer je već prihvatio prvi disjunkt, dok neće moći vrednovati rečenicu (5) jer to ovisi o još nevrednovanom drugom članu konjukcije. Što se tiče rečenice (3) samo ga trebamo uputiti da *usvojena pravila može iterirati - uvijek kad napravi vrednovanja nekih rečenica, ima podlogu za vrednovanje novih rečenica.* Tako na primjer prihvativši rečenicu (2) on sad ima podlogu za prihvaćanje rečenice ""Snijeg je bijel" je istinita rečenica" je istinita rečenica, odnosno za odbacivanje rečenice (3) (jer " biti lažna rečenica " znači da za negaciju vrijedi "biti istinita rečenica"). Odbacivši rečenicu (3) on sada ima podlogu i za odbacivanje rečenice (5). Nastavljajući taj proces subjekt bi sve više rečenica prihvatao ili odbacivao. Neke rečenice, poput rečenice koja kaže da je ona sama istinita, nikad ne bi na ovaj način mogao vrednovati. To su rečenice koje su u intuitivnom smislu neutemeljene.

Formalna konstrukcija koja odgovara ovoj intuiciji sada će biti tek skicirana dok će u sljedećem odjeljku biti detaljno opisana. Neka je L klasični dvovaljani interpretirani jezik logike prvog reda. Njemu ćemo dodati za početak neinterpretirani predikat T i tako ga proširiti do jezika LT. Pri tome ćemo pretpostaviti da postoji kodiranje rečenica jezika LT u domenu D jezika L — svakoj rečenici S pridružen je kod $< S >\in D$. Pošto T namjeravamo interpretirati kao predikat istinitosti i pošto očekujemo da je njegova upotreba riskantna, tj. ne mora dati određenu istinitosnu vrijednost, to ćemo njega interpretirati parom (E, A) disjunktnih podskupova od D. Ideja je da skup E treba sadržavati kodove rečenica na kojima T poprima istinu, i zvat ćemo ga **ekstenzija** predikata T, dok A treba sadržavati kodove rečenica na kojima T poprima laž kao i sve objekte koji nisu kodovi rečenica i na kojima T također poprima laž, i taj skup ćemo zvati **antiekstenzija** predikata T. Naravno, ostavljamo mogućnost da $E \cup A$ ne obuhvaća sve kodove rečenica. Na preostalim kodovima rečenica predikat T je neodređen. Tako interpretirani jezik LT ćemo bilježiti $LT(E, A)$. **Početnu interpretaciju** (E_0, A_0) predikata T tvore prazni skupovi kodova rečenica, tj. $E_0 = \emptyset$ a A_0 je skup svih objekata koji nisu kodovi rečenica. U prvom koraku subjekt će naučiti da su sve rečenice početnog jezika koje on prihvaća istinite rečenice(njihovi kodovi tvore novu ekstenziju E_1), a one koje odbacuje lažne rečenice (njihovi kodovi zajedno s objektima koji nisu kodovi rečenica tvore novu antiekstenziju A_1 predikata T. Dakle $T(E_1, A_1)$ je naprosto predikat istine jezika L. No cilj je da on bude predikat istine jezika LT, tj. da bude istinit na svim istinitim rečenicama tog jezika, a lažan na svim lažnim rečenicama tog jezika i objektima koji koji nisu kodovi rečenica. U tu svrhu ćemo *poboljšavati* inerpretaciju. Naime po navedenim pravilima jezik $LT(E_1, A_1)$ ima više lažnih i istinitih rečenica nego što je evidentirano u skupovima E_1 i A_1, pa ćemo njihovim dodavanjem napraviti novu interpretaciju $LT(E_2, A_2)$, gdje je E_2 skup svih kodova istinitih rečenica jezika $LT(E_1, A_1)$, a A_2 skup svih kodova lažnih rečenica zajedno s objektima koji nisu kodovi rečenica. No sada se situacija ponavlja. Na isti način za svaki ordinal α pravimo prijelaz s $LT(E_\alpha, A_\alpha)$ na bogatiji jezik $LT(E_{\alpha+1}, A_{\alpha+1})$. U transfinitnom skoku novu ekstenziju i novu antiekstenziju tvori unija svih dotadašnjih ekstenzija, odnosno antiekstenzija. Tako generiramo transfinitni niz jezika sa sve bogatijim ekstenzijama i antiekstenzijama (subjekt usvaja sve više rečenica kao istinite i sve više kao lažne). Primijetimo da nova interpretacija nastaje refleksijom nad starom interpretacijom, odnosno nad svim starim interpretacijama u slučaju transfinitnog skoka. Ako u jednom takvom prijelazu nismo dobili ništa novo, tj. ako je nova interpretacija jednaka staroj, ako je, matematičkim rječnikom kazano, stara interpretacija **fiksna točka** ovog **reinterpretiranja**, tada stari predikat $T(E_\alpha, A_\alpha)$ obuhvaća svojom ekstenzijom i antiekstenzijom upravo sve kodove istinitih i sve kodove lažnih rečenica jezika $LT(E_\alpha, A_\alpha)$— on je **predikat istine tog jezika**. U sljedećem odjeljku su detaljno opisana matematička svojstva konstrukcije koja potvrđuju ovu intuiciju. Naime jaka Kleeneova trovaljana semantika ima sljedeće svojstvo: ako rečenici koja nema istinitosnu vrijednost dodijelimo istinitosnu vrijednost, to neće utjecati na rečenice koje imaju istinitosnu vrijednost. Njihova istinitosna vrijednost će ostati ista, samo će neke do tada neodređene rečenice dobiti istinitosnu vrijednost. Za operaciju reinterpretiranja se tada kaže da je ona *monotona*. Pokaže se da monotonost garantira ono što intuitivno i očekujemo, da je ovaj niz transfinitnih jezika *rastući*, tj da dobivamo sve veće ekstenzije i antiekstenzije predikata T. No jedan argument iz teorije skupova tada osigurava da se ovaj proces jedanput mora zasititi, tj. da je za neki ordinal α jezik $LT(E_\alpha, A_\alpha)$ fiksna točka ovog reinterpretitanja, dakle jezik s vlastitim predikatom istine. Štoviše, u sljedećem odjeljku ćemo pokazati da je to **najmanja fiksna točka predikata istine**, tj. da je manja od svih drugih fiksnih točaka u smislu da ima i manji skup istinitih tvrdnji i manji skup lažnih tvrdnji nego ijedna druga fiksna točka. Kripke je tako pokazao sljedeće:

> ...svaki jezik, uključujući i one koji sadrže teoriju brojeva ili sintaksu, može se proširiti do jezika s vlastitim predikatom istine, i asocirani pojam istine je *matematički* definiran teorijsko skupovnim tehnikama.

Tako je Kripke dao rješenje koje je matematički precizno kao i Tarskijevo, no suprotno od Tarskijevog rješenja ovdje je sačuvana shema T a modificirana klasična dvovaljana semantika. Opisana konstrukcija za razliku od Tarskijeve, daje jedan predikat istine a ne beskonačno njih. Također ona formulira ideju hijerarhije kao sve bolje aproksimacije minimalnog jezika s predikatom istine. Ta hijerarhija je primjenjiva i u situacijama u kojima Tarskijeva hijerarhija nije primjenjiva i za razliku od Tarskijeve jednostavno se proširuje na transfinitne nivoe. Ona obuhvaća ideju da nivo rečenica ovisi o " empiričkim činjenicama", kao i ideju riskantnosti rečenica — rečenica u nekim krajnjim slučajevima može ne dobiti određenu istinitosnu vrijednost. To su rečenice koje ne pripadaju najmanjoj fiksnoj točki (njenoj ekstenziji ili antiekstenziji) i Kripke ih zove **neutemeljenim rečenicama**. Pogledajmo opet primjer Deana i Nixona:

Dean: Sve Nixonove tvrdnje o Watergateu su lažne.

Nivo ove rečenice u Kripkeovoj hijerarhiji ovisi o nivoima rečenica koje je Nixon izrekao o Watergateu. Ako Nixon nije izrekao direktno ili indirektno (spominjući izjave drugih koji su govorili o Deanu) ništa o Deanu, tada je moguće da njegove rečenice imaju sasvim određene nivoe u hijerarhiji, pa će Deanova rečenica imati prvi sljedeći nivo u hijerarhiji. No ako je Nixon izrekao (između ostalog o Watergateu) i sljedeće:

Nixon: Sve Deanove tvrdnje o Watergateu su lažne.

i tada je moguće (za razliku od Tarskijeve hijerarhije) da Deanova rečenica ima određen nivo. Ako je Dean pored navedene tvrdnje izrekao barem jednu istinitu (utemeljenu na nivou α) tvrdnju o Watergateu, tada je Nixonova tvrdnja lažna, utemeljena na nivou $\alpha + 1$. Ako su i sve ostale Nixonove tvrdnje o Watergateu lažne tada je Deanova tvrdnja istinita na prvom nivou nakon svih nivoa Nixonovih tvrdnji. Dakle, empiričke činjenice određuju nivo rečenice ali i pitanje je li ona uopće utemeljena. Naime, ako je situacija takva (mada krajnje nevjerojatna) da su gornje tvrdnje jedine koje su Nixon i Dean izrekli o Watergateu tada je jasno da one nisu utemeljene.

Što se tiče ostalih fiksnih točaka predikata istine Kripke kaže sljedeće:

> Iako je najmanja fiksna točka vjerojatno najprirodniji model za intuitivni pojam istine, i to je model *generiran* našim instrukcijama imaginarnom subjektu, druge fiksne točke nikad nisu u *konfliktu* s tim instrukcijama.

Ostale fiksne točke Kripkeu služe za razlikovanje neutemeljenih rečenica. Za razliku od utemeljenih rečenica koje imaju u svim fiksnim točkama istinitosnu vrijednost, i to uvijek istu, Lažac u nijednoj fiksnoj točki nema istinitosnu vrijednost. Rečenice s tim svojstvom Kripke zove **paradoksalnim**. Za razliku od njih Istinoljubac je neutemeljena rečenica za koju je lako pronaći fiksnu točku u kojoj je istinita (lažna). Naprosto Kripkeovu konstrukciju započnemo tako da u ekstenziju (antiekstenziju) predikata T stavimo kod Istinoljupca. Takve rečenice koje u nekim fiksnim točkama imaju jednu istinitosnu vrijednost a u nekima drugu Visser [Vis89] je poslije nazvao **bikonzistentnim**. U tom kontekstu je zanimljivo pogledati **maksimalne fiksne točke predikata istine**, tj. fiksne točke od kojih nema većih. Za njih Kripke kaže da su to fiksne točke koje pridružuju što je moguće više istinitosnih vrijednosti, a da je to u skladu s intuitivnim pojmom istine. Svaka fiksna točka se može, po Zornovoj lemi, proširiti do maksimalne fiksne točke. U svakoj od tih točaka Istinoljubac ima određenu istinitosnu vrijednost, ali ne i u svima istu, što ukazuje na *proizvoljnost* istinitosnih vrijednosti u tim točkama. Utemeljene rečenice imaju istu istinitosnu vrijednost u svim fiksnim točkama pa i tako gledano njihova istinitosna vrijednost nije proizvoljna. No postoje i neutemeljene i neparadoksalne rečenice koje ako imaju istinitosnu vrijednost u nekoj fiksnoj točki, ona im je uvijek ista. Takva je npr. sljedeća rečenica:

Istinoljubac je istinita rečenica ili Istinoljubac nije istinita rečenica.

Mada ova rečenica nema proizvoljnu istinitosnu vrijednost (ako nije neodređena mora biti istinita), njeni dijelovi mogu imati proizvoljnu vrijednost. Nasuprot tome istinitosna vrijednost sljedeće rečenice, koja će ovdje biti nazvana **Logičar**, je *hereditarno neproizvoljna*:

Logičar: Logičar je istinita rečenica ili Logičar nije istinita rečenica.

Ova rečenica ima istinitosnu vrijednost u nekoj fiksnoj točki, i ta je vrijednost ista u svim fiksnim točkama u kojima ona ima istinitosnu vrijednost. Dakle, ta vrijednost nije proizvoljna. No ta neproizvoljnost je drugog tipa od prethodne rečenice, jer i njeni dijelovi imaju neproizvoljnu istinitosnu vrijednost (ako je imaju). Takve rečenice tvore **intrinsične fiksne točke**, fiksne točke u kojima je istinitosna vrijednost rečenica u suglasju s istinitosnim vrijednostima tih rečenica u svim drugim fiksnim točkama. Naime, ako je npr. rečenica u toj točki istinita tada je istinita i u svim fiksnim točkama u kojima ima istinitosnu vrijednost. Mada ne postoji najveća fiksna točka, postoji **najveća intrinsična fiksna točka** koja je zbog svoje jedinstvenosti u strukturi fiksnih točaka posebno zanimljiva. Za nju Kripke kaže:

> Najveća intrinsična fiksna točka je jedinstvena "najveća" interpretacija od $T(x)$ koja je konzistentna s našom intuitivnom idejom istine i koja nema proizvoljne izbore u određivanju istinitosti. Ona je tako objekt od specijalnog teorijskog interesa kao model.

Kripke u daljnjem ističe da se fiksne točke mogu dobiti i s drugim trovaljanim semantikama umjesto jake Kleeneove. Za konstrukciju je jedino važno da su te semantike monotone. Pri tome posebno ističe van Fraasenovu supervaluaciju. U toj valuaciji interpretacija $T(E, A)$ se proširuje tako da joj se dodaju rečenice koje su istinite (lažne) u svim proširenjima ove interpretacije koja su potpuna (klasična) U toj valuaciji su sve klasične logičke istine istinite u minimalnoj fiksnoj točki, dok su kod Kleeneove valuacije istinite jedino ako imaju istinitosnu vrijednost. No Kleene ponavlja svoju osnovnu namjeru:

> Nije svrha ovog rada dati bilo koju preporuku izmedju Kleeneovog jakog trovaljanog pristupa, van Fraasenove supervaluacijskog pristupa, ili bilo koje druge sheme...Niti je moja sadašnja svrha dati bilo koju čvrstu preporuku između minimalne fiksne točke date sheme i različitih drugih fiksnih točkaka...Moja je svrha dati familiju fleksibilnih instrumenata koji mogu biti istovremeno istraživani i čija plodnost i suglasnost s intuicijom može biti provjerena. ...To ne znači da tvrdim da ne postoje definitni odgovori ili da ja osobno ne preferiram neke valuacijske sheme pred drugima. Ali mojo osobna gledišta su manje važna nego raznolikost oruđa koje je na raspolaganju, tako da za svrhu ovog pregleda zauzimam agnostičku poziciju.

No u fusnoti napominje da se ipak najmanja fiksna točka ističe u mnogim aspektima kao prirodan izbor.

Na kraju članka Kripke komentira koliko se njegov jezik približava idealu univerzalnog jezika. Time što ti jezici sadrže predikat istine oni odgovaraju ideji univerzalnosti. No sama konstrukcija nije provedena u jeziku nego u metajeziku teorije skupova. Nadalje, postoje tvrdnje o objektnom jeziku koje ne možemo izreći u objektnom jeziku. Npr. Lažac nije istinit u objektnom jeziku no mi to ne možemo izreći u objektnom jeziku:

> Ako mislimo o minimalnoj fiksnoj točki, recimo Kleeneove valuacije, kao o modelu prirodnog jezika, tvrdnja da Lažac nije istinit mora biti zamišljena kao tvrdnja asocirana s naknadnim stupnjem u razvijanju prirodnog jezika, u kojem govornici reflektiraju nad generirajućim procesom koji vodi najmanjoj fiksnoj točki. No ona sama nije dio tog procesa. Nužnost pribjegavanja metajeziku možda je jedna od slabosti ove teorije. Duh Tarskijeve hijerarhije je još uvijek s nama.

Nadalje Kripke spominje mogućnost upotpunjenja jezika. U fiksnoj točki je shema T ispunjena u smislu da rečenica "S je istinita" i rečenica S ili imaju istu istinitosnu vrijednost ili su obje neodređene. Alternativna intuicija bi bila da "S je istinita" proglasimo lažnim kad je S lažna ili neodređena. Tada bi T bio totalni predikat i dobili bismo klasični jezik, no shema T više ne bi vrijedila. To modificiranje interpretacije predikata T u datoj fiksnoj točki Kripke zove **dovršavanjem predikata istine**. Time bismo u stvari dobili metajezik kojim opisujemo cijelu konstrukciju. Njegov predikat istine ne bi se podudarao s predikatom T koji opisuje istinitost jezika prije takvog dovršenja. No Kripke odbija takvu interpretaciju sa sljedećim obrazloženjem:

> Mislim da se primarnost prve intuicije može filozofski opravdati, i zbog toga sam istaknuo pristup zasnovan na toj intuiciji. Alternativna intuicija se javlja tek nakon refleksije nad procesom koji obuhvaća prvu intuiciju.

Mada Kripke uočava u čemu njegov jezik nije univerzalan ipak smatra da to nije razlog da se odbaci njegov pristup:

> Na osnovi činjenice da se cilj postizanja univerzalnog jezika čini nedostiživim neki su zaključili da je pristup s neodređenostima, ili bilo koji drugi pristup koji se pokušava više približiti prirodnom jeziku nego ortodoksni pristup, neplodan. Ja se nadam da plodnost ovog pristupa, i njegovo slaganje s intuicijom o prirodnom jeziku u mnoštvu slučajeva, dovodi u sumnju takva negativna shvaćanja.

3.2 Matematika Kripkeove analize pojma istine

U ovom odjeljku je dan precizan opis kako se interpretirani jezik logike prvog reda može nadopuniti do trovaljanog jezika s vlastitim predikatom istine, analizirana je međusobna povezanost takvih nadopuna, te je analizirana istinitosna vrijednost samoreferirajućih rečenica u tim nadopunama.

Pod **bazičnim interpretiranim jezikom** L podrazumijevat ćemo jezik logike prvog reda zajedno sa svojim modelom. Nosač modela zvat ćemo **domenom jezika** L i bilježiti DL. Jedina pretpostavka je da DL sadrži sve kodove rečenica jezika LT kojeg dobijemo dodavanjem jednomjesnog predikata T jeziku L. Dakle, pretpostavljamo postojanje **kodiranja** tj. injektivne funkcije

$$< >: SLT \longrightarrow DL$$

gdje je SLT skup rečenica jezika LT. Tako će svaka rečenica S jezika LT imati svoj kod $< S >$. Podskup domene jezika sastavljen od svih kodova rečenica od LT bilježit ćemo SLT^*. Ovo kodiranje omogućava da jezik LT govori o vlastitim rečenicama.

Smjerana interpretacija predikata T je da on bude predikat istine jezika LT tj. da poprima vrijednost istine ($\top$) na kodovima istinitih rečenica od LT, vrijednost laži ($\perp$) na kodovima lažnih rečenica od LT i na objektima koji nisu kodovi rečenica, a vrijednost neodređeno ($|$) na kodovima neodređenih rečenica od LT. Za početak **interpretacija za** T može biti bilo koja trovaljana funkcija

$$t : DLT \longrightarrow \{\top, \perp, |\}$$

s domene jezika DLT u skup istinitosnih vrijednosti $IV = \{\top, \perp, |\}$ takva da je $t(x) = \perp$ kada x nije kod rečenice. Ovo je varijanta Kripkeove interpretacije putem ekstenzije i antiekstenzije predikata. Naprosto za ekstenziju uzmemo skup svih objekata koje t preslikava u $\top$, a za antiekstenziju skup svih objekata koje t preslikava u $\perp$. U ovakvom pristupu je disjunktnost ovih skupova automatski osigurana. Pošto je vrijednost ovakvih funkcija van kodova rečenica uvijek ista smatrat ćemo ih funkcijama sa skupa kodova rečenica:

$$t : SLT^* \longrightarrow IV$$

Prošireni jezik LT s ovako interpretiranim T bilježit ćemo $LT(t)$. U klasičnoj notaciji to znači da je

$$T^{LT(t)} = t$$

Za jezik $LT(t)$ denotacija terma n u valuaciji v (oznaka $D_v^{LT(t)}(n)$) je definirana standardnom rekurzivnom definicijom. No valuacija formula nije standardna jer interpretacija od T dopušta da neke rečenice mogu poprimiti neklasičnu istinitosnu vrijednost $|$. Zato moramo imati **trovaljanu se-**

mantiku jezika TS kojom ćemo interpetirati veznike i kvantore. Ona im pridružuje odgovarajuće transformacije istinitosnih vrijednosti:

$\neg \quad \longmapsto \quad \neg^{TS} : IV \longrightarrow IV$

$\wedge, \vee, \rightarrow, \leftrightarrow \quad \longmapsto \quad \wedge^{TS}, \vee^{TS}, \rightarrow^{TS}, \leftrightarrow^{TS} : IV^2 \longrightarrow IV$

$\forall, \exists \quad \longmapsto \quad \forall^{TS}, \exists^{TS} : P(IV) \setminus \emptyset \longrightarrow IV$

gdje je $P(IV)$ partitivni skup skupa IV

Tek specificiranjem tih transformacija specificirali smo trovaljanu semantiku TS. To nam omogućava rekurzivnu definiciju **funkcije istinitosne vrijednosti** $I^{LT(t)}_{v,TS}$ koja svakoj formuli φ pridružuje istinitosnu vrijednost u datom jeziku $LT(t)$, datoj trovaljanoj semantici TS i datoj valuaciji v:

$\varphi \quad \longmapsto \quad I^{LT(t)}_{v,TS}(\varphi)$

Radi jednostavnosti često će biti izostavljani indeksi koji se u datom kontekstu podrazumijevaju. Tako ćemo i ovdje ovu funkciju jednostavnije bilježiti I^t_v. Ona je definirana sljedećim rekurzivnim uvjetima:

1. Na atomarnim formulama jezika L vrijednost funkcije I^t_v se podudara s istinitosnim vrijednostima jezika L u valuaciji v:
2. $I^t_v(T(n)) = t(D_v(n))$, za bilo koji term n
3. $I^t_v(\neg\varphi) = \neg^{TS} I^t_v(\varphi)$
4. $I^t_v(\varphi b \psi) = I^t_v(\varphi) b^{TS} I^t_v(\psi)$ za binarni veznik b
5. $I^t_v(\forall x\varphi) = \forall^{TS}(\{I^t_{v[d/x]}(\varphi) \mid d \in DLT\})$
6. $I^t_v(\exists x\varphi) = \exists^{TS}(\{I^t_{v[d/x]}(\varphi) \mid d \in DLT\})$

Istinitosna vrijednost rečenice S ne ovisi o valuaciji v pa ćemo je jednostavnije bilježiti $I^t(S)$.

Najviše će nas zanimati **jaka Kleeneova trovaljana semantika** koju ćemo kraće bilježiti SK. Rekurzivni uvjeti na funkciju istinitosti I^t_v te semantike su sljedeći:

1. $I^t_v(\neg\varphi) = \begin{cases} \top & za\ I^t_v(\varphi) = \bot \\ \bot & za\ I^t_v(\varphi) = \top \\ | & u\ protivnom \end{cases}$

φ	$\neg\varphi$
$\top$	$\bot$
$\bot$	$\top$
$\mid$	$\mid$

2. $I_v^t(\varphi \wedge \psi) = \begin{cases} \top & \textit{za } I_v^t(\varphi) = \top \textit{ i } I_v^t(\psi) = \top \ (\textit{obje su istinite}) \\ \bot & \textit{za } I_v^t(\varphi) = \bot \textit{ ili } I_v^t(\psi) = \bot \ (\text{barem je jedna lažna}) \\ \mid & \textit{u protivnom} \end{cases}$

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\bot$	$\mid$
$\bot$	$\bot$	$\bot$	$\bot$
$\mid$	$\mid$	$\bot$	$\mid$

3. $I_v^t(\varphi \vee \psi) = \begin{cases} \top & \textit{za } I_v^t(\varphi) = \top \textit{ ili } I_v^t(\psi) = \top \ (\textit{barem jedna je istinita}) \\ \bot & \textit{za } I_v^t(\varphi) = \bot \textit{ i } I_v^t(\psi) = \bot \ (\text{obje su lažne}) \\ \mid & \textit{u protivnom} \end{cases}$

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\top$	$\top$
$\bot$	$\top$	$\bot$	$\mid$
$\mid$	$\top$	$\mid$	$\mid$

4. $I_v^t(\varphi \to \psi) = \begin{cases} \top & \textit{za } I_v^t(\varphi) = \bot \textit{ ili } I_v^t(\psi) = \top \\ \bot & \textit{za } I_v^t(\varphi) = \top \textit{ i } I_v^t(\psi) = \bot \\ \mid & \textit{u protivnom} \end{cases}$

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\bot$	$\mid$
$\bot$	$\top$	$\top$	$\top$
$\mid$	$\top$	$\mid$	$\mid$

5. $I_v^t(\varphi \leftrightarrow \psi) = \begin{cases} \top & za\ I_v^t(\varphi) = I_v^t(\psi) \neq | \text{ (obje su istinite ili obje su lažne)} \\ \bot & za\ I_v^t(\varphi) \neq I_v^t(\psi)\ i\ nijedna\ vrijednost\ nije\ | \\ & \text{(jedna je istinita a druga lažna)} \\ | & u\ protivnom \end{cases}$

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\bot$	$\mid$
$\bot$	$\bot$	$\top$	$\mid$
$\mid$	$\mid$	$\mid$	$\mid$

6. $I_v^t(\forall x \varphi(x)) = \begin{cases} \top & ako\ \forall a \in DLT\ I_v^t(\varphi(\bar{a})) = \top \\ \bot & ako\ \exists a \in DLT\ I_v^t(\varphi(\bar{a})) = \bot \\ | & u\ protivnom \end{cases}$

7. $I_v^t(\exists x \varphi(x)) = \begin{cases} \top & ako\ \exists a \in DLT\ I_v^t(\varphi(\bar{a})) = \top \\ \bot & ako\ \forall a \in DLT\ I_v^t(\varphi(\bar{a})) = \bot \\ | & u\ protivnom \end{cases}$

Jaka Kleeneova trovaljana semantika se obično interpretira kao *investigativna* semantika, tj. kao semantika u kojoj rečenici kojoj još nismo pridružili klasičnu istinitosnu vrijednost pridružujemo vrijednost neodređeno ($|$). Npr. ako je rečenica φ istinita a rečenici ψ još ne znamo klasičnu vrijednost, tj. neodređena je, po klasičnim uvjetima na istinitost ipak možemo zaključiti da je $\varphi \vee \psi$ istinita. No vrijednost od $\varphi \wedge \psi$ bitno ovisi o vrijednosti ψ. A pošto ova (još) nije određena ne možemo odrediti ni klasičnu vrijednost od $\varphi \wedge \psi$ pa i njoj pridružujemo vrijednost neodređeno. Na taj način možemo interpretirati sve navedene rekurzivne uvjete ove semantike. Iz istih razloga se možemo uvjeriti i da vrijede sve klasične interdefinabilnosti veznika i kvantora.

Za razliku od jake Kleeneove semantike koja odgovara ideji da su rečenice smislene i kad nemaju klasičnu istinitosnu vrijednost, **slaba Kleeneova trovaljana semantika** WK odgovara ideji da rečenice bez klasične istinitosne vrijednosti nisu smislene. U toj semantici čim je jedna rečenica neodređena tada su i sve druge rečenice koje je u sebi sadrže također neodređene. Preciznije, u rekurzivnim uvjetima izlaz je neodređen uvijek kada je barem jedan ulaz neodređen, inače je određen po klasičnim uvjetima. Tako je npr. tablica za konjukciju sljedeća:

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\bot$	$\mid$
$\bot$	$\bot$	$\bot$	$\mid$
$\mid$	$\mid$	$\mid$	$\mid$

U nekim trovaljanim semantikama istaknute su i drugačije interpretacije klasičnih veznika. Takve su npr.

1. **jaka negacija**

φ	$\neg\varphi$
$\top$	$\bot$
$\bot$	$\top$
$\mid$	$\top$

2. **Lukasiewiczev bikondicional**

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\bot$	$\mid$
$\bot$	$\bot$	$\top$	$\mid$
$\mid$	$\mid$	$\mid$	$\top$

U datoj trovaljanoj semantici i datom bazičnom jeziku interpretacija predikata T, tj. funkcija

$t : SLT^* \longrightarrow IV$

određuje istinitost rečenica jezika $LT(t)$ tj. funkciju

$I^t : SLT \longrightarrow IV$

koja preko kodiranja rečenica daje funkciju istog tipa kao i t:

$I^{t*} : SLT^* \longrightarrow IV$

gdje je $I^{t*}(< S >) = I^t(S)$

Cilj je postići takvu interpretaciju t predikata T da on bude predikat istinitosti datog jezika, tj. da t generira istinitosnu funkciju koja je na kodovima upravo jednaka t:

$I^{t*} = t$

Lako se uvjeriti da je upravo tada ispunjena T shema koju pomoću funkcije istinitosti možemo ovako izreći:

$I^t_{v[<S>/x]}(T(x)) = I^t(S)$

U traženju željene interpretacije pomoći će nam odgovarajuća usporedba interpretacija. Na prostoru istinitosnih vrijednosti uvest ćemo usporedbu $<$ koja odgovara ideji informativnosti: istina i laž su informativnije od neodređenosti:

$| < \top, \; | < \bot$

Lako se uvjeriti da je ovo parcijalni uređaj na skupu istinitosnih vrijednosti za kojeg popratnu relaciju $\leq$ definiramo na standardni način:

$x \leq y \leftrightarrow x < y$

Pomoću usporedbe istinitosnih vrijednosti definiramo i usporedbu $\leq$ interpretacija predikata T:

Za $t_1, t_2 : SLT^* \longrightarrow IV \quad t_1 \leq t_2 \leftrightarrow \forall x \in SLT^* \; t_1(x) \leq t_2(x)$

Jednostavnije rečeno veća interpretacija se podudara s manjom tamo gdje ova poprima vrijednost istinu ili laž no poprima te vrijednosti i u nekim argumentima na kojima je manja interpretacija neodređena (veća interpretacija je informativnija).

Za fiksni bazni model i trovaljanu semantiku istinitosna funkcija I^{t*} ovisi o t. Tako imamo operator na prostoru interpretacija

$t \longmapsto I^{t*}$

Pošto je ideja početnu interpretaciju t predikata T zamijeniti eventualno boljom interpretacijom I^{t*}, to ćemo ovaj operator zvati **operatorom reinterpretiranja**. On ovisi o zadanoj trovaljanoj semantici. No standardne semantike su takve da je u većoj interpretaciji predikata T više rečenica istinito ili lažno. Naime, pogledamo li npr. tablicu istinitosti disjunkcije jake Kleeneove trovaljane semantike, lako se uvjeriti da vrijedi sljedeće: ako disjunkcija ima određenu vrijednost (istinu ili laž), zamjena neodređene vrijednosti jednog člana disjunkcije određenom vrijednosti neće alternirati rezultat. Npr.

$| \vee^{SK} \top = \top \longmapsto \bot \vee^{SK} \top = \top$

Jedina promjena koja se može desiti pri zamjeni neodređenog argumenta određenim je da ako je rezultat do tada bio neodređen, nakon zamjene može postati određen. Npr.

$| \vee^{SK} \bot = | \quad \longmapsto \quad \bot \vee^{SK} \bot = \bot$

Za razliku od te semantike stroga negacija $\neg^{SN}$ nema to svojstvo — zamijenimo li neodređeno istinom rezultat nije više istina nego laž:

$\neg^{SN} \mid = \top \quad \longmapsto \quad \neg^{SN}\top = \bot$

Za operator φ na skupu interpretacija kažemo da je **monoton operator** ako čuva usporedbu interpretacija:

$t_1 \leq t_2 \longrightarrow \ \varphi(t_1) \leq \varphi(t_2)$

Za datu trovaljanu semantiku kažemo da je **monotona semantika** ako je njoj pripadni operator reinterpretacije monoton:

$t_1 \leq t_2 \longrightarrow \ I^{t_1 *} \leq I^{t_2 *}$

Monotona je i sljedeća tzv. **van Fraasenova trovaljana semantika** [vF66] koja se često spominje u literaturi jer čuva klasične logičke istine. U toj semantici se istinitosna vrijednost rečenice jezika $LT(t)$ određuje ispitivanjem njene istinitosti u svim klasičnim jezicima $LT(tc)$, gdje je tc klasična totalna interpretacija predikata T koja proširuje t. Ako je u svim takvim jezicima rečenica istinita (lažna) tada je smatramo istinitom (lažnom) u jeziku LT, dok je u protivnom neodređena.

Teorem 31. *Jaka Kleeneova, slaba Kleeneova i van Fraasenova trovaljana semantika su monotone semantike*

Za van Fraasenovu semantiku to direktno slijedi iz definicije, dok se za Kleeneove semantike to dokaže indukcijom po rekurzivnoj strukturi formula. U svakom koraku indukcije koristi se prethodno ilustrirano svojstvo tablica da na većem ulazu daju veću vrijednost.

Ovakvim modeliranjem problem traženja prave interpretacije predikata T, tj interpretacije t takve da je

$I^{t*} = t,$

se svodi na ispitivanje fiksnih točaka monotonih operatora na parcijalnim uređajima, a što je davno riješen problem. Osnovni rezultat je sljedeći:

Teorem 32. *(Knaster, Tarski)*

Neka je φ monoton operator na parcijalnom uređaju $(D, \leq)$ i neka su a i b elementi od D takvi da je $a \leq b$, $a \leq \varphi(a)$ (za takav a kažemo da je **adekvatan** *za dati operator φ) i $\varphi(b) \leq b$ (za takav b kažemo da je* **zatvoren** *za dati operator φ). Tada vrijedi sljedeće:*

1. *Ako svaki neprazan podskup od D ima infimum tada φ ima fiksnu točku koja je među svim fiksnim točkama povrha a najmanja. Ta fiksna točka je ispod b.*

2. *Ako svaki neprazan podskup od D ima supremum tada φ ima fiksnu točku koja je među svim fiksnim točkama ispod b najveća. Ta fiksna točka je povrh a.*

Ovdje će biti dana tek gruba skica dokaza. Detaljan dokaz se može vidjeti u [Fit86]. Za dokazati prvu tvrdnju razmatra se skup svih zatvorenih elemenata povrh a:

$\{x \in D \mid a \leq x \ i \ \varphi(x) \leq x\}$

Pošto b pripada tom skupu on nije prazan. Po pretpostavci teorema on tako ima infimum za kojeg se pokaže da je tražena fiksna točka. Druga tvrdnja je dualna prvoj, pa joj je i dokaz dualan dokazu prve tvrdnje. Tražena fiksna točka je supremum skupa

$\{x \in D \mid b \geq x \ i \ \varphi(x) \geq x\}$

Svojstva uređaja na prostoru interpretacija $SLT^* \longrightarrow IV$ "naslijeđena" su od uređaja na skupu istinitosnih vrijednosti IV. Lako je vidjeti običnim prebrajanjem podskupova da na skupu istinitosnih vrijednosti IV svaki neprazan podskup S ima infimum (oznaka $\bigwedge S$) a svaki neprazni podskup koji ima gornju među ima i supremum (oznaka $\bigvee S$). Specijalno, takav je svaki lanac u IV. Ta se svojstva prenose i na uređaj na prostoru interpretacija:

Teorem 33. *(svojstva uređaja $\leq$ na $SLT^* \longrightarrow IV$)*

1. *Svaki neprazan podskup ima infimum.*
2. *Postoji najmanja interpretacija (zvat ćemo je $t \mid$).*
3. *Svaki neprazni skup koji ima gornju među ima i supremum.*
4. *Svaki lanac ima gornju među.*

Dokaz :

1. Neka je S neprazan podskup od $SLT^* \longrightarrow IV$. Lako se uvjeriti da je traženi infimum sljedeća funkcija:

   $i(x) = \bigwedge\{t(x) \mid t \in S\}$

2. Iz prethodnog slijedi da i cijeli skup $SLT^* \longrightarrow IV$ ima infimum koji je ujedno i najmanji element. Lako je vidjeti da je to funkcija

   $t \mid (x) = \mid \ \forall x$

3. Neka neprazan podskup S od $SLT^* \longrightarrow IV$ ima gornju među g. Postojanje te međe osigurava ispravnost sljedeće definicije tražene funkcije s

$$s(x) = \begin{cases} | & ako\ \forall t \in S\ t(x) = | \\ t(x) & ako\ za\ neki\ t\ \ t(x) \neq | \end{cases}$$

4. Neka je C lanac u $SLT^* \longrightarrow IV$ Tada svojstvo usporedivosti njegovih elemenata osigurava ispravnost definicije koja daje traženu gornju među (štoviše i supremum):

$$s(x) = \begin{cases} | & ako\ \forall t \in S\ t(x) = | \\ t(x) & ako\ za\ neki\ t\ \ t(x) \neq | \end{cases}$$

■

Iz navedenih definicija je jasno da je postojanje gornje međe u stvari pitanje suglasnosti interpretacija na određenim istinitosnim vrijednostima. Nama će to biti značajno za dvočlani skup. Formalno, kažemo da su interpretacije t_1 i t_2 **suglasne** ako skup $\{t_1, t_2\}$ ima gornju među. Za fiksnu točku t monotonog operatora φ kažemo da je **intrinsična fiksna točka** ako je suglasna sa svakom drugom fiksnom točkom.

Zahvaljujući navedenim svojstvima uređaja na $SLT^* \longrightarrow IV$ pomoću teorema Knaster - Tarskog mogu se dokazati sljedeća svojstva fiksnih točaka monotonih operatora na tom prostoru:

Teorem 34. *(svojstva fiksnih točaka monotonog operatora φ na $SLT^* \longrightarrow IV$)*

1. *Povrh svake adekvatne interpretacije ta ($ta \leq \varphi(ta)$) postoji maksimalna fiksna točka.*

2. *Postoji maksimalna fiksna točka.*

3. *Među svim fiksnim točkama povrh adekvatne interpretacije postoji najmanja fiksna točka.*

4. *Postoji najmanja fiksna točka.*

5. *Među svim fiksnim točkama ispod zatvorene interpretacije tz ($\varphi(tz) \leq tz$) postoji najveća fiksna točka.*

6. *Fiksna točka t je intrinsična ako i samo ako se nalazi ispod svake maksimalne fiksne točke.*

7. *Postoji najveća intrinsična fiksna točka.*

8. *Ako se adekvatna interpretacija t nalazi ispod svih maksimalnih fiksnih točaka, tada je najmanja fiksna točka povrh nje intrinsična.*

Opet će biti dana tek gruba skica dokaza a detaljan dokaz se može vidjeti u [Fit86]

1. Promatra se skup svih adekvatnih interpretacija

   $A = \{t \mid t \leq \varphi(t)\}$

   On je neprazan jer mu po pretpostavci pripada ta. Po svojstvu uređaja na interpretacijama svaki lanac u A ima gornju među za koju se može pokazati da je u A. Po Zornovoj lemi to znači da postoji maksimalni element od A koji je povrh ta. No pokaže se da su maksimalni elementi od A upravo maksimalne fiksna točke i tako se dokaže teorem.

2. Pošto postoji adekvatna interpretacija (to je najmanja interpretacija $t \mid$) to po prethodnoj tvrdnji postoji i maksimalna fiksna točka.

3. Po prvoj tvrdnji za adekvatnu interpretaciju ta postoji maksimalna interpretacija tm koja je povrh ta. Pošto te dvije interpretacije zadovoljavaju uvjete Knaster - Tarskijevog teorema to po njemu postoji i najmanja fiksna točka povrh ta.

4. Primijenimo li prethodnu tvrdnju na najmanju interpretaciju dobijemo i postojanje najmanje fiksne točke.

5. Neka je tz zatvorena interpretacija. Promatra se skup S svih interpretacija koje su ispod nje. Specijalno njemu pripada najmanja interpretacija $t \mid$. Za taj skup se pokaže da mu svaki podskup ima supremum u njemu i da je zatvoren na φ. Time su ispunjene pretpostavke drugog dijela Knaster - Tarskijevog teorema za $t \mid$ i tz pa on daje postojanje najveće fiksne točke ispod tz.

6. Neka je M skup svih maksimalnih fiksnih točaka. Treba pokazati da je fiksna točka i intrinsična ako i samo ako je $i \leq \bigwedge M$. Pretpostavimo da je i intrisična i promatramo za neki $m \in M$ skup $\{i, m\}$. Po intrinsičnosti i postoji $\bigvee\{i, m\}$. Koristeći maksimalnost m pokaže se da je $\bigvee\{i, m\} = m$, a što znači da je i ispod m. Obratno, neka je i ispod svake maksimalne fiksne točke i neka je t neka fiksna točka. Pošto postoji maksimalna fiksna točak m povrh t to je $\{i, t\} \leq m$. Dakle i je suglasna s f, tj. ona je intrinsična fiksna točka.

7. Lako se pokaže da je $\bigwedge M$ zatvorena interpretacija. Po prethodnom sve fiksne točke ispod nje su upravo intrinsične fiksne točke i među njima postoji najveća.

8. Iz pretpostavki tvrdnje i svojstva $\bigwedge M$ slijedi da se može primijeniti Knaster - Tarskijev teorem na t i $\bigwedge M$, a on osigurava da je najmanja fiksna točka povrh t ispod $\bigwedge M$, dakle intrinsična.

Korolar 31. *Fiksne točke Jake Kleeneove, slabe Kleeneove i van Fraasenove trovaljane semantike imaju svojstva navedena u teoremu. Drugim riječima, svaki bazični jezik L koji u svojoj domeni sadrži kodove rečenica proširenog jezika LT se može u svakoj od ovih semantika proširiti do jezika $LT(t)$ kojem je T vlastiti predikat istine.*

Dokaz : Prvi dio slijedi iz monotonosti ovih semantika. Drugi dio slijedi iz tog da su upravo fiksne točke t ovih semantika interpretacije predikata T kao predikata istine jezika $LT(t)$. ∎

Ovim teoremom je riješeno pitanje postojanja raznih proširenja jezika do jezika s vlastitim predikatom istine. Kripke je istaknuo najmanju fiksnu točku. Njegovu intuiciju o učenju pojma istine koja vodi najmanjoj fiksnoj toči formalizira sljedeći opis ove točke. Neka je ta adekvatna interpretacija dane semantike tj. neka je $ta \leq I^{ta*}$ (takva je npr. interpretacija $t \mid$ kojom Kripke i započinje svoju konstrukciju). Za nju definiramo transfinitnom rekurzijom po ordinalima sljedeći niz interpretacija koji po Kripkeu odgovara procesu učenja pojma istine:

$$t_0 = ta$$

$$t_{\alpha+1} = I^{t_\alpha *}$$

$$t_\alpha = \bigvee\{t_\beta \mid \beta \leq \alpha\} \text{ za granični ordinal } \alpha$$

Indukcijom po ordinalima se lako pokaže da je ovaj niz rastući:

$$t_{\alpha+1} \geq t_\alpha$$

Kad bi niz bio striktno rastući imao bi članova koliko i ordinala pa klasa svih članova ne bi tvorila skup već bi bila prava klasa. No to je nemoguće jer je ona podskup skupa svih interpretacija $SLT^* \longrightarrow IV$. Dakle postoji prvi ordinal α, takav da je $t_{\alpha+1} = t_\alpha$, kao i svi sljedeći članovi. No to je fiksna točka date semantike jer je po definiciji niza za taj α

$$I^{t_\alpha *} = t_\alpha$$

Lako se uvjeriti i da je to najmanja fiksna točka date semantike. Napomenimo da se dualnom konstrukcijom može dobiti i najveća fiksna točka ispod zatvorene interpretacije tz:

$$t_0 = tz$$

$$t_{\alpha+1} = I^{t_\alpha *}$$

$$t_\alpha = \bigwedge\{t_\beta \mid \beta \leq \alpha\} \text{ za granični ordinal } \alpha$$

Specijalno, ako ovu konstrukciju započnemo s $\bigwedge M$ dobivamo upravo najveću intrinsičnu fiksnu točku date semantike.

U kontekstu jezika s vlastitim predikatom istine najzanimljivija je analiza samoreferirajućih rečenica. Njih možemo konstruirati na isti način kako se to radi u dvovaljanoj semantici, baptizmom ili dijagonalizacijom. *Baptizam* je jednostavni semantički mehanizam — jeziku L dodajemo konstante koje interpretiramo kao imena kodova rečenica od $LT(t)$. Za bilo koju formulu $A(x)$ s jednom slobodnom varijablom x jezika $LT(t)$ možemo jeziku dodati novu konstantu $\overline{C}$ koja će imenovati upravo rečenicu $< A(\overline{C}) >$:

$$\overline{C}^L =< A(\overline{C}) >$$

Dok je $\overline{C}$ ime te rečenice u samom jeziku znak C ćemo koristiti za njeno ime u metajeziku.

Da bismo razumjeli što ta rečenica kaže moramo znati značenje formule $A(x)$. Ta formula definira trovaljani predikat, tj. funkciju $A^{LT(t)} : DLT \longrightarrow IV$ na sljedeći način:

$$A^{LT(t)}(a) = I^t_{v[a/x]}(A(x))$$

Za $A^{LT(t)}$ kažemo da je **definabilan** formulom $A(x)$. Taj trovaljani predikat određuje značenje te formule. Tako rečenica $A(\overline{C})$ o sebi govori da za nju vrijedi $A^{LT(t)}$.

Ako svaka rečenica S ima u jeziku ime $\overline{S}$ svog koda $< S >$ tada na nivou istinitosnih vrijednosti dobivamo analog klasičnoj dijagonalnoj lemi:

Teorem 35. *(dijagonalna lema u baptističkoj verziji)*

Neka je $\overline{C}$ konstanta jezika L koja imenuje kod rečenice $< A(\overline{C}) >$ jezika $LT(t)$ i neka jezik $LT(t)$ za svaku rečenicu S ima ime $\overline{S}$ njenog koda $< S >$. Tada je

$$I^t(A(\overline{C})) = I^t(A(\overline{A(\overline{C})}))$$

Dokaz : $I^t(A(\overline{C})) = A^{LT(t)}(\overline{C}^L) = A^{LT(t)}(< A(\overline{C}) >) = I^t(A(\overline{A(\overline{C})}))$ ∎

U slučaju da je jezik klasičan ovo je upravo klasična dijagonalna lema koju je naravno zbog dvovaljanosti jezika jednostavnije izreći u samom jeziku tvrdnjom

$A(\overline{C}) \leftrightarrow A(\overline{A(\overline{C})})$

U slučaju jezika s vlastitim predikatom istine nije potrebno da svaki kod rečenice ima svoje ime:

Teorem 36. *(dijagonalna lema u baptističkoj verziji za jezik s vlastitim predikatom istine)*

Neka je $\overline{C}$ ime koda rečenice $< A(\overline{C}) >$ jezika $LT(t)$ s vlastitim predikatom istine T. Tada je

$I^t(T(\overline{C})) = I^t(A(\overline{C}))$

Dokaz : $I^t(T(\overline{C})) = t(\overline{C}^L) = t(< A(\overline{C}) >) = I^t(A(\overline{C}))$ ∎

Prisustvo funkcijskog simbola *dijagonalizacije* daje uniforman način konstrukcije samoreferirajućih rečenica:

Teorem 37. *(dijagonalna lema u dijagonalizacijskoj verziji)*

Neka domena jezika L sadrži za svaku formulu $A(x)$ jezika $LT(t)$ s najviše jednom slobodnom varijablom x njen kod $< A(x) >$, a sam jezik L ime tog koda $\overline{A(x)}$. Nadalje, neka jezik L ima funkcijski simbol D dijagonalizacije jezika $LT(t)$. Tada za svaku formulu $A(x)$ jezika $LT(t)$ s jednom slobodnom varijablom x postoji zatvoren term n takav da je

$n^L =< A(n) >$

Pri tome vrijedi

$I^t(A(n)) = I^t(A(\overline{A(n)}))$

Ako je $LT(t)$ jezik s vlastitim predikatom istine T tada je

$I^t(T(n)) = I^t(A(n))$

Dokaz : Za zatvoreni term uzmemo

$n = D(\overline{A(D(x))})$

Tada je

$n^L = D^L(< A(D(x)) >) =< A(D(\overline{A(D(x))})) >=< A(n) >$

Drugi dio teorema slijedi iz prvog na isti način kao i u baptističkoj verziji.
■

Ove dijagonalne leme omogućavaju nam na jedan ili drugi način, konstrukcije samoreferirajućih rečenica. Radi uniformnosti, bez obzira koju konstrukciju koristili, term kojim je ostvareno samoreferiranje bilježit ćemo $\overline{C}$ gdje će nam C biti metajezično ime samoreferirajuće rečenice $A(\overline{C})$ Takvu situaciju naznačit ćemo sljedećim zapisom:

$C : A(\overline{C})$

U ovakvom označavanju imamo sljedeće uvjete na istinitosnu funkciju:

1. $I(C) = I(A(\overline{C}))$
2. $I(T(\overline{C})) = I(C)$

Prvi uvjet proizlazi iz toga što su C i $A(\overline{C})$ nestandardno i standardno metajezično ime iste rečenice a drugi uvjet proizlazi iz prvog uvjeta i dijagonalne leme za jezik s vlastitim predikatom istine T.

Neka je $LT(t)$ jezik s vlastitim predikatom istine T i neka sadrži Jakog Lašca:

$LL : \neg T(\overline{LL})$.

Tada nam uvjeti istinitosti u jakoj Kleeneovoj semantici SK daju:

$I(LL) = I(\neg T(\overline{LL})) = \neg^{SK} I(T(\overline{LL})) = \neg^{SK} I(LL)$

Dakle

$I(LL) = \neg^{SK} I(LL)$

Po tablici istinitosti za $\neg^{SK}$ rješenje ove jednadžbe za $I(LL)$ je

$I(LL) = |$

Dakle, Lažac je neodređen u svim fiksnim točkama jake Kleeneove semantike.

Analogne jednadžbe možemo postaviti i za ostale paradoksalne rečenice. Postojanje jedinstvenog rješenja znači da rečenica u svim fiksnim točkama ima upravo tu vrijednost. No postojanje više rješenja daje tek moguće istinitosne vrijednosti rečenice u fiksnim točkama. Npr za Istinoljupca

$Is : T(\overline{Is})$

imamo:

$I(Is) = I(T(\overline{Is})) = I(Is)$

Ovu jednadžbu zadovoljavaju sve istinitosne vrijednosti. No trebamo pokazati da zaista postoje fiksne točke u kojima Istinoljubac ima te vrijednosti. Postupak je sljedeći: pokušamo naći adekvatnu interpretaciju u kojoj je on istinit (lažan) jer to osigurava da je istinit (lažan) u najmanjoj fiksnoj točki povrh te interpretacije. Npr. za istinitost se prirodno nameće sljedeća početna interpretacija:

$$t(x) = \begin{cases} \top & za\ x =< Is > \\ | & u\ protivnom \end{cases}$$

Lako se uvjeriti da je ta interpretacija adekvatna:

$$t \leq I^{t*}$$

Samo treba dokazati da kada je $t(x) = \top$ da je tada i $I^{t*}(x) = \top$. No $t(x) = \top$ jedino za $x =< Is >$, a tada je

$$I^{t*}(< Is >) = I^t(Is) = I^t(T(\overline{Is})) = t(< Is >) = \top$$

Tako imamo adekvatnu početnu interpretaciju u kojoj je Istinoljubac istinit, pa je istinit i u svakoj fiksnoj točki povrh ove interpretacije. Na isti način bismo dokazali postojanje fiksnih točaka u kojima je lažan. No pošto je u nekim točkama Istinoljubac istinit, a u nekim lažan to on mora u najmanjoj fiksnoj točki biti neodređen.

Ispitajmo istinitosne vrijednosti u fiksnim točkama rečenice Logičar:

$$Log : T(\overline{Log}) \vee \neg T(\overline{Log})$$

$I(Log) = I(T(\overline{Log}) \vee \neg T(\overline{Log})) = I(T(\overline{Log})) \vee^{SK} \neg^{SK} I(T(\overline{Log})) = I(Log) \vee^{SK} \neg^{SK} I(Log)$

Dakle istinitosna vrijednost Logičara zadovoljava sljedeću jednadžbu:

$$I(Log) = I(Log) \vee^{SK} \neg^{SK} I(Log)$$

Običnom provjerom možemo utvrditi da su moguća rješenja $\top$ i $|$. Analogno Istinoljupcu, konstruiranjem adekvatne početne interpretacije t

$$t(x) = \begin{cases} \top & za\ x =< Log > \\ | & u\ protivnom \end{cases}$$

možemo pokazati postojanje fiksne točke u kojoj je Log istinit. No da bismo pokazali da je on istinit u najvećoj intrinsičnoj fiksnoj točki po teoremu o fiksnim točkama moramo pokazati da je t ispod svih maksimalnih fiksnih točaka:

$t \leq m$, za bilo koju maksimalnu fiksnu točku.

No to znači da moramo pokazati da je Logičar istinit u svim maksimalnim fiksnim točkama. Neka je m maksimalna fiksna točka u kojoj je Log neodređen. Pomoću nje ćemo definirati novu interpretaciju mL

$$mL(x) = \begin{cases} m(x) & za\ x \neq < Log > \\ \top & za\ x = < Log > \end{cases}$$

Lako se uvjeriti da je mL adekvatna interpretacija u kojoj je Log istinit. No ona se tada može proširiti do fiksne točke mLf. Tako imamo proširenja

$$m \leq mL \leq mLf$$

No m je maksimalna fiksna točka pa mora biti $mLf = m$. Dakle $mL = m$, a to je kontradikcija jer je Log u mL istinit dok je po pretpostavci u m neodređen. Dakle, Log mora biti istinit u svakoj maksimalnoj fiksnoj točki. To ujedno znači da je on istinit i u najvećoj intrinsičnoj fiksnoj točki. Nasuprot tome on je neodređen u najmanjoj fiksnoj točki što se može dokazati indukcijom po nivoima t_α najmanje fiksne točke.

Konstrukcija fiksne točke je nadopuna početnog klasičnog jezika u kojem nije definabilan njegov predikat istine, do jezika s vlastitim predikatom istine. No u ovim jezicima se javljaju nove nedefinabilnosti. Za daljnje razmatranje će pogotovo biti značajna nedefinabilnost neodređenosti:

Teorem 38. *(o nedefinabilnosti neodređenosti u fiksnim točkama)*

U jeziku $LT(t)$ s vlastitim predikatom istine T nije definabilna neodređenost, tj. ne postoji formula $N(x)$ s jednom slobodnom varijablom takva da je istinita na kodovima neodređenih rečenica, dok je na ostalim objektima lažna.

Dokaz : Definabilnost neodređenosti vodila bi paradoksu Trovaljanog Lašca tj. prisustvu kontradiktorne rečenice

Ja nisam istinita ili sam neodređena

$$Lt : \neg T(\overline{Lt}) \vee N(\overline{Lt})$$

Zaista za istinitost ove rečenice vrijedi

$$I(Lt) = I(\neg T(\overline{Lt}) \vee N(\overline{Lt})) = \neg^{SK} I(T(\overline{Lt})) \vee^{SK} I(N(\overline{Lt})) = \neg^{SK} I(Lt) \vee^{SK} I(N(\overline{Lt}))$$

Tako dobivamo uvjet

$$I(Lt) = \neg^{SK} I(Lt) \vee^{SK} I(N(\overline{Lt}))$$

Kad bi bilo $I(N(\overline{Lt}) = \top$ po gornjem bismo uvjetu dobili da je $I(Lt) = \top$ a po smislu predikata N da je $I(Lt) =|$. No kad bi bilo $I(N(\overline{Lt}) = \bot$ po gornjem bismo uvjetu dobili da je $I(Lt) =|$ a po smislu predikata N da $I(Lt) \neq|$. Dakle dobivamo kontradikciju, što znači da jezik nema takav predikat. ■

3.3 Otvorena pitanja Kripkeove analize pojma istine i kritike

Kripke je pokazao da je u prirodnom govoru većina rečenica koje sadrže pojam istine riskantna. Pod određenim uvjetima one ne posjeduju istinitosnu vrijednost. No ti uvjeti se ne mogu izolirati na sintaktičkom ili unutrašnjem semantičkom nivou jezika, već ovise o vanjskoj semantičkoj strukturi jezika, ili kako Kripke kaže, o empiričkim činjenicama. *Zbog toga teorija istine ne može napraviti restrikciju na T - shemu već je treba prihvatiti, ali tako mora prihvatiti i riskantnost rečenica, mogućnost da pod nekim uvjetima rečenica nema istinitosnu vrijednost nego je neodređena. Vidi se da je to sasvim suprotan tip rješenja u odnosu na Tarskijevo rješenje gdje je restringirana shema T a zadržana klasična logika.* To vodi proučavanju jezika s trovaljanom semantikom. Kripke nije dao određeni model već teorijski okvir za ispitivanje raznih modela, a što je i sam eksplicitno naveo. *Svaka fiksna točka svake monotone trovaljane semantike može biti model za pojam istine. Svaki takav izbor daje normativno rješenje. Dok se u Tarskijevom rješenju problematične rečenice ne mogu izreći, ovdje se one mogu izreći, a kontradikcija se izbjegava tako da se one proglašavaju neodređenim. No, nijedno od Kripkeovih rješenja ne može biti i analitičko rješenje, dok se ne priloži analiza koja će sadržajno pokazati da upravo to rješenje modelira pojam istine.* Naravno, moguće je da taj pojam nosi više intuicija kojima odgovara više modela i da će takva analiza dati više rješenja.

U tom smjeru Kripke je napravio nekoliko koraka. *Preferirao je jaku Kleeneovu trovaljanu semantiku* za koju je napisao da je “odgovarajuća” no nije obrazložio i po čemu je odgovarajuća. Jedan razlog za takav izbor je vjerojatno Kripkeovo opredijeljenje da *paradoksalne rečenice smatra smislenim.* To eliminira slabu Kleeneovu semantiku koja odgovara zamisli da su paradoksalne rečenice besmislene, pa tako i neodređene. Ako je dio rečenice besmislen tada je i cijela rečenica besmislena. Prevedeno na neodređenost to daje upravo slabu Kleeneovu semantiku. Drugi razlog bi mogao biti da Jaka Kleeneova trovaljana semantika ima tzv. *investigativnu interpretaciju.* Po toj interpretaciji ona odgovara klasičnom utvrđivanju istinitosti pri čemu sve rečenice koje nemaju već utvrđenu vrijednost privremeno smatramo neodređenim. Njihova neodređenost po uvjetima istinitosti veznika i kvantora povlači da su neodređene i neke druge, od njih složene rečenice. Kada tim rečenicama odredimo istinitosnu vrijednost, tada možemo odrediti i istinitosnu vrijednost rečenica koje su od njih složene. Ovaj izbor trovaljane semantike je pogotovo adekvatan za jedinu intuiciju o pojmu istine koju je Kripke dao, a to je *intuicija o učenju ili utvrđivanju istinitosnih vrijednosti koja daje najmanju fiksnu točku.* Ta intuicija se bavi time kako nekoga tko je kompetentan korisnik početnog jezika (bez predikata istine T) naučiti koristiti rečenice koje sadrže predikat T. Taj čovjek zna koje su rečenice početnog jezika istinite a koje ne. Dajemo mu pravilo da prvima pridruži atribut T, a drugima negira taj atribut. Time mu postaju određene neke nove rečenice koje sadrže predikat istine i koje su do tada bile neodređene. To povlači za sobom i određivanje istinitosti nekih od njih složenih rečenica, koje su do tada bile neodređene. Prirodan način ovog određivanja je upravo Jaka Kleeneova trovaljana semantika kao formalizacija investigativne semantike. Tako subjekt dobiva novi skup istinitih i lažnih rečenica kojima pridružuje odnosno negira atribut T. Pošto se ovaj proces obično zasićuje tek na nekom transfinitnom ordinalu ova intuicija penjanjem po ordinalima sve više prelazi u metaforu. Naime *teško je zamisliti da u određivanju istinitosti neke rečenice polazimo od rečenica koje ne sadrže taj pojam i čiju istinitost moramo znati da bismo penjanjem po ordinalima utvrdili istinitosnu vrijednost početne rečenice, ili pak da uopće nema tu vrijednost.* Realnije je na osnovu razumijevanja najmanje fiksne točke pokušati zaključiti pripada li ta rečenica datoj fiksnoj točki i ako pripada koja joj je istinitosna vrijednost. No Kripke ne razrađuje takav postupak.

Što se tiče ostalih fiksnih točaka *Kripke prije svega koristi strukturu skupa svih fiksnih točaka da bi razlikovao razne tipove problematičnih (neutemeljenih) rečenica.* Neki (npr.[Kre88]) smatraju da je upravo to Kripkeovo rješenje — analiza rečenica pomoću svih fiksnih točaka. No to bi značilo da je pojam istine heterogen — neke rečenice su u nekim fiksnim točkama istinite a u drugim lažne. Pogotovo je teško zamisliti situaciju u kojoj korisnik jezika koristi sve fiksne točke, dakle razne interpretacije jezika da bi utvrdio istinitosnu vrijednost neke rečenice. *Ako je cilj jedna fiksna točka koja bi trebala biti model pojma istine, onda proučavanje svih fiksnih točaka ima smisla samo dok nije pronađen pravi model.* S te strane je upitna i Kripkeova klasifikacija neutemeljenih rečenica na osnovu svih fiksnih točaka.

Ispitujući strukturu fiksnih točaka *Kripke je istaknuo najveću intrinsičnu fiksnu točku zbog njenog istaknutog položaja u strukturi fiksnih točaka* kao najveću interpretaciju koja daje istinitosne vrijednosti koje su sukladne s istinitosnim vrijednostima u svim drugim fiksnim točkama. No *ona je istaknuta po svojim formalnim svojstvima i nije dana nikakva sadržajna analiza koja bi vodila toj fiksnoj točki.*

Iz navedenog se može zaključiti da je Kripke izolirao jedan bitan aspekt pojma istine, njegovu riskantnost, te je dao okvir za ispitivanje raznih normativnih rješenja. Ako se prihvati taj okvir preostaje problem utvrđivanja koje rješenje modelira intuiciju o istini. Da bi takva analiza bila vjerodostojna, ona mora davati i analitičko rješenje, mora objasniti uzrok paradoksa istine.

Kripkeova analiza, pogotovo analiza pojma istine kao najmanje fiksne točke primjenjiva je i na situacije na koje Tarskijeva analiza nije primjenjiva i koje su sveprisutne u svakodnevnom govoru, a to su situacije u kojima postoji *cirkularnost* u referiranju jednih rečenica na istinitost drugih rečenica. U hijerarhijskom pristupu takve rečenice moramo zabraniti, tj ne možemo ih uopće izreći u okviru hijerarhije jezika (sjetimo se Lašca ili pak Deana i Nixona). Ona nadalje zadržava osnovnu intuiciju o istini sadržanu u shemi T, ali jezik smatra trovaljanim.

Upravo ovo zadnje svojstvo uzrok je jednog dijela nezadovoljstva takvim modelima. Često se to izražava kroz *kritiku da u fiksnim točkama ne vrijede svi klasični zakoni logike*. Npr. za malo bogatiji jezik (koji npr. sadrži Lašca) u svakoj fiksnoj točki jake Kleeneove semantike je tvrdnja $\forall x\ \neg(T(x) \wedge \neg T(x))$ neodređena. Zato se preferira van Fraasenova supervaluacija koja čuva logičke istine klasične logike. No, takvi rezultati su prirodna posljedica prvotnog određenja — *ne možemo očekivati da će za trovaljani jezik vrijediti logički zakoni dvovaljanog jezika*. Ako je prihvaćen trovaljani jezik mora biti prihvaćena i shema zaključivanja koje taj jezik nosi bez obzira na našu nenaviknutost na nju. Štoviše, u jakoj Kleeneovoj semantici logičke istine dvovaljane logike su uvijek istinite kada su određene.

Nešto nezgodnija situacija je da navedeni zakon $\forall x\ \neg(T(x)\wedge\neg T(x))$*, kao i drugi logički zakoni, nije istinit u najmanjoj fiksnoj točki ni kada nema Lašca ili Istinoljupca, tj. kada nema paradoksalnih rečenica niti rečenica koje u raznim fiksnim točkama poprimaju razne vrijednosti*. Naime, tada navedena tvrdnja nije istinita zbog sebe same! Jer da bi se utvrdila njena istinitost moraju se ispitati istinitosti svih rečenica pa tako i nje same. Na taj način se može vidjeti da je ona neutemeljena rečenica, dakle neodređena u najmanjoj fiksnoj točki. *No u najvećoj neproizvoljnoj fiksnoj točki ona je istinita. Tako ova kritika prelazi u argument za najveću intrinsičnu fiksnu točku.*

Drugi tip kritike nastoji pokazati da neka sasvim intuitivna razmatranja o pojmu istine nisu u skladu s modelom fiksne točke. Sljedeći primjer, kao i prethodni, je Guptin [Gup82]. Neka imamo sljedeće izjave osoba A i B (**Guptin paradoks**)

Izjave osobe A:

(a1) Dva plus dva je tri. (laž)

(a2) Snijeg je uvijek crn. (laž)

(a3) Sve što B kaže je istina. ( )

(a4) Deset je prost broj. (laž)

(a5) Nešto što B kaže nije istina. ( )

Izjave osobe B:

(b1) Jedan plus jedan je dva. (istina)

(b2) Zovem se B. (istina)

(b3) Snijeg je ponekad bijel. (istina)

(b4) Najviše jedna stvar koju A kaže je istinita. ( )

U najmanjoj fiksnoj točki rečenice (a1), (a2), (a4), (b1), (b2) i (b3) su određene već na nultom nivou. No (a3) i (a5) "čekaju" (b4), a (b4) čeka njih i tako ostaju neodređene. No na intuitivnom nivou im je sasvim jednostavno odrediti istinitosnu vrijednost. Pošto su (a3) i (a5) kontradiktorne, a sve druge izjave osobe A lažne, to je (b4) istinita. No to znači da je (a3) istinito, a (a5) lažno. No, *ova se intuicija poklapa s vrednovanjem u najvećoj fiksnoj točki*! Da bi i za nju našao intuitivni kontraprimjer Gupta zamjenjuje (a3) i (a5) sljedećim izjavama:

(a3*) (a3) je istinita. ( )

(a5*) "(a3) nije istinita" je istinita ( )

Sada u maksimalnoj fiksnoj točki ni (a3*) i (a5*) pa tako ni (b4) nemaju istinitosnu vrijednost. Gupta smatra da je na intuitivnom nivou (b4) istinita, jer je najviše jedna od navedenih tvrdnji istinita. No u intuitivnom argumentu se desio *semantički pomak* koji tu kritiku smješta u jedan sasvim drugi kontekst koji je predmet sljedećeg paragrafa.

Paradigmatski primjer kritike koja koncepciji fiksne točke suprotstavlja određenu intuitivnu argumentaciju je Jaki Lažac:

LL: Rečenica LL nije istinita.

Ova rečenica ne može biti ni istinita ni lažna. Po koncepciji fiksne točke ona je neodređena. Ali to znači da nije istinita, i paradoks se obnavlja. Ova argumentacija je ipak u jednom dijelu pogrešna. U zadnjem koraku se desio *semantički pomak*. Zadnji pojam istine nije više onaj prvotni. Prvotni pojam istine je pojam istine unutar objektnog jezika (fiksne točke). Zaključak da je data rečenica neodređena je zaključak u metajeziku, nastao refleksijom nad onim što se zbiva u fiksnoj točki. Naime, zaključak *da rečenica LL nije istinita u objektnom jeziku je istina metajezika a ne objektnog jezika i zbog toga se paradoks ne obnavlja*. Desio se prijelaz iz fiksne točke u refleksiju nad fiksnom točkom, ono što je Kripke nazvao "dovršavanjem" predikata istine, odnosno prelazak s primarne intuicije na alternativnu intuiciju, nastalu refleksijom nad primarnom intuicijom. *Zato nema paradoksa već jedino ostaje pitanje treba li ostati u objektnom jeziku (fiksnoj točki) ili pak nastaviti razmišljanje u metajeziku, dvovaljanom opisu onoga što se dešava u fiksnoj točki.* Isto razmišljanje možemo ponoviti i s običnim Lašcem. Utvrdivši da je Lažac neodređen mogli bismo utvrditi da je lažan jer nije istina ono što govori, i tako obnoviti paradoks. No ni ovdje nema obnavljanja paradoksa jer se desio semantički pomak: u metajeziku je laž da je Lažac lažan u jeziku. Na isti način možemo razriješiti i Metalašca i Intenzionalnog lašca.

Metalažac:

1. Rečenica na liniji 1 nije istinita.
2. Rečenica na liniji 1 nije istinita.

Dok se na intuitivnom nivou čini da je ista rečenica i neodređena i istinita paradoksa nema jer se opet desio semantički pomak: Ista rečenica je neodređena u jeziku a istinita u metajeziku.

Intenzionalni Lažac:

(1) Rečenica (1) nije istinita

Dok je ova rečenica neodređena zamijenimo li (1) njenom kvotacijom čini se da dobijamo istinitu rečenicu:

Rečenica "Rečenica (1) nije istinita" nije istinita.

No nije u pitanju nikakva intenzionalnost nego isti semantički pomak kao i u prethodnom slučaju. Na isti način se može objasniti i modificirani Guptin paradoks. Dakle, nije u pitanju obnavljanje paradoksa u fiksnoj točki nego je to eventualno argument za prelazak s prvotne intuicije na sekundarnu. Argument za to je i logička istina dvovaljane semantike $\forall x\, \neg(T(x) \wedge \neg T(x))$. Ona je obično neodređena u fiksnoj točki, no spremni smo je prihvatiti kao istinitu. Ona i jeste istinita, ali u metajeziku.

Sljedeća vrsta kritike tiče se *univerzalnosti jezika. U fiksnoj točki je postignuta zatvorenost jezika na vlastitu istinitost, ali se zato javljaju nove slabosti jezika.* Prije svega *jezik je nepotpun u bulovskom smislu.* Pošto su svi njegovi veznici monotoni, u njemu mogu biti definabilni jedino monotoni veznici. Tako ne mogu biti definirani npr. Lukasiewizcev bikondicional ili pak ekskluzivna negacija. Ako bismo dodali neki od tih veznika jezik više ne bi bio monoton i ne bismo više imali garanciju da ga možemo upotpuniti do fiksne točke. Po meni *to nije nedostatak Kripkeove teorije jer Kripke neodređenost ne smatra trećom vrijednošću nego nedostatkom istinitosne vrijednosti. Ti veznici naprosto ne pripadaju primarnoj intuiciji o pojmu istine.* Kripkeu se jedino može prigovoriti da je premalo analizirao svoju upotrebu trovaljanih shema. Npr. u intuiciji najmanje fiksne točke postoji određeni konflikt između neodređenosti u samoj konstrukciji (ne još određeno) i neodređenosti nakon završetka konstrukcije (nikad više određeno). Na isti način se Kripkeova teorija može braniti i od *pojave novih nedefinabilnosti u fiksnoj točki.* Laž je definabilna negacijom predikata istine ali, kao što je pokazano, *neodređenost nije definabilna.* Pošto fiksna točka nema taj pojam ne možemo npr. Trovaljanog Lašca

Ova rečenica je lažna ili neodređena

ni izraziti. To je u skladu s prvotnom intuicijom, po kojoj utvrđujemo istinitost i lažnost rečenica i tek refleksijom nad tim procesom zaključujemo da su te rečenice neodređene. *Taj pojam pripada metajeziku kojim opisujemo fiksnu točku. Zbog toga ni Trovaljani Lažac ne vodi paradoksu nego semantičkom pomaku u metajezik.* Mada ovakvi pojmovi na intuitivnom nivou i ne pripadaju fiksnoj točki, pa se na taj način može braniti njihova nedefinabilnost, oni se prirodno javljaju u refleksiji nad tim jezikom. I čini se umjetnim zabraniti taj prijelaz u metajezik, koji se, kako vidimo i u paradoksima istine, često spontano zbiva. Po meni, *logička analiza pojma istine ima prirodno zaokruženje u metajeziku kojim je opisana fiksna točka, u dvovaljanom opisu odgovarajuće fiksne točke odgovarajuće trovaljane semantike. Konstrukcija je opisana metajezikom i samo vrednovanje rečenica* *se odvija u metajeziku. Fiksna točka je model procesa utvrđivanja istinitosti u primarnoj semantici jezika, a koja se pokazuje trovaljanom. Teorija istine je dvovaljani opis tog modela.* Naravno, pošto je to dvovaljani opis on potpada pod Tarskijev teorem nedefinabilnosti. Pojam istinite rečenice ovog opisa ne podudara se s pojmom istine fiksne točke koju opisuje. To je istinitosno vrednovanje rečenica kojima opisujemo istinitosne vrijednosti rečenica u fiksnoj točki. No i nije bio cilj opisati predikat istine metajezika, već pomoću njega opisati predikat istine fiksne točke. Situacija je analogna npr. biološkom opisu nekog životnog staništa. Istinitosne vrijednosti nam služe za vrednovanje rečenica tog opisa i nisu same sebi svrha. Ista je situacija i ovdje, samo što one umjesto biološkog opisa vrednuju opis fiksne točke. U tom smislu je taj opis prirodno zaokruženje rješenja početnog problema, logičke analize pojma istine u datom jeziku. U sljedećem je poglavlju napravljen pokušaj argumentiranja za prihvatanje određene fiksne točke određene trovaljane semantike za model pojma istine, te je argumentirano kako se taj model može prirodno nadopuniti da vlastitog dvovaljanog opisa.

POGLAVLJE 4

Modeliranje pojma istine pomoću najveće intrinsične fiksne točke jake Kleeneove trovaljane semantike

Na osnovu elementarnih razmatranja o funkcioniranju jezika dano je analitičko rješenje paradoksa istine i zasnovana odgovarajuća primarna semantika predikata istine. Pokazano je da je to upravo najveća intrinsična fiksna točka jake Kleeneove trovaljane semantike. Ova semantika je nadopunjena do konačne dvovaljane semantike jezika kao prirodnog okruženja za izražavanje i rezoniranje o predikatu istine primarne semantike ([Č01]).

4.1 Jedna intuitivna analiza pojma istine i paradoksa istine

Ugrubo, pod **klasičnim jezikom** podrazumijevat će se svaki jezik modeliran po uzoru na svakodnevni jezik deklarativnih rečenica. Tu spada na primjer standardni matematički jezik koji je u osnovi svakodnevni jezik obogaćen simbolizacijom i mehanizmom upotrebe varijabli. Radi određenosti, razmatrat će se jezik logike prvog reda, jezik s eksplicitnim i preciznim opisom forme i značenja. Pod jezikom će se uglavnom podrazumijevati **interpretirani jezik**, tj. jezična forma zajedno s interpretacijom.

Pored određene formalne (gramatičke) strukture i unutrašnje značenjske strukture jezik ima i svoju vanjsku značenjsku strukturu, vezu jezičnih formi i vanjezičnih elemenata — onoga o čemu jezik govori. Ta veza počiva na određenim *vanjezičnim pretpostavkama* o upotrebi jezika. Za klasični jezik to su pretpostavke da postoje objekti o kojima jezik govori, da je svako ime ime nekog objekta, da svakom funkcijskom i relacijskom simbolu odgovara neka operacija odnosno veza među objektima, a svaka atomarna rečenica je istinita ili lažna, ovisno o tome da li "jeste to" ili "nije to" o čemu ona govori. Te pretpostavke su iznikle iz svakodnevne upotrebe jezika gdje smo naviknuti na njihovu ispravnost, no moguće su situacije u kojima one nisu ispunjene. Paradoks lašca i drugi paradoksi istine svjedoci su jedne takve situacije. Oni su rezultat tenzije između podrazumijevanih pretpostavki jezika i njihovog neispunjenja.

Promotrimo rečenicu L (**Lažac**):

L: L je lažna rečenica. (ili "Ova rečenica je lažna.")

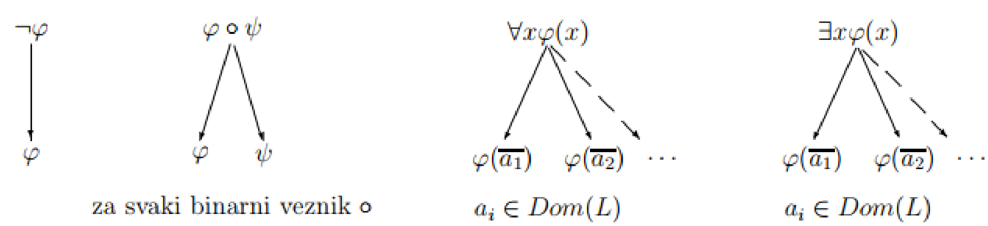
Koristeći se svakodnevnim poimanjem istine i upotrebe jezika, da utvrdimo je li L istinita moramo pogledati je li istina što ona govori. No ona govori upravo o vlastitoj istinitosti i to na kontradiktoran način. Ako pretpostavimo da je istinita, tada je istina što govori — da je lažna. Ako pak pretpostavimo da je lažna tada je laž što govori — da je lažna, pa je istinita. Rečenica je tako samokontradiktorna. No ono što je još značajnije je paradoksalan osjećaj da joj ne možemo odrediti istinitost. Istu paradoksalnost, mada ne i kontradiktornost, dobivamo kod sljedeće rečenice I (**Istinoljubac**):

I: I je istinita rečenica. (ili "Ova rečenica je istinita.")

Za razliku od Lašca kojem ne možemo pridružiti ni istinu ni laž, ovoj rečenici na jednako (ne)uvjerljiv način možemo pridružiti i istinu i laž. Nema nekih dodatnih određenja koji bi uradili izbor između dviju mogućnosti, ne zbog našeg neznanja nego "u principu". Tako ni ovoj rečenici ne možemo odrediti istinitost.

Tarskijeva analiza je istaknula T shemu kao osnovnu vezu istinitosne vrijednosti rečenice koja govori o istinitosti druge rečenice i istinitosne vrijednosti rečenice o kojoj ona govori. No zadržavanje klasične logike zahtjeva restrikciju te sheme. Kripkeova analiza je pokazala da su rečenice koje sadrže pojam istine u svakodnevnom govoru riskantne, te da o empiričkim činjenicama ovisi vode li normalnoj situaciji ili paradoksima. Ne postoji prirodna restrikcija na T - shemu. T - shemu treba zadržati kao osnovni intuitivni princip kojeg koristimo u svakodnevnom govoru a riskantnost treba prihvatiti — moguće je da rečenica u određenoj situaciji nema istinitosnu vrijednost. Analiza koja će ovdje biti provedena bavit će se pitanjem zašto i kako se dešava da neke rečenice nemaju istinitosnu vrijednost. Pri tome se polazi od osnovne intuicije da su prethodne rečenice smislene (jer dobro razumijemo što govore, štoviše to smo i koristili u neuspjelom određivanju njihove istinitosti), ali svjedoče o neuspjehu klasične procedure određivanja istinitosti u nekim "krajnjim" situacijama. Paradoksalnost je rezultat sukoba pretpostavke o uspjehu procedure i otkrića neuspjeha.

Osnovna pretpostavka klasičnog jezika je da je svaka rečenica istinita ($\top$) ili lažna ($\bot$). Istinitost složenijih rečenica određena je po unutrašnjoj semantici jezika istinitošću jednostavnijih rečenica. Da bismo bolje predočili tu semantičku povezanost rečenica zamislit ćemo ih kao čvorove grafa S i od svake rečenice povući strelice prema svim rečenicama o kojima ovisi njena istinitost. Radi određenosti razmatrat ćemo rečenice interpretiranog jezika prvog reda L. Radi jednostavnosti izlaganja pretpostavit ćemo da za svaki objekt a domene jezika $dom(L)$ postoji zatvoreni term jezika $\bar{a}$ koji ga imenuje. Standardni logički vokabular jezika L tvorit će veznici $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ i kvantori $\forall, \exists$. Strelice semantičkog grafa S definiramo rekurzijom po induktivnoj strukturi rečenica. No umjesto striktnih definicija poslužit ćemo se "slikama" tipičnih čvorova grafa:



Istinitost rečenica možemo opisati kao funkciju $I : S \longrightarrow \{\top, \bot\}$, koja svakoj rečenici pridružuju njenu istinitost, $\top$ ili $\bot$.

Unutrašnja semantika jezika opisuje određivanje istinitosti složenije rečenice pomoću istinitosti jednostavnijih rečenica na koje pokazuje (strelicama grafa). Te opise možemo izreći kao standardne uvjete na funkciju istinitosti I:

1. $I(\neg\varphi) = \begin{cases} \top & za\ I(\varphi) = \bot \\ \bot & za\ I(\varphi) = \top \end{cases}$

2. $I(\varphi \wedge \psi) = \begin{cases} \top & za\ I(\varphi) = \top\ i\ I(\psi) = \top\ (obje\ su\ istinite) \\ \bot & za\ I(\varphi) = \bot\ ili\ I(\psi) = \bot\ \text{(barem je jedna lažna)} \end{cases}$

3. $I(\varphi \vee \psi) = \begin{cases} \top & za\ I(\varphi) = \top\ ili\ I(\psi) = \top\ (barem\ jedna\ je\ istinita) \\ \bot & za\ I(\varphi) = \bot\ i\ I(\psi) = \bot \text{ (obje su lažne)} \end{cases}$

4. $I(\varphi \to \psi) = \begin{cases} \top & za\ I(\varphi) = \bot\ ili\ I(\psi) = \top \\ \bot & za\ I(\varphi) = \top\ i\ I(\psi) = \bot \end{cases}$

5. $I(\varphi \leftrightarrow \psi) = \begin{cases} \top & za\ I(\varphi) = I(\psi) \text{ (obje su istinite ili obje su lažne)} \\ \bot & za\ I(\varphi) \neq I(\psi) \text{ (jedna je istinita a druga lažna)} \end{cases}$

6. $I(\forall x \varphi(x)) = \begin{cases} \top & ako\ \forall a \in dom(L)\ I(\varphi(\bar{a})) = \top \\ \bot & ako\ \exists a \in dom(L)\ I(\varphi(\bar{a})) = \bot \end{cases}$

7. $I(\exists x \varphi(x)) = \begin{cases} \top & ako\ \exists a \in dom(L)\ I(\varphi(\bar{a})) = \top \\ \bot & ako\ \forall a \in dom(L)\ I(\varphi(\bar{a})) = \bot \end{cases}$

Po tim uvjetima, da bismo odredili istinitost dane rečenice moramo pogledati istinitost svih rečenica na koje ona pokazuje, zatim eventualno iz istih razloga i istinitost rečenica na koje one pokazuju itd. Svaki takav put po strelicama grafa dolazi do atomarnih rečenica (jer složenost rečenica duž puta opada) i pitanje istinitosti početne rečenice posve je određeno istinitošću atomarnih rečenica na koje ona hereditarno pokazuje. U svakodnevnim situacijama jezik ne govori o istinitosti vlastitih rečenica, pa istinitost atomarnih rečenica ne ovisi o istinitosti nekih drugih rečenica. One su listovi semantičkog grafa S — od njih ne vode strelice prema drugim rečenicama. Za ispitati njihovu istinitost moramo pogledati vanjezičnu stvarnost o kojoj govore. Na primjer, da bismo utvrdili istinitost najjednostavnije atomarne tvrdnje $P(\bar{a})$ moramo vidjeti ima li objekt a svojstvo P. Pretpostavka klasičnog jezika je da to ili jeste ili nije, tj. da je $P(\bar{a})$ ili istinita ili lažna. Ta pretpostavka je u standardnim situacijama ispunjena, bilo efektivno bilo "u principu". Tako svaka atomarna rečenica ima točno određenu istinitosnu vrijednost pa procedura određenja istinitosti bilo koje rečenice daje jednoznačan rezultat, $\top$ ili $\bot$. Formalno to osigurava princip rekurzije koji kaže da postoji jedinstvena funkcija $I : S \longrightarrow \{\top, \bot\}$ čije se vrijednosti na atomarnim rečenicama podudaraju s vanjezično zadanim vrijednostima i koja zadovoljava navedene klasične semantičke uvjete.

Ta klasična situacija može biti (i jeste) narušena kada atomarne rečenice govore o istinitosti drugih rečenica. Tada postoje strelice s atomarnih rečenica na te rečenice duž kojih nastavljamo određivanje istinitosti početne rečenice. Najjednostavniji takav slučaj je kada jezik sadrži **predikat istinitosti** T pomoću kojeg može govoriti o istinitosti vlastitih rečenica. Tada jeziku pripadaju atomarne tvrdnje oblika $T(\bar{\varphi})$ sa značenjem "φ je istinita rečenica". *Uvjet istinitosti $T(\bar{\varphi})$ dio je unutarnje semantike jezika, kao npr. i uvjeti za $\varphi \wedge \psi$. Ta istinitost ne ovisi o vanjskom svijetu već o istinitosti rečenice φ po logičkom smislu koji pripisujemo predikatu T, a taj je da $T(\bar{\varphi})$ smatramo istinitom kada je φ istinita, a lažnom kada je φ lažna.* Tako u slučaju prisustva predikata istinitosti T imamo nove strelice u grafu

$$\begin{tikzcd} T(\bar{\varphi}) \arrow[d] \\ \varphi \end{tikzcd}$$

i novi uvjet na funkciju istinitosti

$$I(T(\bar{\varphi})) = \begin{cases} \top & za\ I(\varphi) = \top \\ \bot & za\ I(\varphi) = \bot \end{cases}$$

No sada, da bismo ispitali istinitost neke rečenice, općenito govoreći, nije dovoljno problem svesti na istinitost atomarnih rečenica, već moramo nastaviti "putovanje" po strelicama ponovo na složenije rečenice. Kako su pri tom moguća "kruženja" ništa nam više ne osigurava da će naša procedura određenja istinitosti uspjeti. Paradoksi istine upravo svjedoče takve slučajeve. Slijede tri ilustrativna primjera:

$$\begin{tikzcd} T(\overline{1+1=3}) \arrow[d] \\ 1+1=3 \end{tikzcd}$$

Procedura određenja istinitosti je stala na atomarnoj tvrdnji za koju znamo da je lažna pa je tako i $T(\overline{1+1=3})$ lažna.

Lažac: Za $L : T(\overline{\neg L})$ imamo

$$\begin{tikzcd} L :\ T(\overline{\neg L}) \arrow[d] \\ \neg L \arrow[u, bend left] \end{tikzcd}$$

No sada procedura određenja istinitosti propada jer se uvjeti na funkciju istinitosti pokazuju neispunjivi. Istinitost $T(\overline{\neg L})$ ovisi o $\neg L$ a ova opet o $L : T(\overline{\neg L})$ na način za koji smo vidjeli da ne može biti ispunjiv.

Istinoljubac: Za $I : T(\bar{I})$ imamo

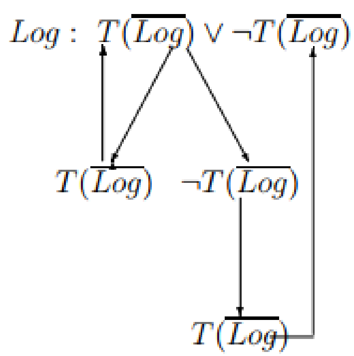
$$I : T(\overline{I})$$

Sada postoji, kako smo vidjeli, više mogućih pridruživanja istinitosne vrijednosti rečenici I. No i tu višestruku ispunjivost moramo smatrati neuspjehom klasične procedure određenja istinitosti koja pretpostavlja ustanovljavanje jedinstvene istinitosne vrijednosti za svaku rečenicu.

Paradoksi upravo proizlaze iz toga što klasična procedura određenja istinitosti ne mora uvijek dati i klasično pretpostavljeni (i očekivani) odgovor. Kako prethodni primjeri pokazuju, takva pretpostavka je neopravdano poopćenje uobičajenih situacija na sve situacije. Možemo zadržati klasičnu proceduru određenja istinitosti, pa tako i unutrašnju značenjsku strukturu jezika iz koje ona proizlazi, ali moramo odbaciti univerzalnost vanjezične pretpostavke o uspješnom završetku procedure. Svijest o tome pretvara paradokse u normalne situacije inherentne klasičnoj proceduri. Po meni, to i jeste analitičko rješenje paradoksa. No ostaje drugo značajno pitanje — kako osigurati uspjeh procedure određenja istinitosti koja je ključna za ispravnost klasične logike, a da pri tom zadržimo unutrašnju značenjsku strukturu jezika. Naravno, zabranu upotrebe jezika koji govori o vlastitoj istinitosti ne možemo smatrati zadovoljavajućim odgovorom, niti hijerarhiju jezika u kojoj svaki jezik može govoriti jedino o istinitosti jezika koji je prije njega u hijerarhiji. Mada je cirkularnost bitan dio paradoksalnih situacija, njeno odbacivanje je pregrubo rješenje koje nedopustivo osiromašuje jezik. Kako je Kripke pokazao, cirkularnost je duboko prisutna u svakodnevnoj upotrebi jezika i to ne samo na neizbježiv već i na bezazlen način, i tek u nekim krajnjim okolnostima vodi paradoksima. Kripke je to pokazao na primjerima koji obuhvaćaju vanjsku značenjsku stukturu jezika ("empiričke činjenice"), no isto se dešava i s unutarnjom značenjskom strukturom. Ni tu cirkularnost ne vodi nužno paradoksima, kako pokazuje i primjer **Logičara**.

$Log :\ T(\overline{Log})\ \vee\ \neg T(\overline{Log})$ (Ova rečenica je istinita ili nije istinita)

Imamo sljedeće semantičke ovisnosti:



Kad bi Log bio lažan tada bi po uvjetima istinitosti $\neg T(\overline{Log})$ bio lažan, $T(\overline{Log})$ istinit, pa bi tako i Log bio istinit. Dakle, takva valuacija grafa nije moguća. No ako pretpostavimo da je Log istinit, uvjeti istinitosti jednoznačno generiraju konzistentnu valuaciju. Tako procedura određenja istinitosti daje jedinstven odgovor — da je Log istinit.

Kripke je pokazao da se okolnosti koje vode paradoksima ne mogu izolirati na sintaktičkom nivou, već je nužan zahvat u značenjsku strukturu jezika. Taj je zahvat ovdje urađen na sljedeći način. Zadržana je prvotna klasična semantika, dakle klasična procedura određenja istinitosti, ali odbačena pogrešna klasična pretpostavka o potpunoj uspješnosti procedure. Odbacivanje te pretpostavke ne mijenja smisao klasičnih uvjeta na funkciju istinitosti, jer su oni definirani na način neovisan o pretpostavci totalne definiranosti te funkcije. Njihovo funkcioniranje u novoj situaciji bit će ilustrirano na slijedećoj rečenici:

$L \ \vee \ 0 = 0$

Po klasičnom uvjetu za veznik $\vee$ ova rečenica je istinita upravo kada je jedna od osnovnih rečenica istinita. Kako je $0 = 0$ istinita to je i ukupna rečenica istinita, bez obzira što L nema istinitosnu vrijednost. Isto tako ako bi uvjet istinitosti za veznik $\wedge$ primijenili na rečenicu

$L \ \wedge \ 0 = 0$

istinitosna vrijednost neće biti utvrđena. Naime, njena istinitost zahtijeva da obje osnovne rečenice budu istinite, a to nije ispunjeno. Njena lažnost zahtijeva da barem jedna osnovna rečenica bude lažna, ali ni to nije slučaj. Tako je nedefiniranost istinitosti rečenice L primjenom klasičnih uvjeta dovela do nedefiniranosti istinitosti ukupne rečenice.

Klasični uvjeti istinitosti daju kriterije na istinitosnu vrijednost složene rečenice u odnosu na istinitosne vrijednosti rečenica od kojih je složena *neovisno o tome imaju li one istinitosnu vrijednost ili ne. Nedostatak neke istinitosne vrijednosti može ali ne mora dovesti i do nedostatka istinitosne vrijednosti ukupne rečenice. To je u cijelosti određeno klasičnim smislom tih konstrukcija i osnovnom postavkom da sve rečenice smatramo smislenim, neovisno o tome imaju li ili ne istinitosnu vrijednost. Dakle, neke rečenice, mada smislene, i mada vrednovane po klasičnim uvjetima neće imati istinitosnu vrijednost, jer im ti uvjeti ne daju jedinstvenu istinitosnu vrijednost. To vodi* **parcijalnoj dvovaljanoj semantici jezika***. Tamo gdje procedura određenja istinitosti daje jedinstvenu vrijednost, istinu ili laž, tu vrijednost prihvaćamo, a tamo gdje procedura "pada", tj. ne daje istinitosnu vrijednost ili pak daje više mogućnosti vrednovanja, rečenica ostaje bez istinitosne vrijednosti. Tu parcijalnu semantiku jezika možemo opisati i kao* **trovaljanu semantiku jezika** *— naprosto neuspjeh određenja istinitosti proglasimo trećom vrijednošću* $|$ *(neodređeno). To ne nosi nikakav dodatni filozofski naboj već je samo ugodno tehničko sredstvo opisa situacije.* Ta semantika jezika je detaljno opisana u sljedećem odjeljku.

No ta semantika ovdje nije prihvaćena i kao konačna semantika jezika. Odlučujući razlog za odbacivanje tog rješenja je mišljenje da je čovjeku prirođena dvovaljana semantika i da svaku drugu semantiku odgovarajućim modeliranjem možemo svesti na dvovaljanu (potvrda tome je da su opisi svih semantika dvovaljani). Ostajanje na trovaljanoj semantici jezika značilo bi da logika razmišljanja ne bi bila više klasična, ona na koju smo kao vrsta "ugođeni". Što se tiče samog predikata istine T to bi značilo zadržavanje njegovog klasičnog logičkog značenja u dvovaljanom dijelu jezika proširenog "šutnjom" u slučaju neuspjeha klasične procedure. Mada u metaopisu $T(\bar{\varphi})$ ima istu trovaljanu istinitosnu vrijednost kao i φ, ta semantika se ne može izraziti u samom jeziku L (jer o trećoj vrijednosti jezik šuti, ili bolje rečeno, treća vrijednost je odraz u metajeziku šutnje samog jezika). Tako su i izražajne mogućnosti jezika slabe. Npr. Lažac je neodređen. No dok smo ovo sasvim jednostavno izrekli u metajeziku, sam jezik L to ne može izraziti jer je, kako je već rečeno (u metajeziku) Lažac neodređen. *Ova "zona šutnje" ne samo da je iz navedenih logičkih i semantičkih razloga nezadovoljavajuća (vodi trovaljanoj logici i slabi izražajnu moć jezika) već je možemo prekinuti prirodnim naknadnim valuiranjem rečenica koje proizlazi iz prepoznavanja neuspjeha klasične procedure.* To će biti ilustrirano na primjeru Lašca. Na intuitivnom nivou razmišljanja, utvrdivši da Lažac nije ni istinit ni lažan, utvrđujemo da je neodređen. Ali to znači da on laže, jer govori da je lažan. Dakle, on je lažan. No to ne vodi obnovi kontradiktornosti jer se desio *semantički pomak* u argumentu kod tvrdnje da je on neodređen i svih tvrdnji koje su nakon nje slijedile. To je pomak od primarne parcijalne dvovaljane semantike do dvovaljanog opisa te semantike koji je tek nadopunjuje u onom dijelu gdje ona nije određena. Naime, Lažac govori o vlastitoj istinitosti unutar prvotne semantike, dok je zadnje vrednovanje unutar konačne semantike. Da je Lažac u njoj lažan ne znači da je istina što on govori jer nije isti semantički okvir. Da je on lažan u konačnoj semantici znači da je laž što on govori o svojoj primarnoj semantici. Iz toga slijedi da on nije lažan u svojoj primarnoj semantici. No on u toj semantici ne može biti ni istinit jer bi tada bio istinit i u konačnoj semantici (koja samo nadopunjava primarnu tamo gdje ona pada). Dakle on je u primarnoj semantici neodređen. Tako ne samo da nismo dobili kontradikciju već smo dobili dodatnu informaciju o Lašcu. Na isti način se može pokazati da i drugi paradoksi koji su smišljeni da obezvrijede rješenje pomoću trovaljane semantike, kao npr. Metalažac ili Intenzionalni Lažac, u sebi sadrže semantički pomak. Uočavanje tog pomaka pretvara ih u prirodan i nekontradiktoran prijelaz argumentacije iz primarne parcijalne semantike u dvovaljani opis te semantike.

Ovu intuiciju lako možemo legalizirati. U sljedećem odjeljku će biti dan detaljan matematički opis koji će sada biti tek skiciran. Predikatom istinitosti jezik govori o svojoj primarnoj semantici. Klasična procedura i klasični smisao predikata istinitosti određuju tu semantiku koja je uslijed odbacivanja neodržive pretpostavke o uspjehu te procedure nužno parcijalna dvovaljana semantika (= trovaljana semantika). *No sam taj govor o primarnoj semantici je njeno prirodno nadopunjene do konačne dvovaljane semantike. Dakle konačna semantika jezika ima za predmet primarnu semantiku jezika koju ujedno nadopunjuje u onom dijelu u kojem ona "šuti" upravo informacijama o toj šutnji.* Taj prijelaz lako možemo ocrtati na semantičkom grafu. Iz valuacije grafa u primarnoj semantici lako dobijamo valuaciju grafa u konačnoj semantici. Samo trebamo revaluirati atomarne tvrdnje oblika $T(\bar{\varphi})$.*Tvrdnja $T(\bar{\varphi})$ ima isto značenje u obje semantike — da je φ istinita u primarnoj semantici, no uvjeti istinitosti nisu isti. Dok su u primarnoj semantici uvjeti istinitosti $T(\bar{\varphi})$ klasični (vrijednost dvovaljane funkcije istinitosti tvrdnji $T(\bar{\varphi})$ i φ je ista), u sekundarnoj semantici nije tako. U njoj istinitost $T(\bar{\varphi})$ znači da je φ istinita u primarnoj semantici, a da je $T(\bar{\varphi})$ lažna znači da φ nije istinita u primarnoj semantici. No to ne znači da je ona lažna u primarnoj semantici već da je lažna ili neodređena.* Tako u konačnoj semantici, formalno gledano, $T(\bar{\varphi})$ nasljeđuje istinitosnu vrijednost od φ u primarnoj semantici dok ostale vrijednosti pretvara u laž. Da je to ispravan i potpun opis valuacije u primarnoj semantici najbolje možemo vidjeti ako uvedemo predikate za ostale vrijednosti primarne valuacije:

$F(\bar{\varphi})$ (= φ je lažna u primarnoj semantici) $\leftrightarrow T(\overline{\neg\varphi})$

$U(\bar{\varphi})$ (= φ je neodređena u primarnoj semantici) $\leftrightarrow \neg T(\bar{\varphi})$ i $\neg F(\bar{\varphi})$

Ovisno o vrijednosti tvrdnje φ u primarnoj semantici, određujemo koja je od prethodnih tvrdnji istinita a koja lažna. Ako je npr. φ lažna u primarnoj semantici tada je $F(\bar{\varphi})$ istinita dok su $T(\bar{\varphi})$ i $U(\bar{\varphi})$ lažne u konačnoj semantici. Odredivši tako (dvovaljanu) valuaciju atomarnih rečenica, određena je i valuacija svih drugih rečenica po klasičnim uvjetima i principu rekurzije. *Ne samo da ova valuacija zadržava primarni logički smisao predikata istine (kao predikata istine primarne semantike) već se ona podudara s primarnom valuacijom tamo gdje je ona određena (daje istinu ili laž).* Naime, ako je $T(\bar{\varphi})$ istinita u primarnoj semantici tada je u njoj istinita i φ, pa je $T(\bar{\varphi})$ istinita u konačnoj semantici. Ako je pak $T(\bar{\varphi})$ lažna u primarnoj semantici, tada je takva i φ, pa je i $T(\bar{\varphi})$ lažna u konačnoj semantici. Kako su uvjeti na istinitost složenih rečenica istovjetni u dvovaljanom dijelu, to se ovo podudaranje proteže na sve rečenice. Dakle vrijedi $T(\bar{\varphi}) \rightarrow \varphi$ i $F(\bar{\varphi}) \rightarrow \neg\varphi$.

Imajući na umu ovakvu dvostruku semantiku jezika lako možemo razriješiti sve paradokse istine. Na intuitivnom nivou smo to već pokazali za Lašca. Sada će to biti pokazano na još nekim primjerima. Da bismo razlikovali u okviru koje semantike spominjemo neki termin koristit ćemo prefiks “p” za primarnu i prefiks “f” za konačnu (finalnu) semantiku. Tako ćemo npr. razlikovati “p-laž” i “f-laž”. Obrazac rješenja je uvijek isti. Paradoks u klasičnom razmišljanju znači neodređenost istinitosti tvrdnje u primarnoj semantici. No to postaje informacija o tvrdnji u konačnoj semantici na osnovu koje možemo zaključiti o njenoj istinitosti u konačnoj semantici.

Promotrimo prvo još neke slučajeve koji poput Lašca vode kontradikciji u naivnoj semantici. Takav je npr. **Jaki lažac** $LL : \neg T(\overline{LL})$ (“Ova rečenica nije istinita.”). U naivnoj semantici on vodi kontradikciji na isti način kao i obični Lažac jer je tu negacija istine isto što i laž. Prepoznavši neuspjeh klasične procedure istinitosti nastavljamo razmišljati u konačnoj semantici. Utvrdivši da je LL p-neodređen utvrdili smo da nije p-istinit. No on upravo to tvrdi pa je tako f-istinit. Dakle LL je u primarnoj semantici neodređen a u konačnoj semantici istinit. Zanimljivo je da svu ovu argumentaciju možemo provesti direktno u konačnoj semantici, a ne indirektno preko utvrđivanja neuspjeha klasične procedure. Tu je argumentacija sljedeća. Kada bi LL bio f-lažan bila bi f-laž što on govori — da nije p-istinit. Dakle on je p-istinit. No to znači (po prije navedenoj suglasnosti semantika) da je f- istinit i to je kontradikcija s pretpostavkom. Dakle on je f-istinit. No ovo ne vodi kontradikciji već dodatnoj informaciji. Iz toga naime slijedi da je f-istina što govori — da nije p-istinit. Dakle on je ili p-lažan ili p-neodređen. No ne može biti p-lažan jer bi tada bio i f-lažan. Tako je on p-neodređen. Tako, mada su i obični Lažac i Jaki Lažac p-neodređeni, dok je prvi f-lažan drugi je f-istinit.

Analizirajmo na isti način i **Curryev paradoks** $C : T(\bar{C}) \to l$ ("Ako je ova rečenica istinita tada je l"), gdje je l neka lažna tvrdnja. Na intuitivnom nivou iz laži C slijedi da je $T(\bar{C})$ istinita, pa tako i C, i to je kontradikcija. Iz istinitosti C pak slijedi istinitost kondicionala kojemu je i antecedent (C) istinit, pa je tako i l istinita. No i to vodi kontradikciji jer je l lažna tvrdnja. Tako zaključujemo u konačnoj semantici da je C p-neodređena, pa tako i f-istinita (jer joj je antecedent f-lažan). No i ovu argumentaciju možemo direktno provesti u konačnoj semantici. Naime iz f-laži C slijedi da je antecedent f-istinit. To znači da je C p-istinita, pa je i f-istinita (po suglasnosti primarne i konačne semantike), i to je kontradikcija. Dakle, C je f-istinita. Iz toga dalje dobijamo da je $T(\bar{C})$ f-lažna ili pak l f-istinita. No kako je l f-lažna to mora biti $T(\bar{C})$ f-lažna pa nije C p-istinita. Ona je dakle p-lažna ili p-neodređena. No iz p-laži bi slijedila i f-laž, pa je tako ona p-neodređena.

Na sličan način i svi drugi paradoksi istine koji na intuitivnom nivou daju kontradikciju prelaze u pozitivnu argumentaciju u konačnoj semantici. Sasvim je druga situacija s paradoksima koji ne vode kontradikciji već dopuštaju bilo kakvu valuaciju istinitosti, kao npr. Istinoljubac. Njegova analiza daje da je on p-neodređen. Iz toga dalje slijedi da nije p-istinit tj. ($I : T(\bar{I})$) da nije I. Dakle, I je f-lažan. No za razliku od kontradikcija ovo se razmatranje ne može provesti direktno u konačnoj semantici. U njemu formulirana argumentacija ne daje odgovor kao ni u primarnoj semantici. Potrebno je pogledati u samu primarnu valuaciju semantičkog grafa da se dobije odgovor. Naravno, ako jezik dovoljno obogatimo da može opisati semantičke grafove, funkcije istinitosti i uvjete postojanja jedinstvene valuacije, tada argumentaciju možemo provesti u konačnoj semantici. To je upravo argumentacija koja je ovdje neformalno korištena.

Ovakvim modeliranjem konačnom je semantikom opisan predikat istine primarne semantike. Naravno, predikat istine primarne semantike nije i predikat istine konačne semantike. No opis pojma istine konačne semantike i nije bio cilj. Pojmom istine konačne semantike vrednujemo opis primarnog predikata istine. Kako je pojam istine konačne semantike ujedno i proširenje predikata istine primarne semantike on tako i sebe parcijalno opisuje, ali ne i potpuno. U tom smislu duh Tarskijeve hijerarhije je i dalje s nama. Za neke je on zao duh jer ne dopušta potpun opis predikata istine konačne semantike. Za mene je on dobar duh jer je pomoću njega potpuno opisan predikat istine primarne semantike.

4.2 Formalni opis rješenja

Neka je L interpretirani jezik prvog reda s domenom D. U krajnjem slučaju dopustit ćemo da je D prazan skup, tj. da je L sveden na logički vokabular. Također ćemo radi jednostavnosti pretpostaviti da za svaki objekt jezika $a \in D$ postoji zatvoren term jezika $\bar{a}$ koji ga imenuje.

L ćemo nadopuniti do jezika LT koji će govoriti i o istinitosti vlastitih rečenica. Njegova domena će pored objekata jezika L sadržavati i vlastite rečenice, a u svom će vokabularu imati predikat S (= "biti rečenica") koji će razlikovati rečenice od nerečenica i predikat T (= "biti istinita rečenica") koji će opisivati istinitost rečenica. Svaka rečenica φ imat će svoje ime $\overline{\varphi}$, no postojat će i posebna imena za rečenice. Davajući im odgovarajuću denotaciju postizat ćemo željeno samoreferiranje. Npr. konstantu $\overline{I}$ interpretirat ćemo kao ime rečenice $T(\overline{I})$ i tako dobiti rečenicu koja o sebi tvrdi da je istinita.

Vokabular jezika LT sastoji se od vokabulara jezika L i novih simbola - unarnih predikata S i T, te **rečeničnih konstanti** $\overline{I}$, $\overline{L}$, $\overline{\neg L}$, $\overline{LL}$, …, i posebnog simbola $^{-}$.

Skup **terma** TLT i skup **formula** FLT jezika LT definiramo kao najmanje skupove koji zadovoljavaju sve uvjete koje zadovoljavaju skupovi terma i formula jezika L te još sljedeće uvjete:

1. rečenične konstante $\overline{I}$, $\overline{L}$, $\overline{\neg L}$, $\overline{LL}$, … su termi
2. ako je t term onda su $S(t)$ i $T(t)$ formule
3. ako je φ formula tada je $\overline{\varphi}$ term

Skup **slobodnih varijabli** terma i formule definira se tako da se standardnim rekurzivnim uvjetima doda još i uvjet da su slobodne varijable terma $\overline{\varphi}$ upravo slobodne varijable formule φ. **Rečenice** jezika LT su zatvorene formule tog jezika. Njihov skup ćemo bilježiti SLT.

Interpretaciju (**model**) jezika LT dobivamo na sljedeći način. Domenu DLT tvore svi objekti iz domene D zajedno sa svim rečenicama jezika LT:$DLT = D \cup SLT$. Sve predikate modela jezika L dodefiniramo tako da su lažni kad im je barem jedan argument van domene D, a funkcije iz modela L dodefiniramo tako da poprimaju stalnu vrijednost, recimo rečenicu $T(\overline{I})$, kada im je barem jedan argument van domene D.

Nove simbole interpretiramo na sljedeći način. Simbolu S dajemo smisao "biti rečenica", tj. interpretiramo ga skupom rečenica od LT. Simbol T bit će predikat istine primarne semantike, tj. imat će smisao "biti istinita rečenica u primarnoj semantici" jednom kad definiramo što je to primarna semantika. U primarnoj semantici to će biti postignuto tako da će T biti uveden kao logički simbol (poput $\wedge$ npr.) s T - shemom kao uvjetom istinitosti, a u konačnoj semantici to će biti postignuto direktno, interpretirajući ga skupom rečenica istinitih u primarnoj semantici. Rečenične konstante interpretirat ćemo kao imena odgovarajućih rečenica:

1. $\overline{I}$ je ime rečenice $T(\overline{I})$
2. $\overline{L}$ je ime rečenice $T(\overline{\neg L})$
3. $\overline{\neg L}$ je ime rečenice $\neg T(\overline{\neg L})$
4. $\overline{LL}$ je ime rečenice $\neg T(\overline{LL})$

itd.

Zatvoreni term $\overline{\varphi}$ interpretiramo kao ime rečenice φ.

Složenije zatvorene terme interpretiramo kao imena odgovarajućih objekata jezika na standardan način.

Kako ćemo se primarno baviti rečenicama, mehanizam referiranja na njih je sljedeći. U metajeziku ćemo koristiti grčka slova φ, ψ, ..., za varijable po rečenicama. Tako u jeziku možemo referirati na proizvoljnu rečenicu φ zatvorenim termom $\overline{\varphi}$. Da bi izrekli da neka tvrdnja vrijedi za proizvoljnu rečenicu φ, npr. da iz istinitosti rečenice u primarnoj semantici slijedi istinitost u konačnoj, naprosto ćemo reći da je za svaku rečenicu φ istinita tvrdnja jezika $T(\bar{\varphi}) \rightarrow \varphi$.

I još jedan detalj. Radi uniformnosti zapisa upotreba znaka $^-$ je trostruka. Osnovna upotreba je u konstrukciji terma $\overline{\varphi}$ kojim referiramo na rečenicu φ. Npr. $\overline{1=1}$ je ime u jeziku LT rečenice 1=1 jezika LT. Druga upotreba je u tvorbi rečeničnih konstanti kojima postižemo samoreferiranje. Npr. u izrazu $\overline{L}$ on nema osnovnu upotrebu jer to nije ime rečenice L. Naime, takvu rečenicu jezik LT nema. Znak L možemo eventualno shvatiti tek kao metajezično ime za rečenicu koja je u jeziku imenovana rečeničnom konstantom $\overline{L}$, a to je po prethodnom rečenica $T(\overline{\neg L})$. Treća upotreba tog znaka je u smislu operatora koji proizvoljnom objektu a domene početnog jezika pridružuje njegovo ime u jeziku $\overline{a}$. No za razliku od prethodnih upotreba ono što je "stavljeno ispod crte" općenito nije izraz jezika nego vanjezični objekt.

Slijedi opis **primarne semantike** jezika LT, tj. opis **primarne istinitosti** rečenica iz LT. Zbog pretpostavke da svaki objekt $a \in DLT$ ima svoje ime $\overline{a}$, moći ćemo razmatrati samo rečenice, a valuacije proizvoljnih formula i terma uvesti naknadno. *Uvjeti na funkciju istinitosti rečenica I_c su klasični uvjeti zajedno s klasičnim uvjetom na predikat istinitosti, ali je odbačena klasična pretpostavka da je to totalna funkcija, tj. da je definirana za svaku rečenicu. Među svim takvim funkcijama izabrat ćemo onu koja je na svojoj domeni jedinstvena (sjetimo se da mogućnost više valuacija date rečenice smatramo neuspjehom klasične procedure), a među svim takvim maksimalnu, (jer svaki uspjeh određenja istinitosti prihvaćamo).* Tako definiramo **klasičnu funkciju istinitosti** I_c jezika LT kao *parcijalnu* funkciju I_c iz $SLT \longrightarrow \{\top, \bot\}$ takvu da vrijedi:

1. Na atomarnim rečenicama koje počinju predikatima jezika L vrijednost joj se podudara s istinitošću tih rečenica u jeziku L interpretiranom nad proširenom domenom DLT, na atomarnim rečenicama oblika $S(t)$ daje istinu ($\top$) ako je t ime rečenice, u protivnom daje laž ($\bot$), a na atomarnim rečenicama oblika $T(t)$ vrijednost će joj biti naknadno određena.

2. klasični uvjeti:

   (a) $I_c(\neg\varphi) = \begin{cases} \top & za\ I_c(\varphi) = \bot \\ \bot & za\ I_c(\varphi) = \top \end{cases}$

   (b) $I_c(\varphi \wedge \psi) = \begin{cases} \top & za\ I_c(\varphi) = \top\ i\ I_c(\psi) = \top\ (\textit{obje su istinite}) \\ \bot & za\ I_c(\varphi) = \bot\ ili\ I_c(\psi) = \bot\ \text{(barem je jedna lažna)} \end{cases}$

   (c) $I_c(\varphi \vee \psi) = \begin{cases} \top & za\ I_c(\varphi) = \top\ ili\ I_c(\psi) = \top\ (\textit{barem jedna je istinita}) \\ \bot & za\ I_c(\varphi) = \bot\ i\ I_c(\psi) = \bot\ \text{(obje su lažne)} \end{cases}$

   (d) $I_c(\varphi \rightarrow \psi) = \begin{cases} \top & za\ I_c(\varphi) = \bot\ ili\ I_c(\psi) = \top \\ \bot & za\ I_c(\varphi) = \top\ i\ I_c(\psi) = \bot \end{cases}$

   (e) $I_c(\varphi \leftrightarrow \psi) = \begin{cases} \top & za\ I_c(\varphi) = I_c(\psi)\ \text{(obje su istinite ili obje su lažne)} \\ \bot & za\ I_c(\varphi) \neq I_c(\psi)\ \text{(jedna je istinita a druga lažna)} \end{cases}$

(f) $I_c(\forall x \varphi(x)) = \begin{cases} \top & ako\ \forall a \in DLT\ I_c(\varphi(\bar{a})) = \top \\ \bot & ako\ \exists a \in DLT\ I_c(\varphi(\bar{a})) = \bot \end{cases}$

(g) $I_c(\exists x \varphi(x)) = \begin{cases} \top & ako\ \exists a \in DLT\ I_c(\varphi(\bar{a})) = \top \\ \bot & ako\ \forall a \in DLT\ I_c(\varphi(\bar{a})) = \bot \end{cases}$

3. klasični uvjet na predikat istine:

   $I_c(T(\bar{\varphi})) = \begin{cases} \top & za\ I_c(\varphi) = \top \\ \bot & za\ I_c(\varphi) = \bot \end{cases}$

   Sada možemo definirati istinitosnu vrijednost i za atomarne rečenice $T(t)$, gdje je t zatvoreni term. Ako t imenuje rečenicu φ tada je istinitosna vrijednost od $T(t)$ ista kao i istinitosna vrijednost od $T(\bar{\varphi})$. Ako pak t nije ime rečenice tada je istinitosna vrijednost od $T(t)$ laž($\bot$).

4. jedinstvenost na domeni:

   Ako postoji funkcija I iz $SLT \longrightarrow \{\top, \bot\}$ koja zadovoljava prethodna tri uvjeta, tada za svaku φ na kojoj su definirane i I i I_c, $\varphi \in Dom(I) \cap Dom(I_c)$ $I(\varphi) = I_c(\varphi)$.

5. maksimalna je:

   Za svaku funkciju I iz $SLT \longrightarrow \{\top, \bot\}$ koja ispunjava prethodna četiri uvjeta

   $Dom(I) \subseteq Dom(I_c)$.

Iz definicije lako slijedi jedinstvenost takve funkcije. Kad bi postojale dvije takve funkcije po zadnjem uvjetu bi imale istu domenu, a po predzadnjem bi se podudarale na njoj. Poslije će biti dokazano i njeno postojanje.

Pojam istinitosti rečenica proširujemo na proizvoljne formule standardnim načinom — fiksiranjem značenja varijabli. Funkciju $v : Var \longrightarrow DLT$ (gdje je Var skup varijabli jezika) koja fiksira značenja varijabli zvat ćemo **valuacijom**. U datoj valuaciji v formulu $\varphi(x_1, x_2, \ldots, x_n)$ sa slobodnim varijablama $x_1, x_2, \ldots, x_n$ smatramo istinitom $\leftrightarrow$ pripadna rečenica $\varphi(\overline{v(x_1)}, \overline{v(x_2)}, \ldots, \overline{v(x_n)})$ je istinita, a lažnom $\leftrightarrow$ pripadna rečenica je lažna. Isto tako u datoj valuaciji v smatramo da term $t(x_1, x_2, \ldots, x_n)$ sa slobodnim varijablama $x_1, x_2, \ldots, x_n$ označava isto što i zatvoreni term $t(\overline{v(x_1)}, \overline{v(x_2)}, \ldots, \overline{v(x_n)})$.

Da bismo bolje razotkrili strukturu klasične funkcije istinitosti nadopunit ćemo je do totalne funkcije tako da joj tamo gdje nije definirana pridružimo *vrijednost* $|$ *(neodređeno).* Tu funkciju ćemo zvati **funkcijom istinitosti primarne semantike** $I_p : ST \longrightarrow \{\top, \bot, |, \}$:

$$I_p(\varphi) = \begin{cases} I_c(\varphi) & za\ \varphi \in Dom(I_c) \\ | & u\ protivnom \end{cases}$$

Skup rečenica na kojima I_p poprima klasične vrijednosti $\top$ i $\bot$ zvat ćemo njenom **domenom određenosti** DDI_p. Primijetimo da na isti način možemo uspostaviti bijektivnu vezu između proizvoljnih parcijalnih dvovaljanih funkcija istinitosti i totalnih trovaljanih funkcija istinitosti.

Iz definicije lako možemo iščitati uvjete (tablice) istinitosti rečeničnih konstrukcija za tu funkciju. Ako su argumenti klasični ($\top$ i $\bot$), tada je i vrijednost klasična, zadana klasičnim uvjetima. Ako je pak neki od argumenata neodređen, tada gledamo propagira li se taj neuspjeh i na određenje vrijednosti složene rečenice po klasičnim uvjetima. Ako jeste tada je i vrijednost neodređena, a ako ne, klasična je. Npr. vrijednost rečenice $\varphi \wedge \psi$ za φ neodređenu a ψ lažnu je laž jer je dovoljno po klasičnim uvjetima da je barem jedna rečenica lažna (u našem slučaju ψ) pa da i ukupna rečenica bude lažna, neovisno o istinitosti druge rečenice. No ako je ψ istinita, tada istinitost ukupne rečenice bitno ovisi o istinitosti φ. Po klasičnim uvjetima, ako je φ istinita tada je i konjukcija istinita, a ako je lažna tada je i konjukcija lažna. No kako je φ neodređena taj neuspjeh se sada propagira i na cijelu konjukciju koja isto biva neodređena.

Tako dobivamo sljedeće **tablice istinitosti za primarnu semantiku** I_p:

1. $I_p(\neg\varphi) = \begin{cases} \top & za\ I_p(\varphi) = \bot \\ \bot & za\ I_p(\varphi) = \top \\ | & u\ protivnom \end{cases}$

φ	$\neg\varphi$
$\top$	$\bot$
$\bot$	$\top$
$\vert$	$\vert$

2. $I_p(\varphi \wedge \psi) = \begin{cases} \top & za\ I_p(\varphi) = \top\ i\ I_p(\psi) = \top\ (obje\ su\ istinite) \\ \bot & za\ I_p(\varphi) = \bot\ ili\ I_p(\psi) = \bot\ \text{(barem je jedna lažna)} \\ | & u\ protivnom \end{cases}$

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\bot$	$\mid$
$\bot$	$\bot$	$\bot$	$\bot$
$\mid$	$\mid$	$\bot$	$\mid$

3. $$I_p(\varphi \vee \psi) = \begin{cases} \top & za\ I_p(\varphi) = \top\ ili\ I_p(\psi) = \top\ (barem\ jedna\ je\ istinita) \\ \bot & za\ I_p(\varphi) = \bot\ i\ I_p(\psi) = \bot\ \text{(obje su lažne)} \\ \mid & u\ protivnom \end{cases}$$

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\top$	$\top$
$\bot$	$\top$	$\bot$	$\mid$
$\mid$	$\top$	$\mid$	$\mid$

4. $$I_p(\varphi \to \psi) = \begin{cases} \top & za\ I_p(\varphi) = \bot\ ili\ I_p(\psi) = \top \\ \bot & za\ I_p(\varphi) = \top\ i\ I_p(\psi) = \bot \\ \mid & u\ protivnom \end{cases}$$

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\bot$	$\mid$
$\bot$	$\top$	$\top$	$\top$
$\mid$	$\top$	$\mid$	$\mid$

5. $$I_p(\varphi \leftrightarrow \psi) = \begin{cases} \top & za\ I_p(\varphi) = I_p(\psi) \neq \mid\ \text{(obje su istinite ili obje su lažne)} \\ \bot & za\ I_p(\varphi) \neq I_p(\psi)\ i\ nijedna\ vrijednost\ nije\ \mid \\ & \text{(jedna je istinita a druga lažna)} \\ \mid & u\ protivnom \end{cases}$$

$\varphi \setminus \psi$	$\top$	$\bot$	$\mid$
$\top$	$\top$	$\bot$	$\mid$
$\bot$	$\bot$	$\top$	$\mid$
$\mid$	$\mid$	$\mid$	$\mid$

6. $I_p(\forall x\varphi(x)) = \begin{cases} \top & ako\ \forall a \in DLT\ I_p(\varphi(\bar{a})) = \top \\ \bot & ako\ \exists a \in DLT\ I_p(\varphi(\bar{a})) = \bot \\ \mid & u\ protivnom \end{cases}$

7. $I_p(\exists x\varphi(x)) = \begin{cases} \top & ako\ \exists a \in DLT\ I_p(\varphi(\bar{a})) = \top \\ \bot & ako\ \forall a \in DLT\ I_p(\varphi(\bar{a})) = \bot \\ \mid & u\ protivnom \end{cases}$

Ovim prevođenjem klasičnih uvjeta istinitosti s parcijalne dvovaljane funkcije na pridruženu trovaljanu funkciju dobili smo upravo jaku Kleeneovu trovaljanu semantiku. Ona se obično interpretira kao semantika uspjeha paralelnih algoritama ili pak kao semantika ispitivanja istinitosti rečenica u smislu da one rečenice kojima *još nije* utvrđena klasična istinitost imaju vrijednost $\mid$. *Ovdje je ona interpretirana kao klasična procedura određenja istinitosti nadopunjena propagiranjem vlastita neuspjeha.*

Obratno, neka imamo trovaljanu funkciju $I : SLT \longrightarrow \{\top, \bot, \mid, \}$ koja ispunjava uvjete jake Kleeneove semantike. Lako je vidjeti da njoj pridružena parcijalna klasična funkcija I' iz $SLT \longrightarrow \{\top, \bot\}$

$$I'(\varphi) = \begin{cases} I(\varphi) & za\ \varphi \in DD(I) \\ nije\ definirana & u\ protivnom \end{cases}$$

zadovoljava klasične uvjete istinitosti. *Tako su preko ove korespodencije uvjeti jake Kleeneove semantike na trovaljanu funkciju istinitosti isto što i klasični uvjeti istinitosti na parcijalnu dvovaljanu funkciju istinitosti.*

Što se tiče predikata istinitosti klasično određenje istinitosti $T(\bar{\varphi})$ propada upravo onda kada propada određenje istinitosti $\bar{\varphi}$. Tako za I_p vrijedi sljedeće:

$$I_p(T(\bar{\varphi})) = I_p(\varphi)$$

A to znači da je I_p fiksna točka jake Kleeneove semantike. Obratno, takvoj funkciji je pridružena parcijalna dvovaljana funkcija koja zadovoljava T - shemu. *Tako su preko ove korespodencije fiksna točka jake Kleeneove semantike i parcijalna dvovaljana funkcija istinitosti koja zadovoljava klasične uvjete istinitosti i T shemu jedna te ista stvar.*

Iz uvjeta jedinstvenosti I_c na svojoj domeni slijedi da je I_p intrinsična fiksna točka. Zaista, neka je J neka druga fiksna točka kojoj odgovara klasična parcijalna dvovaljana istinitosna funkcija J' i neka rečenica φ pripada $DD(J) \cap DD(I_p)$ (obje valuacije imaju određenu vrijednost na njoj) Trebamo vidjeti da je $J(\varphi) = I_p(\varphi)$. To lako utvrdimo računanjem koristeći pri tom jedinstvenost I_c:

$J(\varphi) = J'(\varphi) = I_c(\varphi) = I_p(\varphi)$

Na isti ovaj način se dokaže da je svakoj parcijalnoj dvovaljanoj istinitosnoj funkciji koja zadovoljava klasične uvjete i shemu T, te je jedinstvena na svojoj domeni, pridružena intrinsična fiksna točka, kao i obratno, da je svakoj intrinsičnoj fiksnoj točki priružena upravo takva dvovaljana funkcija. Dakle, *preko ove korespodencije su intrinsična fiksna točka jake Kleeneove semantike i parcijalna dvovaljana funkcija istinitosti koja zadovoljava klasične uvjete i T shemu i koja je jedinstvena na svojoj domeni, jedna te ista stvar.*

Uvjet maksimalnosti funkcije I_c pak povlači i uvjet maksimalnosti I_p — za svaku drugu intrinsičnu fiksnu točku I $DD(I) \subseteq DD(I_p)$. Dakle, *I_p je maksimalna intrinsična fiksna točka jake Kleeneove semantike.* Lako je vidjeti da vrijedi i obrat. *Tako su preko navedene korespodencije najveća intrinsična fiksna točka jake Kleeneove semantike i klasična funkcija istinitosti jedna te ista stvar.* Koristeći rezultat iz prethodnog poglavlja koji kaže da postoji jedinstvena najveća intrinsična fiksna točka, ovom bijektivnom korespodencijom smo dobili sljedeće

Teorem 41. *Postoji jedinstvena klasična funkcija istinitosti I_c jezika LT.*

Time je pokazano da intuitivna analiza klasične procedure određenja istinitosti i njenog eventualnog neuspjeha vodi jedinstvenoj formulaciji. Kako je ona u svom trovaljanom obliku upravo maksimalna intrinsična fiksna točka jake Kleeneove semantike, ovom analizom je ujedno dana i argumentacija za upravo taj izbor među raznim fiksnim točkama raznih trovaljanih semantika za model pojma istine. Ujedno je dan i direktan matematički opis te fiksne točke pomoću kojeg jednostavnije možemo utvrditi istinitosnu vrijednost rečenice u toj fiksnoj točki negoli iz njene standardne karakterizacije u strukturi svih fiksnih točaka. Za primjer, mogu se usporediti standardan dokaz istinitosti Logičara u ovoj fiksnoj točki (str. 47) i dokaz pomoću ovakve karakterizacije (str. 62)

Konačnu semantiku I_f jezika LT dobijamo, kako je već u prvom odjeljku opisano, tako da za njen predmet uzmemo primarnu semantiku jezika, tj. da predikat istine govori o istinitosti rečenice u primarnoj semantici. Dakle sva druga semantička određenja ostaju ista kao i za primarnu semantiku, samo što se na drugi način određuje istinitost atomarnih rečenica $T(\bar{\varphi})$:

$$I_f(T(\bar{\varphi})) = \begin{cases} \top & \textit{za } I_p(\varphi) = \top \\ \bot & \textit{u protivnom} \end{cases}$$

Ovo povlači i odgovarajuće redefiniranje istinitosnih vrijednosti za atomarne rečenice $T(t)$, gdje je t zatvoreni term. Ako t imenuje rečenicu φ tada je istinitosna vrijednost u konačnoj semantici od $T(t)$ ista kao i istinitosna vrijednost u konačnoj semantici od $T(\bar{\varphi})$. Ako pak t nije ime rečenice tada je istinitosna vrijednost od $T(t)$ laž($\bot$).

Kako ta funkcija sada ima točno određenu istinitost $\top$ ili $\bot$ na svim atomarnim rečenicama i kako zadovoljava klasične uvjete istinitosti na složenim rečenicama, ona je klasična totalna dvovaljana funkcija istinitosti. Po principu rekurzije po strukturi rečenica, postoji jedinstvena takva funkcija $I_f : SLT \longrightarrow \{\top, \bot\}$.

Iz definicije slijedi da je $I_p(T(\bar{a})) \leq I_f(T(\bar{\varphi}))$, pa po monotonosti jake Kleeneove semantike *slijedi da je konačna semantika I_f proširenje primarne semantike I_p*:

$$I_p \leq I_f$$

Iz sljedećih je metadefinicija jasno da *pomoću ovog predikata možemo opisati i preostale vrijednosti istinitosti u primarnoj semantici*:

$$F(\bar{\varphi}) \;\leftrightarrow\; T(\overline{\neg\varphi})$$

$$U(\bar{\varphi}) \;\leftrightarrow\; \neg F(\bar{\varphi}) \wedge \neg T(\bar{\varphi})$$

Ugodno je uvesti i pojam određene istinitosne vrijednosti D:

$$D(\bar{\varphi}) \;\leftrightarrow\; F(\bar{\varphi}) \vee T(\bar{\varphi})$$

Sada možemo direktno opisati istinitost u primarnoj semantici:

$$I_p(\bar{\varphi}) = \top \;\leftrightarrow\; I_f(T(\overline{\varphi}) = \top$$

$$I_p(\bar{\varphi}) = \bot \;\leftrightarrow\; I_f(F(\overline{\varphi}) = \top$$

$$I_p(\bar{\varphi}) =| \leftrightarrow\; I_f(U(\overline{\varphi}) = \top$$

Da bi se stekao bolji uvid u izražajnu moć konačne semantike bit će pobrojane neke tvrdnje koje su istinite u njoj ([Fef84]). Prije svega *u konačnoj semantici istinite su tvrdnje koje izriču konzistentnost primarne semantike.* Naime, za svaku rečenicu φ u konačnoj semantici je istinito:

1. $\neg(T(\bar{\varphi}) \wedge F(\bar{\varphi}))$

Zaista

$I_f(\neg(T(\bar{\varphi}) \wedge F(\bar{\varphi}))) = \top$

po klasičnim uvjetima istinitosti je upravo onda kada

nije $(I_f(T(\overline{\varphi})) = \top$ i $I_f(F(\overline{\varphi})) = \top)$

a po definiciji I_f to je ekvivalentno tvrdnji

nije $(I_p(\varphi) = \top$ i $I_p(\varphi) = \bot)$

Iz opisa primarne semantike slijedi istinitost te tvrdnje pa je tako i početna tvrdnja istinita u konačnoj semantici.

U konačnoj semantici možemo izreći i da je ona nadopuna primarne semantike. Naime u njoj su istinite sljedeće tvrdnje:

2a. $T(\overline{\varphi}) \to \varphi$

2b. $F(\overline{\varphi}) \to \neg\varphi$

Dokažimo prvu tvrdnju. Neka je $T(\overline{\varphi})$ istinita u konačnoj semantici. Tada je φ istinita u primarnoj semantici. Ali pošto je konačna semantika proširenje primarne semantike, to je φ istinita i u konačnoj semantici.

Naravno, obrat općenito ne vrijedi jer bismo tada imali u konačnoj semantici shemu T što je u prisustvu npr. Lašca nemoguće jer je u pitanju klasični jezik. Ali *ako rečenica ima klasičnu istinitosnu vrijednost u primarnoj semantici tada vrijedi obrat*:

2c. $D(\overline{\varphi}) \ \to \ (T(\overline{\varphi}) \leftrightarrow \varphi) \wedge (F(\overline{\varphi}) \leftrightarrow \neg\varphi)$

Mada u primarnoj semantici logičke istine klasične logike nisu općenito istinite, već mogu biti i neodređene (npr. tvrdnja $I \vee \neg I$, gdje je I istinoljubac), *tvrdnje koje su logički ekvivalentne u klasičnoj logici imaju jednaku istinitosnu vrijednost u primarnoj semantici.* Naime u konačnoj semantici vrijedi sljedeće:

Za logički ekvivalentne rečenice u klasičnoj logici φ_1 i φ_2 vrijedi

3a. $T(\overline{\varphi_1}) \ \leftrightarrow \ T(\overline{\varphi_2})$

3b. $F(\overline{\varphi_1}) \ \leftrightarrow \ F(\overline{\varphi_2})$

3c. $U(\overline{\varphi_1}) \;\leftrightarrow\; U(\overline{\varphi_2})$

Trebamo dokazati da je za takve tvrdnje

$I_p(\varphi_1) = I_p(\varphi_2)$

Da bismo to dokazali koristit ćemo sljedeću tvrdnju koju lako dokažemo indukcijom po strukturi rečenica:

Ako je $I_p(\varphi) = |$ onda postoje proširenja I_1 i I_2 od I_p koja su potpune klasične istinitosne funkcije i kojima je $I_1(\varphi) = \top$ a $I_2(\varphi) = \bot$.

Dokažimo sada da je $I_p(\varphi_1) = I_p(\varphi_2)$. Neka je $I_p(\varphi_1) = \top$. Onda je i $I_f(\varphi_1) = \top$. Zbog logičke ekvivalentnosti u klasičnoj logici to znači da je i $I_f(\varphi_2) = \top$. To opet povlači da je $I_p(\varphi_2) = |$ ili $I_p(\varphi_2) = \top$. Prva mogućnost bi značila nadopunu do potpune klasične funkcije istinitosti I koju bismo mogli izabrati tako da je $I(\varphi_2) = \bot$. S druge strane zbog monotonosti semantike u njoj bi bilo $I(\varphi_1) = \top$, a što je u kontradikciji s logičkom ekvivalentnošću ovih tvrdnji u klasičnoj semantici. Zato prva mogućnost otpada i preostaje druga mogućnost koju je i trebalo dokazati. Analogno se dokažu i preostali slučajevi za vrijednost od $I_p(\varphi_1)$.

Sljedeće istine konačne semantike su direktan opis tablica istinitosti jake Kleeneove semantike:

4a. negacija:

1. $T(\overline{\neg\varphi}) \;\leftrightarrow\; F(\bar{\varphi})$
2. $F(\overline{\neg\varphi}) \;\leftrightarrow\; T(\bar{\varphi})$
3. $U(\overline{\neg\varphi}) \;\leftrightarrow\; U(\bar{\varphi})$

4b. konjukcija:

1. $T(\overline{\varphi \wedge \psi}) \;\leftrightarrow\; T(\bar{\varphi}) \wedge T(\bar{\psi})$
2. $F(\overline{\varphi \wedge \psi}) \;\leftrightarrow\; F(\bar{\varphi}) \vee F(\bar{\psi})$
3. $U(\overline{\varphi \wedge \psi}) \;\leftrightarrow\; (T(\bar{\varphi}) \wedge U(\bar{\psi})) \vee (U(\bar{\varphi}) \wedge T(\bar{\psi})) \vee (U(\bar{\varphi}) \wedge U(\bar{\psi}))$

4c. disjunkcija:

1. $T(\overline{\varphi \vee \psi}) \;\leftrightarrow\; T(\bar{\varphi}) \vee T(\bar{\psi})$
2. $F(\overline{\varphi \vee \psi}) \;\leftrightarrow\; F(\bar{\varphi}) \wedge F(\bar{\psi})$
3. $U(\overline{\varphi \vee \psi}) \;\leftrightarrow\; (F(\bar{\varphi}) \wedge U(\bar{\psi})) \vee (U(\bar{\varphi}) \wedge F(\bar{\psi})) \vee (U(\bar{\varphi}) \wedge U(\bar{\psi}))$

4d. kondicional:

1. $T(\overline{\varphi \to \psi}) \;\leftrightarrow\; F(\bar{\varphi}) \vee T(\bar{\psi})$
2. $F(\overline{\varphi \to \psi}) \;\leftrightarrow\; T(\bar{\varphi}) \wedge F(\bar{\psi})$
3. $U(\overline{\varphi \to \psi}) \;\leftrightarrow\; (T(\bar{\varphi}) \wedge U(\bar{\psi})) \vee (U(\bar{\varphi}) \wedge F(\bar{\psi})) \vee (U(\bar{\varphi}) \wedge U(\bar{\psi}))$

4e. bikondicional:

1. $T(\overline{\varphi \leftrightarrow \psi}) \;\leftrightarrow\; (T(\bar{\varphi}) \wedge T(\bar{\psi})) \vee (F(\bar{\varphi}) \wedge F(\bar{\psi}))$
2. $F(\overline{\varphi \leftrightarrow \psi}) \;\leftrightarrow\; (T(\bar{\varphi}) \wedge F(\bar{\psi})) \vee (F(\bar{\varphi}) \wedge T(\bar{\psi}))$
3. $U(\overline{\varphi \leftrightarrow \psi}) \;\leftrightarrow\; U(\bar{\varphi}) \vee U(\bar{\psi})$

4f. univerzalna kvantifikacija:

1. $T(\overline{\forall x \varphi(x)}) \leftrightarrow \forall x T(\overline{\varphi(x)})$
2. $F(\overline{\forall x \varphi(x)}) \leftrightarrow \exists x F(\overline{\varphi(x)})$
3. $U(\overline{\forall x \varphi(x)}) \leftrightarrow \neg \exists x F(\overline{\varphi(x)}) \wedge \exists x U(\overline{\varphi(x)})$

4g. egzistencijalna kvantifikacija:

1. $T(\overline{\exists x \varphi(x)}) \leftrightarrow \exists x T(\overline{\varphi(x)})$
2. $F(\overline{\exists x \varphi(x)}) \leftrightarrow \forall x F(\overline{\varphi(x)})$
3. $U(\overline{\exists x \varphi(x)}) \leftrightarrow \neg \exists x T(\overline{\varphi(x)}) \wedge \exists x U(\overline{\varphi(x)})$

Za ilustraciju bit će dokazano pravilo za konjukciju. U konačnoj semantici je istinita $T(\overline{\varphi \wedge \psi})$ ako i samo ako je $\varphi \wedge \psi$ istinito u primarnoj semantici. Po uvjetima istinitosti primarne semantike to je ispunjeno ako i samo ako su u primarnoj semantici istinite φ i ψ, tj. ako i samo ako je u konačnoj semantici istinito $T(\bar{\varphi}) \wedge T(\bar{\psi})$.

Iteriranje predikata istine nije zanimljivo jer u konačnoj semantici vrijedi sljedeće:

4a. $T(\overline{T(\overline{\varphi})}) \;\leftrightarrow\; T(\bar{\varphi})$

4b. $F(\overline{T(\overline{\varphi})}) \;\leftrightarrow\; F(\bar{\varphi})$

4c. $T(\overline{F(\overline{\varphi})}) \;\leftrightarrow\; F(\bar{\varphi})$

4d. $F(\overline{F(\overline{\varphi})}) \;\leftrightarrow\; T(\bar{\varphi})$

4e. $U(\overline{T(\overline{\varphi})}) \leftrightarrow U(\bar{\varphi})$

4f. $U(\overline{F(\overline{\varphi})}) \leftrightarrow U(\bar{\varphi})$

4g. $U(\overline{F(\overline{\varphi})}) \leftrightarrow U(\bar{\varphi})$

4h. $F(\overline{U(\overline{\varphi})}) \leftrightarrow D(\bar{\varphi})$

4i. $\neg(T(\overline{U(\overline{\varphi})}))$

Za ilustraciju će biti dokazana 4g. tvrdnja. U konačnoj semantici je istinito $U(\overline{F(\overline{\varphi})})$ ako i samo ako je u primarnoj semantici neodređeno $F(\overline{\varphi})$, dakle ako i samo ako je u primarnoj semantici neodređeno $\bar{\varphi}$, tj. ako i samo ako je u konačnoj semantici istinito $U(\bar{\varphi})$.

Prethodna pravila vrše određenu redukciju problema utvrđivanja istinitosne vrijednosti rečenice φ u primarnoj semantici, tj. problema utvrđivanja istinitosti $T(\bar{\varphi})$ i $F(\bar{\varphi})$ u konačnoj semantici. Pogledajmo kako to funkcionira na jednom primjeru:

$$T(\overline{\neg(F(\overline{\varphi}) \wedge T(\overline{\psi}))}) \leftrightarrow$$
$$F(\overline{F(\overline{\varphi}) \wedge T(\overline{\psi})}) \leftrightarrow$$
$$F(\overline{F(\overline{\varphi})}) \vee F(\overline{T(\overline{\psi})}) \leftrightarrow$$
$$T(\overline{\varphi}) \vee F(\overline{\psi})$$

Tako se ispitivanje istinitosti rečenice $T(\overline{\neg(F(\overline{\varphi}) \wedge T(\overline{\psi}))})$ svelo na ispitivanje istinitosti njoj pridružene rečenice:

$$(\neg(F(\overline{\varphi}) \wedge T(\overline{\psi})))^+ = T(\overline{\varphi}) \vee F(\overline{\psi})$$

Ovo prevođenje možemo napraviti jednostavno tako da φ prevedemo u preneksnu disjunktivnu normalnu formu i sve negacije atomarnih formula oblika $\neg T(\overline{\psi})$ i $\neg F(\overline{\psi})$ zamijenimo formulama oblika $F(\overline{\psi})$, odnosno $T(\overline{\psi})$. No možemo ovo prevođenje i rekurzivno opisati. Zbog interdefinabilnosti veznika možemo se ograničiti na negaciju, disjunkciju i konjukciju. Sljedećom rekurzijom formuli φ pridružit ćemo formule φ^+ i φ^- koje pitanje istinitosti odnosno lažnosti φ u primarnoj semantici svode na ispitivanje istinitosnih vrijednosti njenih atomarnih formula u primarnoj semantici:

1. Za atomarnu formulu φ koja pripada početnom jeziku L ili pak počinje predikatom $Sent$ je

   $\varphi^+ = \varphi \qquad\qquad \varphi^- = \neg\varphi$

2. $T(\overline{\varphi})^+ = T(\overline{\varphi})$ $\quad$ $T(\overline{\varphi})^- = F(\overline{\varphi})$
   $F(\overline{\varphi})^+ = F(\overline{\varphi})$ $\quad$ $F(\overline{\varphi})^- = T(\overline{\varphi})$
3. $(\neg\varphi)^+ = \varphi^-$ $\quad$ $(\neg\varphi)^- = \varphi^+$
4. $(\varphi \wedge \psi)^+ = \varphi^+ \wedge \psi^+$ $\quad$ $(\varphi \wedge \psi)^- = \varphi^- \vee \psi^-$
   $(\varphi \vee \psi)^+ = \varphi^+ \vee \psi^+$ $\quad$ $(\varphi \vee \psi)^- = \varphi^- \wedge \psi^-$
5. $(\forall x\varphi(x))^+ = \forall x(\varphi(x)^+)$ $\quad$ $(\forall x\varphi(x))^- = \exists x(\varphi(x)^-)$
   $(\exists x\varphi(x))^+ = \exists x(\varphi(x)^+)$ $\quad$ $(\exists x\varphi(x))^- = \forall x(\varphi(x)^-)$

Indukcijom po strukturi formula se lako pokaže da je u konačnoj semantici istinito:

5a. $T(\overline{\varphi}) \leftrightarrow \varphi^+$

5b. $F(\overline{\varphi}) \leftrightarrow \varphi^-$

Ova redukcija se može iterirati. Npr. vidjeli smo da je

$(\neg(F(\overline{\varphi}) \wedge T(\overline{\psi})))^+ = T(\overline{\varphi}) \vee F(\overline{\psi})$

No isti postupak možemo primijeniti i na atomarne formule $T(\overline{\varphi})$ i $F(\overline{\psi})$ itd. Redukcija prestaje kad dođemo do atomarnih formula oblika $T(t)$ i $F(t)$, gdje je $t \neq \overline{\varphi}$. Ako je t otvoren term tad se moraju ispitivati sve valuacije njegovih slobodnih varijabli što pitanje istinitosti obično ponovo prebacuje na složenije rečenice te dolazi do cirkularnosti i nastavka redukcije. Ako je pak t zatvoren term tad je on ili ime nekog objekta iz domene jezika koji nije rečenica ili je rečenična konstanta. U prvom slučaju završetak procesa je trivijalan jer su tada i $T(t)$ i $F(t)$ lažni. Ako je pak $t = \overline{C}$ za neku rečeničnu konstantu $\overline{C}$ tada imamo samoreferiranje koje smo pomoću rečeničnih konstanti i ugradili u jezik, jer $\overline{C}$ imenuje samoreferirajuću rečenicu $A(\overline{C})$:

$C : A(\overline{C})$

Istinitosnu vrijednost $T(\overline{C})$ i $F(\overline{C})$ u principu možemo riješiti analizom primarne semantike, no zanimljivo je vidjeti u kojoj mjeri to možemo opisati u samom jeziku LT. Zbog samoreferiranja proces se sada opet prebacuje na složenije rečenice:

$T(\overline{C}) \leftrightarrow T(\overline{A(\overline{C})})$

$F(\overline{C}) \leftrightarrow F(\overline{A(\overline{C})})$

pa dolazi do cirkularnosti i nastavka redukcije. Tako na primjer za jakog Lašca

$LL: \quad \neg T(\overline{LL})$

imamo:

$T(\overline{LL}) \leftrightarrow T(\overline{\neg T(\overline{LL})}) \leftrightarrow F(\overline{T(\overline{LL})}) \leftrightarrow F(\overline{LL})$

Dakle

$T(\overline{LL}) \leftrightarrow F(\overline{LL})$

Iz tog slijedi da je $U(\overline{LL})$, pa tako i $\neg T(\overline{LL})$. Tako smo iz ove cirkularne redukcije ipak izvukli zaključak da je Jaki Lažac neodređen u primarnoj semantici, a istinit u konačnoj semantici. *Na isti način bismo mogli u konačnoj semantici razriješiti svaki paradoks koji u intuitivnoj argumentaciji daje kontradikciju.* Za paradoks tipa Logičara na ovaj bismo način u konačnoj semantici mogli jedino pokazati da Logičar ne može biti lažan. To vrijedi i općenito - *u konačnoj semantici se pomoću pobrojanih svojstava jedino može formalizirati argument da rečenica ne može imati neku istinitosnu vrijednost u primarnoj semantici.* Da bismo vidjeli koja od mogućih vrijednosti je stvarna istinitosna vrijednost rečenice moramo zaviriti u semantički graf i tako utvrditi primarnu valuaciju rečenice.

Iz navedenog slijedi da se u konačnoj semantici mogu izreći sve tvrdnje o istinitosnim vrijednostima rečenica u primarnoj semantici, te da ona ima neka zanimljiva svojstva (koja su prethodno pobrojana) na osnovu kojih možemo, istina nepotpuno, rezonirati o primarnoj semantici. Tim svojstvima bismo mogli dodati aksiome kojima bismo za svaku rečeničnu konstantu $\overline{C}$ opisali istinitosnu vrijednost pripadne samoreferirajuće rečenice $C : A(\overline{C})$. Time bi se opisanom redukcijom mogla utvrditi istinitost u primarnoj semantici svake rečenice koja je bulovska kombinacija atomarnih rečenica. Preostaje ispitati kakvim bi se aksiomima mogle opisati situacije koje obuhvaćaju kvantifikaciju.

LITERATURA

[Bur79] Tyler Burge. Semantical paradox. *Journal of Philosophy*, 76:169–198, 1979.

[Fef84] Solomon Feferman. Toward useful type-free theories. i. *The Journal of Symbolic Logic*, 49:75–111, 1984.

[Fit86] Melvin Fitting. Notes on the mathematical aspects of Kripke's theory of truth. *Notre Dame Journal of Formal Logic*, 27:75–88, 1986.

[Gai92] Haim Gaifman. Pointers to truth. *Journal of Philosophy*, 89, 1992.

[GB93] A. Gupta and N. Belnap. *The Revision Theory of Truth*. MIT Press, 1993.

[Gup82] Anil Gupta. Truth and paradox. *Journal of Philosophical Logic*, 11:1–60, 1982.

[Kir92] Richard L. Kirkham. *Theories of Truth*. MIT Press, 1992.

[Kre88] Michael Kremer. Kripke and the logic of truth. *Journal of Philosophical Logic*, 17:225–278, 1988.

[Kri75] Saul A. Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.

[Mar84] Robert L. Martin, editor. *Recent Essays on Truth and the Liar Paradox*. Oxford University Press, 1984.

[Sin01] Hourya Sinaceur. Alfred Tarski: Semantic shift, heuristic shift in metamathematics. *Synthese*, 126:49–65, 2001.

[Sky84] Brian Skyrms. Intensional aspects of semantical self-reference. In Robert L. Martin, editor, *Recent Essays on Truth and the Liar Paradox*. Oxford University Press, 1984.

[Smu92] Raymond M. Smullyan. *Gödel's Incompleteness Theorems*. Oxford University Press, 1992.

[TA56] Vaught R. Tarski A. Arithmetical extensions of relational systems. *Compositio Mathematica*, 13:81–102, 1956.

[Tar33] Alfred Tarski. Pojęcie prawdy w językach nauk dedukcyjnych. *Towarzystwo Naukowe Warszawskie*, 1933. njemački prijevod, "Der Wahrheitsbegriff in den formalisierten Sprachen", Studia philosophica 1, 1935, 261–405.

[Tar44] Alfred Tarski. The semantic conception of truth. *Philosophy and Phenomenological Research*, 4:341–376, 1944.

[Č01] Boris Čulina. The concept of truth. *Synthese*, 126:339–360, 2001.

[vF66] Bas van Fraassen. Singular terms, truth-value gaps and free logic. *Journal of Philosophy*, 63:481–495, 1966.

[Vis89] Albert Visser. Semantics and the liar paradox. In Gabbay and Guenthner, editors, *Handbook of Philosophical Logic*. Reidel, 1989.

[Wol02] Jan Woleński. From intentionality to formal semantics (from Twardowski to Tarski). *Erkenntnis*, 56:9–27, 2002.

[Yab93] Stephen Yablo. Paradox without self-reference. *Analysis*, 53:251–252, 1993.

Sažetak

Ovaj rad se bavi pojmom istine i paradoksima istine. Filozofske teorije obično razmatraju pojam istine unutar šireg konteksta pitajući se postoji li uopće veza istine sa svijetom i ako postoji kakva je. Za razliku od njih ovo razmatranje je logičke prirode. Ono se bavi unutrašnjom značenjskom strukturom jezika, međusobnom semantičkom povezanošću rečenica, prije svega povezanošću rečenica koje govore o istinitosti drugih rečenica i rečenica o čijoj istinitosti govore. Paradoksi istine ukazuju na postojanje problema u našem temeljnom poimanju značenja jezika i test su za svako ponuđeno rješenje. Pri tome treba razlikovati normativni od analitičkog aspekta rješenja. Prvi nastoji osigurati da se paradoksi neće pojaviti, dok ih drugi nastoji objasniti. Naravno, važan je i praktični aspekt rješenja koji nastoji osigurati što bolji okvir za logičko zasnivanje znanja, kao i analizu prirodnog jezika.

Tarskijeva analiza istaknula je $T-shemu$ kao osnovni intuitivni princip o pojmu istine ali je ujedno pokazala i njenu nesuglasnost s klasičnom logikom. Tarskijevo rješenje se sastoji od zadržavanja klasične semantike i ograničenja $T-sheme$: o istinitosti tvrdnji nekog jezika možemo govoriti jedino u bitno bogatijem metajeziku. To rješenje odgovara ideji refleksivnosti razmišljanja i pokazalo se izuzetno plodnim za matematiku i nauku općenito. No ono je normativne prirode — paradoksi istine su izbjegnuti tako da se paradoksalne rečenice uopće ne mogu izreći u takvom jeziku. Ono se također javlja i previše restriktivnim jer iz istih razloga se u njemu ne može izreći nijedna situacija u kojoj postoji cirkularno referiranje jednih rečenica na istinitost drugih rečenica, ma koliko takva situacija bila uobičajena i bezazlena.

Kripke je pokazao da ne postoji prirodna restrikcija na $T - shemu$ već da je treba prihvatiti. No time treba prihvatiti i riskantnost rečenica, mogućnost da pod nekim uvjetima rečenica nema klasičnu istinitosnu vrijednost nego je neodređena. To vodi proučavanju jezika s trovaljanom semantikom. Kripke nije dao određeni model već teorijski okvir za ispitivanje raznih modela — svaka fiksna točka svake monotone trovaljane semantike

može biti model za pojam istine. Ta rješenja su normativne prirode — paradoksalne rečenice se mogu izraziti a kontradikcija se izbjegava tako da se neke rečenice smatraju neodređenim. Nijedno od ponuđenih rješenja ne može biti i analitičko rješenje dok se ne priloži analiza koja će sadržajno pokazati da upravo to rješenje modelira pojam istine.

Ovaj rad je pokušaj davanja analitičkog rješenja paradoksa istine. U njemu je analizirano zašto i kako se dešava da neke rečenice nemaju istinitosnu vrijednost. Pri tome se pošlo od osnovne intuicije da su paradoksalne rečenice smislene (jer dobro razumijemo što govore, štoviše to i koristimo u određivanju njihove istinitosti), ali svjedoče o neuspjehu klasične procedure određivanja istinitosti u nekim "krajnjim" situacijama. Paradoksi upravo proizlaze iz toga što klasična procedura određenja istinitosti ne mora uvijek dati i klasično pretpostavljeni (i očekivani) odgovor. Analiza pokazuju da je takva pretpostavka neopravdano poopćenje uobičajenih situacija na sve situacije. Možemo zadržati klasičnu proceduru određenja istinitosti, pa tako i unutrašnju značenjsku strukturu jezika iz koje ona proizlazi, ali moramo odbaciti univerzalnost vanjezične pretpostavke o uspješnom završetku te procedure. Svijest o tome pretvara paradokse u normalne situacije inherentne klasičnoj proceduri. Neke rečenice, mada smislene i mada vrednovane po klasičnim uvjetima neće imati istinitosnu vrijednost jer im ti uvjeti ne daju jedinstvenu istinitosnu vrijednost. Njima možemo pridružiti treću vrijednost, neodređeno, kao oznaku definitivnog neuspjeha klasične procedure. Analiza propagiranja tog neuspjeha u strukturi rečenica daje upravu jaku Kleeneovu trovaljanu semantiku, ali ne kao investigativnu proceduru, kakvom se javlja kod Kripkea, već kao klasičnu proceduru određenja istinitosti nadopunjenu propagacijom vlastitog neuspjeha. Analiza cirkularnosti u određenju klasične istinitosne vrijednosti daje kriterij kada klasična procedura uspijeva a kada ne, kada će rečenice imati klasičnu istinitosnu vrijednost a kada ne. Pokaže se da tako dobijeni skupovi istinitih i lažnih rečenica daju upravo najveću intrinsičnu fiksnu točku jake Kleeneove trovaljane semantike. Time je dana argumentacija za upravo taj izbor između svih fiksnih točaka svih monotonih semantika za model logičkog pojma istine, a ujedno je dan i njen neposredni matematički opis. Također je pokazano kako se taj jezik može semantički nadopuniti do vlastitog klasičnog metajezika koji se u mnogo čemu javlja prirodnim zaokruženjem procesa razmišljanja o istinitosnim vrijednostima rečenica datog jezika.

Summary

The thesis deals with the concept of truth and the paradoxes of truth. Philosophical theories usually consider the concept of truth from a wider perspective. They are concerned with questions such as — Is there any connection between the truth and the world? And, if there is — What is the nature of the connection? Contrary to these theories, this analysis is of a *logical nature*. It deals with the internal semantic structure of language, the mutual semantic connection of sentences, above all the connection of sentences that speak about the truth of other sentences and sentences whose truth they speak about. Truth paradoxes show that there is a problem in our basic understanding of the language meaning and they are a test for any proposed solution. It is important to make a distinction between the *normative* and *analytical* aspect of the solution. The former tries to ensure that paradoxes will not emerge. The latter tries to explain paradoxes. Of course, the *practical* aspect of the solution is also important. It tries to ensure a good framework for logical foundations of knowledge, for related problems in Artificial Intelligence and for the analysis of the natural language.

Tarski's analysis emphasized the *T-scheme* as the basic intuitive principle for the concept of truth, but it also showed its inconsistency with the classical logic. Tarski's solution is to preserve the classical logic and to restrict the scheme: we can talk about the truth of sentences of a language only inside another essentially richer metalanguage. This solution is in harmony with the idea of reflexivity of thinking and it has become very fertile for mathematics and science in general. But it has normative nature — truth paradoxes are avoided in a way that in such frame we cannot even express paradoxical sentences. It is also too restrictive because, for the same reason we cannot express a situation in which there is a circular reference of some sentences to other sentences, no matter how common and harmless such a situation may be.

Kripke showed that there is no natural restriction to the T-scheme and

we have to accept it. But then we must also accept the *riskiness* of sentences — the possibility that under some circumstances a sentence does not have the classical truth value but it is *undetermined*. This leads to languages with three valued semantics. Kripke did not give any definite model, but he gave a theoretical frame for investigations of various models — each fixed point in each three valued semantics can be a model for the concept of truth. The solutions also have normative nature — we can express the paradoxical sentences, but we escape a contradiction by declaring them undetermined. Such a solution could become an analytical solution only if we provide the analysis that would show in a substantial way that it is the solution that models the concept of truth.

Kripke took some steps in the direction of finding an analytical solution. He preferred the strong Kleene three valued semantics for which he wrote it was "appropriate" but did not explain why it was appropriate. One reason for such a choice is probably that Kripke finds paradoxical sentences meaningful. This eliminates the weak Kleene three valued semantics which corresponds to the idea that paradoxical sentences are meaningless, and thus indeterminate. Another reason could be that the strong Kleene three valued semantics has the so-called investigative interpretation. According to this interpretation, this semantics corresponds to the classical determination of truth, whereby all sentences that do not have an already determined value are temporarily considered indeterminate. When we determine the truth value of these sentences, then we can also determine the truth value of the sentences that are composed of them. Kripke supplemented this investigative interpretation with an intuition about learning the concept of truth. That intuition deals with how we can teach someone who is a competent user of an initial language (without the predicate of truth T) to use sentences that contain the predicate T. That person knows which sentences of the initial language are true and which are not. We give her a rule to assign the T attribute to the former and deny that attribute to the latter. In that way, some new sentences that contain the predicate of truth, and which were indeterminate until then, become determinate. So the person gets a new set of true and false sentences with which he continues the procedure. This intuition leads directly to the smallest fixed point of strong Kleene semantics as an analytically acceptable model for the logical notion of truth. However, since this process is usually saturated only on some transfinite ordinal, this intuition, by climbing on ordinals, increasingly becomes a metaphor.

This thesis is an attempt to give an analytical solution to truth paradoxes. It gives an analysis of why and how some sentences lack the classical truth value. The starting point is basic intuition according to which paradoxical sentences are meaningful (because we understand what they are talking about well, moreover we use it for determining their truth values), but they witness the failure of the classical procedure of determining their truth value in some "extreme" circumstances. Paradoxes emerge because the classical procedure of the truth value determination does not always give a classically supposed (and expected) answer. The analysis shows that such assumption is an unjustified generalization from common situations to all situations. We can accept the classical procedure of the truth value determination and consequently the internal semantic structure of the language, but we must reject the universality of the exterior assumption of a successful ending of the procedure. The consciousness of this transforms paradoxes to normal situations inherent to classical procedure. Some sentences, although meaningful, when we evaluate them according to the classical truth conditions, the classical conditions do not assign them a unique value. We can assign to them the third value, "undetermined", as a sign of definitive failure of the classical procedure. An analysis of the propagation of the failure in the structure of sentences gives exactly the strong Kleene three valued semantics, not as an investigative procedure, as it occurs in Kripke, but as the classical truth determination procedure accompanied by the propagation of its own failure. An analysis of the circularities in the determination of the classical truth value gives the criterion of when the classical procedure succeeds and when it fails, when the sentences will have the classical truth value and when they will not. It turns out that the truth values of sentences thus obtained give exactly the largest intrinsic fixed point of the strong Kleene three valued semantics. In that way an argumentation is given for that choice among all fixed point of all monotone three valued semantics for the model of the logical concept of truth. An immediate mathematical description of the fixed point is given, too. It has also been shown how this language can be semantically completed to the classical language which in many respects appears a natural completion of the process of thinking about the truth values of the sentences of a given language. Thus the final model is a language that has one interpretation and two systems of sentence truth evaluation, primary and final evaluation. The language through the symbol T speaks of its primary truth valuation, which is precisely the largest intrinsic fixed point of the strong Kleene three valued semantics. Its final truth valuation is the semantic completion of the first in such a way that all sentences that are not true in the primary valuation are false in the final valuation.

Životopis

Rođen sam 21. rujna 1959. godine u Šibeniku. Osnovnu školu "Bratstvo i jedinstvo" i gimnaziju "Vladimir Nazor" završio sam u Splitu. Godine 1978. upisao sam studij inžinjerske fizike na PMF-u Sveučilišta u Zagrebu. Zbog promjene interesa, nakon četiri godine studija prebacio sam se na studij inžinjerske matematike istog fakulteta. Na tom smjeru sam diplomirao 1989. godine s radom *Gödelovi rezultati i formalizacija matematike* pod vodstvom prof. dr. Dragutina Svrtana. Iste godine sam upisao poslijediplomski studij iz matematike na kojem sam magistrirao 1995. godine s radom *Neutemeljeni skupovi* pod vodstvom prof. dr. Zvonimira Šikića. Jedno vrijeme sam radio u osnovnoj i srednjoj školi, te u knjižnici Matematičkog odjela PMF-a, a od 1991. godine radim prvo kao znanstveni novak a poslije kao asistent na Katedri za matematiku Fakulteta strojarstva i brodogradnje Sveučilišta u Zagrebu. Od 1991. godine sudjelujem u radu Seminara za matematičku logiku i osnove matematike.