

Problem slijeđenja pravila u mišljenju i jeziku

Babič, Barbara

Undergraduate thesis / Završni rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Croatian Studies / Sveučilište u Zagrebu, Fakultet hrvatskih studija**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:111:382567>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-30**



Repository / Repozitorij:

[Repository of University of Zagreb, Centre for Croatian Studies](#)





SVEUČILIŠTE U ZAGREBU
FAKULTET HRVATSKIH STUDIJA

Barbara Babič

**PROBLEM SLIJEĐENJA PRAVILA U
MIŠLJENJU I JEZIKU**

ZAVRŠNI RAD

Zagreb, 2024.



SVEUČILIŠTE U ZAGREBU
FAKULTET HRVATSKIH STUDIJA
ODSJEK ZA FILOZOFIJU I KULTURU

Barbara Babič

**PROBLEM SLIJEĐENJA PRAVILA U
MIŠLJENJU I JEZIKU**

ZAVRŠNI RAD

Mentor: doc. dr. sc. Tomislav Janović

Zagreb, 2024.

Sadržaj

Sažetak.....	2
Abstract.....	3
1. Uvod.....	4
2. Wittgensteinovo shvaćanje značenja u <i>Filozofskim istraživanjima</i>	5
2.1. Značenje.....	5
2.2. Slijeđenje pravila i jezične igre.....	6
3. Kripkeovo skeptičko osporavanje značenja.....	8
3.1. Kripkeov skeptički argument.....	9
4. Konstitucijski i normativni zahtjev teorije značenja.....	10
4.1. Redukcionističke i neredukcionističke teorije.....	11
5. Upotrebna teorija značenja Paula Horwicha.....	12
6. Mogu li chatbotovi znati značenja?.....	15
6.1. Način funkcioniranja ChatGPT modela.....	15
6.2. Sličnost ljudskog mehanizma učenja jezika mehanizmu chatbota.....	16
7. Zaključak.....	17
8. Literatura	19

Sažetak: Ovaj rad problematizira fenomen značenja (*meaning*) u mišljenju i jeziku te njegovu vezu s fenomenom slijeđenja pravila. Polazi se od Wittgensteinovog odmaka od tradicionalnog shvaćanja značenja i njegovog novog razumijevanja njegove biti. Wittgenstein odriče mentalnim stanjima ključnu ulogu u uspostavljanju i razumijevanju značenja. Umjesto toga ističe važnost upotrebe jezika i analize različitih konteksta – različitih „jezičnih igara“ – u kojima riječi poprimaju značenja. Originalnom interpretacijom odgovarajućih dijelova Wittgensteinovih *Filozofskih istraživanja* Saul Kripke je postavio poznati skeptički argument čiji je zaključak da značenje ne postoji. Taj će argument biti izložen u trećem dijelu rada, dok će u četvrtom dijelu biti objašnjeni uvjeti koje teorija značenja mora zadovoljiti da bi izbjegla kripkeovskom prigovoru. Jednu takvu teoriju je predložio Paul Horwich pod nazivom „upotrebna teorija značenja“ (*Use Theory of Meaning* ili UTM). Ta će teorija biti tema petog dijela rada. U šestom, završnom dijelu razmatra se zamisao da aktualni sustavi umjetne inteligencije poznati kao *Large Language Models* (LLM) ili chatbotovi operiraju značenjima, pod pretpostavkom da se značenja riječi mogu svesti na njihovu upotrebu u skladu s kontekstualno osjetljivim pravilima.

Ključne riječi: Wittgenstein, značenje, jezik, skeptični argument, Kripke, Horwich, teorija značenja kao upotrebe (UTM), umjetna inteligencija.

Abstract: This thesis deals with the phenomenon of meaning in thought and language, and its relation to the phenomenon of rule-following. It starts from Wittgenstein's departure from the traditional conception of meaning and his new understanding of its origin. Wittgenstein rejects the idea that mental states play a decisive role in establishing and understanding of meanings. Instead, he emphasizes the importance of language use and of analyzing different contexts — different “language games”—in which words acquire their meanings. Offering an original interpretation of the relevant parts of Wittgenstein's *Philosophical Investigations*, Saul Kripke has set up a well-known skeptical argument whose radical conclusion is that meanings do not exist. This argument will be presented in the third part of the paper. In the fourth part, the conditions that a theory of meaning must satisfy in order to avoid Kripke's skepticism will be explicated. One such theory was proposed by Paul Horwich under the label “Use Theory of Meaning” (UTM). This theory will be presented in the fifth part of the thesis. The sixth, final part considers the idea that current artificial intelligence systems known as Large Language Models (LLM) or chatbots operate with meanings, provided that meanings of words can be reduced to their use by application of context-sensitive rules.

Keywords: Wittgenstein, Language, Meanings, Skeptical Argument, Kripke, Horwich, Use Theory of Meaning (UTM), Artificial Intelligence.

1. Uvod

Analitička filozofija jedan je od smjerova moderne filozofije koji se razvijao ponajviše u anglosaksonskim zemljama od početka 20. stoljeća. Ono na čemu su ustrajali analitički filozofi prošlog stoljeća prvenstveno su jasnoća i preciznost izraza. Središnja točka njihovog interesa je analiza jezika. Temelje takvoj analizi su postavljali u logici, a naglasak su stavljali na logičku jasnoću analiziranih iskaza. Kao posebna grana analitičke filozofije razvila se filozofija jezika koja problematizira prirodu i strukturu jezika, njegovu funkciju i njegov odnos prema stvarnosti. Pored pitanja značenja, ključna pitanja filozofije jezika jesu kako jezik predstavlja stvarnost i kako razumijemo i interpretiramo jezične iskaze.

Pristup problemima jezika koji je prakticirao Ludwig Wittgenstein u svojim utjecajnim djelima *Tractatus logico-philosophicus* i *Filozofska istraživanja* značajno je utjecao na razvoj filozofije jezika. Kao i ostali analitički filozofi, i sam je Wittgenstein smatrao da se filozofski problemi mogu riješiti, ili barem razjasniti, analizom jezika u kojem se ti problemi izražavaju. Fokus ovog rada će biti na jednom od problema filozofije jezika, problemu značenja – onako kako ga je postavio Ludwig Wittgenstein u svojim *Filozofskim istraživanjima*. U vrijeme kada je Wittgenstein došao na Cambridge, smatralo se kako su značenja predeterminirani, apstraktni entiteti. On je takvom metafizičkom poimanju značenja doskočio novom pretpostavkom. Odmakom od karakteriziranja značenja kao nečeg nadiskustvenog, ponudio je radikalno drukčiji pristup tom problemu: značenje je izjednačeno s upotrebom riječi. Za razumijevanje Wittgensteinovog shvaćanja tog kompleksnog problema važno je prvo izložiti njegovu kritiku tradicionalne (zdravorazumske) teorije značenja od koje se odmiče. Ta teorija polazi od pretpostavke da značenja riječi učimo usvajanjem pravila za njihovu upotrebu, to jest da ih upotrebljavamo tako da slijedimo odgovarajuća (semantička) pravila. Rješenje problema značenja koje Wittgenstein sugerira je naišlo na mnoge kritike, među njima i skeptičko tumačenje Wittgensteinova prijedloga. Najviše pažnje je privukao skeptički argument Saula Kripkea koji, uzimajući u obzir Wittgensteinove nagovještaje u *Filozofskim istraživanjima*, odriče mogućnost postojanja značenja. Među teorijama koje su pokušale izbjeći implikacije Kripkeovog argumenta redukcijom značenja na neke izvanmentalne entitete je i teorija Paula Horwicha, poznata pod nazivom „upotrebna teorija značenja“ (*Use Theory of Meaning* ili UTM).

Ključna razlika između Wittgensteinova i tradicionalnog shvaćanja značenja bit će sažeto prikazana u drugom dijelu rada. Predmet trećeg dijela je Kripkeov skeptički argument

protiv postojanja značenja. U četvrtom će dijelu biti izloženi uvjeti koje teorija značenja mora zadovoljiti da bi izbjegla tom argumentu. U kojoj mjeri je u tome uspjela Horwicheva teorija, pokazat će se u petom dijelu rada. U šestom, završnom dijelu razmatra se zamisao da aktualni sustavi umjetne inteligencije poznati kao *Large Language Models* (LLM) ili chatbotovi operiraju značenjima, pod pretpostavkom da se značenja riječi mogu svesti na njihovu upotrebu u skladu s određenim, kontekstualno osjetljivim pravilima.

2. Wittgensteinovo shvaćanje značenja u *Filozofskim istraživanjima*

2.1 Značenje

Već je na početku *Filozofskih istraživanja* naznačena spona između značenja i upotrebe koja predstavlja temelj Wittgensteinovog razrješenja problema značenja. Drugim riječima, proučavanje konkretne upotrebe jezika put je prema razumijevanju njegovog najvažnijeg aspekta – činjenice da jezični iskazi imaju značenje. Sam je Wittgenstein jasno objasnio svoje viđenje značenja: „Značenje neke riječi je njezina upotreba u jeziku.“¹ Značenje nije fiksno i nepromjenljivo, nije neraskidivo vezano uz riječ, nego proizlazi iz naše upotrebe riječi, uvijek u odgovarajućem kontekstu.

Wittgensteinovu filozofiju jezika je obilježio snažan odmak od psihologizma, tumačenja koje svodi probleme logike i filozofije na psihološke fenomene. Smatrao je da se ti problemi, među kojima je i problem značenja, ne mogu objasniti psihološkim procesima. Točnije, značenja ne ovise o individualnim mentalnim stanjima. Slično njegovom prethodniku Fregeu, vjerovao je da je izvor filozofskih problema pokušaj njihovog rješavanja proučavanjem unutarnjih stanja. Saul Kripke naglašava kako se Wittgensteinova filozofija uma često smatra biheviorističkom zato što se Wittgensteinovi stavovi tumače kao da su negativno nastrojeni prema unutarnjim (mentalnim) stanjima.² Sam Wittgenstein, govoreći o problemima filozofije, zaključuje da se oni rješavaju uvidom u rad našeg jezika, tako da se taj rad spozna.³ Ukazujući na važnu ulogu analize jezika, Wittgenstein navodi: „Filozofija je borba sredstvima našeg jezika protiv začaranosti našeg razuma.“⁴ Iz tog razloga mentalne procese prepušta psihologiji, dok se sam fokusira na objektivnu logičku strukturu jezika i stvarnosti, rješavajući filozofske probleme detaljnom analizom jezika. Kritizirajući mentalizam, koji stavlja naglasak na čovjekove

¹ Wittgenstein, 1998: 20.

² Kripke, 1982: 14.

³ Wittgenstein, 1998: 47.

⁴ *Ibid.*

unutarnje psihološke procese, isticao je da je značenje javno i očituje se u načinu upotrebe jezika. Proučavanju jezika i razumijevanju značenja pristupio je izučavajući praktičnu upotrebu jezika u kojoj se značenje oblikuje kroz slijeđenje jezičnih pravila unutar različitih situacija i konteksta – različitih „jezičnih igara“.

2.2. Slijeđenje pravila i jezične igre

Slijeđenje pravila je tema koju Wittgenstein problematizira u središnjem dijelu svojih *Filozofskih istraživanja* (FI), najizravnije u paragrafima 138–242. Na temelju sistematičnosti jezika zaključujemo kako moraju postojati neka pravila koja reguliraju upotrebu jezika. Wittgenstein tu upotrebu naziva jezičnom igrom i uspoređuje ju s bilo kojom drugom igrom koju smo naviknuli igrati. U svima njima vrijedi pravilo: ako nema pravila, nema igre.⁵ Taj zaključak prenosi i na jezik i zaključuje kako moraju postojati pravila koja uvjetuju našu upotrebu jezika. Ta nam pravila pomažu standardizirati i uniformirati izražavanje. Ako se svi ponašamo u skladu s pravilima, naše su radnje predvidljivije i time razumljivije okolini.⁶ U 54. paragrafu FI naznačena je veza između igranja jezičnih igara i slijeđenja pravila, koje se odnosi na upute za djelovanje unutar određene jezične igre. Čini se kako ljudi slijede određena jezična pravila koja se razlikuju u zavisnosti od jezične igre koja se igra. Stoga možemo zaključiti kao su jezične igre temelj funkcioniranja zajednice, učenje tih igara olakšava slijeđenje pravila. Kasnije ta pravila nastavljamo slijediti automatski, bez razmišljanja o njima. Kako sam Wittgenstein kaže, „Igra se uči tako što se promatra kako je drugi igraju. Ali mi kažemo da se igra prema tim i tim pravilima, jer promatrač može razabrati ta pravila iz prakse igre, kao neki prirodni zakon koji djelatnosti igre slijede.“⁷

Wittgenstein uzima Augustinovu teoriju učenja značenja kao primjer tradicionalne teorije. Augustin pretpostavlja pokazno ili ostenzivno učenje kao prvi stupanj učenja značenja, no naglašava da je to tek prvi korak ovladavanja jezikom. Kao djeca, mi prve riječi učimo tako što nam odrasli ukazuju na predmete, izgovarajući riječi kojima se određene stvari nazivaju. Wittgenstein nadograđuje tu tradicionalnu teoriju uvođenjem konteksta u razumijevanje značenja. Sama gramatička pravila koja naučimo još u ranoj dobi su nedostatna; potrebni su i primjeri upotrebe, situacije u kojima se određeno pravilo primjenjuje.

⁵ Vidi: Sluga, 2011: 114.

⁶ Sluga, 2011: 119.

⁷ Wittgenstein, 1998: 26.

Proces učenja značenja riječi i kasnija upotreba riječi slijeđenjem pravila u različitim situacijama mogu se ilustrirati sljedećim primjerom. Pukim pokazivanjem na predmet koji nazivamo „duda“ uz izgovaranje riječi „duda“, malo će dijete, koje još nije spremno govoriti, ali razumije što mu se govori, u toj ranoj dobi naučiti povezivati riječ koju čuje s predmetom koji vidi. Ipak, tek kada majka djetetu pruži dudu kao utjehu u situacijama kada je dijete nervozno, pritom izgovarajući riječ „duda“, dijete će shvatiti značenje te riječi. Kada dijete progovori, svoju će potvrdu ispravnosti korištenja i razumijevanja riječi „duda“, kada želi predmet za koji ta riječ stoji, dobiti u interakciji s drugim govornicima. Ako riječ koristi ispravno, povratna informacija okoline će biti poticajna. U tom trenutku dijete počinje slijediti jezična pravila. U početku će izgovarati riječ samu za sebe, no postepeno će shvatiti da će, koristeći riječ „duda“, postići što ono želi – da mu netko iz okoline pruži dudu. Kasnije će upotreba riječi postati kompleksnija, učit će pravila jezika koja reguliraju korištenje riječi u rečenici te početi samo sklapati rečenice. U početku će one biti kratke, poput „Želim dudu“. S vremenom će dijete biti sudionik sve većeg broja situacija u kojima se riječ koristi. Nakon što bude svjedok većem broju kontekstualne upotrebe riječi „duda“, moći će, sljedeći pravila, koristiti riječ i u kompleksnijim rečenicama, poput „Duda mi pomaže da lakše zaspi“ ili „Duda je za bebe“. Ovim jednostavnim primjerom ilustrirana je praksa učenja značenja u različitim situacijama ili jezičnim igrama, slijedeći određena pravila.

Problem koji se nameće je pitanje načina slijeđenja jezičnih pravila. Jezične igre igramo slijedeći određene obrasce, ali tako da o njima u pravilu ne razmišljamo niti ih odlučujemo svjesno slijediti kada se nađemo u novoj situaciji. Pravila usvajamo još kao djeca, a nakon što ih usvojimo, slijedimo ih nesvjesno, bez posebne odluke, što nam postaje prirodno. Wittgenstein to sažima na sljedeći način: „Ako slijedim pravilo, ne biram. Pravilo slijedim slijepo!“⁸ No, ako je slijeđenje pravila nesvjesno i „slijepo“ (bez posebne odluke i odabira), postavlja se pitanje je li riječ o stvarnom slijeđenju pravila (*rule-governed behavior*) ili tek o ponašanju koje možemo opisati odgovarajućim pravilima (*rule-described behavior*).⁹

Pravila značenja nisu objektivna, poput prirodnih zakona, nego su nastala u zajednici i uvjetovana su načinom upotrebe unutar zajednice. To da značenje nije objektivno, u ovom smislu znači da ne postoji po sebi, da nije metafizički neovisan entitet, nego da proizlazi iz načine upotrebe riječi unutar zajednice. Wittgenstein mnogobrojnim primjerima jezičnih igara pokazuje kako je značenje riječi određeno njezinom upotrebom, ovisno o kontekstu u kojem se

⁸ Wittgenstein, 1998: 85

⁹ Searle, 2004: 252

riječ izgovara ili pojavljuje. Osim jezičnog konteksta (rečenica i većih jezičnih cjelina) ključan je i okolišni kontekst (situacija u kojoj se govornici nalaze).

Važnost konteksta za razumijevanje značenja Wittgenstein pokazuje primjerom dvojice zidara koji igraju jezičnu igru traženja i davanja zidarskog alata. U toj igri zidar B zidaru A javlja broj ploča na kupu: „Takva jedna obavijest mogla bi, dakle, glasiti; 'Pet ploča'. Što je sad razlika između obavijesti, ili tvrdnje 'Pet ploča' i naredbe 'Pet ploča!'? Pa uloga koju izgovaranje tih riječi igra u jezičnoj igri.“¹⁰ Ovdje se hoće upozoriti na to da potpuno isti iskazi mogu imati različito značenje. Zidar može riječ „ploča“ razumjeti kao konstataciju (izjavnu rečenicu), ali ju može razumjeti i kao naredbu da donese ploču. Takvo različito razumijevanje mu je omogućeno razumijevanjem konteksta i vrste jezične igre koja se u danom slučaju igra.

3. Kripkeovo skeptičko osporavanje značenja

Koliko god nam se upravo opisano shvaćanje značenja, koje uključuje razumijevanje upotrebe i konteksta, činila plauzibilna, ona može biti izvor skepticizma u pogledu značenja. Tako je Saul Kripke uočio da mi, s obzirom na mnoštvo različitih situacija, što znači i mnoštvo različitih upotreba određenog jezičnog iskaza, nikada ne možemo sa sigurnošću reći što neki iskaz u danom trenutku znači. Razlog tome je kompatibilnost određene upotrebe (u Kripkeovom primjeru: određenog nastavka brojevnog niza) s različitim mogućim pravilima koja određuju „ispravno“ značenje (u Kripkeovom primjeru: „ispravan“ nastavak brojevnog niza).

Kripkeovu skepsu pobudio je ponajviše Wittgensteinov 201. paragraf koji glasi: „Naš je paradoks bio ovaj: Neko pravilo ne bi moglo odrediti način djelovanja budući da se svaki način djelovanja treba dovesti u suglasnost s pravilom, onda ga treba dovesti i u sklad s protuslovljem. Otuda ovdje ne bi bilo ni sklada ni protuslovlja.“¹¹

Poznato je da Wittgenstein u svojim djelima ne donosi nikakve konačne zaključke niti postavlja neku konkretnu teoriju. Njegova filozofija ima više terapeutsku namjenu ili, kao što tvrdi Hans Sluga, „[nj]egov je cilj bio da naše mišljenje skrene s filozofskog teoretiziranja i vrati ga na ono 'zdravorazumsko'“. ¹² Nije sigurno da je sam Wittgenstein imao namjeru završiti u skepsi po pitanju značenja, stoga je tumačenje navedenog paragrafa i povijesno-filozofska

¹⁰ Wittgenstein, 1998: 10.

¹¹ *Ibid*, 81.

¹² Sluga, 2011: 117.

ispravnost Kripkeove interpretacije Wittgensteina dvojben. No ona je svejedno filozofski zanimljiva, što pokazuje činjenica da je privukla ogromnu pažnju filozofa.

3.1 Kripkeov skeptički argument

Kripkeova skeptička interpretacija Wittgensteinovog shvaćanja značenja završava skeptičkim odgovorom: s obzirom na to da se stalno susrećemo s novim situacijama koje nas navode na to da određene riječi koristimo na način na koji do sada to nismo činili, te s obzirom na beskonačnost mogućih situacija u kojima bismo mogli koristiti neko pravilo koje do sada nismo koristili, od kojih ogromnu većinu nikada nećemo iskusiti, mi ne možemo uspostaviti kontinuitet korištenja riječi koji bi nam pružio uporište za uspostavljanje njezinog „pravog“ značenja. Ne postoji nikakva čvrsta činjenična evidencija (*a matter of fact*) o tome što neka riječ u danom trenutku stvarno znači. Takav način rezoniranja vodi nas do radikalnog zaključka koji glasi da značenja ne postoje.

Svoj skeptički argument Kripke je ilustrirao čuvenim primjerom slijeđenja pravila u aritmetici. Ako smo usvojili pravilo zbrajanja i ako to pravilo ispravno primijenimo, rješenje koje ćemo ponuditi za zadatak „ $68+57=?$ “ bit će „125“. No, što nas priječi da odgovorimo „5“ – pita se Kripke. Naime, nije nemoguće da smo usvojili pravilo prema kojem se uobičajeno pravilo zbrajanja primjenjuje samo u slučajevima kada su oba pribrojnika manja od 57, dok je u svim ostalim slučajevima odgovor koji trebamo dati 5. Takvo kompliciranje pravilo Kripke naziva „kvadiranje“ (*quaddition*), a zaključak njegovog argumenta glasi: ne postoji nikakva druga evidencija osim naše intuicije koja bi upućivala na to da smo u danom slučaju, kao i u svim drugim standardnim slučajevima, slijedili pravilo zbrajanja, a ne pravilo kvadiranja.

Kripke se odlučio za matematički primjer, no on je u potpunosti analogan jezičnom. Prema Kripkeu, možemo zamisliti stol koji se nazali u Eiffelovom tornju i taj predmet ne zvati stolom. Razlog tome može biti da smo do tog trenutka stolom nazivali svaki stol osim onog u Eiffelovom tornju. Naziv za tako definirani stol bi mogao biti „tabair“.¹³ Sličan skeptički argument ponudio je Nelson Goodman. Prema tom argumentu, može postojati boja nazvana „zelava“ (*grue*) koja je zelena do određenog trenutka i plava nakon tog trenutka. „zelava“ (*grue*) se definira na sljedeći način: predmet X je zelav ako i samo ako je opažen prije trenutka *t* i zelene je boje ili nije opažen prije *t* i plave je boje.¹⁴ Da bi primjer bio realističan, Goodman

¹³ Kripke, 1982: 19.

¹⁴ Cohnitz, Rossberg, 2024.

nas poziva da zamislimo smaragd s takvim svojstvom. Iako je svaki smaragd koji smo dosad vidjeli zelen, te stoga možemo induktivno zaključiti da će i svaki sljedeći biti zelen, ništa nas ne priječi da zamislimo zelavi smaragd. Ako je to tako onda naša evidencija jednako podržava zaključak da će neopaženi smaragd biti zelen koliko i zaključak da će taj smaragd biti plav.

Ovdje treba uočiti da se Kripke ne bavi epistemološkim pitanjem – kako znamo je li „125“ točan odgovor na „68+57“ – nego metafizičkim pitanjem – postoji li uopće činjenica na temelju koje bismo mogli zaključiti koje smo pravilo slijedili u svim onim slučajevima kada je kao odgovor naveden bilo koji broj manji od 125? Riječ je o metalingvističkom problemu, koji izražava za Kripkea retoričko pitanje o tome postoji li neko nad-pravilo koje bi nam nalagalo koje pravilo primijeniti u kojem slučaju?¹⁵

4. Konstitucijski i normativni zahtjev teorije značenja

Iz tako postavljenog problema značenja proizlaze dva temeljna zahtjeva koja se postavljaju pred svaku teoriju značenja koja bi željela izbjeći skeptički odgovor na Kripkeov skeptički argument – odgovor koji značenje uzima kao temeljnu, nesvodivu činjenicu (*brute fact*) našeg jezičnog ponašanja.

Prvi je zahtjev konstitucijski: Što je to što konstituira ili tvori značenje riječi? Na koju činjenicu se možemo pozvati kako bismo pokazali da smo sve vrijeme slijedili isto pravilo ili da smo uvijek mislili na istu stvar koristeći određenu riječ? Naime, činjenica je da smo se u životu našli samo u određenom (konačnom) broju situacija, te stoga, iako nam se čini da u svakoj od njih slijedimo isto pravilo, ne možemo biti sigurni bismo li to pravilo slijedili i u nekoj novoj situaciji. Točnije, ne možemo sa sigurnošću reći da smo koristeći određenu riječ uvijek mislili isto, da smo je upotrebljavali „u istom značenju“. Da bismo u to bili sigurni, morali bismo imati neko nadpravilo koje nam to jamči. Primijenjeno na Kripkeov primjer, treba nam pravilo koje bi odredilo koje pravilo smo koristili znakom „plus“. Problem je što takvim pravilom ne raspolažemo. Jedino što zamjećujemo je uniformnost upotrebe – činjenicu da i mi, kao i ljudi oko nas (jezična zajednica), imamo sklonost upotrebljavati riječi i simbole (naprimjer simbol „+“) na određeni, a ne na neki drugi, način. Ta je činjenica neobjašnjiva, ireducibilna empirijska činjenica (*brute empirical fact*) upotrebe jezika koja je, prema Kripkeovoj interpretaciji Wittgensteina, dio naše „forme života“.

¹⁵ Khani, (2024).

Drugi je zahtjev normativan: Što je to što određuje ispravnu upotrebu riječi u budućnosti? Čak i ako nam se čini da smo ispravno upotrebljavali riječi u prošlosti, ta činjenica ne može normativno djelovati na buduću upotrebu riječi. Norma ne može slijediti iz činjenice. Postoji samo upotreba iz koje ne možemo izvesti obavezu daljnje ispravne upotrebe. Odgovor na ovaj zahtjev bi, dakle, trebala biti neko preskriptivno, a ne deskriptivno objašnjenje. Taj odgovor nam mora reći kako trebamo koristiti riječi, a ne samo ponuditi opis prošle ili sadašnje upotrebe.

Ako teorija ispuni oba zahtjeva, uspješno odgovara na skeptički izazov. Nemogućnost ispunjavanja jednog ili obaju zahtjeva ruši temelje teorije značenja. Teorija koju Wittgenstein uzima kao primjer takve teorije jest tradicionalna, augustinovska teorija značenja. Prema toj teoriji, mi riječi učimo ostenzivno, tako što u djetinjstvu stvaramo vezu između predmeta na koji nam se pokazuje i riječi koju čujemo. To je objašnjenje Wittgenstein smatrao neuvjerljivim, naglašavajući važnost konteksta upotrebe riječi u učenju njihova značenja. Iako priznaje ulogu pokazivanja u učenju jezičnih značenja, naglašava da to ne bi bilo moguće bez određene obuke. Ovdje pod obukom misli na upotrebu koja se događa u zajednici i koju zajednica prešutno prihvaća kao ispravnu. Imenovanje riječi tek je priprema za upotrebu riječi¹⁶, prvi korak u procesu razumijevanja i primjene značenja. Pokazna definicija riječi, koja se zapravo svodi na imenovanje, može dovesti do razumijevanja značenja samo ako je od prije poznato što nam izgovorena riječ treba predstavljati. Wittgenstein to pokazuje na primjeru igranja šaha. Kada nekome kažemo „Ovo je kralj“, uz pokazivanje na figuru, izraz ima značenje tek ako su osobi pravila igre poznata od prije. Također, upozorava na ograničenja pokazne definicije koja se ne može primijeniti na učenje značenja osjeta (ili drugih mentalnih stanja), brojeva ili pokaznih zamjenica. Ukratko, tradicionalna teorija značenja ne može zadovoljiti ni konstitucijski ni normativni zahtjev.

4.1. Redukcionističke i neredukcionističke teorije značenja (protiv Kripkevog skeptika)

Postoje razni načini kojima bi se moglo doskočiti skeptiku i izbjeći pesimističan zaključak o nepostojanju značenja, to jest nepostojanju semantičkih pravila čijim slijeđenjem riječi zadobivaju stabilna i trajna značenja. Drukčije rečeno, postoje razne činjenice na koje se mogu navesti kao odgovor na pitanje zašto slijedimo ovo pravilo, a ne ono, odnosno na temelju čega tvrdimo da neka riječ znači ovo, a ne ono. Te se činjenice dijele u dvije kategorije: one koje su

¹⁶ Wittgenstein, 1998: 12.

nesvodive na nešto drugo i one koje to nisu. Sukladno tome, to jest ovisno o tome na koje činjenice se pozivaju, teorije značenja dijele se na redukcionističke i neredukcionističke.

Redukcionističke teorije obično se pozivaju na činjenice o prijašnjem ponašanju govornika, dispoziciju govornika za određeno ponašanje te princip ekonomičnosti objašnjenja (*Parsimony Principle*). Promatrajući govornika koji u istoj (ili sličnoj) situaciji reagira na isti (ili sličan) način, čini nam se pouzdanim zaključiti da on slijedi isto (ili slično) pravilo. Možemo reći da on ima dispoziciju za odgovarajuće (verbalno) ponašanje – sklonost prema određenoj vrsti reakcije na određeni podražaj (ako su određeni uvjeti zadovoljeni). Prema tome, značenje proizlazi iz samog ponašanja govornika, odnosno iz njegove upotrebe riječi. Ili: značenje je identično njegovoj dispoziciji da u odgovarajućim situacijama (koje uključuju i jezični i izvanjezični kontekst) upotrebljava odgovarajuću riječ. Ipak, skeptik bi i dalje mogao tvrditi da naše ponašanje, odnosno način na koji je navedena dispozicija realizirana, može biti inicirana nekim drugim pravilom, bez obzira na sličnost uvjeta u kojima se javlja.

Neredukcionističke teorije oslanjaju se na neko ireducibilno mentalno stanje govornika ili jednostavno tvrde da su činjenice o značenju primitivne, *sui generis*, to jest nesvodive na neke druge činjenice. Kao što u mnogim drugim znanstvenim teorijama mora postojati neka konačna, nesvodiva ontološka razina na koju se teorija poziva, isto vrijedi i za semantičke teorije (teorije značenja). Činjenica da pri upotrebi riječi mislimo na nešto, a ne na nešto drugo primitivna je, ireducibilna, aksiomska činjenica koju se ne treba niti može objašnjavati pozivanjem na neku drugu činjenicu. Svaki put kad koristimo riječ mi intuitivno znamo što ona znači, to jest koristimo tu riječ na točno određeni način bez nekog posebno razloga.

Ipak, skeptik bi i dalje mogao tvrditi da bi se shvaćanjem značenja kao nesvodivog ili njegovim svodenjem na neku daljnju nesvodivu činjenicu možda mogao ispuniti konstitucijski, ali ne i normativni zahtjev. Drugim riječima, mi bismo možda mogli dati odgovor na pitanje u čemu se značenje neke riječi sastoji, na što se ono svodi, ali ne bismo imali odgovor na pitanje zašto bi to bio kriterij prema kojem se svaka daljnja upotreba riječi treba ravnati. Značenje ostaje misteriozan entitet.

6. Upotrebnna teorija značenja Paula Horwicha

Kripkeovim argumentom zapadamo u prilično bezizlaznu situaciju. Čini se nemoguće odgovoriti na Kripkeov izazov teorijom značenja koja bi zadovoljila konstitucijski i normativni zahtjev. Ipak, mnogi su filozofi pokušavali naći izlaz iz te slijepe ulice. Ovdje ću se posvetiti

jednoj od takvih teorija. Riječ je o teoriji Paula Horwicha koju on sam naziva „upotrebnom teorijom značenja“ (*Use Theory of Meaning* ili UTM).

Poput Wittgensteina, Horwich značenje poistovjećuje s načinom upotrebe riječi.¹⁷ Njegova teorija (UTM) može se definirati kao komunitarijanski dispozicijski redukcionizam. Budući da se značenje poistovjećuje s upotrebom, komunitarijanski aspekt je posve prirodan – zajednica igra ključnu ulogu u uspostavljanju značenja. Horwich svoju teoriju temelji na objašnjenju doslovnog značenja (*literal meaning*), pretpostavljajući da se sve vrste nedoslovnih značenja mogu objasniti nekim dodatnim teorijama koje će nedoslovna značenja svesti na doslovna. Korelaciju između doslovnih i nedoslovnih značenja objašnjava ovako: „Naravno da bi bilo poželjno obraditi sve dodatne oblike značenja kao što su: propozicijski sadržaj, uvjeti istinitosti, referencije, govornikovo namjeravano značenje, konvencionalni pragmatički sadržaj značenja, konverzacijsku implikaturu, metaforu, ironiju itd. Ali UTM sam po sebi nema teorije o tim vrstama značenja. Najviše što se može plauzibilno tvrditi jest da ukoliko osiguramo odgovarajuće objašnjenje doslovnog semantičkog značenja utoliko ćemo biti u povoljnom položaju smisliti teorije s njom povezanih fenomena.“¹⁸

Horwich UTM gradi na pretpostavci da se na osnovi toga što ljudi očigledno koriste iste riječi na isti način može zaključiti kako postoji neki zajednički čimbenik koji regulira uvjete prihvatljivosti neke riječi, a to je ono što zovemo značenje. Značenje je, dakle, ono što je zajedničko svim onim situacijama u kojima se neka riječ *de facto* pojavljuje (u govornom, mentalnom ili grafičkom obliku), a koji onda određuje njezinu ispravnu upotrebu. Horwich dalje pretpostavlja da toj upotrebi odgovara neko idealno pravilo koje ju opisuje. U skladu s tim, daje sljedeću formalnu definiciju značenja riječi:

Značenje riječi *w* proizlazi iz njezinog nesemantičkog svojstva koje objašnjava njezinu cjelokupnu upotrebu. To je svojstvo prihvaćanja (*acceptance property*) koje ima sljedeći oblik: „da se takve-i-takve rečenice koje sadrže *w* redovito prihvaćaju u tim-i-tim okolnostima“ je idealni zakon koji upravlja upotrebom riječi *w* (uz pomoć odgovarajućih stručnjaka i uz uvjet da različite druge riječi imaju odgovarajuća značenja).¹⁹

¹⁷ Horwich, 2005: 32.

¹⁸ Horwich, 2005: 28–29.

¹⁹ *Ibid*, 28.

Treba naglasiti da Horwich svoju teoriju značenja širi i na mentalni ili „unutarnji jezik“ (u smislu neizgovorenog govora). Naime, on pretpostavlja da se javni jezik odražava u jeziku misli, te da ono što vrijedi za značenja javnog jezika stoga vrijedi i za značenja jezika misli.

Dakle, značenje riječi proizlazi iz njezine upotrebe koja određuje uvjete prihvatanja te riječi. Upotreba riječi je nesemantičko dispozicijsko svojstvo – svojstvo prihvaćenosti (*acceptance property*). Reduciranje značenja na upotrebu je ono što teoriju čini redukcionističkom. Prihvatanje rečenice u kojoj se dana riječ pojavljuje Horwich definira kao

psihološku (nesemantičku) relaciju subjekta prema rečenici koja se manifestira u našem oslanjanju na tu rečenicu u teorijskom i praktičnom zaključivanju. Slikovit način objašnjenja te ideje jest da se kaže da S prihvaća rečenicu samo ako se ta rečenica, ili njezin mentalni korelat, nalaze u S-ovom skupu vjerovanja (*belief box*).²⁰

Ako i drugi govornici na sličan način prihvaćaju druge rečenice u kojima se dotična riječ pojavljuje, uspostavlja se određena zakonolika pravilnost njezine upotrebe (*law-like regularity of usage*), a ta pravilnost ima određene normativne učinke – „sugerira“ da bi buduća upotreba trebala biti u skladu s tom pravilnošću, s „idealnim zakonom koji upravlja upotrebom riječi“. Drugim riječima, takav idealan zakon za Horwicha ima status jezičnog pravila koje članovi jezične zajednice ne samo slijede, nego ga i trebaju slijediti:

Idealni zakoni koji upravljaju značenjem riječi djeluju unutar osobe kao posljedica, barem dijelom, korektivnog usmjeravanja od zajednice (*corrective reinforcement by the community*). Stoga nije neprimjereno o tim zakonima govoriti kao o „implicitno slijeđenim pravilima”.²¹

Mogu li se ta implicitna pravila formulirati i opisati te kako bi taj opis izgledao, Horwich jasno ne kaže. Niti jasno kaže kako se točno ta pravila „slijede“, to jest u čemu se sastoji njihova normativna snaga. On jednostavno pretpostavlja da jezična norma slijedi iz jezične upotrebe: ako se riječ upotrebljava na određeni način, odnosno unutar odgovarajućih rečenica i u odgovarajućim okolnostima, a u skladu s odgovarajućim idealnim zakonom, onda iz takve aktualne upotrebe slijedi naša obveza istovjetne buduće upotrebe. Prema Horwichu, stoga, komunitarijanski dispozicijski redukcionizam može zadovoljiti i normativni zahtjev teorije značenja, čime daje odgovor na skeptički izazov.

²⁰ Horwich, 2005: 40–41.

²¹ *Ibid*, 50.

7. Može li chatbot znati značenja?

Problemi filozofije jezika koji su obilježili rad Ludwiga Wittgensteina, intrigirali su mnoge njegove prethodnike i nasljednike. Čini se da u današnjem pragmatičnom svijetu ta pitanja ne oduševljavaju istom mjerom kao što su to činila u prošlom stoljeću. Iz tog razloga ću pokušati ponuditi novi pogled na pitanje značenja, i na taj način re-aktualizirati ovu relativno iscrpljenu temu. Točnije, pokušat ću povezati Wittgensteinov problem značenja i Horwichovu upotrebnu teoriju značenja s vrlo aktualnom temom umjetne inteligencije, točnije s njezinom primjenom u vidu chatbota – konverzijskih programa univerzalne namjene.

7.1. Način funkcioniranja ChatGPT modela

Za početak, treba u kratkim crtama objasniti kako suvremeni chatbotovi, od kojih je najpoznatiji ChatGPT, funkcioniraju.²² Takvi programi se nazivaju velikim jezičnim modelima (*large language model* – *LLM*), a osnova njihova softwarea su takozvani transformeri – vrsta neuronske mreže koja koristi mehanizam nazvan „*self-attention*“ koji omogućava modelu da procijeni relevantnost različitih riječi u rečenici s obzirom na aktualni kontekst. Model poput ChatGPT-a zahtijeva više slojeva transformera kroz koje se model usavršava. Procesiranje podataka kroz slojeve transformera se naziva treniranje. Model u početku dobiva ogromne količine tekstualnih podataka iz raznih izvora, nakon čega se podaci „tokeniziraju“, dijele na manje jedinice. Inicijalnim inputima (unesenim podacima) pridružene su odgovarajuće težinske vrijednosti (*weights*) koji određuju relevantnost određenog inputa za dani output. Oni su u početku postavljeni nasumično, a kroz različite cikluse treniranja se ažuriraju. Model „uči uz nadzor“ (*supervised learning*), to jest pokušava predvidjeti sljedeću riječ u rečenici na temelju prethodnog „iskustva“ – nizova slova i riječi kojima je bio opskrbljen. U fazi treniranja model se pušta ograničenom broju korisnika („beta testera“) kako bi se dobile povratne informacije iz stvarnog svijeta. Nakon što se model pusti u upotrebu, on više ne „uči“.

Za lakše razumijevanje te kompleksne strukture, funkcioniranje ChatGPT modela ću objasniti na primjeru funkcioniranja logističke regresije. Logistička regresija predstavlja mali model neuronske mreže koji se koristi za modeliranje binarnih ovisnih varijabli. To znači da se koristi kada je rezultat koji pokušavamo predvidjeti određen kao jedan od dva moguća ishoda, obično označenih kao 0 i 1. Pretpostavimo da želimo stvoriti sustav koji će prepoznati životinju

²² „How ChatGPT Works Technically: ChatGPT Architecture“, 2023.

na slici, točnije pretpostavimo da od našeg programa tražimo da odredi je li životinja na slici ljenjivac ili nije. Inputi kojima ćemo opskrbiti našu logističku regresiju će biti neke fizičke karakteristike životinja, poput „krzno“, „nokti“, „boja krzna“, „uši“ itd. Težinske vrijednosti koje ćemo pridružiti inputima odgovaraju tome koliko je dano svojstvo relevantno za naš konačni cilj – prepoznavanje ljenjivca. Dakle, pretpostavlja se da ćemo veću vrijednost pridati karakteristikama koje su tipične za ljenjivca. Kroz fazu treniranja sustav prilagođava svoje težinske vrijednosti da bi u fazi predikcije mogao što uspješnije odgovoriti na izazov – odgovoriti je li životinja ljenjivac ili nije.

7.2. Sličnost ljudskog mehanizma učenja jezika mehanizmu chatbota

Sada možemo prijeći na relevantnost funkcioniranja modela poput ChatGPT-a za razumijevanje Wittgensteinovog problema značenja i Horwichove upotrebne teorije značenja, točnije na uspostavljanje analogije između učenja kod jezika chatbota i ljudskog subjekta. Kao što je pokazano, ChatGPT uči tako što se „hrani“ podacima, nizovima znakova. Iako kod chatbota izostaje vizualna komponentna učenja, riječi kao nizovi znakova su ipak kontekstualno određene: uklopljene su u rečenice kao veće nizove znakova, a ove su uklopljene u odlomke kao još veće nizove znakova itd. Na taj način, opskrbljen mnogobrojnim i raznolikim primjerima upotrebe riječi, model sam nastavlja proizvoditi nizove znakova, koristeći naučene obrasce iz mnoštva tekstualnih podataka na kojima je treniran. Kreiranje tih nizova determinirano je statističkim vezama između riječi i fraza u skupu podataka za treniranje. Također, u fazi učenja, model prolazi fazu utjecaja „stranog“, ljudskog faktora, pri čemu je njegova reakcija statistički pojačana reakcijom „sugovornika“, i to u onoj mjeri u kojoj se riječi koriste u skladu s uvriježenim normama. Možemo primijetiti da je taj proces učenja vrlo sličan ljudskom. Od najranije dobi ljudi bivaju opskrbljeni određenim brojem informacija u određenom kontekstu. Nakon toga treniraju svoju upotrebu riječi s kojima su već imali doticaja, uzimajući u obzir reakciju okoline. Kao i chatbot, čovjek koristi samo riječi koje se nalaze u njegovoj „bazi riječi“ koju je stekao „treniranjem“. Na osnovi tih poznatih nizova riječi uči proizvoditi nove rečenice, uključujući one koje nikada prije nije proizveo. To ne znači da je proces učenja jezika kod ljudi potpuno istovjetan procesu učenja jezika kod stroja, nego da postoje određene sličnosti u mehanizmu učenja i njegovoj osjetljivosti na kontekstualne informacije. Unutarnji, mentalni procesi, koji se tradicionalno i zdravorazumski smatraju ključnima za učenje jezika, ovdje su zanemareni – kao što je to upravo slučaj i kod Wittgensteina koji se usredotočio na bihevioralne aspekte učenja i upotrebe jezika.

Vratimo se sada Horwichovoj teoriji prema kojoj je značenje izjednačeno s upotrebom riječi. Ako značenje riječi doista proizlazi iz njezinog nesemantičkog svojstva, iz „svojstva prihvatanja“ (*acceptance property*) koje proizlazi iz složenih socijalnih uvjeta pod kojima se dotična riječ prihvata u skladu s nekim „idealnim zakonom“, onda nema razloga tvrditi da su jedino ljudi sposobni za znanje značenja. Sam ChatGPT na pitanje „Što misliš, koliko je način na koji ti učiš jezik analogan ljudskom načinu učenja jezika?“ odgovara kako se način na koji jezični modeli i ljudi uče jezik bitno razlikuje. Kao jedan od primjera navodi razumijevanje: „Ja nemam stvarno razumijevanja ili svijesti o jeziku ili kontekstu. Moji odgovori su generirani na temelju obrazaca u podacima, bez stvarnog semantičkog razumijevanja.“ Međutim, ako se značenja svode na upotrebu, onda ni čovjek ne mora imati „stvarno semantičko razumijevanje“ da bi mogao raspolagati značenjima riječi. Štoviše, upravo je nesemantičko svojstvo upotrebe ono ključno za postojanje značenja.

Mogli bismo zaključiti da je chatbot poput ChatGPT-a čak uspješniji od čovjeka u uspostavljanju i upotrebi značenja, u primjeni „idealnog zakona“ koji spominje Horwich, jer ima pristup dalekom većem broju primjera upotrebe riječi u različitim kontekstima. Iako takvim shvaćanjem značenja ne možemo sasvim otkloniti Kripkeov skeptički argument, sa stajališta upotrebe taj argument postaje nevažan jer značenje riječi prestaje biti ovisno o slijeđenju bilo kojeg drugog pravila osim statističkog pravila aktualne upotrebe od članova jezične zajednice. Štoviše, moglo bi se tvrditi da su modeli poput ChatGPT-a bliži idealnom značenju od nas ljudi jer uzimaju u obzir sve relevantne kontekste, ili veliku većinu njih, u kojima se neka riječ može pojaviti, što nijedan pojedinac ne može.

7. Zaključak

Čak i ako se ne vodimo ambicioznom nakanom rješavanja filozofskih problema rješavanjem problema jezika, pitanje značenja ostaje intrigantno. Evidentno je da se komunikacija odvija bez poteškoća. Ljudi se uglavnom dobro razumiju iz čega smo skloni zaključiti da značenja moraju postojati. Ipak, kada bismo trebali točno definirati što je to što tvori (konstituirati) značenje, to nam ne polazi za rukom. S druge strane, također se čini protuintuitivnim prihvatiti skeptički odgovor kao konačno rješenje tog problema. Stoga ne čudi da su predložene različite teorije koje su pokušale dati odgovor na skeptički izazov. Ipak, snaga skeptičkog argumenta je upravo u tome što, svaka teorija koja ga pokušava neutralizirati ima neke nedostatke zbog kojih ne uspijeva ispuniti dva zahtjeva uspješne teorije značenja – konstitucijski i normativni zahtjev.

U ovom radu je u osnovnim crtama predstavljena jedna takva teorija – Horwichova upotrebna teorija značenja (UTM) – čija je namjera neutralizirati skeptički prigovor svođenjem značenja riječi na nesemantičko svojstvo njezine prihvaćanosti u jezičnoj zajednici, uzimajući u obzir najširi mogući skup jezičnih i izvanjezičnih konteksta njezine upotrebe. Prema Horwichovoj zamisli, svojstvo prihvaćanosti ima normativni učinak, što znači da njegova teorija ujedno zadovoljava normativni zahtjev teorije značenja .

Ako Horwichovu teoriju prihvatimo kao plauzibilnu teoriju značenja, onda nema razloga da sposobnost znanja značenja ne pripišemo i sustavima umjetne inteligencije, na primjer suvremenim chatbotovima. Tu implikaciju navedene teorije možemo tumačiti i negativno – kao kritiku upotrebne teorije značenja. Naime, intuicija nam govori da je raspolaganje značenjima ekskluzivno ljudska sposobnost, pa njezino pripisivanje drugim bićima ispada kao nedostatak teorije. No čini se da osim intuicije ne postoji ništa na što bismo se mogli pozvati kako bismo znanje značenja ograničili na nas ljude, pa tako ni tumačenje navedene implikacije kao kritike upotrebne teorije značenja nije opravdano.

Ne možemo sa sigurnošću reći hoće li filozofski problemi jezika biti ikada riješeni. Još je teže reći hoće li temeljni filozofski problemi biti ikada riješeni. Ipak, postavljanjem skeptičkih izazova, kao i novim pokušajima odgovora na te izazove ipak se, makar sitnim koracima, približavamo rješenju velikih i naizgled nerješivih pitanja kao što je pitanje značenja i slijeđenja pravila.

8. Literatura

- Cohnitz, D., Rossberg M. (2024.) „Nelson Goodman“, *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/spr2024/entries/goodman/> (stranica posjećena: 17. lipnja 2024.).
- „How ChatGPT Works Technically: ChatGPT Architecture“, audiovizualni materijal, Youtube, 24.4. 2023, <https://youtu.be/bSvTVREwSNw?si=vm77b8wGBA1Yr6Kb> (stranica posjećena: 17. lipnja 2024.)
- Horwich P. (2005.) *Reflections on Meaning*, New York: Oxford University Press.
- Khani, A.H. (2024.) „Kripke's Wittgenstein“, *The Internet Encyclopedia of Philosophy*, <https://iep.utm.edu/> (stranica posjećena: 13. travnja 2024.).
- Kripke, S. A. (1982.) *Wittgenstein on Rules and Private Language*, Cambridge: Harvard University Press.
- Searle, J. R. (2004.) *Mind: A Brief Intorduction*, New York: Oxford University Press
- Sluga, H. (2011.) *Wittgenstein*, Chichester: Wiley-Blackwell.
- Wittgenstein, L. (1998.) *Filozofijska istraživanja*, prev. I. Mikecin, Zagreb: Globus.