

KORPUSNE METODE IN NJIHOV ODSEV V JEZIKOSLOVNIH TEORIJAH 20. STOLETJA

Simon KREK

Amebis, d.o.o., Kamnik

Institut "Jožef Stefan", Laboratorij za umetno inteligenco

Krek, S. (2013): Korpusne metode in njihov odsev v jezikoslovnih teorijah 20. stoletja. Slovenščina 2.0, 1 (1): 4–23.

URL: http://www.trojina.org/slovenscina2.0/arhiv/2013/1/Slo2.0_2013_1_02.pdf.

V 20. stoletju se je strukturalizem vzpostavil kot osrednja jezikoslovna teorija, v prvi polovici stoletja predvsem s svojim začetnikom Ferdinandom de Saussurjem, v drugi polovici pa z likom Noama Chomskega. Zadnji je vztrajno zavračal smiselnost analiz obsežnih količin besedil, ki jih je obravnaval kot nezanimive v primerjavi z jezikovno intuicijo naravnega govorca. Vzporedno s strukturalizmom so se množile tudi jezikoslovne smeri, ki so opozarjale na nezadostnost prevladujoče jezikoslovne paradigme in na teoretske uvide, ki jih je omogočila šele sistematična analiza velikih količin besedil. Prispevek obravnava dileme, ki izhajajo iz navedene dihotomije in umešča t. i. korpusno jezikoslovje v širši jezikoslovni kontekst.

Ključne besede: korpusi, jezikoslovje, strukturalizem, semantika, skladnja, Saussure, Chomsky

1 UVOD

"/R/azlika med besedoslovnim in skladenjskim dejstvom lahko izgine /.../"¹
(Ferdinand de Saussure [1916])

"/I/skanje semantično zasnovane definicije 'gramatičnosti' bo jalovo /.../"²
(Noam Chomsky [1957])

¹ Citirano po: de Saussure 1997: 151-152.

² Citirano po: Čomski 1984: 24, prev. S. K.

"Tradicionalno ločevanje med besediščem in slovnico je napačno."³
(Ray Jackendoff [2005])

Jezikoslovje kot veda, katere znanstveni objekt je polje človeške govornice,⁴ je v 21. stoletje vstopilo razdeljeno v množico nasprotnih taborov z enakim številom med sabo različnih teorij, ki se strinjajo takorekoč le o svojem osnovnem objektu, dojetem v najširšem smislu. Kdorkoli v tem času želi uporabljati pojme, kot so besedilni korpus, jezikovni podatek, slovnica, slovar itd., se mora vprašati, kam se v tej konkurenci znanstvenih konceptov in metod uvršča t. i. korpusno jezikoslovje in se tako ali drugače opredeliti glede alternativ. V prispevku nikakor ne želimo podajati pregledne sheme razvoja jezikoslovja v dvajsetem stoletju, temveč bomo od izbranih teorij izpostavili le dele, ki so relevantni v okviru umestitve in razdelave predpostavk korpusnega jezikoslovja. V okviru tega vprašanja se kaže predvsem velik prepad med strukturalnim jezikoslovjem druge generacije z osrednjim likom Noama Chomskega in izpeljankami iz njegovega štirideset let starega osnovnega programa (minimalistični program, optimalnostna teorija), ter jezikoslovnimi smermi, ki jih je težje opredeliti z enim samim atributom, vsekakor pa lahko rečemo, da jih opredeljuje močno opiranje na empirične jezikovne podatke, dajanje prednosti pojmu jezika kot (družbenega) diskurza ter temeljno nepristajanje na interpretacijo jezikovnih pojavov z izpostavljanjem ene jezikovne ravnine na račun drugih.

2 PRVA GENERACIJA STRUKTURALIZMA

Dvajseto stoletje se je na polju humanistike (in družboslovja) začelo z več revolucionarnimi preobrati, ki so radikalno spremenili pogled na znanost o jeziku. Pojem znanstvenosti izpostavljamo namerno, kajti razmišljanja o jeziku so stara toliko kot pisni viri in sistematična obravnava jezika je stara

³ Citirano po: Jackendoff, Culicover 2005: 26, prev. S. K.

⁴ Čeprav Saussure odločno zatrjuje, da govornica (*langage*) v pomenu človeške zmožnosti sporazumevanja z jezikovnimi znaki ne more biti znanstveni objekt, pojem uporabljamo prav v tem pomenu.

vsaj 2.500 let, zanimanje za vprašanja o naravi in izvoru jezika pa še starejše in je zaposlovalo takorekoč vse vidnejše mislece od začetkov razumske refleksije človeškega bivanja dalje. Jezikoslovje je svojo znanstveno opredelitev v smislu comtovskega znanstvenega pozitivizma dobilo pozno, šele v devetnajstem stoletju, s historičnim oz. primerjalnim jezikoslovjem, ki si je za privilegirani objekt izbralo genealoški razvoj sorodnih jezikov, pri čemer je ob tej znanstveni disciplini seveda ves čas obstajala starejša v luči ideološkega formiranja evropskih nacionalnih narcizmov od renesanse naprej spodbujana jezikovna dejavnost opisovanja konkretnih jezikov v slovnica, slovarjih, ter še starejše poučevanje retorike, učenje tujih jezikov, prevajanje in druge jezikovne prakse, torej dejavnosti, ki so se opirale na sholastično tradicijo razmišljanj o jeziku. To se zdi pomembno, kajti tako kot druga področja na polju humanistike je jezikoslovje v primerjavi s trdimi znanostmi že od začetka v domala shizofrenem stanju opredeljevanja svojega teoretskega objekta v smislu dihotomije teoretsko/aplikativno, čeprav iz samega dejstva, da humanistika raziskuje človeške prakse, izhaja njena neprimerljivost z metodami trdih znanosti, ki najdevajo svoj objekt v nevtralnem od človeka neodvisnem zunanjem svetu naravnih pojavov.⁵

Ko je torej jezikoslovje našlo klasifikabilni objekt, ki je bil s stališča racionalizma 18. in 19. stoletja dojet kot ustrezen, ker je spominjal na naravne procese, se je lahko samoopredelilo kot znanost in začelo živeti svoje znanstveno življenje. V tako celo stoletje obdelovano znanstveno polje je vstopal Saussure. Tudi on pa je kot dodobra izučen v historični primerjalni metodi iskal odgovor na vprašanje znanstvenosti svojega področja, vendar izrazito nezadovoljen s preteklimi rešitvami.⁶ Tako je sam opredelil svoj

⁵ V to, koliko so same ideološke predpostavke naravoslovnih znanosti sporne in kako pogled določa sam objekt opazovanja, se na tem mestu ne želimo spuščati, referiramo pa denimo na problematiko struktur znanstvenih revolucij v (Kuhn 1998).

⁶ "Jezik jim je veljal za posebno področje, četrto kraljestvo narave; zato so se lahko uveljavili načini razmišljanja, ki jih druge znanosti ne bi prenesle. Danes ni mogoče prebrati osem do deset vrstic, ne da bi ostrmeli nad miselnimi bizarnostmi in pojmi, s katerimi so jih upravičevali." (de Saussure 1997: 14)
(*opomba se nadaljuje na naslednji strani*)

znanstveni program s potezo, ki je imela trajne posledice za precejšen del kasnejšega jezikoslovja:

Lahko si torej zamislimo *znanost, ki proučuje življenje znakov v naročju družbenega življenja*; bila bi del socialne psihologije in torej obče psihologije; imenovali jo bomo *semiologija* /.../ Jezikoslovje je le del te splošne znanosti, zakone, ki jih bo odkrila semiologija, bo mogoče uporabiti tudi v jezikoslovju, jezikoslovje pa se bo tako pridružilo jasno določenemu področju v celoti človeških dejstev. /.../ tu naj poudarimo le eno: jezikoslovju smo lahko zdaj prvič določili mesto med znanostmi le zato, ker smo ga pridružili semiologiji. (de Saussure 1997: 27)

Izpostavljanje znakovne narave jezika ter znamenite štiri osnovne dihotomije, ki jih izpostavlja Saussure (jezik/govor, označevalec/označenec, sintagmatika/paradigmatika, sinhronija/diahronija), so na nek način prešile⁷ ne le jezikoslovno, temveč tudi širše humanistično polje. Hkrati pa so potencialnega jezikoslovca postavile pred nekaj dejstev: raziskuje lahko bodisi diahroni razvoj ali sinhrono stanje nekega jezika kot semiotičnega sistema, označevalec je privilegirana ravnina jezikovnega izraza, ker svet "zunajjezikovnih referentov nenehno uhaja eksaktnemu in racionalnemu opisu" (Škiljan 1997: 272), edini možni objekt znanstvene raziskave je jezik (*langue*), ne pa govor (*parole*).

Če je ideja, da je jezik le eden od znakovnih sistemov, kjer je posamezni element določen izključno z razmerji do drugih elementov, v zadostni meri abstrahirala del govornice (*langage*), da je ta – namreč *langue* – lahko postal vreden znanstveni objekt, je hkrati s to potezo celotno polje konkretnih ubeseditev abstraktnega sistema razmerij jezikovnih znakov izključila iz znanstvene interpretacije. *Parole* je tako postal trajno obeležen z oznako

"Vse to se bo proti moji volji končalo s knjigo, v kateri bom brez entuziazma pojasnil, zakaj v lingvistiki ni v rabi niti en sam termin, ki bi mu lahko pripisal kakšen smisel." (pismo A. Meilletu, v: Škiljan 1997: 268)

⁷ V lacanovskem pomenu, da nenadoma isti objekt/situacijo/razmerje zagledamo v popolnoma drugi luči, čeprav se na sebi ni spremenil/a.

neznanstvenosti in večina jezikoslovcev, ki so strukturalno metodo vzeli za svojo, je konkretne ubeseditve enostavno ignorirala in se v celoti zanašala na raziskovanje medsebojnih razmerij prepoznanih elementov znotraj formalnega strukturalnega sistema. To je pomenilo, da se raziskuje tisti del jezika, ki za razliko od kaotičnosti *parole* kaže red in urejenost:

Celotni jezikovni sistem resda temelji na iracionalnem načelu arbitrarnosti znaka, ki bi, če bi ga uveljavljali brez zadržka, peljal v skrajno zapletenost; a duhu uspeva, da v *nekaterih dele znakovne množice* vnese načelo reda in urejenosti, in prav v tem je vloga relativne motiviranosti. Če bi bil mehanizem jezika popolnoma racionalen, bi ga bilo mogoče proučevati v njem samem; ker pa je *zgolj delov popravek po naravi kaotičnega sistema*, se moramo postaviti na gledišče, ki nam ga narekuje sama narava jezika, in zato proučujemo ta mehanizem kot omejitev arbitrarnosti. (de Saussure 1997: 148, poud. S. K.)

S tem jezikoslovcem ni naložil lahke naloge. Iz paradigmatsko/sintagmatske dihotomije izhaja, da na omejitev tradicionalnih slovnice na oblikoslovje in skladnjo ter od nje povsem ločenega "besedoslovja" ni mogoče pristati, temveč "razlika med besedoslovnim in skladenjskim dejstvom lahko izgine /.../" (ibidem), pri čemer je besedoslovcem predpisal tudi način obravnave njihovega objekta, ki predpostavlja predvsem enako metodo, ki je v temelju strukturalna, znakovna: "Na drugi strani pa se tiste besede, ki niso nesestavljene in nerazstavljive enote, bistveno ne razlikujejo od stavčnih členov, od skladenjskih dejstev; nižje enote, ki take besede sestavljajo, se povezujejo po istih temeljnih načelih kot besede, kadar se povezujejo v skupine" (ibid: 151-2). Lahko bi rekli, da se v teh stavkih skriva velik del programa kasnejših strukturalnih jezikoslovnih smeri, ki so po vseh jezikovnih ravnanjih razgrajevale jezik na najmanjše delce in po principu arbitrarne, vendar motivirane kombinatorike sestavljale jezikovne monade⁸ v smiselne nize. Če je Saussure skušal rešiti kaotičnost svojega znanstvenega objekta z razgrajevanjem na jezikovne atome in iskanjem pravil, analognih pravilom

⁸ Monada v širšem pomenu samozadostne enote, ki jo je sploh mogoče dojeti kot enoto. Pojem se opira na starogrško in po njej leibnizevsko tradicijo.

šahovske igre, za njihovo kombiniranje, so druge znanstvene jezikoslovne avtoritete tistega ali malce poznejšega časa, ki ravno tako v širši opredelitvi spadajo v strukturalno jezikoslovje, nad pomensko platjo jezika takorekoč dvignile roke. V eni od vplivnejših jezikoslovnih monografij med obema svetovnjima vojnama, Bloomfieldovi *Language*, tako stoji trditev:

Pomen glasovne oblike lahko natančno definiramo, če je ta pomen povezan z nečim, o čemer kaj vemo na znanstven način. Na primer, v okviru kemije in mineralogije lahko definiramo imena mineralov, tako kot v primeru, če rečemo, da je običajni pomen angleške besede *salt* (sl. *sol*, op. prev.) "natrijev klorid (NaCl)", imena rastlin in živali pa lahko definiramo s pomočjo tehničnih terminov v botaniki in zoologiji, nikakor pa ne moremo natančneje definirati besed, kot so *love* in *hate* (slo. *ljubezen*, *sovrašтво*, op. prev.), ki se tičejo stanj, ki ne morejo biti natančneje razvrščena – in prav teh je velika večina.⁹

Tak radikalni dvom o možnosti znanstvenosti interpretacije pomena jezikovnih znakov ni kompatibilen s saussurjanskim strukturalizmom, vendar kaže na temeljne težave pri odnosu lingvističnega znaka z zunanjo pojavnostjo, označencem,¹⁰ ker ima ta spoznavno naravo v epistemološko filozofskem smislu, od česar pa se želi jezikoslovje kot znanost ograditi, ker se želi ukvarjati zgolj in samo z jezikom in njegovimi notranjimi razmerji. Bloomfield kot deklarirani empirist pa vendarle ni ostal pri dokončnosti zgornje sodbe, temveč je v skladu s svojimi behaviorističnimi izhodišči odgovor iskal v empiričnih podatkih, zato v zgodovini jezikoslovja ostaja z oznako kontekstualista: "jezikoslovci, kot je Bloomfield, so zagovarjali šibkejšo različico kontekstualizma, pri kateri je odnos med sobesedilom in pomenom bolj posreden in ki jo lahko izrazimo s formulo kot 'pomen je v končni konsekvenci izpeljiv iz sobesedila, ki ga je mogoče opazovati' ali 'pomen je v

⁹ Citirano po: Bloomfield 1933: 139, prev. S. K.

¹⁰ Podobno gre razumeti tudi Hjelmslevovo substanco vsebine: lingvistični znak je "entiteta, ki jo generira povezava med izrazom in vsebino" Hjelmslev 1961: 47, prev. S. K. Vsebina je dalje razdeljena na formo in substanco, substanca vsebine pa postane vidna s "projekcijo forme na pomen, tako kot razprta mreža vrže senco na površino" (ibid: 57).

končni konsekvenci mogoče skrčiti na sobesedilo, ki ga lahko opazujemo" (Leech 1981: 64). Ta način razmišljanja se postavlja nasproti mentalizmu, pri katerem je jezik zunanji izraz človekovih mentalnih stanj, idej ali konceptov in svoje korenine vleče takorekoč iz aristotelovskih postulatov,¹¹ svoje odločne zagovornike znotraj jezikoslovja pa ima tudi na začetku 21. stoletja. Razlika med kontekstualisti in mentalisti iz časa med svetovnima vojnama je pomembnejša, kot se morda zdi, saj lahko prav v njej najdemo zametke idejnih razhajanj, ki bodo na koncu 20. stoletja delili jezikoslovje.

Bloomfieldov istočasni evropski pandan je bil J. R. Firth, ki je ravno tako odločno obsodil možnost objektivne interpretacije jezika izven jezikovnih pojavov, ki jih je mogoče opazovati:

Če jezik dojemamo kot "ekspresiven" ali "komunikacijski", s tem impliciramo, da je instrument notranjih mentalnih stanj. In ker o teh vemo tako malo, celo ob kar najbolj pazljivi introspekciji, problem jezika postane tem bolj skrivnosten, kolikor bolj ga skušamo razložiti z naslavljanjem na notranja mentalna dogajanja, ki jih ni mogoče opazovati. Če besede obravnavamo kot dejanja, dogodke, običaje, svojo raziskavo omejimo na to, kar je objektivno v skupnem življenju s soljudmi.¹²

Če zgornjo trditev vzamemo resno, jezikoslovno gledano o mentalnih stanjih, ki generirajo jezikovno komunikacijo, ne moremo reči nič znanstveno zanesljivega, lahko pa objektivno preko beleženja in opazovanja "govornih dejanj" – ker jezik v temelju zadeva "skupno življenje s soljudmi" – deduciramo jezikoslovne generalizacije o teh dejanjih. Na tem izhodišču je kasneje dejansko temeljila iniciativa Firthovih naslednikov z idejo, da o jeziku težko rečemo karkoli brez čim večje množice dejanskih zabeleženih konkretnih govornih dejanj, kar je v končni konsekvenci pripeljalo do prvih

¹¹ "Govorjene besede so simboli miselnih izkustev in pisane besede so simboli govorjenih besed. Tako kot vsi ljudje ne pišejo enako, tako tudi ne uporabljajo enakih zvokov za govor, toda miselna izkustva, ki jih te neposredno simbolizirajo, so skupna vsem, tako kot so skupne stvari, katerih podobe so naša izkustva." (Aristotle, *On interpretation*, Part 1, <http://classics.mit.edu/Aristotle/interpretation.html>, prev. S. K.)

¹² Citirano po: Firth 1964: 173, prev. S. K.

korpusov, namenjenih jezikoslovnim raziskavam.

Zgodovinsko gledano je po radikalnem prvem strukturalnem preobratu, ko je bila izražena želja po novi znanosti, ki bo raziskovala jezik kot sistem znakov, naslednjih štirideset let potekal poskus, kako to znanost izpeljati ne le na ravni fonologije, kjer je sicer dobila izvrstno potrditev in množico konkretnih izpeljav, temveč kot interpretativni aparat na ravni celotnega polja govorice (*langague*), pri čemer je v metodološki doslednosti najdlje verjetno prišel kopenhagenski lingvistični krožek. Pomembno za našo obravnavo je, da je prvotni strukturalizem, ki ga definira čas nekako do petdesetih let prejšnjega stoletja oz. do prvih objav Zeliga Harrisa in Noama Chomskega, brezkompromisno vztrajal prvič, pri nedeljivosti jezikovnega izraza in vsebine, pri čemer je bila nedokončanost naloge apliciranja enake metode na celotno raziskovanje jezika pravzaprav ves čas poudarjena,¹³ ter drugič, pri temeljni družbeni naravi jezika,¹⁴ kjer jezikovni sistem po sebi ne more biti jamstvo za identiteto vrednosti jezikovnega znaka, temveč je "jamstvo v skrajni konsekvenci družba, kolektiv, ki ta sistem sprejema" (Škiljan 1997: 275).

3 DRUGA GENERACIJA STRUKTURALIZMA

Druga generacija strukturalizma deklarativno dekonstruira obe navedeni predpostavki, pri čemer je končni motiv pri tem – ko se je na nek način izkazalo, da je projekt celostnega znanstvenega opisa jezika v strukturalni metodi potencialno neuspešen – prav ohranitev prestižne znanstvenosti

¹³ Dilemo je ubesedil E. Benveniste v govoru na 9. lingvističnem kongresu v Cambridgeu na MIT leta 1962: "S tem pa se nenadoma pojavi problem, ki muči vsi sodobno lingvistiko, problem razmerja oblika : pomen, ki bi ga mnogi lingvisti radi reducirali le na pojem oblike, ne da bi se jim pri tem uspelo znebiti njenega korelata, pomena. Česa vse niso poskušali, da bi se pomenu izognili, ga ignorirali ali izključili! Zaman: ta Meduzina glava je še vedno tu, v središču jezika, in zaslepljuje tiste, ki ga opazujejo" (Benveniste 1988: 137).

¹⁴ "Ko kakšen semiološki sistem postane dobrina skupnosti, ga zaman ocenjujemo zunaj tistega, kar mu daje to kolektivno naravo, in če hočemo spoznati njegovo bistvo, je dovolj, če si ogledamo, kaj je v odnosu do kolektivtete." (Saussure v: De Mauro 1965, citirano po: Škiljan 1997: 275)

jezikoslovja.¹⁵ Prvo predpostavko o inherentni in nedeljivi povezavi pomena in izraza Chomsky zanika takoj na prvih straneh prelomne *Syntactic structures* iz l. 1957: "pojma 'gramatično' ne moremo povezati s 'smiselnim' ali 'signifikantnim' v nekem semantičnem smislu /.../ iskanje semantično zasnovane definicije 'gramatičnosti' bo jalovo" (Čomski 1984: 23, prev. S. K.) ter "/z/di se, da smo prisiljeni zaključiti, da je gramatika avtonomna in da ni odvisna od pomena, ter da nam probabilistični modeli ne dajejo uvida v nekatere od osnovnih problemov sintaktične strukture" (ibid: 25). Zanihanje družbene konstruiranosti jezikovnega znaka je programsko izraženo takorekoč hkrati, na isti strani: "predpostavljamo, da obstaja intuitivno poznavanje gramatičnih stavkov angleškega jezika /.../" (ibid: 21) ter "jasno je, da skupine gramatičnih stavkov ni mogoče identificirati z nekim korpusom dokazov, ki jih zberejo raziskovalci" (ibid: 23). Obe točki za prvotne predpostavke strukturalizma pomenita hud davek, ki pomeni predvsem metodično in strogo razločevanje ter privilegiranje ene jezikovne ravnine (sintakse) na račun drugih (fonologija, semantika, pragmatika), drugi pa je dolgoročno gledano še usodnejši. Pomeni namreč, da smo izgubili temelj družbene, kolektivne konstruiranosti jezikovnega znaka. Ker pa bi bila predpostavka o nasebnem obstoju jezika izven človeškega uma absurdna, se je bilo treba vrniti k mentalističnemu izhodišču in predpostaviti, da pri vsakem govorcu jezika obstaja vrojeni govorni aparat, na katerega se lahko z gotovostjo zanesemo po tem, ko se je oblikoval skozi genetsko programirano in nereflektirano družbeno konstitucijo v procesu kognitivnega razvoja v otroštvu. Kot tak je poslej na voljo v obliki idealnega maternega govorca nekega jezika, ki zlahka razločuje med tem, kar postane temeljni cilj jezikoslovne raziskave: "Osnovna naloga lingvistične analize jezika L je ta, da razločimo med *gramatikalnimi* nizi, ki so stavki v L, od *negramatikalnih* nizov, ki niso stavki, in da razišče strukturo gramatikalnih nizov. Gramatika jezika L bo tako naprava, ki

¹⁵ Prvi stavek v Sintaktičnih strukturah se glasi: "Ta študija obdeluje sintaktično strukturo tako v širšem smislu (nasproti semantiki) kot v ožjem smislu (proti fonologiji in morfolgiji) in predstavlja del poskusa vzpostavljanja in raziskovanja osnove *formalne obče teorije jezikovne strukture*." (Čomski 1984: 15, prev. in poud. S. K.)

generira vse gramatikalne nize v L, ne generira pa negramatičnih nizov." (ibid: 21)

Pri tem zastavku je pomembno dvojje: abstrahiranje pomena iz stavčne strukture neizogibno potisne raziskovanje jezika v smer, ki bi mu lahko rekli matematizacija jezikovnih struktur in dejansko se je Chomsky naslonil na že nekaj časa prisotno vejo analitične filozofije, ki so jo zanimala vprašanja jezika s stališča logike. Chomsky v predgovoru v Sintaktične strukture tako kot vir inspiracije omenja Russillovega učenca V. V. Quinea (Čomski 1984, 16), kar pravzaprav pomeni, da so logično-matematične operacije torej prvi način, kako se semantični element posredno vrača v abstraktno sintaktično strukturo, vendar zgolj v svoji analitično filozofski formi.¹⁶ Hkrati se Chomsky vsaj na začetku svojega raziskovanja ni povsem odrekel možni bodoči "semantizaciji" sintaktičnih struktur:

Pojem 'strukturnega pomena' kot nasprotje 'leksikalnemu pomenu' je videti dokaj sumljiv in vprašanje je, ali se gramatični vzorci v jeziku uporabljajo dovolj konsistentno, da bi se jim lahko pripisovalo neposreden pomen. Povsem naravno je, da kljub temu najdemo precej pomembnih korelacij med sintaktično strukturo in pomenom, ali če rečemo drugače, odkrivamo, da se gramatični vzorci uporabljajo povsem sistematično. Te korelacije bi lahko sestavljale del predmeta precej bolj splošne teorije jezika, ki bi se ukvarjala s sintakso in semantiko in njunimi medsebojnimi povezavami. (ibid: 125, poud. S. K.)

Kasnejši razvoj teorije kaže, da je bila ta misel v radikalnejšem smislu pripuščanja pomenskih kategorij v abstraktni sintaktični model bolj ali manj

¹⁶ Analitična filozofija jezika je imela dolgo tradicijo dilem o tem, ali govorijo o notranjih jezikovnih razmerjih in ali imajo ta razmerja dejansko neko korelacijo – kakršnokoli že – z zunanjo realnostjo. Ponazorjeno s preprostim primerom bi lahko rekli, da iz trditve "vsi ljudje so umrljivi" in "Sokrat je človek" logično izhaja, da je Sokrat umrljiv. Pomembno je, da logična operacija dobi smisel/pomen le v korelaciji z zunanjo realnostjo – od tod potreba po "vseh možnih svetovih" v modalni logiki, pri Saussuru pa ta operacija na nek način postane nesmiselna. Ko privzamemo pogled na jezik, pri katerem je pomen določen izključno z notranjimi razmerji jezikovnih znakov, kakršnakoli gotovost na podlagi povezave z "zunanjim svetom" izgine.

opuščena, ta isti razvoj pa opredeljuje prav nenehen boj z izključenimi ravnmi jezika, za katere se zdi, da hočejo vdirati v čisto sintaktično strukturo. Tako je bilo po *Studies on Semantics in Generative Grammar* (Chomsky 1976) v *Lectures on Government and Binding* (Chomsky 1988) potrebno uvesti koncept jedrnih in perifernih pravil. Jedrna pravila predstavljajo globinske regularnosti jezika, ki jih določajo parametri, periferna pa so "izjeme", kjer ni najti notranjih regularnosti. Raziskovalni program torej mora še bolj omejiti svoj obseg:

Smiselno bi bilo, če se osredotočimo na jedrni sistem in se ne ukvarjamo s pojavi, ki so rezultat historičnih naključij, mešanja dialektov, osebnih idiosinkratičnosti in podobno. Kot pri vsaki drugi empirični raziskavi morajo v njen tek vstopati vprašanja, ki so notranja teoriji, in pričakujemo, da bodo potrebne nadaljnje distinkcije. (Chomsky 1996: 20)

Ločevanje med pomenom in matematizirano sintaktično strukturo – ki za razliko od fonetike in semantike izkazuje kombinacijsko kompleksnost – je ostala ena od dogem osrednjega toka "kembriške šole".¹⁷ Na ravni dogme poleg prepričanja o obstoju jezikovnega organa z vrojeno univerzalno slovnico ostajajo še derivacijska formalna tehnika ter obstoj globinskih oz. prikritih ravni sintakse. Šola je sicer izjemno razvejana in takšne generalizacije nedvomno delajo krivico množici raziskovalcev, ki so prevzeli program, a se v določenih aspektih oddaljili od osrednjega toka. Vendarle pa so navedene lastnosti na nek način razločujoče pri samodefiniranju šole, kdo je "znotraj" in kdo "zunaj". Precej številčna pa postaja skupina raznorodnih teorij, ki z izhodiščem v generativni teoriji skušajo jezik dojemati bolj holistično v primerjavi z osrednjim generativnim programom.¹⁸ Dilemo je najbolj strnjeno

¹⁷ To poimenovanje za jezikoslovni program s centrom na Massachusetts Institute of Technology uporablja J.-C. Milner v *Uvodu v znanost o govorici* (Milner 2001).

¹⁸ V to kategorijo bi lahko spadale denimo Head-Driven Phrase Structure Grammar (Pollard, Sag 1994), Lexical-Functional Syntax (Bresnan 2001), Construction Grammar (Fillmore 1995), Autolexical Syntax (Sadock 1991) in Role and Reference Grammar (Valin, LaPolla 1997).

ubesedil Chomskyjev učenec Ray Jackendoff v svoji *Simpler Syntax*, ki sicer ravno tako izhaja iz generativne tradicije:¹⁹

- ▶ Formalna tehnika je omejitvena (*constraint-based*).
- ▶ 'Skritih ravni' sintakse, sestavljenih iz posameznih sintaktičnih enot, ni.
- ▶ Kombinacijska kompleksnost se neodvisno pojavlja v fonologiji, sintaksi in semantiki.
- ▶ Slovnični pojavi se pojavljajo v kontinuumu od idiosinkratičnih (vključno s posameznimi besedami) do splošnih pravil slovnice.

Oz. bolj specifično na tezo o jedrnosti in periferiji:²⁰

- ▶ Med besedami in pravili ter med perifernimi in jedrnimi fenomeni obstaja kontinuum. /.../
- ▶ Idealizacija "jedra", čeprav je a priori smiselna, se je v praksi izkazala za sistematično zavajajočo.²¹

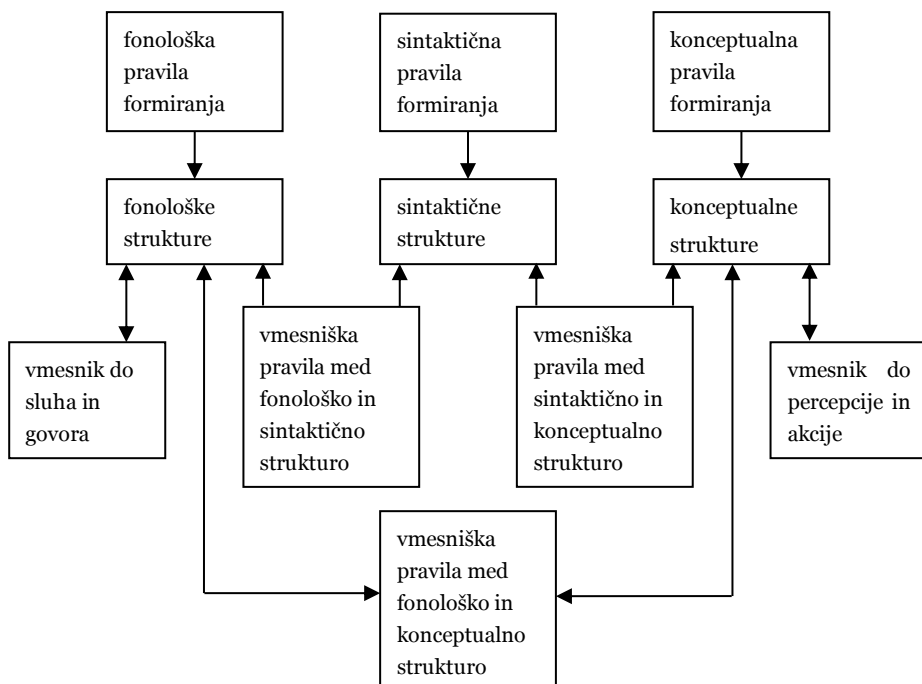
In še najpomembnejši povzetek: "Tradicionalno ločevanje med besediščem in slovnico je napačno" (ibidem). Ob vračanju na izhodiščno nepristajanje na ločevanje jezikovnih ravnin, značilno za prvo generacijo strukturalizma, je druga zelo pomembna sprememba vsebovana v malce kriptični opredelitvi formalne tehnike. Izhodišče "omejitvenosti" namreč pomeni, da smo zavrnili derivacijsko načelo, pri katerem postavljamo za izhodišče nuklearni zgled sintaktične strukture, ki ga potem z derivacijskimi pravili izpeljujemo do kompleksnejših struktur, temveč smo jezik nediskriminatorno glede na ravnine dojeli kot sistem omejitev, ki prav vse vplivajo druga na drugo (formalno je sistem predstavljen na Sliki 1). Sam formalizem predstavitve, s pomočjo katerega Jackendoff shematično ponazarja tudi druge podobne rešitve, ni tako pomemben, kot je pomembno dejstvo, da je pri tem poudarjen

¹⁹ Citirano po: Jackendoff, Culicover 2005: 15, prev. S. K.

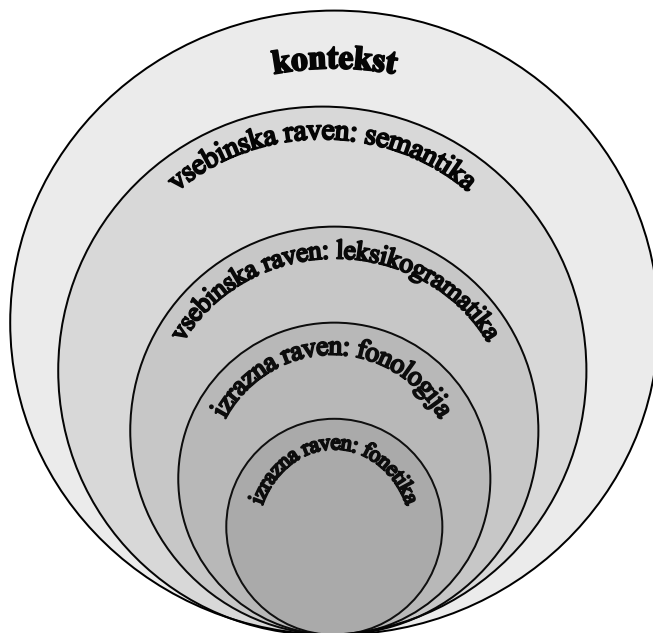
²⁰ Citirano po: ibid: 26, prev. S. K.

²¹ Jedra v smislu jedrnih sintaktičnih pojavov, ki so izpostavljeni in edini pripoznani na račun manj pogostih, obrobni ali idiosinkratičnih.

holističn(ejš)i pogled na jezik v neostrukturalnih formalnih okvirih, kar ta koncept približuje funkcionalističnemu slovniciarstvu, ki je vzporedno s kembriško šolo živelo svoje teoretsko življenje vse od firthejanskih poudarkov na kontekstualni naravi jezika in predvsem nikdar ni odstopilo od družbene konstituirnosti jezika oz. jezikovnega znaka. To stališče je lepo ubesedeno v stavku iz Hallidayeve *An Introduction to Functional Grammar*: "Jezik je serija redundanc, s katerimi povezujemo naše ekosocialno okolje z nenaključnimi motnjami v (zvočnem valovanju) zraka." (Halliday, Matthiessen 2004: 26, prev. S. K.)



Slika 1: Jackendoff – tripartitna paralelna struktura



Slika 2: Halliday - stratifikacija

Poleg družbene konstituiranosti jezika so neofirthejanski funkcionalisti, kot sta John McH. Sinclair in M. A. K. Halliday, ves čas vztrajali tudi pri obravnavi jezika oz. govornice kot celote ter pri apriorni in sistemski nedeljivosti jezikovnih ravnin:

Značilnost pristopa do sistemske teorije, ki ga uvajamo tukaj, je ta, da je *vseobsegajoč*: obravnava jezik v njegovi celovitosti, tako da je vse, kar rečemo o enem vidiku, treba vedno razumeti v navezavi s celotno sliko. /.../ Naš tradicionalni kompozicijski način razmišljanja o jeziku je treba, če že ne nadomestiti, pa vsaj dopolniti s "sistemskim" razmišljanjem, pri katerem skušamo razumeti naravo in dinamiko semiotičnega sistema kot celote. (ibid: 20)

V tem odstavku takorekoč odzvanja Saussurova ideja o celostni teoriji jezika kot le še enem od semiotičnih sistemov in Halliday si dejansko postavlja

natanko tako nalogo. Pri tem je potrebno dodati še *caveat*, ki ravno tako izhaja iz saussurovskega razmišljanja. Če je jezik v sebi zaključen sistem znakov, lahko na ta način opišemo le vsak jezik/sistem posebej, in na neki povsem drugi ravni iščemo primerjave med različnimi sistemi. Tako kot bi lahko na primer opisali le pravila šahovske igre in jih šele nato primerjali s pravili drugih namiznih iger. Halliday glede tega ne okoliši: "Naš namen v tej knjigi je bil opisati in pojasniti vire proizvodjanja pomena v sodobni angleščini" (ibid: 4). Semiotika funkcijskega opisa torej nima aspiracij po opisu "univerzalne slovnice", temveč se v tem konkretnem primeru omejuje na en semiotični sistem, sodobno angleščino. Medtem ko pri obravnavi navaja različne stratifikacijske nivoje te nedeljive celote, prej razdruženo leksiko in sintakso združi pod en sam nivo obravnave, poimenovan leksikogramatika, kot je prikazano na Sliki 2. Znotraj semiotike angleškega jezika je torej kakršnokoli ločevanje med leksiko in sintakso dojeta kot lažno: "slovnica in besedišče nista dve ločeni komponenti jezika – sta le dva konca enega kontinuuma" (ibid: 7) in ta enotni pojav je po svoji naravi v temelju probabilističen:

Na sistem je torej treba gledati kot na nekaj, kar je po svoji naravi probabilistično. Tovrstna privzeta drža je za opazovalce leksike značilna že dober del stoletja; /.../ Opazovalcem slovnice se zdi takšna drža pogosto neokusna, ko da bi pripisovanje probabilitet v slovnici ogrožalo svobodo posameznika. /.../ Glavni teoretični problem paradigmatičnih slovnice, kakršna je "sistemska" slovnica, /.../ je vprašanje, kako probabiliteto vgraditi v vedo o slovnici. (Halliday 2005: 77-78, prev. S. K.)

Pripisati probabiliteto slovničnim pojavom ne pomeni, da podatek o pogostosti npr. trdilne oblike proti vprašalni pomeni le znak, da je nek pojav lahko motiviran ali nemotiviran, s tem merimo na to, da ima skladnja kot drugi pol istega pojava enako lastnost kot besedišče, ki nima nepremičnih in trdnih stalnih vrednosti – pomenski potenciali so vedno izkoristljivi in udejanjeni šele v konkretnem kontekstu in čim več konkretnih ubeseditev zajamemo v analizo, z večjo natančnostjo lahko opišemo sistem.

4 KORPUSNO JEZIKOSLOVJE

S tem pa smo pravzaprav prišli do možne interpretacije teoretskih izhodišč t. i. korpusnega jezikoslovja. Korpus sam na sebi je seveda le ogromna zbirka ostankov govornih dejanj, ki brez jezikoslovne interpretacije ostaja na ravni zbirke metuljev ali podatkov o vremenu, kar je pogosta šala mentalističnega pristopa. Da bi raziskovanje te zbirke dobilo teoretsko osnovo, je ključna odločitev, kako resno vzeti vsako jezikovno dejanje kot pripadajoče in izhajajoče iz nekega sistema jezikovnih znakov, ki je namenjeno (pomenski) interpretaciji in ki tudi ima interpreta. Če lahko katerokoli realno komunikacijsko/jezikovno pojavitev v korpusu zavržemo kot neinterpretabilno s systemskega stališča in s tem opredeljeno kot ne-jezik, neskladnja itd.,²² smo se s tem a priori odpovedali možnosti teoretske refleksije jezika oz. govornice s saussurjevskim pomenu (tj. *langage*). Če pa pristanemo na holistični pristop, jezika ne moremo več dojemati kot inventar formalnih struktur, temveč kot vir ustvarjanja pomena, ki v končni fazi izhaja iz sistemskih vzorcev, ki jih opredeljuje izbirnost. Ti sistemski vzorci niso opredeljeni le kot izključujoča možnost ali nemožnost, temveč so interpretabilni na lestvici od posamezne instanciacije do takorekoč neskončnih ponovitev. Sistem izbir pa je družbeno pogojen le kot potencial, ki ga posamezni govorec lahko izkoristi za svoj – v osnovi komunikacijski – namen. Na ta način ni ovir za preverjanje takorekoč kakršnekoli hipoteze jezikoslovne narave na empiričnem materialu – korpusu.

Pri empiričnem raziskovanju množice besedil pa ponovno z drugimi sredstvi pridemo do matematizacije jezika, ki za jezikoslovca lahko predstavlja druge vrste past. Ko smo soočeni z ogromno množico empiričnih podatkov, za katere vemo, da predstavljajo vzorec nekega sistema, se lahko zazdi, da bomo do teoretskih zaključkov lahko prišli z apliciranjem kompleksnih statističnih

²² "Vsak naravni korpus je nujno izkrivljen. Nekateri stavki se ne bodo pojavili, ker so povsem očitni, drugi, ker so nepravilni, spet nekateri, ker so nevljudni. Korpus, če naj bo naraven, bo tako hudo izkrivljen, da opis ne bo nič drugega kot zgolj seznam" (Chomsky 1962: 159, citirano po Leech 2005: 29-30).

metod, že uspešno uporabljenih na množicah drugih empiričnih podatkov z notranjo strukturiranostjo. Še več, uporaba metod iz trdih znanosti ponovno zbujata vtis ali daje upanje, da iz tega jezikoslovje črpa svojo novo znanstveno upravičenost in *raison d'être*.²³ Dejstvo je, da je t. i. računalniško jezikoslovje od začetka obdelav velikega števila jezikovnih podatkov omogočilo uvid v številne pojave, ki bi sicer ostali neopaženi ali neustrezno interpretirani, vendar je statistika brez jezikoslovne interpretacije v temelju slepa in indiferentna do razlikovanja med, če damo dovolj preprost primer, t. i. stalno in prosto zvezo. Čim uporabimo jezikoslovni metajezik, naj bo to na povsem osnovnem nivoju, kot je recimo "beseda", "črka", "stavek" ali malce višjem: "morfem", "kolokacija", "frazem", "glagolska fraza", "predložna zveza" itd., smo na slepi statistični podatek aplicirali jezikoslovno teorijo in pristajanje na jezikoslovni zdravi razum pomeni le zatiskanje oči pred nevarnostjo, ki jo izpostavlja Milner:

Med lingvisti in tehniki jezika lahko opazimo celo bolj poudarjeno vzajemno ravnodušnost in nepoznavanje, kot je v navadi na drugih področjih. /.../ Ugotoviti je treba le še, kdo odkriva zakone jezikov. Nihče se ne obotavlja, če gre za kemijo, za biologijo ali za atomsko fiziko. Pri jezikih pa so bili vse do sedaj tehniki očitno prepričani, da se lahko obrnejo na kogarkoli. (Milner 2001: 33)

Korpusno jezikoslovje torej ni jezikoslovna teorija na sebi, je metodologija oz. zbirka tehnik, ki se močno opirajo na statistične metode in jih ima jezikoslovec poljubnega teoretskega prepričanja na voljo, da svoje teze preverja na empiričnem jezikovnem gradivu.

LITERATURA

Benveniste, E. (1988): *Problemi splošne lingvistike*. Ljubljana: Znanstveni

²³ Po A. Koyréju je to diskurzivna konfiguracija, ki se je izoblikovala z Galileom in jo označuje kombinacija dveh potez: (I) matematizacija empiričnega; (II) vzpostavitev razmerja s tehniko, tako da je tehnika definirana kot praktična uporaba znanosti (od tod naziv aplikativne znanosti), znanost pa definirana kot teorija tehnike (od tod naziv temeljne znanosti) (Milner 2001: 21).

inštitut Filozofske fakultete.

Bloomfield, L. (1933): *Language*. London: Allen and Unwin.

Bresnan, J. (2001): *Lexical-functional syntax*. Oxford: Blackwell.

Chomsky, N. (1962): A transformational approach to syntax. *Third Texas Conference on Problems of Linguistic Analysis in English*. Austin: University of Texas.

Chomsky, N. (1976): *Studies on semantics in generative grammar*. The Hague, Paris: Mouton.

Chomsky, N. (1988): *Lectures on government and binding*. Dordrecht: Foris Publications.

Chomsky, N. (1996): *The minimalist program*. Cambridge: MIT Press.

Čomski, N. (1984): *Sintaksičke strukture*. Novi Sad: Knjižena zajednica Novog Sada.

De Saussure, F. (1997): *Predavanja iz splošnega jezikoslovja*. Ljubljana: ISH Fakulteta za podiplomski humanistični študij.

Fillmore, C. J. (1995): *Construction Grammar*. Chicago: Chicago University Press.

Firth, J. R. (1964): *The Tongues of Men and Speech*. London: Oxford University Press.

Halliday, M. A. K., Matthiessen, C. M. I. M. (2004): *An introduction to functional grammar*. London: Arnold.

Jackendoff, R., Culicover, P. (2005): *Simpler Syntax*. Oxford: Oxford University Press.

Kuhn, T. (1998): *Struktura znanstvenih revolucij*. Ljubljana: Krtina.

Leech, G. (1981): *Semantics*. London: Penguin Books.

Milner, J.-C. (2001): *Uvod v znanost o govorici*. Ljubljana: Krtina.

Pollard, C., Sag, I. (1994): *Head-driven phrase structure grammar*. Stanford: The University of Chicago Press.

Sadock, J. M. (1991): *Autolexical syntax: a theory of parallel grammatical representations*. Chicago: University of Chicago Press.

Škiljan, D. (1997): *Saussurjev ustvarjalni molk*. V: *Predavanja iz splošnega jezikoslovja*: 261–280. Ljubljana: ISH Fakulteta za podiplomski humanistični študij.

Valin, R. D. Van, LaPolla, R. J. (1997): *Syntax: structure, meaning and function*. Cambridge: Cambridge University Press.

CORPUS METHODS AND THEIR REFLECTION IN LINGUISTIC THEORIES OF THE 20TH CENTURY

In the 20th century structuralism established itself as the central linguistic theory, in the first half mainly through its originator Ferdinand de Saussure, and in the second half with the figure of Noam Chomsky. The latter consistently refused to acknowledge analysis of extensive quantity of texts as a valuable method, and favoured linguistic intuition of a native speaker instead. In parallel with structuralism other trends in linguistics emerged which pointed to the inadequateness of the prevailing linguistic paradigm and to theoretical insights which were only possible after the systematic analysis of large quantities of texts. The paper discusses some of the dilemmas stemming from this dichotomy and places corpus linguistics in a broader linguistic context.

Keywords: corpora, linguistics, structuralism, semantics, syntax, Saussure, Chomsky

To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-
Deljenje pod enakimi pogoji 2.5 Slovenija.

This work is licensed under the Creative Commons Attribution ShareAlike 2.5
License Slovenia.

<http://creativecommons.org/licenses/by-sa/2.5/si/>

