

KORPUSI IN JEZIKOSLOVJE

S prihodom korpusov je v jezikoslovju na voljo več kakovostnih podatkov o jezikovni realnosti kot kadarkoli prej. Z razvojem metodologij gradnje in analize korpusov pa se je razvila tudi metodologija, ki omogoča temeljit vpogled v jezikovno delovanje in učinkovite jezikovne opise. Ob uporabniško usmerjenih gradivnih opisih jezika in njegovega delovanja pa se oblikujejo tudi novi strokovni diskurzi, ki se sopostavljajo tradicionalnim; v današnji družbi znanja se namreč načini ustvarjanja in prenašanja znanja ter intelektualnih veščin v njej oblikujejo v demokratičnem dialogu med ustvarjalci vedenja in tistimi, ki to vedenje potrebujejo.

1 Jezikovna realnost in jezikoslovje

Za izhodišče našega razmisleka o korpusih in jezikoslovju nam bo služila teza G. Sampsona, plenarnega predavatelja na letošnji konferenci *Corpus linguistics* (Lancaster, 28.–31. 3. 2003). Spodbujen s predhodnimi študijami, ki so pokazale, da se je računalniško jezikoslovje v devetdesetih letih popolnoma usmerilo v korpusne podatke, je naredil raziskavo o razmerju med empirično in intuitivno zasnovanimi jezikoslovnimi študijami. Analiziral je razprave v reviji *Language* v zadnjih petdesetih letih in ugotovil, da se delež empiričnih študij ni bistveno povečal, še več, v drugi polovici devetdesetih let prejšnjega stoletja je njihov delež glede na sredino devetdesetih celo upadel; na nek način se je ponovila slika sredine šestdesetih in začetka sedemdesetih, ko je glede na petdeseta bistveno upadel delež empiričnih študij in porasel delež intuitivno zasnovanih.

1.1 Prvi prelom: dva vzporedna diskurza

V zgodnjih petdesetih in v začetku šestdesetih so v reviji *Language* prevladovala izrazito empirično zasnovane fonetične in morfološke študije, z vstopom generativne slovnice pa so v šestdesetih močno porastle intuitivno zasnovane.

Čeprav se s časovne distance prelom v generativno slovnico ne zdi tako usoden, saj

je tudi drugemu delu jezikoslovja bilo omogočeno razvijanje svojih jezikoslovnih konceptov, pa se je diskusija med predstavniki jezikoslovja, ki temelji na delu z obsežnim načrtno zbranim gradivom, in predstavniki generativne slovnice zaostрила v šestdesetih letih 20. stoletja do te mere, da sta področji živeli ločeno in oblikovali tudi dva popolnoma ločena strokovna diskurza. Šola okrog Chomskega je zavračala veliko količino jezikovnih podatkov za temelj jezikoslovnih raziskav (Malmkjær 1996: 74), prvi poskusi gradnje korpusa pa so bili označeni kot »popolna izguba časa«, saj naj bi govorec določenega jezika v desetih minutah produciral za jezikoslovce več ilustrativnega gradiva, kot ga je na voljo v večmilijonskih besedilnih korpusih (Biber in Finegan 1991: 204). Jezikoslovje, ki je utemeljevalo principe korpusnega jezikoslovja v šestdesetih letih, se je tako v veliki meri organiziralo prav kot jezikoslovni pristop v opoziciji do jezikoslovnega dela Chomskega ne le glede gradiva kot osnove jezikovega opisa, ampak tudi v razmerju do uporabne vrednosti raziskovalnih rezultatov – korpusni opisi so se od samega začetka usmerjali v področje uporabnosti in se potrjevali v stiku z uporabniki jezikovnega znanja in ne le znotraj ozko zaprtega jezikoslovnega kroga.

Vprašanje absolutne zadostnosti korpusa ali absolutne zadostnosti intuicije, ki jo predvideva Chomsky, je s časovne distance gledano umetno ustvarjena dilema. Delo s korpusom res temelji na veliki količini zbranih besedil, ki so podlaga za jezikovno analizo, a se ne odreka intuiciji. Šestdeseta leta tudi v slovenskem prostoru izpostavijo jezikovne opise, temelječe na načrtno zbranem gradivu, in v slovarskih priročnikih zavrnejo možnost opisa jezikovnih elementov, ki nimajo podlage v jezikovni realnosti:

Slovenci smo navajeni, morda bolj kakor drugi narodi, da zaradi narodnostne ogroženosti zelo pazimo, da se v knjižni jezik ne vnaša preveč tujega, oz. tega, česar ne izkazuje literarna tradicija. Zdaj bo v slovarju registriranega mnogo več: to, kar je bilo priznано kot dobro, manj dobro in tudi to, kar je veljalo za slabo. Hoteli smo prikazati knjižni jezik v najširšem pomenu besede: živ, poln, z dubletami, notranjimi nasprotji, vzporednimi istočasnimi normami, jezik sredi zagona in razvoja. /.../ Slovar bo registriral dejansko stanje v jeziku, torej osnove njegove norme, s kvalifikatorji in kvalifikatorskimi pojasnili pa bodo vstavljene v ta okvir posebnosti, dvojnosti in izjeme (Suhadolnik 1968: 4–5).

Jezikovni podatki v korpusu ne morejo biti smiselno interpretirani brez interpretativnih zmožnosti raziskovalcev. Pri uporabi korpusnih podatkov je dragocena tako intuicija govorcev določenega jezika kot jezikoslovcev. Vendar je pri informatorjih možna veliko večja manipulacija, povezana s (samo)spraševanjem, »Ali lahko tako rečem?« in odgovori »Da, lahko to rečem.« in v nadaljevanju »Ampak tega nikoli ne bi rekel.« (McEnery in Wilson 1996: 12). Kritika Chomskega, da korpusni podatki niso jezikoslovno relevantni, saj se nekateri stavki ne bodo pojavili, ker »so povsem očitni, drugi, ker so nepravilni, spet nekateri, ker so nevljudni« (Kennedy 1998: 23), kažejo na dva v osnovi res drugačna pristopa v jezikoslovju, povezana v veliki meri z normativnim vrednotenjem jezika, pa tudi vprašanjem stilistične vrednosti besedil oz. dobrega/slabe jezika. Seveda pa vprašanje o kakovosti jezikovnih podatkov ostaja eno ključnih vprašanj korpusnega jezikoslovja:

Vprašanje referenčnosti /korpusa/ je v svoji globalni razsežnosti pomembno jezikovno-filozofsko vprašanje, povezano z začetkom in razvojem sodobnega korpusnega jezikos-

lovja. Kakšna količina izpričanih besedil in stavkov je potrebna, da lahko ustrezno ponazarja jezikovno zmožnost chomskyjevskega idealnega rojenega govorca? Zlasti ob upoštevanju enega od osnovnih teoremov tvorbeno-pretvorbne slovnice, da jezikovna oziroma slovnicična zmožnost, ki jo sestavlja omejeno število pravil različnega reda, govorcem omogoča tvorbo in razumevanje neskončnega števila stavkov oziroma izjav? Popolnoma jasno je, da noben korpus ne more vsebovati neskončnega števila slovnicično pravih (in komunikacijsko ustreznih) stavkov. Vprašanje je le, katero končno število lahko vzamemo za metonimijo neskončnega in kako se raziskovalci pri uporabi korpusa zavedajo, da ima ne glede na svojo končno velikost vsak korpus poleg vseh prednosti, ki na marsikaterem jezikoslovnem področju izpričano omogočajo ustrežnejše in učinkovitejše raziskovanje in uporabnost, na koncu vendarle status metonimičnosti (Stabej 1998).

Prav zanašanje generativne slovnice na intuicijo rojenega govorca je spodbudilo razvoj korpusnega jezikoslovja, v veliki večini vezanega na angleški jezik, v okoljih, kjer je imela anglistika močno tradicijo, a govorcem ni bila prvi jezik, npr. Belgija, Nizozemska, Norveška, Švedska. Korpusno gradivo je v teh okoljih pomenilo pravi razcvet anglistike in v veliki meri pripomoglo k še večji metodološki raznolikosti angleškega jezikoslovja.

Zanašanje Chomskega na intuicijo rojenih govorcev je bilo izzvano predvsem z novimi smermi jezikoslovja, ki so se začele dinamično pojavljati v petdesetih letih 20. stoletja, predvsem sociolingvistike, psiholingvistike, pragmatike in analize diskurza. Razvoj uporabnega jezikoslovja je zahteval podatke o jezikovni rabi, in sicer tako naravnih govorcev nekega jezika kot prvega kakor tudi kot drugega ali tujega jezika. Posebej je ta segment podatkov o jezikovni rabi relevanten za študije učenja jezika in raziskave nekega jezika kot tujega jezika. V angleškem okolju je bilo to področje tudi zaradi razvoja v pomembno industrijsko panogo, ki je spodbujala in financirala tovrstne raziskave, že zgodaj izjemno razvit samostojni segment uporabnega jezikoslovja.

Zanašanje na intuicijo je v generativni slovnici privedlo tudi do definiranja idealnega govorca. Predvsem sociolingvistične raziskave in utemeljitev funkcijskega pristopa v jezikoslovju so pokazale na nevzdržnost takega izhodišča. Takoj ko je bil v izhodišče postavljen vidik nemonolitnosti jezikovne pojavnosti, je idealni naravni govorec nezadosten, saj njegovo jezikovno vedenje ne more zajeti celotne jezikovne pojavnosti in njene variabilnosti. Danes predvsem referenčni korpusi skušajo zajeti čim večjo paleto jezikovne pojavnosti, seveda z jasnim zavedanjem, da je zajetje jezikovne različnosti v celoti nemogoče (Čermák 2002: 269). Ali, kot je bilo tudi v našem prostoru že opozorjeno:

/O/staja dejstvo, da tudi referenčni korpus ne more zajeti vseh jezikovnih možnosti, kar bi lahko zajeli v naslednjo formulacijo: v jeziku je lahko tudi tisto, česar v korpusu ni. K temu sodi tudi zrcalna trditev: vse, kar je v korpusu, je tudi (bilo) v jeziku (Stabej 1998).

Analize korpusov angleščine so pokazale na vrsto jezikovnih rab, ki bi jih z vidika jezikovne intuicije označili kot nemogoče, z normativnega vidika pa označili kot napake. Tovrstne analize so opozorile na včasih prevelike posplošitve v skladu z željo po sistematizaciji. V šestdesetih letih se je tudi v slovenskem prostoru skušalo načrtno presegati normativistični pogled na jezik, ki ni imel opore v realnih jezikovnih podatkih; gre za še vedno aktualne ideje, ki tudi danes živo zadevajo slovensko jezikoslovno realnost:

Pri nas smo bili vajeni razlikovati besede zelo pavšalno po tem, ali so pravilne ali nepravilne, in smo imeli za to ocenjevanje v večini primerov neprimerne, navadno nejezikovne kriterije. Zgodilo se je, da je bila posamezna beseda ali zveza že dolgo v splošni rabi, pa je ta ali oni odkril, da gre za kalk ali vsaj paralelo v tujem jeziku, in že je bila obsojena kot germanizem ali romanizem ali izposojenka. Strah pred resničnimi in navideznimi napakami je bil tako velik, da tudi v primeru, če je bila beseda z mnogimi nespornimi dokazi rehabilitirana, nikoli več ni bila nevtralna. Zato pri nas pravzaprav ni bilo človeka, ki bi o sebi lahko rekel, da obvlada slovenski knjižni jezik. In izrazne možnosti, ki so bile v omenjenih razmerah že tako majhne, so se še zmanjševale. /.../ Slovar bo o vsaki besedi, o vsakem pomenu in obliki povedal, kdaj, kje in kako jo je mogoče uporabljati, da bo zvenela nevtralno, ali pa bo nakazal smer do nevtralnega knjižnega izraza. /.../ Vse to bo, upamo, pripomoglo k prenehanju preganjanja izoliranih jezikovnih napak in utrdilo zavest o normalnosti govorjenega oz. pisanega jezika povprečnega izobraženca, istočasno pa poglobilo resnično, široko in poglobljeno zanimanje za slovensko besedo v celoti (Suhadolnik 1968: 5, 6).

Nenazadnje je navezava procesiranja naravnih jezikov kot potencialnega prihodnjega sistema analize in sinteze besedil naravnih jezikov na realne korpusne podatke nujna, zato da bi v končni fazi tovrstne aplikacije res lahko učinkovito delovale. Podporo jezikovnotehnološkim rešitvam lahko nudijo le korpusni podatki. Jezikovni opisi, ki so za jezikovne tehnologije funkcionalni, namreč vsebujejo veliko več podatkov o besedilnem okolju, ki jih omogočajo le obsežni korpusi (Taubert 1995: 109–110).

1.2 Drugi prelom: dva vzporedna diskurza nekoliko drugače

Če je bilo v šestdesetih zmanjšanje empiričnih raziskav povezano s pojavom izredno močne in vplivne jezikoslovne šole, pa lahko spremembo v drugi polovici devetdesetih pripisujemo predvsem odporu tradicionalnega jezikoslovja do spremembe znanstvenega diskurza in diskurza sodobne družbe sploh; diskurz je v informacijski družbi postal bistveno bolj demokratičen, odprt za različnost metodoloških pristopov, z močno interdisciplinarnostjo strok pa tudi na prepihu ne le ozko specializirane strokovne javnosti, ampak javnosti sploh. Ob tako spremenjenem znanstvenem diskurzu pa je v tradicionalnih diskurzih prišlo do še večjega odpora do novih pogledov in do zapiranja vase.

Pri omenjenih procesih gre za značilnost sodobne družbe, načine ustvarjanja in prenašanja znanja ter intelektualnih veščin v njej. Na ravni visokošolskega prenosa znanja lahko opazujemo paralelne procese kot pri raziskovalni dejavnosti; strnjeno gre za sledeče:

Tradicionalni diskurzi v visokem šolstvu so – čisto samoumevno – izhajali iz **perspektive tistega, ki vrednost ustvarja in posreduje**, poenostavljeno rečeno: iz perspektive profesorja. Tej »Arhimedovi točki« akademskega znanja so bile podrejene vse študijske strukture. /N/astajajoči novi diskurzi preobračajo prav to točko in jo na novo utemeljujejo na **perspektivi tistega, ki vednost potrebuje in išče**. / (Zgaga 2003: 2).

Tako je v diskusiji na konferenci *Corpus linguistics* velik del publike na plenarnem zasedanju zavrnil relevantnost revije *Language* kot indikatorja stanja na področju

celotnega jezikoslovja. Opozorjeno je bilo, da gre za revijo, ki s svojo uredniško politiko goji tradicionalni jezikoslovni diskurz, prav uporabno jezikoslovje, ki je korpus v svoj metodološki aparat že celovito vgradilo, pa ga že dolgo načrtno pre-sega, hkrati pa je s svojo »uporabnostjo« nenehno izpostavljeno presoji uporabni-kov jezikovnega znanja. Ob tem in prav zato si ne dovoljuje sodbe z apriorno vzvišene pozicije enega pogleda, ki svoj tip znanstvenega diskurza ocenjuje kot edino relevantnega, relevantnost dosega znotraj ozkega strokovnega kroga, vse osta-le pa hierarhično vrednoti ali jih celo izključuje s področja »pravega« znanstvene-ga diskurza.

2 Korpusi v jezikoslovju in korpusno jezikoslovje

Delo v zvezi s korpusi je danes povezano z različnimi aktivnostmi; oblikovanje kakovostnih korpusov pa zahteva učinkovito sodelovanje jezikoslovcev z razisko-valci s področja družboslovnih ved in računalništva, predvsem tistega dela, ki se ukvarja z naravnimi jeziki. Popolnoma odprta diskusija enakovrednih partnerjev lahko pripelje do visokih sinergijskih učinkov, ki jih pri interdisciplinarnem delu pričakujemo.

2.1 Gradnja korpusov združuje tako jezikoslovce kot računalniške strokovnjake, saj je delo na področju oblikovanja korpusa vezano tako na načela njihove gradnje kot na zagotavljanje ustreznega elektronskega zapisa, ki v nadaljevanju omogoča ana-lizo jezikovnih podatkov (Kennedy 1998: 9). Pri načelih gradnje korpusov je naj-prej potrebno pripraviti okvirni načrt gradnje, ki zajema serijo premislekov in odlo-čitev. V osnovi bi jih lahko strnili v naslednje sklope (Atkins et al. 1992: 2):

- specifikacija korpusa in njegova oblika,
- strojna in programska oprema,
- zajem besedil in označevanje korpusnih dokumentov,
- procesiranje zbranega gradiva,
- končna oblikovanost korpusa in povratne informacije v zvezi z njim.

V največji meri so jezikoslovni premisleki v zvezi s korpusom vezani na prvo ali-nejo, torej specifikacijo korpusa in njegovo obliko. Temeljni premislek je vezan na tip korpusa, ki ga želimo graditi; ta za seboj potegne odločitve v zvezi z jezikom besedil (enojezični ali večjezični korpus), s časovnim zajemanjem besedil (sinhro-ni ali diahroni korpus), premislek o zajemu besedil glede na prenosnik (pisni ali govorni korpus) itd.

Izhodiščni jezikoslovni premislek pri osrednjem tipu korpusa, tj. referenčnem kor-pusu, ki želi predstaviti določen jezik v čim širšem obsegu njegove pojavnosti, pa je vezan tudi na določitev parametrov za uravnoteženost v korpusu zajetih besedil na eni strani (Biber 1993: 243) ter njihovo jezikoslovno označenostjo v korpusu na drugi (Atkins et al. 1992: 7–8). Prav to so parametri, ki dvigajo kakovost jezikov-nim podatkom v korpusu, saj je razumljivo, da korpus, ki zajema samo leposlovna

besedila ali samo besedila enega časopisa, ne more biti kakovosten vir za npr. referenčne jezikovne priročnike.

Da bi z gradnjo sploh lahko začeli, je potrebna tehnična podpora, ki mora od samega začetka slediti zahtevam tako glede strojne kot programske opreme ter biti sposobna oblikovati orodja za procesiranje zbranega gradiva. Prav pri procesiranju podatkov se je potrebno odločati tako, da jezikovnim podatkom zagotovimo čim večjo uporabnost, izmenjavo ter trajnost, kar v zadnjem času omogočajo standardi za prenos in zapis jezikovnih podatkov.

Čeprav se razmislek v zvezi s postopki zajemanja besedil zdi dokaj trivialen, pa so se korpusi prav na tem nivoju velikokrat znašli pred nerešljivo težavo: kako sploh organizirati zbiranje besedil ter prepričati besedilodajalce, da za namene korpusa svoja besedila odstopijo. Prav zaradi nepredvideno zapletenih postopkov se je pri mnogih korpusih njihova gradnja precej zavlekla, tako da se danes vsi zavedajo zahtevnosti in zamudnosti zbiranja besedil (Atkins et al. 1992: 3).

S pridobivanjem besedil je povezano še eno temeljno vprašanje, ki ga mora vsak resno zastavljen korpusni projekt rešiti pred začetkom gradnje, tj. zagotavljanje varovanja avtorskih pravic. Potrebno je poznavanje področja varovanja avtorskih pravic, in sicer tako na mednarodni kot državni ravni, ter v skladu s tem oblikovanje ustreznih rešitev (Atkins et al. 1992: 4). Prav izkušnje pri starejših korpusih, ki vprašanja avtorskih pravic niso zadovoljivo rešile, tako da danes tovrstnih podatkov sploh ni mogoče uporabljati, so oblikovalce kasnejših korpusov prisilile v razmislek ter iskanje ustreznih rešitev.

Pri končni obliki korpusa je z vseh vidikov smiselno spremljati odzive na rešitve, jih sistematično obdelati ter razmisleke v zvezi z gradnjo revidirati ter tako pri njegovi nadgradnji dosegati večjo kakovost ter prijaznost do uporabnikov.

Čeprav se zdi samoumevno, pa je mogoče vendarle potrebno izpostaviti potrebo po gradnji korpusa tako, da lahko sproti sledimo jezikovnim spremembam, torej zasnovati delo na način, da lahko novo besedilno gradivo v korpus nenehno vključujemo. Šele korpusni podatki nam lahko dajo pravo sliko o jezikovnem razvoju in težnjah jezikovnega razvoja; podatki, zbrani po npr. listkovni metodi za starejše slovarje, za tovrstne raziskave zaradi omejenosti informacij niso primerni (Čermák 2002: 268). Ob tem seveda ne gre za metodološka vprašanja, ampak največkrat za vprašanja financiranja projektov, ki imajo za cilj vedno samo nek vmesni cilj brez pravega končnega cilja.

2.2 Razvoj orodij za delo s korpusom je vezan predvsem na področje računalniškega jezikoslovja (Kennedy 1998: 9), danes pa nekateri široko dostopni programi omogočajo tudi računalniško manj izobraženim uporabnikom korpusa samostojen razvoj manj kompliciranih programskih orodij, ki si jih za svoje specifične potrebe lahko oblikujemo sami; ob tem se vzporedno razvijajo tudi postopki in metode korpusne analize.

2.3 Ob delu s korpusi se razvijajo druga področja jezikovnih tehnologij, ki za svoje izhodišče potrebujejo kvalitetne jezikovne vire; sem sodi razvoj črkovalnikov, različnih slovničnih pregledovalnikov, elektronskih slovarjev in tezavrov pa tudi sinteze in analize govora ipd.

2.4 V jezikoslovju sodijo na področje korpusnega jezikoslovja raziskave za potrebe opisnega jezikoslovja, kamor tradicionalno sodijo različne vrste slovničnih in leksikalnih analiz (Kennedy 1998: 9) – predvsem leksikologija in leksikografija sta področji, ki že po tradiciji izhajata iz sistematično zbranega gradiva (McEnery in Wilson 1996: 90); prav slovarji, narejeni na podlagi korpusov, pa so tudi prvi res široki javnosti dostopni rezultati jezikoslovne uporabe korpusov.¹ S pojavitvijo večjega števila širši javnosti dostopnih različnih tipov korpusov pa se njihova uporabnost širi predvsem na tradicionalno gradivno usmerjena jezikoslovna področja, kot so besediloslovje, prevodoslovje, sociolingvistika, stilistika ipd. (McEnery in Wilson 1996: 98–101, 111).

2.5 Korpusno jezikoslovje trenutno pomeni zbirni pojem za niz različnih aktivnosti v zvezi s korpusi, pri tem pa je zredišče analize velika količina načrtno zbranega avtentičnega gradiva in empirična analiza dejanskih vzorcev jezikovne rabe (Biber et. al. 1998: 5); šele pojav računalnikov in oblikovanje metod zbiranja ter gradnje korpusov je omogočilo pridobiti veliko količino relevantnih aktualnih jezikovnih podatkov za jezikoslovne študije; sodobni jezikovni opisi tako lahko temeljijo na empirični analizi zares velike količine avtentičnih besedil (Biber et. al. 1998: 9–10). Gre za značilnosti jezikovnih podatkov, ki jih starejšim zbirkam jezikovnih podatkov ne moremo pripisati (Čermák 2002: 265).

Uporaba računalnika z avtomatičnimi in interaktivnimi tehnikami omogoča analizo velike količine jezikovnih podatkov. Z razvojem tehnik korpusne analize količina jezikovnih podatkov ne predstavlja ovire, pač pa prav obratno, z večjo količino podatkov lahko pridemo do novih informacij; za določene postopke korpusne analize pa so sploh primerni samo izjemno veliki korpusi. Z uporabo tako kvantitativnih kot kvalitativnih analiz se lahko gradijo novi jezikovni opisi; kvantitativne analize se pojavljajo v izhodišču vsakega korpusnega pristopa, pomenijo relevanten jezikovni podatek, ki pa ga je seveda potrebno še interpretirati in razložiti (Biber et. al. 1998: 5, 8–9; Gorjanc 2002: 90–106).

Ko govorimo o korpusnem pristopu v jezikoslovju, srečamo pojma delni in popolni korpusni pristop (Tognini-Bonelli 2001). Popolni korpusni pristop sugerira uporabo korpusa neodvisno od uveljavljenih jezikoslovnih interpretacij, kar v bistvu pomeni, da bi bil mogoč le s »surovim« besedilnim gradivom brez kakršnihkoli jezikoslovnih oznak. Šele opazovanje »surovih« besedil in njihova analiza brez sklicevanja na obstoječe jezikoslovne modele bi v lahko oblikovalo zares neodvisni in neobremenjeni korpusni pristop. Prav zaradi zavedanja obremenjenosti analiz z obstoječimi jezikoslovnimi koncepti se danes raje govori o korpusu kot viru za preverjanje hipotez na eni in viru za gradnjo hipotez na drugi strani.

Ob korpusu kot viru za gradnjo hipotez o jeziku se oblikuje jedro korpusnega jezikoslovja, korpus je pri tem v izhodišču jezikoslovnih analiz, oblikovana je metodologija gradnje korpusov in korpusne analize, na tej osnovi se gradijo jezikovni opisi in oblikujejo jezikovnotehnološke aplikacije. Ob tem pa področje ostaja odprto za druge veje jezikoslovja, ki korpus uporabljajo le kot klasično gradivo zgolj za preverjanje svojih hipotez.

¹ V slovenskem prostoru je prvi pravi korpusni slovarski projekt priprava velikega angleško-slovenskega slovarja Oxford-DZS; o njem natančneje poročamo v tej številki.

3 Sklep

Korpusi so v jezikoslovno delo vnesli besedilno gradivo, ki je količinsko in kakovostno absolutno preseglo predračunalniške gradivne zbirke, hkrati pa pokazalo na njihove omejitve in izpostavilo pomanjkljivosti. Vendar pa pojav korpusov v jezikoslovju pomeni mnogo več kot zgolj gradivo za jezikoslovno analizo. Korpusno jezikoslovje je ob metodologiji gradnje korpusov oblikovalo tudi metodologijo korpusne analize in novih jezikovnih opisov; ti v svojem izhodišču prisegajo na jezikovno realnost in tudi ob nepričakovanih rezultatih ne podlegajo intuiciji, vključujejo več podatkov o tipičnem besedilnem okolju ter sploh podatkov o komunikacijski realnosti.

Delo s korpusi je usmerjeno izrazito uporabniško, v svoj strokovni diskurz korpusno jezikoslovje tako izrazito vključuje dialog z uporabniki vedenja, ki se ustvarja in posreduje. Kljub programski odprtosti področja za vse veje jezikoslovja pa je komunikacija s tradicionalnimi jezikoslovnimi šolami in strokovnimi praksami velikokrat otežena, še posebej takrat, ko tradicionalni diskurz vrednostno hierarhizira pojavljajoče se nove strokovne diskurze, ki jih narekujejo spremembe v sodobni družbi.

Literatura

Atkins, Sue, Clear, Jeremy in Oster, Nicholas, 1992: Corpus Design Criteria. *Literary and Linguistics Computing* 7/1. 1–16.

Biber, Douglas in Finegan, Edward, 1991: On the exploitation of computerized corpora in variation studies. Aijmer, Karin in Altenberg, Bengt (ur.): *English Corpus Linguistics*. London in New York: Longman. 204–220.

Biber, Douglas, 1993: Representativeness in Corpus Design. *Literary and Linguistics Computing* 8/4. 243–257.

Biber, Douglas, Conrad, Susan in Reppen, Randi, 1998: *Corpus Linguistics. Investigating Language Structure in Use*. Cambridge: Cambridge University Press.

Čermák, František, 2002: Today's corpus linguistics. Some open questions. *International Journal of Corpus Linguistics* 7/2. 265–282.

Gorjanc, Vojko, 1999: Korpusi v jezikoslovju in korpus slovenskega jezika FIDA. Kržišnik, Erika in Lokar, Meta (ur.): *35. seminar slovenskega jezika, literature in kulture*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovanske jezike in književnosti Filozofske fakultete 47–59.

Gorjanc, Vojko, 2002: *Jezikoslovna načela gradnje računalniških besedilnih zbirk strokovnih jezikov*. Doktorska disertacija. Mentorica prof. dr. Ada Vidovič Muha. Ljubljana: Univerza v Ljubljani, Filozofska fakulteta.

Kennedy, Graeme, 1998: *An Introduction to Corpus Linguistics*. London: Longman.

Malmkjær, Kirsten, 1996 (1991): *The Linguistic Encyclopedia*. London, New York: Routledge.

- McEnery, Tony in Wilson, Andrew, 1996: *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Sampson, Geoffrey, 2003: Are we nearly there yet, Mum? Archer, Down, Rayson, Paul, Wilson, Andrew in McEnery, Tony (ur.): *UCREL Technical Paper number 16. Special issue. Proceedings of the Corpus Linguistics 2003 conference, Lancaster, 28–31 March 2003*. Lancaster: Lancaster university (UK). 678. <http://www.grsampson.net/Aawn.html>
- Stabej, Marko, 1998: Besedilnovrstna sestava korpusa FIDA. Kačič, Zdravko (ur.): *Uporabno jezikoslovje* 6. Tematska številka »Jezikovne tehnologije«. 96–106.
- Suhadolnik, Stane, 1968: Koncept novega slovarja slovenskega knjižnega jezika. 4. seminar slovenskega jezika, literature in kulture. Predavanja iz jezika. 1–11.
- Teubert, Wolfgang, 1995: Language Resources: The Foundations of a Pan-European Information Society. Rettig, Heike s sodelovanjem Júlie Pajzs in Gáborja Kissa (ur.): *TELRI: »Language Resources for Language Technology«*. Proceedings of the First European Seminar, Tihany, September 15 and 16. 105–128.
- Tognini-Bonelli, Elena, 2001: *Corpus Linguistics at Work*. Amsterdam, Philadelphia: John Benjamins.
- Váradi, Tamás, 2000: Corpus Linguistics – Linguistics or Language Engineering? Erjavec, Tomaž in Gros, Jerneja (ur.): *Informacijska družba, Jezikovne tehnologije*. Ljubljana 17.–19. okt. 2000. Ljubljana: Inštitut Jožef Stefan. 1–5.
- Zgaga, Pavel, 2003: *Bolonjski proces terja mnogo več kot »mehansko« spremembo strukture študija*. Teze za posvet »Modeli študija in uresničevanje bolonjskega procesa«, Ljubljana, 4. 6. 2003.