

JEZIKOVNOTEORETSKA NAČELA V KORPUSNEM JEZIKOSLOVJU

Darinka VERDONIK

Fakulteta za elektrotehniko, računalništvo in informatiko Univerze v Mariboru

Verdonik, D. (2015): Jezikovnoteoretska načela v korpusnem jezikoslovju. Slovenščina 2.0, 3 (1): 1–27.

URL: http://www.trojina.org/slovenscina2.0/arhiv/2015/1/Slo2.0_2015_1_01.pdf.

Predmet razprave so teoretsko-metodološka načela, ki so se razvijala v krogih t. i. novofirthijancev, kjer se od vsega začetka opredeljujejo za korpusno analizo, čim manj obremenjeno s predhodnimi jezikoslovnimi teorijami. V prispevku najprej pregledamo dela teh avtorjev, iz katerih izhajajo med drugim slovnica vzorcev (angl. pattern grammar), teorija leksikalnega proženja (angl. lexical priming), teorija konvencij in invencij (angl. theory of norms and exploitations) ter teorija kontekstne prozodije (angl. contextual prosodic theory). Nato povzamemo njihove zaledne predpostavke in stališča o jeziku kot rezultatih raziskovanja. Tako definiramo šest skupnih načel: predmet raziskovanja je jezikovna raba, tj. jezik v »kontekstu situacije«, raziskovalni fokus se obrne k temu, kar je običajno, raziskovalčeva intuicija je v vlogi evalvacije avtentičnih jezikovnih rab, jezikovne ravni (slovnica in slovar) so razumljene kot prepletene, na jezikovni sistem se gleda kot visoko dinamičen, v teoretskih temeljih pa zavzame eno osrednjih mest jezikovni vzorec. Nazadnje opozorimo tudi na nekatere omejitve, s katerimi se soočamo pri korpusnem pristopu.

Ključne besede: korpus kot teorija, popolni korpusni pristop, korpusno jezikoslovje, teorija jezika

1 UVOD¹

Preverjanje jezikoslovnih dejstev in hipotez v korpusu je postala v jezikoslovju,

¹ Avtorica se iskreno zahvaljujem vsem, ki so po branju prvih različic članka s svojimi komentarji in predlogi pripomogli h kvalitetnejši končni različici, še posebej obema recenzentoma.

ne samo korpusnem, splošno sprejeta praksa oz., kot zatrjuje Teubert (2005: 1): »Dandanes je korpus osnovni vir skoraj vsakogar, ki se ukvarja z jezikoslovjem. Nobena introspekcija ne more dosegati veljave brez preverjanja v avtentičnem jezikoslovnem gradivu. Korpusna analiza je postala ključni element skoraj vseh jezikoslovnih študij.«

Ob tem pa so krogi, ki v korpusih ne samo preverjajo jezikovna dejstva, ampak se ukvarjajo tudi z načrtovanjem in gradnjo korpusov ter razvojem postopkov in orodij korpusne analize, znani pod imenom korpusno jezikoslovje. Gre za eno najmlajših področij jezikoslovja. Razvijati se je začelo potem, ko so bili na voljo prvi zmogljivejši računalniki, od šestdesetih let 20. stoletja naprej, s koreninami v angleškem jezikovnem prostoru. Prvi in še danes močni centri so npr. University College London, Lancaster University, University of Birmingham, University of Nottingham in ameriška Northern Arizona University, zunaj anglosaškega sveta pa velja omeniti vsaj Karlovo univerzo v Pragi.

V slovenskem prostoru lahko za začetke korpusnega jezikoslovja štejemo aktivnosti, povezane s korpusom FIDA, ki segajo v leto 1997 (Erjavec idr. 1998), par let kasneje je sledil korpus Nova beseda (Jakopin 2003). Skupaj z izdelavo prvih korpusov so nastale tudi prve objave, ki so predstavljale področje dela in z njim povezana spoznanja in poglede na jezik. V tesni povezavi s tehnološkim »bratom« korpusnega jezikoslovja, jezikovnimi tehnologijami, jih je veliko objavljenih v zbornikih konference Jezikovne tehnologije, ki od leta 1998 do danes poteka vsaki dve leti, pomembnejšo prelomnico pa predstavljata tudi prva monografija (Gorjanc 2005) in zbornik (Gorjanc, Krek 2005) v slovenskem prostoru, ki področje pregledno predstavita. Tema so kasneje sledile danes že kar številne monografije (Gantar 2007; Zemljarič Miklavčič 2008; Vintar 2010; Arhar Holdt 2011; Šorli 2012; Logar et al. 2012; Logar 2013; itd.).

Korpusno jezikoslovje je veliko energije usmerilo v vprašanja, povezana z gradnjo korpusov (zasnova, način računalniškega zapisa, načini in sheme

označevanja, zbiranje besedil in avtorske pravice ipd.) in korpusno analizo (vključno z razvojem orodij), ob tem pa so se zlasti v jezikoslovnih krogih, tako ali drugače povezanih z birminghamsko univerzo (John Sinclair, Michael Stubbs, Michael Hoey, Wolfgang Teubert, Susan Hunston, Patrick Hanks idr.), že od začetkov opredeljevali za korpusno analizo, ki naj bo čim manj obremenjena s predhodnimi jezikoslovnimi teorijami in ki naj izhaja predvsem iz samega besedila. Vzporeden premik lahko dokaj sočasno ali pa vsaj kmalu zatem, v devetdesetih letih 20. stoletja, opazujemo tudi v računalniškem jezikoslovju, kjer se začnejo raziskovalci od pristopov, ki so temeljili na jezikovnih pravilih, obračati k statističnim metodam.

Tako je leta 2001 E. Tognini-Bonelli (2001) uvedla ločevanje med delnim korpusnim pristopom (angl. corpus-based approach), ki uporablja korpus, da preverja vnaprej postavljene hipoteze, in popolnim korpusnim pristopom (angl. corpus-driven approach), ki postavlja hipoteze izključno na podlagi rezultatov korpusnih analiz, ne pa na podlagi predhodnih, pogosto teoretsko izpeljanih predpostavk, in se v glavnem nanaša prav na dela z Birminghamom povezanega kroga avtorjev, ki jih obravnavamo v tem prispevku. Vendar je bilo z rabo termina popolni korpusni pristop v praksi kar nekaj zmede (več o tem npr. McEnery, Hardie 2012: 150–151), zato McEnery in Hardie predlagata termina korpus kot metoda (angl. corpus-as-method) proti korpus kot teorija (angl. corpus-as-theory), ki ju prevzamemo tudi tukaj.

Prav slednje, korpus kot teorija, bo predmet razprave v tem prispevku. Ker gre za zbir precej raznolikih teorij o jeziku, nas bo zanimalo, ali in kakšni so njihovi skupni teoretsko-metodološki temelji in zaledne predpostavke o jeziku in raziskovanju jezika, na katerih temeljijo. Čeprav v slovenskem prostoru posamezne teoretske poglede korpusnega jezikoslovja² nekatera predhodna dela že zelo usmerjeno predstavljajo (zlasti Krek 2013; Gantar 2008; Šorli 2013), pa še niso bili sistematično pregledani in pretreseni s stališča njihovih

² Tukaj in povsod naprej v prispevku vedno, ko uporabljamo zvezo »korpusno jezikoslovje«, mislimo predvsem novofirthijansko korpusno jezikoslovje, kot je predstavljeno v razdelku 2.

skupnih zalednih teoretsko-metodoloških načel o jeziku.

2 NOVI POGLEDI NA JEZIK V KORPUSNEM JEZIKOSLOVJU

Krog korpusnih jezikoslovcev, povezanih z birminghamsko univerzo, je znan po sledenju Firthovim pogledom na jezik (Firth 1962), tako se jih je prijelo ime novofirthijanci (McEnery, Hardie 2012). J. R. Firth (za več informacij o njem glej Chapman, Routledge 2005) je bil verjetno eden najbolj vplivnih angleških jezikoslovcev sredi 20. stoletja. Njegovi študenti so bili mnogi znani jezikoslovci, med njimi M. A. K. Halliday, avtor sistemske funkcijske slovnice, in John Sinclair, eden osrednjih jezikoslovcev novofirthijanskega kroga korpusnega jezikoslovja.

J. R. **Firth** je zagovarjal stališče, da je treba jezik opazovati v t. i. »kontekstu situacije«, s katerim je po njegovem mnenju pomen tesno povezan. Poudarja predvsem družbeno dimenzijo jezikovne rabe (lahko ga označimo kot antikognitivističnega), kar do določene mere odseva tudi korpusno jezikoslovje, ki vsaj v nekaterih segmentih iz tega razvije stališče, da je diskurz/besedilo zelo pomemben predmet jezikoslovnega raziskovanja in posledično mora korpus vedno vključevati le celotna besedila, ne pa samo posameznih fragmentov. Naslednji Firthov vpliv na korpusno jezikoslovje je oživitev pojmov kolokacija³ in koligacija,⁴ ki postaneta ena osrednjih terminov korpusnega jezikoslovja nasploh. Kolokacije postanejo »absolutni temelj preučevanja leksike in semantike in posledično /.../ temelj celotnega pristopa k jeziku« (McEnery, Hardie 2012: 132). Niz teh pojmov se pri Sinclairju in naslednikih nadaljuje še

³ McEnery in Hardie (2012: 122) opozarjata, da je bil pojem znan vsaj od srede 18. stol. dalje, po Firthu pa ga približno istočasno aktualizirata tako Sinclair kot M. A. K. Halliday. Kolokacije običajno pomenijo razmerje med posameznim izrazom (eno- ali večbesednim) in leksikalnimi enotami, s katerimi se ta v besedilu družijo bolj pogosto, kot bi pričakovali naključno. Hoey (2005) kasneje doda, da gre za psihološko asociacijo med leksikalnimi enotami, ki se manifestira v sopojavljanju v besedilih. Tudi v slovenskem prostoru je bil termin večkrat obravnavan, kot je pregledno predstavljeno v Gantar (2007).

⁴ Koligacije lahko razumemo vzporedno terminu kolokacij, le da se nanašajo na slovnične kategorije, v katerih in s katerimi se posamezni izrazi pojavljajo.

s semantičnimi preferencami/asociacijami in diskurzno/semantično prozodijo.⁵ Tretje, za korpusno jezikoslovje nadvse pomembno stališče Firtha je, da postavi preučevanje pomena (v nasprotju s preučevanjem slovnice) in leksike (ki jo ločuje od semantike) v središče jezikoslovja, v nasprotju s takrat prevladujočim Bloomfieldovim oz. v času vzpona korpusnega jezikoslovja v nasprotju s Chomskyjevim pogledom na jezik, ki oba poudarjata osrednje mesto slovnice.

Johna Sinclairja lahko označimo kot enega osrednjih jezikoslovcev novofirthijanskega kroga. Njegovo delo in posledično pogledi na jezik so med drugim tesno povezani z delom pri slovarju COBUILD, ki je znan kot pomembna leksikografska prelomnica. Za Sinclairjevo (1991; 2004) korpusno jezikoslovje je značilno poudarjanje potrebe po čim večjem obsegu korpusnega gradiva ter zavračanje preveč podrobnega, na teoriji temelječega označevanja gradiva s slovničnimi kategorijami – golo besedilo je zanj središčna točka, na katero se mora opirati jezikoslovec. Preveliko opiranje na slovnične oznake v gradivu po njegovem mnenju hitro odraža teorijo, na podlagi katere so oznake dodane, ne pa gradiva in s tem jezika samega, in posledično le stežka odkrijemo kaj zares novega o jeziku, česar ne predpostavlja teorija.

Naslednje pomembno stališče Sinclairja je, da obstaja tesna povezava med različnimi pomeni besede in strukturami, v katerih se pojavlja (pri tem mu strukture pomenijo kolokacije in sorodne vzorce). Z drugimi besedami, po njegovem mnenju obstaja v jeziku močna tendenca, da pomen in skladnja

⁵ Termina semantične preference (Stubbs 2001) oz. asociacije in diskurzne prozodije (Hoey 2005) sta povezana z raziskovanjem leksikalne semantike in izhajata iz opažanj, da se posamezne leme pogosteje rabljene skupaj z določeno semantično skupino besed (npr. določena lema se praviloma povezuje z besedami, ki označujejo nekaj težavnega). Pojem diskurzne prozodije naj bi pri tem, po mnenju nekaterih avtorjev, ločeval samo med pozitivnimi in negativnimi kategorijami, medtem ko naj bi semantične preference veljale za katerokoli semantično področje, kar pa drugi avtorji problematizirajo. McEnery in Hardie (2012: 138) tako, sledeč Sinclairju, nakažeta, da bi bilo ustrežnejše ločevanje med obema na podlagi semantično-pragmatične lestvice, pri čemer so semantične preference na semantični strani, diskurzna prozodija pa na pragmatični strani. V slovenskem prostoru o problematiki semantične prozodije na več mestih razpravlja M. Šorli (Šorli 2013; 2014).

delujeta soodvisno. Pri iskanju odgovora na vprašanje, ali je pomen tisti, ki pogojuje strukturo, ali obratno, pa zaključí, da niti eno niti drugo, ampak da so celotne strukture (tj. večbesedne enote) tiste, ki jih izberemo, oz. kot pravi njegovo načelo idiomatskosti (angl. idiom principle): »Načelo idiomatskosti pomeni, da ima jezikovni uporabnik na voljo veliko količino napol vnaprej sestavljenih fraz, ki predstavljajo zanj eno samo izbiro, čeprav se zdi, da jih lahko ločimo na segmente.« (Sinclair 1991: 110)

Kjer načelo idiomatskosti odpove, nastopi nasprotno načelo – načelo proste izbire (angl. open-choice principle): »To je način gledanja na besedilo kot rezultat velikega števila kompleksnih izbir. Na vsaki točki, kjer je enota zaključena (beseda, fraza ali stavek), se odpre širok nabor izbir in edina omejitev je slovničnost.« (Sinclair 1991: 109)

Slovnica vzorcev (ang. pattern grammar) (Hunston in Francis 2000) nadgradi izkušnje iz projekta angleškega slovarja COBUILD. Osredotočena je na iskanje slovničnih vzorcev (npr. išče vzorce tipa V n – glagolska skupina, samostalniška skupina) in jih povezuje z leksikalnimi enotami, s katerimi se pogosto družijo. Predmet opazovanja tako ni več posamezna leksikalna enota (beseda, fraza itd.), ampak njeno okolje – vzorec. Vzorec pri tem pomeni besede in slovnične kategorije besed, s katerimi se pogosto druží in ki (kar je prav tako pomembno) hkrati prispevajo k njenemu pomenu. Na tak način slovnica vzorcev dokazuje, da se posamezne besede vežejo pogosteje z določenimi slovničnimi kategorijami kot z drugimi. Slovnične kategorije ločujejo samo zelo površinsko in poenostavljeno, s čimer se skušajo izogniti preveliki odvisnosti od teorije, hkrati pa se tudi izognejo nedoslednostim, ki nastajajo pri označevanju besedila na podlagi tradicionalne oblikoslovne in skladijske klasifikacije. Slovničnih kategorij, ki jih ločijo, je tako skupaj samo devet in so definirane na način: glagolska skupina, samostalniška skupina, stavek, uveden z obliko *-ing*, itd. Pomembno opažanje avtoric slovnice vzorcev je, da so vzorci variabilni in da se ne pojavljajo vedno v istem zaporednem nizu (gl. npr. Hunston, Francis 2000: 59–66). Prav tako vzorci niso vedno povsem

zaporeden niz besed, ampak se lahko včasih v vzorec (tudi v angleščini, kjer je besedni red veliko bolj stabilen od fleksijskih jezikov, kot je slovenski) vrinejo druge besede. Nekateri vzorci so mejni in odprti za interpretacijo.

Michael Hoey je znan kot zagovornik leksikalnega pristopa k analizi teksta in besediloslovju. Skozi povezovanje korpusnih raziskav s psiholingvistiko je razvil v korpusnem jezikoslovju prepoznavno teorijo leksikalnega proženja⁶ (angl. lexical priming) (Hoey 2005). Z njo utemeljuje, da so kolokacije tesno povezane s pojavom, znanim v psiholingvistiki kot angl. priming⁷, kjer ta pojem pomeni, da prožilna beseda vpliva na prepoznavo ciljne besede, npr. prožilna beseda *telo* pripravi bralca/poslušalca, da hitreje prepozna ciljno besedo *srce*. Hoey na proženje gleda nekoliko drugače: medtem ko psiholingvistika opazuje razmerje med prožilno in ciljno besedo, Hoeyja zanima prožilna beseda sama po sebi, tisto, kar ta beseda proži, pa jo samo dodatno osvetljuje. Utemeljuje, da beseda s tem, ko jo srečujemo v različnih okoliščinah in sobesedilih, postane povezana s sobesedili in okoliščinami, v katerih je bila rabljena, in naše vedenje o tej besedi vključuje informacijo, da se sopojavlja z določenimi besedami v določenih vrstah konteksta. Enako velja za besedne nize, ki izhajajo iz te besede: tudi ti postanejo povezani s sobesedili in okoliščinami, v katerih se rabijo. Posledica posameznikovega srečevanja besede v določenih kontekstih je mentalno proženje besede za rabo v diskurzu, in to na vseh ravneh, od besedne do pragmatične in besedilne:

1. beseda je prožena za sopojavljanje z določenimi drugimi besedami (kolokacije),
2. beseda je prožena za pojavljanje v določenih semantičnih nizih,
3. beseda je prožena za pojavljanje v določenih pragmatičnih funkcijah,

⁶ Po našem najboljšem vedenju je to prvi predlog prevoda za Hoeyev »lexical priming« v slovenščino.

⁷ Prevod v slovenščino še ni ustaljen, pojavljajo se termini aktivacija, proženje ...

4. beseda je prožena za pojavljanje v določenih skladenjskih funkcijah in položajih (koligacije),
5. beseda je prožena za rabo v določenih slovničnih kategorijah,
6. beseda je prožena, da se pojavlja v določenih vrstah ali se izogiba določenim vrstam kohezivnih vezi v diskurzu,
7. beseda je prožena, da se pojavlja v določenih semantičnih razmerjih v diskurzu,
8. beseda je prožena, da se pojavlja v določenih položajih ali se izogiba določenim položajem znotraj diskurza.

Najpomembneje pa je, meni Hoey, da so vsa ta proženja krovno opredeljena z žanrom oz. specifično situacijo, kar pomeni, da prožimo besede v prvi vrsti tudi glede na družbeni kontekst jezikovne rabe. Prav povezanost proženja besed s specifičnimi konteksti je po Hoeyu (2005: 13) glavni razlog, da pojav ni bil opažen prej, saj po njegovih besedah jezikoslovje tradicionalno opazuje jezik kot monoliten in tudi korpusno jezikoslovje običajno dela s splošnimi korpusi, medtem ko določene značilnosti postanejo vidne šele, če delamo z bolj specifičnim gradivom.

Nenehna raba besed v novih in novih kontekstih po Hoeyu vpliva na to, da se obstoječi vzorci proženja ali utrdijo ali pa oslabijo, če srečamo besedo v nenavadnem, novem kontekstu. Leksikalno proženje je torej dinamična, ne pa stalna ali nespremenljiva lastnost besed.

Medtem ko Hoey raziskuje konkordanco v povezavi s kognitivno/psihološko dimenzijo jezikovne rabe, pa **Teubert** (2003; 2005; 2010), nasprotno, zagovarja stališče, da kognitivna/psihološka dimenzija jezikovne rabe ni in ne more biti predmet korpusnojezikoslovnega zanimanja. Na konkordanco gleda izključno znotrajdiskurzno ali boljše znotrajbesedilno. Besedilo/diskurz je namreč po njegovem mnenju edina realnost, ki je jezikoslovcu na voljo, in povezava med besedilom/diskurzom in zunajbesedilno realnostjo ni del jezikovnega sistema in posledično ne predmet raziskovanja v korpusnem

jezikoslovju. Po njegovem (Teubert 2010) je na dnu vsakršnega jezikoslovnega raziskovanja vedno vprašanje pomena, pomen, trdi, pa ni neki končen in dostopen mentalni koncept, ni kompleksen konstrukt univerzalnih semantičnih primitivov ali podobno, pač pa je nekaj, o čemer se posamezniki v diskurzu nenehno dogovarjajo in se torej lahko nenehno spreminja: vedno znova gre za vprašanje, kakšen pomen vtisne besedam tisti, ki jih izreče, in kako jih razume naslovnik. Edina jezikoslovna realnost je za Teuberta znotraj diskurza tisto, kar je nekdo v neki točno določeni situaciji izrekel/zapisal. Tudi koncept »jezika«, kot ga razume jezikoslovje, je za Teuberta samo konstrukt nešteti diskurzov o jeziku. Predlaga, da bi ideje, pojme, konceptualizacije, povezane z jezikom, ki jih najdemo v diskurzih, morali primerjati, ponovno interpretirati in tako plast za plastjo dodajati nove možne interpretacije. Bolj kot lahko interpretacije, kaj in kakšen je jezik, temeljijo na podatkih, ki si jih delimo, bolj bodo sprejemljive za družbo. V tem omejenem smislu vidi, da ima korpusno jezikoslovje prednost pred drugimi pristopi k jeziku.

Znotraj besedila išče pomen tudi **Hanks**, prav tako eden vidnejših sodelavcev pri slovarju COBUILD. Že Sinclair (1991) opozori na nenavadno visoko sopojavljanje nekaterih nizov besed in hkrati na presenetljivo mero variabilnosti tudi v stalnih besednih zvezah. Hanks (2013) v svoji teoriji konvencij in invencij (angl. theory of norms and explitations – TNE)⁸ nadgradi prav ta vidik korpusnojezikoslovnega opažanja ter uvede kriterije za sistematično ločevanje med običajnimi, konvencionalnimi vzorci in inventivnimi rabami teh vzorcev. Po mnenju Hanksa sta variabilnost jezikovne rabe in kreativnost uporabnikov pri tem tako velika, da jezikovne rabe ni mogoče razložiti v enem koraku, ampak se je treba najprej osredotočiti na to, kar je najbolj pogosto in običajno. Šele v drugem koraku pa se osredotočimo na

⁸ Hanksova »theory of norms and exploitations« je bila v slovenščino prvič prevedena v Gantar (2008) kot »teorija jezikovnih konvencij in možnosti njihove izrabe«, kar pa je zelo neprikladno za pogosto navajanje, zato Snojeva (2010: 39) predlaga poenostavitev v »teorija pravil in uporabe«, ki pa pomensko ni povsem skladen z izvirnim pojmom. Iz tega razloga smo se po posvetovanju z nekaterimi kolegi odločili ponuditi nov prevod, ki bi bil ustrezno kratek in pomensko čim bolj skladen s teorijo: »teorija konvencij in invencij«.

vse, kar odstopa od običajnih in pogostih vzorcev in pravil, ter to opišemo ne kot odstopa od norme, ampak kot jezikovne rabe, ki sledijo nekemu drugemu naboru pravil, ne istim pravilom, kot smo jih definirali v prvem koraku za običajne rabe. Skladno s tem razlaga, da imajo tudi jezikovni uporabniki dve vrsti kompetenc oz. dva medsebojno povezana sistema pravil jezikovne rabe. Prvi sistem poimenuje angleško »norm«, torej norma, pri čemer mu norma pomeni konvencionalno – uporabnik se mora naučiti konvencionalne rabe jezika. Drugi sistem poimenuje angleško »exploit«, torej izraba oz. invencija – uporabnik se mora naučiti izrabljati konvencijo na različne načine, da lahko pove nove stvari oz. da pove stvari na nov, zanimiv, nenavaden način. Pri tem pa Hanks eksplicitno opozarja, da med konvencijo in invencijo ni jasne, ostre meje, ampak je med njima široko sivo področje, kjer so nekatere rabe manj običajne kot druge.

Naslednje pomembno opažanje o lastnosti leksike, ki prav tako izhaja iz dela Sinclairja, je t. i. **semantična prozodija**. Sinclair pri opazovanju konkordanc za angleški frazni glagol *set in* (Sinclair 1991: 74–75) med drugim opozori, da je njegova najbolj nepričakovana lastnost, da se osebki, ki kolocirajo s tem glagolom, nanašajo večinoma na negativne stvari (*gniloba, propad, bolezen ...*). Za tovrstno kolociranje je Louw (1993) predstavil termin semantična prozodija, pri čemer pa naj bi bil avtor termina dejansko Sinclair (glej Louw, Chateau 2010), medtem ko npr. Stubbs (2001) uporablja termin diskurzna prozodija. Semantična prozodija se torej nanaša na opažanje, da mnoge besede oz. večbesedne enote izražajo določeno vrsto pomena, ki ga lahko pogosto označimo kot pozitivnega ali negativnega in ki se ustvari skozi sistematično zaporedje kolokacij, njegova osnovna vloga pa je, da izraža avtorjev odnos do določene pragmatične situacije (Louw 2000). Mnoge besede ali večbesedne enote torej niso nevtralne, ampak nosijo določen (večinoma samo pozitiven ali negativen) naboj. Pojem lahko vidimo kot enega od ključev, ki omogočajo povezovanje korpusnega jezikoslovja s pragmatično dimenzijo jezika. Nekatera zanimiva opažanja v zvezi s semantično prozodijo so, da so negativne

semantične prozodije pogostejše od pozitivnih, da (v angleščini) semantična prozodija kaže nekakšno povezanost s (glagolsko) prehodnostjo, da namensko izkoriščanje semantične prozodije omogoča efekt ironije, sicer pa razkriva avtorjevo neiskrenost, itd. (glej Louw 2000). Na konceptu semantične prozodije poskuša Louw postaviti teorijo kontekstne prozodije (angl. contextual prosodic theory), ki raziskuje širši pomen koncepta semantična prozodija za razumevanje jezika in življenja. Teorija razlaga semantično prozodijo v povezavi s kontekstom situacije, kot ga razume Firth, in skuša dokazovati, da sta situacijski in besedilni kontekst skladna in povezana. Tako na primer razlaga, da je dejstvo, da se v angleščini zveza *srčna kap* rabi skoraj vedno z nedoločnim členom (*a heart attack*) in skoraj nikoli z določnim (*the heart attack*), razložljivo s tem, da v jeziku in v življenju srčni kapi pogosto sledi kolokacija oz. dogodek: smrt (Louw 2000). Louw torej naredi nadvse pomemben korak za korpusno jezikoslovje, saj prestopi iz znotrajbesedilnega v zunajbesedilni, realni svet, v katerem poteka komunikacija.

Zunaj novofirthijanskega kroga je v korpusnem jezikoslovju manj poudarjanja in iskanja novih teoretskih pristopov k preučevanju jezika. Omenili bomo samo še eno, v procesiranju naravnega jezika zelo vplivno teorijo, tj. **shemska semantika**⁹ (angl. frame semantics) ameriškega jezikoslovca Charlesa Fillmorja (1982). Vendar ta ne izhaja iz čistega korpusnega pristopa, ampak nasprotno izdelava precej obsežen korpusnooznačevalni aparat, znan pod »znamko« FrameNet (gl. Krek 2008; Fillmore idr. 2003). Fillmore (1982) je glavnino svoje raziskovalne pozornosti namenil glagolom. Sprva jih jezikovnoteoretsko opisuje na podlagi pretvorbene in formalne slovnice, kasneje pa jo preko odvisnostne slovnice in teorije valence preoblikuje v nov opis, ki ga imenuje slovnica sklonov, vendar še vedno ni povsem zadovoljen z vsemi opisi. V sedemdesetih se tako seznanijo še s teorijo govornih dejanj, preformativnostjo in pragmatiko na splošno, kasneje pa še s konceptom

⁹ Terminologija, povezana s shemsko semantiko, je bila v slovenski prostor že prenesena (Žele 2001; Perko 2004; Krek 2008), tukaj jo prevzamemo po Kreku (2008), tako kot Može (2009).

prototipa v teoriji prototipov, ki izhaja iz predpostavke, da so nekateri člani kategorije bolj ključni kot drugi. Shemska semantika tako temelji na predpostavki, da obstajajo neke vrste situacijski, kulturni prototipi situacij, v katerih rabimo besede, in te situacije moramo poznati in razumeti, da lahko v celoti razumemo pomen besed. S tem pa se – če se vrnemo v času nazaj – dejansko zelo približa tistemu delu jezikoslovja in diskurzne analize, ki izhaja iz idej Malinowskega in kasneje Firtha, namreč da lahko pomen razumemo le v kontekstu situacije. Razlika, lahko rečemo kar ključna, pa je ta, da Fillmore kontekste situacij posplošuje in prototipizira v neke vrste kognitivno predstavo najbolj običajnih in pogostih lastnosti situacije, v kateri se besede rabijo. Shemska semantika gleda na leksiko preko meja slovnice ter (podobno kot novofirthijanci) poveže leksiko in slovnico, in to po relativno strogo določeni metodologiji. Prav tako ne ostaja znotraj besedilnega sveta korpusa, ampak (podobno kot semantična prozodija) prestopi meje besedila ter gradi opis na podlagi zunajbesedilnega, realnega sveta. Na primer glagol *aretirati* opiše s pomočjo t. i. jedrnih elementov *organ*, *osumljenec*, *obtožba*. To so elementi situacijske sheme, ki morajo biti prisotni v realnem svetu, v besedilu pa je lahko kateri od teh tudi samo implicitno, prek anaforičnega navezovanja ali kako drugače »prisoten«, ne pa nujno leksikaliziran.

Drugih korpusnojezikoslovnih avtorjev in centrov tukaj ne bomo omenjali, saj so za namene članka manj osrednji, velja pa vsaj omeniti, da je zelo močen center v ZDA vsaj še Univerza v Arizoni, kjer deluje Douglas Biber, usmerjen zlasti v korpusnojezikoslovno raziskovanje registrov, v kontinentalni Evropi pa Univerza v Pragi, kjer se ukvarjajo predvsem s prenašanjem odvisnostne slovnice v označevanje korpusov, ter Katoliška univerza Louvian v Belgiji, ki je znana po aplikativnih korpusnojezikoslovnih raziskavah o angleščini kot drugim/tujem jeziku. Z razmahom korpusnega jezikoslovja v zadnjem desetletju se seveda zelo intenzivno razvijajo še mnogi drugi centri.

3 SKUPNE PREDPOSTAVKE KORPUSNEGA JEZIKOSLOVJA O JEZIKU IN NJEGOVEM RAZISKOVANJU

V nadaljevanju korpusnojezikoslovne teorije, predstavljene v predhodnem razdelku, pregledamo s stališča njihovih zalednih skupnih predpostavk in stališč o jeziku kot rezultatih raziskovanja.

3.1 Predmet raziskovanja: od sistema k rabi

Korpusno jezikoslovje usmeri svoj pogled v rabo jezika v kontekstu. Ta usmeritev izhaja iz firthijanskih pogledov na jezik – Firth sodi med prve jezikoslovce, ki izpostavljajo pomen preučevanja jezika v kontekstu situacije; gre za znani termin »context of situation«, ki ga Firth (1964) povzame po Maĳinowskem (1946).

Toda koncept konteksta je zelo kompleksen in razumljen na več načinov (več o tem v Verdonik 2013). Del korpusnega jezikoslovja, predstavljenega zgoraj, se prevladujoče zanima predvsem za kontekst oz. sobesedilo. Teubert (2003) se celo eksplicitno omeji nanj; po njegovem mnenju namreč v besedilu ni več nikakršne zunanje resničnosti, zunanje okoliščine jezikovne rabe torej niso več niti relevantne niti dostopne. Skrčen na sobesedilo se koncept konteksta prevladujoče rabi pri Hunstonovi in Francisovi (2000) ter Hanksu (2013), v veliki meri tudi pri Sinclairju, ki pa eksplicitno izpostavi, da ločnice ne vidi tako ostro: v utemeljitvi rabe tega termina v svojem delu (Sinclair 1991: 171) pravi, da lahko kontekst običajno pomeni okoliški jezik, torej sobesedilo, da pa ne izključuje nujno nejezikovnega okolja. Toda že Stubbs (2001) na primer eksplicitno vključi širši kulturni kontekst v svoje korpusnojezikoslovne razprave, podobno se s povezavo med besedilom in zunanjim, nebesedilnim kontekstom jezikovne rabe veliko ukvarjata Louw (2000) in Hoey (2005).

3.2 Raziskovalni fokus: od mogočega k običajnemu

Osrednji namen uporabe zelo velikih korpusov za raziskovanje jezika je, da se odkrije, kaj je središčno, tipično/običajno in s tem naravno, ne pa, kaj je

mogoče, čeprav ni v korpusu nikoli realizirano ali je realizirano izjemoma. Pogostost določene rabe (ki je v predkorpusnem obdobju niti ni bilo mogoče zanesljivo ugotavljati) je tako izredno pomemben faktor. V kakšnem smislu, zelo dobro ponazarja naslednji Sinclairjev primer: »Zelo zgovorno je, če na primer med nekaj sto pojavitvami osnovne glagolske oblike ni nobena velelniška (op. avt.: govora je o angleških glagolskih oblikah). To ne pomeni, da v eteričnem svetu teoretične šole ali tradicionalne slovnice velelnik tega glagola 'ni mogoč'.« (Sinclair 1987: 108; povzeto iz Hunston, Francis 2000: 17)

V tem duhu je tudi korpusnojezikoslovno gledanje, po katerem enkratna pojavitev ne predstavlja pojava, ki ga je smiselno raziskovati; šele vsaj dva- ali večkratna pojavitev je lahko pokazatelj, da morda ne gre le za slučajnost, ampak sistemskost, ne le individualnost, ampak naravnost v jeziku.

3.3 Vloga raziskovalčeve intuicije: od kreacije k evalvaciji

Jezikoslovje kot dokaze, da so razlage pravilne, pogosto (nikakor ne vedno in ne na vseh področjih v enaki meri) uporablja (oz. je uporabljalo, saj s prihodom korpusov v vedno večji meri posega po primerih iz njih) umetno ustvarjene ali prilagojene primere jezikovne rabe, ki so po mnenju raziskovalca možni in verjetni oz. se jih »spomni«. Sinclair (1991: 39) poimenuje tako vlogo introspekcije kreacija.

Korpusno jezikoslovje ostro zavrne možnost, da si jezikovne primere izmišljamo, trdeč, da si noben resen znanstvenik ne more privoščiti, da podatke najprej sam kreira, potem pa jih razlaga (Hanks 2013: 20), in da je jezikoslovje do zdaj temeljilo na spekulaciji, ne na pravih dokazih (Sinclair 2004: 9). Prav tako primerjava z dejansko jezikovno rabo, zabeleženo v korpusih, dokazuje, da ljudje nismo preveč dobri pri poročanju o svojem lastnem jezikovnem obnašanju, ker si, med drugim, veliko bolj zapomnimo tiste rabe, ki so nenavadne in izstopajoče, kot najbolj vsakdanje in pogoste (Hanks 2013: 20–22). Korpusno jezikoslovje tako metodološko izhaja iz analize velike količine primerov realne (nikoli za namene analize ustvarjene, kreirane) jezikovne rabe.

Prvi korak analize je praviloma kvantitativni (predvsem pogostost pojavitve, tudi v smislu ključnih besed, besedne povezovalnosti ipd.), v drugem koraku pa se gradivo analizira kvalitativno in iščejo razlage za jezikovne pojave – šele tukaj nastopi raziskovalčeva introspekcija, katere naloga pa je samo evalvirati in razlagati dokaze, ne pa kreirati jih.

3.4 Razumevanje jezikovnih ravni: od razmejenega k prepletenemu

Splošno gledano se jezikovni opis pogosto deli na dve temeljni osi: slovar in slovnico. Čeprav nikakor ne moremo trditi, da ju jezikoslovje vidi ločeni, pa je vendarle pogosto namenilo več skrbi temu, da je ločilo predmeta njunega preučevanja (gl. še Hoey 2005: 1–2), kot pa usmerjenemu raziskovanju njunega prepletanja.

V korpusnem jezikoslovju postane preplet obeh temeljnih osi, slovarja in slovnice, veliko bolj kompleksen, kot je običajno v jezikoslovju, in na mestih se slovar in slovnica skorajda zlijeta v eno. Pomemben pojem, ki vodi k prepoznavanju globine prepletanja, je koligacija, torej slovnične kategorije ali slovnični nizi, v katerih se posamezna beseda pogosto pojavlja. Od tu je samo še korak do opažanj, kot jih najdemo v slovnici vzorcev (Hunston, Francis 2000), da se namreč posamezne skladijske strukture uresničujejo večinoma skozi določeno vrsto leksikalnih enot (in seveda tudi obratno, da se posamezne leksikalne enote realizirajo večinoma skozi določen tip skladijskih struktur). Leksika in skladnja sta torej izbrani vzajemno, ena z drugo, sta povezani in ju ne moremo opazovati ločeno. Še vedno pa pri slovnici vzorcev v ozadju obstaja distinkcija slovnica – slovar.

Hoey (2005) v teoriji leksikalnega proženja naredi še korak dlje in slovnico razloži kot rezultat pogostega skupnega pojavljanja ne samo posameznih leksikalnih enot, ampak tudi nižjih, kot so zlogi in glasovi. Slovnične kategorije so po njegovem samo priročna oznaka, ki jo damo kombinaciji (nekaterih) za posamezno besedo najbolj značilnih in od žanra neodvisnih proženj (Hoey 2005: 154). Slovnica je zanj zbir in preplet proženj najpogostejših glasov, zlogov

in besed jezika. Je zbir kolokacij, koligacij in semantičnih asociacij besed, kot so *je, in, v, se*, zlogov, kot so *en, il, ti, no*, in nekaterih glasov (Hoey 2005: 159), oz. drugače povedano – slovnica se oblikuje kot posledica proženj glasov, zlogov oz. besed in je torej sekundarna. (Hoey 2005: 159)

3.5 Razumevanje jezikovnega sistema: od stabilnega k dinamičnemu

V korpusnem jezikoslovju se na jezik ne gleda kot na sistem, ki se realizira v vsakokratnih rabah, ampak je jezik vrsta vsakokratnih rab in vzorci teh rab se nalagajo ter tako sestavljajo le okviren, ohlapen, večno spreminjajoč se, visoko dinamičen sistem (Hunston in Francis 2000: 17). V takem sistemu ni nič natančno določeno; v njem sicer lahko prepoznavamo tipično, pogosto oz. običajno, vendar obstaja še cel spekter netipičnega in nenavadnega, meje med enotami sistema so zabrisane in nejasne. Opisovanje jezika se posledično od pripisovanja definicij premakne k pripisovanju namigov o rabah, npr. o pomenu besed (gl. npr. Hanks 2013: 87), oz. k opisovanju rab v smislu pogostosti in posploševanja, ne pa tudi omejevanja (torej določanja mej pojavov). Opozarja se na individualno izkušnjo posameznika z jezikom, ta vodi v oblikovanje delno individualnih vzorcev rabe in posledično je »jezikovni sistem« pri vsakem posamezniku delno individualen (gl. npr. Hoey 2005: 159–160).

3.6 Teoretski temelji: od pravil k vzorcem

V korpusnem jezikoslovju postane osrednja enota raziskovanja vzorec (angl. pattern). Pomen pojma vzorec ni vnaprej podan, ampak se nekako spontano izoblikuje v predmet raziskovanja skozi empirične analize, in čeprav je v korpusnem jezikoslovju široko in pogosto uporabljan, je le ponekod eksplicitno naveden kot pomemben termin, npr. v indeksu (Sinclair 2004; Stubbs 2001), ali definiran in eksplicitno izpostavljen kot predmet raziskovanja (Hunston, Francis 2000; Hanks 2013). Do preučevanja vzorcev je prišlo prek stališča, da besedam določa pomen njihov kontekst, in opažanja, da »besede vstopajo v pomenske povezave z drugimi besedami okrog njih« (Sinclair 2004: 25) –

primeri so tako pogosti, da jih ni mogoče ignorirati: večbesedne slovnične enote, kot so npr. povratni glagoli ali sestavljeni glagolski časi, zloženke in sklopi, kjer celo norma niha pri odločitvi, ali so ena beseda ali dve ali več besed ..., poleg teh pa seveda še frazemi, pregovori, stalne besedne zveze in termini, na meji pa pogoste kolokacije ... Iz tega je sledilo stališče, da enota pomena ni beseda, ampak je v pretežni meri večbesedna, in glavna korpusnega jezikoslovja se je usmerila v raziskovanje povezanosti besed z drugimi, okoliškimi besedami v besedilu. Pri tem je bil privzet pojem vzorca za označevanje predmeta raziskovanja. Poglejmo inherentne lastnosti pojma vzorca, kot izhajajo iz njegove rabe v korpusnem jezikoslovju:

1. vzorec pomeni neko jezikovno strukturo, torej je sestavljen iz manjših enot; te enote lahko obstajajo na različnih jezikovnih ravneh: leksikalni, semantični, skladenjski itd.,
2. vzorec se v jezikovni rabi velikokrat pojavlja bolj pogosto, kot bi pričakovali na podlagi naključnosti, torej je lahko izstopajoče pogost, vendar to ni nujni pogoj,
3. vzorec definira pomenska povezanost – vsi elementi vzorca na neki način prispevajo k njegovemu pomenu v aktualnem kontekstu; to je nujni pogoj, da nekaj prepoznamo kot vzorec,
4. vzorec ni stabilen, ampak je variabilen: variira lahko tako v strukturi (npr. na ravni zaporedja elementov ali prisotnosti/odsotnosti istih ali alternativnih elementov) kot v pomenu (od konteksta do konteksta lahko dobiva vedno nove pomene),
5. vzorec lahko obstaja na različnih jezikovnih ravneh, do zdaj so bile izpostavljene predvsem leksikalna (tipične kolokacije), leksikosemantična (vzorec kot zaporedje tipičnih semantičnih kategorij, povezanih s posamezno besedo), leksikoslovnična (vzorec kot zaporedje tipičnih slovničnih kategorij, povezanih s posamezno besedo), leksikopragmatična (vzorec kot pripisovanje

negativnih/pozitivnih nabojev posamezni besedi), slovnicična (vzorec kot pogosto in prepoznavno zaporedje slovnicičnih kategorij).

Pojem vzorca se tako širi na številne pojave, ki jih jezikoslovje pogosto opisuje s pojmom pravil, morda do neke mere celo zaseda mesto, ki ga je tradicionalno jezikovni znak, čeprav dejansko pojma jezikovnega vzorca ne moremo vzporejati s pojmom jezikovnega znaka, kot je definiran zlasti v strukturalizmu. Slednji temelji na predpostavki, da obstajajo ideje, prototipi vsega: ideja drevesa, ideja stola ... Jezikovni znak zajema oboje, idejo in jezikovno obliko, ki idejo označuje. Korpusno jezikoslovje ne sledi tej smeri razmišljanja o jeziku (in svetu).

Zgoraj smo že nakazali, da se pojavi, ki smo jih tradicionalno navajeni opazovati kot jezikovna pravila, v korpusnem jezikoslovju opisujejo kot vzorci. Vseeno se pojem pravil občasno še pojavlja tudi v korpusnem jezikoslovju, npr.: »Teorija, ki jo predlagamo v tej knjigi, pravi, da mora vsak govorec naravnega jezika usvojiti ne enega, ampak dva sistema pravil (poudarek avt.) jezikovnega obnašanja.« (Hanks 2013: 8) Toda verjetno gre pri tem bolj za posledico navade in jezikovni vzorec govorjenja o stvareh, ki še ni bil zavestno prepoznan, saj rezultati korpusnih analiz, o katerih avtorji v korpusnem jezikoslovju pišejo, pojma pravil v resnici ne podpirajo. Tako tudi ta isti avtor samo dve strani naprej trdi, da je naravni jezik »v veliki meri sistem analogij, verjetnosti in preferenc« (Hanks 2013: 10). Za analogije, verjetnosti, preference ... pa je pojem pravila preveč determinističen, tog in statičen.

4 MEJE KORPUSNEGA JEZIKOSLOVJA

Ob pregledovanju jezikovnoteoretskih premikov se moramo zavedati tudi omejitev, s katerimi se soočamo pri korpusnem pristopu k raziskovanju jezika. Le-te med drugimi na kratko reflektira Hanks (2013: 80–81), ko opozarja, da v resnici ne vemo, ali pomen besed izhaja iz sobesedila ali iz področja rabe ali iz našega védenja o svetu ali česa drugega, in če jezik ne bi v tolikšni meri sledil tipičnim vzorcem, kot jim, korpusni pristop ne bi imel prav velikega dometa. V

par besedah na nevarnost omejevanja izključno na korpusnojezikoslovni pristop k preučevanju jezika (kot tudi na nevarnost izključevanja korpusnega pristopa) opozarja tudi Johansson (1991: 6). Dodatno kritično refleksijo korpusnojezikoslovne metodologije skušamo izdelati v nadaljevanju.

4.1 Problem »praznega iskalnega okenca«

Novinec v korpusnem jezikoslovju se najprej sreča s problemom praznega iskalnega okenca. Ta je tesno povezan z vprašanjem ortografskega znaka kot edinega ključa za iskanje po korpusu. Če hočemo v korpusu nekaj najti, moramo najprej vedeti, kaj bomo iskali, in sicer moramo to vedeti v obliki besede ali dela besede ali pa (opcijsko) slovnične kategorije, če je ta označena. Korpusno jezikoslovje se torej začne z iskanjem nizov črk, besed ... Toda kateri niz, katere besede so tiste, ki naj nas zanimajo? Če ne izhajamo iz neke teorije, imamo v korpusnem pristopu k raziskovanju jezika samo možnosti, da si ali najprej izdelamo seznam besed, ga opazujemo in iz njega izločimo kandidate za iskanje, ali pa izberemo besede za iskanje naključno in/ali na podlagi raziskovalčeve jezikovne intuicije. Na ročno pregledovanje besedil (po vzoru računalniške terminologije poimenujmo to kar brskanje¹⁰), ki bi prav tako lahko vodilo k določanju kandidatov za iskanje, se korpusno jezikoslovje praviloma ne opira. V praksi se to manifestira skozi dejstvo, da korpusni konkordančniki ponujajo samo iskanje, ne pa tudi brskanja – tudi tam, kjer razlog za to ni varovanje avtorskih pravic.¹¹ Toda če najprej ročno prebiramo (in poslušamo) gradivo, lahko tako na podlagi empiričnega opazovanja gradiva določimo besede oz.

¹⁰ Na tem mestu bi želeli opozoriti na teorijo, ki se vsaj v dveh vidikih prekriva s korpusnim jezikoslovjem, vendar ima ravno obratno metodološko pomanjkljivost kot korpusno jezikoslovje: temelji na »brskanju«, le redko pa ga nadgradi z iskanjem. To je konverzacijska analiza. Tako kot korpusno jezikoslovje utemeljuje, da je treba preučevati realno jezikovno rabo. Tako kot korpusno jezikoslovje pristopa k analizam brez kakršnih koli predhodnih teoretskih predpostavk, teoretske zaključke dela šele naknadno, na podlagi opažanj analiz. Vendar so tam analize ročne, s prebiranjem transkripcij in poslušanjem posnetkov, ne strojne, so kvalitativne, ne kvantitativne (za več podrobnosti o metodološkem postopku konverzacijske analize glej ten Have (1990)).

¹¹ Da pa ideja brskanja po korpusu le ni povsem nenavadna, kot izjema dokazuje konkordančnik michiganskega korpusa govornjene angleščine v akademskem okolju – MICASE, ki podpira tovrstno funkcijo.

vzorke, ki jih bomo v drugem koraku po korpusu iskali in podrobneje analizirali. Pristop k analizam (samo) prek besednih seznamov namreč ni uporaben za vse vrste raziskovalnih vprašanj. Za primer navedimo raziskovanje izrazov tipa *in tako naprej* v govoru. S korpusnojezikoslovno analizo večbesednih izrazov se nekaj najbolj pogostih tovrstnih izrazov pojavi na seznamu najpogostejših večbesednih izrazov (gl. npr. Adolphs in Carter 2013: 29–31). Toda šele če s kombinacijo brskanja in iskanja ključnih besed dodatno preiščemo korpusno gradivo (in seveda postopek zaključimo z ročno kvalitativno analizo), se nam pokaže cel spekter variabilnosti te skupine besed in njihova polna podoba (kot npr. v Verdonik in Kačič 2012), kar bi sicer ostalo neopaženo.

4.2 Vrata odpira samo ortografski znak

Korpusnojezikoslovna metoda je omejena na površinsko besedilo, ki ga (izvorno, brez dodatnih oznak, ki pa so praviloma že odvisne od neke predhodne jezikovne teorije) sestavljajo črke, ločila in presledki – in prav nič drugega. Ortografski znak je edino sredstvo, edini ključ, ki nam v korpusni metodi odpira vrata do jezikovne rabe. Korpusno jezikoslovje ne more brez komplementarnih metodoloških postopkov iskati, kje je nastal nesporazum ali konflikt, na kak način se skozi besedilo menjavajo teme in še vrste drugih, zelo relevantnih jezikoslovnih in diskurzivnih vprašanj. Vsako vprašanje mora predhodno pripraviti v tako obliko, da lahko prek prave kombinacije ortografskih znakov, njihove vrste in/ali zaporedja najde relevantne primere rabe. Zaradi te omejitve se raje osredotoča na tista vprašanja, ki so bolj neposredno povezana z ortografskim znakom.

4.3 Meje mojega besedila so meje mojega sveta

Korpusnojezikoslovno iskanje poteka znotraj besedila. S tem hočemo reči, da ne iščemo po komunikacijski situaciji in vsem, kar je z njo povezano – ne iščemo torej po dejanski jezikovni rabi, ampak samo po jezikovno »materializiranem« rezultatu le-te. Omejitev ni nujno problematična, če se je zavedamo in rezultate analiz interpretiramo s tem zavedanjem. Interpretacija je namreč nujen

sestavni del korpusne analize, ki sledi kvantitativnim analizam in le-te osmisli. Nasprotno, da lahko tudi z opazovanjem samo rezultatov komunikacijskih dogodkov povemo zelo veliko o samih dogodkih, dokaže Hoey (2005) s teorijo leksikalnega proženja oz. tudi Louw (2000) v teoriji kontekstne prozodije. Toda po drugi strani zasledimo tudi zavestno zapiranje raziskovalčevega zornega kota v meje besedila, kot na primer pri Teubertu (2005). Njegovi argumenti pri tem dejansko imajo težo: z drugimi, zlasti npr. sociolingvističnimi metodami lahko namreč povemo o komunikacijskih situacijah in vsej kompleksnosti družbenih razmerij, v katerih potekajo, več in še marsikaj drugega kot samo s korpusnojezikoslovno metodo.

5 ZAKLJUČEK

V prispevku smo predstavili poglede na jezik, ki se razvijajo v t. i. novofirthijanskem krogu korpusnega jezikoslovja, ter skušali sintetizirati njihove zaledne predpostavke in stališča o jeziku kot rezultatih raziskovanja. Na ta način smo definirali šest pomembnih skupnih načel v metodi in teoriji. Čeprav smo ta načela predstavili kot »premike« od prejšnjega k novemu razumevanju, je bilo to predvsem v namene ilustracije, v kakšni smeri se po našem razumevanju zgodovinsko gledano giblje razvoj jezikoslovne misli. To ne pomeni, da podobnih »premikov« ne najdemo tudi v kakšnih drugih vejah jezikoslovja, prav nasprotno. Za ilustracijo: jezikovno rabo štejejo za predmet preučevanja mnoge discipline znotraj analize diskurza, o povezovanju slovnice in slovarja govorijo zelo poudarjeno tudi v funkcijskem in kognitivnem jezikoslovju itd. Toda iskanje vzporednic z drugimi vejami jezikoslovja je lahko šele naslednji korak v razpravljanju o teh vprašanjih.

V zvezi s korpusnim jezikoslovjem se pogosto odpre vprašanje in diskusija o njegovi vlogi v jezikoslovju. Krog korpusnih jezikoslovcev, ki smo jih predstavili v tem prispevku, zavzema stališče, da je korpusno jezikoslovje teoretska paradigma sama po sebi. Toda zdi se, da so pri tem v manjšini in prevladujoče mnenje kaže v prvo smer (gl. npr. McEnery in Gabrielatos 2006; Teubert 2010).

Dejstvo je pač, da korpusi in orodja za delo z njimi dejansko omogočajo jezikoslovcu »poljubnega teoretskega prepričanja /.../, da svoje teze preverja na empiričnem jezikovnem gradivu« (Krek 2013: 20), in dobro je, da je tako. Za ožji krog korpusnega jezikoslovja, predstavljen v tem prispevku, pa vseeno lahko zaključimo, da ima izdelana teoretsko-metodološka načela in predstavlja precej samostojno in samosvojo jezikoslovno vejo ... kolikor je te danes še mogoče zares ločevati. Meddisciplinarnost (oz. znotraj disciplin sodelovanje in komplementarna uporaba različnih teoretskih modelov) je vse pogostejše glavna gonilna sila razvoja v znanosti in na področju jezikoslovja je to morda še najbolj izrazito prav pri korpusnem jezikoslovju v njegovem najširšem pomenu.

LITERATURA

Adolphs, S., Carter, R. (2013): *Spoken Corpus Linguistics: From Monomodal to Multimodal*. New York, London: Routledge.

Arhar Holdt, Š. (2011): *Luščenje besednih zvez iz besedilnega korpusa z uporabo dvodelnih in tridelnih oblikoskladenjskih vzorcev*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Chapman, S., Routledge, C. (2005): *Key Thinkers in Linguistics and the Philosophy of Language*. Edinburgh: Edinburgh University Press.
Dostopno prek: <http://www.lel.ed.ac.uk/homes/patrick/firth.pdf>
(3. 6. 2014).

Erjavec, T., Gorjanc, V., Stabej, M. (1998): Korpus FIDA. V T. Erjavec, J. Gros (ur.): *Zbornik Jezikovne tehnologije za slovenski jezik*: 124–127. Ljubljana: Institut "Jožef Stefan".

Fillmore, C. J. (1982): *Frame semantics. Linguistics in the Morning Calm: Selected Papers from SICOL-1981*. Seul: Hanshin Publishing Company.

Fillmore, C. J., Johnson, C. R., Petruck, M. R. L. (2003): Background to FrameNet. *International Journal of Lexicography*, 16 (3): 235–250.

Firth, J. R. (1962): A synopsis of linguistic theory. *Studies in Linguistic*

- Analysis*: 1–32. Oxford: Basil Blackwell.
- Firth, J. R. (1964): *The Tongues of Men & Speech*. London: Oxford University Press.
- Gantar, P. (2008): (Slovenska) leksika med leksikonom in slovnico. *Jezik in slovastvo*, 53 (5): 19–35.
- Gorjanc, V. (2005): *Uvod v korpusno jezikoslovje*. Domžale: Izolit.
- Gorjanc, V., Krek, S., ur. (2005): *Študije o korpusnem jezikoslovju*. Ljubljana: Krtina.
- Hanks, P. (2013): *Lexical Analysis: Norms and Exploitations*. Cambridge, London: The MIT Press.
- ten Have, P. (1990). Methodological issues in conversation analysis 1. *Bulletin de Méthodologie Sociologique*, 27: 23–51.
- Hoey, M. (2005): *Lexical Priming: A New Theory of Words and Language*. London, New York: Routledge.
- Hunston, S., Francis, G. (2000): *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam, Philadelphia: John Benjamins.
- Jakopin, P. (2003): Nekaj zanimivosti iz besedilnega korpusa Nova beseda. *Jezikoslovni zapiski*, 9 (2): 145–152.
- Johansson, S. (1991): Computer corpora in English language research. V S. Johansson in A. B. Stenstrom (ur.): *English Computer Corpora: Selected Papers and Research Guide*. Berlin: Mouton de Gruyter.
- Krek, S. (2013): Korpusne metode in njihov odsev v jezikoslovnih teorijah 20. stoletja. *Slovenščina 2.0*, 1 (1): 4–23.
- Krek, S. (2008): Framenet in slovenščina. *Jezik in slovastvo*, 53 (5): 37–54.
- Logar, N. (2013): *Korpusna terminografija: Primer odnosov z javnostmi*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za

družbene vede.

- Logar, N., Grčar, M., Brakus, M., Erjavec, T., Arhar Holdt, Š., Krek, S. (2012): *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: Gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.
- Louw, B. (1993): Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. V M. Baker, G. Francis, E. Tognini-Bonelli (ur.): *Text and Technology: In Honour of John Sinclair*: 157–176. Amsterdam: John Benjamins.
- Louw, B. (2000): Contextual prosodic theory: Bringing semantic prosodies to life. V C. Heffer, H. Sauntson (ur.): *Words in Context: A Tribute to John Sinclair on his Retirement*: 48–94. Birmingham: University of Birmingham.
- Louw, B., Chateau, C. (2010): Semantic prosody for the 21st Century: Are prosodies smoothed in academic context? A contextual prosodic theoretical perspective. V S. Bolasco, I. Chiari, L. Giuliano (ur.): *Statistical Analysis of Textual Data: Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles*: 755–764. Dostopno prek:
http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0755-0764_129-Louw.pdf (23. 6. 2015).
- Malinowski, B. (1946/1923): The problem of meaning in primitive languages. V C. K. Ogden, I. A. Richards (ur.): *The Meaning of Meaning: A Study of the Influence of Language upon Thought*: 296–336. New York: Harcourt, Brace & World, Inc.
- McEnery, T., Gabrielatos, C. (2006): English corpus linguistics. V B. Aarts, A. McMahon (ur.): *The Handbook of English Linguistics*: 33–71. Oxford: Blackwell.
- McEnery, T., Hardie, A. (2012): *Corpus Linguistics: Method, Theory and*

Practice. Cambridge: Cambridge University Press.

Može, S. (2009): Semantično označevanje korpusa slovenščine po modelu FrameNet. V M. Stabej (ur.): *Infrastruktura slovenščine in slovenistike*: 265–269. Ljubljana: Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik.

Perko, G. (2004): *Razločevanje prevodnih ustreznici v dvojezičnem uvezovalnem slovarju: Predlogi za slovensko-francoski slovar*. Doktorska disertacija. Ljubljana: Filozofska fakulteta.

Sinclair, J. (1987): Grammar in the dictionary. V J. Sinclair (ur.): *Looking up: An Account of the COBUILD Project in Lexical Computing*: 104–115. London, Glasgow: Collins ELT.

Sinclair, J. (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (2004): *Trust the Text: Language, Corpus and Discourse*. London, New York: Routledge.

Snoj, J. (2010): *Metafora v leksikalnem sistemu*. Ljubljana: Založba ZRC.

Stubbs, M. (2001): *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford, Malden: Blackwell Publishers.

Šorli, M., ur. (2012): *Dvojezična korpusna leksikografija. Slovenščina v kontrastu: novi izzivi, novi obeti*. Ljubljana: Trojina, zavod za uporabno slovenistiko

Šorli, M. (2013): Forms of encoded pragmatic meaning: semantic prosody: A lexicographic perspective. *Lingue e linguaggi*, 10: 95–111. Dostopno prek: <http://siba-ese.unisalento.it/index.php/lingueilinguaggi/article/view/13459> (19. 12. 2014).

Šorli, M. (2014): Pragmatic meaning in lexicographical description: Semantic prosody on the go. V A. Abel, C. Vettori, N. Ralli (ur.): *The User in Focus: Proceedings of the XVI EURALEX International Congress*:

477–493. Bolzano: Institute for Specialised Communication and Multilingualism.

Teubert, W. (2003): Writing, hermeneutics and corpus linguistics. *Logos and Language*, 4 (2): 1–17.

Teubert, W. (2005): My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10 (1): 1–13.

Teubert, W. (2010): Rethinking corpus linguistics. V A. Sanchez, M. Almela (ur.): *A Mosaic of Corpus Linguistics*, 19–42. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang GmbH.

Tognini-Bonelli, E. (2001): *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Verdonik, D. (2013): Koncept konteksta v jezikoslovnih in diskurzivnih teorijah. *Slavistična revija*, 61 (4): 631–649.

Verdonik, D., Kačič, Z. (2012): Označevalci odprte propozicije v govorjenem diskurzu. *Slavistična revija*, 60 (1): 79–94.

Vintar, Š., ur. (2010): *Slovenske korpusne raziskave*. Ljubljana: Znanstvena založba Filozofske fakultete.

Zemljarič Miklavčič, J. (2008): *Govorni korpusi*. Ljubljana: Znanstvena založba Filozofske fakultete.

Žele, A. (2001): *Vežljivost v slovenskem jeziku*. Ljubljana: ZRC SAZU.

PRINCIPLES IN LINGUISTIC THEORY OF CORPUS LINGUISTICS

The paper discusses theoretical and methodological principles developed by the circle of linguists known as the neo-Firthians, i.e. a group of linguists that argue for a pure empirical approach to corpus data, ignoring linguistic probabilistic theories as much as possible. First, an overview of the seminal works and findings of the neo-Firthian authors is made, including the J.R. Firth's view of language, John Sinclair's contribution to corpus investigation, the pattern grammar by Hunston and Francis, Hoey's theory of lexical priming, Hanks' theory of norms and exploitations, and Teubert and Louw's investigations on semantic prosody. In the second part, we summarize their core ideas of and approaches to language research, thus defining six principles: the subject of research is language use, i.e. language in the 'context of situation'; the focus of research is what is typical in language; the researcher's intuition is employed only for the evaluation of authentic language data; lexis and grammar are interconnected; the language system is highly dynamic; language patterns are central to language theory. Finally, we outline some limitations of using corpus approach in language research.

Keywords: corpus as theory, corpus-driven approach, corpus linguistics, theory of language

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/>

