

Korpusni pristop kot metoda raziskovanja in učenja tujega strokovnega jezika

Neva Čebroň

Fakulteta za humanistične študije, Univerza na Primorskem
neva.cebron@fhs.upr.si

Prispevek prikazuje postopke zbiranja in raziskovanja dveh primerljivih strokovnih korpusov. Njuna primerjava prikazuje poenostavljeno večdimenzionalno analizo jezikovnih sprememb s primerjavo list besed, ključnih besed in z opisom kolokacijskih in sintagmatskih vzorcev rabe besed. Za raziskavo uporabi brezplačne računalniške programe in utemelji rabo strokovnih korpusov kot ustrezno orodje za pomoč učiteljem tujega strokovnega jezika pri luščenju ključnega strokovnega in polstrokovnega besedišča.

Ključne besede: tuji strokovni jezik, pedagoškodidaktični pristopi, korpusno jezikoslovje

The present paper examines various methods of compiling and analysing two LSP corpora. By means of comparing frequency lists and keywords, a basic multidimensional analysis is carried out in order to illustrate language variation in different text types and genres from the two specialist domains in terms of collocational and colligational patterns of language use. A number of freely available corpus analysis tools are explored to support the claim that corpus-based research can be helpful in identifying key specialist and semi-specialist vocabulary and thus enhance LSP teaching courses.

Keywords: LSP, teaching approaches, corpus linguistics

Korpusno jezikoslovje je mlado in zelo živahno področje jezikovnih raziskav, saj se je v zadnjih desetletjih umestilo kot osrednji raziskovalni pristop v leksikografiji in leksikologiji (npr. Sinclair 1982 in 1991, Biber, Conrad in Reppen 1998, Carter 1998, McEnery in Wilson 2001, Gorjanc 2005), posledično pa tudi v uporabnem jezikoslovju, še posebej pri uvajanju novih pristopov poučevanja splošnega in strokovnega tujega jezika (npr. Burnard in McEnery 2000, Flowerdew 2002, Sinclair 2004, Gavioli 2005, Granger et al. 2009). Korpusna metoda jezikovnega raziskovanja opazuje in analizira jezikovno rabo ob podpori zbirke avtentičnih besedil v računalniško berljivi obliki. Omogoči nam, da pregledujemo večje, načrtno zbrano jezikovno gradivo, kar nam pomaga analizirati

in opredeliti pojavnost jezikovne rabe na osnovi večjega vzorca primerov resničnega ubesedovanja. Na osnovi empiričnega pristopa k preučevanju avtentičnih rab jezika je korpusna metoda preusmerila jezikovne raziskave od intuitivnega opredeljevanja slovničnih struktur k luščenju izrazja in analizam leksikogramatičnih paradigem na osnovi podatkov v besedilnih korpusih (Sinclair 1991, McEnery, Xiao in Tono 2006, Hyland in Bondi 2006).

Korpusni raziskovalni pogled na novo opredeli tudi vlogo in mesto strokovnega jezika, saj v svojih raziskavah dokazuje, da se strokovni jeziki ločijo od splošnega jezika na več ravneh (Biber 1995, Bhatia 2004, Halliday et al. 2004, Flowerdew 2005, Paltridge in Starfield 2013). Ugotavlja skladske, slogovne, besedilne, pragmatične

in semantičnoleksikalne posebnosti, ki označijo posamezni strokovni jezik, da v njem lažje upovedujemo specializirana področja vedenja in znanja. Hkrati opozori tudi na razlike pri upovedovanju v različnih besedilnih zvrsteh. V času, ko sporazumevanje v delovnem okolju pogosto poteka v tujem jeziku, posebno v angleščini, so ugotovitve korpusnega raziskovanja strokovnih jezikov pomembne usmeritve pedagoškega dela. Pouk tujega strokovnega jezika jih mora upoštevati, da smiselno nadgradi študentovo splošno jezikovno znanje in ga pripravi za razumevanje tujega strokovnega jezika, kakor tudi za razvijanje sporazumevalne zmožnosti in rabo tujega strokovnega jezika v delovnem okolju (npr. Lewis 1997 in 2000; Willis 1990, Hyland 2004 in 2006, Nesi 2013).

Postavlja se torej vprašanje, kako naj analiziramo in raziskujemo posamezni strokovni jezik, da bomo prepoznali vsaj pomembnejša odstopanja od splošne rabe in jih vgradili v pouk tujega strokovnega jezika na nadaljevalni izobraževalni ravni (natančneje, na ravni B1 do C1 po SEJO). Prispevek predlaga več pristopov in orodji za analizo korpusov strokovnih in polstrokovnih besedil, ko uvaja metodologijo iskanja in luščenja osrednjega izrazja in terminologije, ki naj pomaga študentu pri sporazumevanju v strokovnem okolju. Nadalje raziskava uvaja vpogled, ki jih daje primerjava strokovnih korpusov s področja humanistike (kulturologije – v nadaljevanju KK) in naravoslovno-tehničnih ved (prometnih ved – v nadaljevanju KP), kakor tudi medjezikovna primerjava podatkov iz specializiranih korpusov s podatki v splošnih korpusih (BNC, English Web 2015). Osredotoči se na rabo strokovnega in polstrokovnega jezika v različnih besedilnih žanrih vsakega strokovnega področja posebej. Na osnovi teh spoznanj skuša opredeliti pristope k poučevanju tujega strokovnega jezika na nadaljevalni stopnji.

V članku najprej predstavimo korpusni raziskovalni pristop in orodja korpusne analize. Teoretične vpogled, nato ponazorimo na primeru dveh strokovnih korpusov, ki smo jih zbrali za namen raziskave: korpus humanističnih besedil

(KK) in korpus besedil prometne stroke (KP). Izsledki, ki nam jih ponudi računalniška obdelava korpusov, tvorijo izhodišča za raziskavo strokovnega in polstrokovnega besedišča, prepoznavanje kolokacij in večbesednih enot, kot tudi za vpletanje ugotovitev korpusne analize v učna gradiva tujega strokovnega jezika. V zaključni razpravi opredeljujemo, utemeljujemo in ovrednotimo pedagoške pristope za uvajanje vpogledov korpusne metode v učenje angleščine kot strokovnega jezika.

Pregled literature

Jezikovni korpusi in vse bolj izpopolnjena računalniška orodja za manipulacijo z njimi omogočajo vrsto jezikovnih raziskav, zato se uveljavljajo v leksikografiji, leksikologiji, pri terminoloških in semantičnih raziskavah (npr. Hockey 2000, Meyer 2002, Reppen, Fitzmaurice in Biber 2002, Killgarriff 2005), pri razčlenjevanju slovničnih vpogledov, pri raziskavah besedilnih zvrsti, raziskavah na področju pragmatike ter celo v sociolingvističnih in medkulturnih analizah (npr. Biber 2006, Bahtia 2004, Carter in McCarthy 2006). Novi vpogledi, ki jih daje korpusna metoda, vplivajo tudi na pedagoškodidaktično metode poučevanja tujega jezika (npr. Granger 1998, Conrad 2000, Lewis 2000, Johns 2002, Flowerdew 2005, Gavioli 2005, Hyland 2006, Nation 2006).

Ko jezik raziskujemo s korpusnim pristopom, izhajamo iz dognanj posameznih vej jezikoslovja, računalniška obdelava podatkov pa nam pomaga hitreje in obširneje pregledovati dokumente, ki dajejo empirične podatke o jezikovni rabi. Marsikatera pretekla izpeljava se tako pokaže v novi luči, saj jo ta metoda dopolni in nadgradi. Pri raziskavah jezikovnih korpusov zaključke namreč izpeljujemo na osnovi podatkov o udejanjeni rabi velikega števila vzorcev (McEnery in Wilson 2001). Gre torej za zamik fokusa jezikovne analize od preučevanja struktur k preučevanju njihove rabe (Biber et al 1998). Ta perspektiva raziskuje, kako govorci in pisci uporabljajo jezikovna sredstva svojega jezika. Znotraj polja raziskav uporabnega jezikov-

slovja korpusna metoda obdelave jezika spodbuja preskok od normativne razlage jezika k funkcionalnemu opisu jezikovne rabe, k obravnavi jezika kot celovitega sistema (Dixon 2005). Svoje raziskave jezika tako jezikoslovci usmerijo k povezanosti oblike in pomena, v leksikogramatični opis jezika (Sinclair 1991, Carter in McCarthy 2006) ali k povezanosti leksikogramatične ravni jezika z vrsto upovedovanja, torej z govornimi položaji, besedilno zvrstjo ter govoreno ali zapisano besedo. Medsebojni vpliv različnih nivojev jezikovne analize pomeni tudi to, da od preučevanja posamezne besede oziroma termina, preidemo k opazovanju besednih zvez in besedilnega okolja, torej k določanju sintagmatskopomenonosnih jezikovnih enot (Halliday et al 2004, Hunston 2002) oziroma od skladijskih raziskav stavkov k pomenskoskladijski analizi njihove vloge v diskurzu (Biber 2006, Flowerdew 2002).

Statistični raziskovalni pogled na novo opredeli tudi vlogo in mesto strokovnega jezika, saj dokazuje, da se strokovni jeziki ločijo od splošnega jezika na več ravneh (Bhatia 1993, Biber 1995, Hyland 2004 in 2006, Flowerdew 2013). Ugotavlja skladijske, slogovne, besedilne, pragmatične in semantičnoleksikalne posebnosti, ki označijo posamezni strokovni jezik, da v njem lažje upovedujemo specializirana področja vedenja in znanja. Strokovni jezik tako poimenuje za jezikovno podvrsto in opazuje funkcionalne razlike pri izbiri jezikovnega registra od poljudnoznanstvenega do strokovnega. Zato ga opredeljuje glede na nivo strokovnosti oziroma glede na nivo formalnosti upovedovanja ali v kontekstu besediloslovja.

Sorazmerno enostavno lahko torej raziščemo, ali je določena raba pogosta v jeziku nasploh ali le v nekaterih zvrsteh in podzvrsteh besedil. Korpusna analiza jezika nam omogoči dovolj zanesljivo opredelitev osrednjega besedišča, torej besedišča, ki ga v nekem jeziku ali zvrsti jezika najpogosteje uporabljamo. Nadalje nam izostri vpogled v pomene, ki so za izbrani termin najpogostejši v strokovnih podzvrsteh določenega jezika.

Korpusna analiza je opozorila tudi na razlike med pristopi pri učenju tujega jezika, ki jih predlagajo učbeniki za angleški jezik, in podatki o jezikovni rabi, ki jih dajejo korpusne analize. Pri strokovnem jeziku poudari nujno, da študente izpostavimo sodobnemu strokovnemu jeziku, kakršnega najdemo v resnični, naravni rabi, da preusmerimo poudarek od slovničnih struktur k leksikalnim oziroma k leksikogramatičnemu opisu jezika in strokovni terminologiji, pri tem pa upoštevamo tudi pogostost jezikovnih elementov in njihovo razpršenost v besedilnih zvrsteh (Tognini-Bonelli 2001, Nesi 2013).

Gre torej za radikalno prevrednotenje pristopov k poučevanju tujega jezika in Willis (1990) ugotavlja, da je najprej potrebno zastaviti korpusno raziskavo ter nanjo opreti naše odločitve o tem, kateri tipični elementi jezikovne rabe najbolj predstavijo ubesedovanje v posamezni stroki. V središče učnih pristopov moramo torej postaviti jezikovno rabo in ne teorijo o jezikovnih strukturah. V leksikalnem učnem načrtu, dopolnjuje Hunstonova (2002), je glavni poudarek na najpogostejših besedah, osrednjih vzorcih rabe in kolokativnih zvezah, ki jih najpogostejše besede tvorijo med sabo. Te zveze vključujejo tudi najpogostejše slovnične strukture.

Poglobljeno uzaveščanje jezikovnih možnosti in omejitev je še posebej pereč problem na nadaljevalnih stopnjah učenja jezika (Lewis 1993 in 2000, Granger 1998, Schmitt 2000). Vzrok, da mnogi študentje tujega jezika zastanejo na nekakšni srednji stopnji in ne napredujejo več, je preprosto ta, da niso navajeni opazovati, kako besede tvorijo »besedne vzorce« in katere besede se med seboj družijo. Čeprav poznajo slovnično in veliko posameznih besed, ne znajo teh besed in struktur povezati v kolokativne zveze, ki bi jim pomagale izraziti misli, ko pišejo ali govorijo. Če ne uporabijo kolokacijskih ali koligacijskih zvez, ki pomagajo pomen izraziti enostavno in natančno, se izražajo zapleteno in uporabljajo kompleksne slovnične strukture in nenavadne besede. Ravno takrat, ko poskuša študent prilagoditi tuji jezik učinkoviti sporazumevalni rabi, naredi največ napak, saj uporablja jezi-

kovne strategije, ki jih pozna iz maternega jezika (Widdowson 1979). V tem smislu je pomembno opozorilo korpusnih raziskav strokovnega jezika predvsem, da moramo biti pozorni na polstrokovno besedišče oziroma na kolokacijske zveze, ki jih vsaka stroka prilagodi svojim potrebam upovedovanja. Zgolj učenje strokovnih terminov namreč študentom ne omogoča natančno razumevanje in upovedovanja strokovnih vsebin.

Pionirsko delo pri določanju osrednjega besedišča različnih strok v angleščini sta opravila Nation (1990) in Coxhead (2000). Nation (2006) je razvil računalniški program Range, ki naj učitelju tujega strokovnega jezika pomaga analizirati in izbrati ključno strokovno terminologijo kot nadgradnjo splošnega jezika. Čeprav so liste besednih družin v znanstvenem jeziku (*Academic Word Lists* – v nadaljevanju AWL) (Coxhead 2000) nekoliko kontroverzne, pogosto usmerjajo opredeljevanje osrednjega strokovnega in polstrokovnega besedišča kasnejših raziskovalcev (npr. Schmitt 2000, Hyland 2006, Min Zhang 2013). Ti opozarjajo, da se strokovni jeziki namreč pomembno razlikujejo med seboj in tudi od splošne rabe, zato nam šele manjši, strokovno usmerjeni korpusi razkrijejo, katera raba in pomen termina je v posamezni stroki najpogostejša ali ustaljena.

Predlagajo torej kontrastivno analizo referenčnih korpusov (npr. BNC, English Web 2015) s strokovno usmerjenimi korpusi oziroma primerjavo korpusov za posamezne stroke ali celo segmentov posameznih korpusov med seboj (Biber 1995, Hyland 2006). Opozarjajo še, da je za tovrstne raziskave bistvena harmonizacija in reprezentativnost korpusa (McEnery et al. 2006), da mora torej raziskovalec pozorno izbrati besedila, ki jih sestavi v korpus, saj bo tako lažje razlagal nove uvide, izpeljeval zaključke ali pomagal študentom ozaveščati strokovno besedišče. Zato Flowerdew (2013), Conrad (2000), Biber in Finegan (1989), Bowker in Person (2002) ter Stubbs (2002) poročajo o rabi lastnih manjših korpusov, ki obsegajo le nekaj deset tisoč besed, vendar vsebujejo besedila ene same, ozko določene stroke ali celo besedilne zvrsti znotraj določene stro-

ke. Šele taki korpusi dajejo relevantne podatke o jezikovni rabi na nekem strokovnem področju. Če te podatke primerjamo s statistikami v referenčnih korpusih, ugotovimo, da so velike razlike tako pri pojavnosti in pomenskih niansah najpogostejšega besedišča, kakor tudi pri največkrat rabljenih slovničnih strukturah.

Učitelj strokovnega jezika lahko uvide, ki jih ponuja korpusna metoda, uvaja v učni proces z neposrednim ali posrednim pristopom (Johns 2002, Charles 2007, Römer 2011, Jablonkai in Čebren 2017): podatke iz korpusne analize uporabi za izdelavo novih učnih gradiv ali vodi študente pri samostojnemu raziskovanju strokovnih korpusov. Obe metodologiji ozaveščata opis strokovnega jezika, ki poleg tipičnih terminov posamezne stroke, izpostavi tudi splošno besedišče, ki ga stroka za svojo rabo opredeli s tipičnimi podpomeni ali sprejme kot značilno stilsko orodje pri svojih opisih (Oakey 2002). Novosti, ki jih je korpusna metoda izpostavila, sta: sopomenke se najpogosteje družijo z različnimi besedami, vendar različni podpomeni termina izberejo različne, a ustaljene fraze in kolokacijske zveze; raba terminov pa ni enako pogosta v vseh njihovih pomenih in podpomenih. Prav v tej smeri velja iskati posebnosti strokovnih jezikov.

Učitelji tujih strokovnih jezikov smo običajno jezikoslovci in ne strokovnjaki za posamezne stroke, zato smo zaznali pomen strokovnega izražanja in od nekdanj zbirali strokovno terminologijo. Korpusna metoda analize jezika nam torej omogoči, da postavimo poučevanje tujega jezika na raven, ki preplete učenje tujega jezika s stroki prilagojenim prevzemanjem ubesedovanja.

Cilji raziskave in raziskovalna vprašanja

Izhajajoč iz predhodnih raziskav in ugotovitev avtorjev iz prejšnjega poglavja, želi pričujoča raziskava odgovoriti na vprašanje, kako lahko smiselno uporabljamo računalniško podprte analize besedilnih zbirk v KK in KP v angleščini za sledeče namene uporabnega jezikoslovja:

- pri luščenju ključne terminologije posameznih strok v angleškem jeziku,

- pri določanju strokovnega besedišča, ki naj bi ga uvajali pri poučevanju angleškega strokovnega jezika kulturologije in prometne stroke,
- pri primerjalnem opredeljevanju semantičnih polj ključnih terminov v teh dveh strokah,
- pri primerjavi shematičnih frazeoloških paradigem v obeh strokovnih domenah,
- pri izbiri didaktične metode za poučevanje angleškega strokovnega jezika.

Zaradi ključne vloge strokovnih jezikov za natančno izražanje pomenov v sodobni znanosti je pomembno, da posamezno jezikovno področje tudi natančneje opredelimo in ga ustrezno uvedemo v učno prakso. Nujno pa je tudi razmisliti, kako statistični jezikoslovni podatki pomagajo učitelju, da opredeli besedišče, ki naj ga smiselno povzame v pedagoškodidaktične svrhe. Študentje morajo v prvi vrsti usvojiti ustrezno besedišče, da razumejo pisna besedila in zvočne zapise vsebin svoje stroke, saj se pri študiju srečujejo s številnimi viri v angleščini. Jezikovne vpogleda mora učitelj nato še prilagoditi jezikovnim zahtevam študentov posamezne stroke na nadaljevalni ravni znanja angleščine (raven B1 ali višja), da bi jim kar najbolj učinkovito pomagal premostiti težave pri strokovnem izražanju lastnih pomenov v tujem jeziku.

Raziskovalni pristopi

Raziskovanje jezikovne rabe ob podpori korpusne metode izpostavi niz možnosti obdelave korpusnih podatkov, ker pa zbirke besedil ponudijo veliko količino besedilnih podatkov, moramo iz njih izluščiti tiste, ki so relevantni za usmeritev naše raziskave. Spletne strani ponujajo veliko računalniških orodji za pomoč pri korpusni jezikovni obdelavi. Pričujejoča raziskava se opira na naslednja orodja za raziskovanje in izpraševanje korpusnega gradiva: AntConc (Anthony 2019), Range Programme (Paul Nation 2006), Sketch Engine (Lexical Computing CZ s.r.o. 2020).

Za luščenje in primerjave ključnega strokovnega besedišča izkoriščamo tudi računalniške obdelave podatkov z računalniške platforme Sketch Engine, in sicer oddelek, ki vsebuje Britanski nacionalni korpus (BNC) in English Web 2015.

Utemeljitev izpeljujemo na osnovi analize jezikovnih podatkov, pridobljenih s primerjavami rab v dveh korpusih strokovnih besedil, natančneje v KK in KP, ki smo jih zbrali za namen te raziskave. Jezikovne podatke v obeh korpusih nadalje primerjamo z angleškimi referenčnimi korpusi BNC in Web 2015. Ob tem moramo opozoriti tudi na omejitve korpusnega pristopa: še tako velik korpus (najsodobnejši korpusi v angleščini obsegajo več sto milijonov besed) ne more predstaviti vseh možnosti rabe v živem jeziku, temveč daje le podatke o rabi v besedilih, ki jih korpus obsega. Kot poudarja Sinclair (1991: 38), je vsak korpus le droben vzorec rabe jezika, zato izsledki teh raziskav nikakor niso dokončni.

Sestava korpusa

Če naj nam korpus služi za pomenoslovno, leksikološko, terminološko analizo pa tudi skladenjsko, slogovno, besedilno in pragmatično raziskavo ter primerjavo osrednjega strokovnega upovedovanja na področju humanističnih in prometnih ved v angleščini, ima ključno vlogo sestava dveh usmerjenih korpusov besedil.

Osnovna struktura je pri obeh korpusih podobna. Razdeljena sta v oddelke in pododdelke glede na zvrst besedila in glede na strokovna področja in podpodročja. Vsak korpus te raziskave obsega približno milijon besed in vključuje besedila različnih avtorjev. Korpus sestavljajo večinoma pisna besedila, nekaj pa je tudi prepisov posnetkov govora. KK in KP vključujeta žanrsko raznovrstna besedila od poljudnoznanstvenih do znanstvenih člankov, kot prikazuje tabela 1.

Tabela 1: Sestava KK in KP po besedilnih zvrsteh

Stroka: kulturologija	Število dokumentov	Število besed	Stroka: prometne vede	Število dokumentov	Število besed
časopisni članki	85	275.543	časopisni članki	120	241.842
znanstveni članki	50	271.979	znanstveni članki	60	253.058
enciklopedični članki	55	286.813	enciklopedični članki	38	252.128
prepis zvočnega zapisa	30	165.735	uradni dokumenti	32	253.307
Korpus	220	1.000.070	Korpus	250	1.000.335

Tabela 2: Sestava KK in KP po strokovnih temah

Stroka: kulturologija	Število dokumentov	Število besed	Stroka: prometne vede	Število dokumentov	Število besed
Zgodovina	52	252.732	letalski promet	50	200.081
Antropologija	53	251.375	cestni promet	50	200.060
Mitologija	57	247.556	železniški prom.	50	200.158
Kulturologija	58	248.370	ladijski promet	50	200.019
			poštni promet /logistika	50	200.018
korpus	220	1.000.070	Korpus	250	1.000.335

Glede na zvrst besedil oba pisna korpusa sestavljajo po štirje približno enakovredni osnovni deli: znanstveni članki in deli strokovnih knjig, enciklopedični članki, članki iz poljudnih revij ali spletnih strani. KK vsebuje še oddelek prepisov zvočnih dokumentov, predvsem dokumentarnih filmov in radijskih oddaj, ki jih prištevamo v sklop govornih poljudnoznanstvenih besedil. KP pa kot četrto komponento obsega uradne dokumente, kot so prometni predpisi in določbe. Ta oddelek smo vključili, ker menimo, da nudi vpogled v pomemben segment besedišča, s katerim se bodo študentje srečevali pri svojem delu, pa tudi vsled pomanjkanja zvočnih zapisov, ki bi nam bili na razpolago za prometno stroko. Takšno sestavo korpusa utemeljujemo predvsem s pedagoškega vidika, saj prinaša študentom prilagojeno, postopno uvajanje polstrokovnega in strokovnega besedišča in struktur.

Korpusni dokumenti so označeni tako, da jih lahko pregledujemo po žanrih ali po strokovnih temah. Strokovne teme vsakega korpusa smo opredelili na osnovi lastnih izkušenj pri pouče-

vanju teh strokovnih jezikov in razdelav v učbenikih posameznih strok, kot prikazuje tabela 2.

Strokovna področja KK obsegajo besedila o zgodovini kultur in civilizacij, antropologiji, mitologiji (oziroma religiji), filozofiji (oz. estetiki) in o kulturologiji nasploh. Zadnja kategorija vsebuje predvsem besedila, ki združujejo tematske opise več področij, in jih je bilo zato težko razvrstiti v eno samo kategorijo.

KP obsega dokumente z naslednjimi temi: letalstvo, cestni promet, železniški promet, ladijski promet in poštni promet z logistiko. Kategorija »logistika« obsega besedila, v katerih se prepleta obravnava različnih področij prometa.

Gre torej za besedila, ki obsegajo razpon od popolnoma strokovnega/znanstvenega sloga do poljudnoznanstvenega. Oddelek s strokovnimi besedili tvorijo v obeh korpusih znanstveni in strokovni članki, objavljeni v revijah v tiskani ali elektronski obliki, in poglavja iz strokovnih knjig in učbenikov. V poljudnoznanstveni oddelek KK so uvrščeni enciklopedični članki o različnih humanističnih temah, medtem ko so

v oddelku s poljudnimi besedili zbrani članki iz različnih revij (npr. *National Geographic*, *Aviation Week*) in primerljivi članki s spletnih strani. Govorjena besedila so zapisi dokumentarnih oddaj, zabeleženih na programih *Discovery* in *National Geographic*, torej besedila, ki bi jih po zvrsti še najlažje uvrstili med poljudnoznanstvena, le da hkrati kažejo tudi značilnosti akademskih predstavitev.

Za potrebe raziskave sta oba korpusa uravnotežena kot celota in v posameznih pododdelkih. Dokumenti v besedilnih korpusih so označeni tako, da jih lahko primerjamo po temah in po besedilnih zvrsteh, ali pa besedilne korpuse kot celoto med seboj ali z referenčnimi korpusi.

Raziskovalna orodja korpusne analize: AntConc, Range, SketchEngine

Čeprav nekateri raziskovalci razvijejo svoja orodja za korpusno analizo (npr. Hyland in Tse 2007, Granger in Paquot 2015), se velika večina opira na številna brezplačna orodja, ki so dostopna na spletu. Ta raziskava pregleduje korpuse z brezplačnimi orodji AntConc (Anthony 2019), Range Programme (Paul Nation, 2006), Sketch Engine (Lexical Computing CZ s.r.o., 2020), saj vsako od teh orodji ponuja nekoliko drugačne vpogled, predvsem pa ne zahtevajo oblikoskladnega označevanja gradiva v korpusu.

AntConc omogoča enostavno gradnjo in pregledovanje neoznačenega korpusa v celoti ali po delih. Ponudi pregledovanje frekvenčne liste, konkordančnih nizov, večbesednih skupov, razpršenosti besedišča po dokumentih in list ključnih besed v primerjalnih korpusih. Range primerja besedila v korpusu s korpusom znanstvene angleščine, ki tvori osnovo za liste AWL (Coxhead, 2000). Izpostavi nam nabor ključnega znanstvenega besedišča v našem korpusu, saj ključne besede in besedne družine razporedi v frekvenčne liste po pogostosti in razpršenosti v korpusu. S tem računalniškim programom namreč razporedimo ves korpus v najmanj tri in največ osemnajst segmentov besedišča glede na pogostost pojavnosti v referenčnih korpusih, ki jih ta računalniški program vključuje in tvorijo

liste znanstvenega besedišča - AWL. Hkrati program izloči najpogostejših 2000 različnic v Korpusu znanstvene angleščine (Coxhead 2011), za katere pričakujemo, da jih študentje že poznajo, ker jih lahko opredelimo kot splošno besedišče. Posvetimo se torej bolj strokovnemu besedišču, specifičnemu za vsako stroko.

Sketch Engine dovoljuje enostavno gradnjo in izpraševanje lastnega korpusa, pri tem pa besedilno gradivo tudi samodejno lematizira (t. j. združi vse oblike pojavnosti v osnovno obliko besede) in besednovrstno označi. Podatke iz korpusa prikaže v kolokacijskih skicah posamezne besede z najpogostejšimi kolokatorji. Na osnovi primerjave korpusnega gradiva z referenčnimi korpusi sestavi tudi liste ključnega besedišča in večbesednih enot.

Orodja za analizo korpusov uporabljamo za luščenje osrednjega polstrokovnega in strokovnega besedišča, kar nam omogoči izdelavo usmerjenih učnih gradiv. Lahko pa korpuse vpletimo v učni proces tudi z neposrednim pristopom, kot samostojno učenje in raziskovanje besedil z rabo gornjih orodji.

Primerjava korpusov

Čeprav je korpusna jezikovna analiza vse bolj uveljavljen pristop k raziskovanju jezikov, k preverjanju hipotez o jeziku in k snovanju novih hipotez, pa Kilgarriř (2005) opozarja, da je sorazmerno malo primerjav korpusov. Korpusne raziskave namreč pokažejo, da podatki, ki jih dobimo v enem korpusu, niso zanesljivi za druge korpuse, saj je pomembno, kakšna besedila tvorijo korpus.

Pionirsko delo je na tem področju opravil Douglas Biber (1989, 1995, 1998) z razvojem metode večdimenzionalne analize jezikovnih sprememb v različnih zvrsteh besedil, različnih strokovnih jezikih oziroma za primerjanje različnih jezikov. V raziskavah Biber (1995) razvrsti opazovanje rabe jezika v več dimenzij, ki združujejo več jezikovnih značilnosti, s temi dimenzijami pa potem primerja korpuse v različnih jezikih ali korpuse različnih strok ali besedilnih zvrsti. Gre za poskus sistematizacije jezikovnih in be-

sedilnih značilnosti, kar privede do zelo kompleksnih statističnih primerjav. O primerjavi korpusov različnih strokovnih jezikov z namenom, da izluščijo ključno izrazje in strukture poročajo številni raziskovalci (npr. Crosthwaite in Cheung 2019, Granger in Paquot 2015, Nelson 2006). Ugotovijo, da tvorijo »ključno besedišče tiste besede, ki so nenavadno pogoste v določenem besedilu« (Scott in Tribble 2006: 36) v primerjavi z rabo v drugih besedilih ali referenčnih korpusih. Za namen te raziskave smo uporabili nekatere vpoglede iz gornjih analiz, saj prispevek poroča o vpogledih, ki jih da primerjava več dimenzij jezikovne rabe v KK in KP, kakor tudi v njihovih podkorpusih. Strokovne korpuse primerjamo tudi z angleškimi referenčnimi korpusi na spletni platformi Sketch Engine.

Razmerje med pojavnicami in različnicami v obeh strokovnih korpusih

Tabela 3: Razmerje med različnicami in pojavnicami

Korpus	Število pojavnic	Število različnic	Razmerje
KK	1.000.000	38.187	1: 26,19
KP	1.000.000	25.511	1: 39,20

Osnovno orodje, ki ga uporabljajo korpusne raziskave, so liste besed, saj nam te izpostavijo temeljne statistične podatke o besednem gradivu vsakega korpusa posebej, torej nabor vseh pojavnic (angl. *token*) v korpusu podeljenih v različnice (angl. *type*). V naši raziskavi analiziramo lematizirane korpuse, ki so označeni besednovrstno, torej so vse oblike pojavnic združene in štete kot osnovna oblika različnice. Seštevek pojavnic v korpusu nam pove, kako obsežno besedišče sestavlja posamezni korpus. Razmerje med različnicami in pojavnicami nam dovoljuje, da med seboj primerjamo različne korpuse in kaže leksikalno variabilnost besedišča oziroma leksikalno gostoto. Za večjo jasnost primerjave smo podatke v tabeli 3 normalizirali na povprečje v korpusu z 1.000.000 pojavnicami.

Primerjava izpostavi precejšnje razlike med korpusoma, zato lahko sklepamo, da je izrazje v KK bistveno bolj raznoliko od tistega v KP. Podatek je sicer le okvirjen, saj na število različnic močno vpliva sestava korpusa pa tudi njegova velikost. Hkrati nas velika razlika v razmerjih opozori, da se besede oziroma različnice v KP veliko pogosteje ponavljajo kot v KK. To bi lahko nakazovalo na večjo ustaljenost rabe terminologije in skladenjskih vzorcev. Nadalje nam razmerje kaže na višjo leksikalno gostoto v KK, kar bi pomenilo, da je pojavnost polnopomenskih besed v KK večja v primerjavi s slovničnopomenskimi besedami (Scott in Tribble 2006: 65). Ravno obraten pojav pa lahko predvidevamo v KP. Leksikalna gostota je pomembna, ko ocenjujemo berljivost besedil, saj z njo poskušamo ugotoviti delež polnopomenskega besedišča v besedilu ali korpusu. Lažje berljiva so namreč besedila z nižjo leksikalno gostoto, ker imajo manj polnopomenskih različnic. Ker sta KK in KP razvrščena v oddelke po besedilnih zvrsteh in bi žanrske razlike kaj lahko vplivale tudi na razmerje med različnicami in pojavnicami, si v tabelah 4 in 5 oglejmo še to primerjavo.

Tabela 4: Razmerje med različnicami in pojavnicami v oddelkih KK

KK	Število pojavnic	Število različnic	Razmerje
strokovna besedila	250.000	13.035	1: 19,18
poljudnoznanstvena besedila	250.000	13.550	1: 18,45
poljudna besedila	250.000	26.882	1: 9,30
govorjena besedila	250.000	18.939	1: 13,20

Čeprav se razmerja med različnicami in pojavnicami v oddelkih korpusov precej znižajo, kar lahko pripišemo predvsem manjšemu obsegu besedil, je očitno, da se besede najbolj ponavljajo v izrazito strokovnih besedilih, leksikalna gostota le-teh je torej najnižja. Če zadevo nekoliko poenostavimo, si lahko predstavljamo, da se vsa-

Tabela 5: Razmerje med različnicami in pojavnicami v oddelkih KP

KP	Število pojavnic	Število različnic	Razmerje
strokovna besedila	250.000	10.060	1: 24,85
pravna besedila	250.000	8.512	1: 29,37
poljudnoznanstvena besedila	250.000	13.463	1: 18,57
poljudna besedila	250.000	19.763	1: 12,65

ka različnica v uradovanih besedilih o prometu ponovi kar 30-krat, v znanstvenih člankih in strokovnih knjigah o prometu 25-krat, medtem ko je izrazje v strokovnih besedilih o kulturoloških temah bolj raznoliko, vsaka različnica naj bi se ponovila le 19-krat v celem podkorpusu. Poljudnoznanstvena besedila v obeh korpusih tvorijo v veliki meri besedila povzeta iz enciklopedij in očitno je, da sledijo žanrsko in stilistično podobnim vzorcem zapisa, saj je razmerje ponavljanja v obeh primerih 1: 18,5. Najbolj odprt nabor besednega gradiva opazimo v obeh korpusih pri publicističnih besedilih, torej v člankih iz revij in biltenov v elektronski obliki. Tako v KK kot v KP so vrednosti razmerja polovico nižje v publicističnih kot v strokovnih besedilih. Tudi govorena besedila, torej zapisi dokumentarnih oddaj o kulturoloških temah, se približujejo vrednostim razmerij v publicističnih besedilih. Primerjava s temi podatki nakazuje, da naša strokovna korpusa nikakor nista ozko usmerjena v strokovni jezik, ampak zajemata širok nabor izrazja okvirno usmerjenega v vsako stroko posebej. Zato se zdita ustrezna osrednjemu namenu raziskave, torej predvsem analizi rabe polstrokovnega besedišča navedenih strokovnih področjih.

Določanje ključnih besed s primerjavo korpusnih list besed

Izbor različnic, ki naj bi jih opredelili kot ključne besede posameznega strokovnega korpusa, se izkaže za zahtevno nalogo predvsem zato, ker nam vsa orodja za samodejno raziskavo korpusov dajejo zelo dolge spiske, nize tisoče besed, kar pa težko smiselno pregledujemo. Za luščenje ključ-

nega izrazja posameznega korpusa uporabimo torej različna računalniška orodja.

Računalniški program AntConc nam ponuja orodje KeyWords, ki izloči najpogostejše besede nekega korpusa tako, da primerja listi besed dveh korpusov. Podobno orodje ponuja tudi platforma Sketch Engine, ki ponudi tudi samodejno luščenje ključnih besed s primerjavo list besed pridobljenih iz korpusov BNC ali English Web 2015. Računalniški program Range lahko izdelava »profil leksikalne gostote«, saj razporedi ves korpus v najmanj tri in največ osemnajst segmentov besedišča glede na pogostost pojavnosti v referenčnih korpusih. Primerjava z listo besed AWL izloči najpogostejše različnice, ki jih lahko opredelimo kot splošno besedišče, da se posvetimo izrazju, specifičnemu za določeno stroko.

Tudi z gornjimi orodji pridobljene liste ključnih besed so obsežne, zato je tu prikazan le vzorec najpogostejših 100 različnic liste ključnih besed KK in KP. V obeh korpusih so v tem segmentu najpogostejše samostalniške različnice, kar gre pripisati visoki stopnji nominalizacije v strokovnih besedilih (Biber 1995). Kar nekaj je tudi pridevnikov, sorazmerno redki pa so glagoli, kot prikazuje tabela 6.

Tabela 6: Primerjava najpogostejših različnic nekaterih besednih vrst v obeh korpusih

korpus	samostalniške r.	pridevniške r.	glagolske r.	število različnic
KK	69	23	8	100
KP	80	11	9	100

Čeprav liste ključnih besed obeh korpusov tvorijo večinoma samostalniki, gre opozoriti tudi na pogosto dvoumnost teh različnic, saj najdemo v KK med njimi veliko besed, ki jih lahko rabimo kot samostalnik ali pridevnik ali kot glagol, na primer besede »evil«, »human«, »male«, »Roman«, »Christian«, »desire« ali »reign«. Pri razvrščanju v spodnje tabele smo upoštevali, kateri besedni vrsti različnica najpogosteje pripada v naših korpusih. Ključne samostalniške različnice v KK opisujejo življenje v

družbi, različne civilizacije in kulture preko preučevanja družbene ureditve, religije, obredov, zgodovine in tradicije:

Tabela 7: Ključne samostalniške različnice v KK

art, battle, being, behaviour, book, brother, century, church, civilization, clan, cult, culture, death, dynasty, earth, empire, ethos, faith, family, father, god, history, honor, house, human, king, knowledge, language, life, literature, man, male, marriage, mother, myth, mythology, people, philosophy, reign, reason, religion, ritual, sacrifice, scholars, society, son, story, sun, temple, theory, tomb, tradition, village, wife, woman, world

Številni so tudi samostalniki, ki poimenujejo zgodovinska ali geografska imena ali označujejo zgodovinska obdobja in različne vere: *Alexander, Buddhism, Buddhist, Christian, Christianity, Han, Islam, Jesus, Jews, Muslims, Ming, Rome, Romans*. Med pridevniškimi različnicami je na listi velik delež pridevnikov izpeljanih iz samostalnikov, ki poimenujejo krajevna imena, narodnosti ali označujejo različne vere: *ancient, Arabic, Aztec, Chinese, collective, cultural, divine, eastern, Egyptian, emotional, ethnic, evil, Greek, imperial, Islamic, Jewish, Macedonian, mimetic, native, orthodox, Persian, political, social*. Glagolov je v tem naboru le 8 in so pretežno v pretekli obliki: *believe, come, desire, find, know, live, reign, worship*. Med njimi so glagolske različnice, ki bi jih pravzaprav šteli v splošno besedišče, na primer: *come, find, know, live*. Hkrati pa lista ključnih glagolskih različnic nanaša take, ki potrjujejo osnovno usmeritev korpusa v opis družbenih dejavnosti, na primer: *believe, worship, reign*.

Povzamemo lahko, da segment najpogostejših ključnih različnic jasno kaže strokovne teme, ki jih KK vključuje, hkrati pa velik del tega besedišča ne moremo prištevati v strokovno terminologijo (npr. poimenovanje družinskih članov), temveč v splošno besedišče.

Tudi v naboru 100 ključnih besed KP tvorijo samostalniške različnice veliko večino, kar 80 % vseh različnic (Tabela 8).

Tabela 8: Ključne samostalniške različnice v KP

air, aircraft, airline, airport, article, automobile, aviation, bus, business, capacity, car, cargo, carriage, carrier, commission, company, congestion, container, contract, country, convention, damage, delivery, destination, directive, driver, distribution, engine, facility, flight, freight, fuel, goods, highway, industry, information, infrastructure, kilometre, liability, loss, luggage, mail, management, market, member, mode, navigation, network, operation, operator, passenger, population, port, price, project, rail, railroad, railway, rate, reason, requirement, road, route, rule, safety, sector, security, service, ship, shipping, speed, state, system, terminal, traffic, train, transport, transportation, user, vehicle

Naj opozorimo tudi na številne dvoumne različnice, saj najdemo med njimi veliko besed, ki jih lahko rabimo kot samostalnik ali kot glagol, kar je značilno za angleščino kot analitični jezik. Skoraj vse najpogostejše glagolske različnice na naši listi lahko rabimo tudi kot samostalnike: *transit, cost, post, charge, use, travel, measure, reason*.

Izrazito malo najdemo različnic, ki jih lahko opredelimo le kot glagole: *access, cost, charge, measure, operate, post, transit, travel, use*. Malo je tudi pridevnikov: *European, federal, financial, intermodal, international, maritime, national, postal, public, regional, uniform*. Povzamemo torej lahko, da lista ključnih besed KP opisuje ožje strokovno področje, da je njeno izrazje bolj specifično, kot tisto, ki smo ga izluščili v najpogostejših ključnih različnicah KK. V segment splošnega besedišča bi sicer lahko šteli različnice, ki označujejo nekatera prevozna sredstva (npr. *bus, car, ship, train*), vendar lahko pričakujemo, da v besedilih tega korpusa vstopajo v strokovno označene kolokacije. Oba korpusa med najpogostejših 100 ključnih besed vključita le eno skupno različnico in sicer samostalnik oziroma glagol *reason*. Zato bomo na straneh 58 in 59 podrobneje raziskali kolokacijsko umeščanje te besede z analizo konkordančnih nizov.

Razslojevanje besedišča od splošnega do polstrokovnega in specialističnega

Čeprav nam je analiza list ključnih besed posameznih korpusov dala zanimive in uporabne rezultate, opazimo, da nabor ključnih besed

Tabela 9: Statistika list besed KK

lista besed	Pojavnice	%	različnice	%	družine
1	766281	72,62	4190	11,15	999
2	103891	8,87	3809	9,25	983
3	40327	3,46	2437	6,65	938
4	30875	2,75	2161	5,73	887
5	17807	1,66	1742	4,72	826
6	12638	1,23	1351	3,77	749
7	9615	0,89	1102	3,16	701
8	6912	0,66	940	2,77	642
9	5988	0,59	730	2,50	619
10	3832	0,40	722	2,24	598
11	3956	0,41	670	2,12	584
12	2116	0,26	606	1,71	496
13	2983	0,33	608	1,72	514
14	1329	0,20	406	1,23	391
15	15050	1,63	2065	5,01	2065
16	123	0,01	21	0,05	4
ni na listi	45150	4,04	14636	36,22	????
skupaj	1000070		38187		9927

zajema velik del splošnega besedišča, oziroma besedišča, za katerega pričakujemo, da ga študentje na nadaljevalni stopnji tujega jezika že poznajo. Hkrati tudi opazimo, da je med visokopogostim besediščem le malo specialističnega strokovnega besedišča.

S programom Range porazdelimo besedišče celega korpusa avtomatsko po stopnjah od najpogostejšega, torej najsplošnejšega ali najbolj vsakdanjega, do različnih stopenj zahtevnosti ali strokovnosti. Vsaka od Rangevih list obsega po tisoč besednih družin, ki jih lahko opišemo kot termin z vsemi tvorjenkami. KK zajema kar 999 družin iz prve liste, kar pomeni, da vsebuje vsaj po eno obliko besede iz vsake od teh družin. To pomeni, da je delež splošnega ali najobičajnejšega besedišča v tem korpusu zelo obsežen, ker zajema 72,62 % vseh pojavnice, vendar le 11,15 % vseh različnic. Podobno velik delež različnic, 9,25 %, se uvršča v drugo listo besed, ki še vedno vsebuje

splošno besedišče, le manj običajno besede. Nato pa ujemanje besedišča v KK s tistim v posameznih listah postopno upada do šestnajste liste, ki vključuje osebna lastna imena. Veliko različnic, kar 36,22 %, pa se uvrsti še med besede, ki jih liste nimajo, torej v pretežno specialistično terminologijo kulturologije. Tabela 9 prikazuje statistično porazdelitev KK v vseh šestnajst besednih list in pa delež besedišča, ki ga nima nobena od list.

V KP ni tako popolnega ujemanja z besednimi družinami s prve liste, pač pa je delež ujemanja različnic še višji, kar 14 % vseh različnic v korpusu je s prve liste in kar 73,87 % vsega besednega gradiva. Višje kot v KK je tudi ujemanje z drugo, tretjo in četrto listo, potem pa so deleži ujemanja nižji. Sklepamo lahko, da je besedišče KP zgoščeno v bolj natančno definiranem segmentu izrazja, medtem ko je besedišče v KK razpršeno po celem spektru zahtevnosti izrazja. To dokazuje tudi visoko število besednih dru-

Tabela 10: Statistika list besed KP

Lista besed	Pojavnice	%	Različnice	%	Družine
1	741604	73,87	3909	14,00	990
2	110298	11,14	3015	11,30	956
3	34491	3,48	1812	7,16	865
4	30594	3,09	1571	6,34	797
5	16501	1,67	1255	4,76	660
6	8931	0,90	856	3,58	596
7	6467	0,65	619	2,61	481
8	3923	0,40	535	2,32	449
9	2573	0,26	381	1,80	361
10	2244	0,23	365	1,58	351
11	1552	0,16	258	1,35	312
12	1225	0,12	215	1,17	276
13	1276	0,13	150	0,92	225
14	636	0,06	92	0,68	172
15	10606	1,07	1333	5,81	1713
16	67	0,01	8	0,03	3
ni na listi	27347	2,76	9107	34,60	?????
skupaj	1000335		25511		7491

žin, kar 9927, ki jih zajema KK, medtem ko jih v KP naštejemo le 7491.

Do podobnih zaključkov so raziskave privedle tudi druge raziskovalce (npr. Granger in Paquot, 2009 in Hylandin Tse, 2007), ki iz teh podatkov izpeljujejo, da AWL nikakor ne predstavlja spiska najprimernejših besed za pedagoško uvajanje strokovnih jezikov nasploh, temveč lahko služijo le kot priporočila in usmeritve. Pričakujemo namreč lahko, da študentje na nadaljevalni stopnji poznajo približno 25 % nabora različnic v naših strokovnih korpusih, nadaljnjih 30 % ali 40 % gradiva pa moramo pozorno preučiti, ker vsebuje ključno strokovno terminologijo.

Program Range služi tudi za pregledovanje, katere oblike določene besedne družine se pojavljajo pogosto v posameznem korpusu. Tabela 10 prikazuje razčlenitev v vse elemente besedne družine dveh različnic s prve liste in sicer *organisation* in *man*. Obe različnici se namreč uvrstita

med prvih sto različnic na listi ključnih besed. V navedenem primeru primerjamo umestitev različnic v KK (frekvenca pojavnice 1) z različnicami v KP (frekvenca pojavnice 2).

Primerjava pokaže, da sta različnici *organisation* in *man* pogosti v obeh korpusih, vendar vsak od korpusov izpostavi pogostejšo rabo drugih oblik. V KK so izrazito pogostejše oblike *organise*, *organised*, *organisers* in *organising*, torej predvsem glagolske oblike, v KP pa oblike *organisation*, *organisational*, *organisationally*, *organisations*, torej samostalniške, pridevniške in prislovne oblike. Ravno obratno pa je z različnico *man*, ki se v KK pojavlja predvsem v samostalniških oblikah *man*, *manhood*, *mankind*, v KP pa v glagolskih oblikah *manned*, *manning*.

Tabela 11: Primerjava besednih oblik različnic »organisation« in »man« v KK in KP

Družine liste 1	Frekv. različnice	KK	KP
		Frekv. pojavnic 1	Frekv. pojavnic 2
organisation	284	17	267
organisational	10	0	10
organisationally	1	0	1
organisations	81	0	81
organise	10	2	8
organised	34	16	18
organiser	2	0	2
organisers	1	0	1
organising	6	2	4
organization	300	137	163
organizational	39	9	30
organizations	148	80	68
organize	38	29	9
organized	203	133	70
organizer	8	4	4
organizers	10	8	2
organizes	2	2	0
organizing	51	43	8
man	2299	2195	104
manhood	31	31	0
mankind	90	90	0
manliness	3	3	0
manly	11	9	2
manned	18	5	13
manning	10	3	7

Konkordančni nizi in besedne skice

Besedišče vsakega jezika je vpeto v mrežo povezav, te tvorijo besedilne vzorce glede na vlogo, ki jo besede imajo v različnih besednih zvezah ali leksikalnih enotah. Korpusni pristop nam omogoči samodejno odkrivanje pojmovnih razmerij med besedami z rudarjenjem besedil (angl. *text mining*) (Gorjanc in Vintar, 2007). Za namen

rudarjenja po naših korpusih nam služi konkordančni orodja AntConc, ki samodejno zbere konkordančne nize kolokacij, v katere vstopa ključna beseda, kot tudi platforma Sketch Engine, ki mrežo povezav ključne besede v določenem korpusu izoblikuje v besedno skico. Konkordančni nizi namreč zberejo vse primere rabe različnice kot konkordančnega jedra v korpusu, ki ga raziskujemo. Tako analiza nabora konkordančnih nizov omogoča, da prepoznamo tipične vzorce rabe, ki bi jih sicer lahko spregledali. Kot primer preučevanja rabe v konkordančnih nizih v tem oddelku preučujemo rabo različnice *reason*, ki jo najdemo v obeh korpusih tako pogosto, da se je uvrstila med prvi sto ključnih besed (glej tudi zgoraj str. 54). Hkrati je to beseda, ki jo računalniški program Range (glej tudi zgoraj str. 55) uvršča na prvo listo, torej v osnovno splošno besedišče.

Konkordančni označi besede v neposredni okolici konkordančnega jedra in s tem pomaga raziskovati pomen, ki jih ima beseda v rabi. S pregledovanjem izpostavljenega besedišča lažje opazimo vzorce ponavljajočih se kolokacijskih zvez, torej sobesedja, v katerega beseda pogosto vstopa v rabi. Vsaka kolokacijska zveza ali kolokacijski sklop se običajno navezuje na poseben pomen besede v konkordančnem jedru, čeprav je isti pomen besede lahko izražen v različnih kolokacijskih zvezah.

Beseda *reason* je v naših korpusih skoraj dosledno rabljena kot samostalnik, čeprav ima lahko tudi glagolski pomen. V besedilih KK namreč najdemo različnico *reason* rabljeno kot glagol petkrat, v KP dvakrat, kot prikazujejo avtentični primeri na tabelah 12 in 13.

Tabela 12: Primeri glagolske rabe glagola »to reason« v KK

Solomon is renowned for *his ability to reason with people*. Wise he was, but what is less occur. This brings *the ability to reason by hypothesis*. The child's logic is now concerned language could have served as a means of communication, *he reasons*, its first function must
From recognition of basic similarities *one might reason to* a sort of monotheism, by the light
though you are angry with Philip, *you might reason that*, to restore your honour and gain favor
the other deities and spirits. *They reasoned that* this concept of a creator came from metaph
„School teaches children,“ *reasoned Kay*, „that is why we send children to school.“ What
We can surmise how people *may have felt, reasoned or reacted*. But this requires a thorough „School teaches children,“ *reasoned Kay*, „that is why we send children to school.“ What
We can surmise how people *may have felt, reasoned or reacted*. But this requires a thorough

Tabela 13: Primer glagolske rabe glagola »to reason« v KP

information flows. Derived from Greek *logistikos (to reason logically)*, the word is polysemic
At this point people are unforgiving and beyond *being reasoned with*; their thoughts revolve

V vseh gornjih primerih je zastopan pomen, ki ga dvojezični Veliki angleško-slovenski slovar (v nadaljevanju VASS) razlaga kot:

Reason²

1. *vi* razmišljati (*about, of, on o*), modrovati, rezonirati, umovati, razumno ali logično misliti; sklepati (*from iz*); soditi, razsojati (*on o*); diskutirati, debatirati (*with z*);
2. *vt* diskutirati, debatirati, razpravljati (*o čem*), pretresati, razlagati; utemeljiti, motivirati; (logično) premisliti (često *up*); razumno, logično izraziti ali formulirati; sklepati, priti do sklepa; z argumenti koga odvrniti (*out od*); z razlogi napeljati, pregovoriti koga (*into k*); (VASS)

Tudi besedna skica izpostavi predvsem besedne skupe. V KK so to predložni skupi: *to reason with, to reason to, to reason that*; v KP pa so: *to reason with, to reason logically*. Veliko pogostejša je samostalniška raba različnice *reason*. V KK jo

izsledimo 381 krat, v KP pa 215 krat. Pri tako obsežnem številu primerov je smiselno najprej preučiti besedno skico, ki jo izluščimo iz vsakega korpusa, in šele nato pregledovati konkordančne nize, torej vlogo besede v širšem kontekstu.

Platforma Sketch Engine z orodjem Besedna skica izoblikuje listo kolokatorjev za vrsto skladijskih razmerij, v katere vstopa samostalniška različnica *reason* v posameznem korpusu. Našteje pridevniško rabljene besede, ki različnico *reason* določajo, glagole, katerih predmet ali osebik je, predvsem pa naniza veliko predložnih zvez oziroma večbesednih skupov.

Tabeli 14 in 15 povzemata podatke take besedne skice za KK in KP.

Tabela 14: Besedna skica za različnico »reason« v KK

pridevnik + reason	unknown, superstitious, good, uncertain, different, theoretical, practical, secondary, political, simple, deep, universal, fundamental, only, basic, various, important, historical
glagol + reason	suggest, relate, give, be, have, see
reason + glagol	be, have
večbesedni skupi z "reason"	For (just) this/that/the same reason For these/two/three/several/many/a variety of reasons For some (strange) reason For the (simple/very good) reason that For moral/political/practical/obscure etc. reasons Part of the/The / principal most important reason/s for There is /was no / a good reason for/why/ that There are two/many reasons to/for This / That /It is (perhaps/probably) / may be the reason why/ that One of the reasons why/for the/ a critique / standards/ arguments/ exercise/ power / limits/ virtue/ notion/ god/ age of reason

Tako pridevniške kot glagolske kolokacije različnice *reason* se razlikujejo v vsakem od korpusov in nanizajo drugačno pomensko mrežo tipičnih zvez za vsako strokovno ubesedovanje. Slovarska razlaga v VASS ustrezno opredeli samostalniško rabo različnice za večino naših primerov:

Tabela 15: Besedna skica za različnico »reason« v KP

pridevnik + reason
several, stated, likely, safety, structural, environmental, energy-saving, inane, prestige, confidentiality, compelling, trivial, evacuation, credible, legitimate, apparent, protection, practical, critical, economic, primary, administrative, principal, main, financial, political, personal, important, key, similar, security, technical, general, commercial, different, major, only,
glagol + reason
explain, state, advance, cite, overcome, explore, signal, list, specify, discuss, relate, set, say, determine, connect, provide, give, be, have, see
reason + glagol
hamper, invite, cite, be
večbesedni skupi z "reason"
For this/that/the same reason
For these/two/three/several/many/a variety of reasons
The main/ A/ One /Another reason for
Part of the reason for
There is no/a/good reason why/to/for
There are two/several reasons to/for
by reason of

Reason¹ n razlog, vzrok, povod, motiv; argument, utemeljitev; um, razum, razumnost, razsodnost, uvidevnost, razumevanje, logika; *jur* pravica (za kaj); glavni razlog; kar je prav in pošteno, upravičenost, usmerjenost; *log* premise nekega dokaza; sposobnost ustvarjanja zaključkov, sklepov na podlagi premis; (VASS)

Če pa upoštevamo vse kolokacijske zveze, ki se pojavijo v obeh korpusih več kot desetkrat, se izkaže, da samostalniška raba različnice *reason* najpogosteje tvori leve in desne kolokacijske zveze s slovničnopomenskimi besedami, torej večbesednimi skupi. Pomembno je tudi, da za različnico *reason* zelo pogosto najdemo ločilo, saj ima fraza kataforično vlogo, zato stoji na začetku povedi in ji sledi razlaga ali naštevanje »vzrokov, razlogov«.

V KK se ob množinskem konkordančnem jedru *reasons* nanizajo primeri podobne fraze, ki vključuje vrstne pridevnike, kot so *fundamental, important, moral, obscure, personal, political, practical, social, specific, superstitious, symbolic*. Te lahko tudi poudarimo s prislovi, kot so *rather, purely, very, some, no*. Za razliko od KK je

v besedilih o prometu beseda *reason* običajno rabljena v ozkem naboru fraz, te pa se najpogosteje ponavljajo v osnovni obliki in vključujejo predlog *for*. V primerjavi z rabami v KK se v fraze sorazmerno redko vrinejo dodatni pridevniki kot leve kolokacijske zveze ključne besede. Najpogosteje je na tem mestu rabljen pridevnik *several*, pa še pridevniki *economic, environmental, good, historical* in pridevniško rabljeni samostalniki, npr. *safety, security*.

Zanimivo je opazovati razlikovanje rabe v besedilnih zvrsteh. Ker taka raziskava presega namen tega prispevka, omenimo le, da v obeh korpusih največ primerov vseh teh rab v absolutnem in povprečnem seštevku najdemo v znanstvenih in upravnih besedilih, nato v govornih besedilih, sorazmerno manj v poljudnoznanstvenih in le redko v publicističnih besedilih. Zaključimo lahko, da gre za rabe, ki jih najpogosteje najdemo v znanstvenem jeziku, pisnem ali govornem. Presenetljivo je, da te najpogostejše rabe v obeh strokovnih korpusih ne zabeleži ne obsežni angleški slovar Collins Cobuild, ne VASS.

Edino v KK (Tabela 14) najdemo rabo samostalnika *reason*, ki je ustaljena zgolj v edninski obliki in jo ponazorijo fraze *the/ a critique/ standards/ arguments/ limits/ exercise/ power/ virtue/ notion of reason*, oziroma *the god/ age of reason*. Gre za fraze, kjer abstraktni samostalniki določajo lastnosti besede *reason*, fraze s samostalnikom *god* in *age* pa nakazujejo metaforični pomen. Ta pomen slovar VASS razlaga kot »um, razum«. Gre za pomen, ki je pogost le v strokovnih besedilih KK, saj ga tu najdemo kar triintridesetkrat v pisnih strokovnih besedilih in enkrat v govornih besedilih, medtem ko ga ne najdemo v KP. Tudi v KP (Tabela 15), oziroma v oddelku tega korpusa z uradovalnimi obvestili in pravnimi dokumenti, se pojavi specifično ubesedovanje, fraza, ki se zdi značilna le za to besedilno zvrst. Gre za frazo *by reason of*, ki jo v naših korpusih najdem le v enem oddelku, čeprav jo slovar Collins Cobuild uvršča med fraze, ki so v rabi v formalnem jeziku.

Gornji primeri korpusne raziskave ponazarjajo, kako konkordančnik pomaga pri rudarjenju besednih pomenov. Opozorijo, da opazovanje kolokacij različnice, ki bi se na prvi pogled zdela nezanimiva, ker se uvršča v splošno besedišče, naniza spekter pomenov od zelo pogostih do ozko usmerjenih, specifičnih le za določeno strokovno področje ali besedilno zvrst. Potrjuje tudi domnevo številnih korpusnih jezikoslovcev (npr. Grange in Paquot 2009, Prinsloo 2009), da je bodočnost korpusne rabe odvisna od tega, v kakšni meri bo besedno rudarjenje pripomoglo k celoviti in strnjeni predstavitvi vzorcev obnašanja posamezne besede v različnih kontekstih.

Razprava in zaključek

Raziskava strokovnih jezikovnih korpusov utemeljuje rabo korpusne metode kot pristopa za pridobivanje novih vpogledov o leksikalnih, skladijskih in besedilnih značilnostih iz strokovnega besednega gradiva. Ti naj bi nas vodili pri uresničevanju pedagoških ciljev pri pouku angleškega jezika posamezne stroke na nadaljevalni izobraževalni ravni. Prispevek prikazuje postopke zbiranja in raziskovanja dveh primerljivih strokovnih korpusov usmerjenih v humanistične (kulturologijo) in naravoslovne (promet) vede. Primerjava korpusov prikazuje poenostavljeno večdimenzionalno analizo jezikovnih sprememb v različnih zvrsteh besedil in različnih strokovnih jezikih, nato izpelje primerjavo list besed in ključnih besed in zaključni z opisom uporabe posameznih besed v obeh korpusih, pri čemer izpostavi razlike in podobnosti v kolokacijskih in sintagmatskih vzorcih rabe. Ob podpori računalniških programov AntConc, Range in Sketch Engine so se strokovni korpusi izkazali za ustrezno orodje za jezikovne analize, ki smo si jih zastavili. Omogočili so nam prepoznati pomembne razlike med rabo angleščine v različnih strokah in tudi opozorili na razlike pri upovedovanju med različnimi besedilnimi zvrstmi posamezne stroke. Prav ti vpogledi nakazujejo rešitev pogoste dileme učiteljev tujih strokovnih jezikov, ki se sprašujemo, koliko in kakšno strokovno terminologijo naj uvajamo

pri pouku. Pridobivanje jezikovnih podatkov s korpusno metodo nam namreč dovolj natančno izlušči ključne besede posameznih strok in sopojavnic. To nam pomaga osredotočiti pouk na segmente izrazja, ki so za študente različnih strok najbolj zanimivi, in hkrati potrjuje nujnost analize konkordančnih nizov in besednih skic, če naj mrežo besednih pomenov natančno opredelimo in zaznamo razlike rabe v posamezni stroki. Raziskava izpostavi velike razlike v naboru ključnega strokovnega besedišča med KK in KP. Hkrati analiza besedišča opozori, da se pri obravnavi tujih strokovnih jezikov ne moremo zanašati na splošne liste znanstvenega jezika, kot jih za angleščino ponuja AWL, saj bi zanemarili pretežni del strokovne terminologije v primeru kulturoloških in prometnih ved. Nadalje osvetli, kako različne stroke izrabljajo možnosti ubesedovanja, ki jih ponuja jezik. Ustaljena raba v avtentični komunikaciji namreč le delno izkoristi možnosti, ki jih ponuja jezikovni sistem, saj celo pri zelo pogostem splošnem besedišču (npr. različnica *reason*) izrabi ozek nabor leksikalnih ali skladijskih možnosti, s tem pa utrdi tipične rabe in pomene posamezne stroke. Korpusna metoda analize jezika torej omogoči, da postavimo poučevanje tujega jezika na raven, ki preplete učenje tujega jezika s stroki prilagojenim prevzemanjem ubesedovanja v tujem jeziku. Učitelj strokovnega jezika lahko uvide, ki jih ponuja korpusna metoda, uvaja v učni proces z neposrednim ali posrednim pristopom. Podatke iz korpusne analize uporabi za izdelavo novih učnih gradiv na osnovi avtentične rabe jezika v določeni stroki (npr. učenje in opazovanje besedišča in sopojavnic prilagojenih za izbirni tip nalog, alternativni tip nalog, dopolnjevanje in povezovanje delov povedi, določitev pravičnega besednega reda v povedi) ali vodi študente pri samostojnemu rudarjenju po strokovnih korpusih in zbiranju kolokacijskih in sintagmatskih vzorcev rabe. Šele korpusni prikazi jezikovne rabe namreč jasno pokaže mrežo kolokativnih vezi besed ter tako študentu ozaveštuje razlike med ubesedovanjem v maternem in tujem jeziku.

Povzetek

Prispevek utemeljuje rabo korpusne metode kot pedagoškodidaktičnega pristopa k uvajanju besedilnih značilnosti strokovnih jezikov na več jezikovnih ravneh in sicer predvsem pri ugotavljanju semantičnoleksikalnih značilnosti, pa tudi za analizo skladenjskih, stilističnih, pragmatičnih in žanrskih posebnosti. Teoretske uvide in predloge korpusnih jezikoslovcev poskuša predstaviti s primerom večplastne analize in primerjave dveh strokovnih korpusov z avtentičnimi besedili v angleščini, natančneje korpusa humanističnih ved (kulturologije) in korpusa naravoslovno-tehničnih ved (prometnih ved). Članek predstavi uporabo različnih računalniških orodij, ki so dostopna na spletu brezplačno, zato lahko služijo učitelju tujega strokovnega jezika za uvajanje novih pedagoškodidaktičnih pristopov, torej za neposredno ali posredno vključevanje korpusnih podatkov o dejanski avtentični rabi tujega jezika znotraj posamezne stroke, s tem pa tudi za ozaveščanje razlik med materinim in tujim jezikom pri ubesedovanju strokovnih vsebin.

Summary

The paper presents a corpus-based approach to dealing with LSP at different levels of language analysis, namely, in introducing typical lexicosemantic features and grammatical structures, as well as in discussing style, genre and register variation. The theoretical insights advanced by a number of corpus linguists are described as a multi-dimensional analysis by means of a comparison of two specialist corpora of authentic texts in English, i.e., a corpus of humanities (cultural studies) and a corpus of natural sciences and technology (transport and logistics). The study suggests a number of corpus-research tools, which are freely available on the Internet, and can help LSP teachers introduce direct or indirect corpus-based learning methods by researching authentic language use of various scientific domains. Such a teaching approach can help raise awareness with regard to the differences in wording of specialist topics between mother tongue and a foreign language.

Viri in literatura

Anthony, L. 2017. "Corpus Linguistics and Vocabulary: A Commentary on Four Studies." *Vocabulary Learning and*

Instruction 6, no. 2: 79–87. doi: 10.7820/vli.v06.2.Anthony

Bhatia, V. 2004. *Worlds of Written Discourse: A Genre - Based View*. London: Continuum.

Bhatia, V. K. 1993. *Analysing genre: language use in professional settings*. London: Longman.

Biber, D. 2006. *University Language*. Amsterdam: John Benjamins.

Biber, D., S. Conrad in R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: CUP.

Biber, D. 1995. *Dimensions of Register Variation – A Cross-Linguistic Comparison*. Cambridge: CUP.

Biber, D. in E. Finnegan. 1989. "Style of Stance in English: Lexical and Grammatical Marking of Evidentiality and Affect." *Text* 9: 93–124.

Bowker, L. in J. Pearson. 2002. *Working with Specialised Language. A practical guide to using corpora*. London in New York: Routledge.

Burnard, L. in T. McEnery. 2000. *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt am Main: Peter Lang.

Carter, R. in M. McCarthy. 2006. *Cambridge Grammar of English*. Cambridge: CUP.

Carter, R. 1998. *Vocabulary*. London in New York: Routledge.

Charles, M. 2007. "Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions." *Journal of English for Academic Purposes*, 6(4): 289–302.

Conrad, S. 2000. "Will corpus linguistics revolutionize grammar teaching in 21st century?" V S. Hunston in G. Thompson. (ur.), *2nd North American Symposium on Corpora and Language Teaching*, 43–57. North Carolina: North Carolina University Press.

Coxhead, A. 2011. "The Academic Word List Ten Years on: Research and Teaching." *Implications. TESOL Quarterly*, 45(2): 355–362.

- Coxhead, A. 2000. "A new Academic Word List." *TESOL Quarterly*, 34 (2): 213-38.
- Crosthwaite, P. in L. Cheung. 2019. *Learning the Language of Dentistry*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Dixon, R. M. W. 2005. *A Sematic Approach to English Grammar*. Oxford, New York: OUP.
- Flowerdew, L. 2013. "Needs Analysis and Curriculum Development in ESP." V B. Paltridge in S. Starfield (ur.) *Handbook of English for specific purposes*, 325-346. Oxford: Blackwell.
- Flowerdew, L. 2005. "An Integration of Corpus-based and Genre-based Approaches to Text Analysis in EAP/ESP: Countering Criticisms against Corpus-based Methodologies." *English for Specific Purposes*, 24: 321-332.
- Flowerdew, L. 2002. "Corpus-Based Analyses in EAP". V J. Flowerdew (ur.) *Academic Discourse*, 95-114. Harlow: Longman.
- Flowerdew, J. 2002. *Academic Discourse*. Harlow: Longman.
- Gavioli, L. 2005. *Exploring corpora for ESP learning*. Amsterdam: John Benjamins.
- Gorjanc, V. in Š. Vintar 2007. "Korpusna analiza vloge označevalcev medleksemskih razmerij v organizaciji besedila." *Jezik in slovnstvo*, 52(3-4): 117-129.
- Gojanc, V. 2005. *Uvod v korpusno jezikoslovje*. Domžale: Založba Izolit.
- Granger, S. in M. Paquot. 2015. "Electronic Lexicography Goes Local. Design and Structures of a Needs-driven Online Academic Writing Aid." *Lexicographica - International Annual for Lexicography / Internationales Jahrbuch für Lexikographie*, 31(1): 118-141.
- Granger, S., E. Dagneaux in F. Meunier. 2009. *International Corpus of Learner English. Version 2*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Granger, S., in M. Paquot. 2009. "In Search of a General Academic Vocabulary: A Corpus-driven Study." V K. Katsampoxaki-Hodgetts (ur.), *Options and Practices of LSP Practitioners*, 94-108. University of Crete Publications.
- Granger, S. 1998. *Learner English on Computer*. London in New York: Addison Wesley Longman.
- Halliday, M. A. K., W. Teubert, C. Yallop in A. Čermakova. 2004. *Lexicology and Corpus Linguistics*. London in New York: Continuum.
- Hockey, S. 2000. *Electronic Texts in the Humanities: Principles and Practice*. Oxford: Oxford University Press.
- Hunston, B. 2002. *Corpora in Applied Linguistics*. Cambridge: CUP.
- Hyland, K. in P. Tse. 2007. "Is there an »Academic Vocabulary«?" *TESOL QUARTERLY*, 41(2): 235-253.
- Hyland, K. 2006. "Disciplinary differences: language variation in academic discourses." V K. Hyland in M. Bondi (ur.), *Academic Discourse across Disciplines*, 32-57. Frankfurt: Peter Lang.
- Hyland, K. in M. Bondi. 2006. *Academic discourse across disciplines*. Frankfurt: Peter Lang.
- Hyland, K. 2004. *Genre and second language writing*. Ann Arbor, MI: University of Michigan Press.
- Jablonkai, R. R. in N. Čebroň. 2017. "Corpora as tools for self-driven learning." V N. A. Alias in J. E. Luaran (ur.). *Student-driven learning strategies for the 21st century classroom*, 274-298. Hershey, PA: IGI Global. <https://doi.org/10.4018/978-1-5225-1689-7.ch018>.
- Johns, T. 2002. "Data-driven learning: the perpetual challenge." V B. Kettemann in G. Marko (ur.). *Teaching and learning by doing corpus analysis*, 107-117. Amsterdam in Huston: Rodopi.
- Kilgariff, A. 2005. "Language is never, ever, ever, random." V *Corpus Linguistics and Language Theory*, 1(2): 263-275. Herndon, VA: Walter de Gruyter.

- Lewis, M. 2000. *Teaching Collocation: Further Developments in the Lexical Approach*. Hove: LTP.
- Lewis, M. 1997. *Implementing the Lexical Approach. Putting Theory into Practice*. Hove: LTP.
- Lewis, M. 1993. *The Lexical Approach. The State of ELT and a Way Forward*. Australia, Canada, Mexico, Singapore, Spain, UK: Thomson in Heinle.
- McEnery, T., R. Xiao in Y. Tono. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. Abingdon: Routledge.
- McEnery, T. in A. Wilson. 1996/2001. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Meyer, C. F. 2002. *English Corpus Linguistics: An introduction*. Cambridge: CUP
- Min, Z. 2013. "The Application of Corpus Tools in the Teaching of Discipline Specific Academic Vocabulary: A Case Study for Information Engineering Undergraduates." V *International Journal of Computer-Assisted Language Learning and Teaching*, 3(4): 33-47.
- Nation, P. 2006. "Learning vocabulary in another language." *Canadian Journal of Linguistics*, 14: 477-485. Toronto: University of Toronto Press Incorporated.
- Nation, P. 1990. *Teaching and learning vocabulary*. Boston: Heinle in Heinle.
- Nelson, M. 2006. "Semantic associations in Business English: A corpus-based analysis." *English for Specific Purposes*, 25: 217-23.
- Nesi, H. 2013. "ESP and corpus studies." V B. Paltridge in S. Starfield (ur.), *The Handbook of English for Specific Purposes*, 407-426. Oxford: Wiley-Blackwell.
- Oahey, D. 2002. "Formulaic language in English academic writing: A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines." V R. Reppen, S. M. Fitzmaurice in D. Biber (ur.), *Using Corpora to Explore Linguistic Variation*, 111-130. Amsterdam in Philadelphia: John Benjamins Publishing Company.
- Paltridge, B. in S. Starfield. 2013. *The Handbook of English for Specific Purposes*. Oxford: Wiley-Blackwell.
- Prinsloo, D. J. 2009. "The role of corpora in future dictionaries." V S. Nielsen in S. Tarp (ur.), *Lexicography in the 21st Century*, 181-206. Amsterdam: Benjamins.
- Reppen, R., S. M. Fitzmaurice in D. Biber. 2002. *Using Corpora to Explore Linguistic Variation*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Römer, U. 2011. Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31: 205-225.
- Schmitt, N. 2000. *Vocabulary in Language Teaching*. Cambridge: CUP.
- Scott, M. in C. Tribble. 2006. *Textual patterns: Keyword and corpus analysis in language education*. Amsterdam: Benjamins.
- Sinclair, J. 2004. *How to Use Corpora for Language Teaching*. Amsterdam: John Benjamins.
- Sinclair, J. 1991. *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.
- Sinclair, J. 1982. "Reflections on Computer Corpora in Linguistic Research." V S. Johansson (ur.) *Computer Corpora in English Language Research*, 1-6. Bergen: Norwegian Computing Centre for the Humanities.
- Stubbs, M. 2002. *Words and Phrases. Corpus Studies of Lexical Semantics*. Malden, Oxford: Blackwell Publishing.
- Tognini Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Widdowson, H. 1979. *Explorations in Applied Linguistics*. Oxford: OUP.
- Willis, D. 1990. *The Lexical Syllabus: A new Approach to Language Teaching*. London in Glasgow: Collins E.L.T.

Internetni viri

- Anthony, L. 2019. *AntConc (Version 3.5.8)* [računalniški program]. Tokyo, Japan: Waseda University. Dostop: <http://www.laurenceanthony.net/software/antconc>.
- Nation, P. 2006. *Range and Frequency*. Dostop: http://www.vuw.ac.nz/lals/staff/Paul_Nation.
- Sketch Engine*. 2003-2020. Lexical Computing Ltd. Brighton, UK. Dostop: <http://www.sketchengine.co.uk/>

Priročniki

- Grad, A., R. Škerlj in N. Vitorovič. 1997. *Veliki angleško-slovenski slovar*. (Elektronska izdaja). Ljubljana: DZS d.d.
- Sinclair, J. 1997. *English Language Dictionary*. (Elektronska izdaja). London in Glasgow: Collins.