

VLOGA BESEDILNEGA KORPUSA PRI POSTAVITVI GESELSKEGA ČLANKA V ENOJEZIČNEM SLOVARJU

Članek ponazori uporabo besedilnega korpusa pri postavitvi geselskega članka v enojezičnem slovarju. V uvodu opiše, kako korpusi prispevajo h kakovosti leksikografskega dela, v nadaljevanju pa se osredotoči na kognitivni profil, ki ga izdelamo s pomočjo analize korpusnih konkordanc. Iz kognitivnega profila je moč razbrati predvsem dva tipa podatkov: pomene iztočnice, ki so izpričani v korpusu, in slovnično-pomenske vzorce, ki se vežejo na posamezne pomene. Članek preveri, kako lahko tako pridobljene podatke uporabimo kot ogrodje za sestavo geselskega članka, in zaključuje s predvidevanji o (delni) avtomatizaciji postopka.

1 Uvod

Korpusi, zbirke besedil v elektronski obliki, izbranih po jasnih, vnaprej določenih kriterijih predvsem sociolingvistične narave, so se zaradi raziskovalnih možnosti, ki jih ponujajo, v zadnjih dveh desetletjih uveljavili v različnih jezikoslovnih disciplinah. S korpusi se je jezikoslovje v svojem raziskovalnem pristopu bolj kot kdajkoli prej približalo naravoslovnim znanostim, saj jezikoslovec na jezikovno gradivo, izluščeno iz korpusa, gleda kot na znanstvene podatke, ki jih je potrebno analizirati in obdelati, podobno kot bi fizik obdelal vrednosti, ki jih je izmeril pri znanstvenem poskusu. Zagovorniki korpusnega pristopa (npr. Gorjanc 1999: 50–51) trdijo, da le-ta v jezikoslovne raziskave vnaša večjo verodostojnost, saj analize jezika ne temeljijo več na intuiciji in jezikovni introspekciji idealnega domačega govorca, ampak na empirični analizi velike količine avtentičnih besedil. Prednost analize velikega števila dejanskih vzorcev jezikovne rabe naj bi bila predvsem v tem, da omogoča izpostavitve v jeziku tipičnega in globlje razumevanje celotnega delovanja jezika.

2 Korpusi v leksikografiji

Kljub uveljavitvi korpusnega pristopa v tako različnih jezikoslovnih disciplinah, kot so stilistika, prevodoslovje in metodika poučevanja tujih jezikov, pa lahko trdimo,

da so korpusi (vsaj veliki referenčni ali reprezentativni korpusi, ki jih za jezikoslovne raziskave uporabljamo zadnjih nekaj let) najtesneje povezani z leksikografijo. Pravzaprav so gradnjo vedno večjih korpusov, ki bi čimbolj natančno odražali dejansko stanje v jeziku, spodbujale ravno potrebe moderne (v začetku predvsem britanske) leksikografije, zlasti spremenjeno razumevanje jezikovne norme (norma postane tisto, kar je v jeziku tipično) in s tem pogojene zahteve po jezikovnih virih, ki bi temeljili na živem, dejansko uporabljanem jeziku in se ne bi zanašali na že uveljavljena spoznanja o jeziku. Slovarske založbe so torej finančno podprle gradnjo slovarskih megakorpusov (Kennedy 1998: 45), gigantov kot sta npr. Bank of English (BoE), ki je ob zadnji objavi januarja 2002 štel 450 milijonov besed, in British National Corpus (BNC, 100 milijonov besed). Tak korpus je tudi FIDA, korpus slovenskega jezika, v katerem so zbrana sodobna slovenska besedila v skupnem obsegu nekaj nad 100 milijonov besed.

Zbirke besedil v leksikografiji niso novost. Večina slovarjev, ki so nastali v predračunalniški dobi, se je opiralo na kartotečne listke (Krek 1999: 3), zbirke citatov, izpisanih iz besedil. Tudi te ogromne zbirke po abecednem redu urejenih citatov, ki so zrasle po slovarskih delavnicah (pri nas je bil to Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, ki še vedno hrani listkovno gradivo, uporabljeno pri pripravi SSKJ), so bile neke vrste korpusi in so bile ključnega pomena pri snovanju slovarjev, v dobi megakorpusov pa se njihova uporaba za slovarske namene zdi vprašljiva. Klasične zbirke citatov se od sodobnih slovarskih korpusov namreč razlikujejo v dveh ključnih točkah. Prva bistvena razlika je v načinu vključevanja citatov oz. besedil v zbirko. Za klasične citate težko rečemo, da so bili, vsaj s stališča sodobne leksikografije, izbrani po objektivnih merilih. Medtem ko si danes sestavljalci korpusov za cilj postavijo prikaz jezika v vseh njegovih pojavnih oblikah, od zelo pogovornih do literarnih in znanstvenih, so nekdaj izpisovalci citatov črpali iz virov, ki so po merilih prevladujoče jezikoslovne skupnosti veljali za jezikovno ustrezne. Tako pri svojem delu niso zajeli jezika v celoti, ampak le določen, običajno literarni, segment, v slovarju pa so se kot norma lahko pojavile sintagme, ki so bile v jeziku obrobne, izjemne, tiste najbolj običajne, tipične in vsakdanje pa so umanjale. Nasprotno pa slovarji, ki temeljijo na korpusni analizi, predvsem pedagoški slovarji za tuje govorce, težijo k izpostavitvi v jeziku tipičnega. Moderni leksikografski korpusi omogočajo ravno to, saj zagotavljajo objektivnost na začetni stopnji raziskovanja, ki ni bila dana pri kartotečnih listkih, tj. objektivnost pri izbiri jezikovnega gradiva za analizo. Osnovna zahteva pri slovarskih megakorpusih je namreč po uravnoveženosti. Korpus mora kot metonimija za celotno jezikovno produkcijo v določenem časovnem obdobju (Stabej 1998: 97) čimbolj natančno odražati dejansko stanje v določenem jeziku, kar dosežemo tako, da v korpus zajamemo količinsko in zvrstno ter besedilnovrstno uravnovežena besedila. Poleg tega je pri raziskovanju celotno jezikovno gradivo, ki je zajeto v korpus, dostopno kot celota. To pomeni, da pri analizi poljubnega leksema iz celotnega korpusnega gradiva ne izberemo le tistih citatov (konkordanc), ki se nam zdijo zanimivi, ampak vse citate, v katerih se ta leksem pojavi.

Druga sprememba, ki so jo v leksikografijo prinesli korpusi, pa je obrtniške narave, saj je povezana z načinom dela in orodji za analizo. Ogromna količina jezikovnega

gradiva, zbranega v korpusu, ne bi bila bistveno bolj koristna in uporabna kot zbirke citatov, če korpusi ne bi bili opremljeni z zmogljivimi računalniškimi orodji za iskanje. V tem je dodatna prednost elektronske oblike pred klasično: omogoča hitrejšo in učinkovitejšo raziskovanje. Leksikograf ima v vsakem trenutku možnost hitrega dostopa do jezikovnih sredstev v korpusnem gradivu. Poleg tega ni omejen le na iskanje po ciljni besedi (kot pri kartotečnih listkih), ampak lahko išče po različnih kriterijih, npr. po besedah, ki se nahajajo v bližnji okolici ciljne besede, kar je pri kartotečnih listkih nemogoče ali vsaj preveč zamudno, da bi bilo izvedljivo v praksi. Leksikograf tako na relativno lahek način pride do različnih podatkov o kontekstualnih omejitvah in pomenskih preferencah določenega leksema: ugotovi na primer, katere zveze so pogostejše in s tem boljši kandidati za posebno izpostavitve v geselskem članku.

Mogoče se zdi, da pri delu s korpusom vse naloge opravi računalnik sam. Prva stopnja raziskovanja je res delno avtomatizirana. Računalniška orodja »prečešajo« celoten korpus in izločijo ter sortirajo enote, ki ustrezajo danim kriterijem (ki jih je postavil raziskovalec). S tem je, kot že rečeno, zagotovljena objektivnost pri izbiri jezikovnega gradiva. Na drugi stopnji pa nastopi leksikograf, ki se na te podatke odzove. Pomembno je torej, kako podatke, izluščene iz korpusa, interpretiramo. Korpus kljub svoji neizpodbitni uporabnosti ni nezmožljiv – včasih tudi zavaja, na primer s korpusnim šumom, »ponavljajočimi se deli jezika, ki s slovarskega stališča neupravičeno dvigujejo pomembnost določenim leмам, npr. naslovi oddaj v radijskih ali televizijskih programih, rubrik v časopisih itd« (Krek 2003: 46).

3 Korpusni pristop

Ne da bi se pregloboko spustili v razpravo o korpusnem pristopu, pa moramo vendarle bolj natančno pojasniti, kako pojem razumemo, preden pokažemo, kako se korpusni pristop uresničuje v moderni leksikografiji. Korpusni pristop je krovni termin, ki pokriva dva temeljna metodološka principa. To sta delni korpusni pristop (*corpus-based approach*) in popolni korpusni pristop (*corpus-driven approach*) (Tognini Bonelli 2000: 1–2). Poglejmo najprej, kakšne so značilnosti prvega.

Delni korpusni pristop korpusa ne uporablja za odkrivanje novih spoznanj o jeziku, temveč predvsem kot dokazno gradivo, s katerim raziskovalec dokaže že obstoječe teorije ali najde primere za vnaprej postavljene jezikovne argumente. V okviru tega pristopa korpus ni sredstvo za postavljanje novih hipotez o jezikovni rabi, temveč le inventar primerov, iz katerega jezikoslovec pač izbere tiste, ki mu ustrezajo in ki njegovo tezo podpirajo, z ostalimi pa se niti ne ukvarja. To pa je lahko zavajajoče, saj je izbor gradiva subjektiven in ne odraža nujno dejanskega jezikovnega stanja. Nasprotno pa pri popolnem korpusnem pristopu teorija nastane **iz** korpusa, se pravi, da je korpusno inducirana. Raziskovalec nima vnaprej postavljene hipoteze in pristopi h korpusu kot celoti ter v svojem opisu zajame vse podatke na določeno temo:

Popolni korpusni pristop za razliko od delnega korpusnega pristopa predstavlja metodologijo, ki korpusa ne uporablja zgolj kot izbor primerov za dokazovanje jezikoslovnih argumentov ali vrednotenje teoretičnih izjav. Raziskovalec pri delu zajame celoto podat-

kov, pri opisu korpusnih dokazov na določeno temo pa se trudi biti čimbolj izčrpen in ne selektiven. Korpus ni skladišče primerov za potrditev vnaprej postavljenih teorij. Teoretične izjave, pa tudi komentarji in priporočila, izvirajo neposredno iz korpusnih dokazov in odslikavajo korpusno gradivo (Tognini Bonelli 2000: 2).

Hipoteza je torej posledica korpusne analize. Postavimo jo korak za korakom, ne na osnovi analize minimalnega besedilnega gradiva, temveč, kot smo že omenili, z opazovanjem celotnega korpusnega gradiva, ki ustreza iskalnim pogojem.

4 Vloga korpusa pri slovarski analizi leksema

Sestavljalec geselskega članka v enojezičnem slovarju ima pri slovaropisni analizi leksema pred očmi jasno načrtan cilj: identificirati mora pomene izbranega leksema, jih v geselskem članku razvrstiti po določenem zaporedju, sestaviti jasne, nedvoumne definicije in izbrati kar najbolj informativne primere rabe, s katerimi nakaže kontekstualne omejitve ali pomenske preference iztočnice ali/in izpostavi najbolj pogoste zveze. Vsak leksikografski vajenec, ki je že kdaj poskušal postaviti geselski članek, ve, kako zahtevna, pa vendar skrajno intelektualno stimulativna je taka naloga. Razdeliti pomensko polje leksema, določiti meje med pomeni in jih ukalupiti (tj. napisati definicije), izpostaviti ponavljajoče se strukture, iz obilja primerov rabe izbrati najbolj ilustrativne – teža te naloge jezikoslovca navda s ponižnostjo do jezika, hkrati pa ga sili v globinsko analizo jezikovne rabe in odkrivanje prefinjenih jezikovnih mehanizmov.

Prvi pogoj za uspešno slovaropisno analizo leksema je zagotovitev zadostne količine jezikovnega gradiva. V predračunalniški dobi je to pomenilo branje ogromnih količin besedil, ročno izpisovanje citatov in urejanje le-teh po abecedi, kar je bilo garaško in izredno zamudno delo. Prvo izdajo OED so npr. pripravljali več kot 70 let (Winchester 1999), od l. 1859, ko so člani Londonskega filološkega društva na enem svojih rednih srečanj uradno začeli s projektom, do l. 1928, ko je izšel zadnji, 12 zvezek. Monumentalno delo (ki je prvič izšlo še pod imenom *New English Dictionary*) je vsebovalo 414 825 definicij in 1 827 306 primerov rabe, s katerimi so sestavljalci želeli ilustrirati prav vsak pomen in – glede na to, da je bil OED diahroni slovar – pokazati predvsem to, kako so se pomeni skozi stoletja spreminjali. Sestavljalci tega cilja gotovo ne bi mogli uresničiti brez prave armade neplačanih prostovoljcev, jezikovnih navdušencev, ki so po navodilih dolgoletnega glavnega urednika, Jamesa Murrayja, brali angleška (in v manjši meri ameriška) literarna dela, časopise in revije, izpisovali citate in jih pošiljali v t. i. skriptorij, kjer so listke sortirali po abecedi, besedni vrsti in pomenu in jih tako pripravili za slovarsko analizo.

OED in ostala temeljna leksikografska dela (tudi SSKJ), ki so nastajala v podobnih pogojih, so brez dvoma izjemni in za sodobnega leksikografa skorajda nepojmljivi dosežki. Ravno ob zgodbah, kot je zgodba o nastajanju OED, se pravzaprav najbolje zavemo, kakšno pridobitev za leksikografijo predstavljajo besedilni korpusi. Zajemanje in opisovanje besedišča je s pomočjo računalniškega korpusa neprimereno lažje in hitrejšo. Korpus je za študij besede v okolju naravnost idealen, kar se

hitro pokaže, če pod drobnogled vzamemo eno temeljnih analitičnih metod, ki so se uveljavile v praktičnem slovaropisju, to so konkordance, se pravi korpusne pojavitve izbrane besedne oblike. Pri delu najpogosteje uporabljamo konkordančne vrstice (KWIC concordances¹; npr. McCarthy 1990: 67, Sinclair 1991: 32–33), kjer je za vsako pojavitev opazovanega leksema v korpusu izpisana vrstica besedila, v katerem se ta leksem pojavi, in sicer tako, da je opazovani leksem (iztočnica) izpisan na sredini vrstice, levo in desno od tega leksema pa tako imenovani minimalni kontekst, npr.:

he, he, he, ste se vsi *izkašljali*? Kaj je? Ste me res odpisali?
 rob. Zaradi sabotaže se bolje počutijo: »se *izkašljajo*«. Takšna dejanja so ponavadi spontana in nenačrtovana.
 je ostal isti. V tistem se je profesorju uspelo *izkašljati*, pljunil je v robec, ki ga je držal
 potem sem *izkašljal* čigumi v dlan in ga, ne da bi vedel
 ačkinih pljučih. Okužena mačka ga pogosto *izkašlja* in se ga tako znebi, vendar so na voljo
 je bil Rupsi še župan, je vedno brez težav *izkašljal* stotaka. Ko so ga za kak praznik prosili za
 ki so na omenjeno linijo že poklicali in se *izkašljali*. Izkazalo se je, da so bile njihove domneve

Konkordančne vrstice leksema izkašljati iz korpusa Fida

Konkordančne vrstice lahko abecedno urejamo po besedah, ki se pojavljajo levo ali desno od iztočnice. Ta dokaj preprost in hiter postopek nam omogoča vpogled v frazeološko obnašanje iztočnice. Pri bolj poglobljeni analizi pomenov in rabe iztočnice pa si pomagamo s širšim sobesedilom in bolj usmerjenim iskanjem. Konkordance resda ponujajo le minimalni kontekst (kar se izpiše na računalniški monitor), kar vedno ne zadostuje. Večina korpusov (tudi FIDA) ima zato orodja, s katerimi s klikom na izbrano vrstico prikličemo več sobesedila. To je potrebno predvsem takrat, ko moramo razjasniti različne dvournosti (npr. reference, elipse ipd.).

Na osnovi konkordanc (ali daljših besedilnih segmentov, kjer je to potrebno) slovaropisec določi pomen izbranega leksema. Če je zagovornik popolnega korpusnega pristopa, pomen določi, ne da bi prej pogledal, kako so ta leksem obravnavali v že obstoječih slovarjih. Prej omenjene »že obstoječe teorije« so v slovaropisju torej slovarske obravnave določenega leksema v obstoječih slovarjih, v slovenskem okolju je to predvsem SSKJ.

Že v prvem delu članka smo nakazali, da lahko iz korpusa izluščimo vrsto zanimivih podatkov. Predvsem britanska leksikografija si je za potrebe EFL slovarjev (pedagoških slovarjev za tuje govorce) vedno prizadevala najti in v slovarjih izpostaviti tipične zveze, v katerih se leksemi pojavljajo, se pravi kolokacije. Leksikografe je torej že pred pojavom korpusov zanimalo, kaj se pojavlja neposredno pred in za iztočnico, korpus pa je omogočil še globljo analizo tipičnega ubesediteljenja. Korpusna analiza se ne ustavi pri dvo- ali trobesednih segmentih, ampak išče daljše enote in manj očitne regularnosti. Raziskovalno polje se torej z dvo- ali trobesedne kolokacije razširi na vzorce (Hunston, Francis in Manning 1997). Značilno za vzorce je, da niso popolnoma slovnične niti popolnoma pomenske enote, ampak so rezultat prepletanja slovničnih in pomenskih preferenc določenega pomena, nekakšen slovnično-leksikalni konglomerat. Hunston, Francis in Manning

¹ KWIC je kratica za *key word in context*, se pravi iztočnica v kontekstu.

(1997), ki so delali s takrat 250 milijonskim BoE, ugotavljajo, da ima vsak leksem enega ali več vzorcev. Za lekseme z več pomeni je tipično, da ima vsak pomen svoj vzorec in ga ta vzorec pravzaprav identificira (vzorec služi kot indikator pomena).

Pomeni leksemov torej vstopajo v zapletena slovnično-pomenska razmerja, produkti teh procesov pa so večdelni konstrukti, vzorci, ki jih razumemo kot funkcionalne enote in so kot take na voljo domačemu govorcu pri procesu verbalizacije. Zato se pri analizi konkordanc ne osredotočimo le na idiomatično jedro – kolokacijo – ampak na regularnosti (ponavljajoče se enote) v širšem okolju. Za ilustracijo smo po principu popolnega korpusnega pristopa analizirali konkordance leme *izkašljati* iz korpusa FIDA (konkordance so v celoti izpisane v Prilogi) in sestavili kognitivni profil (*cognitive profile*; Hanks 2000) leksema *izkašljati*, delno formaliziran zapis pomenov leksema in vzorcev, ki se vežejo na posamezne pomene.

4.1 Korpusno inducirani kognitivni profil leksema *izkašljati*²

Število zadetkov za lemo *izkašljati* v korpusu Fida (po izločitvi enakih): 47

1. 'izločiti s kašljem'

I. vzorec: *izkašljati se* (4 zadetki)

- ✗ je ostal isti. V tistem se je profesorju uspelo *izkašljati*, pljunil je v robec, ki ga je držal
- ✗ anatomske razmere in nesposobnost malih otrok, da bi se *izkašljali* in izpraznili nos in žrelo. Pogoji za zdrava ušesa
- ✗ (se lahko včasih brez vsakega strahu, da me bodo pod pretvezo nevarnosti onečiščenja avtobusa z bruhanjem izkrcali na kakšni obskurni postajici), mirno in debelo *izkašljam* ter sem pa tja lagodno in globoko rignem. Šoferji
- ✗ otrokovega življenja, ker se tak otrok že itak slabo *izkašlja*. Sicer pa velja pravilo, da skušamo suh kašelj

II. vzorec: *izkašljati* + T³ [izloček, sluz, prah, predmet, voda] (17 zadetkov)

- ✗ Potem sem *izkašljal* čigumi v dlan in ga, ne da bi vedel
- ✗ in dozori v mačkinih pljučih. Okužena mačka ga pogosto *izkašlja* in se ga tako znebi, vendar so na voljo (ga = pljučnega črva)
- ✗ izločka, ki postane manj lepljiv in ga je lažje *izkašljati*. Pogosto jih uporabljamo v inhalacijah.
- ✗ ga spremljajo dodatni simptomi, npr. vročina ali krvav *izkašljani*⁴ izloček.

² V kognitivnem profilu so podčrtani izrazi, ki nakazujejo regularnosti v širšem okolju iztočnice; glej poglavje 4.2.

³ Predmet v 4. sklonu.

⁴ *Izkašljan* analiziramo kot pretekli deležnik, se pravi kot glagolsko obliko, zato ga iz kognitivnega profila ne izločimo.

- ✕ (Okužimo se) z vdihanjem bacilov tuberkuloze, ki jih *izkašlja* bolnik z aktivno boleznijo. Širi se lahko tudi z
- ✕ utekočinjenje in sprostitvev sluzi, ki jo bolnik tako lažje *izkašlja*. V hujših primerih mora bolnik ostati v bolnišnici,
- ✕ sapniku ali pa mehčajo sluz, da jo bolnik lažje *izkašlja*. Kašelj praviloma pomirjamo le ponoči, da se bolnik
- ✕ Otroku prigovarjajte, naj predmet *izkašlja*.
- ✕ Če otrok predmeta ne *izkašlja*, ga položite čez koleno in nekajkrat udarite po predelu
- ✕ otrok je sposoben sluz, ki mu zaliva dihala, *izkašljati* in izpljuniti, dojenček pa vso to nesnago požira,
- ✕ ; nato prerežejo popkovino ter ga očistijo krvi. Dojenček *izkašlja* vodo, ki se mu je nabrala v pljučih,
- ✕ glavo navzdol ter vas udaril po zadnjici, da bi *izkašljali* vodo, ki se vam je nabrala v pljučih.
- ✕ dojenček rodi, mora začeti tudi dihati. Potem ko *izkašlja* vodo, ki se mu je nabrala v pljučih
- ✕ (Odspotikala se je prek različne ropotije, ki) so jo ljudje odlagali v priročni grajski deponiji, in *izkašljala* dober kilogram prahu. Knedlina torba in vrečka s špecerijo
- ✕ (potegnil dolg dim iz velikanske cigarete in ga dolgo dolgo držal v pljučih, hoteč, da se mi možgani do zadnjega atoma napolnijo s tehacejem), se dolgo dolgo zadrževal, da nisem vse skupaj *izkašljal* in potem sem bil do amena zmuden. Konfuzen
- ✕ (temna megla se dviga dalje in ko pride do grla, se začnem) daviti in dušiti, čutim silno potrebo, da bi *izkašljala* vse to. Uboga Milena, ki mi sledi
- ✕ Ko prekriža noge in se vživlja v svojo vlogo, *izkašljata* fotelja izpod rjuh veliko prahu. Ko je končno le

2. 'izročiti denar'

I. vzorec: *izkašljati* + T (4 zadetki)

- ✕ je bil Rupsi še župan, je vedno brez težav *izkašljaj* stotaka. Ko so ga za kak praznik prosili za
- ✕ (občinstvo je videti nekam nemo, vendar pa na koncu simpatičen tip v črnem, Gogi ga očitno dobro pozna,) saj mu prijateljsko pravi »gospod profesor«, le *izkašlja* zahtevano vsoto. Sledi dobrodušen aplavz, jaz pa se
- ✕ homoseksualnosti, na koncu zmaga prvi, ki mora zanjo *izkašljati* 2000 tolarjev. Cena? Sitnica!« Potem (zanjo = za fotografijo)

- ✘ (Rene grozi, da bo javnosti postregel z) umazanimi podrobnostmi iz življenja Jacksonov, če Janet ne bo *izkašljala velikodušne odpravnine*. Rene, ki je bil z Janet

II. vzorec: izkašljati se (1 zadetek)

- ✘ (In če bogati ne sodelujejo v igri), jih ponavadi tako ali drugače prisilijo, da se *izkašljajo*.

3. 'povedati svoje mnenje'

vzorec: izkašljati se (7 zadetkov)

- ✘ (dodali so še zapise mnenj ljudi), ki so na omenjeno linijo že poklicali in se *izkašljali*. Izkazalo se je, da so bile njihove domneve
- ✘ (si je pridelal veliko izkušenj. Ve, kaj mori) ljudi, ve, kako jih zagrabit, da se *izkašljajo*, zna speljati pozornost na sončnejše plati senčnega življenja,
- ✘ (Pred kar) zavidljivim številom znanih političnih osebnosti se je kot prvi govorac *izkašljal* politolog Toni Kuzmani. Zadržal se je pri trendu izginjanja (politike iz javnega življenja)
- ✘ (Saj sem rekla, da se morata pogovoriti, potem bo) bolje. Grem stavit, da sta se oba dobro *izkašljala*. Če bi eden od njiju odšel od doma,
- ✘ (Lahko jo nekam pošljem, in takoj bo ocenila, ali me mora poslati nazaj ali pa naj bo tiho. Je) pravi amortizer. Ko si v krizi, se moraš *izkašljati*. Težko je, če si prepuščen samo tujim ljudem
- ✘ (Gre) za taktiko v slogu: vržeš bombico, sindikati se *izkašljajo*, vlada preveri javno mnenje, in če to ni (ugodno, ukrepom obrusi najostrejša robove)
- ✘ kot napake nekoga drugega. V nedeljo se bodo ljudje *izkašljali*, potem pa bo že na prihodnji seji določil le (tiste občine, ki bodo lahko preživele)

4. 'dati informacije'

vzorec: izkašljati se (6 zadetkov)

- ✘ (če nočeš nazaj v panamsko keho, od koder si prišel, in to na podlagi obtožb, da) z neznanci snuje, kar pač snuje, se takoj *izkašljaj*.
- ✘ kar pa snuje, se takoj *izkašljaj*. Če se izkašljaš, te čaka vreča denarja, če pa ne,
- ✘ Za koga dela? Na dan z imeni! *Izkašljaj se!*
- ✘ do njega. Zašepetal je: Tipa sta se *izkašljala*. Rekla sta, da je prodajala svojo pičko, (kadar je bila plonk)
- ✘ v kateri je stanoval. Eden od njih se je *izkašljal* policajem, da bi umora ne obesili njim. Vasquez
- ✘ (in pijančkom obljubljal, da jih ne bom premikastil, če) mi bodo postregli s pravimi imeni. Če se niso *izkašljali*, sem jih aretiral, če so se, sem (jih vseeno aretiral)

5. 'grajati'

vzorec: izkašljati se + na + T (2 zadetka)

- ✗ (to je bila prava groza) za Kistiakowskega, na katerega so se zdaj kar po vrsti *izkašljali* Oppenheimer, general Groves, Vannevar Bush in tako naprej (po položaju)
- ✗ (Je med igralci prišlo do nekih trenj in dvakrat zapored smo izgubili. Glavni trener nas) je povabil na večerjo. Že dobro nadelani smo se *izkašljali* drug na drugega – do dna. Zjutraj smo bili

6. 'izmisлити si'

vzorec: izkašljati + T (2 zadetka)

- ✗ značilnosti: njihova imena so pristala v dnevnem časopisju in *izkašljali* so alibi-je, ki so jih izločili iz kroga osumljencev
- ✗ naložila v prejšnjem pismu! Če ne bom zdaj hitro *izkašljala* prepričljive razlage, mi bo Steve prišel na sled in (potem se bo usulo)

7. 's težavo povedati' (1 zadetek)

- ✗ Prosim?! sem *izkašljaj* in ga debelo gledal. Mislim, iskrenost pa taka

8. Neanalizirano (3 zadetki)

- ✗ He, he, he, ste se vsi *izkašljali*? Kaj je? Ste me res odpisali?
- ✗ rob. Zaradi sabotaže se bolje počutijo: »se *izkašljajo*«. Takšna dejanja so ponavadi spontana in nenačrtovana.
- ✗ od dobre volje. Jasno, z odgovorom sem bil *izkašljaj* tudi iver. Srečal sem dva, tri znanke in

4.2 Od kognitivnega profila do geselskega članka

Na osnovi analize konkordanc leme *izkašljati* iz korpusa FIDA (48 zadetkov) smo pomenski prostor glagola *izkašljati* razdelili na 7 kategorij. Pogosto lahko že na osnovi analize minimalnega konteksta preciziramo pomen, ki je izpričan v konkordanci, npr. *Dojenček izkašlja vodo, ki se mu je nabrala v pljučih*, kjer je popolnoma jasno, da gre za pomen 'izločiti s kašljem'. Nekatere konkordance pa so glede interpretacije bolj odprte in v takih primerih potrebujemo več sobesedila. V konkordancah v kognitivnem profilu so deli besedila, ki jih prvotno ni bilo na monitorju in smo jih dodali iz razširjenega sobesedila, v oklepaju. Težko rečemo, ali je *izkašljati* v tem pogledu tipičen ali izjemen – posploševanja bodo mogoča šele, ko bo na voljo večje število kognitivnih profilov. Predvidevamo pa, da se več dvoumnosti pojavi pri leksemih, kjer so pomeni slabo ločljivi in delno prekrivni. Tipičen primer dvoumnosti je naslednji segment:

... se dolgo dolgo zadrževal, da nisem vse skupaj izkašljaj in potem sem bil do amena zmeden. Konfuzen.

V gornji konkordanci ni jasno, na kaj se nanaša zaimek *vse*, zato ne vemo, ali je *izkašljati* rabljen v konkretnem pomenu 'izločiti s kašljem' ali figurativnem pomenu 'dati informacije'. Ko imamo na voljo več sobesedila, lahko razvežemo navezovalno razmerje (*vse* = *dim*), in ugotovimo, da je *izkašljati* rabljen v konkretnem pomenu, torej 'izločiti s kašljem'.

(potegnil dolg dim iz velikanske cigarete in ga dolgo dolgo držal v pljučih, hoteč, da se mi možgani do zadnjega atoma napolnijo s tehacejem), se dolgo dolgo zadrževal, da nisem vse skupaj izkašljal in potem sem bil do amena zmeden. Konfuzen

Podobno besedilno ambigviteto najdemo tudi v naslednjem izseku, kjer ne poznamo nanašalnice zaimka *ga*.

... in dozori v mačkinih pljučih. Okužena mačka *ga* pogosto izkašlja in se *ga* tako znebi, vendar so na voljo

Ta kratek besedilni segment vsebuje precej signalov, ki kažejo na to, da je *izkašljati* rabljen v konkretnem pomenu in da se *ga* nanaša na prenašalca bolezni, vendar je navezovalno razmerje kljub temu smiselno razvezati, saj je lahko nanašalnica kolokator ali pa pomaga zožiti pomensko polje predmetov, ki se vežejo na ta pomen glagola *izkašljati*.

Od sedmih pomenov leksema *izkašljati*, ki smo jih določili s pomočjo korpusne analize, je najpogostejši konkretni pomen 'izločiti s kašljem', ostali pa so bolj ali manj metaforične izpeljave iz tega osrednjega pomena. Za vsak pomen smo navedli njegove vezljivostne pogoje – pomensko-slovnične vzorce, ponavljajoče se strukture, ki jih opazimo v konkordancah. Pri vsakem vzorcu smo kot ilustrativno gradivo navedli vse konkretne manifestacije tega vzorca iz pregledanih konkordanc. Pojavljata se v glavnem dva vzorca: refleksivna raba *izkašljati se* in povezava *izkašljati* s predmetom v tožilniku, in sicer obe v konkretni in figurativni rabi. Posebej velja izpostaviti vzorec *izkašljati se na koga*, ki je vezan izključno na pomen 'grajati'. Razmerja 1:1 so zelo zanimiva, saj tu oblika služi kot indikator pomena.

Z natančnejšo analizo konkordanc pa lahko odkrijemo še bolj subtilne jezikovne mehanizme. Poglejmo torej, katere vzorce smo odkrili pri posameznih pomenih leksema *izkašljati*.

Pri prvem pomenu (izločiti s kašljem) se s precej visoko frekventnostjo pojavljajo izrazi z modalnim nabojem, za katere lahko v grobem rečemo, da spadajo v pomensko polje 'sposobnost' (modalni izrazi so v kognitivnem profilu podčrtani). Ti izrazi kažejo na to, da mora agens pri izkašljevanju premagati določeno oviro oz. da za izkašljevanje porabi energijo. Če pogledamo, kaj običajno izkašljamo, se kot predmet v tožilniku pojavi *izloček*, *bacil*, *pljučni črv*, *sluz*, *prah* itd. Tem leksemom težko dodelimo status kolokatorja, saj bi bilo za to potrebnih več sopojavitev z jedrom kolokacije (za vsak predmet imamo samo po eno). Lahko pa omejimo pomensko polje – izkašljamo kaj slabega, kar je običajno povezano z boleznijo. Zanimiva sta predmeta *voda* (ki se pojavi trikrat) in *predmet* (ki se pojavi dvakrat). Na prvi pogled se zdi, da sta to najpogostejša kolokatorja, vendar nas tu korpus

zavaja. Če pogledamo distribucijo zveze izkašljati + voda, vidimo, da so vse tri konkordance, kjer se pojavi *voda*, iz istega besedila in isto velja za *predmet*. Če se zveza večkrat pojavi v istem besedilu, pa ima kolokacijsko manjšo težo, saj gre lahko za idiolekt enega avtorja, kot če ima širšo distribucijo in se pojavlja v različnih besedilih različnih avtorjev.

Zanimivi sta konkordanci 18 in 29, ki predstavljata metaforični odmik od vzorca. V 18. konkordanci se metaforičnost kaže v agensu, ki je netipičen, ker je neživ (fotelj). Tu gre torej za prenos lastnosti živega subjekta (sposobnost izkašljevanja) na neživo naravo. V 29. konkordanci pa predmet, ki ga agens izkašlja, ni nekaj konkretnega (kot v vseh drugih primerih). Ohrani pa negativni naboj, tipičen za »izkašljanca« – *temna megla, ki davi in duši*, je gotovo nekaj slabega. V 18. in 29. konkordanci imamo torej prenos s konkretnega v bolj abstraktna stanja, kar je v jeziku eden najbolj tipičnih mehanizmov za ustvarjanje novih pomenov. Zaradi svoje inovativnosti in kreativnosti bi bili 18. in 29. konkordanca slaba primera rabe za ilustracijo pomena *izločiti s kašljem* v slovarju, hkrati pa imamo premalo zadetkov za metaforično rabo, da bi si v slovarju zaslužili samostojen pomen in posebno izpostavitve.

Pri drugem pomenu (izročiti denar) se desno od glagola pojavljajo denarni izrazi: *stotak, zahtevana vsota, preživnina, 2000 tolarjev*. To glede na pomen ni nepričakovano. Vendar se tudi levo od glagola pojavljajo izrazi, ki konvergirajo v isto pomensko polje: *brez težav, le, morati*, pogojni odvisnik v povezavi z grožnjo, izraženo v prvem delu povedi. Vsi ti izrazi imajo modalni naboj in nakazujejo, da izkašljanje denarja ne poteka v pretirano prijateljskih okoliščinah; oseba izkašlja denar, ko jo k temu prisilijo zunanji dejavniki. To hipotezo podpira tudi vzorec II., kjer je element prisile eksplicitno izražen: *jih prisilijo, da se izkašljajo*.

Pri tretjem pomenu (povedati svoje mnenje) nismo odkrili nobenega ponavljajočega se vzorca. Mogoče bi se ti pokazali, če bi za ta pomen imeli več zadetkov. Zanimiva pa je rahla modifikacija osrednjega pomena, in sicer v naslednjih treh konkordancah:

(Saj sem rekla, da se morata pogovoriti, potem bo) bolje. Grem stavit, da sta se oba dobro izkašljala. Če bi eden od njiju odšel od doma,

(Lahko jo nekam pošljem, in takoj bo ocenila, ali me mora poslati nazaj ali pa naj bo tiho. Je) pravi amortizer. Ko si v krizi, se moraš izkašljati. Težko je, če si prepuščen samo tujim ljudem

si je pridelal veliko izkušenj. Ve, kaj mori) ljudi, ve, kako jih zagrabit, da se izkašljajo, zna speljati pozornost na sončnejše plati senčnega življenja,

V teh treh primerih *izkašljati se* pravzaprav ne izraža racionalno izražanje mnenja, ampak je bolj izpostavljen čustveni aspekt: nekoga nekaj teži, potem pa se izkašlja, tj. pove, kaj mu leži na duši. Lekseme, ki govorijo v prid tej interpretaciji, smo podčrtali. Tudi tu zaenkrat še ne bomo uvedli samostojnega pomena, verjetno pa je tej modifikaciji pomena smiselno prilagoditi definicijo v slovarju ali pa s primeri nakazati odstopanje od osrednjega pomena.

Četrty pomen (dati informacije) je soroden drugemu, saj osebo spet zunanji (običajno človeški) dejavnik prisili k delovanju, v tem primeru posredovanju informacij.

Element prisile je tu izražen v obliki grožnje, velelnika ali pogojnega razmerja – če se ne izkašljaš, potem se bo zgodilo kaj slabega (glej podčrtane izraze pri pomenu 4).

Pri petem pomenu (grajati) se pojavi nov vzorec (nekdo se izkašlja na nekoga), ki je vezan izključno na ta pomen in torej služi kot indikator pomena.

Pri šestem pomenu (izmisliti si) sta zanimiva oba predmeta, *alibije, ki so jih izločili iz kroga osumljencev in prepričljive razlage*, ki nakazujeta tipično situacijo: alibije in prepričljive razlage izkašljamo takrat, ko nam grozi nevarnost in hočemo odvrniti neprijetne posledice.

Za sedmo pomensko kategorijo imamo samo en zadetek, kar odpira zanimivo in aktualno vprašanje: koliko sopojavitev je potrebnih, da dobimo vzorec? Sinclair (1998: 5) predlaga, da sta potrebni vsaj dve sopojavitvi, kjer je pomembno tudi to, da se citata pojavita neodvisno drug od drugega, se pravi ne v istem besedilu. Če pa imamo več sopojavitev (moteče je, da ta *več* ni preciziran), pa lahko že govorimo o prototipični povezavi. Sedme kategorije torej ne bomo ustoličili kot pomen, ampak bomo nanjo gledali kot na modifikacijo oz. figurativno rabo prvega pomena. V prid tej interpretaciji govori dejstvo, da se vzorca prekrivata: v obeh primerih gre za prehodno rabo glagola *izkašljati*. Mogoče je, da se v slovenskem jeziku oblikuje nov pomen glagola *izkašljati*. Dokler ni na voljo obnovljen/razširjen korpus FIDA, lahko to preverimo na svetovnem spletu, tako da leksem vtipkamo v enega od brskalnikov. V času pisanja tega prispevka tudi konkordance na svetovnem spletu ne kažejo na to, da bi *izkašljati* tvoril nov pomen.

Tri konkordance smo uvrstili v kategorijo neanalizirano. V to kategorijo spadajo zadetki, kjer imamo premalo sobesedila, da bi lahko določili pomen ključne besede, ali pa so preveč inovativni/kreativni, da bi bilo mogoče z zanesljivostjo interpretirati pomen. Iz kognitivnega profila so zato izključeni.

S pomočjo korpusne analize smo odkrili vrsto slovnično-pomenskih lastnosti glagola *izkašljati*. Za konec smo preverili, kako uporabni so ti podatki v praktični leksikografiji in na osnovi kognitivnega profila postavili poskusno slovarsko geslo leksema *izkašljati*:

izkašljati prehodni glagol **1** izločiti s kašljem; **okužimo se z vdihanjem bacilov tuberkuloze, ki jih izkašlja bolnik**; (*refleksivno*) **nesposobnost malih otrok, da bi se izkašljali**; **2 pog.** izročiti denar, zlasti pod prisilo; **na koncu simpatičen tip v črnem le izkašlja zahtevano vsoto**; **3 pog.** (*refleksivno*) povedati svoje mnenje; zaupati skrbi; **ko si v krizi, se moraš izkašljati**; **4 pog.** (*refleksivno*) posredovati informacije, zlasti pod prisilo; **na dan z imeni! Izkašljaj se!**; **5 pog.** izmisliti si; **če ne bom hitro izkašljala prepričljive razlage, mi bo Steve prišel na sled**; **6 pog.** (*refleksivno*) grajati; **izkašljati se na koga**

Pri postavitvi geselskega članka nas je zanimalo predvsem to, kako lahko rezultate korpusne analize uporabimo pri razdelitvi pomenskega polja leksema (koliko pomenov mu pripišemo), pri sestavljanju definicij in izboru primerov rabe, zato drugih elementov gesla članek ne navaja. Vedno obstaja več možnosti razporejanja pomenov, (ne)ločevanja konkretnih in abstraktnih pomenov itd. Odločitve o tem, katere podatke želimo v geslu bolj izpostaviti in katere manj, se pravi, kakšen pri-

kaz je optimalen, morajo biti rezultat temeljitega premisleka predvsem o tipu slovarja (za dekodiranje ali enkodiranje) in ciljni publikli (domači govorci ali tujci).

V geselskem članku smo poskusili opredeliti tudi stilno raven posameznih pomenov. Pomene, za katere menimo, da se pojavljajo v pogovornih zvrsteh jezika, smo opremili s kvalifikatorjem 'pogovorno' (pog.). Pri določanju stilne ravni smo si pomagali s širšim sobesedilom in s podatki o avtorju in mestu objave.

V poskusnem geselskem članku pomeni niso razporejeni po pogostosti pojavitve v korpusu, ki je dokaj pogosta in tudi smiselna tehnika razvrščanja pomenov, ampak je bil kriterij mogoče bolj ohlapna »osrednjost pomena«, kar pomeni, da kot prvi pomen navedemo najbolj konkretni pomen, potem pa iz njega izpeljane abstraktne pomene. To tehniko razvrščanja pomenov so npr. dokaj uspešno uporabili uredniki NODE (Hanks 1998). To je seveda le eden od načinov razvrščanja pomenov v slovarskem članku in z njim so povezane določene težave, vendar je problematika preveč kompleksna, da bi jo lahko obdelali v tem članku. V našem geslu imamo na prvem mestu najbolj konkreten 'izločiti s kašljem', ki je hkrati najpogostejši pomen, sledi mu 'izročiti denar', kjer še vedno »izločimo« nekaj konkretnega, temu pa sledijo 'izločanja' bolj abstraktnih predmetov.

Glede na to, da je bil glavni namen naše korpusne analize odkriti vzorce, ki se vežejo na posamezne pomene leksema *izkašljati*, se poraja vprašanje, v kateri element mikrostrukture geselskega članka te podatke najbolj učinkovito vključimo. Najbolj logičen odgovor bi bil, da v definicijo, in včasih lahko podatke o tipični situaciji res vpletemo v definicijo, npr. pri 2. in 4. pomenu. Vendar morajo biti definicije kratke, jedrnat in jasne, zato jih običajno ne moremo obremenjevati s podatki o tipični rabi leksema. Za prikaz vzorcev so idealni primeri rabe, ki definicijo dopolnijo in nadgradijo. V poskusnem geslu smo iz našega inventarja primerov poskušali izbrati primere, ki dobro ilustrirajo pomen in hkrati nakažejo tipično okolje, v katerem se leksem pojavlja. Primere iz korpusa smo spremenili minimalno (v primeru *če ne bom zdaj hitro izkašljala prepričljive razlage, mi bo Steve prišel na sled* smo izločili besedo *zdaj*, ker je redundantna), saj je ena glavnih prednosti slovarja, narejenega po korpusnem pristopu, ravno v tem, da uporabniku ponudi žive primere rabe, ki kažejo, kakšna je dejanska raba jezika. Kakor hitro primere prirejamo ali jih krajšamo, je to že kršenje korpusnega pristopa in vračanje k introspekciji. Če so primeri iz korpusa neprimerni kot ilustrativno gradivo za slovar (so predolgi ali prezapleteni), se korpusni jezikoslovec raje kot za posege v primere odloči za prikaz tipične strukture, kot smo naredili pri šestem pomenu, oz. si pomaga z drugimi besedilnim viri.

5 Sklep

Korpusno inducirani kognitivni profili so učinkovito sredstvo za odkrivanje zakonitosti jezikovne rabe in dobro izhodišče za slovaropisno delo. Kljub temu pa geselskih člankov slovenskih-x (x = katerikoli jezik, vključno s slovenskim) slovarjev ne gre postavljati samo s pomočjo kognitivnih profilov. Korpus FIDA je opremljen z orodji za izračun različnih statističnih vrednosti, s pomočjo katerih lahko dobimo

veliko podatkov o skladenjskih vzorcih, kolokatorjih in frazeologiji.⁵ Poleg tega noben slovar ne ignorira precedenčne literature, v našem primeru predvsem SSKJ. Verjetno je, da bi se geselski članki, narejeni s pomočjo korpusa FIDA, delno prekrivali s članki v SSKJ, kot je tudi zelo verjetno, da bi se s pomočjo korpusa odkrile posebnosti rabe, ki so se sestavljalcem SSKJ (zaradi »ročnega« načina dela) izmuznile. Korpus FIDA torej ne spodbija avtoritete SSKJ, ampak ga dopolnjuje in nadgrajuje. S tem v zvezi je zanimiva že primerjava našega geselskega članka za *izkašljati* in geselski članek istega leksema v SSKJ, ki jo v celoti prepuščam bralcem.

V leksikografski praksi je zanimivo tudi vprašanje, koliko konkordanc je potrebno analizirati, da lahko na osnovi pregledanega materiala izdelamo kognitivni profil, identificiramo vzorce in v končni fazi postavimo geselski članek. V našem primeru smo analizirali vse zadetke, vendar ima *izkašljati* v korpusu FIDA le 48 zadetkov. So pa tudi leksemi, ki imajo po več tisoč zadetkov. Če bi pregledali vse te konkordance, bi se sestavljanje potencialnega slovarja ne potegnilo le na 70 let (kot pri OED), ampak v neskončnost. V praksi se zato določi meja – pregleda se x število naključno izbranih primerov, ki se vzamejo kot celota. Število x mora biti dovolj veliko, da je vzorec reprezentativen, hkrati pa ne preveliko, da projekt preveč ne preseže vnaprej določenega časovnega okvira. Po ustnem zagotovitlu Patricka Hanksa, glavnega urednika NODE, so pri sestavljanju NODE pri vsaki besedi (tudi pri najbolj pogostih besedah v jeziku) začeli s 300 konkordancami, kar naj bi pri veliki večini besed zadostovalo. Le za pomensko zelo zapletene lekseme so uredniki analizirali dodatnih 300 konkordanc, v izjemnih primerih pa tudi do tisoč. Dejstvo je, da se vsak slovar dela pod velikim časovnim pritiskom in težko rečemo, kaj se dejansko dogaja v slovarskih delavnicah, ko bi slovar že moral biti končan, pa še ni in založniki pritiskajo na urednike. Tristo konkordanc se zdi obvladljiva številka, čeprav si v slovenskem okolju zelo verjetno ne bi mogli privoščiti, da bi za vsako geslo analizirali tristo (ali po potrebi več) konkordanc. Sama sem za kognitivni profil *izkašljati* in poskusno geslo porabila več časa, kot ga imajo na voljo uredniki, ki delajo na slovarskih projektih. Po drugi strani pa se zdi, da s pomočjo kognitivnih profilov postane vidna marsikatera lastnost leksemov, ki je, če konkordance le preletimo, že zaradi omejene zmogljivosti pomnjenja človeških možganov, ne odkrijemo. Treba je torej najti učinkovito ravnovesje med kakovostjo in hitrostjo dela. Rešitev iz te zagate mogoče že obstaja, vsaj v angleško govorečem okolju. Angleška korpusna leksikografija eno svojih glavnih nalog v 21. stoletju vidi v razvoju programskih orodij, ki bi omogočala izkoriščanje celotnega potenciala vedno obsežnejših korpusov, se pravi, da bi pri analizi upoštevala vse pojavitve iztočnice (in ne le naključni izbor konkordanc) in hkrati leksikografu ponudila več podatkov o rabi iztočnice kot le spisek najpomembnejših kolokatorjev. Na Univerzi v Brightonu so tako razvili avtomatski postopek (Kilgariff in Rundell 2002), s katerim lahko diagnosticirajo pomene iztočnice in kolokatorje, ki se vežejo na posamezne pomene. V prihodnosti lahko morda pričakujemo popolnoma avtomatiziran postopek izdelave kognitivnih profilov, kar bi bilo leksikografu predvsem pri pomensko kompleks-

⁵ Statistične metode raziskovanja okolja iztočnice opisujeta npr. Gorjanc in Krek (2001).

nih geslih v veliko pomoč, saj kognitivni profil, kot smo poskušali pokazati, predstavlja že kar ogrodje geselskega članka.

Literatura

- Gorjanc, Vojko, 1999: Korpusi v jezikoslovju in korpus slovenskega jezika FIDA. Kržišnik, Erika in Lokar, Meta (ur.): *35. seminar slovenskega jezika, literature in kulture*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovanske jezike in književnosti Filozofske fakultete. 47–59.
- Gorjanc, Vojko in Krek, Simon, 2001: A corpus-based dictionary database as the source for compiling Slovene-X dictionaries. *6th Conference on Computational Lexicography and Corpus Research*. »Computational Lexicography and New EU Languages. Birmingham: Centre for Corpus Linguistics. 41–47.
- Hanks, Patrick, 2000: Immediate context analysis: distinguishing meanings by studying usage. Heffer, Chris in Sauntson, Helen (ur.): *Words in Context. A Tribute to John Sinclair on his Retirement*. University of Birmingham ELR discourse analysis Monograph No. 18.
- Hunston, Susan, Francis, Gill in Manning, Elizabeth, 1997: Grammar and Vocabulary: Showing the Connections. *ELT Journal* 51/3. 208–215.
- Kennedy, Graeme, 1998: *An Introduction to Corpus Linguistics*. London & New York: Addison Wesley Longman.
- Kilgarriff, Adam in Rundell, Michael, 2002: Lexical Profiling Software and its Lexicographic Applications – a Case Study. Braasch, Anna in Povlsen, Claus (ur.): *Proceedings of the Tenth EURALEX International Congress*. Copenhagen: CST. 807–818.
- Krek, Simon, 1999: Računalniški korpusi v slovaropisju. *Razgledi* 13 (23. junij 1999). 8–9.
- Krek, Simon, 2003: Sodobna dvojezična leksikografija. *Jezik in slovstvo* XLIII/1. 45–60.
- McCarthy, Michael, 1990: *Vocabulary*. Oxford: OUP.
- Sinclair, John, 1998: The Computer, the Corpus and the Theory of Language. Azzaro, Gabriele in Ulrych, Margherita (ur.): *Transiti linguistici e culturali II. XVIII AIA Congress Anglistica: metodi e persorsi comparatistici delle lingue, culture e letterature di origine europea*. Trieste: EUT. 1–15.
- Sinclair, John, 1991: *Corpus, Concordance, Collocation*. Oxford: OUP.
- Stabej, Marko, 1998: Besedilnovrstna sestava korpusa FIDA. Štrukelj, Inka (ur.): *Jezik za danes in jutri. Zbornik referatov na II. kongresu Društva za uporabno jezikoslovje Slovenije*. Ljubljana: Društvo za uporabno jezikoslovje Slovenije. 97–106.
- Tognini Bonelli, Elena, 2000: Functionally Complete Units of Meaning across English and Italian: Towards a Corpus-driven Approach. Granger, Sylviane in Altenberg, Bengt (ur.): *Lexis in Contrast*. Amsterdam/Philadelphia: Benjamins. 1–15.
- Winchester, Simon, 1999: *The Surgeon of Crowthorne. A Tale of Murder, Madness and the Oxford English Dictionary*. Harmondsworth: Penguin Books.

Slovarji

NODE – Hanks, Patrick (ur.), 1998: *The New Oxford Dictionary of English*. Oxford: OUP.

OED – 1989. *The Oxford English Dictionary*. 2. izd. Oxford: Clarendon Press. (OED – Simpson, J. A., Weiner, E. S. C., ured. 1989: *The Oxford English Dictionary*. 2. izd. Oxford: Clarendon Press.)

SSKJ – 1997. *Slovar slovenskega knjižnega jezika*. 2. ponatis. Ljubljana: Inštitut za slovenski jezik ZRC SAZU, DZS.

Spletne strani

BoE – Bank of English <<http://titania.cobuild.collins.co.uk>>

BNC – British National Corpus <<http://info.ox.ac.uk/bnc/>>

FIDA – Korpus slovenskega jezika <<http://www.fida.net>>

Priloga

Konkordance leme *izkašljati* iz korpusa Fida

1. He, he, he, ste se vsi *izkašljali*? Kaj je? Ste me res odpisali?
2. rob. Zaradi sabotaže se bolje počutijo: »se *izkašljajo*«. Takšna dejanja so ponavadi spontana in nenačrtovana.
3. je ostal isti. V tistem se je profesorju uspelo *izkašljati*, pljunil je v robec, ki ga je držal
4. Potem sem *izkašljal* čigumi v dlan in ga, ne da bi vedel
5. in dozori v mačkinih pljučih. Okužena mačka ga pogosto *izkašlja* in se ga tako znebi, vendar so na voljo
6. je bil Rupsi še župan, je vedno brez težav *izkašljal* stotaka. Ko so ga za kak praznik prosili za
7. ki so na omenjeno linijo že poklicali in se *izkašljali*. Izkazalo se je, da so bile njihove domneve
8. z neznanci snuje, kar pač snuje, se takoj *izkašljaj*. Če se *izkašljaš*, te aka vreča denarja,
9. kar pa snuje, se takoj *izkašljaj*. Če se *izkašljaš*, te čaka vreča denarja, če pa ne,
10. od dobre volje. Jasno, z odgovorom sem bil *izkašljal* tudi iver. Srečal sem dva, tri znanke in
11. izločka, ki postane manj lepljiv in ga je lažje *izkašljati*. Pogosto jih uporabljamo v inhalacijah.
12. ga spremljajo dodatni simptomi, npr. vročina ali krvav *izkašljani* izloček.
13. z vdihanjem bacilov tuberkuloze, ki jih *izkašlja* bolnik z aktivno boleznijo. Širi se lahko tudi z
14. š utekočinjenje in sprostitve sluzi, ki jo bolnik tako lažje *izkašlja*. V hujših primerih mora bolnik ostati v bolnišnici,
15. so jo ljudje odlagali v priročni grajski deponiji, in *izkašljala* dober kilogram prahu. Knedlina torba in vrečka s špecerijo
16. ljudi, ve, kako jih zgrabiti, da se *izkašljajo*, zna speljati pozornost na sončnejše plati senčnega življenja,
17. Kistiakowskega, na katerega so se zdaj kar po vrsti *izkašljali* Oppenheimer, general Groves, Vannevar Bush in tako naprej

18. Ko prekriza noge in se vživlja v svojo vlogo, *izkašljata* fotelja izpod rjuh veliko prahu.
Ko je končno le
19. mirno in debelo *izkašljam* ter sem pa tja lagodno in globoko rignem. Šoferji
20. Prosim?! sem *izkašljal* in ga debelo gledal. Mislim, iskrenost pa taka
21. se dolgo dolgo zadrževal, da nisem vse skupaj *izkašljal* in potem sem bil do amena zmeden. Konfuzen.
22. otrok je sposoben sluz, ki mu zaliva dihala, *izkašljati* in izpljuniti, dojenček pa vso to nesnago požira,
23. otrokovega življenja, ker se tak otrok že itak slabo *izkašlja*. Sicer pa velja pravilo, da skušamo suh kašelj
24. sapniku ali pa mehčajo sluz, da jo bolnik lažje *izkašlja*. Kašelj praviloma pomirjamo le ponoči, da se bolnik
25. zavidljivim številom znanih političnih osebnosti se je kot prvi govorec *izkašljal* politolog Toni Kuzmani. Zadržal se je pri trendu izginjanja
26. anatomske razmere in nesposobnost malih otrok, da bi se *izkašljali* in izpraznili nos in žrelo. Pogoji za zdrava ušesa
27. naložila v prejšnjem pismu! Če ne bom zdaj hitro *izkašljala* prepričljive razlage, mi bo Steve prišel na sled in
28. bolje. Grem staviti, da sta se oba dobro *izkašljala*. Če bi eden od njiju odšel od doma,
29. daviti in dušiti, čutim silno potrebo, da bi *izkašljala* vse to. Uboga Milena, ki mi sledi
30. Za koga dela? Na dan z imeni! *Izkašljaj* se!
31. pravi amortizer. Ko si v krizi, se moraš *izkašljati*. Težko je, če si prepuščen samo tujim ljudem
32. Otroku prigovarjajte, naj predmet *izkašlja*.
33. Če otrok predmeta ne *izkašlja*, ga položite čez koleno in nekajkrat udarite po predelu
34. saj mu prijateljsko pravi »gospod profesor«, le *izkašlja* zahtevano vsoto. Sledi dobrodušen aplavz, jaz pa se
35. homoseksualnosti, na koncu zmaga prvi, ki mora zanjo *izkašljati* 2000 tolarjev. Cena? Sitnica!« Potem
36. za taktiko v slogu: vržeš bombico, sindikati se *izkašljajo*, vlada preveri javno mnenje, in če to ni
37. nato prerežejo popkovino ter ga očistijo krvi. Dojenček *izkašlja* vodo, ki se mu je nabrala v pljučih,
38. glavo navzdol ter vas udaril po zadnjici, da bi *izkašljali* vodo, ki se vam je nabrala v pljučih.
39. dojenček rodi, mora začeti tudi dihati. Potem ko *izkašlja* vodo, ki se mu je nabrala v pljučih,
40. umazanimi podrobnostmi iz življenja Jacksonov, če Janet ne bo *izkašljala* velikodušne odpravnine. Rene, ki je bil z Janet
41. je povabil na večerjo. Že dobro nadelani smo se *izkašljali* drug na drugega – do dna. Zjutraj smo bili
42. do njega. Zašepetal je: Tipa sta se *izkašljala*. Rekla sta, da je prodajala svojo pičko,
43. značilnosti: njihova imena so pristala v dnevnem časopisju in *izkašljali* so alibije, ki so jih izločili iz kroga osumljencev
44. v kateri je stanoval. Eden od njih se je *izkašljal* policajem, da bi umora ne obesili njim. Vasquez
45. mi bodo postregli s pravimi imeni. Če se niso *izkašljali*, sem jih aretiral, če so se, sem
46. jih ponavadi tako ali drugače prisilijo, da se *izkašljajo*.
47. kot napake nekoga drugega. V nedeljo se bodo ljudje *izkašljali*, potem pa bo že na prihodnji seji določil le