
SESTAVLJANJE ENOJEZIČNEGA SLOVARJA ZA DOMAČE GOVORCE

V članku primerjamo vlogo slovarja in leksikalne baze in obravnavamo problematiko slogovne zaznamovanosti in jezikovne pravilnosti v slovarjih. Nadalje se ukvarjamo z različnimi tipi razvrstitev pomenov v slovarjih, z zgodovino besed in z načeli izbire slovarskih gesel. Korpus navajamo kot bistven vir za obravnavo pomena, kolokacij in sintagmatike, ukvarjamo pa se tudi z načini interpretacije korpusnih podatkov – korpusnim profiliranjem iztočnic. Zaključujemo z mislijo, da slovar prikazuje osrednji jezikovni standard, pri čemer morajo vsa izražena jezikoslovna mnenja v njem temeljiti na korpusnih dokazih.

Gljučne besede: leksikografija, korpusno jezikoslovje, računalniško jezikoslovje, analiza diskurza

Uvod: slovar in podatkovna baza

Za začetek bi se rad zahvalil organizatorjem za čast, da smem otvoriti konferenco s plenarnim predavanjem. Poskusil vam bom povedati nekaj malega o angleških izkušnjah s sestavljanjem enojezičnih slovarjev, namenjenih v prvi vrsti domačim in ne tujim govorcem jezika. Žal mi je le, da z nami zaradi slabega zdravja ne more biti bivši kolega, genialni Ramesh Krishnamurthy z univerze Aston, ki bi nam lahko kaj več povedal o korpusih in računalniški tehnologiji v leksikografiji. Prihodnost uporabe računalnikov pri sestavljanju slovarjev za domače govorce je zaradi razvoja korpusnih virov v današnjem času tako obsežna in zanimiva tema, da je v svojem govoru ne bom niti poskusil obravnavati. Mislim, da je dr. Krishnamurthy nameraval govoriti ravno o tem. Podrobnejša razprava o tehnoloških vidikih

sestavljanja novih slovarjev, o tem, kakšni bi lahko bili oziroma bi morali biti v dobi interneta, ter v kakšno pomoč nam bo računalniška tehnologija, bo morala počakati na drugo priliko.

Začel bom s primerjavo vlog slovarja in leksikalne baze ter povedal nekaj besed o problematiki slogovne zaznamovanosti in jezikovne pravilnosti. Nato bom na kratko spregovoril o besedah, zgodovini besed, leksikografskih raziskavah in izbiri slovarskih gesel – čemur je pokojni Dwight Bolinger rekel »nabiranje besed«. Posvetil se bom tudi makro- in mikrostrukturi slovarja.

Leksikalna baza je osnovni, temeljni vir za mnoge pomembne jezikoslovne izdelke, med drugim slovarje, učbenike ter računalniške programe za obdelavo naravnega jezika. Veliki enojezični slovar ima drugačno funkcijo: združuje govorce nekega jezika, ima družbeno povezovalno funkcijo, kajti izpostavlja osnovne podatke o besedah, pomenih in rabi, na čemer temeljijo različne rabe jezika. Besede imajo pomene – oziroma natančneje rečeno imajo potencial, da ustvarjajo pomene, če jih postavimo v neko sobesedilo – ti pa so povezani z določenimi nizi sintagmatskih vzorcev, ki se jih da odkriti s temeljito korpusno analizo. Toda besede imajo tudi svojo slogovno vrednost, niso vse enako primerne v vseh okoliščinah. Nekatere besede in slovnične strukture spadajo v sleng, lahko so umestne le v govorjenem kontekstu, značilne le za določene regije ali dialekte. Spet druge se uporablja le v uradnih pravnih dokumentih ali pa le v romanih ali poeziji. Nekatere so pomensko nedvoumne in jasne, a se jih ne sme uporabiti v uglajeni družbi. V leksikalni bazi so te stvari posredno razvidne iz vrste besedil, v katerih se beseda pojavlja. Dober slovar pa vse to jasno izpostavlja. Ni le inventar besed, njihovih pomenov in sintagmatskih vzorcev, temveč navaja tudi podatke, kot je slogovna vrednost besed – družbeni odnos do »pravilne rabe«. Od slovarja se pričakuje, da bo podal sodbo, kaj je v različnih kontekstih pri rabi pravilno in kaj ne. Pomembno pa je, da sodbe v slovarju temeljijo na empirični analizi dejanske rabe v besedilih dobrih piscev, ne zgolj na osnovi preferenc in predsodkov nekaj novinarjev, akademikov in samozvanih strokovnjakov. Potrebna je torej interakcija med znanstveno zasnovano leksikalno bazo in slovarjem kot družbenim artefaktom.

Etimologija in zdrava pamet

Kakšna je vloga slovarja pri raziskovanju in navajanju etimologije in zgodovine besede? Nekateri menijo, da je to edina funkcija slovarja, jaz pa bom zagovarjal trditev, da je zdravorazumsko prepoznavanje konvencij glede pomena in rabe besed, na katere se opirajo pripadniki jezikovne skupnosti in na podlagi katerih se sporazumevajo, celo pomembnejša funkcija. Treba je torej razlikovati med dvema večjima vrstama enojezičnih slovarjev za domače govorce. Na eni strani so tradicionalni večji slovarji, ki temeljijo na zgodovinskih principih in navajajo zgodovino besede in etimologijo. Vrhunski primer te vrste slovarja je veliki, v

več knjigah izdani *Oxford English Dictionary* (OED). Nastal je med letoma 1878 in 1928, druga izdaja je izšla leta 1986. Zdaj ga na založbi Oxford University Press revidirajo pod vodstvom urednika Johna Simpsona. Revidirana gesla si je moč sproti ogledati na spletu in rezultati raziskave za vsako besedo so na voljo strokovnjakom in javnosti takoj, ko je končana. Novo izdajo OED dobivamo torej po kosih, v skladu s potekom projekta in brez omejitev abecednega reda. V prejšnjih časih smo na objavo rezultatov leksikografskih raziskav v velikem historičnem slovarju morali čakati tudi do 50 let. Zdaj nam je na razpolago v nekaj tednih po izdelavi. OED je slovar, ki temelji na historičnem načelu. Na prvem mestu je etimologija besede, za njo pa najstarejši znani pomen besede. Najnovejši razvoj je predstavljen čisto na repu gesla. Ker je pomen besede nestabilen in se precej hitro spreminja, se zato v takem slovarju zdajšnji pomen besede nahaja na zadnjem oziroma skoraj zadnjem mestu, pred njim pa je eden ali več zastarelih oziroma redkih pomenov. Slovar, narejen po historičnem principu, vam bo na primer povedal, da je *camera* majhen obokan prostor in takoj za tem, da je zakladnica rimske kurije. Malo naprej boste izvedeli, da je *camera* zatemnjena soba (*camera obscura*) na vrhu hiše z odprtino v strehi, nad katero je ogledalo. *Camera obscura* odseva podobe okolice (mesta ali podeželja) na svetlobno mizo v sobi. Primerek *camere obscurae* se je kot turistična zanimivost ohranil v Edinburghu. Ta obskurni izraz je pomemben kot povezava med modernim pomenom besede *camera* in historičnim pomenom 'majhen prostor'. Historični slovar šele na koncu gesla omeni, da je *camera* priprava za fotografiranje ali snemanje filmov.

Druga vrsta enojezičnega slovarja je slovar, izdelan po sinhronem načelu. Sinhroni slovar v osnovi obrne vrstni red pomenov in najnovejši pomen umesti na prvo mesto. Tak slovar vam tako najprej pove, da je *camera* priprava za fotografiranje in snemanje filmov. Šele nato razloži, od kod beseda izvira in kako so se razvili sodobni pomeni. Sinhroni slovar se ukvarja predvsem s konvencionalnimi pomeni in rabo ter ne toliko z zgodovinskimi in etimološkimi raziskavami. Med angleškimi slovarji, izdelanimi po sinhronem načelu, sta *Collins English Dictionary* (CED; 1979), ki sem ga zasnoval in uredil v sedemdesetih letih prejšnjega stoletja – založnikom se je sicer kasneje zdelo primerno, da z naslovnice odstranijo moje ime – in *New Oxford Dictionary of English* ((N)ODE; 1998), ki sva ga oblikovala in redigirala z Judy Pearsal v devetdesetih letih in na katerem se je moje ime ohranilo tudi v novejših izdajah, kljub spremembi naslova (iz naslova druge izdaje so odstranili *New*). V Ameriki je Houghton Mifflin izdal *American Heritage Dictionary* (AHD; 1969, zdaj je izšla njegova četrta izdaja), ki svoj nastanek dolguje ogorčenju Jamesa Partona, izdajatelja revije *American Heritage*, nad neuspešnim spopadom Merriam Webstrovega *Third New International Dictionary, unabridged* (1961) z vprašanjem slogovne zaznamovanosti in pravilne rabe besed. Parton se je odločil, da bo naročil izdelavo lastnega slovarja in nastal je AHD. Poleg razlag in primerov AHD vsebuje mnogo krajših prispevkov o spornih točkah rabe besed, v katerih primerjajo in analizirajo mnenja več kot sto strokovnjakov za slogovna vprašanja. AHD temelji na sinhronih načelih. Glavni avstralski slovar, *Macquarie Dictionary* (1981, izšla je njegova četrta izdaja) prav tako temelji na načelu sinhronosti.

In kako je s historičnimi slovarji, ki so izšli v eni knjigi? Človek bi mislil, da jih sploh ni, ker je slovar v eni knjigi nadvse praktično orodje, v resnici pa jih je kar nekaj in nekateri prav dobro uspevajo. Britanski primer historičnega slovarja je bil *Chambers 20th Century Dictionary*, ki je vztrajal skoraj stoletje, dokler se leta 1993 založnik iz tržnih razlogov ni odločil, da ga bo opustil in nadomestil s slovarjem *Chambers English Dictionary*, ki jezik opisuje po načelu sinhronosti.

Presenetljivo v Ameriki še vedno prevladujejo slovarji po historičnem principu, celo med enozvezkovniki. Najbolj priljubljen slovar v Ameriki, po prodanih izvodih sodeč, je *Merriam Webster Collegiate*, ki je historični slovar. Prvič je bil objavljen leta 1898, zadnja po vrsti je njegova 11. izdaja. Ni prav verjetno, da se kdo od njegovih kupcev in uporabnikov zaveda temeljne razlike med njim in bolj zdravorazumskimi sinhronimi slovarji. Ni izključeno celo, da kdo pogleda v slovar in si misli, da je pravi pomen *camere* prvi naštet: 'majhen prostor' ali 'vatikanska zakladnica', in da je naprava za fotografiranje zgolj novejši, pogovorni razvoj pomena besede nižje slogovne vrednosti, ki naj se ga previdni pisci ogibajo. To vsekakor ne bi bilo neumestno sklepanje glede na razporeditev pomenov, čeprav seveda povsem napačno. Le kako naj uporabnik takega slovarja razbere sodobni pomen polisemične besede, če ga še ne pozna?

Nekatere besede niso spreminjale pomena, odkar je angleščina nastala, druge so pomen spremenile več kot le enkrat. V angleščini bi lahko našteli nešteto primerov spremembe pomena. *Sock* je bil prvotno 'lahek čevljev', *dope* je najprej pomenil 'lak', šele kasneje 'droga'. *Silly* je nekoč pomenil 'vesel' in 'nizkega stanu'.¹ Zakaj je angleški *magazine* očitno soroden francoskemu *magazin*, čeprav imata zelo različne pomena? V angleščini je *magazine* periodična publikacija ali del pištole, francoski *magazin* pa označuje veleblagovnico, kamor greš po nakupih. Združuje ju dejstvo, da obe izvirata iz arabske besede za skladišče. Našteli bi lahko še na tisoče primerov spremembe pomenov besede.

Razporeditev in predstavitev takih informacij v slovarju je lahko precej tvegan in zahteven podvig. *Size* v sodobni angleščini pomeni 'velikost', 'dimenzijo', 'razsežnost', torej je osnoven pojem. Mislili bi, da je beseda *size* v angleščini že od nekdaj, a temu ni tako. V jezik je prišla v poznem srednjem veku, precej presenetljivo, zaradi goljufivih navad srednjeveških pekov. Če v 15. stoletju nisi bil zadovoljen s štruco, ki si jo kupil pri peku, če se ti je zdela premajhna in se je to nenehno ponavljalo, so peka odvedli pred krajevno sodišče, t. i. *assizes*, in ga kaznovali zaradi nepoštenega trgovanja. *Size loaf* je bila štruca s strani sodišča odobrenih dimenzij in velikosti. Iz te ozke zasnove se je pomen besede razširil in danes pomeni dimenzijo ali velikost česarkoli. Ključno pri tem je, da je sprememba pomena besed nepredvidljiva, a pogost in bistven del jezikovnega razvoja. Današnja **razširitev** (ang. *exploitation*) besednega pomena bo morda jutri že **standard** (ang. *norm*).

¹ *Sock* – 'nogavica', *silly* – 'neumen, trapast'. Op. prev.

Videli smo, da je besedni pomen nestalen in se lahko pogosto spremeni. Nekateri, vključno z dvojezičnimi slovaropisci, kot je Sue Atkins, sploh zanikajo obstoj besednega pomena. Tega ni težko razumeti. Če pomislite na besedo *fire*, kaj pomeni? Je nekaj, kar nenadzorovano divja na polju ali v gozdu, nekaj, kar lepo in mirno gori v vaši hiši, ali pa je povezano z orožjem? Pomeni, da izgubiš službo? Ima kaj opraviti z lončarstvom? Ali nekaj, kar vzbuja vnetje, navdušenje?² Odgovor je, da je *fire* v osami vse to in še mnogo več, oziroma ima potencial, da pomeni vse to, ko ga uporabimo. Iz tega razloga sem trdil, da ima, strogo vzeto, izraz v osami le pomenski potencial, nima pa pomena. V enojezičnem slovarju torej najdete seznam **pomenskih potencialov**, ne pa pomenov v strogem pomenu besede. Kateri pomen se z uporabo aktivira, se pokaže šele v sobesedilu. In da bi vedeli, v kakšnih sobesedilih se besede navadno pojavljajo, potrebujete korpuse in korpusno analizo. Introspekcija kot vir podatkov ali raziskovalna tehnika enostavno ne deluje. V petdesetih letih generativnega jezikoslovja smo se morali zgrda naučiti, da samoopazovanje ne da zanesljivih podatkov o rabi besed v vsakdanjem jeziku. Eno izmed najpomembnejših spoznanj korpusne lingvistike je, da niti izobraženim jezikoslovcem poročanje o lastnem jezikovnem vedanju ne gre dobro od rok. Introspekcija popači, morda zato, ker se ljudje ob intuitivnem razmisleku navadno spomnijo le nenavadnih primerov, ki ilustrirajo meje možnega, namesto vsakodnevne rabe. Kaže, da je običajna raba zakopana tako globoko v podzavesti, da jo ljudje težko priključijo v zavest in točno opišejo.

Nabiranje besed

Prva dolžnost leksikografa je »nabiranje besed«. Urednik slovarja za domače govorce mora v slovar vključiti široko paleto besed, kajti zelo pogosto bodo ljudje iskali ravno nenavadne in redke besede ter pomene. Urednik slovarja za tuje govorce pa bo moral biti bolj selektiven ter izbrati in pojasniti le tiste besede, ki bodo po njegovi presoji prišle prav tujemu govorcu.

Leta 1857 je Richard Chenevix Trench, eden od utemeljiteljev slovarja *Oxford English Dictionary*, slovaropisce opisal kot »administrativne popisovalce jezika«. Opis je povsem primeren. Popisovanje se morda zdi preprost poklic, a je v primeru leksike naloga, polna zanimivih izzivov, kot mi bo, upam, uspelo prikazati. Morda se zdi, da bi moralo biti sestavljanje seznama vseh besed v jeziku in njihovih pomenov lahka naloga, a je v resnici zelo zahtevna in težka. Z introspekcijo ne moremo priti do seznama vseh besed v jeziku. Delno zato, ker jih je težko priklicati v spomin, kot sem že omenil, delno pa zato, ker nihče ne ve vsega. Celo po štiridesetih letih leksikografskega dela naletim na besede, ki mi niso znane. Slovaropisec je včasih tudi v zagati, kaj sploh šteje za besedo. Osnovno besedišče standardnega jezika se preliva v mnoge smeri – v tehnični žargon, regionalno omejeno narečje, sleng, zastarelo besedišče, pesniške skovanke in tako naprej. Bi moral slovar vsebovati

² *To fire a gun* – 'ustreliti s strelnim orožjem', *to fire* – 'odpustiti (iz službe)', *to fire* – 'žgati (glino)', *to fire somebody* – 'navdušiti koga'. Op. prev.

lastna imena? Kje naj leksikograf potegne črto? In ko sprejme vse te odločitve, ostane vprašanje meja besede. Angleščina ni aglutinacijski jezik kot na primer turščina, ali nekateri bi rekli nemščina, kjer je veliko težje najti mejo besede. A kljub temu imamo tudi v angleščini včasih zanimive težave z določanjem meja besede, kar bom ponazoril z dvema primeroma.

V angleščini je dobro znan problem sestavljenih glagolov. Jasno je, da *take off*, dejanje letala, ko se dvigne od tal in začne leteti, ni ista beseda kot osnovna beseda *take*. Rečeno je bilo celo, da ima *take off* prav toliko skupnega s *take* kot *disease* z *ease*. Po tej logiki bi morali omenjena glagola v slovarju obravnavati kot ločeni leksikalni enoti. Naj torej v slovar vključimo vse frazne glagole? Kaj pa na primer glagoli, kot je *finish up*?³ Naj bi bili tudi ti samostojna gesla v slovarju? Klasičen odgovor je ne, kajti imajo sestavljen pomen. V omenjenem glagolu *finish* še vedno pomeni končati, *up* pa ima zgolj dopolnilno, poudarno vlogo. Toda po drugi strani obstajajo tudi frazni glagoli, kakršen je *break up*, pri katerem ima *up* lahko idiosinkratično ali dopolnilno, poudarno vlogo.

Na primer: politični protestniki se lahko razidejo (ang. *break up*) ali pa jih razžene policija (ang. *are broken up by the police*). Zakonca ali par se lahko razideta (ang. *break up*); v ameriški angleščini pa se izraz *break up* uporablja tudi v primeru, kadar ljudi preplavijo čustva in se začnejo smejeti ali jokati. Obstaja še več pomenov, pri katerih morata biti prisotna glagol *break* in členek *up*. Obstajajo pa tudi pomeni, kjer členek ni obvezen. Na primer, če polomimo mizo, lahko glagol *break* uporabimo brez *up*, lahko pa tudi skupaj z njim. V tem primeru členek poudari, okrepi pomen glagola *break*, pove, da se miza ni prelomila le na pol, temveč je razpadla na več delov. Bi morali v slovarju navesti vse ali le idiosinkratične pomene takih glagolov?

Še bolj problematični so večbesedni izrazi, kakršna sta na primer *fire engine* in *fire extinguisher*. *Fire engine* je kamion z opremo za gašenje požarov. Jasno je, da pomen tu ni sestavljen. *Engine* ne pomeni 'kamion'. Besedo *fire engine* bi torej morali uvrstiti v slovar. Ameriški izraz *fire truck* je bolj ali manj sestavljen. Tudi pomen izraza *fire extinguisher* je bolj ali manj sestavljen. Je kos opreme za gašenje požarov. Toda ker je izraz, ki označuje razred edinstvenih predmetov, *fire extinguisher* ponavadi uvrstijo med slovarska gesla. Kaj pa večbesedni izrazi, kot sta *wood fire* in *forest fire*? V angleščini obstaja na stotine tisočev takih večbesednih izrazov, gre za izredno produktivno področje jezika. Ves čas nastajajo novi večbesedni izrazi. Slovarji pri njihovem beleženju slabo opravljajo svoj posel. Običajno opravičilo za nenavajanje je, da so to zloženke s prozornim pomenom. *Wood fire* je ogenj, pri katerem gori les, *forest fire* pa ogenj, pri katerem gori gozd. Na prvi pogled se to zdi prepričljivo, a se na koncu vendarle izkaže za nezadovoljivo. *Wood* in *forest* sta sopomenki⁴ in če bi bil pomen resnično prozoren,

³ *Take off* – 'vzleteti', *take* – 'vzeti', *disease* – 'bolezen', *ease* – 'lahkota', *finish up* – 'končati'. Op. prev.

⁴ *Wood* v angleščini pomeni tako 'les' kot tudi 'gozd'. Torej je *wood fire* mogoče interpretirati tudi kot 'ogenj, pri katerem gori gozd'. Op. prev.

bi morala imeti *wood fire* in *forest fire* enak pomen. A ga nimata. *Wood fire* (snovno ime) je kurjenje drv na ognjišču ali v taboru pod nadzorom človeka, *forest fire* pa je gorenje gozda ali gozdov (števni samostalnik), torej gozdni požar, ki divja brez nadzora. Omenjena izraza sta bila konvencionalizirana vsak po svoje, v vsakem od njiju je pri osnovnih komponentah izrabljen drug pomen. To bi bilo v slovarju potrebno razložiti, a izrazov ne boste našli nikjer.

Politiko odločanja o tem, kaj šteje kot leksikalna enota in kje zarisati mejo, spremlja vprašanje, kako in kje najti besede. V zadnjih 150 letih je bil branju tekstov in zbiranju navedkov namenjen t. i. *Oxford Reading Programme*. V svojih zlatih časih, v poznem 19. stoletju in na začetku 20. stoletja, je v njem sodelovalo mnogo prostovoljnih bralcev, med katerimi je bil velik odstotek angleških duhovnikov, nekaj plemičev, ki niso imeli boljšega dela, in vsaj en zločinski blaznež.⁵

Pred korpusi je torej navedke, na katerih je temeljil OED, prispevalo na stotine prostovoljcev. Zbiranje navedkov je dober način zbiranja podatkov, sploh podatkov o redkih in nenavadnih besedah. Toda bralci morajo znati presoditi, kaj je vredno zbirati. Nihče ne pošlje navedkov za vse možne rabe besed, kot so *come, go, up in down*.⁶ Čeprav je *Oxford Reading Programme* bil in še vedno je čudovit in občudovanja vreden projekt, je pri njem nujno prišlo do izkrivljanja pri izbiri gradiva. Ni mogel zagotoviti statistično zanesljivih dokazov o rabi. James Murray, ustanovni urednik OED-ja, se je v svojem predsedniškem nagovoru društvu Philological Society leta 1878 pritožil (takrat se je šele lotil velikanske naloge urejanja navedkov in zapisovanja slovarskih gesel na njihovi osnovi): »Za *abusion* imamo petdeset navedkov, za *abuse* pa manj kot pet.«

Bralci navadno zberejo navedke za nenavadne besede, kot je *triskaidekaphobia* (nerazumen strah pred številko 13) in za nenavadne pomene. Računalniki, na drugi strani, pa ne presojajo. Če računalniku damo nalogo, naj poišče vse pojavitve besede *of* v nekem korpusu, bo to v nekaj milisekundah naredil. Lahko jih razvrsti v konkordančni niz, ki mu pravimo tudi indeks KWIC (*key word in context* – ključna beseda v kontekstu), ki ga slovaropisec lahko vzorči in razišče, ter primerja vzorce, s katerimi je posamezna ključna beseda povezana. Murray bi v trenutku dojel pomen obsežnega elektronskega korpusa.

Še en vir besed so obstoječi slovarji. Slovaropisje nalaga, **kopiči**. Slovar nastane na osnovi drugega slovarja. Ne začenjamo vsi z ničle. Avtorje slovarja včasih obtožijo plagiatorstva drugih slovarjev, a če dobro pomislimo, je vsaka slovarska definicija majhna hipoteza. Noben znanstvenik ne bi objavil hipoteze brez posveta z deli

⁵ Zgodba o dotičnem zločinskem blaznežu je lepo opisana v knjigi Simona Winchestra. Dr. Minor je bil kirurg v ameriški državljanski vojni. Bil je shizofrenik in je trpel za paranoidnimi blodnjami. Po umoru moškega v Londonu so ga zaprli v Broadmoor, umobolnico za neprištevne zločince. Tam je mnogo let produktivno in učinkovito bral in zbiral navedke za OED in njegova bolezen ga ni pri tem prav nič ovirala.

⁶ *Come* – 'priti', *go* – 'iti', *up* – 'gor', *down* – 'dol'. Op. prev.

svojih predhodnikov. Torej je smiselno pogledati druge slovarje in ovrednotiti njihovo vsebino. Mehanično prepisovanje pa je seveda nesmiselno. Ravno v tem je nevarnost, kajti (v nasprotju s katerokoli drugo vrsto raziskovanja) slovaropisje zahteva, da slovaropisec pove nekaj o čisto vsem. V poznih večernih urah, ko je leksikografska ekipa v zaostanku z delom in si član ekipe želi obdelati le še eno geslo več, je prepisovanje resnična skušnjava. Toda tej nevarnosti se je treba pri spodobnih projektih zoperstaviti. Presoja definicij v obstoječih slovarjih v luči novih dokazov je eno, mehanično prepisovanje teh definicij pa nekaj povsem drugega.

Torej obstajajo trije viri za izbiro slovarskih gesel: branje navedkov, obstoječi slovarji in korpusni dokazi. Elektronsko preiskovanje korpusnih podatkov je do sedaj prineslo le omejeno bero novih besed in novih pomenov. Delno zaradi težavnosti odločanja, kaj šteje kot beseda ali pomen in delno zaradi dejstva, da je korpus le vzorec jezika. Iskanje po internetu se bo morda izkazalo za bolj produktivno, če bomo zmogli najti način, kako dosledno opredeliti obstoječi nabor in definirati, kaj šteje za »leksikalno enoto«. Kot smo videli, je prepoznavanje večbesednih izrazov s samostojnim pomenom naloga, ki je še posebej težavna. Ena izmed predlaganih metod je merjenje statistično pomembnih sprememb v frekvenci besed in kolokacij. Tako sva imela pred dvajsetimi leti, ko je bila korpusna lingvistika še v povojih, Ken Church in jaz priliko izmeriti nenaden porast uporabe besede *greenhouse* in opaziti nov par z njo povezanih večbesednih izrazov *greenhouse effect* in *greenhouse gas*.⁷ Ta dva izraza sta zdaj standardni gesli v enojezičnih slovarjih.

Napotki o rabi besed

Nekoliko pobljše si oglejmo potrebe uporabnikov slovarjev. Na začetku sem rekel, da ima slovar družbeno povezovalno vlogo, kar sicer drži, vendar le do neke mere. Obstaja nekaj dobrih študij o dvojezičnih slovarjih in slovarjih za tuje govorce, ne poznam pa nobene študije o uporabi slovarjev med domačimi govorniki, zato bo vse, kar bom povedal v nadaljevanju, precej nepreverjeno, temelji pa na odzivu tržnih oddelkov založb Collins v sedemdesetih in osemdesetih letih prejšnjega stoletja ter Oxford University Press v devetdesetih letih. Založniki slovarjev, ljudje, ki imajo večinoma v rokah finance in nadzirajo leksikografske proračune, pogosto vztrajajo, da mora biti novi slovar »prilagojen tržišču«. To je precej nevarno in je recept za skrajno konzervativnost, kajti dokler izdelek ni narejen, javnost – potencialno tržišče – nikakor ne more vedeti, ali si bo novega izdelka želela ali ne.

Zdi se, da pri angleškem jeziku ljudje slovarje v glavnem uporabljajo za preverjanje **pravilnega črkovanja**, na veliko žalost slovaropiscev, ki se tako trudijo z razlagami in slovnico. To skoraj gotovo velja za slovarje angleškega jezika, ki

⁷ *Greenhouse* – 'rastlinjak', *greenhouse effect* – 'učinek tople grede', *greenhouse gas* – 'toplogredni plin'. Op. prev.

nima fonetičnega zapisa, temveč je v pisavi mnogo nepravilnosti in posebnosti. Za slovenščino ne bi mogel reči. Tujcu se pisava zdi precej podobna izgovorjavi. Vi bi mi znali povedati, ali bi tudi Slovenci potrebovali napotke za črkovanje. Morda je problem vseeno manjši kot v angleščini. Po drugi strani pa v slovenščini morda **pregibanje besed** (morfologija) in **razlike med narečji** predstavljajo večji problem kot v angleščini. Slovaropisci morajo ponuditi napotke, kjer so pač potrebni.

Ljudje se na slovarje ne obračajo le po pomoč glede črkovanja in pregibanja, temveč tudi glede **pravilne rabe in izbire besed**. Se pravilno reče *uninterested* ali *disinterested*,⁸ je med obema kakšna razlika? Dandanes (na mojo veliko nejevoljo, priznam, a kot leksikograf moram biti objektiven) lahko slišimo ljudi reči: »*I totally refute that*«. V tem pomenu se glagola *refute* še do nedavna ni moglo uporabiti. Rečemo lahko »*I deny that*«, kajti *deny* je performativni glagol, kot npr. *promise*.⁹ Nekaj lahko zanikamo ali obljubimo zgolj z izjavo (ali vam bodo verjeli, je že nekaj drugega). *Refute* pa do nedavnega ni bil performativni glagol. Da bi lahko uporabili *refute* (v pomenu ovreči), moramo svoje trditve učinkovito argumentirati, ne samo nekaj performativno izjaviti. Zdaj pa se glagol *refute* uporablja kot sinonim glagola *deny*. »*I totally refute that*« v jeziku politikov pomeni, da »nočem priznati, da je res«. Slovar bi moral uporabnike opozoriti, da bi morali natančni pisci in govorniki še vedno razlikovati med pomenoma omenjenih besed.

Še en primer tega, kakšen tip napotkov bi slovar moral ponujati, so slovnici dopolnila. Se reče *bored with* ali *bored of*?¹⁰ Vse več ljudi reče »*I'm bored of that*«. Jaz, na primer, bi uporabil »*I'm bored with that*«. Toda moje osebne preference, tako kot preference kateregakoli strokovnjaka, sploh niso pomembne. Bistvo je, ali je prav ali narobe. Rabe vedno večjega števila izobraženih govorcev angleščine ne moremo ignorirati. Proti uporabi *bored of* ni logičnega argumenta. Izbira predloga poteka iz arbitrarnega konvencionaliziranega niza. Ta primer se razlikuje od drugih pogostih rab, ki jim lahko ugovarjamo na logičen način. Na primer: stavka »*He could of done it*« in »*He should of done it*«, čeprav pogosta, sta napačna.¹¹ Jasno je, da je pomožnik *have* tukaj nujen. Nadomestil ga je predlog *of*, ki je v hitrem govoru njegova enakoglasnica, kar je posledica neznanja slovnice.

Podobno je z *between you and I*, ki je še bolj pogost.¹² Taki rabi je mogoče ugovarjati na logični osnovi, ker angleški predlogi zaimkom določajo »predmetni« sklon (ang. *object-case*) (v nobeni verziji standardne angleščine ne boste slišali »*He gave it to I*« ali »*she came home with I*«). Obstaja povsem primerna oblika zaimka in to je *me*.¹³

⁸ *Uninterested* ali *disinterested* – 'ravnodušen'. Op. prev.

⁹ *I totally refute that* – 'to odločno zanikam', *I deny that* – 'to zanikam', *promise* – 'obljubiti'. Op. prev.

¹⁰ *Bored with* ali *bored of* – 'naveličan (česa)'. Op. prev.

¹¹ *He could of done it* – 'lahko bi naredil to', *he should of done it* – 'moral bi narediti to'. Op. prev.

¹² *Between you and I* – 'med nama rečeno'. Op. prev.

¹³ Dobesedno 'dal je jaz' namesto 'dal je meni', 'domov je prišla z jaz' namesto 'domov je prišla z mano', *me* – 'mene, meni'. Op. prev.

Napaka je posledica izumiranja slovničnega sklona v angleščini (obdržal se je le pri nekaterih zaimkih) v povezavi s hiperkorektnostjo. Zgodilo se je namreč sledeče: šolarje učijo, da je napak reči »*me and*« v vlogi osebk (npr. v stavku »*Me and my friends are going on holiday.*«), narave slovnične napake pa ne razumejo povsem, zato zaradi hiperkorekcije uporabijo »*and I*« namesto »*me and*« v vseh primerih, ne glede na sklon.¹⁴ Tako je »*Between you and I*« postala ustaljena formula med ljudmi, ki nimajo občutka za slovnični sklon. Postavlja pa se vprašanje, ali bi moral slovar take napake priznati kot del angleščine in če je tako, kaj naj o njih pove?

Še en primer so t. i. »razdeljeni« nedoločniki (ang. *split infinitives*). V angleščini že vsaj tristo let poteka živahna razprava – še posebej med ljudmi, ki so prepričani, da je angleščina pravzaprav zakrinkana latinščina – ali je sprejemljivo razdeliti nedoločnik z vrinjenjem prislova med nedoločniški prosti morfem *to* in glagol. Se sme reči »*to boldly go [where no man has been before]*«?¹⁵ Zoper to ni logičnega ugovora, a konzervativni samooklicani strokovnjaki taki rabi ugovarjajo. Slovarji bi morali podati sodbo.

Leksikalna in paraleksikalna vsebina

V nekaterih akademskih krogih je moderno razlikovati med »leksikalno sematiko« in »enciklopedičnimi podatki«. Toda navadni uporabniki enojezičnega slovarja tega ne razlikujejo tako natančno. Ljudje hočejo dobiti instantne informacije o kulturnih referencah in jih ne zanima, ali so ti podatki opredeljeni kot semantični ali enciklopedični. Našel bom nekaj primerov, po katerih sprašujejo angleški govorniki in pričakujejo odgovor v slovarju:

- »Kakšno je znanstveno ime za drozga?«
- »Je *scapula lopatica* hrbtenica ali ključnica?«
- »Katero je glavno mesto Čila?«
- »Zakaj se madrigalu reče madrigal?«
- »Kaj pomeni *nook-shotten*?« (Shakespeare govori o Angliji kot o otoku, ki je *nook-shotten*. Kaj to pomeni?)
- »Kaj je plenilec?« zanima ljudi. »Je pingvin plenilec?« (Saj lovijo ribe, ne?)
- »Kaj so *chinos*?«
- »Kaj je *ohm*? Kaj je *joule* in zakaj se mu tako reče?«
- »Je *aa* angleška beseda?«

Ljudje slovarje uporabljajo za *Scrabble*[®] (črkovna družabna igra) – nikar ga ne podcenjujte – in za križanke. Pravzaprav *aa* je angleška beseda. Označuje vrsto

¹⁴ *Me and* – 'jaz in', *me and my friends are going on holiday* – 'jaz in moji prijatelji gremo na počitnice'. Pravilen stavek bi se glasil: *My friends and I are going on holiday*. V stavku je zaimek »*I*« del osebk in torej v »imenovalniku«, v zvezi »*between you and I*« predlog *between* določa sklon – v angleščini le en, »object case« – temu, kar sledi, tam bi torej morala nastopati sklonska oblika zaimka »*I*« = »*me*«. Op. prev.

¹⁵ *To boldly go [where no man has been before]* – 'pogumno se odpraviti nekam, kjer še nihče ni bil'. Op. prev.

lave in je havajskega izvora. V vsakodnevnem govoru in branju ni ravno pogosta, pri *Scrabblu* pa je neverjetno uporabna.

Poleg tega ljudje hočejo, da je slovar avtoritativen popis njihovega jezika, čeprav se praši na njihovih policah in ga nikoli ne odprejo. Na polici ga želijo imeti za vsak primer, če bodo morali nekoč morda kaj preveriti v njem.

Če ljudje slovar vendarle odprejo, želijo včasih samo brskati po njem. Zato hočejo, da so v njem tudi zabavne besede. Naštel bom na primer nekaj izrazov za različne vrste zločincev iz različnih obdobj in virov: *cutpurse* (žepar, poulični tat, izraz je uporabil Shakespeare), *mosstrooper* (*mosstroopers* so bili roparji na meji med Anglijo in Škotsko v 16. in 17. stoletju). Kaj je *yegg*? Kaj je *snakehead*? Kaj je *tsotsi*? Kaj je *rudeboy*? Kaj je *grifter*? Vse to so izrazi za različne vrste zločincev z različnih koncev angleško govorečega sveta. In predvsem, bo rekel tržni oddelek, hočemo *nove besede*. »Kajti potem bodo novinarji pisali o našem slovarju.«

Korpusni dokazi in primeri

Korpus kaže, kako je beseda rabljena. Ne kaže neposredno, kaj posamezna beseda pomeni, priskrbi pa dokaze, na osnovi katerih je mogoče sklepati o pomenu. V časih, ko še ni bilo na razpolago korpusnih dokazov, sem bil snoalec in urednik prvih izdaj dveh obsežnih enojezičnih slovarjev angleščine (Hamlyna leta 1971 in Collinsa leta 1979). Nato sem imel to srečo, da so me leta 1983, v zgodnjih dneh korpusne leksikografije, imenovali za odgovornega urednika prve izdaje Cobuildovega slovarja, pri katerem sem sodeloval z Johnom Sinclairjem in drugimi. Odkrili smo, da mnoge posplošitve v slovarjih iz časa pred korpusi niso ustrezne, čeprav so možne. Reči hočem, da se ob primerjanju s korpusnimi dokazi niso najbolje odrezale. Zato sem nadvse goreč spreobrnjenec. *New Oxford Dictionary of English* (1998), pri katerem sem bil glavni urednik, je bil prvi (in zaenkrat edini) slovar za domače govorce angleščine, narejen na podlagi korpusnih dokazov in navedkov iz bralnega programa.

Korpusni dokazi so bistven vir podatkov za kolokacije in sintagmatiko, ki jih je potrebno preučiti statistično, da razumemo odnos med rabo in pomenom besede. To slovarju zagotovi strukturo ali okvir. V korpusih lahko odkrijemo vzorce rabe besed, vendar nam ti ponujajo zgolj namige, asociacije, nekaj verjetnega o rabi in pomenu, ne pa nekaj gotovega. Pravzaprav kažejo dlje od leksikografije, in sicer na potrebo po novem, na leksiki temelječem pristopu k lingvistični teoriji, a to je že tema, ki ne spada v današnje predavanje.

Bistvena lastnost naravnega jezika je, da je poln negotovosti. Razlog je v tem, da so kategorije naravnega jezika zgrajene okoli prototipičnih »najboljših zgledov« in imajo meje, ki niso ostro definirane, temveč zabrisane. Kar je za leksikografe seveda velik problem. V angleščini je preprost primer tega beseda *hazard*.

Prototipična raba glagola *hazard* je s samostalnikom *guess* v vlogi neposrednega predmeta.¹⁶ »*Hazard a guess*« je frazem, ki predstavlja več kot petdeset odstotkov rabe tega glagola v vseh korpusih, ki sem jih pregledal. Pomen fraze kot celote je 'nekaj izreči brez pravega prepričanja, da je resnično'. Prototipični neposredni predmet, *guess*, je samostalnik, ki označuje govorno dejanje ali nek pojem. Če si поблиžje ogledamo korpus (npr. *British National Corpus*), bomo našli vse vrste samostalnikov, ki označujejo govorno dejanje ali nek pojem. Ne le skoraj sinonimni *conjecture*, temveč tudi *inference*, *opinion* in celo *definition*.¹⁷ Najde se celo nekaj primerov, ko glagol uvaja odvisni in premi govor, kot v prvem in drugem primeru. Gre za mejne primere, slovnično nepravilne v večini variant ameriške angleščine.

1. I would **hazard** that the ratio of real balances to total private sector net worth is less than 1% ...
2. "My uncle," said Wendy, expanding further on her family, "was Provost of Dumfries; he had a rather odd name – 'Chicken'." "Not Hen Chicken?" I **hazarded**, as this humorous diminutive was part of my family mythology.¹⁸

Leksikograf je v takih primerih pred dilemo, ali predstaviti in razložiti prototipični primer (v tem primeru *hazard a guess*) ali zastaviti širše (*hazard something*), pri čemer obstaja nevarnost, da se bo izgubila tipična fraza in njen pomen. Za pričujočo dilemo ni ene same pravilne rešitve. Je stvar presoje in izbire, pri kateri je treba upoštevati verjetne potrebe uporabnikov slovarja. Uporaben kompromis je uporaba besede *typically*.¹⁹ Na primer v razlagi prvega pomena *hazard*: 'to state a proposition, typically a guess, without any great confidence that it is true' – 'nekaj izreči, tipično domnevo, brez posebnega prepričanja, da je resnično'.

Če se pri definiciji omenjenega pomena odločimo za širšo varianto, lahko tipično frazeologijo ponazorimo s pomočjo slovarskega zgleda. Iz tega in drugih razlogov bi se moral slovaropisec upreti skušnjavi izbire nenavadnih, inovativnih, ustvarjalnih, mejnih zgledov in namesto tega izbrati osrednje, tipične, običajne primere rabe, čeprav se mu bodo morda zdeli rahlo dolgočasni. Cilj, ki bi ga morali imeti pred očmi pri izbiri, bi moral biti ponazoritev običajne rabe, ne meja predstavljenih možnosti. Na žalost pa imamo ljudje iz neznanih razlogov prirojeno težnjo, da se osredotočamo na mejne primere in nenavadne rabe, zato je potrebno mlade leksikografe izuriti v izbiranju običajnih, morda celo dolgočasnih primerov.

Interpretiranje dokazov

Primer glagola *hazard* kaže, kako lahko korpusni dokazi nadgradijo in celo nadomestijo intuicijo. Pomen, ki prvi pade na pamet mnogim angleškim govorcem,

¹⁶ *Hazard* – 'tvegati', *guess* – 'ugibanje, domneva'. Op. prev.

¹⁷ *Conjecture* – 'ugibanje, domneva', *inference* – 'sklep, zaključek', *opinion* – 'mnenje', *definition* – 'definicija, razlaga'. Op. prev.

¹⁸ *To hazard that ...* – primer odvisnega govora z glagolom »*hazard*«, "Not Hen Chicken?" *I hazarded* – primer premege govora z glagolom »*hazard*«. Op. prev.

¹⁹ *Typically* – 'tipično'. Op. prev.

ko brez prednosti objektivnega dokaza intuitivno razmislijo o njem, je 'ogroziti'. Vsekakor ima omenjeni glagol tudi ta pomen, toda le v dvajsetih odstotkih rabe, če je verjeti korpusnim dokazom. Korpus spodbudi leksikografa, da prepozna jezikovna dejstva, ki »po občutku« niso tako očitna.

Primer, kako naj se ne uporablja dokazov in primerov, bomo po drugi strani našli v internetnem *Wiktionaryju*, v enojezični angleški verziji, kjer je glagol *hazard* razložen kot:

1. To expose to chance; to take a risk: *I'll hazard a guess.*
2. To incur or venture.²⁰

Kaže, da sta razlagi z manjšimi spremembami prepisani iz drugega slovarja, brez premisleka o tem, kako je beseda dejansko rabljena. Primer je pripet k prvemu pomenu, čeprav dejansko ponazarja drugega. Ne gre za osamljeno napako, pravzaprav je precej tipična za enojezični angleški slovar *Wiktionary*. Razlika med *Wikipedijo* in *Wiktionaryjem* je naravnost dramatična. *Wikipedia* je velik uspeh. Je antologija enciklopedičnih člankov, ki so jih napisali ljudje, ki menijo, da se spoznajo na določeno temo. Če se izkaže, da tega znanja nimajo in je prispevek napačen, potem – če je kakorkoli zanimiv za javnost – ljudje, ki res imajo znanje s tega področja, prispevajo boljšega ali novega. Model enciklopedije kot kolektivno spisane antologije se je izkazal za dobrega. Toda tega modela se ne da prenesti na slovar. Slovar ni antologija. *Wiktionary* je poln slabo prepisanih gesel, pogosto napačno razloženih in z napačnimi primeri. Za zanesljive podatke o besedah in pomenih nekega jezika morajo dokaze o dejanski rabi – korpusne ali navedke – interpretirati dobro podkovani in izurjeni ljudje na podlagi konsistentnih načel.

Pisanje razlag

Kot sem nakazal že v prejšnjem razdelku, je glavna skrb enojezičnega leksikografa oblikovanje posameznega slovarskega gesla, in sicer s pisanjem razlag, ki odražajo dejansko rabo, pri čemer izbira srednjo pot med upoštevanjem zgolj prototipičnih rab besede in upoštevanjem vseh možnih rab.

Na drugem mestu je tehnična točnost razlag. Pisanje razlag tehničnih izrazov je za enojezičnega leksikografa problem zase. Da bi izraz razumel in pojasnil – definicije bi morale stremeti tudi k temu, da pojasnjujejo, ne zgolj definirajo – mora dejansko uporabljati izraz, ki ga pojasnjuje. To ne velja le za znanstvene definicije, temveč tudi za druga področja, na primer šport. Vsak, kdor je kdaj igral kriket, ve, kako nerodne znajo biti v ameriških slovarjih razlage izrazov iz kriketa. Na primer, v enem izmed novejših ameriških slovarjev nam povedo, da je v kriketu *bowler* 'the player who throws the ball to the batsman' – 'igrallec, ki vrže žogo odbijalcu'. Ne

²⁰ Dobesedni prevod: 1. prepustiti naključju; izpostaviti (se) tveganju; tvegal bom domnevo. 2. tvegati ali drzniti si. Op. prev.

omenjajo obvezne iztegnjene roke, ki razlikuje »*bowling*« od meta igralca v polju pri kriketu (»*throwing*«) in od meta v baseballu (»*pitching*«). Gre za pomembno razliko, kajti športne rubrike angleških časopisov izven Severne Amerike pogosto vsebujejo sledeče stavke:

3. Brett Lee, Australia's answer to Shoaib Akhtar, is the latest fast bowler to be accused of throwing.²¹ – Simon Briggs, Daily Telegraph, London, 22. februar 2009.

Bralec, ki bi si skušal razložiti ta stavek s pomočjo ameriške definicije izraza *bowler*, bi bil prej zmeden kot razsvetljen.

Brez dvoma tudi Američani naletijo na podobno nerodne britanske razlage izrazov iz baseballa, ameriškega nogometa, ki mu pravijo kar nogomet, ali hokeja, ki mu mi pravimo hokej na ledu. Avtorji razlag morajo biti uporabniki izrazov, ki jih razlagajo, da so sposobni oceniti pomembnost določenih komponent pomena. Toda profesionalni uporabniki izrazov, ki niso izurjeni leksikografi, imajo ponavadi še večje težave z razlago in pojasnjevanjem, zato je potrebna interakcija med tehničnim svetovalcem – znanstvenikom ali športnikom, odvisno od primera – in leksikografom, ki je spreten pri razlaganju.

Pri nekaterih besedah pride do razkola med pomeni, ki jih uporablja znanstvena srenja, in tistimi, ki jih uporabljajo navadni ljudje. Kadar navadni ljudje rečejo »*wait a second*«,²² s tem ne mislijo »počakaj obdobje enako trajanju 9 192 631 770 period sevanja, ki ustreza prehodu med dvema hiperfinima nivojema osnovnega stanja atoma cezija 133«, kar je definicija sekunde kot osnovne časovne merske enote, kakršno so dorekli na Generalni konferenci za uteži in mere, na kateri se občasno sestanejo v Parizu pod okriljem *Système International d'unités* (SI enote). Resen slovar mora narediti oboje: podati znanstveno definicijo in pojasniti rabo v vsakdanjem jeziku.

Podobna težava nastane pri razlagah mnogih splošno razširjenih bitij in predmetov, npr. pri razlagi besede *pajek*. Oglejte si sledeč izpis iz slovarja *Oxford Dictionary of English*:

an eight-legged predatory arachnid with an unsegmented body consisting of a fused head and thorax and a rounded abdomen. Spiders have fangs which inject poison into their prey, and most kinds spin webs in which to capture insects. Order Araneae, class Arachnida.

pajkovec z osmimi nogami, nečlenjenim trupom, ki ga sestavljata glavoprsje in okrogel zadek. Pajki imajo v čeljusti strupno žlezo, s katero v plen vbrizgajo strup. Večina vrst splete mrežo, v katero lovijo žuželke. Red Araneae, razred Arachnida.

²¹ Brett Lee, avstralski odgovor na Shoaiba Akhtarja, je zadnji v vrsti hitrih bowlerjev, obtožen napačnega metanja. Op. prev.

²² *Wait a second* – 'počakaj trenutek'. Op. prev.

Prvi del razlage se trudi definirati – postaviti meje znotraj klasifikacije – in ne razložiti. Čemu omenjati »nečlenjen trup« in »glavoprjsje«? To dvoje je omenjeno, da bi ločili pajke od žuželk, ki pripadajo povsem drugemu razredu. Šele drugi stavek, ki ni del uradne definicije, se nekoliko približa razlagi. Primerjajte zgornje z razlago iz Cobuilda:

A spider is a small creature with eight legs that looks like an insect. Most types of spider make webs in which they catch insects for food.

Pajek je majhna žival z osmimi nogami, ki je podobna žuželki. Večina vrst plete mreže, v katere lovijo žuželke, ki so njihova hrana.

Očitno je, da Cobuild želi v prvi vrsti tujim govorcem podati razlago, ne znanstvene definicije. Definicije v enojezičnih slovarjih pa stremijo k podajanju obojega.

Z uporabo korpusov, kakršen je *British National Corpus*, ter s pomočjo orodij za korpusno analizo, kot je *Sketch Engine*, je možno izdelati lingvistični profil izrazov, kakršen je *pajek*:

- Znanih je več tisoč vrst pajkov (pajki lijakarji, pajki, ki gradijo mreže, ki predejo kroglaste mreže, ki jedo ptiče, talni, orjaški, ogromni, veliki, drobni, strupeni, črne vdove, kamelji pajki, avstralske črne vdove, zaklopničarji, volkci, belorepi pajki, rakovičarji, tarantele itn.).
- Nekatere vrste pajkov *lovijo plen*.
- Pajki *pičijo*.
- Nekatere vrste pajkov so *strupene*.
- Mnoge vrste pajkov *predejo mreže* iz močnih *svilenih niti*.
- Pajki *preživijo* sredi svojih *mrež*.
- Pajki *nadzirajo*, kaj se dogaja na njihovih *mrežah*.
- Pajki imajo *osem nog*.
- Imajo *tanke*, kosmate in glede na trup *dolge noge*.
- Pajki imajo *osem oči*.
- Pajki veliko časa *prebijejo* povsem *negibni*.
- *Gibanje* pajkov je *nenadno*.
- Pajki *lezejo*.
- Pajki *hitijo*.
- Pajki so *urni* in *spretni*.
- Pajki lahko *lezejo po stenah*.
- Mnogo ljudi je *groza (strah)* pajkov.
- Ljudje *pobijajo* pajke.
- Angleži se pogosto trudijo *prepoditi* pajke iz *kadi*.

Primer 1: *Korpusni lingvistični profil besede spider.*

Prepričan sem, da tak profil povzema to, kar večina ljudi – vsaj v Angliji – meni o pajkih. (V zvezi z zadnjo točko na listi hitro dodajam, da – čeprav v nekaterih britanskih korpusih obstaja rahlo izpostavljena povezava med leksikalnima

enotama *spider in bath* – v resnici ne mislim, da ta predstavlja nesporno dejstvo, ki ga je potrebno znanstveno raziskati.)

Usklajenost zaključenih skupin

Ko je s pomočjo korpusnih dokazov določen okvir za vse navadne besede, je potrebno vključiti še druge informacije. Načelo pokrivanja izrazov z vseh področij človeške dejavnosti, vključno s športnimi, vodi k še enemu načelu enojezične leksikografije, in sicer **usklajenosti zaključenih skupin**. Vsi izrazi v neki skupini – kemijski elementi na primer, ali človeški organi – bi morali biti definirani usklajeno, v podobnem slogu, ne glede na pogostnost, tudi če so nekateri člani skupine morda tako redki, da jih v korpusu sploh ni. Enako načelo drži za terminologijo dejavnosti, kakršne so kriket, snooker, curling in avstralski nogomet, ki so šele v zadnjem času postali pomembni, v glavnem po zaslugi televizije. Če ljudje, ki po televiziji gledajo snooker, slišijo komentatorja reči »*There is a possible plant here*,«²³ bodo morda v slovarju poiskali definicijo relevantnega pomena izraza *plant*. *The (New) Oxford Dictionary of English* ga definira kot 'a shot in which the cue ball is made to strike one of two touching or nearly touching balls with the result that the second is potted'.²⁴ Iz česar posledično izhaja, da mora imeti slovar zadovoljivo razlago pomena iz snookerske terminologije tudi za *cue ball in to pot*.²⁵

Torej, ko slovaropisec bere časopis ali gleda televizijo, ga pogosto manj zanima vsebina povedanega, kot to, **kako je bilo nekaj povedano**. Kakšne besede so uporabili? Bi jih bilo potrebno uvrstiti v slovar?

Urednik enojezičnega slovarja mora odločiti, kako daleč se bo šlo na tehničnih področjih. Bi bilo kot iztočnice potrebno vključiti *strobilus*, *strobila*, *strobilation*?²⁶ Domači govorci, ki ne poznajo omenjenih besed, bodo pričakovali, da so v slovarju. Toda slovar ni terminološka baza. Znanstvena terminologija, pa tudi terminologija tehnoloških dejavnosti je tako obsežna in specializirana, da večji del ne spada v slovar, temveč v projekte, kot je IATE (*Inter-Active Terminology for Europe*), terminološka podatkovna baza Evropske unije, v kateri zbirajo in hranijo tehnično težko razumljivo terminologijo s stipulativnimi definicijami.²⁷

Vse to skupaj to pomeni, da materni govorci kot uporabniki enojezičnega slovarja pričakujejo, da bo seznam besed v slovarju popoln, leksikograf pa mora najti način, da izpolni pričakovanja kljub dejstvu, da je popis živega jezika dinamičen,

²³ *There is a possible plant here* – lahko, da se bo igralec odločil za vrsto udarca 'plant'. Op. prev.

²⁴ Udarec, kjer udarna krogla zadene eno od dveh (skoraj) dotikajočih se krogel, tako da druga krogla pade v luknjo. Op. prev.

²⁵ *Cue ball* – 'udarna krogla', *to pot* – 'potopiti, poslati v luknjo'. Op. prev.

²⁶ *Strobil*, *strobile* – 'strobilacija' – vrsta nespolnega razmnoževanja. Op. prev.

²⁷ Stipulativna definicija je definicija, kjer si izberemo poljuben nov izraz, ki nam nato označuje nek poljuben pojem. Op. prev.

da so meje zabrisane in da se ves čas pojavljajo nove besede in izrazi – na novo izmišljeni ali izposojeni.

Novе besede

Večina leksikografskih založnikov ob vsaki novi izdaji večjega slovarja objavi knjižico z novimi besedami, da bi vzbudili zanimanje novinarjev. *Macmillan English Dictionary for Advanced Learners* (MEDAL) ni izjema. Je slovar za tuje govorce, toda založniki dobro razumejo prednosti reklame, zato so izdali knjižico, ki je zastoj, in vključuje »nove besede«, kot so *blogosphere* (blogosfera), *chav* (slabšalni naziv za agresivnega najstnika, ki običajno izhaja iz delavske družine), *air kiss* (poljub, ki ga pošlješ po zraku), *career gapper* (nekdo, ki si je privoščil premor v karieri), *Chelsea tractor* (veliko, močno vozilo, ki naj bi ga vozili ljudje iz precej bogatega londonskega predela Chelsea). *Chick lit* je literarna zvrst, napisana za mlade ženske, tipično namenjena branju med ležanjem na plaži; *civil partnership* običajno označuje homoseksualni zakon. *Designer baby* – no, prav neverjetno je, kaj vse se da dandanes narediti z genetiko.

Treba se je lotiti dela

Sestavljanje avtoritativnega enojezičnega slovarja za materne govorce je strah zbujajoča naloga. Vsak naravni jezik je sestavljen iz nešteto leksikalnih enot, od skrajno običajnih, kot so funkcijske besede in pomensko oslabljeni glagoli, do redkih tehničnih izrazov in besednih zvez. Glavni urednik mora izdelati jasno politiko glede vseh vprašanj, ki sem jih omenil v svojem predavanju, in še glede marsikaterih drugih, ter zagotoviti, da jo vsi, ki prispevajo k slovarju, dosledno upoštevajo. Ne le kot ekipa, temveč kot »en sam kolektivni avtor«. Celotne politike ni mogoče izdelati, preden se ne začne dejanska obdelava gesel. Na začetku je treba začrtati širše smernice, ki jih v teku projekta poglobljamo in prilagajamo v skladu z vprašanji, ki se pojavljajo.

Vestni pisci definicij so nagnjeni k temu, da si razbijajo glavo, kako bi čim bolj natančno zajeli pomen vsake besede. Toda pretirano travmiranje je kontraproduktivno. Včasih se zgodi, da je prva formulacija, ki jo leksikograf zapiše, povsem ustrezna, a si razbija glavo še naprej in definicija postaja vse slabša in vse manj razumljiva, ko skuša zajeti vse možne pomene. Strah pred napako je še en dejavnik, ki ljudi upočasni in ne prinese nobene opazne koristi. Na žalost tudi počasni avtorji, ki se bojijo, da bodo delali napake, naredijo ravno toliko napak kot tisti, ki delajo hitro v skladu z ustreznim načrtom.

Zato je potrebno pri sestavljanju slovarja postaviti nek sistem, znotraj katerega vsakdo geslo obdela po svojih najboljših močeh in gre hitro naprej. Sistem pravi: »Naj te ne skrbi, ali si naredil napako, nekdo bo pregledal, kar si napisal«.

Glavni urednik bo poskrbel za povratno informacijo ter popravek očitnih napak in naključnih spodrslijajev. Slovaropisje je timski »šport«, ekipa mora imeti neko strukturo; leksikografi morajo brati in preverjati delo svojih kolegov v kooperativnem okolju.

Medij

V sodobnem svetu se pojavi vprašanje, na katerem mediju izdati novi enojezični slovar. Tradicionalni medij knjige izpodriva internet. To pa v leksikografiji poraja pomembna vprašanja:

- Je internet prav tako varen in trajen medij kot tiskana stran? Bodo bralci v prihodnosti, na primer čez 500 let, lahko pogledali v elektronski slovar na internetu tako kot danes lahko bralec prelista staro knjigo v knjižnici?
- Je nov slovar možno izdelati kot investicijo znotraj kapitalističnega sistema? Ali pa morajo vse nove slovarje financirati centralizirane agencije, ki jih obvladuje vlada?
- V primeru, da slovar financira komercialni investitor, recimo založnik, kakšen je ustrezen poslovni model? Ljudje so se po zgledu *Wikipedie* navadili, da se da zastonj dobiti informacije, ki jih prispevajo prostovoljci. Toda tak model za slovar, kot smo argumentirali zgoraj, ni primeren.

Zaključek: dokazi in interpretacija

Velik sodoben enojezični slovar kateregakoli jezika ima v skupnosti pomembno vlogo. Osnovni predpogoj je leksikalna podatkovna baza, ki je sestavljena na podlagi analize korpusnih podatkov, ki pa hkrati odraža tudi odnos družbe do jezika. Korpus izpostavlja vzorce rabe besed, za terminologijo, nenavadne besede, imena, zgodovino besed, mnenja glede pravilnosti in drugo pa so potrebne dodatne raziskave. Toda raziskati je treba podatke, ne pa promovirati mnenj samozvanih strokovnjakov. V samem jedru leksikografije je torej korpus. Slovar mora odražati družbeni odnos do jezika in poskrbeti za napotke glede pomena in etimologije. Da pa bi bil avtoritativen, morajo vsa izražena mnenja v njem temeljiti na dokazih. O odnosu javnosti do pravilne rabe je treba poročati in ga ovrednotiti, ne pa pomesti pod preprogo. Naloga slovarja je, da poroča o vseh osrednjih, običajnih rabah in pomenih. Če slovar skuša pokriti vse predstavljive možnosti rabe, obstaja velika nevarnost, da bo postal nerazumljiv. Jezik bo pohodil leksikografa, ki ima prevelike ambicije, kajti prilagodljivost pomena in rabe besed je neskončna. Slovar skuša prikazati osrednji standard, ki ga upoštevajo govorci jezika v medsebojnih pogovorih, ne pa bujne fantazije jezikoslovcev, ki pridejo na dan z vsemi mogočimi mejnimi primeri.

*Prevedla Nataša Hrovat Jerlah
tgraycat@gmail.com*