
Priročnik praktične leksikografije založbe Oxford

Mija Michelizza

B. T. Sue Atkins – Michael Rundell, *The Oxford Guide to Practical Lexicography*, Oxford – New York: Oxford University Press, 2008, XII + 540 str.

V prvem delu se knjiga osredotoča na t. i. predleksikografijo: stvari, ki jih je treba vedeti, preden se začne sestavljati slovar. Avtorja opozarjata na opravila, ki se jih izvaja, in na vire, ki jih je treba zbrati pred začetkom same izdelave slovarja. V drugem delu priročnik kaže, kako v besedilnih virih iskati pomene iztočnic in kako beležiti leksikografsko relevantne podatke o vsaki teh enot. Tretji del pa predstavlja sestavljanje geselskega članka oz. redakcijsko delo tako za enojezične kot tudi za dvojezične slovarje.

The Oxford Guide to Practical Lexicography

The first part of this book focuses on “pre-lexicography”; that is, things that one should know before beginning to compile a dictionary. The authors draw attention to the tasks carried out and the sources that must be gathered before beginning actual production of a dictionary. The second part of the guide shows how to look for the meanings of headwords in text sources and how to record lexicographically relevant information about each of these units. The third part presents the composition of a dictionary entry and the editing work involved in both monolingual and bilingual dictionaries.

1 Uvod

Priročnik praktične leksikografije založbe Oxford *The Oxford Guide to Practical Lexicography* je razdeljen v tri sklope: prvi je t. i. predleksikografija, drugi del predstavlja analizo podatkov, v tretjem delu pa je podrobneje predstavljeno sestavljanje geselskega članka.

Prvo poglavje je (1) Uvod (*Introduction*, str. 1) z uvodnimi pojasnili avtorjev, drugo poglavje je naslovljeno (2) Tipi slovarjev in uporabniki slovarjev (*Dictionary types and dictionary users*, str. 17), sledijo poglavja (3) Leksikografsko pričevanje (*Lexicographic evidence*, str. 45), (4) Metode in viri (*Methods and resources*, str. 97), (5) Jezikoslovna teorija in leksikografija (*Linguistic theory meets lexicography*, str. 130), (6) Načrtovanje slovarja (*Planning the dictionary*, str. 160) in (7) Načrtovanje geselskega članka (*Planning the entry*, str. 200). V drugem delu sta poglavji (8) Gradnja leksikalne baze (1): pomeni besede (*Building the database* (1):

word senses, str. 263) in (9) Gradnja leksikalne baze (2): slovarski pomen besede (*Building the database (2): the lexical unit*, str. 317). V tretjem delu so še tri poglavja, in sicer: (10) Gradnja enojezičnega geselskega članka (*Building the monolingual entry*, str. 385), (11) Faza prevajanja (*The translation stage*, str. 465) in (12) Gradnja dvojezičnega geselskega članka (*Building the bilingual entry*, str. 484).

V nadaljevanju izpostavljam nekatere glavne poudarke in ideje priročnika. Knjiga je uvod v izdelavo slovarja – od začetnih faz (načrtovanja in zbiranja gradiva) pa do sestavljanja slovarja. Avtorja se zavedata, da je načinov za izdelavo slovarja več. Opisujeta način, ki je dobro funkcioniral za potrebe slovaropisnih skupin, v katerih sta sodelovala, in menita, da so opisane leksikografske tehnike v večini primerov neodvisne od jezika. Dodajata, da uporabniki slovar pogosto pojmujejo kot avtoritativni zapis, torej da obstaja mnenje, da gre za preskriptivno besedilo, vendar Atkinsova in Rundell pojmujeta slovar kot opis besedja, ki ga uporabljajo pripadniki neke govorne skupnosti, torej pojmujeta slovar kot del deskriptivne leksikografije. Zavedata se, da so vsi slovarji nepopolni.

2 Predleksikografija

Drugo poglavje se osredotoča na začetne faze načrtovanja slovarskega dela in na merila glede na tip slovarja. Podana je klasifikacija slovarjev: glede na jezik(e) slovarja (eno- ali večjezični), glede na zajeto besedje slovarja (splošni jezik, terminološki slovar ipd.), glede na velikost, glede na medij (tiskana, elektronska ali spletna oblika), glede na organizacijo slovarja (od besede do pomena ali semantične povezave z drugimi besedami), glede na uporabnikov jezik (uporabnikove jezike), glede na uporabnikovo strokovno znanje (jezikoslovci, otroci, odrasli ipd.), glede na to, za kaj uporabnik uporablja slovar – za dekodiranje ali za enkodiranje. Pri nastajanju novega slovarja je treba upoštevati dve načeli: prvič, da je prostor omejen in ga je treba zato razumno uporabiti, in drugič, da je slovar kot ekosistem: odločitev o vsebini, predstavitvi in obliki se ne da sprejemati izolirano, saj spremembe v enem delu vplivajo na preostanek slovarja. Postavlja se tudi vprašanje, ali je z nastopom elektronskih slovarjev obseg sploh še relevanten. V tej fazi je treba izoblikovati tudi »profil uporabnika« – opis tipičnega uporabnika slovarja z oceno njegovih potreb in jezikovnega znanja. Zavedati pa se je treba, da ni mogoče predvideti vseh uporabniških zahtev – podan je primer nogometnega zvezdnika Davida Beckhama, ki si je leta 2002 poškodoval metatarzalno kost (angl. *metatarsal bone*) – pridevnik *metatarsal* se je kar na vsem lepem začel pojavljati v vseh medijih, noben slovar pa takrat še ni vseboval tega leksema.

V nadaljevanju sledi predstavitev virov za sestavo slovarja: natančneje je obravnavano oblikovanje, zbiranje in procesiranje besedilnega korpusa za gradnjo slovarskega dela. Ko iščemo po besedilnem korpusu, moramo biti pozorni, saj mora biti tipična jezikoslovna lastnost, ki je zanimiva za slovaropisje, dovolj pogosta in hkrati dobro razpršena po različnih tipih besedil. V predkorpusnem obdobju se je slovarje delalo s pomočjo citatov oz. iztržkov iz besedil, ki pa so predstavljali le malo podatkov, očiten je bil tudi subjektivni element pri takšnem delu. Navkljub

vsemu se je treba zavedati, da je povsem enostavno z računalniškim programom zaznati nove besede (kot npr. *blogosfera*), vendar pa je velik delež novih leksemov še vedno večbesednih zvez. Pri ugotavljanju novih pomenov že obstoječih besed ima človek še vedno prednost pred računalniškimi programi. Za leksikografijo popoln korpus ne obstaja – pojavi se problem reprezentativnosti. Pri naravnih jezikih je težko definirati, kaj je to celotna populacija, saj populacija neprestano raste. Delna rešitev je stratificirano vzorčenje, vendar pa univerzalno sprejete klasifikacije besedilnih tipov nimamo. Avtorja menita, da je treba sprejeti dejstvo, da oblikovanje uravnoteženega korpusa nikoli ne bo znanstveno delo, saj oblikovanje idealnega vzorca zahteva preveč subjektivnih odločitev, tu pa imajo vpliv še finančni in praktični dejavniki. Vseeno pa nabor dobrih meril lahko vpelje uporabno tipologijo besedilnih tipov, čeprav je to eden kočljivejših problemov. Nove »hibridne« besedilne zvrsti na spletu še dodatno zapletajo problem, po drugi strani pa je treba poudariti, da ravno spletna besedila lahko »rešujejo« pogost problem vključevanja govornega besedila v besedilni korpus, saj se na spletu že najde marsikatera transkribirana oddaja ipd. Količine razpoložljivih besedil na spletu so mnogo prevelike za potrebe leksikografskega dela. Za leksikografijo je bolje, da splet vidimo kot vir besedil, iz katerega je lahko izdelan leksikografski korpus. Veliko korpusov je že nastalo delno tudi iz spletnih besedil. Prvi leksikografski korpus, ki je narejen samo iz spletnih besedil, pa je *The Oxford English Corpus* iz leta 2006. Seveda so viri s spleta tudi pomanjkljivi (malokaj npr. vemo o avtorju in o nastanku besedila), a vseeno nekatere raziskave kažejo, da so zadetki v korpusu *BNC* primerljivi z zadetki v korpusu spletnih besedil. Res pa je, da imajo spletna besedila še več šuma (npr. okvirje, naslove, slike, različne vrste seznamov ipd.) in je zato potrebno »čiščenje« tovrstnih besedil. Danes so tehnologije tudi na tem področju že napredovale in zato je tovrstno »čiščenje« do določene mere že avtomatizirano. V splošnem imajo leksikografi raje velike količine podatkov kot manjše, prečiščene korpuse. Končno je treba izpostaviti tudi problem avtorskih pravic – mnenje A. Kilgarriffa je, da so korpusi spletnih besedil manjša podzvrst iskalnikov, kot je npr. Google, in torej s tega vidika ne bi smeli biti problematični.

Glavne faze v procesu izdajanja slovarja so predstavljene v četrtem poglavju: od analize korpusa in nastajanja leksikalne baze (opisana je programska oprema tako za korpusno poizvedovanje kot tudi za pisanje gesel), pri dvojezičnih slovarjih je treba upoštevati še fazo prevajanja, do zaključenih slovarskih gesel. Predstavljen je slogovni priročnik, dokument, ki natančno opisuje geselske članke v slovarju in opozarja na prikaz kategorij, kot so npr. množinski samostalniki, deležniški pridevniki, (stalne) besedne zveze, izpeljanke, kratice in njihove razvezave, razlike v črkovanju (npr. med angleško in ameriško varianto).

Sledijo, po mnenju avtorjev, nekatere najuporabnejše ideje teoretičnega jezikoslovja za leksikografsko delo. Atkinsova in Rundell ponazarjata glavna pomenška razmerja med leksemi, ki so leksikografu lahko v pomoč pri njegovem delu, predstavljena je Fillmorova teorija semantičnih okvirov, ki je za leksikografa lahko zanimiva s stališča prikaza vezljivosti. Sledi tudi predstavitev Melčukovih leksikalnih funkcij idr.

V nadaljevanju izdelave slovarja se je treba odločiti, kaj vse bo slovar vseboval. Predstavljeni so tipi geselskih člankov, ki jih lahko najdemo v različnih slovarjih. Obravnavani so tudi izdelava geslovnika in različni tipi iztočnic – od enobesednih (enostavnih) iztočnic, krajšav, delnih besed (npr. pripone) do večbesednih zvez.

V poglavju Načrtovanje geselskega članka gre za obravnavo mikrostrukture slovarja, kjer sta predstavljeni tako oblika kot vsebina geselskega članka: med drugim poglavje obravnava, kako ločevati in kako hierarhično prikazati posamezne dele geselskega članka ipd. Razmah korpusnega jezikoslovja je vplival tudi na spremembe v strukturi slovarjev. Nekateri slovarji (npr. *Macmillan English Dictionary for Advanced Learners*), ki so nastali na podlagi korpusa, označujejo pogostnost določene besede (npr. z od nič do tremi zvezdicami), nekateri slovarji celo ločujejo pogostnost glede na to, ali se beseda pojavlja v govornem ali zapisanem jeziku (npr. *Longman Dictionary of Contemporary English*). V tem poglavju sta se avtorja dotaknila tudi razlik med elektronskim in tiskanim slovarjem.

3 Analiza podatkov

Prvi korak v procesu gradnje leksikalne baze je identifikacija in registracija pomenov večpomenskih besed in besednih zvez. Redkejša kot je beseda, manj možnosti je, da bo semantično in skladdenjsko kompleksna. Velja pa tudi obratno. Čim pogostejša je beseda, več je možnosti, da bo imela več pomenov. Slovarji pomene delijo in vsakega posebej označijo s številko, vendar če si pogledamo gradivo, vidimo, da meje pogosto niso tako jasne, kot so prikazane v slovarjih. V gradivu lahko opazimo, da med pomeni obstajajo različni odnosi: npr. različna pomena iste besede sta lahko popolnoma nepovezana (*punch* 'udarec s pestjo' : *punch* 'pijača'), beseda nastopa v različnih besednih vrstah, vendar izraža enako idejo (*laugh* (glagol) : *laugh* (samostalnik)), pomena sta (zgodovinsko) povezana, povezava med njima pa ni več jasna (*broadcast* 'sejati' : *broadcast* 'oddajati (na radiu)') itd. Kot teoretična pomoč iz jezikoslovja je predstavljena »prototipska« teorija pomena (angl. »*prototype*« *theory of meaning*) Eleanor Rosch. Avtorja opozarjata, da so slovarji tradicionalno razlikovali med polisemijo in homonimijo. Slednja se s stališča uporabnika sinhronnega slovarja ne zdi zelo relevantna, zato se tudi homonimne besede v nekaterih slovarjih ne pišejo več pod različnimi iztočnicami. Seveda je čisto drugače pri zgodovinskih slovarjih. V nadaljevanju so predstavljene še teorije D. A. Crusea s tremi koncepti, Apresjanova teorija regularne polisemije in druge. Rosamund Moon je prepričana, da »kontekst razdvoumlja«. Kontekst, v katerem se beseda pojavi, je najbolj zaupanja vreden indikator pomena. Ugotovljeno je bilo, da je kontekst, ki ga potrebujemo za razumevanje pomena, presenetljivo majhen. Indikatorje konteksta pa avtorja delita na zunanje (tematika besedila, narečje, subkulture ...) in notranje (skladnja ...). En slovar ima lahko pri določeni iztočnici štiri pomene in drugi šest pomenov, pa nobeden izmed njiju zaradi tega ne bo »boljši«.

Sledi registracija slovarskih pomenov – predstavljene so vrste podatkov, ki naj bi bile zabeležene, in predstavljen je način, kakor naj se te podatke vnaša v leksikalno bazo (prikazano s primeri iz korpusa in s primeri vnosov za angleščino). Pri gradnji

leksikalne baze obstaja geslo *When in doubt don't leave it out* (»Ko podvomiš, raje ne izpusti«). Ko ugotavljamo, kaj je iztočnica, je to v veliki večini primerov neproblematično, so pa primeri (npr. v angleščini samostalniki na *-ing*, množinski samostalniki idr.), kjer se zatakne in kjer morajo biti leksikografi še posebej pozorni. Opis pomena v leksikalni bazi mora biti dovolj razlikovalno opredeljen, ni pa potrebno, da je izoblikovana razlaga za v slovar. Splošno pravilo je, da naj bi vsak podatek, ki je vnesen v leksikalno bazo, imel svojo potrditev s primerom iz korpusa. In kakšni naj bodo ponazarjalni primeri? Kratki, kontekst naj pomaga razumeti besedo, v primeru naj ne bi bilo besed, ki so težje za razumevanje kot iskana beseda, primeri naj ne vsebujejo lastnega imena resnične osebe (žive ali pokojne), naj ne vsebujejo izposojenk, citatnih besed ali zvez ipd. Na tem mestu se pojavlja še vprašanje določanja osnovne (slovarske) oblike večbesednih zvez oz. vprašanje, pod katero iztočnico uvrstiti večbesedno zvezo. Tu obstajajo različne možnosti. Glede na nekatere raziskave naj bi govorci različnih jezikov iskali večbesedne zveze v slovarju na različne načine. Tako naj bi npr. Nemci pogosteje iskali določeno večbesedno zvezo najprej pri glagolu. Tudi to je pri sestavljanju slovarja treba upoštevati.

4 Sestavljanje geselskega članka

Avtorja obravnavata prehod iz leksikalne baze v enojezični slovar in razglabljata o možnostih za predstavitev in razporeditev različnih kategorij informacij, ki sestavljajo geselski članek. Z novimi tehnologijami in možnostmi elektronske oblike slovarja obstaja skušnjava, da bi vključili v slovar najrazličnejše vrste informacij, saj omejitev s prostorom nenadoma ni več. Vendar nima smisla vključevati tistega, za kar vemo, da za našega uporabnika ne bo imelo nikakršnega pomena. Avtorja opozarjata na ločevanje lastnosti, ki se nanašajo samo na določen pomen oz. na lastnosti, ki se nanašajo na celoten geselski članek. Slovnice informacije so v novejših slovarjih večinoma prikazane: kot prikaz vzorca (npr. *remember to do sth*), kot primeri povedi (npr. *Did you remember to lock the door?*) ali kot definicije (npr. *expire: When something such as a contract, deadline, or visa expires, it comes to an end or is no longer valid (COBUILD, 5. izdaja, 2006)*). Kako lahko besedilni korpusi vplivajo na predstavitev informacij v slovarju, je prikazano na primeru *promise someone to do something*, ki je v takšni obliki zabeležen v predkorpusnih slovarjih. S korpusom se je pokazalo, da se v takšni obliki le redko pojavlja, zato v naslednjih izdajah zasledimo obliko *promise to do something*. Največ pozornosti je v tem poglavju namenjene pisanju razlag. Samuel Johnson je predlagal termin razlaga (angl. *explanation*), vendar avtorja vseeno uporabljata v angleško govorečem okolju bolj uveljavljen termin definicija (angl. *definition*) – tule uporabljamo termin *razlaga* namesto ustrežnejšega prevoda v knjigi uporabljenega termina *definicija*, saj je *razlaga* v slovenščini ustaljen leksikografski termin –, čeprav nakazuje na nekaj dokončnega, kar pa slovarske razlage niso. Treba je razlikovati razlage za dekodiranje in enkodiranje – tudi glede na to se slovarji razlikujejo. Razlage morajo biti razumljive: jezik, uporabljen v razlagah, mora biti prilagojen uporabniku; če razlaga vsebuje besede, ki so večpomenske, naj te ne bodo v drugotnih ali netipičnih

pomenih; razlage naj bodo takšne, da uporabniku ne bo treba ponovno poizvedovati, da izve pomen prve iskane besede ali besedne zveze; besedilo in strukture naj bodo razumljivi in naj ne bi silili uporabnika, da se nauči posebnih konvencij za določen slovar. Zavedati se je treba, da včasih več (pri razlagi) ne pomeni tudi bolje. Treba je razlikovati med informacijami, ki so resnične, in informacijami, ki so relevantne za slovar. Če gre za uporabnika slovarja, ki je iz drugačnega socialnega okolja, je treba pojasniti določene sociokulturne pomene besede ali besedne zveze. Ko je dobesedni pomen razviden iz razlage, nima smisla dodajati kvalifikatorjev, kot npr. slabšalno. Že sama razlaga pove, kako se beseda ali besedna zveza uporablja. Posebno pozornost je treba posvetiti besedam in besednim zvezam, ki se nanašajo na etnično ali rasno pripadnost, telesno in duševno motenost, spolno usmerjenost, starost, spol. Kočljive so pri pisanju razlag v slovarju tudi besede s področij različnih verskih ali političnih prepričanj. Leksikograf se mora zavedati, da nevtralnost ni vedno mogoča. Pri razvrstitvi informacij v razlagi se postavlja vprašanje, kaj predstaviti prej. Treba je zadovoljiti uporabnika, ki si želi čim prej priti do iskane informacije – torej navadno pomena besede, ki ga ne razume. Slovarji so se začeli pojavljati tudi na mobilnih telefonih in kratkost razlag v tem mediju pride še kako prav. Predvsem za šolske slovarje se uporablja razlagalno besedje (angl. *defining vocabulary*). Pogosta napaka pri pisanju razlag je osredotočanje na etimologijo namesto na pomen.

Dvojezični slovarji so obravnavani posebej. Predstavljen je proces iskanja ustreznic v ciljnem jeziku za vse vrste izrazov izvornega jezika. Avtorja opozarjata, da obstajajo pomembne razlike med prevajanjem za slovar in besedilnim prevajanjem. Obravnavani so vsi pomembni dejavniki, ki jih morajo kasneje upoštevati uredniki dvojezičnih gesel. Predstavljeni sta uporabi korpusa ciljnega jezika in vzporednega korpusa za iskanje prevodov, s primeri iz angleščine in francoščine je tudi ponazorjeno, kako naj bi bilo to registrirano v leksikalni bazi. Ko je končana faza prevajanja, moramo iz zbranega materiala oblikovati dvojezični geselski članek. Na primerih različnih slovarjev so prikazani načini, kako se da razporediti informacije znotraj geselskega članka, in opravila, ki se izvajajo za sestavljanje geselskega članka (od različnih možnosti za prikaz prevodov do izbire primerov); na koncu so še predlogi strategij za pomoč uporabnikom, da izberejo primeren izraz za svoje potrebe v ciljnem jeziku. Pri tem je treba biti pozoren tudi na to, da so pomenska pojasnila namenjena predvsem govorncu izvornega jezika – govorec ciljnega jezika jih redko potrebuje. Ko je geselski članek končan, ga mora leksikograf ponovno prebrati, najbolje čez teden ali dva, ko se neobremenjen lahko prepriča, da je res razumljiv.

5 Zaključek

Priročnik *The Oxford Guide to Practical Lexicography* je sistematičen pregled že več desetletij uveljavljenih leksikografskih metod, dopolnjen z novimi možnostmi, ki jih ponuja hiter tehnološki razvoj – predvsem z vidika korpusnega jezikoslovja. Gre za širok pregled leksikografskega dela tako pri eno- kot tudi pri dvojezičnih slovarjih s ponazoritvami iz različnih že obstoječih slovarjev. Knjiga lahko služi kot dober uvod neizkušenim leksikografom, že izkušenim pa približa nove metode, brez katerih si danes leksikografskega dela ne moremo več predstavljati.