

KOMPJUTORIZACIJA LEKSIKOGRAFIJE¹; SEMANTIČKE RESTRIKCIJE ILI REVERZIBILNOST TEHNOLOŠKOG PROCESA

JURE ŠONJE

(Leksikografski zavod »Miroslav Krleža«, Zagreb)

UDK 03:681.3

Izvorni znanstveni članak

Primljen: 15. IX. 1995.

SAŽETAK. Oblikovanje leksikografske baze podataka hrvatskoga kao polaznoga jezika te ruskoga, engleskoga, njemačkoga, francuskoga, talijanskoga, španjolskoga i latinskoga kao ciljnih jezika radi tiskanja *Osmojezičnoga enciklopedijskoga rječnika* i mnogostruke znanstvene uporabe posebnih ispisa iz građe sadržane u cjelokupnoj leksikografskoj bazi podataka. Da bi se postigla oba cilja, rječnički se članak hrvatskoga jezika upisuje i ispisuje u osam segmenata – 1. leksem, 2. gramatička odrednica, 3. leksikografska odrednica, 4. sinonim/i, 5. definicija, 6. značenja i primjeri uz značenja, 7. primjeri, 8. izričaji, a stranih jezika u četiri segmenta – 1. ekvivalenti leksema, 2. ekvivalenti značenja i ekvivalenti primjera uz značenja, 3. ekvivalenti primjera, 4. ekvivalenti izričaja.

Tim leksikografskim softverom stvara se egzaktna metodologija stručnoga rada koja ne dopušta odstupanja, a računalni programi rade se po zahtjevima leksikografa ili drugih lingvista koji ovako strukturiranu bazu podataka mogu iskoristiti kao građu za svoja posebna istraživanja.

Međusobna uvjetovanost leksikografsko-računalnog postupka

Informatičari obično smatraju da računalo u leksikografiji ima funkciju multiciljanog pisaćeg stroja te ističu u prvom redu ekonomičnost računalne obradbe podataka, datoteke i pripreme magnetske trake za tiskanje knjige računalnim fotoslogom, a leksikografi koji nisu upućeni u rad s pomoću računala drže da računalo samo može obaviti velik dio poslova leksikografa. Teško je meritorno reći koliko je ekonomična primjena računala u leksikografiji, ali je rukovanje rječničkom građom unesenom u računalo neusporedivo lakše negoli onom pisanom na pisaćem stroju, pa je rječnik po svojoj naravi idealan za računalnu obradbu. To, međutim, nije predmet ovoga rada.

¹ Rad je napisan i prihvaćen za prvi broj *Lexisa*, neostvarenoga projekta znanstvenoga časopisa tadašnjega Jugoslavenskoga leksikografskoga zavoda. U novoj redakciji napravljene su prijeko potrebne adaptacije sadašnjemu vremenu. U računalnoj obradbi podataka *Osmojezičnoga enciklopedijskoga rječnika* došlo je u drugom svesku do stanovitih promjena, ali je stručno-leksikografski dio softvera sa strukturom rječničkih članaka i grafičkih znakova kao model *Rječnika* radi redakcijske pripreme njegova sadržaja za obradbu u računalu prema informatičkim programima isti kao i u prvom svesku, koji je opisan u ovom radu.

Prema našem iskustvu nije sama ekonomičnost računala u leksikografiji najvažnija u potrebi da se leksikografija od pisućega stroja i tiskarskoga sloga preorijentira na računalnu obradu podataka. Razlozi su u višestrukim mogućnostima što ih računalo pruža leksikografiji, koje se još posve ne mogu ni predvidjeti, jer se tek otkrivaju u eksperimentalnom istraživačkom radu.

Računalo, dakako, ne može obaviti stručne poslove leksikografa. Sva pamet stroja ograničena je na obradu pameti koju je u nju unio leksikograf, a informatičar popratio izradbom programa za obradu toga sadržaja. To znači da leksikograf mora u prvom redu egzaktno definirati globalni sadržaj i potanko razraditi predmet svoga rada da bi ga zajedno s informatičarom mogao formalizirati do tančina, jer računalo ne prima i ne daje ništa što nije u sustavu programa, a predviđene forme moraju biti takve i tolike da stručnoleksikografskoj obradbi pružaju mogućnost rješenja svakog zahtjeva struke na pretpostavljeno isti način. To znači da formalizacija rječničkoga članka mora sadržavati sve stručno i leksikografsko iskustvo te opća pravila uređenja sadržaja leksikografskog djela, i to da se stručnost obradbe mora jedinstveno osigurati za svaki rječnički članak posebno i za leksikografsku strukturu u cjelini.

Formalizacija strukture rječničkoga članka: Osmojezični enciklopedijski rječnik

Ako se rade redakcijske pripreme za kompjutorizaciju enciklopedija, leksikona i rječnika, onda je odmah jasno da leksikografska formalizacija za računalnu obradu podataka ovisi o strukturi izdanja koja se želi ostvariti, jer su strukture članaka enciklopedija, leksikona i rječnika vrlo različite. Struktura npr. biografskih enciklopedijskih članaka razlikuje se i po tipu biografija i po tipu leksikografskog djela, a segmentiranje strukture leksikografskoga članka mora biti potpuno i konzistentno da bi obradba jedinice i rad na svim fazama bili mogući i isplativi s obzirom na upis i ispis podataka. Tako npr. za *Opću enciklopediju* struktura biografske jedinice mogla bi imati ove segmente: 1. prezime, 2. transkripciju prezimena, 3. ime, 4. kvalifikativ, 5. datum rođenja, 6. mjesto rođenja, 7. datum smrti, 8. mjesto smrti, 9. tekst, 10. djela. Tih deset segmenata obuhvaća u načelu sve što sadržava biografija *Opće enciklopedije*, ali to nije sve što treba da bude formalizirano. Segmenti moraju dobiti i druge odrednice: 1. moguć ili obavezan broj članova, 2. grafičke znakove, 3. pravopisne znakove. Grafički i pravopisni znakovi moraju biti formalizirani u svim segmentima osim u tekstu, u kome se oni samo predviđaju kao mogući. Unos podataka svih faza redakcije teksta mora se raditi po riječima, rečenicama, grafičkim i pravopisnim znakovima, a ne u cjelini, pa tekst u tom smislu treba da bude i determiniran². Osim toga, tekstu se može odrediti najmanji i najveći broj redaka. Radi jasnoće leksikografskoga izražavanja rečenici se može odrediti i najveći broj grafičkih znakova. Ta ograničenja teksta leksikografi bi najteže prihvatili, ali bi dobro planiranje mogućih redaka teksta i grafičkih znakova rečenice odredilo približnu jednolikost leksikografskih članaka prema njihovu tipu i značenju i jasnoću leksikografskog stila, što bi pridonijelo stvaranju čvrste leksikografske strukture. Trebalo bi,

² Više godina nakon pisanja ovoga rada kompjutoriziran je *Hrvatski biografski leksikon*. Za taj posao – radi znanstvenoistraživačkih zahtjeva kojima mogu odgovoriti podaci sadržani u HBL-u – raščlanio sam strukturu biografskih članaka po ovom modelu, s tim da je struktura tog biografskoga članka tražila i segment bibliografije, ali informatičari Leksikografskoga zavoda nisu po tim segmentima kompjutorizirali *Leksikon*, tako da se cijeli biografski članak upisuje i ispisuje u cjelini, a ne po segmentima.

dakako, segmentirati strukturu svakog tipa enciklopedijskoga članka ovisno o njezinu sadržaju. Na takvu formalizaciju sadržaja *Opće enciklopedije* silila bi leksikografe računalna obradba podataka, što znači da bi se leksikograf u svom radu morao podvrgnuti radu stroja kome je odredio sadržaj rada, ali bi zauzvrat dobio egzaktan sustav rada i egzaktnu strukturu proizvoda. Valja napomenuti da je veći dio takve formalizacije proveden u *Općoj enciklopediji* u općim pravilima za pisanje članaka edicije, ali su u nekompjutorskom radu moguća odstupanja od strukture, a u kompjutorskom nisu moguća. *Opća enciklopedija*, kao i druge enciklopedije Leksikografskoga zavoda, radit će se s pomoću računala u nekom od idućih izdanja, no naše se iskustvo temelji isključivo na *Osmojezičnom enciklopedijskom rječniku (OER)*³.

Bivši Jugoslavenski leksikografski zavod, koji je počeo pripremu OER-a, odlučio se na računalnu obradbu podataka u ovom izdanju 1. zbog dugoročne orijentacije na primjenu računala i fotosloga; 2. zbog istraživanja mogućnosti elektroničke obradbe podataka u leksikografiji; 3. zbog goleme količine podataka što ih sadržava OER; 4. zbog pretpostavke da bi rukovanje tim podacima pisanim pisaćim strojem bilo teško, da bi se na tom poslu morao zaposliti velik broj ljudi i da bi se pojavljivale mnoge pogreške u podacima; 5. stoga što svaki podatak u rječniku sadržava manji broj znakova nego u bilo kojem drugom leksikografskom djelu. Svi su razlozi bili, dakle, praktične naravi, i na početku istraživanja – koje je trajalo oko dvije godine te je završeno iscrpnom leksikografsko-računalnom projektnom dokumentacijom kao osnovom za izradbu programa – nije se moglo pretpostaviti da će praktična vrijednost ovako koncipirane obradbe podataka biti višestruka. Ona se ne ograničava samo na proizvodnju *Rječnika* nego omogućuje vrlo široku stručnu i znanstvenu upotrebu podataka.

Stupac hrvatskoga jezika izborom leksema, strukturom rječničkoga članka i obradbom strukture rječničkoga članka određuje sadržaj *Rječnika*, koji tako ima karakter sedam dvojezičnih rječnika. Na temelju te konstatacije moglo bi se zaključiti da bi bilo dovoljno kad bi se građa tiskala u obliku sedam dvojezičnih rječnika, ali nije tako. Ne samo da osmojezičnim sadržajem *Rječnik* dobiva poliglotsku upotrebnu praktičnu, stručnu i znanstvenu vrijednost od polaznoga jezika prema svima ostalima, nego se i svaki od stranih jezika stavlja u odnos prema ostalim jezicima u svakom osmojezičnom gnijezdu, a u indeksu svaki ekvivalent označen brojem osmojezičnoga gnijezda dobiva funkciju leksičke jedinice kao polazni stupac, te se tako dobiva novih sedam dvojezičnih rječnika u kojima su strani jezici na lijevoj strani i svaki se ekvivalent-leksem svakog stranoga jezika nađe u odnosu prema svim ostalim leksemima, značenjima i ekvivalentima.

³ *Osmojezični enciklopedijski rječnik* prvo je kompjutorizirano izdanje Leksikografskoga zavoda. *Rječnik* se sastoji od ovih jezika: hrvatski, ruski, engleski, njemački, francuski, talijanski, španjolski i latinski. Na lijevoj je strani hrvatski jezik, a na desnoj sedam jezika koji su u prošlosti bili ili su sada svjetski jezici. Posljednji svezak sadržavat će abecedni indeks stranih jezika. Broj podataka koji će se tako dobiti ne može se ni predvidjeti ni izračunati sve do završetka rada na *Rječniku*, pa će odluka o opsegu indeksa ovisiti o dobivenom broju ekvivalenata stranih jezika, ali se već sada pouzdano zna da u indeks neće moći ući svi ekvivalenti stranih jezika, jer se svaka hrvatska natuknica i svako njezino značenje povećava u indeksu brojem ekvivalenata svakog stranog jezika i brojem stranih jezika. Redukcija će, prema tome, biti prijeko potrebna, tako da će unos ekvivalenata u indeks biti ograničen na određeni broj ekvivalenata svake natuknice i njezinih značenja.

Valja odmah napomenuti da na taj način ciljni jezici dobivaju ograničenu i nepotpunu funkciju polaznih jezika stoga što su njihovi podaci dobiveni u satelitskom odnosu prema polaznom jeziku, ali se ti podaci ipak mogu korisno upotrijebiti u projektima u kojima bi svaki od stranih jezika bio na lijevoj strani, a hrvatski jezik na desnoj. Tako složena struktura OER-a s višestrukim skrižanim vezama multiplicira obavijesti te višestrukom praktičnom, stručnom i znanstvenom uporabom opravdava golem stručni rad potreban da se projekt dovrši i da se kompenziraju velika početna novčana sredstva.

Baza podataka koja će se dobiti završetkom projekta bit će iskorištena ne samo tiskanjem OER-a nego i naknadnim tiskanjem zasebnih dvojezičnih rječnika. Osim toga malim naknadnim redakcijskim radom uz određene programe moći će se tiskati različiti tipovi manjih praktičnih rječnika za različite svrhe, što znači da baza podataka dobivena na ovaj način ima vrlo visoku uporabnu vrijednost. Da se *Rječnik* radi pisačim strojem, mogao bi se upotrijebiti jednokratno i svako dalje iskorištavanje dobivene leksikografske građe na ovom projektu tražilo bi i velik rad i znatna novčana sredstva.

Karakter dvojezičnih i višejezičnih rječnika polaznom jeziku određuje i namjenu i ulogu u rječniku. Polazni jezik, ovisno o praktičnoj namjeni, leksikografski registrira jezik u ograničenom totalu. Koncipiran je ambicioznije nego što je ostvaren u drugim dvojezičnim hrvatskim rječnicima: 1. po broju natuknica, 2. po stanovitom enciklopedijskom karakteru, 3. po složenosti strukture rječničkih članaka. Broj natuknica bio je globalno određen u projektom elaboratu, ali je u radu na redakcijsko-računalnoj projektnoj dokumentaciji proširen stoga što se želi dobiti veća količina podataka radi višekratne upotrebe. Tako se uz lekseme jezičnoga standarda i supstandarda više registriraju i regionalizmi. Stupac se još proširuje izborom riječi općega enciklopedijskog karaktera, pa tako sadržava i znatan broj historizama i stručnih termina. Poslužiti će tako i prosječnom i višem obrazovnom sloju. Prije prijelaza na računalnu obradbu podataka bilježio se i naglasak, ali se od toga odustalo na upozorenje informatičara da bi im težak problem bio programiranje našeg četveroakcenatskoga sustava, pa tako kompjutorizacija novoštokavskoga naglasaka ostaje otvorena za neki drugi projekt.⁴ Mi smo se zadovoljili time što donosimo velike količine drugih obavijesti.

Funkcija segmenata strukture rječničkoga članka

Rječnički članak mogao bi se upisivati u terminal kao jedna cjelina, ali bi tada nastale mnoge teškoće u radu. Podaci se unose s kartica, na kojima su stručno-redakcijski obrađeni, ali se na ispisima radi konačna redakcija i upotpunjuje čitavo slovo. Razbijanje strukture rječničkoga članka na segmente otvorilo je mogućnost višekratne uporabe baze podataka, te smo se odlučili za sustavno razbijanje strukture po funkciji svakoga segmenta u rječničkom članku. Sadržaj strukture bio je već određen, a njezino je segmentiranje bilo jednostavan posao kad su se shvatile mogućnosti upotrebe baze podataka koja se dobiva potpunim segmentiranjem strukture. Evo tih segmenata: 1. leksem, 2. gramatička odrednica, 3. leksikografska odrednica, 4. sinonim/i, 5. definicija, 6. značenja i primjeri uz značenja, 7. primjeri, 8. izričaji.

⁴ Novoštokavski naglasak kompjutoriziran je u projektu izradbe *Rječnika hrvatskoga jezika* Leksikografskoga zavoda i Školske knjige.

Potanko raščlanjivanje svakog segmenta iz projektne dokumentacije nije predmet ovoga rada, ali se o tome nešto ipak mora reći kako bi se shvatila i struktura i funkcija segmenta u strukturi i u računalnom sustavu, koji otvara velike mogućnosti uporabe. Prije toga valja konstatirati da su segmentima određene sigle za upis i ispis te grafički i pravopisni znakovi. Isto tako određena je i tehničko-grafička i pravopisna forma pisanja rječničkih članaka i osmojezičnoga gnijezda, u koju se programiranim automatizmom na kraju redakcijskog rada od pojedinačnih segmenata označenim posebnim znakovima stvara jedinstven tekst jedinice i gnijezda. To znači da svi koji rade bilo stručne bilo tehničke poslove moraju sustav poznavati u cjelini i u pojedinostima. To i jest najveća teškoća u radu: jer svako odstupanje od leksikografskoga i računalnog sustava remeti strukturu jedinice i *Rječnika*, a podaci koji se ne unose striktno po strukturi ne mogu se obrađivati u računalu. Kad se struktura u cijelosti shvati, onda ta teškoća odjednom postane prednost u radu. Uostalom, nema ni stručnoga ni bilo kakva kreativnoga rada u kome ne postoje mnoga ograničenja, a računalna obradba nameće još i dodatno ograničenje, koje se mnogim humanistima čini neprevladivim i smetnjom u stručnom radu. Oni koji sustav svladaju shvate odmah da im je ostala sva stručna sloboda i da kvaliteta njihova rada ovisi isključivo o njihovoj stručnoj sposobnosti i radu. Računalo je samo stroj koji stručnjaku služi u onoj mjeri u kojoj ga zna zaposliti i iskoristiti da sam ne radi one poslove koje mu on može obaviti programiranim automatizmom. U tom uzajamnom odnosu stroja i čovjeka stroju pripada samo mehanika, a čovjeku ostaje sve što je u tom radu umno, te o umnosti stroja i degradaciji stručnosti čovjeka nema ni govora. Sve operacije stroja predvidio je stručnjak stručnim sadržajem i računalnim rasporedom, koji informatičar programira za računalnu obradbu podataka.

Multiplikacija znanstvenog sadržaja

Da leksem i gramatička odrednica nisu razdvojene u dva segmenta, nego da se upisuju, vode i ispisuju kao jedan segment, što se moglo učiniti, dobivena baza podataka mogla bi se iskoristiti samo u ovom rječniku. Bilježenjem gramatičke odrednice svakom klosemu u obliku posebnoga segmenta mogu se na temelju odgovarajućih programa dobiti ispisi svake vrste riječi *Rječnika* posebno. Tako se npr. iz baze podataka mogu praviti ispisi samo imenica, ili samo imenica ženskoga roda, ili samo imenica muškoga i ženskoga roda, ili samo imenica muškoga, ženskoga i srednjega roda, ili samo množinski oblici po rodovima. Glagoli su još dragocjeniji podaci u bazi. Ako nekom znanstvenom radniku treba samo ispis glagola koji su i tranzitivni i intransitivni, ili refleksivni i recipročno-refleksivni, ili bilo koji drugi glagolski fenomen u jeziku s obzirom na gramatički podatak vrste riječi – može programiranjem dobiti ispis koji želi. Koliko znanstveni radnik time gubi vrijeme na prikupljanju podataka? Upravo ništa. Sav rad sastoji mu se u znanstvenoj obradbi podataka što ih je dobio. Kao što se prema tabeli oznake vrste riječi u svaku leksičku jedinicu unosi segment gramatičke odrednice sa svim njezinim posebnostima, tako se u segmentu leksikografske odrednice prema tabeli unose leksikografski i stručni podaci o leksemima, pa se posebnim programiranjem mogu ispisati npr. svi etnici, ili svi pejorativi, ili svi leksemi označeni jednom od niza stručnih kratica.

Mi, međutim, nismo naručili programe za znanstvenu upotrebu baze podataka, nego samo za jednokratnu uporabu, ali smo stoga stvorili mogućnost da se izrade ti programi,

tj. da ih prema potrebi naruče oni kojima budu trebali u radu. Svi koji se bave jezikoslovnim studijima mogu radi prikupljanja posebne građe iz baze podataka naručiti ispise prema vlastitoj potrebi.

Svim dvoznačnim i višeznačnim leksičkim jedinicama određena su značenja: 1. sinonimom/nimima, 2. kratkom definicijom. Premda se hrvatski stupac radi za višejezični rječnik, u kome se značenja hrvatske leksičke jedinice definiraju ekvivalentima na stranim jezicima, a ne hrvatskim jezikom, kao u jednojezičnom rječniku, sedmojezičnost desne strane nametnula je i uporabu jedinstvenoga definiranja svih značenja leksičkih jedinica lijeve strane, ali ne i jednoznačnih leksičkih jedinica, pa tako lijevi stupac ima i dvojezični/višejezični (po obradbi jednoznačnih riječi) i jednojezični karakter (po obradbi dvoznačnih i višeznačnih riječi). Budući da su značenja definirana na jedan od dva načina ili kombinacijom definicije i sinonima, i da se sinonimika navodi i jednoznačnim leksičkim jedinicama, a katkada po potrebi i definicija, lijevi stupac pruža ne samo ograničenu jednojezičnu upotrebu nego i mogućnost programiranja ispisa po sinonimici. Posebno se mogu npr. programirati ispisi svih europeizama koji imaju ili onih koji nemaju istoznačnicu hrvatskoga podrijetla. Isto tako radi praćenja semantičkoga kružnog toka mogu se programirati ispisi rječničkih članaka u kojima se javljaju naznačene leksičke jedinice i istoznačnice bilo samostalno bilo zajedno s definicijama ili samo s definicijama. Koliko ti podaci mogu biti dragocjeni i u leksikografskom i u znanstvenom radu ne treba ni spominjati. Neslućene su mogućnosti što ih može pružiti ovako segmentirana struktura rječničkoga članka.

Od osam segmenata strukture rječničkoga članka samo su dva obavezna: leksem i gramatička odrednica. Svi su ostali segmenti mogući. Ako se dogodi da se leksičkoj jedinici ne upiše jedna od predviđenih gramatičkih odrednica, računalo takvu strukturu ne prima. To je dakako nadzorni sustav što ga računalo vodi mjesto čovjeka, ali ono prima svaku stručnu pogrešku koja može biti unesena u taj segment ako je pogreška označena predviđenim znakovima tog segmenta. Računalo, dakle, nema niti pameti niti znanja da razlikuje pridjev od priloga i u tom smislu zna raditi nezgodne rasporede. Tako npr. priloge koji imaju pridjevski oblik može abecedirati ispred pridjeva ako nije posebno programiran da ih abecedira iza pridjeva.⁵

Definiranje značenja i semantičke restrikcije

Određivanje značenja leksičkim jedinicama nije samo najteži dio leksikografskoga rada nego i najsloženiji segment s mnogim upisnim oblicima. S obzirom na semantički karakter, značenja se dijele na jednostavna i složena. Kad su značenja bliska, definiraju se u složena značenja, kojih prema raspoloživom polju može biti do devet, a kad su udaljeni, određuju se kao jednostavna značenja, kojih isto tako prema raspoloživom polju može biti do devet. Budući da je polje ograničeno brojem devet, složenim se značenjima onim semantičkim strukturama koje imaju više od devet značenja osigurava prostor od devet puta devet, kako se ne bi moralo osigurati skupo polje od devedeset i devet. To, međutim, ima i svoju suprotnu stranu. Leksikograf je u svom radu ograničen brojem devet pa mora tražiti stručno najpovoljniji način u definiranju jednostavnih i složenih značenja, jer najsloženije semantičke strukture mora smjestiti u taj sustav, katkada i protiv svoje

⁵ Kompjutorizacijom naglasaka taj se problem lako riješio jer pridjev na kraju ima duljinu, a prilog nema.

stručne savjesti. Unatoč tom ograničenju, najsloženijim strukturama dana je mogućnost da budu obrađene u cijelosti. To se, međutim, ne može ustvrditi za izričaje, koji se iz istih razloga bilježe i unose samo u izboru od broja devet, i žrtvovanje mnogih izričaja i jest glavni nedostatak ovakve računalne obradbe podataka ne samo u *Rječniku* nego i u formiranju baze za druge potrebe.

Upisnih oblika značenja ima vrlo mnogo, onoliko koliko ih je bilo potrebno predvidjeti da se svaka leksikografska situacija riješi stručno najpovoljnije. Značenje zapravo ima funkciju jednoznačne leksičke jedinice, pa se upisni oblici prema potrebi formiraju s pomoću nekih segmenata, ili kombinacijama segmenata, npr. s pomoću sinonima, sinonima s oznakom vrste riječi, definicije, različitih forma leksikografske odrednice itd. U segmentu značenje javlja se i primjer u funkciji podsegmenta, ali je njegova upotreba u ovoj funkciji višestruko češća nego kod jednoznačnih leksičkih jedinica, gdje je primjer u funkciji segmenta. Ta formalna hijerarhija znači samo to da se u segmentu drugi segment može pojaviti samo kao podsegment, ili pak kao upisna forma segmenta. Jesu li segmenti u svim varijantama mogli ostati samostalni? Ta struktura bila bi jednostavnija za redakciju.

Segment *leksikografska odrednica* složene je strukture. U nju ulaze sve kratice i znakovi koji se u leksikografiji upotrebljavaju radi pobližeg označivanja leksičke jedinice ili njezina upućivanja. Obavijesti dobivene ovim segmentom odnose se na nekoliko jezičnih i leksikografskih slojeva i neotuđivi su sadržaj svakoga rječnika, a o egzaktном bilježenju tih podataka umnogome ovisi i kakvoća rječnika. Ti su podaci ne samo leksikografske determinante nego i sastavni dijelovi semantičke strukture jedinice. Ostručivanjem leksičkih jedinica bilježi se sloj koji na bilo koji način ima ograničenu uporabnu vrijednost ili čestotnost, a uputnicom se označuje supstandardni i regionalni (uz oznaku reg) sloj u jeziku, što omogućuje i opsežnije bilježenje tih slojeva. Sinonimika bilo standardnoga bilo nekog posebnoga sloja ne bilježi se uputnicom, a standardnim i stručnim tudicama navode se sinonimi hrvatske etimologije po standardnosti i frekvenciji. Ne upućuju se ni riječi posebnih socijalnih grupa i zajednica nego se bilježe posebnom oznakom. I znak izvedenosti nije samo grafički znak, kao ni uputnica, nego je vrlo čvrst element semantičke strukture, te se s pomoću njega određuju i značenja. Leksikografske odrednice funkcioniraju uvijek na isti način bilo da se bilježe kao poseban segment ili kao upisni oblik značenja.

I primjeri imaju istu funkciju kad se navode kao segmenti jednoznačne leksičke jedinice i kao podsegmenti značenja. To znači da su segmenti koji služe kao upisni oblici ili kao podsegment značenja u različitom hijerarhijskom odnosu samo iz programskih razloga te da je njihova leksikografska upotreba uvijek ista. Primjeri nisu obavezni, ali ih za razliku od naših dvojezičnih rječnika ima dosta. Primjeri se stavljaju kad god sastavljač lijevoga stupca vidi potrebu da 1. ilustrira, učvrsti ili upotpuni značenje ili čitavu višeznačnu strukturu, 2. ilustrira gramatičke podatke ili naznači rekciju, 3. registrira upotrebu ili karakterističan aspekt uporabe. Njihova je funkcija, dakle, višeznačna, i oni zajedno s izričajima daju život riječima, što inače nedostaje onim leksičkim jedinicama koje nemaju primjera i izričaja. Kao elementi semantičke strukture primjeri mogu sužavati značenja i uporabu kao uostalom i definicije, jer riječi u živoj uporabi imaju slobodniji prostor nego u leksikografiji, koja neprekidno čini stanovito nasilje nad jezikom da bi ispunila korisnu funkciju ne samo regulatora jezičnih fenomena nego i egzaktного semantičkoga i upotrebnog regulatora tih istih fenomena.

Problem reverzibilnosti tehnološkoga procesa

Odnos lijevoga stupca i desnoga ili desnih stupaca odnos je potpune zavisnosti desne od lijeve strane. U lijevom stupcu određen je ne samo čitav sadržaj *Rječnika* nego on ima dosta sadržaja koji ne postoji na desnoj strani, što ga čini autonomnim dijelom rječnika. On, dakle, ima višak obavijesti prema desnim stupcima. Mali rječnici mogu se raditi bez viška obavijesti lijeve strane jer je njihova uporaba ograničena na obavijesti o ekvivalentima vrlo ograničenog broja leksičkih jedinica lijevoga stupca, ali se u velikim rječnicima, koji nastoje što potpunije i što egzaktnije leksikografski snimiti jezik polaznoga stupca, ne može raditi tako a da se ne žrtvuje kakvoća i uporabna vrijednost rječnika. Tako npr. u ekvivalentima stranih jezika OER-a nema uputnica i ne prevodi se ni definicija ni sinonimika, nego se navode jednočlani ekvivalenti, a višečlani, koji mogu imati i oblik definicije, samo kad nema jednočlanih, kojih može biti do devet, tako da redosljed ovisi o semantičkoj bliskosti. Reducirano je bilježenje vrsta riječi i leksikografskih odrednica. Mogao bi se, dakako, raditi i takav rječnik koji bi i na desnoj strani bilježio one gramatičke podatke koji su na lijevoj strani, i ti bi podaci bili dragocjeni.

Struktura leksičke jedinice stranih jezika sastoji se od ovih segmenata: 1. ekvivalenti leksema, 2. ekvivalenti značenja i ekvivalenti primjera uz značenja, 3. ekvivalenti primjera, 4. ekvivalenti izričaja. Svi segmenti mogu imati leksikografsku odrednicu, a ekvivalenti leksema i ekvivalenti značenja još i gramatičku odrednicu (isključivo imenice). Predviđeni su i deskriptori kao autonoman segment stranih jezika radi dopunjavanja baze podataka stranih jezika možebitnim leksičkim jedinicama drugih rječnika kojima bi polazni bili strani jezici, ali se brzo uvidjelo da nemaju praktične vrijednosti jer se strani jezici ne mogu formirati kao cjelovite leksikografske baze podataka za uporabu na lijevoj strani. Stoga podaci stranih jezika mogu biti samo potpuniše iskorišteni u dvojezičnim projektima s polaznim stranim jezikom.

Svaki stupac radi se samostalno. U radu na bilježenju ekvivalenata stranih jezika polazni se stupac neprekidno stavlja u odnos prema sedam stranih jezika, zbog čega nastaju leksikografski problemi uzrokovani osmojezičnom kontrastivnom lingvistikom. Kod dvojezičnih rječnika ti problemi su ne samo sedam puta manji nego sedam puta sedam, jer problem koji se pojavi u određivanju ekvivalenta u jednom stranom jeziku ograničava ne samo autonomiju polaznoga stupca nego i mogućnosti prevođenja drugih jezika stoga što osmojezično gnijezdo mora biti jedinstvene strukture. Zbog toga je u redakciji osmojezičnoga gnijezda dopuštena – prema stvarnoj potrebi, a ne radi simplifikacije rada – ponovna redakcija leksičke jedinice početnog stupca. To je predviđeno i u sustavu računalne obradbe podataka. Redakcijski i kvalitativno najveći nedostatak višejezične leksikografije sastoji se u problemima višejezične kontrastivne lingvistike, ali je taj nedostatak zanemariv prema količini obavijesti i upotrebnoj vrijednosti takva djela. Kod jezika na kojima su objavljeni opsežni rječnici i leksikoni i ti su problemi manji, jer se lako nađu ekvivalenti. To su engleski, njemački, francuski i talijanski, a najveći su problemi u latinskom jeziku, jer je on mrtav jezik.

Čovjek, leksikografija i stroj

Rad na modernom rječniku sa svim civilizacijskim slojevima u jeziku traži polidisciplinarni rad sveden na jedinstvenu razinu – leksikografsku. Leksikografska obradba leksičkih jedinica koje u jeziku pripadaju različitim slojevima uvijek je jednaka. Nema dobre znanstvene definicije npr. u fizici koja nije ujedno i dobra leksikografska definicija, i obratno, a stupanj jasne leksikografske apstrakcije imanentan je semantičkoj apstrakciji samoga jezika. Računalna obradba leksikografskih podataka, koja leksikografa oslobađa nekih tehničkih poslova namećući mu svoj strogi sustav u radu, ne oslobađa čovjeka leksikografskoga rada od prikupljanja podataka do oblikovanja klasičke jedinice i rječnika. Računalu treba pripremiti svu građu i dati mu stručno završen posao kako bi ono moglo raditi planirane tehničke rasporede. Od njega se ne može dobiti ništa što mu se nije dalo, ali računalo i ne uskraćuje ništa od onoga što mu se zaista dalo. Čovjek i stroj u istom su radnom procesu na način da čovjek i dalje radi leksikografsko djelo akribijski i mukotrpno, gotovo onako kao što je radio i u srednjem vijeku. Isto toliko vrijeme mora utrošiti da obavi stručni dio posla, pa i više od toga. Stroj se javlja u funkciji struke dodatnim zahtjevima time što čovjeku nameće strogo poštovanje utvrđenoga sustava da bi mu zatim služio vlastitim operacijama multiplikacije efekata sustava, štedeći mu vrijeme za izvorni stručni rad. Leksikograf nema razloga da se boji računala, jer ga ono ne može istisnuti iz rada, niti mu smanjiti kakvoću proizvoda, niti mu oduzeti stvaralačku radost, ako on zna sve što treba znati da bi ispitao mogućnosti računala te ga punio podacima. Računalno postaje poslušno kad se njime ovlada. Otpor prema računalu ne spada u racionalno ponašanje. To je otpor prema promjeni vlastita ponašanja i strah od učenja novih metoda rada.

Može li se računalo upotrijebiti i u prikupljanju građe za rječnik? U toj funkciji ono bi umnogome olakšalo rad leksikografu, ali valja imati na umu da ispis sadržajno uvijek odgovara upisu, te bi ispisi mogli poslužiti samo kao izvori za prikupljanje leksičkih podataka, a tu funkciju imaju tiskani tekstovi.

Dosljednost tehničkoga sustava leksičkih jedinica i osmojezičnoga gnijezda računalo može osigurati u onoj mjeri u kojoj redakcija ispravno bilježi podatke, tj. u kojoj se u terminalu ispravno interpretiraju i upisuju te u računalu ispravno obrađuju. Budući da se upis radi u redakciji, svi leksikografski poslovi spadaju na rad organiziran u redakciji. Na tom poslu radi velik broj suradnika, svaki na svom dijelu posla, a rezultat uvijek mora biti jedinstven i stručno i tehnički. Svi koji rade moraju sustav razumjeti i dobro znati, ali se teško može postići to da dva čovjeka primjenjuju pravila na isti način a kamoli da to čini mnoštvo ljudi. Stoga je tehnička redakcija, koju moraju raditi leksikografi, prijeko potrebna kao i na svim drugim leksikografskim poslovima, što je još jedna potvrda da u stručnim poslovima računalo ne može zamijeniti čovjeka. Loše rađena tehnička redakcija može umnogome usporiti rad i dovesti u pitanje rezultate stručno dobro obavljena posla, a dobra tehnička redakcija primijenit će sustav u cijelosti i upozoriti na možebitnu stručnu nekonzistentnost primjena sustava.

Laici i neobaviješteni stručnjaci mogu misliti da se podaci negdje nađu, upišu, dobiju natrag i evo gotove knjige. Neznalicama je svaki posao jednostavan i lak, a samo iskusni znalci znaju i teškoće i težinu svakog posla. U računalnoj obradbi leksikografskih

podataka od početnog redakcijskog rada do redakcijskih gotovih proizvoda nema manje rada nego u pripremi leksikografskoga proizvoda neračunalnom metodom pripreme proizvoda. Ostaju sve redakcijske faze rada, sva kolacioniranja i sve korekture, samo nova tehnologija traži i novu tehnološku organizaciju, koja je također sastavni dio projektne dokumentacije, ali u ovom radu ima marginalno značenje. Računalna obradba podataka ne proizvodi, međutim, knjigu, nego osigurava tehnički sustav da se knjiga redakcijski dovrši.

Računalo, dakle, pripada redakciji u najširem smislu riječi. U leksikografsko-računalnom proizvodu uz redakciju rade još programeri, unositelji i obrađivači podataka. Programeri i obrađivači rade u računalnom sustavu. Ni lako ni brzo programeri ne mogu riješiti leksikografske i lingvističke složene zadatke, nego im je u svakom novom zadatku potreban istraživački rad. Informatičari imaju, međutim, metodologiju rada koja im omogućuje da riješe složene zadatke baš kao i leksikografi i lingvisti. Leksikografi i lingvisti nalaze se s informatičarima na istom poslu, kome pristupaju svaki sa svog stručnoga aspekta. Budući da je informatika potrebna i modernoj lingvistici i modernoj leksikografiji, dobro je kad računalni sustavi u svojim istraživačko-programerskim ekipama imaju lingviste. S lingvistima informatičarima leksikografi mogu istražiti mnoga neistražena područja i time pomaknuti informatičko-lingvističko-leksikografske spoznaje u praktičnom i teorijskom smislu.

I najsavršenije tehničko sredstvo funkcionira u onoj mjeri u kojoj funkcionira ljudski rad koji se služi tim sredstvom. U ovom slučaju to nije samo rad redakcije nego i rad programera i obrađivača podataka. Jedna od osnovnih funkcija računala sastoji se u potpunoj pouzdanosti pa se i rad mora organizirati tako da on uvijek bude doista pouzdan. Samo ljudski rad može tehničko sredstvo učiniti savršenim. To znači da leksikografija za svoje potrebe može optimalno iskoristiti elektroničko računalo u jedinstvenom radnom trojstvu leksikograf-računalo-informatičar tako da osigura nadzor metoda i efekta rada. Razumijevanjem takva karaktera i uloge računala u leksikografiji postaju zanemarivi humanistički otpori primjeni i svladavanju tehničkoga sustava.⁶

COMPUTERIZATION OF LEXICOGRAPHY – SEMANTIC RESTRICTIONS OR REVERSIBILITY OF THE TECHNOLOGICAL PROCESS

SUMMARY. The author discusses questions of meaning and defining within lexicographic structure, especially in the process of making dictionaries. The practical question is whether such a structure can survive when faced with the possibility of technological restrictions. Computerization brings advantages in a sense of multiplication of forms of lexical references and enables a kind of technological reversibility which demands precise lexicographic structure, whereby it contributes to the analytical efforts of its structuring.

⁶ U vezi s računalnom obradbom *Osmojezičnoga enciklopedijskog rječnika*, vidi također:

Gospodnetić, R., M. Šimunović, M. Đurović: *Upotreba računala u izradi Osmojezičnog rječnika*, Bled, 1982;

Gravis Delić A., M. Đurović: *Osmojezični enciklopedijski rječnik, oblikovanje teksta za fotoslog*, 7. međunarodni simpozij »Kompjuter na Sveučilištu«, Cavtat 1985., SRCE, Zagreb, 1985.