

## POGOSTOSTNI SLOVAR RUSKEGA JEZIKA\*

Pogostostni slovar ruskega jezika je sestavljen na podlagi izbora 1 056 382 besednih rab in obsega 39 268 besed iz besedil izbora. To je doslej najpopolnejši tovrstni slovar ruskega jezika; posvečen je spominu B. A. Larina.

Izdajanje prvih pogostostnih slovarjev sega v drugo polovico preteklega stoletja, ko sta nastala angleški in nemški slovar z namenom, olajšati tiskarsko tehniko in sistem stenografije, izpopolniti pravopis in metode pri poučevanju branja slepih<sup>1</sup>. Prvi »šolski« slovar pa je nastal šele l. 1911<sup>2</sup>. Prvi pogostostni slovar ruskega jezika je sestavil F. Mališ, sledili so slovarji H. H. Josselona, E. A. Štejnfeldtove, N. P. Vakarja — in obravnavani.

Temeljna naloga pogostostnih slovarjev je seznanjanje bralcev s stopnjo rabljenosti besed v jeziku. Ti slovarji so rezultat statističnih raziskav določenega besednega gradiva ter kažejo pogostost besed in količino besedil, iz katerih so besede vzete. Obravnavani slovar je rezultat dognanj, do katerih je prišla sovjetska leksikostatistika, tesno povezana s slovaropisjem in avtomatizacijo.

Pri nastajanju slovarja je sodelovalo okrog 100 ljudi s Filološke fakultete leningrajske univerze in iz Laboratorija za semiotiko Znanstvenoraziskovalnega inštituta za uporabno matematiko in kibernetiko gorkovske univerze. Pobudnik in vodja dela je bila L. N. Zazorina, ki je tudi obdelala teoretične osnove in praktična navodila za obdelavo besednega gradiva. Zazorina je napisala tudi načrt slovarja. Analitično slovnico in Uvod. Obdelavo gradiva na elektronskih računalnikih sta vodila V. A. Agraev in V. V. Borodin. Načrt za odpravo napak glosarja na elektronskih računalnikih sta napravila V. V. Borodin in V. M. Muratova. Poglavlje Statistični parametri slovarja je napisal E. V. Tisenko.

Slovar ima štiri dele: 1. Poročilo sestavljalcev; 2. Uvod; 3. Pogostostni slovar in 4. Dodatek.

V Poročilu se seznanimo s številnim timom znanstvenikov rusistov, matematikov in inženirjev, ki so sodelovali pri nastajanju slovarja.

Prva dva dela Uvoda govorita o namenu slovarja, njegovih virih, izboru računske enote in obdelavi besedil. Pomembna je ugotovitev, da v besedoslovlju do danes ni rešeno vprašanje razmejitev živega besednega sestava sodobnega ruskega jezika. Slovar je bil narejen zato, da se določijo meje aktivnega slovarskega sestava žive besedne rabe sodobnega izobraženca. Glavni namen slovarja je, dati zadovoljive podatke o sodobnem besedju glede na zvrstnostno razločevanje. Slovar zajema štiri vrste besedilnih virov: umetnostno prozo, dramatiko, znanstvena in publicistična, časopisna in revialna besedila ter kaže na različna področja, kjer bi ga lahko s pridom uporabili. Ta pogostostni slovar prvič na široko zajema znanstveno-publicistični in poslovni jezik, tako značilna za področje množičnega občevarja, sicer pa se ravna po normalih besednega sestava ruskega jezika, kakor jih je obdelal leksikolog B. A. Larin. Na podlagi njegove razdelitve na štiri samostojna področja besedne rabe<sup>3</sup> obsega slovar za vsako

\* Častotnyj slovar' russkogo jazyka. Sestavljalci V. A. Agraev, V. V. Borodin, L. N. Zazorina, V. M. Muratova, Z. V. Tisenko. Moskva, Russkij jazyk, 1977. 955 str.

<sup>1</sup> Inostrannye jazyki v škole, 1967, št. 3.

<sup>2</sup> R. C. Eldridge, Six Thousand Common English Words, Niagara Falls, 1911.

<sup>3</sup> Umetnostni, praktično sporočevalni, znanstveni in časnikarski jezik ter poslovni jezik.

skupino po 1/4, tj. 250 000 besednih rab. Ker avtorji nimajo dovolj zapisov iz praktičnega sporazumevalnega občevalnega jezika, so za obdelavo vzeli gradivo sodobne realistične dramatike in besedila javnega obveščanja, v katerih so prisotne tudi prvine občevalnega jezika. Poudarjajo, da poleg leposlovnega v občevalni jezik prodira tudi znanstveni slovar. Zato so v kartoteko virov vključili vidne sovjetske znanstvenike s področja fizike, kemije, biologije, mineralogije in zgodovine. Zajeli so tudi članke in govore Lenina, Kalinina in kongresno gradivo KPSZ, nenehno prisotno v tisku, sestavke pisateljev časnikarjev iz obdobja druge svetovne vojne (npr. A. N. Tolstoj, Leonov, Šolohov), od povojnih publicistov so zajeli članke N. M. Gribačeva. Viri za obdelavo časnikarskega jezika so bili časopisi Pravda, Izvestija, Sovetskaja Rossija (izdaja enega dne) idr.; vsak časopis so izpisali v celoti. Med revialnim tiskom so zajeli revije: Vokrug sveta, Nauka i žizn', Novoe vremja, Novyj mir in Junost'. Pri izpisovanju niso zajeli leposlovne proze, obdelali pa so reportaže. Slovar ni zajel jezika poezije in eksaktnih ved, zajema pa obdobje od Lenina in Gorkega do 60. let 20. stoletja. V vsaki izmed štirih funkcijskih zvrsti jezika je zajeto od 11 do 15 avtorjev. Tretji del Uvoda ima naslov Enote besedila in prvine slovarja.

Sledi Analitična slovnica slovarja. V nji so podana pravila za analizo samostalnikov, pridevnikov, števnikov, zaimkov, glagolov, prislovov, predlogov, veznikov, členkov in medmetov. Na koncu Uvoda so navedeni viri slovarja. Naj omenimo samo nekaj imen iz sodobne ruske besedne umetnosti: Antropov S. P., Gajdar A. P., Ovečkin V. V., Paustovskij K. G., Prišvin M. M., Sobolev L. S., Fedin K. A., Šolohov M. A., Arbuzov A. N., Bulgakov M. A., Lavrenev B. A., Rozov V. S., Simonov K. M., Štejn A. P. idr.

Tretji del obravnavanega slovarja, Pogostostni slovar, je sestavljen iz treh enot.

Abecedno pogostostni glosar je najvažnejši in najboljšežnejši njegov del. V njem je 39 268 različnih besed. Vsaka leksikalna enota glosarja ima naslednje količinske oznake: 1. pogostost: a) splošno pogostost glede na ves izbor; b) pogostost glede na število rab besede v vsaki izmed štirih funkcijskih skupin besedil (časopisno-revialna, dramska, znanstvena in publicistična, umetniško-prozna); 2. količino besedil glede na omenjene štiri skupine, v katerih je prisotna posamezna beseda.

Pogostostni glosar ima besede s pogostostjo 10 in več, vsega skupaj 9044 enot, razvrščenih glede na upadanje pogostosti. Najbolj pogost je *v* (*vo*) — 42 854, sledijo *i* — 36 266, *ne* — 19 228, *na* — 17 262, *ja* 13 839, *byl'* — 13 307, *čto* — 13 185, *on* — 13 145, *s* (*so*) — 12 975, *a* — 10 719, *kak* — 7425, *čto* — 6940 idr. Pogostost 100 imajo *vozrast*, *gromko*, *dym*, *želanie*, *zabota* idr., največjo pogostost 10, imajo *abscissa*, *bes*, *varen'e*, *denek* idr. Pogostost npr. 178 imajo *desjatok*, *krov'*, *otdel'nyj*, *pozдно*, *svoboda*, *xvata'*, *čuť*. V pogostostnem glosarju niso navedene besede s pogostostjo 1, skupaj 13 379 enot. Take besede so npr. *autozapodec*, *balkanskiy*, *vkleit'*, *gazetčik-professional*, *darmoed*, *židovski*, *zavod-avtomat* idr. Besed s pogostnostjo 1 je kar precej, preseneča pa, da so omenjene in še druge podobne v sodobnem ruskem jeziku tako redke.

V poglavju Statistična struktura slovarja je tabela s sedmimi kolonami. Najprej je zaporedno število stopnje, sledijo pa kolone absolutna pogostost, število besed z dano pogostostjo, nakopičena absolutna pogostost, nakopičeno

število besed (dva dela: absolutno in relativno število besed), relativne pogostosti in nakopičena relativna pogostost. Iz navedenih števil se dajo razbrati statistične informacije o vrstnem redu skupin besed z določeno pogostostjo, vsoto pogostosti posameznih besednih skupin idr.

V zadnjem delu *D o d a t e k* imamo seznama homografov, slovarskih in slovnčnih homonimov, popis nekaterih besed, ki niso zajete v slovarjih, npr. *zubator*, in vir, od kod je beseda vzeta. Iz Dodatka povzemamo, da je izbira besed dovolj reprezentativna in izraža lastnosti celote, tj. ruskega jezika. V tem delu imamo naveden tudi koeficient besedne raznolikosti slovarja. Količnik določa razmerje števila različnih besed (V) in števila vseh besednih rab (N):

$$C = \frac{V}{N} = \frac{39\,268}{1\,056\,582} = 0,0371.$$

Iz podatkov se vidi, da je veličina C odvisna od dveh faktorjev: od obsežnosti izbire besed in obsežnosti različnih besed, na katere vplivajo funkcijske in avtorske posebnosti izbranih besedil. Zaključni del navaja še štiri preglednice in pet diagramov. Iz 5. preglednice npr. razberemo pogostostne podatke za nekaj besed iz različnih funkcijskih zvrsti jezika: tako ima glagol *byt'* splošno pogostost 13 307, v časopisno-revialnih besedilih 2521, v dramatiki 5178, v znanstvenopublicističnih 2480 in v umetnostni prozi 3128. Tabela 4 govori npr. o pogostosti besednih vrst. Največjo pogostost ima samostalnik 26,65 — sledi glagol s 17,12 — najmanjšo samostalniško rabljeni pridevniki — 0,053.

Na podlagi obravnavanega slovarja so ruski jezikoslovci določili, katere besede se v vseh jezikovnih skupinah besedil uporabljajo največkrat. Število takih besed je 6440, kar predstavlja 868 577 besednih rab. Njihovo razporeditev po slovnčnih kategorijah prikazuje tabela 6. Zanimiv je podatek, da predstavljajo besede s frekvenco 10 in več 92 % vseh besednih rab. Diagram 2 kaže npr. količino različnih besed v odstotkih. Največ je besed z začetno črko P, najmanj pa na E in JU. Dodatek tudi govori o gradivu, ki ni zajeto v slovarju, a je že narejeno, tako: 1. abecedno pogostostna glosarja: znanstveno-publicistični in časopisno-revialni; 2. abecedno-pogostostni glosarji iz različnih besedil dramatike in proze; 3. abecedno-pogostostni glosar splošno rabljenih besed vseh štirih skupin besedil in 4. odzadnji slovar vsega zbranega gradiva.

Obravnavani slovar je pomemben zlasti zato, ker omogoča določevati rabo osnovnega ruskega besedja glede na različne vidike sodobnega slovaropisja. Prav bo prišel sestavljalcem učbenikov in slovarjev minimumov, s pridom ga bodo rabili pri strojnem prevajanju, transkripciji, prečrkovanju, avtomatičnem tisku, branju črkovnih besedil in še drugje. Morebiti je pomanjkljivost slovarja v tem, da besed ne razlaga, ne navaja naglasa, ne razločuje besed kot enot frazeološke in nefrazeološke rabe, kar pa je vse najti v 17-tomnem Slovarju sodobnega ruskega knjižnega jezika, ki ima 150 000 leksemov v najrazličnejši besedno-stilistični rabi. Pogostostni slovar ruskega jezika je nedvomno velik dosežek sodobne rusistike.

*Jurij Rojs*

Pedagoška akademija v Mariboru