
BENDRINĖS LIETUVIŲ KALBOS LEKSIKOS KOMPIUTERIZAVIMO NAUJOVĖS

Daiva Murmulaitytė

Lietuvos kalbos institutas, P. Vileišio g. 5, LT-2055 Vilnius

1. Prieš keletą metų pradėjus rašyti „Bendrinės lietuvių kalbos žodyną“ (toliau – BŽ), iš karto galvota ir apie kompiuterinį jo variantą (toliau – BŽ(k)). Buvo sukurtas ir kalbinei visuomenei pristatytas¹ BŽ kompiuterizavimo projektas, kuriame aptarta kompiuterinio žodyno struktūra – 3 skirtingos jo versijos, skirtos įvairaus lygio vartotojams ir numatytos tenkinti skirtingus jų poreikius, smulkiau aprašytas įsivaizduojamas vienos jų, vadinamojo darbinio BŽ(k), kūrimo kelias.

Tuo pačiu metu buvo pradėti rašyti bandomieji BŽ tekstai, mėginant iš karto diegti kai kurias naujoves, reikalingas tolesniam kompiuterizavimui. Dabar apžvelgiant per tą laiką padarytus darbus ir iškilusius sunkumus, galima teigti, kad viltys iš karto kurti abu BŽ variantus – poligrafinį ir kompiuterinį – dėl įvairių objektyvių ir subjektyvių priežasčių nepasiteisino. Viena svarbiausių yra ta, kad minėtasis darbinis BŽ(k) buvo įsivaizduojamas kaip be galo išsami duomenų bazė, numatyta naudoti iš esmės naujo BŽ ir (ateityje) tolesnių jo leidimų bei kitų žodynui rengimui. Daug dėmesio skirta visapusiškos duomenų paieškos šioje bazėje galimybėms numatyti. Be abejo, neįmanoma iš karto numatyti visko, kas gali parūpti mokslininkams kad ir po keliolikos ar keliasdešimt metų, kokie leksikos tyrimo aspektai juos tuomet domintų, juolab kad ir programinė kompiuterių įranga šiais laikais kinta itin sparčiai. Tačiau pradžioje buvo svarbu suregistruoti visus šiuo metu aktualius ir įmanomus tyrimo aspektus bei suformuluoti galimas paieškos užduotis. Akivaizdu, kad duomenų paieška bazėje turi būti labai įvairi, pačiu paprasčiausiu atveju kompiuteris turi ieškoti nurodytų žodžių – tiek antraštinių, tiek vartojamų definicijose ir iliustraciniuose pavyzdžiuose. Be to, jis turi „mokėti“ kaityti žodžius, kad, tarkim, skaitant kurio nors žodžio apibrėžimą ir norint pasitikslinti, ką reiškia jame vartojami žodžiai, arba tiesiog patikrinti, ar jie apskritai yra bazėje kaip antraštiniai, nereikėtų

¹ 1999 m. spalio 7 d. J. Jablonskio konferencijoje „Normų kaita ir vartoseną“ šio straipsnio autorės skaitytas pranešimas „Bendrinės lietuvių kalbos žodyno“ kompiuterizavimas“; taip pat žr.: D. Murmulaitytė, „Bendrinės lietuvių kalbos žodyno“ kompiuterizavimo klausimu, *Lituanistica*, 1999, Nr. 3(39), p. 92–98.

nustatinėti pagrindinių jų formų (vardažodžių vienaskaitos vardininko ar veiksmožodžių bendraties), užtektų tiesiog juos pažymėti ir nurodyti kompiuteriui ieškoti jų duomenų bazėje. Taip pat automatiškai turi būti sudarinėjami duomenų sąrašai pagal kalbos dalis, kirčiuotes, morfologinius, leksinius, sintaksinius, vartojimo sričių ir daugybę kitų požymių bei jų derinių. Visa tai reikėjo turėti galvoje rašant naujojo BŽ tekstą, kad jis būtų iš karto tinkamas perkelti į kompiuterinę duomenų bazę. Tačiau tai ganėtinai apsunkina patį BŽ teksto rašymą ir reikalauja vos ne antra tiek laiko tinkamam jo įforminimui – reikia nuolat galvoti ne tik apie knygą skaitysiantį žmogų, bet ir apie duomenis apdorosiantį kompiuterį, surašyti „nematomas“, t. y. tik kompiuteriui skirtas, pažymas, labai griežtai laikytis instrukcijoje numatytos žodyninio straipsnio struktūros reikalavimų ir kt. Taigi ilgainiui paaiškėjo, kad tradicinis BŽ rašymas ir vienalaikis jo rengimas kompiuterizavimui trukdo vienas kitam. Ir tai natūralu, nes tiek rengėjo, tiek vartotojo darbas su knyga ir su kompiuteriu vis dėlto iš esmės skiriasi, kritiniais atvejais bandant juos žūtbūt suderinti ir rasti abiem pusėms priimtinius sprendimus, neretai daugiau prarandama, nei laimima. Seniai aišku, kad kompiuterinės duomenų paieškos ir medžiagos tvarkymo galimybės yra nepalyginamai didesnės už tradicines ir stambesnės apimties nekompiuterizuoti darbai dabar jau sunkiai įsivaizduojami, tačiau kompiuteris reikalauja formos požiūriu duomenis pateikti maksimaliai tiksliai. Šis formalus tikslumas rengiant ir skaitant įprastą žodyno teksto variantą neretai trukdo, atrodo neracionalus ar net apskritai nepriimtinas, ir tai ima varžyti tiek BŽ teksto autorių, tiek poligrafinio BŽ vartotoją. Pavyzdžiui, vos ne kiekvienas žodyninis straipsnis papildomas minėtomis „nematomomis“ pažymomis, neretai jų prireikia ne vienos. Autoriui nėra lengva nepraleisti nė vienos reikalingos pažymos, tai blaško dėmesį ir atitraukia nuo pagrindinio teksto rašymo. Redaktoriui irgi reikia nuolat turėti galvoje, kad jos skirtos tik kompiuteriui ir poligrafinio BŽ skaitytojams jų nematys. Tarkim, vartojimo sričių pažymos naujajame BŽ dedamos daug dažniau, nei buvo „Dabartinės lietuvių kalbos žodyne“ (toliau – DŽ), tačiau jei vartojimo sritis aiški iš definicijos, tai jos pažyma yra „nematoma“, ji reikalinga tik kompiuterinei paieškai. Jei redaktorius nuspręstų pakeisti definiciją, atitinkamai turėtų patikrinti ir pažymą, galbūt po redagavimo ji turėtų būti matoma ir skaitytojui.

Šiuo metu atrodo tikslinga minėtus du darbus atskirti. BŽ ir „Bendrinės kalbos leksikos duomenų bazės“ (toliau – BLDB) kūrimas galėtų eiti dviem lygiagrečiais keliais, o ne vienu, kuriame nuolat išliktų nemaža tikimybė susidurti ir įstrigti. Be to, kurtina būtent išsami visos bendrinės lietuvių kalbos leksikos duomenų bazė, o dabar rašomas BŽ galėtų būti tik vienas jos komponentų, taigi šiek tiek laisviau galima būtų traktuoti jo struktūros dalykus, nes akivaizdu, kad BLDB turėtų apimti daug daugiau ir įvairesnės informacijos, jos struktūra būtų kita, ir BŽ duomenys į ją jei ir būtų perkeltami automatizuotai, tai tik po papildomo specialaus, kompiuterizavimo reikmėms skirto, redagavimo.

2. Kokia galėtų būti ši BLDB ir kuo ji skirtųsi nuo anksčiau projektuoto BŽ(k)? BLDB būtų skirta:
 - 1) leksikografinių duomenų kaupimui;

- 2) pagalbinės lingvistinės informacijos iš įvairių kalbotyros leidinių kaupimui;
- 3) duomenų tvarkymui ir moksliniam tyrimui;
- 4) įvairios paskirties ir apimties žodynų leidybai (sukūrus atitinkamas papildomas programas).

Iš esmės tai būtų pakoreguotas darbinis BŽ(k), pagrindinį dėmesį skiriant leksikografinės informacijos kaupimui. Atsisakius kitų dviejų anksčiau numatytų funkcijų (BŽ teksto rašymo ir redagavimo), naujosios duomenų bazės struktūra šiuo požiūriu turėtų būti gerokai paprastesnė ir lankstesnė, nes nebeliktų anksčiau planuotų penkių BŽ teksto variantų (autorius, redaktorių, vyr. redaktoriaus, svarstomojo ir galutinio). Taip būtų išvengta tam tikro kartojimosi, nes vieno žodžio duomenys būtų kaupiami ir laikomi vienoje vietoje – anksčiau buvo manyta, kad šalia pagrindinės išsamios duomenų bazės ilgainiui (parašius ir aprobavus visą BŽ tekstą) atsiras kita, mažesnė bazelė, apimanti visa tai ir tik tai, kas bus poligrafiniame BŽ variante.

Kalbant apie duomenų įvedimą (jei nebūtų rasta kitų, greitesnių ir racionalesnių, būdų) ir informacijos pateikimą vartotojui, atrodo tikslinga išlaikyti tą patį anksčiau numatytą kartotekos principą, kai visi tam tikro antraštinio žodžio duomenys (pagrindinės formos, morfologiniai, leksiniai, sintaksiniai, vartojimo sričių ir kt. požymiai, kirčiuotės, definicijos, vartojimo pavyzdžiai ir t. t.) tam tikra griežta tvarka išdėstomi jo „kortelėje“, t. y. kiekvienos rūšies duomenys turi joje savo vietą (lauką). Vartotojas gali skaityti „kortelę“ ir gauti visą jį dominančią informaciją apie tą ar kitą žodį. Be abejo, kompiuteris taip pat gali pagal tam tikrus požymius ar jų derinius grupuoti medžiagą ir pateikti vartotojui įvairius duomenų sąrašus, kuriuose būtų ne visa jo turima informacija, o tik ta, kuri aktuali vartotojui. Pavyzdžiui, fotografijos terminais besidomintis vartotojas galėtų pagal vartojimo srities pažymą fot. susidaryti visų šios srities žodžių sąrašą ir nurodyti kompiuteriui ekrane pateikti (ar į atskirą failą įrašyti) tik, tarkim, pačius terminus (paprastuosius ir sudėtinius), jų apibrėžimus bei sinonimus, jei kiti duomenys (pagrindinės formos, kirčiavimas, vartosenos pavyzdžiai ir t. t.) jam tuo tarpu nėra svarbūs².

3. Skirtinguose „kortelės“ laukuose būtų išdėstyti įvairių rūšių duomenys:

1. Homonimo ar homografo ženklas.
2. Leksinis vienetas (tolygus žodyninio straipsnio antraštei):
 - 2.1. žodis;
 - 2.2. sudėtinis terminas ar sudėtinis pavadinimas;
 - 2.3. frazeologizmas.
3. Pagrindinės antraštinio žodžio formos.
4. Semantinių grupių pažymos³.

² Čia galima būtų priminti, kad analogiškas galimybes teikė ir kompiuterinis DŽ variantas.

³ Šis laukas (jo struktūra, vieta bazėje, ryšiai su kitais laukais) kol kas kelia daugiausia neaiškumų. Tačiau neabejotina, kad jis yra būtinas. Tuo įsitikinta naudojantis kompiuteriniu DŽ variantu, kuriame semantinės analizės galimybė nebuvo realizuota ir tai gerokai apsunkino žodžių atranką ir sisteminimą rengiant BŽ.

5. Gramatinės pažymos:
 - 5.1. kalbos dalies;
 - 5.2. skaičiaus;
 - 5.3. giminės;
 - 5.4. linksnio;
 - 5.5. kitos.
6. Kirčiuotė(s).
7. Reikšmės eilės numeris.
8. Sintaksinės pažymos:
 - 8.1. junglumo;
 - 8.2. kitos.
9. Vartojimo srities pažymos.
10. Stilistinės pažymos.
11. Tarminės pažymos.
12. Daryba:
 - 12.1. pamatinis žodis;
 - 12.2. pamatinio žodžio reikšmės numeris;
 - 12.3. darybos kategorija.
13. Etimologija:
 - 13.1. etimonas;
 - 13.2. etimono kalbos pažyma;
 - 13.3. etimono reikšmė.
14. Definicija:
 - 14.1. tekstas;
 - 14.2. lotyniškas pavadinimas (zool., bot., gal ir anat., med. terminų);
 - 14.3. komentarai.
15. Nuorodos:
 - 15.1. kalbos kultūros ir pan. požiūriais teiktinesni variantai;
 - 15.2. dažniau vartojami žodžiai;
 - 15.3. kitkas.
16. Sinonimai:
 - 16.1. žodžiai;
 - 16.2. sudėtiniai terminai ir pavadinimai;
 - 16.3. frazeologizmai.
17. Antonimai:
 - 17.1. žodžiai;
 - 17.2. sudėtiniai terminai ir pavadinimai.
 - 17.3. frazeologizmai.
18. Iliustracijos:
 - 18.1. tekstas;
 - 18.2. vartojimo srities pažyma;
 - 18.3. paaiškinimai;
 - 18.4. komentarai;
 - 18.5. indeksai.

19. Vartojimo dažnio rodikliai⁴.
20. Indeksai⁵.
21. Pastabos iš kalbotyros skelbinių:
 - 21.1. eilės numeris;
 - 21.2. tekstas;
 - 21.3. autorius;
 - 21.4. šaltinis;
 - 21.3.1. leidinio pavadinimas;
 - 21.3.2. (periodinio) leidinio numeris;
 - 21.3.3. leidimo vieta;
 - 21.3.4. leidimo metai;
 - 21.5. pastabos komentaras.

Šis duomenų rūšių sąrašas kol kas, kol programiniai bazės kūrimo darbai dar nepradėti, tėra preliminarus. Dar galima pasvarstyti, kokie duomenys galėtų būti BLD bazėje, ir jį išplėsti. Tarkim, galbūt vertėtų išplėsti darybos duomenų laukų struktūrą ir pan. Arba gal būtų verta, kad kompiuteris akivaizdžiai rodytų, kiek reikšmių turi žodis, ir tam rodikliui būtų skirtas atskiras laukas. Duomenys jame turėtų keistis automatiškai, įrašius naują reikšmę ar išbraukus kurią nors buvusią, kai dėl įvairių priežasčių bendrinėje kalboje ji nebevartojama (pvz., paseno) ar nevertotina. Taip pat dar reikėtų labiau apmąstyti ir apsvarstyti kai kurių laukų (pvz., semantikos, sinonimų) vidinę struktūrą, vietą ir tarpusavio ryšius.

Kalbant apie informacijos, rastos įvairiuose kalbotyros skelbiniuose (21-asis laukas), kaupimą, minėtinas ir kitas šio klausimo sprendimo būdas: galima būtų tiesiog susieti BLDB su kompiuterine „Duomenų iš kalbotyros skelbinių“ kartoteka, pritaikyta ir naujiems žodžiams, jų vartosenos pavyzdžiams, definicijoms, neteiktinų žodžių pakaitams ir kt. informacijai kaupti. Tiesa, ši kartoteka dar nebaigta kurti – kol kas galima tik įvesti duomenis, o išsamios, įvairiapusės paieškos uždaviniai tik numatyti, bet dar nerealizuoti. Programiškai suderinus BLDB ir šią kartoteką, ateityje būtų galima dirbant su viena, naudotis ir kitos galimybėmis. Be to, minėtoje kartotekoje būtų daugiau ir įvairesnės informacijos nei viename BLDB laukų.

⁴ Šiame lauke būtų pateikiami keletas skirtingų šaltinių (tekstynų, dažninių žodynų) duomenys. Ateityje, kai tų duomenų bus pakankamai, kad būtų galima daryti apibendrinančias išvadas, žodžius (ar net atskiras jų reikšmes?) galima būtų žymėti atskirais indeksais (tūkstančių vienetų tikslumu), pagal kuriuos galima būtų orientuotis, ar žodis visuotinai vartojamas, ar retai pasitaiko. Panaši praktika taikoma kai kuriuose anglų kalbos aiškinamuosiuose žodynuose.

⁵ Šiame lauke būtų galima žymėti įvairius nenumatytus atvejus. Tarkim, išsirikęs žodžius pagal kuriuos nors anksčiau išvardytus požymius ar jų derinius, tyrėjas pageidautų jų sąrašą susiaurinti koku nors nauju aspektu. Kompiuteris negalėtų to padaryti, nes paieška pagal tą požymį nebuvo numatyta. Skaitant jo išrinktą sąrašą, dominančius žodžius ar jų reikšmes galima būtų žymėti savo indeksu, pagal kurį vėliau kompiuteris jau galėtų sudaryti ir siauresnį sąrašą. Atskirai turėtų būti sudaroma ir nuolat pildoma indeksų rodyklė, kuria remdamasis kompiuteris perspėtų, jei tyrėjas norėtų kitiems tikslams vartoti jau esantį indeksą.

4. Dar svarstytinas klausimas, ką laikyti pagrindiniu BLDB vienetu. Anksčiau manyta, kad atskirą „kortelę“ turėtų turėti kiekvienas antraštinis BŽ žodis. Tačiau projektuojant pačios „kortelės“ struktūrą, stengiantis sukurti universalią, visoms kalbos dalims tinkančią „kortelę“, neretai paaiškėdavo, kad parankiau „pildyti korteles“ ne žodžiams, o sememoms (atskiroms jų reikšmėms), jei žodis yra daugiareikšmis. Čia galima būtų priminti kitur jau ne kartą minėtą parankų pavyzdį: keturias reikšmes turintį žodį *kamšà*, kuris žymi veiksmo vietą (1. „žabais, virbais išklotas kelias“), įrankį (2. „kamšalas“, dar laikomą ir veiksmo rezultato pavadinimu), veiksmą (3. „kamšatis, spūstis“) bei veikėją (4. b. šnek. „kamšlys“). Taigi kiekviena reikšmė turėtų turėti vis kitą semantinę pažymą. Kitais atvejais visas daugiareikšmis žodis gali priklausyti vienai semantinei grupei, tad žymėti ta pačia pažyma kiekvieną reikšmę iki šiol, dirbant tradiciniais metodais, atrodė neracionalu.

Be to, kai kuriais atvejais BŽ (kaip ir DŽ, „Lietuvių kalbos žodyne“ (toliau – LKŽ)) gali skirtis to paties žodžio sememų giminė ar skaičius. Pvz.: *gyvátė* dkt. (2) 1. zool. ilgas, žvynuotas, lankstus roplys nuodingais dantimis, angis (*Vipera berus*) <...>; 2. bendr. niek. bjaurybė, nenaudėlis <...>; *íetininkas, íetininkė* dkt. (1) 1. sportininkas, ieties metikas; 2. vyr. ist. ietimi ginkluotas karys; *basútė* dkt. (2) 1. lengvas atviras vasarinis batas: *Baslės diržėlis nutrúko*. 2. dgs. tokių batų pora: *Nusipirkaū basutės*. Tokiais atvejais reikėtų numatyti gramatinių (giminės, skaičiaus) pažymų lauką kiekvienai reikšmei, o ne visam žodžiui.

5. Galutinis BŽ tekstas, prieš įvedant jo duomenis į BLDB, turėtų būti dar kartą peržiūrimas grynai formalaus medžiagos pateikimo požiūriu ir kartais atitinkamai pertvarkomas, nes kai kuriuos dalykus BLDB reikėtų pateikti kitaip, nei tai daroma BŽ.

1. BŽ bus straipsnių, turinčių kelis antraštinius žodžius. Taip bus pateikiami: a) kirčiavimo variantai: *kībiras* dkt. (3^b), *kibīras* (2); b) morfologiniai variantai: *ligónis, ligónė* dkt., *ligónis* bendr.; c) vyriškosios ir moteriškosios giminių vardažodžiai (ir įvardžiai): *darbiniŋkas, darbiniŋkė* dkt.; *gėras, gerà* bdv., *trėčias, trečią* sktv. klnt., *kurīs, kurī* įv. klaus. BLDB kiekvienas žodis turėtų savo atskirą „kortelę“, jo duomenys nebūtų maišomi su kitų žodžių duomenimis. Tad minėtais atvejais įvedant BŽ medžiagą analogiški duomenys, straipsnyje priklausantys visiems antraštiniais žodžiams (definicija, vartojimo srities pažymos, iliustracijos ir kt.), turėtų būti kartojami kiekvieno antraštinio žodžio BLDB „kortelėje“.

2. BŽ gali būti nepažymėtų homonimų ir homografų – pvz., tais atvejais, kai jie nėra pirmieji antraštiniai žodžiai (jei straipsnis jų turi kelis), pvz., *draūgas, draūgė* dkt. ir *draugė* dkt. arba *stiklīnis, stiklīnė* bdv. ir *stiklīnė* dkt.⁶ BLD bazėje moteriškosios giminės daiktavardis **1 draūgė** ir būdvardis **1 stiklīnė** turėtų atskiras korteles su savo definicijomis (šiuo atveju sutampančiomis su atitinkamų vyriškosios giminės

⁶ Tokių atvejų buvo ir DŽ bei kituose žodynuose, tačiau ten jie ne taip krito į akis, nes moteriškosios giminės būdvardžiai ir daiktavardžiai juose buvo pateikiami kaip atitinkamų vyriškosios giminės būdvardžių ir daiktavardžių pagrindinės formos, po tildės rašant tik dalį žodžio.

žodžių definicijomis), iliustracijomis ir kt. duomenimis ir būtų pažymėtos homonimu ženklu – arabišku skaitmeniu. Tik nuosekliai žymint visus homonimus ir homografus galima tikėtis, kad, suformulavus kompiuterinės paieškos pagal šiuos požymius užduotį, visi šios rūšies žodžiai bus automatiškai išrinkti ir prieinami tirti.

3. Rašant BŽ tekstą stengtasi nuosekliau nei iki šiol kituose žodynuose pateikti sinonimus, tačiau įvairovės išvengti vis dėlto nepavyko. Kai kurie jų dabar rašomi po definicijos ir žymimi vartotojui nematoma pažyma sin., kiti (teiktinesni, dažniau vartojami ir pan.) pateikiami kaip nuorodos be minėtosios pažymos, dar kiti rašomi pačioje definicijoje ir pan. BLD bazėje sinonimams turėtų būti skiriamas atskiras laukas. Jis turėtų būti nuosekliai pildomas sinonimų ir kt. šaltinių medžiaga nepriklausomai nuo to, ar tie patys žodžiai vartojami definicijose, nuorodose ir pan.

Tas pats pasakytina ir apie antonimus, kurie iki šiol aiškinamuose žodynuose retai vartojami kitiems žodžiams aiškinti. BLDB jiems taip pat turėtų būti skirtas atskiras laukas.

4. BŽ numatyta kai kuriuos vedinius, kurie DŽ dažniausiai buvo pateikti pamatinių žodžių straipsniuose (veiksmažodiniai priesagų *-imas / -ymas, -tojas, -ėjas, -ikas*, būdvardiniai priesagų *-umas, -ybė* ir pan. vediniai), pateikti ne aiškinant, o nurodant jų pamatinį žodį su atitinkamų reikšmių, kuriomis vedinys remiasi, numeriais. Tokia praktika buvo LKŽ, vienas kitas atvejais pasitaikė ir DŽ. Tačiau tokiu būdu dažnai iškreipiamas daugiareikšmio žodžio struktūros vaizdas. Pvz., kaip pirmoji reikšmė pateikiami visi „gryniesi“, neleksikalizuoti, veiksmo pavadinimai, kurie remiasi skirtingomis pamatinio veiksmažodžio reikšmėmis, pvz.: **kabaldāvimas** dkt. (1) → k a - b a l d u o t i 1–4; **tánkinimas** dkt. (1) 1. → t a n k i n t i 1, 3. 2. → t a n - k i n t i 4; *Betòno, gruñto tánkinimas*. Paskui jau kaip atskiros reikšmės gali būti rašomi sukonkretėję, specializuoti ir pan. pavadinimai. Taigi veiksmažodžio abstraktas turi daugiau reikšmių, nei rodo reikšmių numeracija jo straipsnyje, ir norint sužinoti realią padėtį, reikia skaičiuoti pamatinio veiksmažodžio reikšmių numerius. Be to, radus tinkamą iliustraciją, kai kurie transpoziciniai vediniai iš bendros jų grupės, pateikiamos kaip pirmoji reikšmė, yra išskiriami ir rašomi atskira reikšme bei iliustruojami. Taigi įvairuoja ir pačių „grynujų“ veiksmažodžių abstraktų pateikimas.

Kitas pavyzdys – pagal ankstesnių žodynų tradiciją, aiškinant žodį nuorodiniu būdu, daugiareikšmio nuorodinio žodžio reikšmių numeriai nerašomi, jei abiejų žodžių reikšmės sutampa. Tas pats pasakytina ir apie deminutyvų pateikimą, pvz.: **lélýtė** dkt. (1) dem. lèlè, plg. **lèlè** turi 3 reikšmes – 1. mažo vaiko (ppr. mergaitės) pavidalo žaislas; 2. figūra (ppr. žmogaus ar gyvulio) teatro vaidinimams; 3. šnek.[sin.]⁷ naujagimis, kūdikis. Kompiuteris tokiu būdu pateiktą deminutyvą traktuotų kaip vienareikšmį. Dar sudėtingiau, kai žodžio daugiareikšmiškumas slypi pamatinio žodžio, vartojamo vedinio apibrėžime, semantikoje, plg.: **legiònas** dkt. (2) 1. [ist.] stambus senovės Romos kariuomenės dalinys: *Cèzario legiònai*. 2. ypatingas kai kurių kraštų kariuomenės dalinys: *Svetimšàlių legiònas*; **legionièrius** dkt. (2) legiono karei-

⁷ Laužtiniuose skliausteliuose pateikiamos „nematomosios“ pažymos – poligrafiniame BŽ jų nebus. Šie duomenys būtų skirti tik kompiuterinei paieškai ir analizei.

vis; **legionininkas** dkt. (1) LEGIONIERIUS. Tiek *legionierius*, tiek *legionininkas* gali būti vartojami ir istorine, ir šių laikų kareivio reikšme, taigi iš tiesų abu jie yra dvireikšmiai⁸, o formalus jų pateikimas jokiū būdu to neatspindi. Tokie atvejai ypač sudėtingi todėl, kad, pirma, tokius pavyzdžius itin sunku išrinkti (kompiuteris čia menkas pagalbininkas), antra, kiekvieną jų nagrinėti ir semantinius ryšius tirti reikia atskirai.

5. Sudėtiniai terminai ir sudėtiniai pavadinimai BŽ pateikiami atskirais straipsniais. Tačiau didelė jų dalis, kaip DŽ ir kituose žodynuose, pateikiama ir kaip vartojimo pavyzdžiai vieno jų komponento straipsnyje, pvz.: **atmintinė** dkt. (2) <...> 2. [komp.] kompiuterinės sistemos dalis – elektroninis įrenginys, kuriame duomenys yra įrašomi, laikomi ir tvarkomi: *Vidinė atmintinė. Išorinė atmintinė*. Taip pateikus duomenis BLDB, nereikia tikėtis visavertės sudėtinių terminų ar sudėtinių pavadinimų paieškos, nes šios rūšies leksika įvairiais atvejais būtų įrašyta į skirtingus laukus. Taigi įvedant BŽ duomenis, iliustracijose pavartotus sudėtinius terminus ir sudėtinius pavadinimus, matyt, reikės papildomai žymėti atitinkamomis pažymomis, pagal kurias kompiuteris atpažintų čia esant ne iliustracijas, o būtent sudėtinius terminus ar pavadinimus. Galbūt bus rasta kokia kita išeitis, tačiau akivaizdu, kad tos pačios rūšies duomenys negali būti išskaidyti po kelis duomenų bazės laukus ir maišomi su kitos rūšies duomenimis.

6. Bene nuosekliausiai parengiamasis BŽ duomenų kompiuterizavimo darbas dirbamas vienu aspektu – „nematomomis“ ir „matomomis“ pažymomis žymint leksikos vartojimo sritis. Tačiau daugiau ar mažiau netolygumų vis dėlto išlieka, nes BŽ tekstą rašo ne vienas autorius ir vartojimo sričių pažymų dėliojimas nėra vienintelis ir svarbiausias dalykas formuojant žodyno straipsnį. Be to, neatmestina galimybė BLD bazėje vartojimo sričių žymėti gerokai daugiau. Tokiu atveju gali būti koreguojama ir iš BŽ imama medžiaga.

7. *Bendrinės lietuvių kalbos žodynas* bus norminis. Jame nebus neteiktinų žodžių ir reikšmių. Tačiau šią leksikos dalį planuojama atskirai pateikti kaip žodyno priedą. Tokiu būdu kai kurie daugiareikšmiai žodžiai būtų suskaidyti: norminės jų reikšmės būtų pateiktos pačiame žodyne, o taisytinės – jo priede. Tuo tarpu BLDB turėtų apimti visą bendrinės kalbos leksiką, taigi čia būtų kaupiami tiek norminiai, tiek nenorminiai dalykai. Bazės duomenys turėtų akivaizdžiai parodyti tokių daugiareikšmių žodžių struktūrą, iš jų galima būtų daryti išvadas apie norminių ir nenorminių reikšmių santykį, ryšius, vartojimo polinkius, dažnį ir t. t.

Tai, kas pasakyta apie BŽ pritaikymą BLDB, neturėtų būti traktuojama kaip BŽ teksto kraipymas ir kažkokio kito jo varianto kūrimas. Tiesiog BLDB galėtų būti pradėta kurti BŽ pagrindu. Laikui bėgant ji, be abejonės, būtų plečiama įtraukiant naujus, BŽ nepateiktus žodžius ar jų reikšmes, naujus, BŽ nebuvusius, vartojimo

⁸ Pastaruoju metu, beje, dažniausia yra trečioji *legionieriaus* reikšmė – „sandomas svetimšalis sportininkas“, tačiau V.Vitkausko ir V.Rudaitienės „Vakarų kalbų naujuosiuose skoliniuose“ (V., 1998) ji laikoma neteiktina (nors naujausias „Tarptautinių žodžių žodynas“ (Vilnius: Alma littera, 2001) ją pateikia).

pavyzdžius, prireikus koreguojant definicijas ir t. t. Joje būtų ir papildomos informacijos, pvz., vartojimo dažnio rodikliai, semantinių laukų pažymos, pastabos iš kalbotyros šaltinių su atitinkamais bibliografiniais pastarųjų duomenimis, įvairūs pagalbiniai indeksai ir kitkas, ko nebuvo BŽ ir kas gali būti aktualu kalbotyros specialistui, o ne eiliniam žodyno skaitytojui.

Be abejonės, BLDB vartotojui turėtų būti suteikta galimybė naudotis ir tokiu BŽ variantu, kuris bus išleistas spausdintu pavidalu, t. y. matyti ne BLDB kortelę, o įprastą žodyninį tekstą ir pan. Ir tai pasakytina ne tik apie BŽ, bet ir apie kitus jau išėjusius ar dar tik išleisimus žodynus, galbūt (tolimesnėje ateityje) ir kai kurias enciklopedijas. Kitaip tariant, BŽ neišskirtinas iš kitų žodynų, kurie palaipsniui turėtų būti nuskaityti ar kitu būdu įvesti į BLDB ir prieinami kompiuteriu dirbančiam leksikografo ar kitos srities kalbininkui. Itin svarbu būsimąją BLDB programos ir kt. požiūriais suderinti su dabar kuriamu terminų banku, kad jo informacija būtų techniškai ir kitaip prieinama BLDB vartotojui.

IŠVADOS

1. Šiuo metu tikslinga sukurti išsamią lietuvių „Bendrinės kalbos leksikos duomenų bazę“. Ji galėtų remtis tuo pačiu kartotekos principu, kuriuo anksčiau buvo numatyta kurti kompiuterinį „Bendrinės lietuvių kalbos žodyno“ variantą.

2. Šioje duomenų bazėje būtų kaupiami, tvarkomi ir mokliškai tiriami visokeriopi leksikografiniai duomenys bei kita su jais susijusi informacija.

3. Dabar rašomo „Bendrinės lietuvių kalbos žodyno“ (poligrafinio jo varianto) duomenys galėtų sudaryti vieną šios bazės dalį. Taip pat čia būtų kaupiami ir kiti duomenys, kurių minėtasis žodynas nesirengia pateikti, pvz., žodžių etimologijos, semantinių grupių pažymos, vartojimo dažnių rodikliai, taip pat nenorminė leksika ir kt.

4. Kai kuriais atvejais „Bendrinės lietuvių kalbos žodyno“ duomenys minėtoje bazėje būtų pateikiami kitaip (homonimai, sinonimai, sudėtiniai terminai, kai kurie daugiareikšmiai vediniai ir kt.), todėl, prieš įvedant juos į bazę, žodyno tekstas formos požiūriu būtų kiek pertvarkomas.

5. Kompiuterinės programos požiūriu „Bendrinės kalbos leksikos duomenų bazė“ turėtų būti suderinta su kitais dabar kuriamais ar dar tik atsirasiančiais elektroniniais leidiniais – terminų banku, įvairiais žodynais, enciklopedijomis, taip pat su „Duomenų iš kalbotyros skelbinių“ kartoteka.

Gauta 2002 05 10

Sutrumpinimai

BLDB – Bendrinės kalbos leksikos duomenų bazė.

BŽ – Bendrinės lietuvių kalbos žodynas.

BŽ(k) – kompiuterinis BŽ variantas.

DŽ – Dabartinės lietuvių kalbos žodynas, Vilnius, 1993.

LKŽ – Lietuvių kalbos žodynas, t. 1–19, Vilnius, 1941–1999.

Daiva Murmulaitė

**INNOVATIONS OF THE LITHUANIAN STANDARD VOCABULARY
COMPUTERIZATION**

S u m m a r y

The article suggests developing a comprehensive *Lexical Database of the Standard Lithuanian Language*, intended for accumulation, administration, and scientific research of lexicographical data as well as other related information. In future, it would be possible to prepare dictionaries of various purpose and scope on the basis of this database. *Dictionary of the Standard Lithuanian Language*, which is being developed, could make a part of the database. The database could be compiled on the basis of card file principle. In such a case, part of the data (homonyms, synonyms, polysemous derivatives, compound terminology, etc.) would be presented in a slightly different manner as regards the form. Besides, the database would contain a lot of other information, which will not be covered by the mentioned dictionary (lexical unit usage frequency indicators, etymology, semantic groups' references, etc.). As far as the computer application is concerned, the database should be consistent with other electronic lexicographical publications.