

Maija BALTIŅA, Daina VUCĀNE  
*Ventspils Augstskola*

## AUTORTEKSTS KĀ NACIONĀLĀ KORPUSA UN RAKSTNIEKA VALODAS VĀRDNĪCAS RESURSS

Nenoliedzami mūsdienās teju visās dzīves sfērās aizvien lielāka nozīme ir modernajām tehnoloģijām, tāpēc liekas tikai loģiski, ka arī valodniecībā aizvien lielāka nozīme ir jaunajām iespējām, ko piedāvā datori un datorprogrammas. Jaunās informācijas tehnoloģijas atstāj iespaidu uz tādām profesijām kā terminologs, tulks, tulkotājs, pasniedzējs, leksikogrāfs (Veisbergs 2003, 163). Šajā rakstā pievērsīsimies valodas korpusa lomai vārdnīcu sastādīšanā, kā arī rakstnieku valodas izpētes nepieciešamībai un rakstnieku valodas vārdnīcu nepieciešamībai Latvijā.

Jāpieskaras vienam no jaunākajiem valodniecības pētniecības virzieniem, jeb valodniecības apakšnozarei, kas analīzei izmanto elektroniski uzkrātu tekstu kopumu, proti, korpusa lingvistikai. Ar vārdu *korpus* šajā gadījumā apzīmēts pietiekami apjomīgs valodas datu kopums, kas aprīkots ar atbilstošu programnodrošinājumu un ko papildina lingvistiskās analīzes rīki (Baltiņa 2006, 15).

Korpuslingvistikā un datorlingvistikā par korpusu tiek uzskatīts reprezentatīvs rakstīta teksta vai transkribētas runas kopums elektroniskā formā, ko izmanto valodas (dialekta, valodas stila) analīzē un aprakstā.

Ieskatoties korpusa lingvistikas attīstības gaitā, iespējams izdalīt trīs posmus šī jēdziena izpratnē:

- 1) 20.gs. 50. gadi, kad ASV valodnieki nāca klajā ar atzinumu, ka valodniecībā pētījumu pamatā jābūt apjomīgiem valodas datu vākumiem. Tie pārsvarā sastāvēja no manuāli veidotām tekstu kolekcijām. Taču N. Čomskis šāda veida korpusus uzskatīja par „melīgiem”, jo tajos kā nevērtīgi tika atnesti triju veidu valodas dati – tie, kas ir pašsaprotami, tie, kas ir neskaidri un tie, ko pieņemts vērtēt kā nepieklājīgus.
- 2) Otrā posma korpusa lingvistika attīstās datorizētā vidē un rit saskaņā ar ar jaunu uzdevumu – veidot korpusu kā runātās un rakstiskās valodas datu vākumu, pēc R. Kvirka 1959. gadā izstrādātā plāna britu angļu valodas korpusa *Survey of English Usage* (SEU) izveidei.

3) Trešais posms korpusa lingvistikā aizsākas 1975. gadā, kad Lundas universitātē J. Svirka vadīta valodnieku grupa aizsāk runātās angļu valodas datorizētu apstrādi, izmantojot SEU datu bāzi (Baltiņa 2006, 15).

Ko tad īsti dod šāds valodas korpuss? Korpuss ļauj pētīt *īstu* valodu, proti, valodu, kāda tā tiek patiesībā lietota. Tajā tiek pārstāvēti visi valodas līmeņi, ne tikai literārā un zinātnes valoda, tāpēc tā ne vienmēr ir gramatiski pareiza, ne viss pakļaujas pieņemtajām valodas normām. Korpuss arī ļauj atklāt lietas, ko bez programmatūras palīdzības ir grūti noteikt un pamanīt. Tas ļauj sarežģēt realitāti pretstatā pieņemtajai patiesībai (Chafe 1992, 82).

Apstrādājot tekstus ar speciāliem lingvistiskās analīzes rīkiem un tos indeksējot, iespējams iegūt papildus informāciju, proti, vārdformu indeksu, vārdformu lietojuma biežuma kvantitatīvos datus, vārdu savienojumu analīzi, gramatisko attiecsmju analīzi u. c.

Neviens korpuss nesaturēs visu informāciju par cilvēkus interesējošiem valodas jautājumiem, taču, jebkurš korpuss, neatkarīgi no tā lieluma, palīdzēs iegūt faktus, ko nekādā citā veidā nebūtu iespējams uzzināt” (Fillmore 1992, 35).

Korpusa izveide nenoliedzami ir dārgs un laikietilpīgs process, tāpēc nepieciešams, lai tā izmantošanas iespējas attaisnotu ieguldīto darbu un līdzekļus. Minamas vairākas korpusa izmantošanas iespējas, piemēram, gramatikas un citu valodniecības jomu izpētē (skatot gan runāto gan rakstīto valodu), salīdzināmās valodas analīzē gan laika šķērsgriezumā, gan pēc funkcionālajiem stiliem, gan arī viena tekstu veida kopumā (teiksim, daiļliteratūras apakškorpuss), tāpat arī terminoloģijas izstrādē, valodas mācīšanās, tulkošanas studijās un tulku, tulkotāju apmācībā, mašīntulkošanā, dabīgās valodas apstrādē (piemēram, risinot morfoloģijas, sintakses daudznozīmības jautājumus) un vēl citās sfērās (Svartvik 1992, 8–10).

Pievēršoties rakstnieku lietotās valodas izpētei un rakstnieku valodas vārdnīcu sastādīšanas jautājumam, nenoliedzami jāatzīmē korpusa izmantošanas iespējas leksikogrāfijā, jo bez korpusa un korpusa rīkiem mūsdienās sastādīt labu vārdnīcu ir grūti. Tas palīdz iegūt tādu nepieciešamu informāciju kā statistikas datus, vārdu savienojumu analīzi utt.

Kā jau minēts, ar valodas korpusa rīku palīdzību iespējams pētīt ne tikai valodu kopumā, bet arī atsevišķus tās slāņus, piemēram, poētisko valodu. Nenoliedzami, katra cilvēka lietotā valoda ir zināmā mērā atšķirīga, tai piemīt savas īpatnības, kuras nereti ir vērts izpētīt. Kā īpaši aizraujošu un savdabīgu iespējams izcelt rakstnieku valodu. Lai būtu iespējams šo leksiku uzskatāmi

parādīt, analizējot autortekstus, tiek veidotas rakstnieku valodas vārdnīcas.

Šādas vārdnīcas ir pēc noteiktas metodikas apkopoti, analizēti un komentēti vārdu krājumi. Šādu vārdnīcu esamība uzskatāma par kultūras faktu, kas liecina par konkrētās kultūrvides līmeni, humanitāro zinātņu briedumu, kā arī datorresursu jaudām.

Jāatzīme, ka valodnieks J. Endzelīns jau 1933. gadā aicinājis sastādīt rakstnieka valodas vārdnīcas. Konstantīns Karulis uzsvēris, ka šādām vārdnīcām ir trejāda nozīme, proti, tās ne vien sniedz sistematizētu materiālu rakstnieka leksikas pētīšanai, bet ir arī svarīgs materiāls latviešu valodas tēzaura (pilnīgās vārdnīcas) un vēsturiskās vārdnīcas sastādīšanai. Tās arī sniedz iespēju precīzāk konstatēt paņēmienus latviešu literārās valodas attīstībai, kas var būt būtisks palīgs arī nākotnē, jo valodas veidošanās process vēl nav beidzies un vēl aizvien ir daudz nozīmīgu risināmu problēmu (Karulis 1966, 3).

Aplūkojot rakstnieku valodas vārdnīcu veidošanas tradīcijas pasaulē, novērojami vairāki rakstnieku valodas vārdnīcu tipi jeb veidi. No tām pašas senākās veidotas manuāli un ir sastopamas rokrakstu veidā. Jaunākās, savukārt, tapušas datorizētās valodas vidē ar programmatūru palīdzību.

Visā pasaulē tiek veidotas rakstnieka valodas vārdnīcas lielākajiem literāturas dižgariem. Minamas vairākas nozīmīgas cittautu rakstnieku valodas vārdnīcas, piemēram, krievu valodā tāda ir „Puškina valodas vārdnīca” (*Словарь языка Пушкина* 4, Maskava, 1956–1961), ukraiņu valodā „Ševčenko valodas vārdnīca” (*Словник мови Шевченка*, Kijeva, 1964), angļu valodā – „Dž. Milтона angļu dzejas darbu vārdnīca” (L. Lockwood, *Lexicon to the English poetical works of John Milton*, Ņujorka, 1907), spāņu valodā „Servantesa vārdnīca” (C. Gomez, *Vocabulario de Cervantes*, Madride, 1962) (Roze 1982). Lietuvā izdota Kristijona Donelaiša valodas vārdnīca (Karulis 1966, 3).

Kā viena no pilnīgākajām minama Šekspīra valodas vārdnīca (D. Crystal, B. Crystal, *Shakespeare's words: A glossary and language companion*, 2004), kas, attīstoties programmatūrai un iespējām, tiek nepārtraukti papildināta un ir pieejama arī internetā adresē <http://www.shakespeareswords.com>.

Aplūkojot situāciju Latvijā rakstnieku valodas vārdnīcu veidošanā, jākonstatē, ka laika gaitā vārdnīcu izveides mēģinājumos izmantoti gan rokraksta, gan iespiedraksta, gan arī datorprogrammatūras paņēmieni.

Kā nozīmīgs manuskripts minams 17./18. gs mijā Rīgas Jāņa baznīcas mācītāja Liborija Depkina (*Liborius Depkin*) sastādītais latviešu-vācu vārdnīcas manuskripts trīs sējumos, pie kura sagatavošanas izdošanai kopā ar profesoru baltistu Pēteri Vanagu strādājis Dienvidaustrālijas Flindersa universitātes pro-

fesors Trevors G. Fennells. Šī darba mērķis – ne vien padarīt unikālo manuskriptu pieejamu nākamajiem pētniekiem, bet arī „garantēt arī to, ka šis darbs nepazudīs mitros pagrabos”. Šajā manuskriptā atrodams Vecās Derības vārdu rādītājs. Lai arī šajā gadījumā runājam par viena darba, ne autora vārdu rādītāju, tomēr vārdnīcas izveides principi ir līdzīgi.

Kā jau minēts, rakstnieka valodas vārdnīcas nereti tiek veidotas tautu literatūras dižgariem. Arī latviešiem tādu netrūkst. Jau vairāk kā pusgadsimtu tiek apspriesta nepieciešamība pēc daudzuprāt lielākā latviešu literatūras dižgara, proti Raiņa, valodas vārdnīcas. Šīs vārdnīcas izveidi centušies virzīt Konstantīns Karulis, Alise Laua, Kārlis Egle, Laimdots Ceplītis u. c. K. Karulis (1966, 3) min, ka arī pašam rakstniekam nav bijusi sveša doma par šādas viņa leksikas vārdnīcas sastādīšanu. Gunas Eglītes izstrādātajā disertācijā ieskicētas idejas par to, kādai būtu jābūt Raiņa rakstnieka valodas vārdnīcai.

Taču darbs nerit tik raiti kā gribētos. Kāpēc? Kā viena no nozīmīgākajām problēmām vienmēr tiek uzsvērts darbietilpīgais kartotēkas veidošanas process. Agrāk vārdi un tos ietverošais teksts tika izrakstīti ar roku. Mūsdienās lielu daļu darba iespējams paveikt ar datorprogrammu palīdzību, taču vēl aizvien pastāv nozīmīga darba daļa, ko nepieciešams veikt cilvēkam. Jautājums – kur rast darbaspēku?

Zināms, ka ir vēl vairākas iestrādes rakstnieka valodas vārdnīcu veidošanai, taču nevienā gadījumā nav panākts pilnīgs gala rezultāts, proti, nav izdevies izveidot taustāmu rakstnieka valodas vārdnīcu.

Tā kā visā pasaulē šādas vārdnīcas jau pastāv, taču Latvijā šādi centieni līdz šim nav snieguši taustāmu rezultātu, nesen Ventspils Augstskolas Lietišķās valodniecības centra telpās radusies iecere un pēc Dr. philol. Maijas Balčiņas iniciatīvas uzsākts darbs pie vienas šādas vārdnīcas izveides. Iepriekš minētās citviet pasaulē publicētās vārdnīcas noder kā paraugs vārdnīcas izveidē, taču šajā projektā paredzēts datorizētās valodas vidē ar programmrīku palīdzību izveidot savu, oriģinālu vārdnīcu.

Protams, var rasties jautājums par šādas vārdnīcas nepieciešamību. Nereti gadās saskarties ar jautājumu, vai šādas vārdnīcas izveide nav zinātne zinātnes pēc. Šajā sakarā vēlētos atsaukties uz rakstnieka Knuta Skujenieka teikto, proti, autors norāda, ka „visā pasaulē pastāv, piemēram, autoru vārdnīcas, kurās var redzēt, ar kādu valodas diapazonu autors strādā. Mums ir kādi studentu pētījumi, bet reālu grāmatu nav. Tas būtu ļoti svarīgi, jo šobrīd literatūra ir aktīva latviešu valodas veidotāja. Varbūt no literatūras valodas var pasmelt tīri praktiski pielietojamas idejas. Mēs aptuveni zinām, kāda ir lietišķā valoda, bet

paralēli taču pastāv arī literatūra, dzeja, atdzeja, kas izmanto ļoti plaša diapazona valodu” (Radzobe 2008).

Iecerētā rakstnieka valodas vārdnīca tiek veidota kā pilotprojekts datorizētās valodas vidē un ar programmrīku palīdzību. Darba pirmajā posmā tiek analizēti noteikta žanra, konkrētajā gadījumā – poētiskie teksti.

Ņemot vērā autora piekrišanu un izteikto vēlmi uzzināt vairāk par sevis lietoto valodu, ir uzsākts darbs pie Zviedrijā dzīvojošā latviešu dzejnieka un tulkotāja Jura Kronberga dzejas tekstu izpētes. Autors kopš 1970. gada sarakstījis 13 dzejoļu krājumus, kas arī kalpo par pamatu vārdnīcas sastādīšanai nepieciešamā dzejnieka valodas korpusa izveidei.

Lai būtu iespējams uzsākt darbu pie vārdnīcas izveides, nepieciešams visus aplūkojamus tekstus iegūt elektroniskā formātā. Ļoti labi, ja nepieciešamie teksti šādā formātā jau ir pieejami. Taču, ja nav, tad tie tiek ieskenēti un ar teksta atpazīšanas programmas palīdzību iegūtie teksti tiek pārveidoti teksta formātā. Tekstam tiek pievienoti metadati – autora vārds, uzvārds, dzejoļa nosaukums (ja tāds ir), krājuma nosaukums, gads, lappuses. Šī papildniformācija palīdz vēlāk atrast vārda *adresi* jeb atrašanās vietu, tādā veidā atvieglojot tekstu izpēti. Tālāk iegūtie teksti tiek apstrādāti ar Latvijas Universitātes Matemātikas un informātikas institūta Mākslīgā intelekta laboratorijas izstrādāto latviešu valodas tekstu korpusu morfoloģiskās un sintaktiskās marķēšanas rīku. Šis marķētājs ir pusautomātisks programmrīks, kas paredzēts latviešu valodas tekstu morfoloģiskai un sintaktiskai analīzei. Ar tā palīdzību katrai vārdformai tiek dots morfoloģiskais raksturojums un, ja iespējams, noteikta arī vārda sintaktiskā loma teikumā. Tas brīvi pieejams adresē <http://www.semti-kamols.lv>.

Nenoliedzami cilvēka iesaistīšanās marķēšanas procesā ir un būs vajadzīga, jo tehnoloģijas var atklāt tikai daļu patiesības par valodu. Programmnodrošinājums sniedz atbildes tikai uz tiem jautājumiem, kurus ir paredzējis cilvēks, tāpēc datorizētā vide nav spējīga valodas datu analīzes posmā darboties bez cilvēka klātbūtnes (Baltiņa 2006, 15). Cilvēka iesaistīšanās nepieciešama, piemēram, paralēlo vārdformu atpazīšanā, mākslinieciskās izteiksmes līdzekļu izdalīšanā, u. c.

Šī darba daļa ir ļoti laiktietilpīga. Līdzīgi kā Raiņa vārdnīcas entuziastiem, nākas meklēt iespējas, kur rast darbaspēku. Nozīmīgu palīdzību esam atraduši studentu lokā, lūdzot tiem samarķēt nelielus dzejoļu krājumņus. Šādā veidā pa mazumiņam savācas kopā samērā liels marķēto tekstu kopums. Pie tam studentiem darboties ar programmrīkiem ir aizraujoši, īpaši ja tiem ļauj

strādāt grupās. Tiesa, nevar vienmēr akli paļauties uz to, ka studenti nepieļaus kļūdas, tāpēc vēlams marķētos tekstus pārbaudīt.

Vārdnīcu sastādot tekstu analīze tiek veikta trijos līmeņos – datorlingviskā, filoloģiskajā un kultūrsemiotiskajā.

Datorizētajā vidē tiek iegūti precīzi rakstnieka vārdu krājuma kvantitatīvie raksturojumi. Šajā gadījumā var runāt par vārdšķirām (konkrētajā gadījumā tiek aplūkoti lietvārdi, darbības vārdi, īpašības vārdi un apstākļa vārdi), vārdlietojumu, vārdsavienojumiem utt. Citiem vārdiem sakot, analīze ļauj noskaidrot, cik reizes, kādos locījumos un kādā kontekstā vārds tiek lietots. Iegūtais materiāls ir viegli pārskatāms, ļaujot viegli un ātri meklēt un pārskatīt nepieciešamo informāciju.

Filoloģiskajā aspektā vārdu krājuma analīzē tiks izmēģināti programmrīki, ar kuru palīdzību iespējams parādīt rakstnieka valodas resp. vārdu krājuma un vārdu lietojuma stilistiskos raksturojumus (piem., metaforas, sinonīmi, salīdzinājumi u. c.). Jāatzīmē, ka datorizētās analīzes priekšrocība šajā analīzes līmenī ir fakts, ka tā palīdz iegūt *pilnīgāku* priekšstatu par konkrētā rakstnieka valodu, padarot to pārskatāmāku un vieglāk apstrādājamu. Cits datorizētās analīzes ieguvums tiek saistīts ar ieceri – gūt *pilnīgāku* un *aptverošāku* priekšstatu par konkrētām rakstnieka valodas stila iezīmēm (piem., kontekstuālie sinonīmi, uz paradoksiem balstītās metaforas, okazionālie sinonīmu pāri, virknes un rindas). Jāsaka, ka, lai arī datorizētā analīze palīdz iegūt kvantitatīvos datus, stilistiskās izteiksmes līdzekļu identificēšanai pašlaik tomēr nepieciešams cilvēka darbs, jo dators nav spējīgs noteikt, vai vārds lietots tiešā vai pārnestā nozīmē, dators nav spējīgs atpazīt tēlainos izteicienus.

Konkrētajā gadījumā cilvēka līdzdalība vārdnīcas sastādīšanā ir diezgan nozīmīga, jo Kronbergs ir metaforas meistars, nozīmju pārcēlājs. 2006. gadā Ieva Kolmane konferencē *Ventspils Paralēle. Mantojums un mūsdienas* uzstājās ar referātu „Metafora: Kronbergs”, kas pievērsās tieši šī mākslinieciskās izteiksmes līdzekļa lietojumam dzejnieka valodā. Tika norādīts, ka Kronbergs strādā kā nozīmju daudzkārstotājs un pasaules paplašinātājs vienas valodas ietvaros, kā arī prot būvēt tiltus, kas nozīmēs ļauj dalīties vairākām valodām.

Pievērsoties topošās rakstnieka valodas vārdnīcas šķirklā uzbūves jautājumam, jānorāda, ka šīs rakstnieka valodas vārdnīcas šķirklis atšķiras no literārās valodas vārdnīcas šķirklā un tā veidošanas principiem. Šajā gadījumā galvenā ir tieši konkrētā rakstnieka valoda un vārdu kopums, balstoties uz ko šķirklis tiek veidots.

Vārdi tiek sakārtoti pēc nozīmēm, taču netiek sniegts literārās valodas vārdnīcas skaidrojums. Tā kā dzeja ir poētiskais valodas līmenis, tad vārda skaidrojums tiek sniegts ar paša dzejnieka tekstu, proti, tiek citēta dzejas rinda vai pants. Tāpat šķirkļa ietvaros tiek atsevišķi izdalītas metaforas, salīdzinājumi, sinonīmi, idiomas. Šajā gadījumā runa iet par sinonīmiem, kurus lieto pats rakstnieks, nevis par tiem, kas atrodami sinonīmu vārdnīcā. Nereti autora sniegtie sinonīmi būs oriģināli un nebūs atrodami sinonīmu vārdnīcā.

Salīdzinājumam kā piemēru varam aplūkot vienu no Kronberga biežāk lietotajiem lietvārdiem *ceļš*. Latviešu literārās valodas vārdnīcā sniegtās lietvārda *ceļš* nozīmes:

- 1) Speciāli izveidota (parasti ar segumu noklāta) zemes strēle sauszemes transportlīdzekļu, arī gājēju kustībai.
- 2) Noejams vai nobraucams posms, attālums.
- 3) Ceļojums, brauciens, gājiens.
- 4) Cilvēka eksistence, darbība, darbības virziens; dzīves gaita, gājums.
- 5) Veids, paņēmieni, pieeja.
- 6) inf. Maršruts starp jebkuriem diviem tīkla mezgliem.
- 7) inf. Maršruts, kas operētājsistēmai jānoiet, lai atrastu izpildāmo programmu apakšdirektorijā.

Rakstnieka valodas vārdnīcas gadījumā tiek aplūkotas tikai tās vārda nozīmes, kuras izmantojis pats rakstnieks. Šajā gadījumā, *ceļš* lietots 17 reizes 5 dažādās nozīmēs. Nozīmes tiek kārtotas pēc to lietojuma biežuma, skaidrotas ar rakstnieka paša dzejas rindām. Tiek sniegti arī piemēri, ja tādi atrodami, kur vārds lietots kā metafora, ja tas lietots salīdzinājumā, standartfrāzē, kā arī tiek sniegti autora darbos izmantoties vārda sinonīmi.

1.

- *Diena vēl nav apstājusies*  
*Diena vēl ir ceļā* [Diena vēl nav gatava, Tagadnes, 1985]
- *kuģis Wasa nogrima savā*  
*uzvaras gājienā: ceļā uz Ap-gāni-stānu* [Caurstrāvotājs (Austrumeiropa), Laiks, 1994]
- *manu vecāku jaunības sapņi,*  
*sakropļoti pēdējā karā,*  
*ceļā no Nacionālā teātra*  
*uz Švarca kafejnīcu* [Travel Document, Tagadnes, 1981]

- *tā ir mūsu lidojošā valoda. **Ceļā**  
uz citiem kontinentiem  
tā zaudēs  
daļu sava skaistuma* [Travel Document, Tagadnes, 1981]
- *dzied par zvaigznēm  
kuru gaisma tevi vēl nav  
sasniegusi,  
kas vienmēr ir **ceļā**  
no kāda cita cilvēka acīm  
uz tavām* [Mēs, Tagadnes, 1985]

2.

- Eju pa Eiropas **ceļu** nr. 4  
ar pārtikas saini katrā rokā [Travel Document, Tagadnes, 1981]
- *Tevi  
lieto, lai nosegtu  
nelīdzenumus un bedres  
**ceļos**, pa kuņģiem mēs traucam* [Asfalts, Tagadnes, 1981]
- *Lēnītēm atausa **ceļi** un lauki un meži un sniegi* [Vilks Vienacis, 2008]
- *Vecajā viesnīcā, kas senāk dēvēta  
tās „visi **ceļi** ved uz...” vārdā,* [Reālisms Rīgā, Tagadnes, 1981]

3.

- *Viņš nesaredz **ceļu** no šodienas mezglā  
Viņš ir mezglis viss pārējais mezglis  
no kura tam neizkļūt laukā* [Vilks Vienacis, 2008]
- *Esmu aprakts pats sevī neatrodu  
**ceļa** no šīs  
filantropiskas pašapmierinātības.* [Caurstrāvotājs (Austrumeiropa),  
Laiks, 1994]

4.

- *Liktenis turpina savu nosprausto **ceļu**  
[No dienasgrāmatas rudenī, Tagadnes, 1981]*
- *Aizsāktais **ceļš** tika pārtraukts  
gals būs citāds* [Vilks Vienacis, 2008]



5.

- *šis laikmets sola mums briesmas,  
Tu mūsos degsi kā liesma  
kas silda,  
kas **ceļu** rāda  
caur zudušu dienu tumsu* [Oda Pagrabam, Tagadnes, 1981]

Metaforas:

- *skudru kolonijas klīst  
man cauri, es jūtos vajadzīgs; es  
nu īsākais **ceļš**  
uz skudru ciltij solīto zemi*
- *Izraudātās acis zvēroja kā tuksneši  
Uz vaigiem divi sarkani asaru **ceļi*** [Sapnis, Tagadnes, 1985]

Sinonīmi:

- *ielā, šoseja, taka*
- *izeja*
- *virziens*

Nobeigumā jāatzīmē, ka ne vienmēr viss ir tik vienkārši, kā sākumā varētu likties. Rakstnieki savos darbos mēdz būt asprātīgi un nereti negribot (vai varbūt tišuprāt) sarežģī valodas pētnieku darbu. Ieva Kolmane referātā „Metafora: Kronbergs” pat teikusi, ka „veselais saprāts liedz jelko, kas ir Juris Kronbers vai viņa daļa, padarīt par akadēmiskas izpētes priekšmetu”. Šeit noteikti jāpiemin autora 2005. gadā izdotsis dzejoļu krājums „Peti-šu”, kur īpaši atklājas, ka Kronbergs ir dzejnieks, kam patīk spēlēties ar vārdiem, padarot tekstu analīzi manāmi sarežģītāku. Kronbergam patīk vārdus darināt, ar tiem spēlēties, eksperimentēt, un ieinteresēt. Šajā gadījumā varam runāt ne vien par citvalodu vārdu vai izteicienu izmantošanu tekstos, kas Kronbergam ļoti raksturīgi, bet arī par citiem autora *āķīgajiem darinājumiem*, kur nereti rodas problēmas jau sākot ar vārdšķiras noteikšanu (piem.: *heming vējš, kūkar pipas, ticio ai netici* u. c.). Lai arī šajā gadījumā jārunā par vārdu spēlēm, tās nenoliedzami ir nozīmīga autora vārdu krājuma sastāvdaļa, kuru nevar atstāt neievērotu. Ar korpusa rīku palīdzību iespējams šos speciālos gadījumus izdalīt un attiecīgi komentēt, pievēršot uzmanību šiem īpašajiem valodas slāņiem.

## AUTHOR'S TEXT AS A RESOURCE FOR THE NATIONAL CORPUS AND THE WRITER'S LANGUAGE DICTIONARY

### *Summary*

A dictionary of a writer's language is a vocabulary that has been methodically collected, analyzed and commented upon and represents the language as used by a particular author. The existence of such dictionaries indicates the level of the specific cultural environment, the stage of development of humanities and the capacity of computer resources.

In the first stage of the project, the corpus of the writer's language is created by collecting all the necessary texts of a selected author in a digital form. In the second stage, linguistic analysis of the corpus of the writer's language is carried out using the existing and improved technologies (software and software tools). And, finally, in the third stage, on the basis of the corpus of the writer's language a dictionary of the writer's language is created according to set thematic, philological and stylistic intentions. The analysis is carried out on three levels: computational linguistic, philological and cultural semiotic.

The computerized environment provides a precise quantitative characterization of the writer's vocabulary (part of speech, word use, collocations etc.). From the philological point of view, vocabulary is analyzed with the help of software tools that help identify and demonstrate the stylistic characterization of the writer's language, namely, the writer's vocabulary and word use (e.g. metaphors, synonyms, similes etc.). Another benefit of computerized analysis is related to the intention to acquire a more complete and comprehensive idea about specific features of the writer's language style.

### LITERATŪRA

Baltiņa Maija 2006, No vārdu sarakstiem līdz korpusa lingvistikai, *Linguistica Lettica* 15, 77–84.

Ceplītis Laimdots 1967, Kā sagatavot Raiņa valodas vārdnīcu, *Literatūra un Māksla* (1967), 3.

Chafe Wallace 1992, The importance of corpus linguistics to understanding the nature of language, in Jan Svartvik (ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, New York: Mouton de Gruyter, 79–98.

Egle Kārlis 1967, Kādai vajadzētu būt Raiņa vārdnīcai, *Literatūra un Māksla* (1967 01 14), 13.

Fennell Trevor G. 2006, Liborius Depkin's fidelity to his sources, *Baltu Filoloģija* 15(1/2), Rīga, 17–26.

Fillmore J. Charles 1992, "Corpus linguistics" or "Computer-aided armchair linguistics", in Jan Svartvik (ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, New York: Mouton de Gruyter, 35–60.

Francis W. Nelson 1992, Language Corpora B.C., in Jan Svartvik (ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, New York: Mouton de Gruyter, 17–35.

Halliday Michael Alexander Kirkwood 1992, Language as system and language as instance: The corpus as a theoretical construct, in Jan Svartvik (ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, New York: Mouton de Gruyter, 61–79.

Karulis Konstantīns 1966, Par Raiņa valodas vārdnīcu, *Literatūra un Māksla* (1966), 3.

Laua Alise 1966, Raiņa leksikas vārdnīca – darbs, kuru nevar atlikt, *Literatūra un Māksla* (1966 12 03), 3.

Radzobe Zane 2008, Visu latviešu valodu vienā programmā, <http://www.diena.lv/lat/izklaide/kd/visu-latviesu-valodu-viena-programma> (2008 09 01).

Roze Liene 1982, Senās un mūsdienu enciklopēdijas, in Liene Roze, *Pasaule vārdnīcu skatījumā*, Rīga: Zinātne 1982, 7–26.

Svartvik Jan 1992, Corpus linguistics comes of age, in Jan Svartvik (ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, New York: Mouton de Gruyter, 7–14.

Vanags Pēteris 2004, Saīsinājumi *Anon., Anonym., Anon. Voc.* Liborija Depkina vārdnīcas *Lettifches Wörterbuch* manuskriptā, *Baltistica* 39(2), 269–277.

Veisbergs Andrejs 2003, Elektroniskās vārdnīcas un datorizētie vārdnīcu resursi, *Linguistica Lettica* 12, 163–175.

Maija BALTIŅA, Daina VUCĀNE

Lietišķās valodniecības centrs

Ventspils Augstskola

Inženieru 101a

LV-3601 Ventspils

Latvia

[[maija.baltina@venta.lv](mailto:maija.baltina@venta.lv)]

[[daina.vucane@venta.lv](mailto:daina.vucane@venta.lv)]