

Korpusu un individuālā vākuma salīdzinājums: ģenitīva un nominatīva konkurence saistījumā ar adverbu¹

A comparison of corpora and individual collection: Genitive and nominative competition in connection with an adverb

Linda Lauze

Humanitāro un mākslas zinātņu fakultāte
Liepājas Universitāte
Kūrmājas prospekts 13, Liepāja, LV-3401, Latvija
E-pasts: linda.lauze@liepu.lv

Ilze Auziņa

Mākslīgā intelekta laboratorija
Matemātikas un informātikas institūts
Latvijas Universitāte
Raiņa bulvāris 29, Rīga, LV-1459, Latvija
E-pasts: ilze.auzina@lumii.lv

Rakstā aplūkotas korpusa datu un individuālā vākuma priekšrocības un trūkumi. Dažādu gramatiski marķētu latviešu valodas korpusu pieejamība nodrošina arvien plašākus korpusa datus balstītus gramatikas pētījumus. Savukārt individuālajam vākumam ir bijusi liela nozīme valodniecības attīstībā, un tas ir senāks praktiskā materiāla ieguves veids. Tomēr mūsdienu tehnoloģiju laikmetā ir radies jautājums par individuālā vākuma lietderību.

Abu datu ieguves veidu praktiskam salīdzinājumam izvēlēta mūsdienu latviešu valodas lietojumā izplatīta parādība – ģenitīva un nominatīva konkurence (saistījumā ar adverbu), kas konstatēta gan individuālajā vākumā, gan korpusu datos.

Šajā pētījumā ir atlasīti trīs adverbi – *daudz* (tā supletīvā pārkāps pakāpes forma *vairāk*), *maz, cik* –, kuri sīkāk tiek analizēti teikuma gramatiskajā centrā saistījumā ar lietvārda ģenitīvu vai nominatīvu. Individuālo vākumu veido samērā spontāns nerediģēts latviešu valodas lietojums mutvārdos un rakstveidā – ar katru adverbu 100 teikumu. Korpusa datu analīzei izmantoti četri latviešu valodas korpusi – „Līdzsvarotais mūsdienu latviešu valodas tekstu korpus” (LVK2018), „Latviešu valodas sintaktiski marķētais korpus” (LVTB), „Latviešu valodas runas atpazīšanas korpus” (LVR2013) un „Pandēmijas dienasgrāmatas” (PanDi). Gan korpusu, gan individuālā vākuma materiālā dominē vārdu savienojumi ar ģenitīvu. No avotu viedokļa, nominatīvs biežāk

¹ Publikācija izstrādāta Valsts pētījumu programmas „Letonika latviskas un eiropiskas sabiedrības attīstībai” projektos „Mūsdienu latviešu valodas izpēte un valodas tehnoloģiju attīstība” (Nr. VPP-LETONIKA-2021/1-0006) un „Mūsdienu latviešu valodas lietojums” (Nr. VPP-LETONIKA-2022/1-0001).

konstatēts „Latviešu valodas runas atpazīšanas korpusā” (LVR2013), bet analizēto trīs adverbu grupā – biežāk saistījumā ar adverbu *cik*.

Atslēgvārdi: latviešu valoda; individuālais vākums; valodas korpusi; sintakse; ģenitīva un nominatīva konkurence; saistījums ar mēra adverbu.

levads

Attīstoties datorlingvistikai un kļūstot pieejamiem dažādiem korpusiem, latviešu valodas gramatikas pētījumos arvien plašāk tiek izmantoti korpusos atlasīti dati. Analizējot lielus un strukturētus valodas korpusus, pētnieki iegūst objektīvus datus par valodas lietojumu, tāpēc valodas korpusi ir neatsverami dažādās valodas sinhroniskās un diahroniskās izpētes jomās. Korpusa dati parasti satur strukturālu, morfoloģisku, sintaktisku, semantisku vai cita veida marķējumu, kas palīdz valodas datus analizēt dažādos aspektos. Arī latviešu valodai ir izstrādāti dažādi korpusi – strukturēti un marķēti rakstītās un runātās valodas dati. Korpusi teorētiskajai lingvistikai piedāvā gan apstiprinājumu kādas valodas parādības eksistencei, gan informāciju par (valodas parādības) biežumu un izplatību (Andronova, Auziņa 2022).

Individuālajam vākumam ir bijusi liela nozīme valodniecības attīstībā, un tas ir senāks praktiskā materiāla ieguves veids. Tomēr mūsdienu tehnoloģiju laikmetā ir radies jautājums par individuālā vākuma lietderību. Abu datu ieguves veidu praktiskam salīdzinājumam izvēlēta mūsdienu latviešu valodas lietojumā izplatīta parādība – ģenitīva un nominatīva konkurence saistījumā ar mēra adverbu, kas konstatēta gan individuālajā vākumā (turpmāk pie piemēriem norāde – IV), gan korpusu datus (turpmāk pie piemēriem norāde – KD).

Raksta mērķis ir, pirmkārt, raksturot latviešu valodas korpusu datu un individuālā vākuma priekšrocības un trūkumus, otrkārt, analizēt ģenitīva un nominatīva konkurenci saistījumā ar adverbu, izmantojot korpusu un individuālā vākuma datus un salīdzinot iegūto rezultātu attiecību.

Latviešu valodā ir vairāki sintaktiskā saistījuma gadījumi, kuros literārās valodas norma nosaka ģenitīva lietojumu, bet praksē ir sastopams arī paralēls nominatīva izmantojums, piemēram:

- 1) subjekta pozīcijā pie eksistenciālā vai piederības nozīmē lietota verba *nebūt*;
- 2) subjekta pozīcijā pie adverbiāla izteicēja, kas izsaka daudzumu;
- 3) dalāmā ģenitīva konstrukcijās.

Jāpiebilst, ka tādos izteikumos, kur dalāmā ģenitīva konstrukcija lietota objekta pozīcijā, sarunvalodā novērots arī akuzatīva izmantojums (1).

(1) ***Daudz piparus neber.*** (IV – viešņa, runas situācija virtuvē, LL 03.07.2023.)

Tā kā saistījumā ar daudzuma apzīmējumu lietvārdus lielākoties izmanto daudzskaitļa formā (izņemot vienskaitļniekus), sieviešu dzimtes lietvārdu daudzskaitļa nominatīva un akuzatīva homoformas apgrūtinātu objektīvu skatījumu uz locījumu konkurenci, galvenā uzmanība pētījumā tiek pievērsta tām dalāmā ģenitīva konstrukcijām, kas lietotas subjekta funkcijā.

Latviešu valodas gramatikas pētniecībā līdz šim jau ir raksturots gan subjekta saistījums ar eksistenciālā vai piederības nozīmē lietota verba *būt* nolieguma formu, gan partitīva konstrukciju sintaktiskās īpatnības (Lokmane 2017; Kalnača, Lokmane 2022) un abu locījumu konkurence (Kalnača 2002, 2007, 2019), arī saistījumā ar mēra adverbu (Smiltnece 2013). Sarmīte Lagzdiņa (1997) ir analizējusi verbu *nebūt* gan pierastajā vienpersonas, gan trīspersonu funkcijā, uzsvērdama semantiskās atšķirības teikumos, kur ir nepieciešams piesaistīt nominatīvu, piem., procesuālajā nozīmē. Andra Kalnača un Ilze Lokmane (2018a, 2018b; sk. arī Kalnača, Lokmane, Metslang 2019) ir aplūkojušas korelāciju starp teikuma priekšmeta locījumu un vārdu secību eksistenciālsteikumos. Sturla Bergs-Olsens (*Sturla Berg-Olsen* 2005, 2009) ir pētījis ģenitīva un nominatīva lietošanu senajos rakstos un mūsdienu mutvārdu un rakstveida tekstos un novērojis, ka latviešu valodā ģenitīvam jau ilgstoši samazinās izmantojums gan dalāmajās konstrukcijās, gan saistījumā ar verbu (resp., adverbālais lietojums). Arī sarunvalodā konstatēts nominatīva izmantojums ģenitīva vietā (Lauze 2004, 85–87). A. Kalnača (2022, 92) pauž līdzīgu viedokli, ka „neatkarīgi no saziņas veida un teksta satura latviešu valodā mēs faktiski esam liecinieki plašākām teikuma sintaktiskās struktūras pārmaiņām, kur teikuma priekšmets pretstatā literārās valodas normām tiecas vienmēr būt nominatīvā”.

1. Praktiskais materiāls un izmantotās metodes

Individuālo vākumu veido samērā spontāns nerediģēts latviešu valodas lietojums mutvārdos un rakstveidā, kopā 400 izteikumu – adverbam *daudz* 200 lietojuma gadījumu (100 pamatformā un 100 tā supletīvajā pārākajā pakāpē *vairāk*) vienkāršā teikumā vai salikta teikuma daļā un pa 100 piemēriem adverbiem *maz* un *cik*. Jāuzsver, ka materiāls nav vākts, koncentrējot uzmanību uz valodas kultūras ķlūdām, bet pēc kārtas internetā, sociālās tīklošanās un mikrologošanas vietnēs (*Facebook*, *Twitter*), radio pārraidēs (konkrētās radiostacijas sk. saīsinājumu sarakstā), dažādās runas situācijās akadēmiskajā vidē universitātē (nodarbībās, sanāksmēs), ikdienā sabiedriskās vietās u. c. fiksējot valodas lietojuma piemērus, kuros konstatēts lietvārda vai vietniekvārda saistījums ar adverbu nominatīvā vai ģenitīvā. Fiksētie runas fakti ir literārajā valodā un sarunvalodā – formālā un neformālā saziņā.

Korpusa datus veido izteikumi no četriem dažādiem latviešu valodas korpusiem (sk. 1.2.), un atrasto un analizēto piemēru skaits ir atkarīgs gan no korpusa veida, gan apjoma, tomēr kopējais piemēru skaits pārsniedz četrus tūkstošus.

1.1. Individuālais vākums

Sociolingvistikā, lai iegūtu dabiskas, spontānas runas faktus, tiek risināts jautājums, kādā veidā pārvarēt Viljama Labova (*William Labov* 1970) izvirzīto novērotāja paradoksa (*observer's paradox*) problēmu: kā, pētniekam nelūdzot, lai cilvēki atklāti runā mikrofonā, noskaidrot, kā viņi runā, kad viņi nav sistemātiska pētījuma objekts. Sarunvalodas izpētē „mākslīga vide, kurā runas paraugi tiek iegūti pēc pieprasījuma,

atrautībā no ierastā draugu un paziņu loka, visai reti ļaus konstatēt detaļas varianta izvēlē, koda maiņu u. tml.” (Baltiņš, Druviete 2017, 147). Tādējādi autentisku sarunvalodas piemēru ieguvē ir ārkārtīgi svarīgi, lai runas faktu vākšana notiktu, runātājiem neziņot, tajā pašā laikā nodrošinot respondentu identitātes aizsardzību un zinātnieka ētikas principu ievērošanu (plašāk sk. Lauze 2008, 7–14). Jāpiebilst, ka, fiksējot sociālo mediju tekstus, kas lielākoties ir publiski pieejami, rakstītāja konta lietotājvārds tiek paturēts.

Tiek izmantota tiešo novērojumu metode, kur pētnieks vēro runas situāciju no malas, un līdzdalīgā novērojuma metode, kur pētnieks ir saziņas dalībnieks, kuram ir pieejama šaurāka vai plašāka papildu informācija par tiem runātājiem, viņu izglītību, profesiju, interesēm u. tml., kas pieder pie viņa kolēģu, draugu, radu, paziņu loka. Piem., izmantotajiem *Facebook* un *Twitter* kontiem ir daudz sekotāju ar augstāko izglītību, arī filoloģisko, pasniedzēji, tulkotāji, literāti, studenti, kas ļautu domāt, ka tāpēc ir vairāk piemēru, kas atbilst literārās valodas normai. Līdzdalīgā novērojuma priekšrocība ir iespēja konstatēt valodas lietojuma variantus viena cilvēka runā (2) un (3), ja piemēri ir ņemti no vairākiem avotiem. Papildinot kartotēku no viena avota, ģenitīva (4) un nominatīva (5) paralēlu izmantojumu vienā un tajā pašā pozīcijā var viegli konstatēt arī nepazīstama runātāja vai rakstītāja tekstā.

- (2) ***Daudz, daudz vēl ir dziesmu, kuras vajadzētu nodziedāt.*** (IV – Mārtiņš Bergmanis, Facebook „LustīgsKoronBalle” straumēšana 05.02.2022.)
- (3) ***Arī mūsu radiostacijā ir daudz viņa dziesmas, daudz viņa dziesmas.*** (IV – Mārtiņš Bergmanis, „LustīgsRadio” 31.01.2022.)
- (4) ***Cik te varētu būt to folkloristu?*** (IV – žurnālists, LR1 Tiešraide no starptautiskā folkloras festivāla „Baltica 2022” noslēguma 10.07.2022.)
- (5) ***Cik šie kolektīvi vispār iesoļo?*** (IV – žurnālists, LR1 Tiešraide no starptautiskā folkloras festivāla „Baltica 2022” noslēguma 10.07.2022.)

Analizējot līdzdalīgā novērojuma izmantojumu jauniešu valodas sociolingvistiskajā izpētē, Vineta Ernstsonē un Laura Tidriķē (2006, 18–19) ir uzsvērušas šīs metodes priekšrocības un trūkumus. Ilglaicīgs pētījums sniedz diezgan pilnīgu ainu par saziņu kādā sociālajā grupā neformālās situācijās, bet par galveno nepilnību atzīts nelielais novēroto cilvēku skaits. Var pievienoties Bernarda Spolska (*Bernard Spolsky* 2003, 12) viedoklim, ka šādi piemēri nav paredzēti statistiskai analīzei, tomēr spilgti gadījumi, kurus pētnieks uzskata par vērtīgiem, balstoties savā novērojumu pieredzē, parasti reprezentē vispārēju noteikumu, nevis normai neatbilstošu gadījumu. Daudzu gadu laikā, kas veltīti latviešu sarunvalodas piemēru individuālam vākumam, vairākkārt ir bijušas runas situācijas, kurās vēlreiz apstiprinās kāda iepriekš praktiskajā materiālā jau konstatēta un teorētiski aprakstīta parādība, kas ir raksturīga neformālā mutvārdu saziņā.

1.2. Korpusu datu atlase un analīze

Lai salīdzinātu korpusu datus un individuālo vākumu un analizētu nominatīva un ģenitīva konkurenci saistījumā ar adverbu, izraudzīti četri atšķirīgi Nacionālajā korpusu kolekcijā (NKK), kas pieejama platformā „Korpuss.lv” (Saulīte et al. 2022; Saulīte,

Darģis, Auziņa 2022), iekļauti mūsdienu latviešu valodas korpusi – „Līdzsvarotais mūsdienu latviešu valodas tekstu korpus (LVK2018)”, „Latviešu valodas sintaktiski marķētais korpus (LVTB)” (Rituma et al. 2023), „Latviešu valodas runas atpazīšanas korpus (LVR2013)” un „Pandēmijas dienasgrāmatas (PanDi)”. LVK2018, LVTB un LVR2013 ir vispārīgi, reprezentatīvi korpusi, kas sniedz vispārīgu priekšstatu par latviešu valodu un nav ierobežoti tematiski vai pēc kādas citas pazīmes, savukārt PanDi ir specializētais korpus, kas ietver kovida laikā tapušās dienasgrāmatas, kurās dokumentēta ikdienas dzīve pandēmijas laikā.

Visi analīzei izraudzītie korpusi ir automātiski morfoloģiski marķēti, izmantojot Latvijas Universitātes Matemātikas un informātikas institūta izstrādāto latviešu valodas automātiskās analīzes rīkkopu LV-PIPE (<http://nlp.ailab.lv/>) (Znotiņš, Bārzdiņš 2020; Grūzītis, Znotiņš 2018), kas nodrošina padziļinātu valodas izpēti. LVTB automātiskais morfoloģiskais marķējums ir manuāli pārbaudīts, un datiem (17 127 teikumiem) ir pievienots arī sintaktiskais marķējums – teikumu sintaktiskās struktūras attēlošanai izmantojot hibrīdu atkarību-frāžu gramatikas modeli (Rituma et al. 2019). Savukārt LRK2013 ir iekļauti audioieraksti un to atšifrējumi ortogrāfiskajā transkripcijā.

Analizējot korpusu datus, tiek izmantotas konkordances, kas iegūtas ar pirms tam rūpīgi formulētu CQL (angļu val. *Corpus Query Language* ‘korpusa vaicājuma valoda’) vaicājumu. Izmantotas arī kvantitatīvās analīzes metodes, skatot atrasto konstrukciju izplatības biežumu korpusos (McEnery, Hardie 2012).

Tā kā LVTB ir pievienots arī sintaktiskais marķējums, kas ļauj daudz precīzāk atlasīt teikumus, kuros adverbs tiek lietots kopā ar lietvārdu nominatīvā vai ģenitīvā, turklāt ņemot vērā arī sintaktisko saistījumu un sintaktiskās lomas, meklējot LVTB, izmantoti vaicājumi, kuros norādīta atkarību struktūra (teikuma loceklis, no kura atkarīgs cits teikuma loceklis) – tiek meklēti adverbi *daudz*, tā supletīvā pārākās pakāpes forma *vairāk*, *maz*, *cik*, kas atkarīgi no lietvārda nominatīvā vai ģenitīvā:

```
a-node [ m/tag ~ '^n...(n|g)', a-node [ m/form = "daudz" ] ]
a-node [ m/tag ~ '^n...(n|g)', a-node [ m/form = "vairāk" ] ]
a-node [ m/tag ~ '^n...(n|g)', a-node [ m/form = "cik" ] ]
a-node [ m/tag ~ '^n...(n|g)', a-node [ m/form = "maz" ] ]
```

Tā kā korpusa apjoms ir salīdzinoši neliels, atrasto piemēru nav daudz – ar *daudz* ir atrasti 85 teikumi, no kuriem tikai divos lietots lietvārds nominatīvā (6)–(7), ar *vairāk* – 24 teikumi, ar *maz* – 4 teikumi un ar *cik* – 13 teikumi. Teikumos, kuros *vairāk*, *cik*, *maz* atkarīgs no lietvārda, lietvārds ir sastopams tikai ģenitīvā.

- (6) *Labprātāk reizi pa reizei pasēžu mājās ar draugiem, iedzeru vīnu, parunājos, pasmējos, ne eju uz vietu, kur skaļa mūzika un daudz sveši cilvēki.* (KD – LVTB)
- (7) *Interesējas arī citi jaunie vecāki – ir jau tikpat daudz bērni, kuri tiek gatavoti – nāk piecgadīgi un pat divi četrgadīgi.* (KD – LVTB)

Otrs vaicājuma tips, kas izmantots, meklējot LVTB konkrēto adverbu un lietvārda lietojumu, ļauj atrast teikumus ar adverbīālu izteicēju, kurā viens no komponentiem ir adverbs *daudz*, *vairāk*, *maz* vai *cik* un otrs ir lietvārds teikuma priekšmeta funkcijā, piem.,

```
a-node [ role ="pred", a-xinfo [ xtype = "xPred", tag ~ '\\[adv', a-node
[ m/form = "daudz" ] ], a-node [ role = "subj", m/tag ~ '^n...(n|g)' ] ]
```

Atrasto piemēru nav daudz: ar *daudz* ir atrasti astoņi teikumi, ar *vairāk* – seši teikumi, ar *maz* – divi teikumi un ar *cik* – viens teikums. Tātad lielākajā daļā atrasto teikumu lietvārds teikuma priekšmeta funkcijā ir ģenitīvā (8), (9). Tikai vienā teikumā, kurā adverbīāla izteicēja komponents ir vārdforma *vairāk* (10), un teikumā, kurā adverbīāla izteicēja komponents ir adverbs *cik*, teikuma priekšmets ir nominatīvā.

- (8) *Sabilē brūno plāksnišu ar uzrakstiem par nozīmīgām vietām ir daudz, un šaurajās ieliņās tās pamanāmas vēl jo labāk.* (KD – LVTB)
- (9) *Jūs taču visi esat pamanījuši, ka cilvēku ar bārdām kļūst arvien vairāk?* (KD – LVTB)
- (10) *Vēl jo vairāk prieks, ja bērniņš ir gaidīts un vesels, atzīst Daugavpils reģionālās slimnīcas vecmāte Iveta Žilvinska [..]* (KD – LVTB)

LVK2018, LRK2013, PanDi korpusos, kas pieejami no *SketchEngine* korpusu pār-lūkprogrammā, atlasītas konkordances, kurās adverbs *daudz*, *vairāk*, *maz* un *cik* lietots saistījumā ar lietvārdu nominatīvā vai ģenitīvā. Lai atrastu adverbis *daudz*, *vairāk*, *maz* un *cik*, kas atkarīgi no lietvārda nominatīvā vai ģenitīvā un atrodas tieši blakus lietvārdam, izmantoti CQL vaicājumi, piem.,

```
[word="daudz"] [tag="n...g.*"]
[word="daudz"] [tag="n...n.*"]
```

Tā kā LVK2018, LRK2013, PanDi nav sintaktiski marķēti, ar izmantotajiem vaicājumiem tiek atrasti arī pētījumam nederīgi teikumi. Tādēļ visi atrastie piemēri ir manuāli pārskatīti, kvantitatīvai analīzei izraugoties tikai tos teikumus, kur savienojuma pirmais komponents ir adverbs *daudz*, *vairāk*, *maz* vai *cik* un otrais komponents – lietvārds nominatīvā vai ģenitīvā. Šajos korpusos nav lietderīgi izmantot CQL vaicājumu, lai meklētu piemērus ar adverbīālu izteicēju, jo nav iespējams formulēt precīzu vaicājumu.

2. Individuālā vākuma un korpusa datu kvantitatīvais salīdzinājums

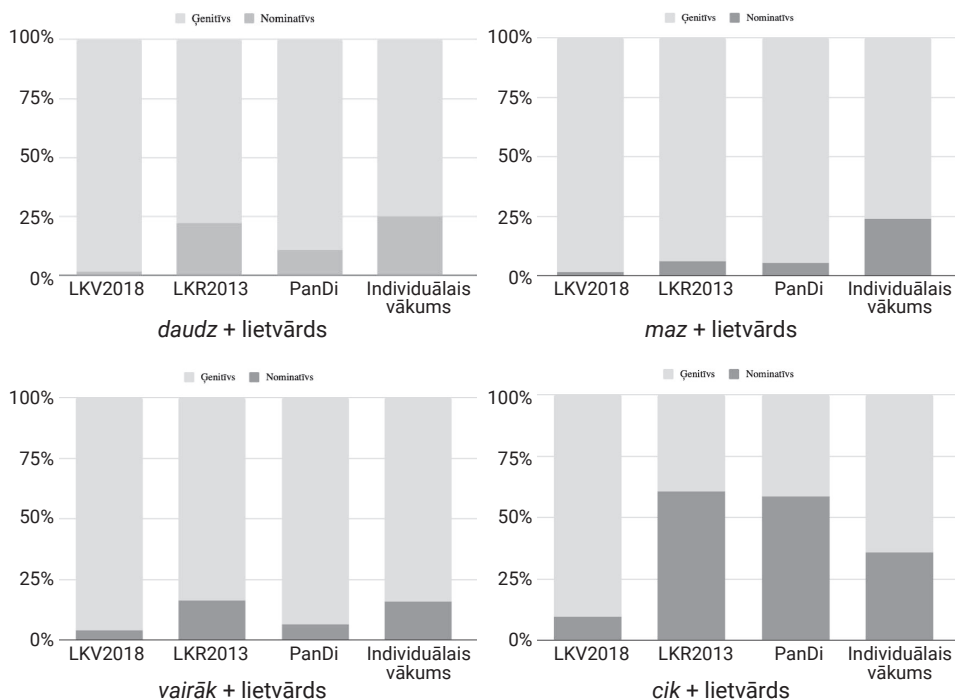
Vienkārša statistiskā analīze, nešķirojot, vai teikumā ir dalāmā ģenitīva konstrukcija (piem., *parādījušās daudz tēmas*) vai adverbīāls izteicējs kopā ar teikuma priekšmetu (piem., *ir ļoti daudz darba*), parāda, ka kopumā gan individuālajā vākumā, gan korpusa datus adverbis *daudz*, tā supletīvā pārākās pakāpes forma *vairāk*, *maz*, *cik* saistījumā ar lietvārdu dominē ģenitīvs, tomēr LVR2013 un PanDi korpusos, kuros iekļauts vairāk sarunvalodas datu, lietvārds nominatīvā gaidāmā ģenitīva vietā sastopams biežāk (sk. 1. attēlu). Izņēmums ir adverbs *cik* – tas LVR2013 un PanDi tomēr nedaudz vairāk lietots kopā ar lietvārdu nominatīvā (11)–(12).

- (11) *[..] tagad protams tas jau automātiski ir ka domā cik kalorijas apmēram tai ēdienā ir [..]* (KD – LVR2013)

- (12) *Laiks ir izjucis nav saprotams cik dienas ir pagājušas, cik dienas vēl priekšā.*
(KD – PanDi)

Analīzē izmantoto piemēru skaits atšķiras (individuālajā vākumā ar katru adverbu apkopoti 100 piemēri, savukārt korpusos atrasto un analizēto piemēru skaits ir cieši saistīts ar korpusa apjomu un vairākos gadījumos pārsniedz pat 1000 piemēru), tāpēc 1. attēla diagrammās norādīts adverba un atbilstošā lietvārda locījuma lietojums procentos. Tā kā izmantoto korpusu apjoms atšķiras (LVK2018 ir 10 milj. vārdlietojumu (12 milj. tekstvienību), LVR2013 ir 1,1 milj. tekstvienību, PanDi ir 565 000 vārdlietojumu (709 000 tekstvienību)), atšķirīgs ir arī no tiem atlasīto un analizēto piemēru skaits.

Likumsakarīgi, ka visvairāk piemēru atlasīti un analizēti no LVK2018: ar *daudz* ir 1957 teikumi (lietvārds ģenitīvā – 98,3 % teikumu (1923 teikumi), lietvārds nominatīvā lietots 1,7 % teikumu (34 teikumi)), ar *vairāk* – 1109 teikumi (ar lietvārdu ģenitīvā – 96,1 % teikumu (1066 teikumi), lietvārds nominatīvā – 3,9 % teikumu (43 teikumi)), ar *maz* – 208 teikumi (ar lietvārdu ģenitīvā – 98,6 % teikumu (205 teikumi), ar lietvārdu nominatīvā – 1,4 % teikumu (3 teikumi)), ar *cik* – 197 teikumi (ar lietvārdu ģenitīvā – 90,4 % teikumu (178 teikumi), ar lietvārdu nominatīvā – 9,6 % teikumu (19 teikumi)). Daudz mazāk analizējamo piemēru iegūts no PanDi un LVR2013, piem., ar adverbu *daudz* no PanDi iegūti un analizēti 283 teikumi (ar lietvārdu ģenitīvā 89 % teikumu (252 teikumi) un ar lietvārdu nominatīvā – 11 % teikumu (31 teikums), no LVR2013 – 284 teikumi (ar lietvārdu ģenitīvā 77,8 % teikumu (221 teikums) un ar lietvārdu nominatīvā – 22,2 % teikumu (63 teikumi)).



1. attēls. Ģenitīva un nominatīva konkurence saistījumā ar adverbiem *daudz*, *vairāk*, *maz*, *cik*

Kvantitatīvā datu analīze parāda, ka gan individuālajā vājumā, gan korpusu datos adverbu *daudz* (arī vārdformas *vairāk*) un *maz* lietojumā kopā ar lietvārdu dominē ģenitīvs. Nedaudz vairāk lietvārds nominatīvā vērojams LVR2013 un individuālajā vājumā, jo šajās datu kopās dominē spontānie mutvārdu teksti. Savukārt, aplūkojot ģenitīva un nominatīva konkurenci lietvārda saistījumā ar adverbu *cik*, LVR2013 un PanDi biežāk sastopams nominatīvs.

3. Sintakses pētījuma rezultāti un diskusija

Šajā pētījumā ir atlasīti trīs adverbi – *daudz* (tā supletīvā pārākās pakāpes forma *vairāk*), *maz*, *cik* –, kuri sīkāk tiek analizēti teikuma gramatiskajā centrā saistījumā ar lietvārda ģenitīvu vai nominatīvu divējādi:

- 1) kā salikts teikuma priekšmets, ko veido adverbs ar daudzuma nozīmi un lietvārds;
- 2) kā teikuma priekšmets pie adverbiāla izteicēja.

Tā kā abos gadījumos literārās valodas norma paredz ģenitīva izmantojumu (<https://mlvv.tezaurs.lv/daudz:1>; <https://mlvv.tezaurs.lv/maz:1>), tam vajadzētu būt atspoguļotam vienkārša teikuma minimālā modeļa klasifikācijā. Kaut arī ir izstrādāti vairāki minēto modeļu saraksti, sintakses speciālistu viedokļi atsevišķos jautājumos atšķiras. Maigone Beitīņa (2009, 92) uzskata, ka „latviešu valodas vienkāršu teikumu predikatīvo modeļu pilnīga saraksta nav; pilnīga inventarizācija nav vēl paveikta”. Atkarībā no teikuma modeļa komponentu vispārinājuma pakāpes minētajos sarakstos tiek piedāvāts dažāds pašu modeļu skaits, kas katrā valodā ir ierobežots. Jāapzinās, ka „modelis atspoguļo tikai teikuma paradigmas pamatformu” (Lokmane 2013, 726). Tādējādi jārēķinās, ka salikts teikuma priekšmets ir zemāka līmeņa sintakses parādība, kas teikuma modelī var neparādīties.

Tomēr, pievienojoties Intas Freimanis uzskatiem (1985, 51), „Latviešu valodas gramatikā” ir atzīmēts verbāla divkomponentu vienkārša teikuma minimālā modeļa realizācijas variants, ko veido nomens ģenitīvā un verbu *nebūt*, *(ne)trūkt*, *(ne)pietikt* 3. personas finītā forma. Tam atbilst teikumi ar saliktu teikuma priekšmetu, ko veido daudzuma apzīmējums, to vidū analizējamie adverbi, kopā ar lietvārdu ģenitīvā. Tādos gadījumos izteicēja pozīcijā ir pieļaujami arī jebkuri citi pilnnozīmes verbi (Lokmane 2013, 728). Piem., individuālajā vājumā un korpusu materiālā konstatētie teikumi, kuros salikts teikuma priekšmets veidots, izmantojot adverbu un lietvārdu ģenitīvā (13)–(15) (teikuma priekšmets ir pasvītrots ar vienu svītru, izteicējs – ar divām).

- (13) *Ļoti maz cilvēku patiesībā iegulda nervus izaugsmei.* (IV – Mārtiņš Krusts @Martinskrusts, Twitter 02.11.2022.)
- (14) *Vai palielinot ministriju skaitu valstij vairāk naudas radīsies?* (IV – Ramona Petraviča, Facebook 03.11.2022.)
- (15) *Tajā iesaistījās maz pagasta iedzīvotāju, galvenokārt tādi, kuru īpašumi atrodas ļoti tuvu pašreizējai ražotnei [...] (KD – LVK2018)*

Adverbiāla divkomponentu vienkārša teikuma minimālā modeļa tips ar teikuma priekšmetu ģenitīvā (16)–(18) ir iekļauts visos apskatītajos sintakses pētnieku darbos ar nelielām modeļa tehniskā atspoguļojuma atšķirībām: $S_{GEN} V_{FIN} COP ADV_{QUANT}$ (Freimane 1985, 60; Beitiņa 2009, 92; Lokmane 2013, 731; Kalnača, Lokmane 2021, 428). Tikai vienā divkomponentu vienkārša teikuma minimālo modeļu sarakstā tādi piemēri ar adverbiālu predikatīvo komponentu kā *Grāmatu ir daudz* un *Palīgu ir maz* atzīti par nomināla divkomponentu vienkārša teikuma minimālā modeļa $N_{GEN} COP_{FIN} COP N_{NOM}$ realizācijas variantiem (Valdmanis 1989, 73), jo sintaktiski semantiskā funkcija nomeniem un mēra adverbiem ir tā pati un, iespējams, morfoloģiskā vārdšķiru atšķirība nav bijis pietiekams pamats atsevišķa papildu teikuma modeļa izdalīšanai.

- (16) *Cilvēku bija ļoti daudz.* (IV – radoraidījumā intervējamā persona, LR1 „Reģionu krustpunktā” 19.01.2022.)
- (17) *Nebūtu jau slikti, ka te laipnības būtu vairāk.* (IV – Vents A Krauklis @VentsAKrauklis, Twitter 13.11.2022.)
- (18) *Šīs tradīcijas cienītāju Daugavpilī kļūst aizvien vairāk.* (KD – LVTB)

Praktiskajā materiālā abos gadījumos ir konstatēti piemēri ar nominatīvu – gan salikta teikuma priekšmeta sastāvā (19)–(21), gan teikuma priekšmeta pozīcijā pie adverbiāla izteicēja (22)–(24).

- (19) *Vairāk smiltis ir savestas.* (IV – radoraidījumā sportists, „Radio SWH” „Vakara sporta stunda” 03.10.2022.)
- (20) *Es domāju drīzāk, ka maz cilvēki pārzina savas tiesības.* (IV – Vaivare @vaivarei, Twitter 20.02.2022.)
- (21) *Ļoti daudz ēdinātāji piedāvā piegādes uz mājām vai līdžņemšanas iespējas.* (KD – PanDi)
- (22) *Tu skaties, ka darba devēji ir ļoti daudz.* (IV – studente, runas situācija augstskolā. LL 19.01.2022.)
- (23) *Vairāk ir tādi, kas saprot, ka uz augšu jādzied vai uz leju.* (IV – radoraidījumā kordirģente, LR1 „Kā labāk dzīvot” 04.11.2022.)
- (24) *Vēl jo vairāk prieks, ja bērniņš ir gaidīts un vesels, atzīst Daugavpils reģionālās slimnīcas vecmāte Iveta Žilvinska [..]* (KD – LVTB)

Sintaktiskajā analizē ir mēģināts apzināt, vai bez raksta ievadā minētā A. Kalnačas apsvēruma par tiekšanos pēc teikuma priekšmeta pamatlocījuma gramatiskajā sistēmā ir vēl kādi papildu faktori, kas ietekmē nominatīva izmantojumu ģenitīva vietā. Individuālā vākuma kartotēkā un PanDi un LVR2013 korpusā ir vairāk spontāni veidotu teikumu, īpaši mutvārdu piemēru sadaļā. Kā zināms, izteikuma dalījums tēmā un rēmā ietekmē vārdu secību, tādējādi lietvārds sintaktiskajā saistījumā ar mēra adverbu var atrasties arī samērā tālu viens no otra. Ne vienmēr attālināta vārdu secība veicina literārās valodas normas neievērošanu. Piem., rakstveida tekstā (25), (26) ir lietots ģenitīvs.

- (25) *Ziņu par notikumiem frontē pagaidām maz.* (IV – Sandra Veinberg, Facebook 16.10.2022.)

- (26) *Sabilē brūno plāksnīšu ar uzrakstiem par nozīmīgām vietām ir daudz, un šaurajās ieliņās tās pamanāmas vēl jo labāk.* (KD – LVTB)

Vairākos piemēros ar adverbu *cik* lietvārds nominatīvā iespējams ir tādēļ, ka valodas lietotājs tiecas saskaņot to ar izteicēju, kurš izteikts ar salikto darbības vārda formu, t. i., ar divdabja formu un palīgdarbības vārdu *būt* personas formā, kuru mēdz arī izlaist (27)–(29).

- (27) *Zinu, tikai to, cik dienas ir atlikušas līdz kursa darba iesniegšanas termiņa beigām, bet ne to, kas par dienu.* (KD – PanDi)
- (28) *Tāda sajūta kā cietumā, kad uz sienām jāskrāpē cik dienas pagājušas...* (KD – PanDi)
- (29) *Es nezinu, cik dienas karantīnā aizvadītas, es tās neskaitu.* (KD – PanDi)

Dažkārt adverba novietojums, kas šīs nelokāmās vārdšķiras vārdiem ir galvenais sintaktisko attieksmju izteikšanas līdzeklis, var maldināt. Tas redzams mutvārdu piemēros ar lietvārdu nominatīvā. Teikuma autors informatīvā ziņā ir uzsvēris mēra adverba saistījumu ar verbu – *maz aizpildīja* (30), *maz zin* (31) – vai adjektīvu sastata izteicēja funkcijā – *ir vajadzīgas vairāk* (32) –, taču vārdu secībā attiecīgā noformējuma trūkst un izskatās adverba saistījums ar lietvārdu. Tādējādi šie teikumi ir interpretējami divējādi.

- (30) *Kāpēc tik **maz** cilvēki aizpildīja?* (IV – Ēriks @ekolsund, Twitter 14.09.2022.)
- (31) *[..] cik **maz** cilvēki zin Hariju Porteru.* (IV – agija kola-kanča @agijakola, Twitter 24.09.2022.)
- (32) *Mums ir vajadzīgas **vairāk** sarunas klātienē!* (IV – alpīnists Kristaps Liepiņš @vx_lv, Twitter 05.11.2022.)

Jāsecina, ka ne vienmēr spontānam mutvārdu tekstam raksturīgās īpatnības vārdu novietojumā nosaka lietvārda locījuma izvēli. Piemēri, kuros var konstatēt sarunvalodai tipisko vārdu secību, ir sastopami gan ar nominatīvu (33), gan ar ģenitīvu (34).

- (33) *Nespēju atcerēties, **cik** tad tie precīzie cipari bija apmeklētāji.* (IV – raidījuma vadītājs, „Radio SWH” „Brīvdienu rīti ar Artūru Krūzkopu” 10.04.2022.)
- (34) ***Daudz** jums istabu kurināmu?* (IV – raidījuma vadītājs, LR2 „Cita lieta” 15.01.2022.)

Iespējams, ka sava loma ir runātāju automatizētajām latviešu valodas prasmēm, respektīvi, runātājs vai rakstītājs iepriekš nepārdomā un nepievērš uzmanību izteikumā nepieciešamā vārda formai, bet to producē automātiski.

Šajā pētījumā ir pievērsta uzmanība arī leksiskajam aspektam un apsvērta doma par ūzusu, proti, vai parastam valodas līdzekļu lietojuma veidam tautas valodā ir kāda nozīme. Regulārs ģenitīva lietojums stabilās vārdkopās, kas ikdienas saziņā tiek automātiski producētas līdzīgi formulām, varētu nostiprināt normas stabilitāti. Piem., visos fiksētajos individuālā vākuma gadījumos nedalāmajā vārdkopā *daudz laimes* atkarīgais komponents ir ģenitīvā. Arī visos korpusos atrastajos piemēros ar vārdkopu *daudz laimes* lietvārds ir ģenitīvā.

- (35) **Daudz laimes, mans Draugs!** (IV – Mārtiņš Bergmanis, Facebook 01.11.2021.)
- (36) **Dāmas, daudz laimes jums vārda dienā!** (IV – radioraidījuma vadītājs, „Radio SWH” „Brīvdienu rīti ar Arturu Krūzkopu” 23.01.2022.)
- (37) **Šovakar pie tējas, kūciņas un sniega būs daudz laimes!** (IV – vetriks^{^^} @AiraVetric, Twitter 27.12.2022.)

Ir vērts pieminēt Jēkaba Graubiņa dziesmu „Daudz baltu dieniņu”, kuru joprojām tautā dzied un kuras tekstā vairākkārt atkārtojas adverba *daudz* saistījums ar lietvārdu ģenitīvā.

Vēlējuma teikumos (35), (36) predikāts mēdz būt sintaktiskajā nullformā. Ja nav verbalizēta darbības vārda *vēlēt* vai *novēlēt* forma izteicēja pozīcijā, saliktajam teikumam loceklim *daudz laimes* ir akuzatīva nozīme, kurā paturēts ģenitīvs atbilstoši literārās valodas normai (plašāk par novēlējumu sintaktiskajiem variantiem sk. Lauze 2012). Tomēr arī netipiskā subjekta pozīcijā (37) ir izmantots ģenitīvs.

Piemēros ar stabilu vārdkopu *vairāk informācijas* (38), (39) nav iespējams noteikt, vai predikāta sintaktiskā nullforma būtu kāda no verbu (*skatīties, meklēt, lasīt* u. tml.) personu formām, kas piesaistītu objektu, resp., minētā vārdkopa būtu lietota akuzatīva nozīmē, vai tajā nav verbalizēta kāda divdabja forma (*pieejams, apskatāms, iegūstams* u. tml.), kas liecinātu par subjekta funkciju.

- (38) **Vairāk informācijas šeit [..]** (IV – Liepājas teātris @Liepajasteatris, Twitter 25.10.2022.)
- (39) Radioraidījuma vadītāja: **Vairāk informācijas zolmaniem.lv.** (IV – „Radio SWH” „Beta beidz nedēļu” 11.11.2022.)
- (40) **Vairāk informācija** [hipersaite] (IV – Mārdadzis, Facebook 18.05.2022.)

Tikai vienā individuālā vākuma gadījumā (40) sociālajos medijos konstatēts piemērs ar nominatīvu, bet ir iespējama arī neuzmanības kļūda, ja pēdējais burts palicis neuzrakstīts. Nominatīvs konstatēts arī trijos LVK2018 piemēros (41)–(43).

- (41) *[..] arī Klientu apkalpošanas speciālistiem, operatoriem ļoti reti ir pieejama vairāk informācija nekā klientam, jo arī viņiem nav piekļuves uzņēmumu norēķinu vai finanšu grāmatvedības sistēmām.* (KD – LVK2018)
- (42) *[..] svētku izdevumiem palīdzēs Naudas plānošanas centra sadarbībā ar GE Money Bank izstrādātais Naudas kontrolieris (vairāk informācija: www.vissparnaudu.lv sadaļā Par mums).* (KD – LVK2018)
- (43) *Ja interesē vairāk informācija par EKO, vai OUTLAST tehnoloģiju EKO zīmola produktos, jautājiet gan šeit māmiņuklubā, gan arī oficiālā pārstāvja ViKrEm mājas lapas kontaktos.* (KD – LVK2018)

Var piebilst, ka sociālajos medijos izmantotie tēmturi ar mēra apstākļa vārdiem var kalpot par paraugu citiem latviešu valodas lietotājiem ģenitīva saglabāšanā, sevišķi tviterī, kas atšķiras no citiem sociālajiem medijiem ar to, ka kopš tā pirmsākumiem „Latvijā aktualizējās prasība pēc tvītu valodiskās kvalitātes, to autoriem aizvien biežāk tika un tiek norādītas viņu kļūdas” (Brice 2019, 77). Piem., tēmtura pamatā ir izmantots kāds sauklis (44), (45) vai restorāna nosaukums (46).

- (44) [pārdod jau notūrītas ogas] *#mazākdarba #vairākprieka* (IV – Inguna Ziemere, Facebook 05.07.2021.)
- (45) *#VairākLētasGaismas* (IV – Ainārs Zelčs @Eraksti, Twitter 13.11.2022.)
- (46) *Liels paldies #vairāksaules par iespēju labi paēst.* (IV – Inga Martinsone @MartinsonEInga, Twitter 25.08.2021.)

Tomēr sintakses skatījumā jāpiekrīt Aksela Holvūta (*Axel Holvoet* 2012, 95) viedoklim, ka primārais ir teikuma strukturālais modelis, kurā ir ietverts lietvārds noteiktā locījumā un „atsevišķas leksēmas nevar pārvaldīt sevis pieprasītos locījumus”.

Secinājumi

Gan korpusu, gan individuālā vākuma datiem ir priekšrocības un trūkumi, tādējādi vislabākos rezultātus var iegūt, apvienojot abas praktiskā materiāla ieguves metodes. Individuālā vākuma autentiskie sarunvalodas piemēri, kas korpusa datus var neiekļūt vispār vai parādīties pēc laika, ir noderīgs, kaut arī apjoma ziņā neliels pētījuma avots.

Gan korpusu, gan individuālā vākuma materiālā daļāmā ģenitīva konstrukcijās un arī adverbiālā izteicējā dominē ģenitīvs. Nominatīvs biežāk konstatēts *Latviešu valodas runas atpazīšanas korpusā* (LVR2013), analizēto trīs adverbu grupā – biežāk saistījumā ar adverbu *cik*.

Nominatīva izmantojumu ģenitīva vietā sekmē gan gramatiskās sistēmas ietekme, kur nominatīvs ir teikuma priekšmeta pamatlocījums, gan vēl pētāmi daži citi papildu faktori – piem., mutvārdu teikumos maldinošs mēra adverbā novietojums līdzās lietvārdam, kaut gan, šķiet, domāts tā saistījums ar verbu vai adjektīvu.

Pētījuma rezultāti ar ģenitīva lietojuma procentuālo pārsvaru ir pārsnieguši autoru negatīvās prognozes. Iespējams, ka regulārs ģenitīva lietojums stabilās adverbiālās vārdkopās, kā arī sociālo mediju tēmturos, kas ikdienas saziņā līdzīgi formulām tiek automātiski producēti, varētu nostiprināt latviešu valodas literārās normas stabilitāti.

Saīsinājumi

ADV	apstākļa vārds
COP	saitiņa
FIN	finīta forma
GEN	ģenitīvs
IV	individuālais vākums
KD	korpusa dati
LL	Lindas Lauzes individuālā vākuma kartotēka
LRK2013	<i>Latviešu valodas runas atpazīšanas korpus</i>
LR1	radiostacija „Latvijas Radio 1”
LR2	radiostacija „Latvijas Radio 2”
LVK2018	<i>Līdzsvarotais latviešu valodas tekstu korpus</i>
LRK2013	<i>Latviešu valodas runas atpazīšanas korpus</i>

LVTB	<i>Latviešu valodas sintaktiski marķētais korpus</i>
N	nomens
NOM	nominatīvs
PanDi	<i>Pandēmijas dienasgrāmata</i>
QUANT	daudzums
Radio SWH	radiostacija „Radio SWH”
S	lietvārds
V	darbības vārds

Avoti

- Andronova, Everita, Auziņa, Ilze. 2022. Korpuslingvistika. *Nacionālā enciklopēdija*. Pieejams: <https://enciklopedija.lv/skirklis/101749-korpuslingvistika>
- Baltiņš, Māris, Druviete, Ina. 2017. *Ceļavējš cilvēku ciltij: valoda sabiedrībā*. Rīga: Latviešu valodas aģentūra.
- Brice, Līva. 2019. #valoda tviterī – 140 zīmju izaicinājums un tiešsaistes konsultācijas. *Valodas prakse: vērojumi un ieteikumi*. 14, 67–78.
- Jērāne, Santa et al. 2023. *Mūsdienu latviešu valodas vārdnīca*. CLARIN-LV digitālā bibliotēka. Latvijas Universitātes Matemātikas un informātikas institūts. Pieejams: <http://hdl.handle.net/20.500.12574/88>
- Kalnača, Andra. 2022. Pārmaiņas un normēšana latviešu gramatikā. *Valodas prakse: vērojumi un ieteikumi*. 17, 91–105.
- Latviešu valodas runas atpazīšanas korpus (LRK2013)*. 2013. Rīga: Latvijas Universitātes Matemātikas un informātikas institūts, SIA Tilde, SIA LETA. Pieejams: <https://korpus.lv/id/LRK2013>
- Levāne-Petrova, Kristīne, Dargis, Roberts. 2018. *Līdzsvarotais mūsdienu latviešu valodas tekstu korpus (LVK2018)*. CLARIN-LV digitālā bibliotēka. Latvijas Universitātes Matemātikas un informātikas institūts. Pieejams: <http://hdl.handle.net/20.500.12574/11>
- Reinsone, Sanita, Ļaksa-Timinska, Ilze, Jaudzema, Justīne. 2022. *Pandēmijas dienasgrāmata. 2020–2022*. CLARIN-LV digitālā bibliotēka. Latvijas Universitātes Matemātikas un informātikas institūts. Pieejams: <http://hdl.handle.net/20.500.12574/48>
- Rituma, Laura et al. 2013. *Latviešu valodas sintaktiski marķētais korpus (LVTB)*. CLARIN-LV digitālā bibliotēka. Latvijas Universitātes Matemātikas un informātikas institūts. Pieejams: <http://hdl.handle.net/20.500.12574/86>

Literatūra

- Beitiņa, Maigone. 2009. *Mūsdienu latviešu literārās valodas sintakse*. Liepāja: LiePA.
- Berg-Olsen, Sturla. 2005. *The Latvian Dative and Genitive: A Cognitive Grammar Account*. Oslo: University of Oslo.
- Berg-Olsen, Sturla. 2009. Lacking in Latvian: Case variation from a cognitive and constructional perspective. *The Role of Semantic, Pragmatic, and Discourse Factors in the Development of Case*. Barðdal, Jóhanna, Chelliah, Shobhana L. (eds.). Amsterdam / Philadelphia: John Benjamins Publishing Company, 181–202. <https://doi.org/10.1075/slcs.108>
- Ernstson, Vineta, Tidriķe, Laura. 2006. *Jauniešu valoda*. Rīga: Latvijas Universitātes Akadēmiskais apgāds.

- Freimane, Inta. 1985. *Vienkāršs teikums un tā paplašināšana*. Rīga: Latvijas Valsts universitāte.
- Grūzītis, Normunds, Znotiņš, Artūrs. 2018. Multilayer corpus and toolchain for Full-Stack NLU in Latvian. *Proceedings of the CLARIN Annual Conference*. Skadiņa, Inguna, Eskevich, Maria (eds.), Pisa: CLARIN, 61–65. Pieejams: https://office.clarin.eu/v/CE-2018-1292-CLARIN2018_ConferenceProceedings.pdf
- Holvūts, Aksels. 2012. *Vispārīgās sintakses pamati*. Sirmule, Lāsme (tulk.). Rīga: Latviešu valodas aģentūra.
- Kalnača, Andra. 2002. Lietvārda locījumu formu konkurence latviešu valodā. *Linguistica Lettica*. 10, 135–150.
- Kalnača, Andra. 2007. Ģenitīva sinonīmijas stilistikās funkcijas mūsdienu latviešu valodā. *Acta Humanitarica Universitatis Saulensis*. 3, 68–78.
- Kalnača, Andra. 2019. Vai latviešu gramatika mainās? *Valsts valodas komisijas raksti*. 10, 259–276. Pieejams: <https://www.president.lv/lv/media/5055/download?attachment>
- Kalnača, Andra, Lokmane, Ilze. 2018a. Teikuma priekšmeta locījums un vārdu secība latviešu valodas eksistenciālteikumos. *Dzīves lingvistika. Veltījums profesoram Jānim Valdmanim*. Vītola, Inita (red.). Rīga: Latvijas Universitātes Akadēmiskais apgāds, 207–221.
- Kalnača, Andra, Lokmane, Ilze. 2018b. Subject case alternation in negated existential, locative, and possessive clauses in Latvian. *Kalbotyra* (71), 42–63. <https://doi.org/10.15388/Kalbotyra.2018.3>
- Kalnača, Andra, Lokmane, Ilze. 2021. *Latvian Grammar*. Riga: University of Latvia Press. <https://doi.org/10.22364/latgram.2021>
- Kalnača, Andra, Lokmane, Ilze. 2022. Partitive genitive constructions and agreement variations in Latvian. *Linguistic Variation*. 23 (1), 75–94. <https://doi.org/10.1075/lv.20019.kal>
- Kalnača, Andra, Lokmane, Ilze, Metslang, Helena. 2019. Subject case alternation in Latvian and Estonian existential clauses. *Eesti Rakenduslingvistika Ühingu aastaraamat / Estonian Papers in Applied Linguistics*. 15, 53–82. <http://dx.doi.org/10.5128/ERYa15.04>
- Labov, William. 1972. The study of language in its social context. *Studium Generale*. 23, 30–87.
- Lagzdīņa, Sarmīte. 1997. Verbs *nebūt* morfoloģiskajā paradigmā un teikumā. *Savai valodai. Latvijas Zinātņu akadēmijas goda loceklim Rūdolffam Grabim veltīts piemiņas krājums*. Blinkena, Aina (red.). Rīga: Latvijas Zinātņu Akadēmijas Vēstis, 165–200.
- Lauze, Linda. 2004. *Ikdienas saziņa: vienkāršs teikums latviešu sarunvalodā*. Liepāja: LiePA.
- Lauze, Linda. 2008. *Izpētes materiāla vākšana sociolingvistikā*. Metodiski norādījumi. Liepāja: LiePA.
- Lauze, Linda. 2012. Novēlējumu sintaktiskie varianti mutvārdu saziņā. *Baltistica. Priedas*. VIII, 79–88. <http://dx.doi.org/10.15388/baltistica.0.8.2112>
- Lokmane, Ilze. 2013. Vienkārša teikuma formālā (strukturālā) organizācija. *Latviešu valodas gramatika*. Nītiņa, Daina, Grigorjevs, Juris (red.). Rīga: Latvijas Universitātes Latviešu valodas institūts, 710–766.
- Lokmane, Ilze. 2017. Partitīva konstrukciju sintaktiskās īpatnības latviešu valodā. *Valoda: nozīme un forma*. 8, 91–99. <https://doi.org/10.22364/VNF.8.7>
- McEnery, Tony, Hardie, Andrew. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Rituma, Laura, Saulīte, Baiba, Nešpore-Bērzkalne, Gunta. 2019. Latviešu valodas sintaktiski marķētā korpusa gramatikas modelis. *Valoda: nozīme un forma*. 10, 200–216. <https://doi.org/10.22364/vnf.10.16>

- Saulīte, Baiba et al. 2022. Latvian National Corpora Collection *Korpuss.lv. Proceedings of the Language Resources and Evaluation Conference*. Calzolari, Nicoletta et al. (eds.) Marseille: European Language Resources Association, 5123–5129.
- Saulīte, Baiba, Dargis, Roberts, Auziņa, Ilze. 2022. Nacionālā korpusu kolekcija *Korpuss.lv. Linguistica Lettica*. 31 (iespiešanā).
- Smiltneice, Gunta. 2013. Lietvārds (substantīvs). *Latviešu valodas gramatika*. Nītiņa, Daina, Grigorjevs, Juris (red.). Rīga: Latvijas Universitātes Latviešu valodas institūts, 324–369.
- Spolsky, Bernard. 2003. *Sociolinguistics*. 4th ed. Oxford: Oxford University Press.
- Valdmanis, Jānis. 1989. Vienkāršs teikums strukturālās sintakses izpratnē. Ceplītis, Laimdots, Rozenbergs, Jānis, Valdmanis, Jānis. *Latviešu valodas sintakse*. Rīga: Zvaigzne, 62–107.
- Znotiņš, Artūrs, Bārzdīņš, Guntis. 2020. LVBERT: Transformer-based model for Latvian language understanding. *Human Language Technologies – The Baltic Perspective*. Utka, Andrius, Vaičenonienē, Jurgita et al. (eds.). Amsterdam: IOS Press, 111–115. <http://dx.doi.org/10.3233/FAIA200610>

Summary

The article describes the advantages and disadvantages of corpus data and individual collection. The availability of various grammatically annotated corpora of the Latvian language ensures more and more extensive grammar studies based on corpus data. On the other hand, the individual collection played a major role in the development of linguistics, and it is an older way of obtaining practical material. However, in today's technological age, the individual usefulness of the collection has come into question.

For a practical comparison of the two data acquisition methods, a common phenomenon in modern Latvian language usage was chosen – genitive and nominative competition (in connection with an adverb), which was found both in the individual collection and in the corpora data. In this study, three adverbs are selected – *daudz* ‘many’ (wordform *vairāk* ‘more’) *maz* ‘few’, *cik* ‘how many’ – which are analysed in greater detail in the syntactic centre of the sentence in connection with the genitive or nominative of the noun. The individual collection consists of relatively spontaneous unedited use of the Latvian language in speech and writing – 100 sentences with each adverb. For corpus-driven data analysis, four corpora of the Latvian language were used: *The Balanced Corpus of Modern Latvian* (LVK2018), *Latvian Treebank* (LVTB), *Latvian Speech Recognition Corpus* (LRK2013), and *Corpus of Latvian Pandemic Diaries* (PanDi). The phrases with the genitive form dominate the material of both the corpora and the individual collection. According to the used sources, nominative is more frequent in the *Latvian Speech Recognition Corpus* (LVR2013), but in the group of three analysed adverbs – more often in connection with the adverb *cik* ‘how many’.

Keywords: Latvian; individual collection; corpus; syntax; genitive and nominative competition.



Rakstam ir Creative Commons Attiecinājuma 4.0 Starptautiskā licence (CC BY 4.0) /

This article is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>)