

# Līdzsvarotais mūsdienu latviešu valodas tekstu korpus 2018

Kristīne Levāne-Petrova

Praktiskas ievirzes pētījumu projekts “Daudzslāņu valodas resursu kopa teksta semantiskai analīzei un sintēzei  
latviešu valodā” nr.1.1.1.1/16/A/219

2018. gada 27. aprīlī

NACIONĀLAIS  
ATTĪSTĪBAS  
PLĀNS 2020



EIROPAS SAVIENĪBA

Eiropas Reģionālās  
attīstības fonds

IEGULDĪJUMS TAVĀ NĀKOTNĒ

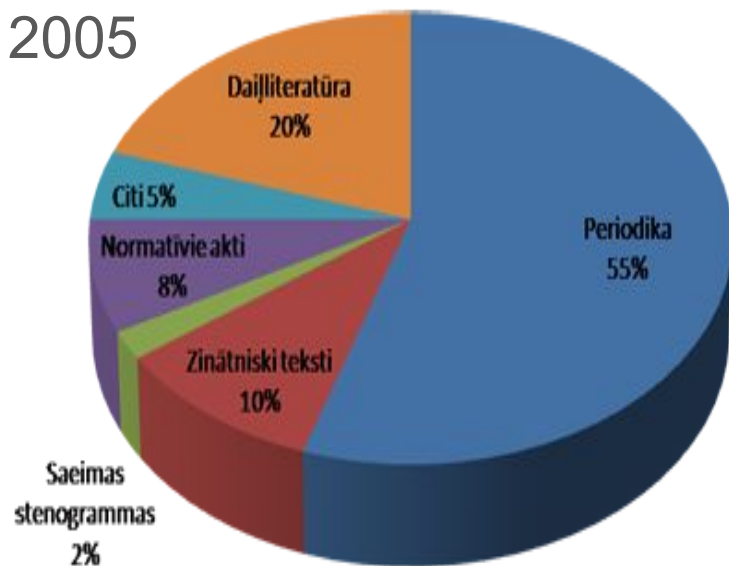


Mākslīgā intelekta laboratorija

LU MII

# Līdzsvarotais mūsdienu latviešu valodas tekstu korpus (LVK)

- Latviešu valodas korpusa koncepcija 2005
- LVK tiek veidots kopš 2007. gada
- 2013. gadā LVK 2013 sasniedza 4,5 miljonus vārdlietojumu (ar LVA atbalstu)

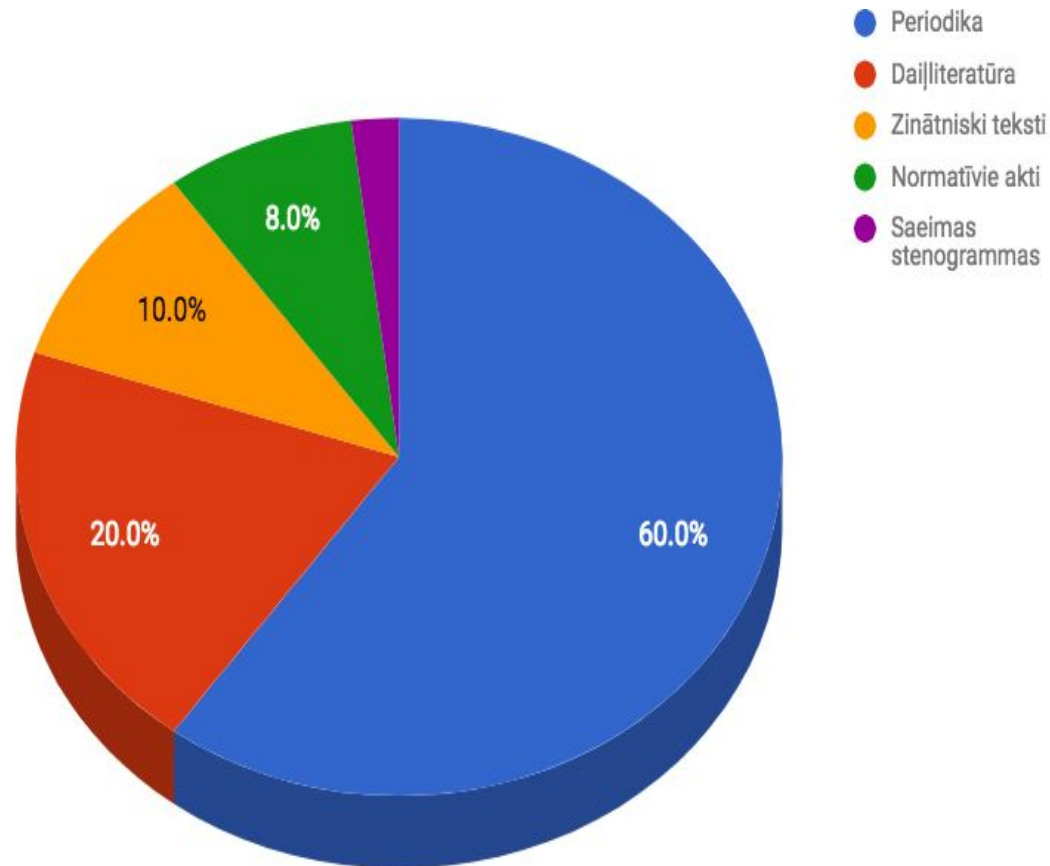


# Līdzsvarotā mūsdienu latviešu valodas tekstu korpusa paplašināšana

- LVK2013 → LVK2018 (4,5 milj. → 10 milj.)
- Tiek veidots, pamatā ievērojot LVK2013 uzbūves principus
- Precizēti iekļaujamie metadati
- Automatizēta tekstu atlase

# Līdzsvarotais mūsdienu latviešu valodas tekstu korpuss (LVK2018)

- Vispārīgs
- Līdzsvarots
- Papildināms
- Pievienoti metadati
- Automātiski morfologiski marķēts



## Tekstu atlases kritēriji (1)

- **Valodas paveids**

Daiļliteratūra, periodika, zinātniski teksti utt.

- **Laiks**

Teksti, kas publicēti kopš 1991.gada.

- **Orīginālteksti**

Korpusā netiek iekļauti tulkojumi.

- **Autentiski teksti**

Teksti tiek iekļauti korpusā tādi, kādi tie ir — bez labojumiem.

## Tekstu atlasē kritēriji (2)

- **Valodas parauga apjoms**

Gan pilni teksti, gan tekstu fragmenti.

Valodas parauga izmērs nedrīkst pārsniegt 5% no korpusa sadaļas apjoma.

- **Pieejamība**

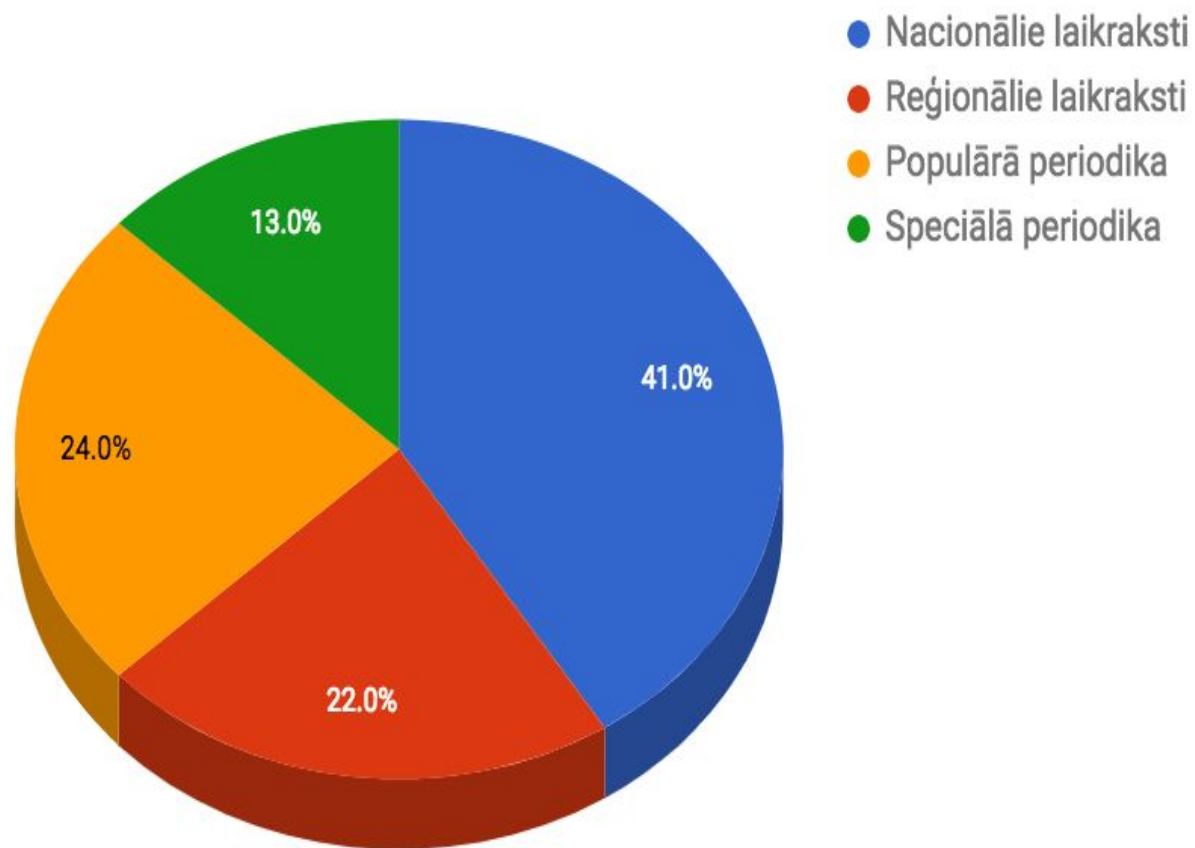
- **Formālas atbilstības princips u.c.**

# LVK2018 periodika

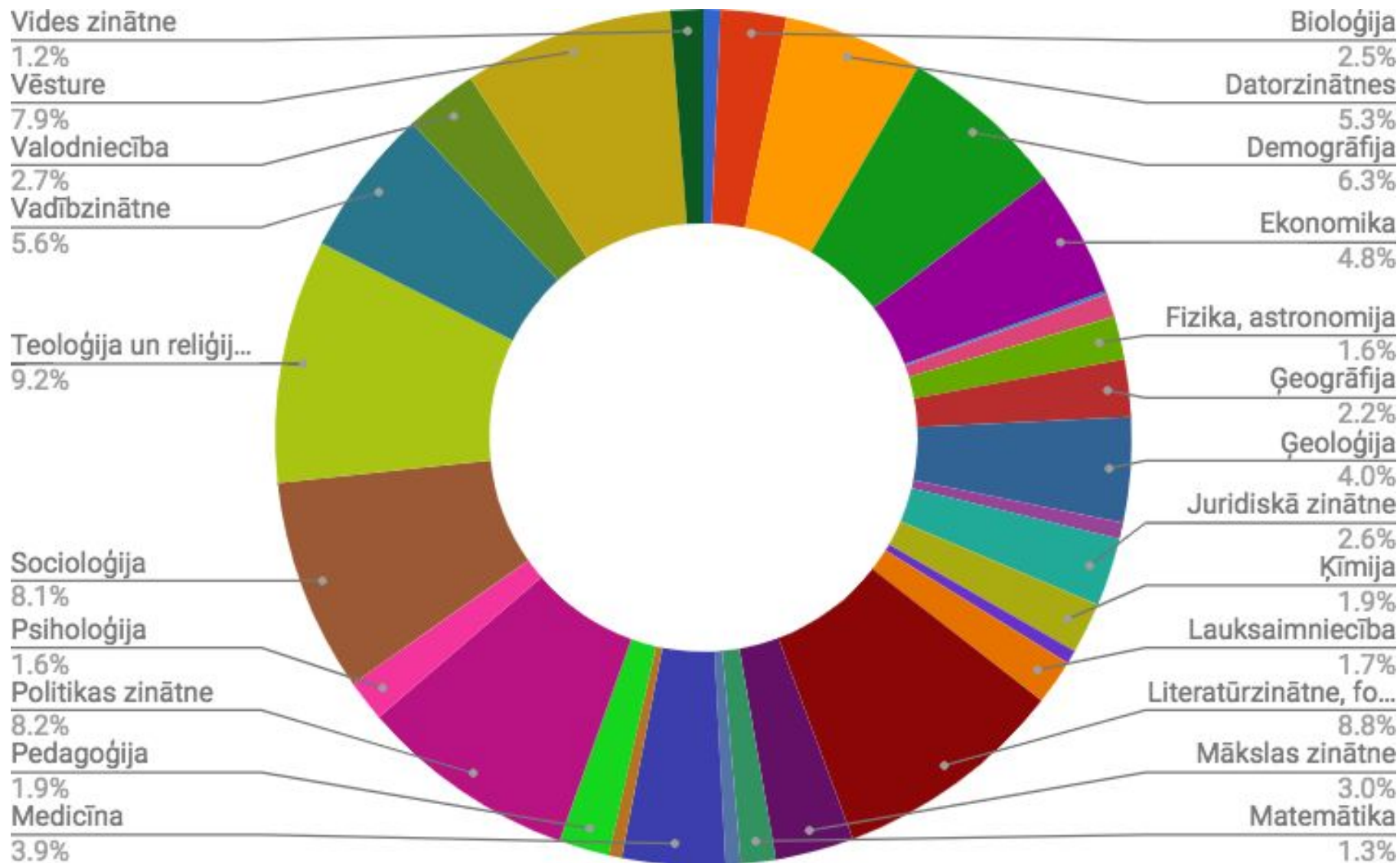
Tematiskais  
daudzveidīgums

Ģeogrāfija

Unikalitāte

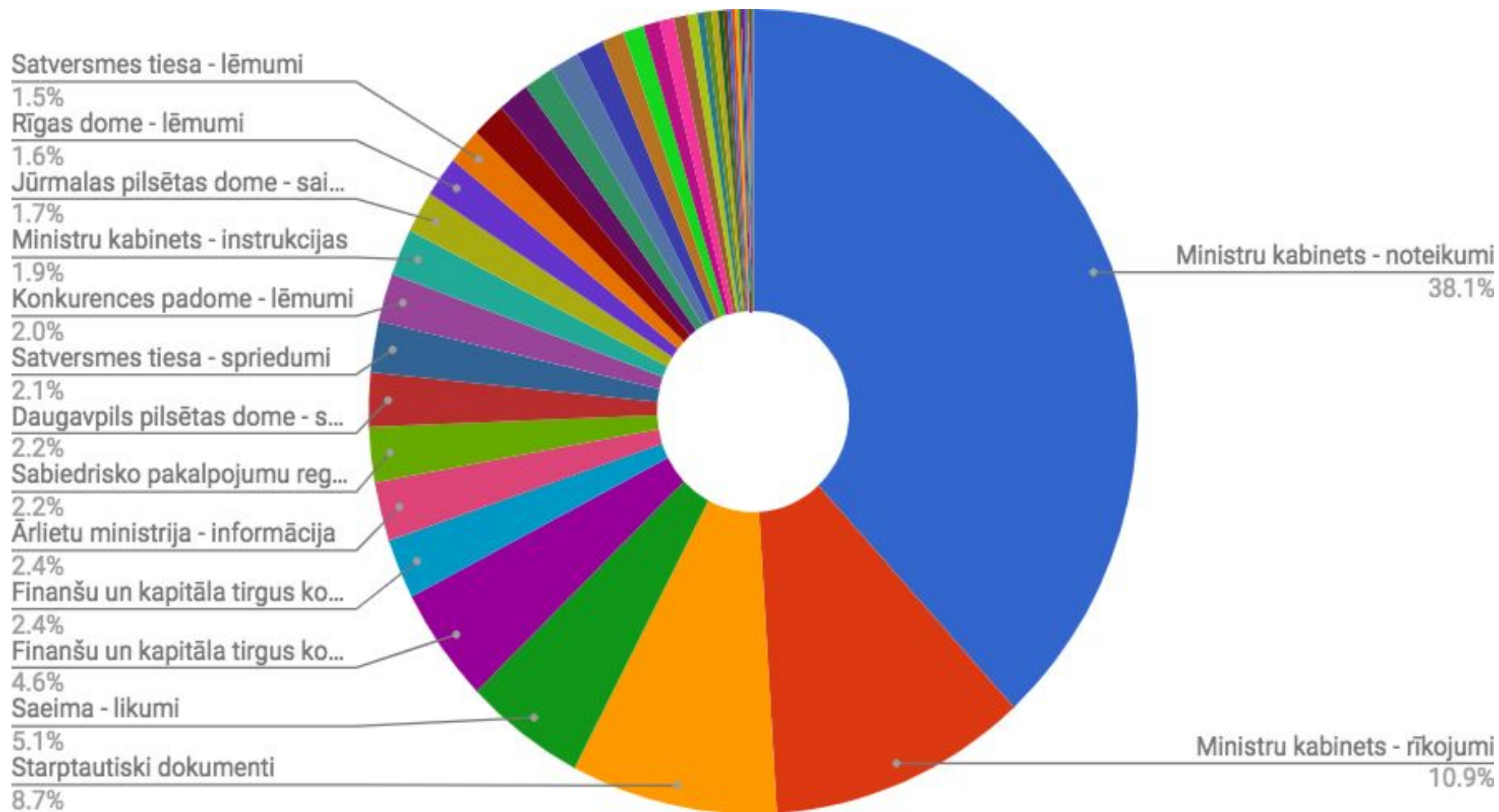


# LVK2018 zinātniski teksti



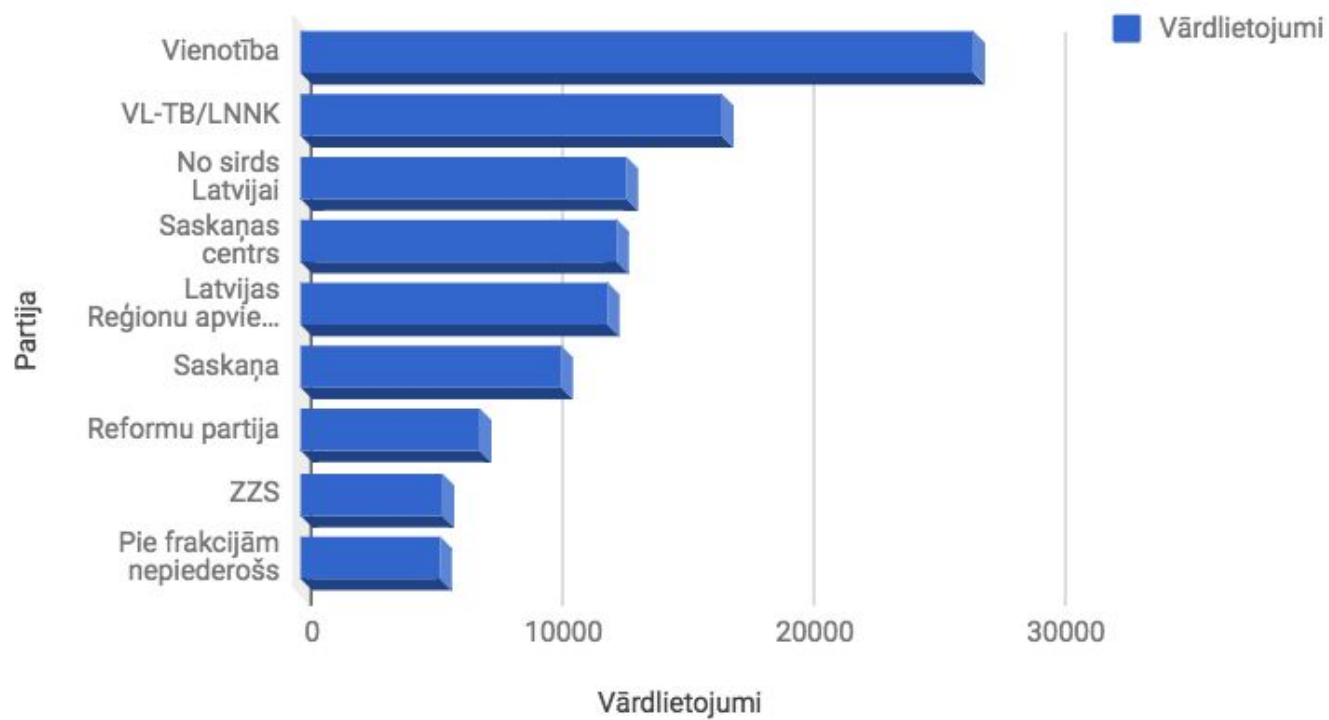


# LVK2018 normatīvie akti



# LVK2018 Saeimas stenogrammas

## Vārdlietojumi un partija



# Metadati

Katram korpusa avotam tiek pievienoti metadati, piemēram,

- Avota autors
- Avota nosaukums
- Izdevējs
- Izdošanas gads

# Kā atsaukties uz korpusu?

Līdzsvarotais mūsdienu latviešu valodas tekstu korpus

LVK2013

LVK2018(alpha)

# Paldies par uzmanību!

www.korpuss.lv

268235	burtu " k" un defisi, aiz defises norādot	korpusa	numuru ( 20.2. apakšpunkts). </p><p> [2.9
z71_0	sagatavot pētījumam tekstu ( arī audio)	korpusu	, tas ir, apkopot, sistematizēt un apstrādāt
z72_0	ģenerālgubernatora pils jaunais rietumu	korpuss	; nozīmīga celtnie rātsnama kompleksā bija
d121_0	tramvaja logu lūkodamās uz tukšajiem fabriku	kopusiem	. </p><p> Kad Millija bija aprakta, Florence
d123_0	, kad sadzirdējām medmāsu kliedzienu no	korpusa	, tad priecīgi saskatījāmies, jo zinājām
d131_0	spokoties pamestu rūpnīcu un neražošanas	korpusu	, nedaceltu būvju, un tas viss kādam piederēja
d165_0	griežas apkārt, cenšoties ieraudzīt rūpnīcas	kopusus	. Taču kārklu un nezāļu puduri ir lielāki
d71_0	sētu, bet caurspīdīgu, vienā pusē slimnīcas	korpuss	, otrā – mežs. Bet blakus placī pastaigājās
p107_0	pielietojums to senlaicīgā un greznā koka	korpusa	dēļ. Tolaik radio korpuss bija no riekstkoka
p107_0	un greznā koka korpusa dēļ. Tolaik radio	korpuss	bija no riekstkoka, vēlāk padomju laikos
d57_0	raibās kleitas. Bodītes lillā krāsotais	korpuss	izrādījās vecs kuģa bunkurs no kniedēta
p1459_0	turpmāk jāpieliek roka – slimnīcas piecstāvu	kopusam	un ambulatorajai daļai nepieciešama siltināšana
p1461_0	Rāviņi visus šos 13 gadus mitinājušies kūts	kopusā	un arī tagad turpina to darīt, tā viņiem
p1482_0	pakalpojumu uzlabošana, slimnīcas trīsstāvu	korpusa	celtniecības uzsākšana. Atbalsts vēlams
p1515_0	Rudzātu speciālās internātskolas mācību	korpusa	atjaunošana, Salenieku pansionātam tiks
p180_0	visticamāk tapis ekstrēmā veidā. " Gar ( rūpnīcas	korpusa	nogruvušās sienas) ārmalu ved kāpnes līdz
p1893_0	iekārtots jaunā, plašā, gaišā telpā jaunā	korpusa	pirmajā stāvā. </p><p> Centra vadītāja Mirdza
p2171_0	gadījumā iespējams salocīt gar automašīnas	korpusa	sāniem. </p><p> Konstruktors apgalvoja, ka
p2510_0	visas sev vajadzīgās paroles. Arī atslēgas	korpuss	ir drošs pret mehāniskiem bojājumiem, jo
p2252_0	izpaušmes pilsētas punktus - Spīķerus un VEF	kopusus	, K. Barona ielu un tās apkārtni. </p><p>