

*“Tas/ kas jau fenn=deeninņam pirmohs laikohs
ghaidiets”*: “SENIE” 20 gados

Everita Andronova
(LU Matemātikas un informātikas institūts
/ Sanktpēterburgas Valsts universitāte)

Latviešu raksti un raksti Baltijā 16.–19. gadsimtā:
pētniecības aktualitātes un problēmas
14.01.2022

Priekšvēsture: nesanāk *it kā ar vienu vārdu (un īsi)*

XX gs. 90.g. LU MII: seno tekstu datu bāze

Korpusa izveide 2002.g., pateicoties LU Pētniecības projektu konkursam

Korpusa dienas un nedienas

2012.g.: jauni horizonti un papildspēki

2017.g.: beidzot Unicode

2021–2022: Aiziešana no tēva mājām – pakāpeniska pārcelšanās uz jaunu platformu

- 1936. g. izskanēja filologu biedrības „Ramave“ ierosinājums latviešu valodas vēsturiskās vārdnīcas labā, **aicinot visus valodas draugus piedalīties vārdu krāšanā**
- 60. gados Velta Rūķe-Draviņa vēlreiz aktualizēja seno tekstu vārdnīcas nozīmi, kas pavērtu **„tālas perspektīvas latviešu valodas vēsturē“** (Rūķe 1961: 11)
- atjaunotā Latvijas Filologu biedrība 1989. g. nolēmj vākt materiālu latviešu tēzauram **no visas pasaules** latviešu avotiem
- par latviešu valodas vēsturisko vārdnīcas ieceri Vispasaules latviešu zinātņu kongresā un citur referēja Juris Balduncīks (Blinkena 1991; Balduncīks 1994), arī Maija Baltiņa (1997; 2001)
- XX gs. ideja netika īstenota: „Diemžēl nav izveidots latviešu valodas tēzaurš, kurā varētu vērot ikviena latviešu valodas vārda rašanās resp. fiksācijas laiku un šo vārdu un formu attīstību vismaz rakstu valodas posmā“ (Blinkena 1996: 42)



Dr.phys. **Andrejs Spektors** kopš 1987.g. LU Lietišķās informātikas daļas vadītājs.

1992.g. LU tika izveidota Mākslīgā intelekta laboratorija (saukta par AILabu).

Līdz 2015.g. to vada A.Spektors, kas vēl šodien turpina strādāt Laboratorijā un ir saistīts ar tezaurs.lv attīstību.

Vadmotīvs: Valoda, kuras nebūs virtuālajā pasaulē, būs lemta iznīcībai

Labvēlīga un rosīga vide: datorzinātņu un filoloģijas studenti, regulāra domu apmaiņa zinātniskos semināros

Atspēriens: 1994.g.rudenī Rīgā organizētais seminārs «Valoda un tehnika Eiropā 2000. Baltijas perspektīva» (turpmākā iesaistīšanās TELRI infrastruktūrā, kas ļāva iepazīties ar valodu tehnoloģijām)

- **1992.–1994.g.** — M. Baltiņas uzraudzībā «Yamahas» datorā sāka ievadīt E. Glika Bībeles grāmatas
- **Kopš 1993 (?)** — LU Filoloģijas fakultātes studentu piesaiste
- **1995.–1996.g.** — seno tekstu datu bāzes izveide (17. gs. iespieddarbi), programmētāja Anita Ozoliņa izveido attiecīgo avotu (Manceļa Sprediķa grāmatas I., II d., 1615.g. katehisms u.c.) vārdformu indeksus
- **1996** — SIA „Tilde“ tika pasūtīti „FrakturaSpecial“ burti
- **1997** — ar Latvijas Zinātnes padomes finansējumu tika turpināta E. Glika Bībeles teksta ievadīšana

2001.g. Liepājā prof. T. Fennels aicina izmantot ne tikai materiālus tradicionālajā grāmatas veidā, bet arī datus elektroniskā formā, kas ilgušus gadus tika uzkrāti LU MII Mākslīgā intelekta laboratorijā (Fennels 2002)

2002.g. Veiksmīga LU MII un LU Filoloģijas fakultātes Baltu valodu katedras sadarbība:

- pilotprojekts – 17.gs. datorvārdnīcas izveide (E.Glika Bībeles materiāls, KKF atbalsts); **P. Vanaga** vadībā izmēģināta metodoloģija vārdnīcas izstrādei
- LU Pētniecības projektu konkurss: *17. gs. latviešu valodas datorvārdnīca.*
- atsaucīgais programmētājs N.Grūzītis
- **2004.–2009. g.** izveidota darba grupa: LU HZF, LU LVI un LU MII (Latvijas Zinātnes padomes atbalsts), projekta vadītājs prof. **P.Vanags**

Valodas korpuss

(strukturēts tekstu vai atšifrētu runas ierakstu kopums, kas paredzēts lingvistiskai analīzei un valodas tehnoloģiju izstrādei. Mūsdienu izpratnē kopš XX gs. 50.g. Ēiropā un ASV):

Vispārīgs

VAI

specializēts

papildināms

VAI slēgts

līdzsvarots

reprezentatīvs

(morfoloģiski, sintaktiski) marķēts VAI nemarķēts («tīrs»)

- tematiski,
- ģeogrāfiski,
- pēc runātāju vecuma,
- pēc laika (diahronisks tekstu korpuss),
- pēc kādas citas pazīmes.

Pirmais brīvi pieejamais latviešu valodas korpuss, kuram tikai izstrādāta arī ērta korpusa platforma ar konkordances programmu, interaktīviem vārdformu indeksiem, meklēšanas iespējām dažādos griezumos (N.Grūzītis)

Veclatviešu rakstītie valodas pieminekļi ir **reprezentatīvi** sava laika valodas atspoguļotāji — rakstu pieminekļi ievietoti **pilnībā**

- Pasaulē ir arī „izlases” korpusa piemēri, piem., ARCHER corpus (A Representative Corpus of Historical English Registers), 1650.–1990. g., sadalīts 7 periodos, katru periodu raksturo desmit ~2000 vārdliet. lielas izlases; kopā 1,7 milj. vārdliet.

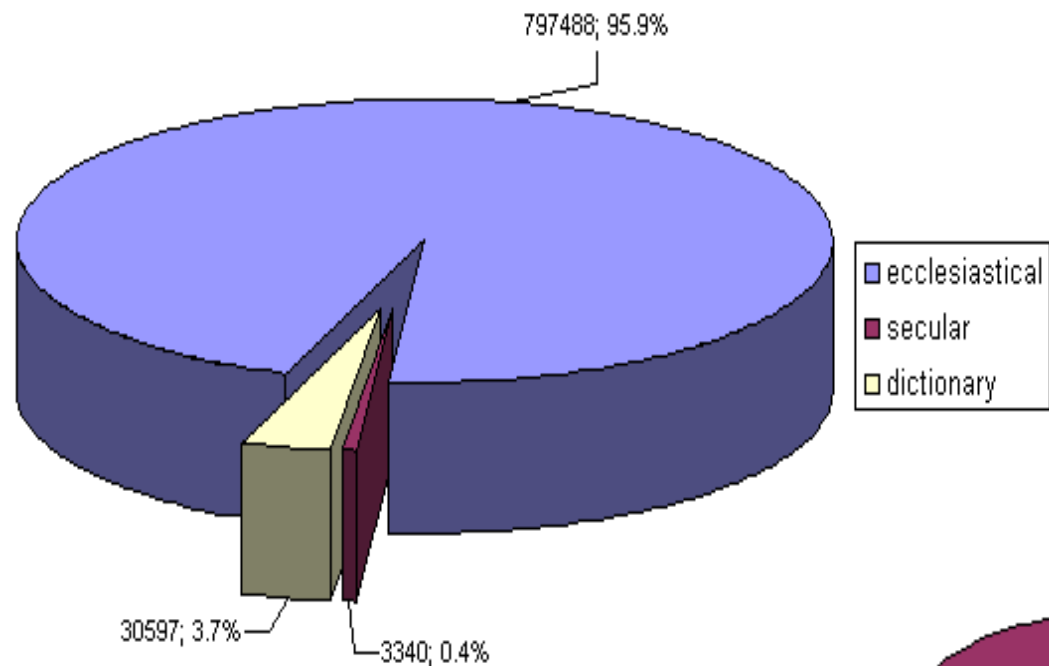
Atšķirībā no mūsdienu latviešu valodas korpusa **daudz** (rakstības, morfoloģijas) **variantu**

- Diahroniskā angļu valodas korpusa daļā (1500.–1600.g.) 60% vārdformu un 35% vārdlietojumu ir vēsturiskie varianti; risinājums — VĀRD (Variant Detector), lai standartizētu rakstības variantus (Baron, Rayson, Archer 2011)

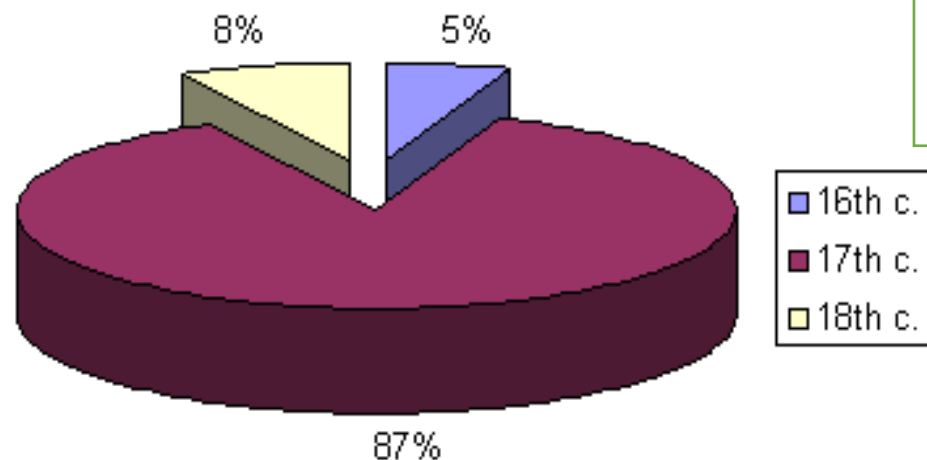
Tekstu atlase, avotu apzināšana – sadarbība ir svarīga!

Tekstu sagatavošana: 90. g.v. – manuāla ievadīšana; 2x pārlasīšana; simbolu vienādošana; no 2002.g. – teksta skenēšana un fraktūras teksta atpazīšana ar ABBYY Fine Reader 6.0; 80-90% precizitāte. Ieskenētie faksimili pieejami internetā

Tekstu apstrāde (strukturālā marķēšana) – būtu bijis labāk uzreiz izmantot kādu standartu



2004 — 16. gs. 3 teksti (no 4)
 2005-2010 — jauni 18. gs. avoti
 2010-2012 — CC1585,
 Dres1682_SBM, 1800.g.



Garīgie teksti – 96 %;
 vārdnīcas – 3,6 %
 un lietišķie teksti – 0,4 %

KKF stipendijas
 LU un Sanktpēterburgas Valsts
 universitātes studentu piesaiste
 jaunu avotu sagatavošanā
 pievienošanai korpusam

Sīkāk: Milčonoka E.
 (2003). Latviešu valodas
 17. gadsimta teksti internetā.
Baltu filoloģija XII (1), Rīga, 2003,
 139–150

Andronova, E. (2007). The Corpus
 of Early Written Latvian: current
 state and future tasks. *Proceedings of
 Corpus Linguistics Conference
 2007*. Birmingham, UK. Pieejams:
[https://www.birmingham.ac.uk/docu-
 ments/college-
 artslaw/corpus/conference-
 archives/2007/245Paper.pdf](https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/245Paper.pdf)

LZP finansēts fundamentālo un lietišķo pētījumu projekts "Korpusā balstīta elektroniska latviešu valodas vēsturiskā vārdnīca" (Nr. 212/2012). Projekta vadītāja **A. Trumpa**.

No 2013. līdz 2016. gadam korpusam tika pievienoti 29 drukātie un rokraksta avoti.

Komandai pievienojas **A. Frīdenberga** (2016.g. aizstāvēja disertāciju «Nominālā vārddarināšana G. Manceļa darbos», kas balstās korpusa datos), **R.Siliņa-Piņķe** (jau kopš 2006.g. raksta īpašvārdu šķirkļus vārdnīcā) un **E.Skrūzmane**

2017. gadā LU Akadēmiskās attīstības projekta "Latviešu valodas seno tekstu korpusa SENIE pāreja uz Unicode" (LU, Nr. AAP2017/63) laikā **73 korpusa avoti** tika sagatavoti *Unicode* formātā.

Neraugoties uz pieradumu lietot korpusā izmantotos simbolus, **pāreja uz Unicode** bija nepieciešama (skat. <http://senie.korpuss.lv/unicode/tabula.pdf>)

19. Myļais Bährns / palleetz labpraht maša` Ghoda` / tas gir labbahk / nhe ka wiß / pehtz ko ta Paßsaule ghada.

	Unicode simbols	Unicode kods	Tabula	Simbols/i seno tekstu korpusā	Piemērs
1.	à	00E0	C1 Controls and Latin-1 Supplement	a`	appakscha` 1 Fuer1650_70_1ms 7.lpp. 23.r.
2.	á	00E1	C1 Controls and Latin-1 Supplement	a´ a´	mihdia`nas VLH1685 13.lpp. 21.r. swa'tas EvTA1753 1.lpp. 7.r.
3.	ã	00E3	C1 Controls and Latin-1 Supplement	a~	exca~ 1 Br1520 PN -- .lpp. 2.r.
4.	â	00E2	C1 Controls and Latin-1 Supplement	a^	a^ 1 Fuer1650_70_1ms 25.lpp. 5.r.
5.	à	0227	Latin Extended-B	a& a´	wara&u& VLH1685_Cat 33.lpp. 32.r. a´ Fuer1650702ms 37.lpp. 8.r.
6.	ä	00E4	C1 Controls and Latin-1 Supplement	a# ä	i\$ra#e&la VLH1685 54{64}.lpp. 17.r. abbâhdena Mancl654 LP2 139.lpp. 14.r.
7.	á	2C65	Latin Extended-C	a, a	La,gs Fuer1650702ms 196.lpp. 2.r. jašchus Fuer1650702ms 197.lpp. 6.r.
8.	b	0180	Latin Extended-B	b/	l/b/. Fuer1650702ms 219.lpp. 7. r.
9.	ć	0107	Latin Extended-A	ć	atnoćie EvTA1753

9. Myļais Bährns / palleetz labpraht mafà Ghodà / tas gir labbahk / nhe ka wiß / pehtz ko ta Paffaule ghada.

Katram ir savas baltās un nebaltās dienas..

- LU MII servera «nobrukšana» īsi pirms pārceļšanās «mākoņos»
- Ilgāku laiku bez korpusa, meklējot kopijas, skenējumus ..
- Atjaunošanā palīdzēja arī lietuviešu kolēģis M.Šimkūns
- **Mācība:** lai cik uzticama un jaudīga ir tehnika, kopijām jābūt vairākām!

SENO izmantošana

- 1) Latviešu valodas vēstures pētījumi gan Latvijā, gan ārpus tās (P.Vanags, A.Trumpa, R.Siliņa-Piņķe, A.Frīdenberga, E.Andronova, E.Kazakēnaite, N.Ostrovskis u.c.)
- 2) Aizstāvēti promocijas darbi, kuru pamatā ir korpusa dati (A.Trumpa (Adjektīvu semantiskā diferenciācija latviešu un lietuviešu valodā (Rīga, 2005)), G.Elksnīte (Nominālās vārdkopas Georga Manceļa tekstos (Liepāja, 2011)), A.Frīdenberga (Nominālā vārddarināšana G. Manceļa darbos (Rīga, 2016) E. Kazakēnaitē („XVI-XVII a. liuteronu latvišķu Biblijos fragmentu sašajos su Glücko Biblijos vertimu (1685–1694)“ (2019)), darbu turpina J.Višņoha (“Verbs E.Glika Bībeles tulkojumā”), varbūt vēl citi (aicinu padalīties, ja izmantojat SENO korpusu)
- 3) Valodas vēstures un korpuslingvistikas kursu programmās (LU, Stokholmas universitātē, Sanktpēterburgas Valsts universitātē, Liepājas Universitātē, Polijā)
- 4) Mērķis – veicināt arī starpdisciplārus pētījumus
- 5) Latviešu valodas vēsturiskā vārdnīca (kopš 2012; 1865 šķirklū) – www.tezaurs.lv/lvvv

SENO veiksmes stāsts

← → ↻ tezaurs.lv/lvvv/

Apps timetable.spbgu https://www.mcafe... 12 Sūtīt SMS [T] Par vārdnīcu Treileris [T] Latviešu valodas vē... Krāsaino zīmāju aus... Vera Singajevska, Tā... » Reading

Latviešu valodas vēsturiskā vārdnīca Par vārdnīcu Šķirkļa uzbūve Meklēšana Avoti Saisinājumi Autori Publikācijas Kontakti

A B C Č D E F G Ģ H I J K Ķ L Ļ M N Ņ O P R S Š T U V Z Ž

Meklēšana [?]

Meklēt visā tekstā

Meklēt

Izvērstā meklēšana


LVVV


Latviešu valodas vēsturiskā vārdnīca (16.–17. gadsimts)
Historical Dictionary of Latvian (16th-17th Centuries)
Historisches Wörterbuch des Lettischen (16.-17. Jahrhundert)

1865 šķirkļi (pēdējo reizi atjaunots 13.01.2022.)
Šķirkļu sagatavošana turpinās

Izdota ar LU Latviešu valodas institūta 2016. gada 5. decembra Zinātniskās padomes lēmumu Nr. 2016/9.
ISBN 978-9984-742-85-4

Autori: Everita Andronova, Anna Frīdenberga, Renāte Siliņa-Piņķe, Anta Trumpa, Pēteris Vanags

 VPP
Valsts pētījumu programma


Latvijas Zinātnes padome

Projekts Nr. 212/2012
„Korpusā balstīta elektroniska latviešu valodas vēsturiskā vārdnīca”

<https://tezaurs.lv/lvvv/?letter=J>

Sīkāk: Andronova E., Siliņa-Piņķe R., Trumpa A., Vanags P. The Electronic Historical Latvian Dictionary Based on the Corpus of Early Written Latvian Texts. *Acta-Baltica Slavica 40. Pogranicze bałtycko-słowiańskie w aspekcie leksykalnym i leksykograficznym*. Warszawa: Instytut Slawistyki Polskiej Akademii Nauk, 2016, 1.–37. lpp. <https://ispan.waw.pl/journals/index.php/abs/article/view/abs.2016.018/2481>

„Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001)

- **SENO** tekstu korpusa modernizācija
 - meklēšanas rīka izstrāde
 - konvertācijas likumu rakstīšana
 - netiks mūsdieniskota morfoloģija
 - **pāreja uz jaunu korpusa platformu NoSketchEngine** (programmētāja L.Pretkalniņa)
 - Pieejams vienuviet ar citiem latviešu valodas vispārīgajiem un specializētajiem korpusiem
 - Jaunu iespēju sniegšana: plašāka konkordances programmas darbība
 - Statistikas datu ērta prezentācija
 - Pielāgošana agrākajām iespējām: vārdformas «adreses» saglabāšana

http://nosketch.korpuss.lv/#dashboard?corpname=senie_unicode

SENO konvertācija mūsdienu rakstībā

Sīkāk sagatavotajās publikācijās:

E.Andronova, A.Frīdenberga, L.Pretkalniņa, R.Siliņa-Piņķe, E.Skrūzmane, A.Trumpa, P. Vanags.
Latviešu valodas senāko rakstu pieminekļu konvertācija mūsdienu rakstībā: iepriekšējā pieredze un automatizācijas mēģinājumi. (Pieņemts publicēšanai «Aktuālas problēmas literatūras un kultūras pētniecībā»).

Andronova E., Frīdenberga A., Pretkalniņa L., Siliņa-Piņķe R., Skrūzmane E., Trumpa A., Vanags P.
Variantums kā konvertācijas izaicinājums: G. Mancēla tekstu transliterācija mūsdienu rakstībā. (Iesniegts recenzēšanai «Letonikā»).

Andronova E., Frīdenberga A., Pretkalniņa L., Siliņa-Piņķe R., Skrūzmane E., Trumpa A., Vanags P.
User-friendly search possibilities for the early Latvian texts: challenges posed by automatic conversion (Pieņemts „Digital Humanities in the Nordic and Baltic Countries 6th Conference”, Upsala, 18.03.2022).

Sagaidāmais rezultāts:

Tas/ kas jau fenn=deeninņam pirmohs laikohs ghaidiets :

Tas/ kas jau sen=dieniņam pirmos laikos gaidīts

2021.-2022.g.: pārēja uz vienotu korpusu platformu *NoSketchEngine* (<http://nosketch.korpuss.lv/#open>)

SENIE

latviešu valodas seno tekstu korpus
korpus früher lettischer texte
the corpus of early written latvian texts

LU Matemātikas un informātikas institūta Mākslīgā intelekta laboratorija
LU Humanitāro zinātņu fakultātes Baltu valodu katedra
LU Latviešu valodas institūts
2002-2018

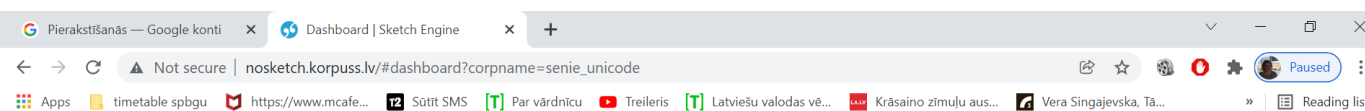


Latvian	LVK2018	9,813,011 words	OPEN
Latvian	Rainis	1,600,249 words	OPEN
Latgalian	MuLa	1,018,962 words	OPEN
Latvian	Tīmeklis2007	99,419,089 words	OPEN
Latvian	Emuāri	6,689,145 words	OPEN
Latvian	Pārspridumi	185,461 words	OPEN
Latvian	LaVA	192,714 words	OPEN
Latvian	Subtitri	5,505,743 words	OPEN
Latvian	LiLa-LV	4,306,984 words	OPEN
Lithuanian	LiLa-LT	4,142,977 words	OPEN
Latvian	SENIE Unicode	2,058,974 words	OPEN
Latvian	UD_Latvian-LVTB	215,547 words	OPEN
Latvian	LRK100 (alpha)	889,343 words	OPEN

Elektroniska bibliotēka? Šobrīd
nekas mainīts vai papildināts
netiek un netiks.

Nākotne rādīs, bet ir vērtīgi
saglabāt faksimilus, indeksus.

NoSketchEngine: http://nosketch.korpuss.lv/



SENIE Unicode senie_unicode

GENERAL INFO

Language: Latvian

COUNTS

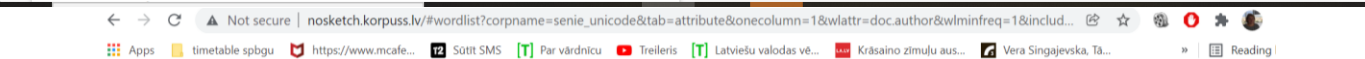
Tokens	2,721,050
★ words	2,058,974
Documents	168

TEXT TYPES

doc (4)	168
★ Autors, doc.author	36
Dokumenta ID, doc.id	168
Gads, doc.year	52
Gadsimts, doc.century	4
page (2)	7,592
para (2)	51,092
language (1)	420,488
block (1)	14,727

LEXICON SIZES

sourceWord	170,614
word	172,308



24	Oto Johans Blūms	321	...
25	Mihaels Ernsts Brīns	266	...
26	Andreass Baumans	200	...
27	Matiāss Stobe	132	...
28	Mihaels Vitenburgs	131	...
29	Karls Ludvigs Tečs	107	...
30	Bruno	73	...
31	Hieronīms Megisers	65	...
32	Andrē Tevē	63	...
33	Johans Hazentēters	58	...
34	Volfgangas Lācijs	54	...

Attribute Value	Tokens	
1 Ernsts Gliks	1,228,897	...
2 Georgs Mancelis	725,863	...
3 nezināms	316,321	...
4 Kristofors Fīrekers	212,497	...
5 Nikolauss Frīdrihs Hespe	99,750	...
6 Martīns Gotlībs Agapetuss Loders	22,064	...
7 Aleksandrs Johans Stenders	21,287	...
8 Georgs Elgers	20,074	...
9 O. F. P. Rīls	18,558	...
10 Gothards Frīdrihs Stenders	13,968	...
11 Georgs Dresels	12,821	...
12 Liborijs Depkins	8,497	...
13 Jānis Reiters	4,243	...



14	Kristians Ravenbergs	3,203	...
15	Karls Frīdrihs Šulcs	2,663	...
16	Nikolauss Kristofers Hāgemeisters	2,644	...
17	Francs Johans Cekels	1,738	...
18	Frīdrihs Vilhelms Cekels	1,623	...
19	Frīdrihs Kazimīrs Urbāns	1,133	...
20	Heinrihs Fūrmans	505	...
21	Mihaēls Vitenburgs	406	...
22	Sila zemnieks Anšs	391	...
23	Johans Kristofs Baumbahs	356	...
24	Oto Johans Blūms	321	...
25	Mihaels Ernsts Brīns	266	...
26	Andreass Baumans	200	...
27	Matiāss Stobe	132	...

SENO korpuss CLARIN krātuvē (<https://www.clarin.eu/resource-families/historical-corpora> - vēl nav pievienots); ir pārstāvēts <https://repository.clarin.lv/repository/xmlui/handle/20.500.12574/12#>

clarin.eu/resource-families/historical-corpora

TABLE OF CONTENTS

Corpus of Old Written Estonian	Estonian	This corpus covers secular and religious texts from the 16th to the 18th century.	Concordancer
Size: 2 million tokens		The corpus is available through a dedicated concordancer.	
Annotation: tokenised, 16-18. century texts have been tagged with contemporary Estonian, morphological and language information. 19. century texts are unannotated.		For the relevant publication, see Kingisepp et al. (2004) .	
Licence: CC-BY			
Classics of Finnish Literature, Kielipankki Version	Finnish	This corpus contains literary texts from 1880 to 1949.	Concordancer
Size: 1.5 million words		The corpus is available through the concordancer Korp.	
Licence: EUPL v.1.1 SA			
Corpus of Old Literary Finnish	Finnish	This corpus contains literary texts from 1543 to 1810.	Concordancer
Size: 4.1 million words		The corpus is available through the concordancer Korp.	
Annotation: MSD-tagged.			

repository.clarin.lv/repository/xmlui/handle/20.500.12574/12#

CLARIN Centre Of Latvian Language Resources And Tools

Authors [Andronova, Everita](#) ; et al.
▼ show everyone
[Andronova, Everita](#) ; [Spektors, Andrejs](#) ; [Vanags, Pēteris](#) ; [Baltiņa, Maija](#) ; [Trumpa, Anta](#) ; [Trumpa, Edmunds](#) ; [Grūzītis, Normunds](#) ; [Siliņa-Piņķe, Renāte](#) ; [Frīdenberga, Anna](#) ; [Skrūzmane, Elga](#) ; [Ķauķīte, Sintija](#)

Item identifier <http://hdl.handle.net/20.500.12574/12>

Project URL <http://senie.korpuss.lv/>

Demo URL <http://senie.korpuss.lv/navigation.jsp?dimension=sources>

Referenced by <http://ceur-ws.org/Vol-2612/short1.pdf>
<https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/245paper.pdf>

Date issued 2002

Type [corpus, text](#)

Size 1500000 tokens

Language(s) [Latvian](#)

Description The Corpus of early written Latvian 'SENIE'(16th-18th cc.) provides access to the texts of written Latvian of the 16th–18th century, and its aim is to facilitate studies of early Latvian in general (e.g. the lexis, morphology and syntax of the texts) and to serve as the basis for the Historical dictionary of the Latvian

Statistics [Login](#)

General Information

- [Deposit](#)
- [Cite](#)
- [Submission Lifecycle](#)
- [FAQ](#)
- [About](#)
- [Help Desk](#)

Starptautiska
aprite

SENIĒ

latviešu valodas seno tekstu korpuss
korpūs früher lettischer Texte
the corpus of early written latvian texts

2021.g.: kamēr pārceļamies uz jaunu vietni...



LATVIEŠU VALODAS SENO (16.–18. GS.) TEKSTU VALODNIECISKO PĒTĪJUMU BIBLIOGRĀFIJA (LĪDZ 2019. GADAM)



„Latviešu valodas seno (16.–18. gs.)
tekstu valodniecisko pētījumu
bibliogrāfija (līdz 2019. gadam)”

<https://dspace.lu.lv/dspace/handle/7/56911>

Starptautiska sadarbība starp augstkolām un bibliotēkām ir ļoti vērtīga (korpuss ir atvērts jauniem avotiem, ko pievienot)

Ekonomiski lietderīgi veidot vienu resursu, kas ir publisks, brīvi pieejams, ar plašām iespējām

Naudai nav īpaši lielas nozīmes, bet ir mierīgāk ap sirdi, ja ir pastāvīgs finansējums iesāktā darba turpināšanai

Ieguvums ir domubiedru komanda

Entuziasms un enerģija ir nedaudz mainīgi lielumi, bet uzkrātā pieredze tikai pieaug 😊



Korpusa izveidē tieši vai netieši roku vai plecu atbalstam pielikuši (saraksts tiek papildināts..)

A.Spektors, M. Baltiņa, A.Ozoliņa, E.Andronova, N. Grūzītis, P.Vanags, A. Trumpa, R.Siliņa-Piņķe, A.Frīdenberga, E.Skrūzmane, S. Kļauķīte, E. Trumpa, L.Leikuma, K.Pokratniece, I.Poikāns, L.Pretkalniņa,

T.Fennels, M.Šimkūns, E.Kazakēnaite, J.Krēsliņš jun., K.Šillere,

LU un Sanktpēterburgas Valsts universitātes studenti, kas uzmanīgi, apņēmīgi un rūpīgi tika galā ar uzticētajiem avotiem, tos pārrakstot un pārļausot.

Paldies par sadarbību **LU MII, LU Latviešu valodas institūtam**, LU Humanitāro zinātņu fakultātei, LNB, LU Akadēmiskajai bibliotēkai, Viļņas Universitātes bibliotēkai, Latvijas Valsts vēstures arhīvam, Tildei.

Par finansiālu atbalstu 20 gadu laikā pateicamies: Sorosa- fondam Latvija, VKKF, LU Zinātiniskās darbības akadēmiskās attīstības projektiem ((LU, Nr. AAP2017/63; Nr. AAP2019/16), IZM, LZP, VPP.

Šobrīd mūsu darbs turpinās projektā un „**Humanitāro zinātņu digitālie resursi: integrācija un attīstība**” (Nr. VPP-IZM-DH-2020/1-0001) Letonika latviskas un eiropiskas sabiedrības attīstībai” projekts „**Latviešu valodas daudzveidība laikā un telpā**” (Nr. VPP-LETONIKA-2021/4-0003).

http://nosketch.korpuss.lv/#dashboard?corpname=senie_unicode

The screenshot shows the CONCORDANCE web application interface. The browser address bar displays the URL: `http://nosketch.korpuss.lv/#concordance?corpname=senie_unicode&tab=basic&keyword=palld*&refs=doc&showresults=1&operation...`. The application header includes the title "CONCORDANCE" and a search box containing "SENIE Unicode". Below the header, a search bar shows the query "simple palld* 18 (6.62 per million)". The interface features a navigation sidebar on the left with various icons, and a main content area displaying search results in a table format. The table has columns for document ID, source text, the search term, and target text. The search term "palldeews" is highlighted in red in the original image. The results are numbered 1 through 13.

Doc ID	Source Text	Search Term	Target Text
1	doc#30 Jck weens tad buhtu paffuddis / Winņs nahk muhš wiŕfo labbo /	Palldeews	/ok faldais JĒfulinņs! Ka tu mums dĒmimis Zillwāhziņņs / Muhs
2	doc#30 ırr nomahkt. 4. To wahrdu buhs teems lıckt ar Meeru / Und nhe	palldeews	dabbuit ./: Tas gir py mums wirf to lauku / Ar fawu Gharru und I
3	doc#30 ts /6. Wahrds Jn nodarriets / Es juhfo deh! Nhelaim beidfis. 14.	Palldeews	ak mieļa Brahļa Sırds! Palldeews ak JĒfu tawa Pirrts / Ka fchki
4	doc#30 rfo deh! Nhelaim beidfis. 14. Palldeews ak mieļa Brahļa Sırds!	Palldeews	ak JĒfu tawa Pirrts / Ka fchkiefta mannu Wainu Es āšmu mehft:
5	doc#30 hkiefta mannu Wainu Es āšmu mehfts Manns Parrahds dfehfts	Palldeews	par tahdu Mainu. 15. Tāhws peeņem mannu Dwehfeliet7. Wah
6	doc#30 as Par wiŕfu labbu patteiz. 13. Ta fazzidams Deews augftakajs /	Palldeews	ak mieļais Radditajs Par tawu Schelafıibu. 14. Zaur Křıtum / ko
7	doc#36 rr nomahkt. 4. To Wahrdu buhs teems lıckt ar Meeru / Vnd nhe	palldeews	dabbuit./:Tas gir py mums wirffon to lauku / Ar fawu Gharru vnd
8	doc#38 =fpeh / tad wings ohtram to Naudu išwiļļ. 9. Vnd fkenneh to par	Palldees	. 10. Dalfch nhe labpraht aisdohd / nhe no ļaunu Prahtu / bett tē
9	doc#41 s/Pullx leelo (Lohpo. Er hat GOtt lob ein gut theil Viehe/winjam	palldeeß	Deewam (ghohtz Thāwam) labba teeffa Lohpo. Er hat kein Glü
10	doc#42 ıex / dohd Brandawıenu par trieß Ghroffchi. Buhß Schkehrft! S.	Palldeeß	. D. gir Ratti ghattawi S. ghattawi gir / tickai ja fwaida. D. gir lab
11	doc#42 Stall / da man die Pferde warten könne. G. Labbs wackars. A.	Palldeeß	. G. Ka klajahß? A. Tadaı / pußlıeds. G. Eß tōw luhdfōhß/ dohd
12	doc#42 ı to Meeffas= ghabbalu / ja nhe buhß ghann / tad praŕfi wehl. J.	Palldeeß	/ deeßghann. G. Maıfe gir tawā Rohkā. J. ghann dabbuıffim. G.
13	doc#42 aimneex / jemm to Meeffas ghabbalu/ dohd Seewai arrıdfan. A.	Palldeeß	Kungs. G. fche Maıfe arrıdfan. Mums nhe gir dauds Maıfes. A. k