

Nacionālā korpusu kolekcija

Baiba Saulīte
Ilze Auziņa
Roberts Dargis

CLARIN praktiskais seminārs
2022. gada 8. decembris

AiLab.lv

LU MII Mākslīgā intelekta laboratorijas galvenie pētījumu **virzieni**:

Dabiskās valodas apstrāde (**NLP**)

- runas **atpazīšana** un **sintēze**
- sintaktiskā un semantiskā **analīze**
- tekstrade

Mašīnmācīšanās

Valodas resursu izveide

- teksta un runas **korpusi**
- mašīnlasāmas **vārdnīcas**



Korpus.lv

Nacionālā korpusu kolekcija

Meklēšana

Materiāli

Par Korpus.lv

sirds*

Meklēt

Korpus	Relatīvais biežums uz 1 miljonu	Absolūtais biežums
Pārspridumi Skolēnu pārspridumu korpus (izstrādātāji: LU MI, LU MII, RTA Mezgli: nosketch.korpus.lv)	2257	510
LatSenRom Latviešu senākie romāni (izstrādātāji: LNB Mezgli: nosketch.inb.lv)	1420	8234
Rainis Raiņa darbu korpus (izstrādātāji: LU MI Mezgli: nosketch.korpus.lv)	1196	2752
IsprozaS Sieviešu isproza (izstrādātāji: LU LMI Mezgli: nosketch.inb.lv)	1083	1272
Senie Latviešu valodas seno tekstu korpus (izstrādātāji: LU LAVI, LU MI, LU HZF Mezgli: nosketch.korpus.lv)	765	2081
LVMED Radioloģisko izmeklējumu transkripciju korpus (izstrādātāji: LU MI, RTA Mezgli: nosketch.korpus.lv)	616	97
MuLa2012 Mācību latgaliešu tekstu korpus 2012 (izstrādātāji: LU MI, RTA Mezgli: nosketch.korpus.lv)	426	565
PanDi Pandēmijas dienasgrāmatas (izstrādātāji: LU LMI Mezgli: nosketch.inb.lv)	408	289
MuLa2022 Mācību latgaliešu tekstu korpus 2022 (izstrādātāji: RTA, LU MI Mezgli: nosketch.korpus.lv)	394	1101
Timekls2020 Latviešu valodas timekls korpus (izstrādātāji: LU MI Mezgli: nosketch.korpus.lv)	334	164 504
LiLa Lietuvišu-latviešu-lietuvišu paralēlo tekstu korpus (izstrādātāji: LU MI, YDU Mezgli: nosketch.korpus.lv)	305	1734
LVK2018 Līdzsvarotais mācību latviešu valodas tekstu korpus (izstrādātāji: LU MI Mezgli: nosketch.korpus.lv)	270	3321
Emuāri Latviešu valodas emuāru korpus (izstrādātāji: LU MI Mezgli: nosketch.korpus.lv)	256	2128



Izstrādātājs: LU LMI
Mezgli: nosketch.inb.lv

Izstrādātāji: LU LAVI, LU MII, LU HZF
Mezgli: nosketch.korpus.lv

Izstrādātāji: RTA, LU MII
Mezgli: nosketch.korpus.lv

CONCORDANCE Sieviešu isproza 1893-2002

simple **sirds*** 1,272 (1 083.32 per million)

Details Left context KWIC Right context

11	docF0 atļu tā caur savu augspārību un aku mīlestību tāpat samānā – par vīps	sirds	nemaz, bet tikai par ārēju izskatu rūpēdams. Kad Laila ko iedomājusies, t
12	docF0 ni vien spriežot laiku pavadijāt, un tu man nemaz tālāk nestāstī, kas tavs	sirdsdrugs	un kādā vietā tas.” Viņš tagad Z. viesnīcā par sulaini.” Anna stāstīja, “bet v
13	docF0 “Ja, kāds valodā vārds tiek nosaukts.” “Iev gan taisnība, ka vārds cilvēka	sirds	un dabu nepārvērš, un pa lielāku daļu tas jau no tēva mantots – jo tālāk, ka v
14	docF0 “bas vārdu bū at viltīgi – tik vien lūdz nedotāt, ka tas zīmētos uz tavu	sirdsdrugu	– Tu jau gan, viņa sirdi un prātu pazīdama un no viņa krietnības pārliecināta,
15	docF0 ik vien lūdz nedotāt, ka tas zīmētos uz tavu sirdsdrugu. Tu jau gan, viņa	sirds	un prātu pazīdama un no viņa krietnības pārliecināta, man par jaunu neper
16	docF0 “ar gējiem.” “Ne, ne, mīļā Trīstī!” sacīja Anna. Trītes roku sāpējušas un pie	sirds	spiezdamā. “Nevāru par to neko vairāk sacīt, – es tikai zinu, ka viņš ir ska
17	docF0 – es tikai zinu, ka viņš ir skaists, laipns un patīkams, un ka es viņu no visas	sirds	mīlu. Nekad neesmu ko dzīslē apdomājis, kā tikai, cik jauki un iatimīgi bij i
18	docF0 i, kura veda uz ārpusi: īs tas sev priekšā ieradzīja kādu sievietes tēlu, viņa	sirds	sāka ātrāk pūkstēt – vai tā varbūt nebija viņa? Šis skaists stāvs tam jau izi
19	docF0 Jūs abi bijāt saruna, kuru neviens dzirdējis, viņa pilnais vairāk, kurš lika manai	sirdi	iet pāri. Es jās neapbēdīju apvainot, jo jūsu persona man sveša un vienaldzī
20	docF0 i, – netaisnīti mani, netaisnīti aizrautos no acīm, netaisnīti manā	sirdi	jūtas, kuras es vairs nespeju apslāpēt. Es esmu dzīvā tik daudz cietusi, tik c

Rows per page: 10 11–20 of 1,272 1 2 >

CONCORDANCE SENIE: 16.–18.gs. teksti

simple **sirds*** 2,081 (24.58 per million)

Details Left context KWIC Right context

31	docF17 : thas padūde preclhan them. Vnde Batoy Itapan 80w / rne dedicez mulla	Syrdicz	enkan mums / Khad has ar mums trellay wūrlon tho Ozelle / Khad has mun
32	docF17 : aridiccan nokunfcherne / Beth es gib yums atkal redcezh / vnde yuile	Syrdicz	Bōw lyxomels / vnd yuile Lyximbē bus neewenan no yums yempt. Vnde py
33	docF17 : / Kur tu no eile ? Beth pcelz tho / ka es to vs yums trellays elme / gir yuile	Syrdicz	puehna nokunfcherne tappuffi. Beth es flacke yums tho telcham. Thas yums
34	docF17 : enne Mhere dome es yums / rne ka tha Pallole dode / dome es yums. Juule	Syrdicz	rhe Ittruccas / vnde rhe ylyftas Bōw. Jus effeth czerdiele / ka es yums
35	docF17 : tho loune / pekaretes tam labbam. Ta Bralige myleflōe Itapan wen otre gyt	Syrdicz	duebeniga. Wens nake them otram ar Godenafcherne preclchan. Nehe effe
36	docF17 : nu abhāngig wipsh stahwu alkahru. Er stehet abhāngig. Tu ne dshwo wis no	Syrdicz	tu alkahru ween dshwo tho arbellest nicht vom herten, du ar arbellest hojye,
37	docF17 : Morgenland. Al-naust,licht werden.Metagh.Sich erholen. Vid. ante fol. sub F.	Syrdicz	atausdas hertz erholst sich Arbellest Spitze von einer Kon Arhesten Akkolpys
38	docF17 : lem Ewigem leben Trachten ringen. Duldbetötet, abgemattet seyn. Badakuffi	Syrdicz	„Ein betötet, untraurigenCur. hertz. Dutzamais, der hinde, hinterste Duhmp
39	docF17 : rasche, die im Kon schreyet Grestsauff und überlegen,aggrest,Umkehrten,	Syrdicz	aggresthache GrestSchneiden,Umkehrten, werden,Tat Nasie agrestschach, u
40	docF19 : peem tinnah, er gehet mit Vah um. Kluckst,wie eine bruthenne glucken	Syrdicz	ta Kluckst. Krust!-hagel Kruet'sene, abolu dshrenesApfel Trank Zussenel

Rows per page: 10 31–40 of 2,081 1 4 >

CONCORDANCE Mācību latgaliešu tekstu korpus 2022

simple **sirds*** 1,181 (333.52 per million)

Details Left context KWIC Right context

21	docF44 : es Kuorsovā, sevi skaitu arī kai varakūoni un cylviku nu Aglyums. Muna	sirds	Latgolē i Varakūoni, Aglynu n dzymtja Kuorsovā. Sovim bārnim kult
22	docF50 : tpa i tai ty vyes ar gošā sauguja. Sīrneiba, vīnukušība, galdēšana, no	sirds	nastys duovony – Latgolē otkon idēvu muzikim jaunus brīnumus, par k
23	docF50 : dzimties kļūdīs. Romans tei raksturots kai iakots na tik Latgolys jau	sirds	i divselis dzēvā, nu arī Latviešu viesuīras krystpunkts – skateļojamys b
24	docF56 : jmar patek bārnim ruodeit, kai lei kasdina var nūstāt šruodotji. Man sup	sirds	ka mes drupel par vāu asam saprotuši sovokūduys školys vajadzību, k
25	docF57 : kō tustestē, 90 gōdis populari beja balderani, nūle nīvīnam narvaga,	sirdsmāteri	padavē. Dabys augus polē padarajom popularus, primāram, idantu – Kon
26	docF70 : sātu, a sastajam narvagei nūstacēt, kas vāinis. Sīrnī alkohots, dzēvotus	sirdsmāteriba	a pats nī per kō nēasayduj. Es pasacuju, ka jī vajag nīvāduot. Vīlu nar, i
27	docF70 : it kō cytu. Interesantuk rakstēt i skaltēt, alkūotji nazymnos faktus. Nu	sirds	vīleju Osyuna pamatškolys nūvōdīntīrēbyys ustōbāt dabūt Jarvīny atir
28	docF70 : busu atbrauksi, ēpabē bālnīvā, sacējūdi: „Es jī jums atjuoju.” Jā! sup	sirds	ka Osyuna vāina nazymta, ka parēnais, esīst latgaliskais fluos vīlys nūsa
29	docF75 : rede jās uzjūzmyus. Par jī to jēlu naidu sajāme. Pats Dymā nūmyra ar	sirdstrikū	Ar sumeti izgojys pasasagot, uzskatī kai suni. Cikoms jīs sovū kuceti g
30	docF75 : re atgūojys de bāberis atsavadeit. A kunga šymā teimurējā nūmyra ar	sirdstrikū	Naspijys puordzēvot, ka juopamat Ruljuna.” Ruljūny luoca Antoneja.

Rows per page: 10 21–30 of 1,181 1 3 >

20+ korpusi, 10+ izstrādātāji, 3+ mezgli



LVK2018

Līdzsvarotais mūsdienu latviešu valodas tekstu korpus
2016–2018, 10 milj. vārdlietojumu (12 milj. tekstvienību)
Izstrādātāji: LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

MuLa2022

Mūsdienu latgaliešu tekstu korpus 2022
2020–2022, 2 milj. vārdlietojumu (2,8 milj. tekstvienību)
Izstrādātāji: RTA, LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

LaVA

Latviešu valodas apguvēju korpus
2018–2021, 192 000 vārdlietojumu (241 000 tekstvienību)
Izstrādātāji: LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

Barometrs

Interneta agresivitātes indekss
2011–2022, 26 milj. komentāru (642 milj. tekstvienību)
Izstrādātāji: RSU, LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

Emuāri

Latviešu valodas emuāru korpus
2014–2015, 6,6 milj. vārdlietojumu (8 milj. tekstvienību)
Izstrādātāji: LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

Hugo.lv

Hugo.lv paralēle korpusi
2018, 10,5 milj. vārdlietojumu
Izstrādātāji: KISC

[Vairāk informācijas](#)

[Meklēt korpusā](#)

LaVA

[Meklēt korpusā](#)

[Vārdu biežuma saraksts](#)

Latviešu valodas apguvēju korpus

Korpusā iekļauti to Latvijas augstākajās mācību iestādes studējošo ārvalstnieku darbi, kuri latviešu valodu apguš kā svešvalodu pirmo vai otro semestri. Teksti ir automatiski morfoloģiski marķēti, un tajos ir manuāli marķētas valodas apguvēju kļūdas.

Publikācija, uz kuru atsaukties:

R. Dargis, I. Auzina, K. Levane-Petrova, I. Kaija

Quality Focused Approach to a Learner Corpus Development

2020

[PDF](#)

[teksta \(26\)](#) [apguvēju \(2\)](#) [morfoloģija \(23\)](#) [kļūdas \(2\)](#) [manuāli pārbaudīts \(4\)](#)

Korpora apjoms	192 000 vārdlietojumu (241 000 tekstvienību)
Izstrādes periods	2018–2021
Izstrādātāji	Latvijas Universitātes Matemātikas un informātikas institūts
Finansējuma avots	LZP Fundamentālo un lietišķo pētījumu programa (Izp–2018/1-0527)
Mājaslapa	http://lava.korpus.lv/lv/
CLARIN	http://hdl.handle.net/20.500.12574/49
Citas publikācijas	I. Kaija and I. Auzina Data collection for learner corpus of Latvian: copyright and personal data protection <i>Selected papers from the CLARIN Annual Conference 2019, 41–47, 2020</i> PDF DOI I. Auzina, I. Kaija, K. Levane-Petrova Mērķhipotēžu izvirzīšana latviešu valodas apguvēju korpusā <i>Valoda: nozīme un forma, 11, 7–26, 2020</i> PDF DOI

[Vairāk informācijas](#)

[Meklēt korpusā](#)

Korpusu iedalījums

vispārīgs (6) apguvēju (2) literārs (4) parlamentārs (1) diahronisks (1) specializēts (13) reprezentatīvs (7) paralēls (1)
teksta (25) runas (5) morfoloģija (23) sintakse (3) semantika (1) kļūdas (2) manuāli pārbaudīts (4) tīmekļa (2)

Pēc valodas **realizācijas** formas

Pēc datu **atlases** principiem

Pēc datu **raksturojuma**

vispārīgs vs specializēts

specializēts: apguvēju, specializēts: literārs;

specializēts: parlamentārs; specializēts: diahronisks

Pēc (lingvistiskā) **marķējuma**

Korpusi ar birku **literārs (4)**

LatSenRom
Latviešu senākie romāni
2019–2022, 4,6 milj. vārdlietojumu (5,8 milj. tekstvienību)
Izstrādātāji: LNB

[Vairāk informācijas](#)

PanDi
Pandēmijas dienasgrāmatas
2020–2022, 565 000 vārdlietojumu (709 000 tekstvienību)
Izstrādātāji: LU LFMI

[Vairāk informācijas](#)

Rainis
Raina darbu korpus
2018, 1,6 milj. vārdlietojumu (2,3 milj. tekstvienību)
Izstrādātāji: LU MII

[Vairāk informācijas](#)

ĪsprozaS
Sieviešu īsproza
2020–2022, 925 000 vārdlietojumu (1,2 milj. tekstvienību)
Izstrādātāji: LU LFMI

[Vairāk informācijas](#)

Likumi

Latvijas Republikas tiesību aktu korpus

Korpusā ievietoti visi tiesību aktu vietnē likumi.lv publicētie Latvijas Republikas tiesību akti (līdz 2022. gada februārim).

teksta (25) **specializēts (13)** morfoloģija (23)

Korpusa apjoms	73,9 milj. vārdlietojumu (116,2 milj. tekstvienību)
Izstrādes periods	2022
Izstrādātāji	Latvijas Universitātes Matemātikas un informātikas institūts
Finansējuma avots	VPP "Humanitāro zinātņu digitālie resursi" (VPP-IZM-DH-2020/1-0001)

NLP-PIPE: Latvian NLP Pipeline as a Service

Rīkkopu izmanto: Korpus.lv,
Latvijas Nacionālā bibliotēka, ziņu
aģentūra LETA, studenti u. c.

Select a sample text

2022. gada 6.-7. oktobrī Rīgā norisinās Baltijas Valodu tehnoloģiju konference, ko rīko LU, LU MII un Tilde, pastāstīja programmkomitejas vadītāja Inguna Skadiņa.

Go

tokenizer x

ner x

morpho x

parser x

☒ Visual ☐ CoNLL ☐ JSON

2022. gada 6.- 7. oktobrī time Rīgā GPE norisinās Baltijas Valodu tehnoloģiju konference event , ko rīko LU organization , LU MII organization un Tilde organization , pastāstīja programmkomitejas vadītāja Inguna Skadiņa person .

INDEX	FORM	LEMMA	UPOSTAG	XPOSTAG	FEATS	HEAD	DEPREL
#text=2022. gada 6.- 7. oktobrī Rīgā norisinās Baltijas Valodu tehnoloģiju konference , ko rīko LU , LU MII un Tilde , pastāstīja programmkomitejas vadītāja Inguna Skadiņa .							
1	2022.	2022.	ADJ	xo	NumType=Ord	2	amod
2	gada	gads	NOUN	ncmsg1	Case=GenlGei	3	nmod
3	6.-	6.-	NUM	xn	NumType=Car	5	amod
4	7.	7.	ADJ	xo	NumType=Ord	5	amod
5	oktobrī	oktobris	NOUN	ncmsl2	Case=LoclGer	7	obl
6	Rīgā	Rīga	PROPN	npfsl4	Case=LoclGer	7	obl
7	norisinās	norisināties	VERB	vmyp_330an	Evident=FhlMc	22	ccomp
8	Baltijas	Baltija	PROPN	npfsg4	Case=GenlGei	9	nmod
9	Valodu	valoda	NOUN	ncfpg4	Case=GenlGei	10	nmod
10	tehnoloģiju	tehnoloģija	NOUN	ncfpg4	Case=GenlGei	11	nmod
11	konference	konference	NOUN	ncfsn5	Case=NomiGe	7	nsubj
12	,	,	PUNCT	zc	—	14	punct
13	ko	kas	PRON	pr000an	Case=AcclPro	14	obj
14	rīko	rīkot	VERB	vmnpt230an	Evident=FhlMc	11	acl
15	LU	LU	PROPN	yp	Abbr=Yes	14	nsubj
16	,	,	PUNCT	zc	—	17	punct
17	LU	LU	PROPN	yp	Abbr=Yes	15	conj
18	MIl	mii	X	y	Abbr=Yes	17	iobj
19	un	un	CCONJ	cc	—	20	cc
20	Tilde	tilde	NOUN	ncfsn5	Case=NomiGe	18	conj
21	,	,	PUNCT	zc	—	7	punct
22	pastāstīja	pastāstīt	VERB	vmnist330an	Evident=FhlMc	0	root
23	programmkomitejas	programmkomiteja	NOUN	ncfsg4	Case=GenlGei	24	nmod
24	vadītāja	vadītāja	NOUN	ncfsn4	Case=NomiGe	26	nmod
25	Inguna	Inguns	PROPN	npmsg1	Case=GenlGei	26	nmod
26	Skadiņa	Skadiņa	NOUN	ncfsn4	Case=NomiGe	22	nsubj
27	.	.	PUNCT	zs	—	22	punct

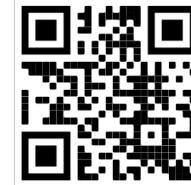
<http://nlp.ailab.lv>

Gramatiskais marķējums

deputātu kandidātiem pietiekami labi	jāzina valsts valoda	jāzina	zināt	VERB	vmnd0t300an
ncmpg1/deputāts ncmpd1/kandidāts rp_/pietiekami rp_/labi	vmnd0t300an/zināt ncfs6/valsts ncfsn4/valoda	valsts	valsts	NOUN	ncfsg6
Drošības dienesta darbiniekam brīvi	jāpārvalda valsts valoda	valoda	valoda	NOUN	ncfsn4
ncfsg4/drošība ncmsg1/dienests ncmsd1/darbinieks rp_/brīvi	vmnd0t300an/pārvaldīt ncfs6/valsts ncfsn4/valoda				
īvo aktu prasībām jāprot un	jālieto valsts valoda	valsts	valodas prasmes	apliecību	apliecinātas
īvts ncmpg1/akts ncfpd4/prasība vmnd0_100an/prast cc/un	vmnd0t200an/lietot ncfs6/valsts ncfsn4/valoda	valsts	valodas	apliecību	apliecinātas
tu absolventi , kurās mācībvaloda	ne bija valsts valoda	valsts	zināja	nepietiekami .	</s> <s> Bija
de ncmpn1/absolvents zc/, pr0p/n/kura ncfsn4/mācībvaloda	vmnis130an/būt ncfs6/valsts ncfsn4/valoda	valsts	zināja	nepietiekami .	</s> <s> Bija
Prasību , ka pašvaldību darbā	jālieto valsts valoda	valsts	varot	izpildīt	ar citiem , efektīvākiem
ncfpg4/prasība zc/, cs/ka ncfpd4/pašvaldība ncms1/darbs	vmnd0t200an/lietot ncfs6/valsts ncfsn4/valoda	valsts	varot	izpildīt	ar citiem , efektīvākiem
, ka pašvaldības domes deputātiem	jāprot valsts valoda	valsts	neesot	diskriminējoša	un atbilstot
zc/, cs/ka ncfs4/pašvaldība ncfs5/dome ncmsd1/deputāts	vmnd0_100an/prast ncfs6/valsts ncfsn4/valoda	valsts	neesot	diskriminējoša	un atbilstot
m ēkas numura zīmē esot	lietojama valsts valoda	valsts	neesot	diskriminējoša	un atbilstot
ncfsg4/ēka ncmsg1/numurs ncfs15/zīme vmnrp_i00an/būt	vmnrp130an/lietot ncfs6/valsts ncfsn4/valoda	valsts	neesot	diskriminējoša	un atbilstot
īvo aktu prasībām jāprot un	jālieto valsts valoda	valsts	neesot	diskriminējoša	un atbilstot
īvts ncmpg1/akts ncfpd4/prasība vmnd0_100an/prast cc/un	vmnd0t200an/lietot ncfs6/valsts ncfsn4/valoda	valsts	neesot	diskriminējoša	un atbilstot
ī sarakstē ar adresātiem iekšzēmē	jālieto valsts valoda	valsts	neesot	diskriminējoša	un atbilstot
ī ncfs15/sarakste spdp/ar ncmpd1/adresāts ncfs15/iekšzeme	vmnd0t200an/lietot ncfs6/valsts ncfsn4/valoda	valsts	neesot	diskriminējoša	un atbilstot
atestācijas komisiju locekļiem obligāti	jāprot valsts valoda	valsts	neesot	diskriminējoša	un atbilstot
ncfsg4/atestācija ncfpd4/komisija ncmpd2/loceklis rp_/obligāti	vmnd0_100an/prast ncfs6/valsts ncfsn4/valoda	valsts	neesot	diskriminējoša	un atbilstot
ieņemamajam amatam un profesijai	jāprot valsts valoda	valsts	neesot	diskriminējoša	un atbilstot
ncmpdmsdppypn/ieņemts ncmsd1/amats cc/un ncfsd4/profesija	vmnd0_100an/prast ncfs6/valsts ncfsn4/valoda	valsts	neesot	diskriminējoša	un atbilstot
74. Domes (padomes) sēdes	notiek valsts valoda	valsts	neesot	diskriminējoša	un atbilstot
xo/74. ncfs5/dome zb/ ncfs5/padome zb/ ncfn5/sēde	vmnpi130an/notikt ncfs6/valsts ncfsn4/valoda	valsts	neesot	diskriminējoša	un atbilstot

CQL vaicājums: [tag="vm.*"] [word="valsts"] [word="valoda"]

Meklēšana 24 korpusos vienlaikus



Nacionālā
korpusu
kolekcija

Meklēšana Materiāli Par NKK

LV ▼



Meklēt

[Meklēšana](#)

Projekti



Nr. VPP-IZM-DH-2020/1-0001
Nr. VPP-Letonika-2021/1-0006



Nr. Izp-2019/1-0464 (2020–2022)



Nr. VPP-IZM-2018/2-0002 (2018–2021)

NACIONĀLAIS
ATTĪSTĪBAS
PLĀNS 2020



EIROPAS SAVIENĪBA
Eiropas Reģionālās
attīstības fonds

IEGULDĪJUMS TAVĀ NĀKOTNĒ



Latviešu valodas aģentūra



Publikācijas

Saulīte, B., Darģis, R., Grūzītis, N., Auziņa, I., Levāne-Petrova, K., Pretkalniņa, L., Rituma, L., Paikens, P., Znotiņš, A., Strankale, L., Pokratniece, K., Poikāns, I., Bārzdīņš, G., Skadiņa, I., Baklāne, A., Saulespurēns, A. & Ziediņš, J. (2022). [Latvian National Corpora Collection – Korpuss.lv](#). *Proceedings of the Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 5123–5129.

Skadiņa, I., Saulīte, B., Auziņa, I., Grūzītis, N., Vasiljevs, A., Skadiņš, R., Pinnis, M. (2022a). [Latvian Language in the Digital Age: The Main Achievements in the Last Decade](#). *Baltic Journal of Modern Computing*, 10(3), 490–503.

Skadiņa, I., Auziņa, I., Darģis, R., Voitkāns, A. (2022b). [CLARIN valodas resursu un rīku pētniecības infrastruktūra humanitārajām un sociālajām zinātnēm](#). *Letonica*, 312–327.

Darģis, R., Levāne-Petrova, K. & Poikāns, I. (2020). [Lessons Learned from Creating a Balanced Corpus from Online Data](#). *Human Language Technologies – The Baltic Perspective*, IOS Press, 127–124.

Znotiņš, A. & Bārzdīņš, G. (2020). [LVBERT: Transformer-Based Model for Latvian Language Understanding](#). *Human Language Technologies – The Baltic Perspective*, IOS Press, 111–115.

Grūzītis, N. & Znotiņš, A. (2018) [Multilayer Corpus and Toolchain for Full-Stack NLU in Latvian](#). *Proceedings of the CLARIN Annual Conference*, 2018, 61–65.

Saulīte, B., Darģis, R., Auziņa, I. Nacionālā korpusu kolekcija *Korpuss.lv*. Pieņemts publicēšanai rakstu krājumā *Linguistica Lettica*.



Ieskaties: <http://ailab.lv>



Seko mums: @AiLab.lv