

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

Mūsdienu latgaliešu tekstu korpus: paveiktais un nākotnes ieceres

PROF.SANITA MARTEŅA, DR.PHIOL.ANTRA KĻAVINSKA
(RĒZEKNES TEHNOLOGIJU AKADĒMIJA)

LIEPĀJAS 27. STARPTAUTISKĀ ZINĀTNISKĀ KONFERENCE
«AKTUĀLAS PROBLĒMAS LITERATŪRAS UN KULTŪRAS PĒTNIECĪBĀ»
10.-11.03.2021.

1. Kas ir mūsdienu latgaliešu tekstu korpuss?
2. Kā tas līdz šim ir izmantots (provizoriskie dati)?
3. Kāds būs mūsdienu latgaliešu tekstu korpuss?
4. Kāda ir mūsu vīzija par tā funkcionalitāti nākotnē?

PRIEKŠLASĪJUMA JAUTĀJUMI

HipLatLit

Lietuviešu
Latgaliešu
Angļu

SĀKUMS PRODUKTI AKTIVITĀTES GALERIJA

Sadarbojoties [Rēzeknes Augstskolai](#), [Vitauta Dižā universitātei](#) un [Latvijas Universitātes Matemātikas un informātikas institūtam](#), ar Latvijas–Lietuvas pārrobežu sadarbības programmas 2007.-2013.gadam atbalstu tika realizēts projekts „HipLatLit” („Humanitārās izglītības pētniecības infrastruktūras izveide Austrumlatvijā, Lietuvā”).

Projekta „HipLatLit” rezultāti:

1. [Mūsdieni latgaliešu tekstu korpuiss](#) (1 miljons vārdlietojumu);
2. [Lietuviešu-latviešu-lietuviešu paralēlo tekstu korpuiss](#) (8 miljoni vārdlietojumu);
3. [Lietuviešu-latviešu-latgaliešu vārdnīca](#) (10 tūkstoši šķirkļu);
4. [Starpdisciplināra doktora studiju programma baltistikā.](#)
5. Datorlingvistikas modulis filoloģijas maģistrantiem Rēzeknes Augstskolā.

- specializēts korpuiss – Latvijā laika posmā no Atmodas (1988) līdz 2012. gadam latgaliešu rakstu valodā publicēti teksti;
- līdzsvarots – literārie, tehniski speciālie (zinātniskie, populārzinātniskie, lietišķie teksti u. c.) un informatīvie teksti (publicistika, ziņas u. c.).
- tekstiem ir pievienoti bibliogrāfiskie metadati;
- avotu veids – drukāti (grāmatas, periodika u. c.) un elektroniski teksti (mājaslapas, blogi).

user: defaults corpus: MuLa

Search volūd* in MuLa[Concordance](#)
[Word List](#)
[Corpus Info](#)[Save](#)
[View options](#)
[KWIC](#)
[Sentence](#)[Sort](#)
[Left](#)
[Right](#)
[Node](#)
[References](#)
[Shuffle](#)[Sample](#)
[Filter](#)
[Overlaps](#)[1st hit in doc](#)[Frequency](#)
[Node forms](#)[Collocations](#)
[ConcDesc](#)
[Visualize](#)Query **volūd.*** 3,605 > GDEX 3,605 (2,730.3 per million)Page 1 of 19 [Go](#) [Next](#) [Last](#)

1 diena.lv	dzeivi i sovys zemis viesturi navarātu runuot sovā	volūdā	. Dūmuot , bez sovu sakņu i ar nasauordeitu identitati
2 diena.lv	stypri i kai muoti , i mōti . Ir labi , ka es jei sovā	volūdā	varu pīraksteit , cik ciši es juos ļubeju . Bet vyspuor
3 bezpvn.lv	Vysys dzīsmis skaidrā , saprūtamā latgalīšu volūdā -	volūdā	, kurā grupa dzid i dūmoj . Albumā var atrast ari
4 bezpvn.lv	laikā . Vysys dzīsmis skaidrā , saprūtamā latgalīšu	volūdā	- volūdā , kurā grupa dzid i dūmoj . Albumā var atrast
5 jureits.wordpress.com	paruoda , ka Latgolys latvišu volūda jeb latgalīšu	volūda	ir sanuoka par pošu valsti . Man juoplikreit dokumentā
6 diena.lv	asu suokuse runuot latgaliski , lai jis nacer , ka	volūdu	byus ispiejams maineit . Tys ir par daudz sarežgeitai
7 raibils.wordpress.com	tū pusi) , cytu fakultati , kur juodazaregistrēj	volūdys	kursim . Tai kai breiws - veromīs kartī , ka juodaīt
8 geoviita.wordpress.com	latgaliski ! Kuri ir nūdrūšynuojuši tū , ka latgalīšu	volūda	ari niule ir dzetva . Man ir lela prīca par vysim
9 jureits.wordpress.com	izaveidoit . Tys faktiski ari paruoda , ka Latgolys latvišu	volūda	jeb latgalīšu volūda ir sanuoka par pošu valsti .
10 diena.lv	tikai slapanai , ar sovejim - kai taidā konspirativajā	volūdā	. Koč Latgola jai ir svātuo zeme . Kur zeme ar dabasim
11 diena.lv	krivim es ni nu šuo , ni nu tuo nasaokšu runuot cytā	volūdā	. Man tys ir par daudz suopeigai . Saīt taidīs psihologisks
12 diena.lv	Rēzeknē nūtiks informatīvs seminars par latgalīšu rokstu	volūdu	i juos nuusacejumim . Vēl es varātu izsturosteit ,
13 geoviita.wordpress.com	Puordeviejs ir jauns puiss , bet narunoj ni angļu , ni krivu	volūdā	. Labi , ka ir rūkys . Nūpierkam gurkus i tynamis
14 jureits.wordpress.com	tymā nav ni vuorda par Latgolys latvišu jeb latgalīšu	volūdu	. Gon latgalīšim , gon Saeimys deputatim man juoatguodynoj

doc.Number	130
doc.Author	saprge (Ilze Sperga)
doc.TextTitle	Viltuota dzeive i zogti uobeleisi
doc.Source	diena.lv
doc.PublicationOfDate	14-Okt-2008
doc.Publisher	SIA Dienas Mediji
doc.Place	http://www.diena.lv/blogi/viltuota-dzeive-i-zogti-uobeleisi-631711
doc.TextType	elektroniskie

http://hipilatlit.ru.lv/bonito/run.cgi/first_form

Word (47 items | 3,343 total frequency)



Word	↓ Frequency ?	Word	↓ Frequency ?	Word	↓ Frequency ?
1 volūdys	794 ...	17 volūdys.	19 ...	33 volūdā:	7 ...
2 volūdu	366 ...	18 volūdom	19 ...	34 seņvolūdas	6 ...
3 volūda	332 ...	19 volūdnīks	19 ...	35 volūdā!	6 ...
4 volūdā	311 ...	20 volūdai.	17 ...	36 volūda"	6 ...
5 volūdas	267 ...	21 volūdom,	16 ...	37 volūdnīceibys	6 ...
6 volūdu,	183 ...	22 volūdai,	15 ...	38 volūdnīkim,	6 ...
7 volūdā.	180 ...	23 volūdas,	13 ...	39 volūdneica	6 ...
8 volūdā,	140 ...	24 volūduos,	13 ...	40 volūdnīceibā	6 ...
9 volūdu.	123 ...	25 volūdas.	12 ...	41 volūdnīkam	6 ...
10 volūda.	95 ...	26 volūduos,	11 ...	42 volūda".	5 ...

<http://www.korpuss.lv/id/MuLa>

Kā MuLa korpusss līdz šim ir izmantots?

Korpusā balstīti pētījumi, piemēram:

- Anna Briška (2013). *Mūsdienu latgaliešu rakstu valodas korpuiss: izveide un izmantošanas iespējas*. Magistra darbs. RTA.
- Nicole Nau (2014). Differential object marking in Latgalian. *Grammatical Relations and their Non-Canonical Encoding in Baltic*. John Benjamins Publishing Company, pp 207 – 255.

Izglītībā, piemēram:

- [Latviešu valodas rokasgrāmata \(valodasrokasgramata.lv\)](http://valodasrokasgramata.lv)
- Nicole Nau (2019). Bringing Latgalian into schools: visions, needs and obstacles. *13th Conference on Baltic Studies in Europe*. <http://naunicol.e.home.amu.edu.pl/wp-content/uploads/2019/09/NauCBSE2019Gdansk.pdf>

Pieredzes stāsts: pētījumi par etnonīmu semantiku

Ethnonym *igauni* in different grammatical forms (72 word tokens in total), *igovins* (one token).

Surname *Igaunis* (6 tokens) and *Ikaunīks* (one token).

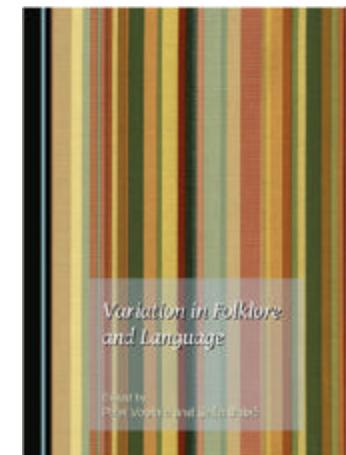
Ethnic nickname *čuhna* (one token).

Collocations: *Ludzys igauni* 'the Estonians of Ludza' (8 tokens); *igaunu volūda* 'Estonian language' (7 tokens) and *igaunu zemnīki* 'Estonian farmers' (2 tokens).

Krīvi (staroveri, pareizticeigī), latgalīši, pūli i igauni Mekši - vysi školā dzīdōja "Боже, царя храни" i vuicejōs četras klases. (MuLa)

Kļavinska, A. (2019). Ethnonyms Denoting Estonians and Their Contextual Meanings in Latgalian Texts. In: Piret Voolaid, Saša Babič (eds.) *Variation in Folklore and Language*. Cambridge Scholars Publishing, p. 7–20.

doc.Number	549
doc.Author	Šķēkurs Uimars
doc.TextTitle	Taldi beja Skolotāji
doc.Source	ru.lv
doc.PublicationOrDate	10-Dec-2009
doc.Publisher	SIA Rēzeknes Vēstis
doc.Place	http://lat.lv/lv/index.php?newsid=2757
doc.TextType	elektroniskie

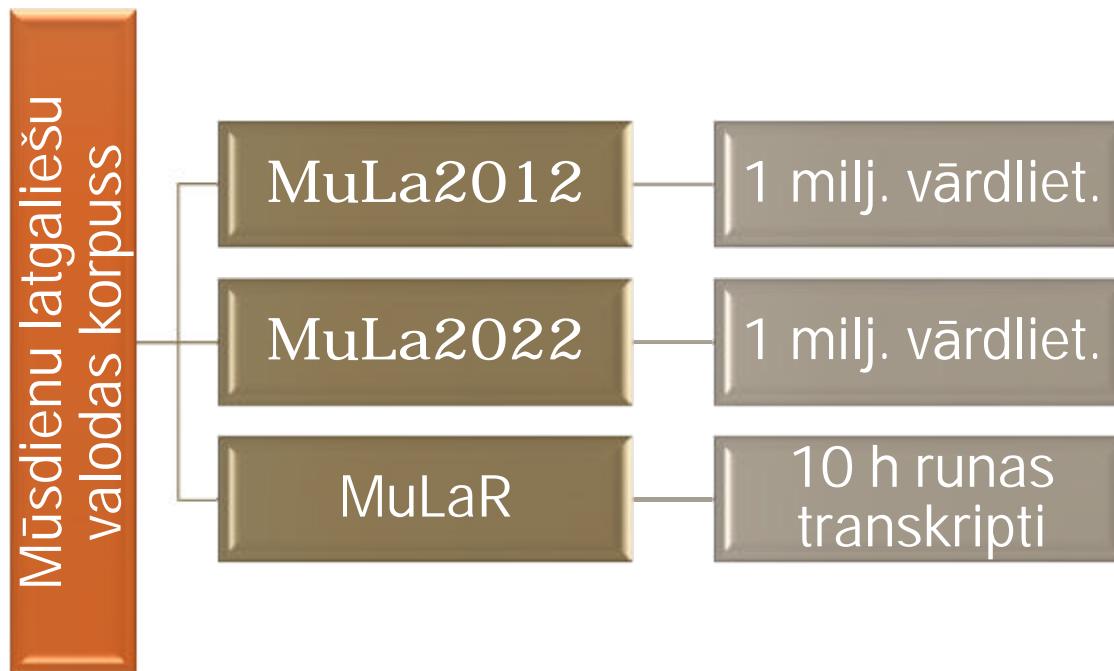


RTA zinātnisko grantu projekts «Austrumu diasporas latgalisko izlokšņu ierakstu datubāzes izveide»

<https://reiti.rta.lv/index.html>

Armands Kociņš (2019). Verbu semantikas īpatnības Sibīrijas latgaliešu mutvārdu tekstos. Bakalaura darbs. RTA

3. Kāds būs mūsdienu latgaliešu tekstu korpuss?



MuLa2022

- Periodika (raksti, intervijas, preses relīzes u.c.).
- Daiļliteratūra (arī tulkotā literatūra).
- Zinātniskie un populārzinātniskie teksti.
- Citi (normatīvie akti, statūti, rezolūcijas).

MuLaR

RUNAS IERAKSTU AVOTI

Folkloras un etnolingvistiskās ekspedīcijas (pēc 2012. gada).

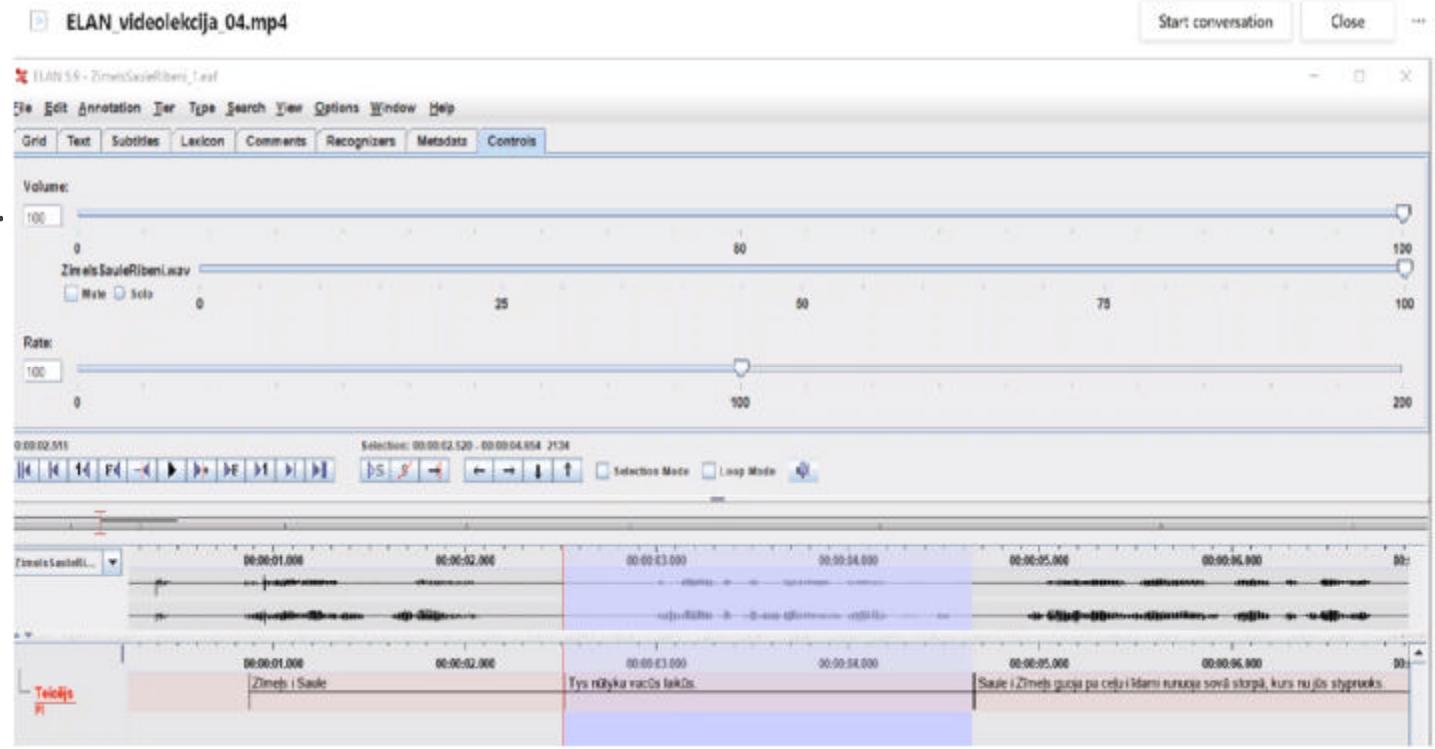
Sibīrijas latgaliešu diasporas ekspedīcijas (2017-2018).

Radio un TV raidījumi.

Lauka pētījums 2021.

TriMCo korpusa dati.

RUNAS IERAKSTU ATŠIFRĒŠANA



4. Vīzija par korpusu izmantošanu plašāk

Uzdevuma komplekts: Rakstības brīvība un rakstības normas

1. Ir veltas un situācijas, kur katrais drīkst rakstīt tā, kā grib, un citas, kur noteikti jāievēro pareizrakstības normas. Diskutējet, kurā grupā ieder dotās situācijas un teksta tipus, ierakstiet tabulā un papildiniet ar citiem piemēriem:

avīze, sludinājums, privāta vēstule, ikumi, komentārs interneta vietnē, drukāts daļliteratūrās daros (piemēram, dzejolu krājums), grāmatas nosaukums uz valka, iekas nosaukumi, īzīpa, eksāmens skoli, zījus interneta portālā, dzejolis kādā interneta mājas lapā

Avar rakstīt, kā grib	B grūti pateikt, atkarīgi no ... (nosaucīet faktorus)	C jāievēro normas



2. Viens no 2018. gada vasaras hitiem ir grupas Latgalīšu Reps dziesma, kas dziedāts daļēji latviešu literārajā un daļēji kādā latgaliešu izloksnē.
Dziesmas nosaukums ir tas pats kā albuma nosaukums, kas izskatās šādi:
[uzraksts: Zārei galī]

No latgaliešu rakstu valodas viedokļa, ņe vārdi ir rakstīti kādaini.

Bet kā būtu pareizi?

Merķauditorija: skolēni 9. klases un vidusskolā

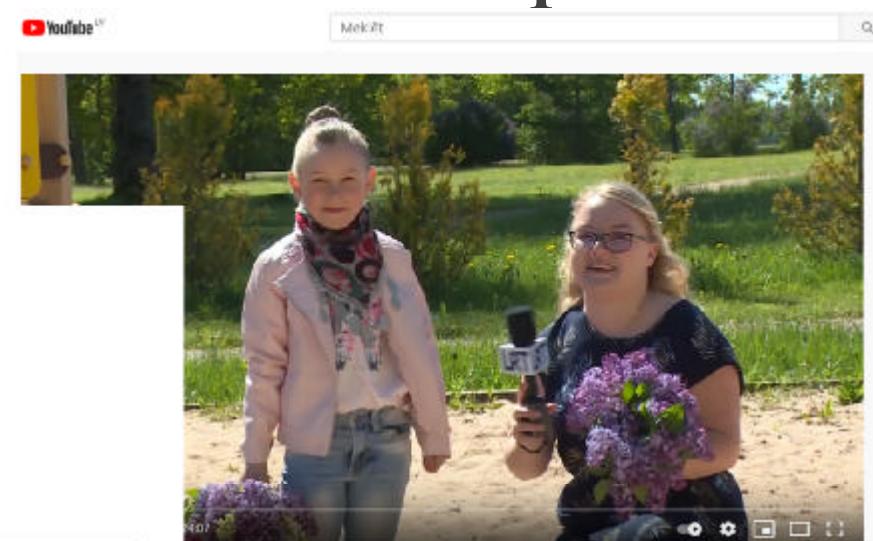
Materiāls: Grupas "Latgalīšu Reps" 2018. gada dziesma

Uzdevums

- Uzdevums PDF formātā
- Uzdevums Word formātā
- Komentāri un atbildes (Word)

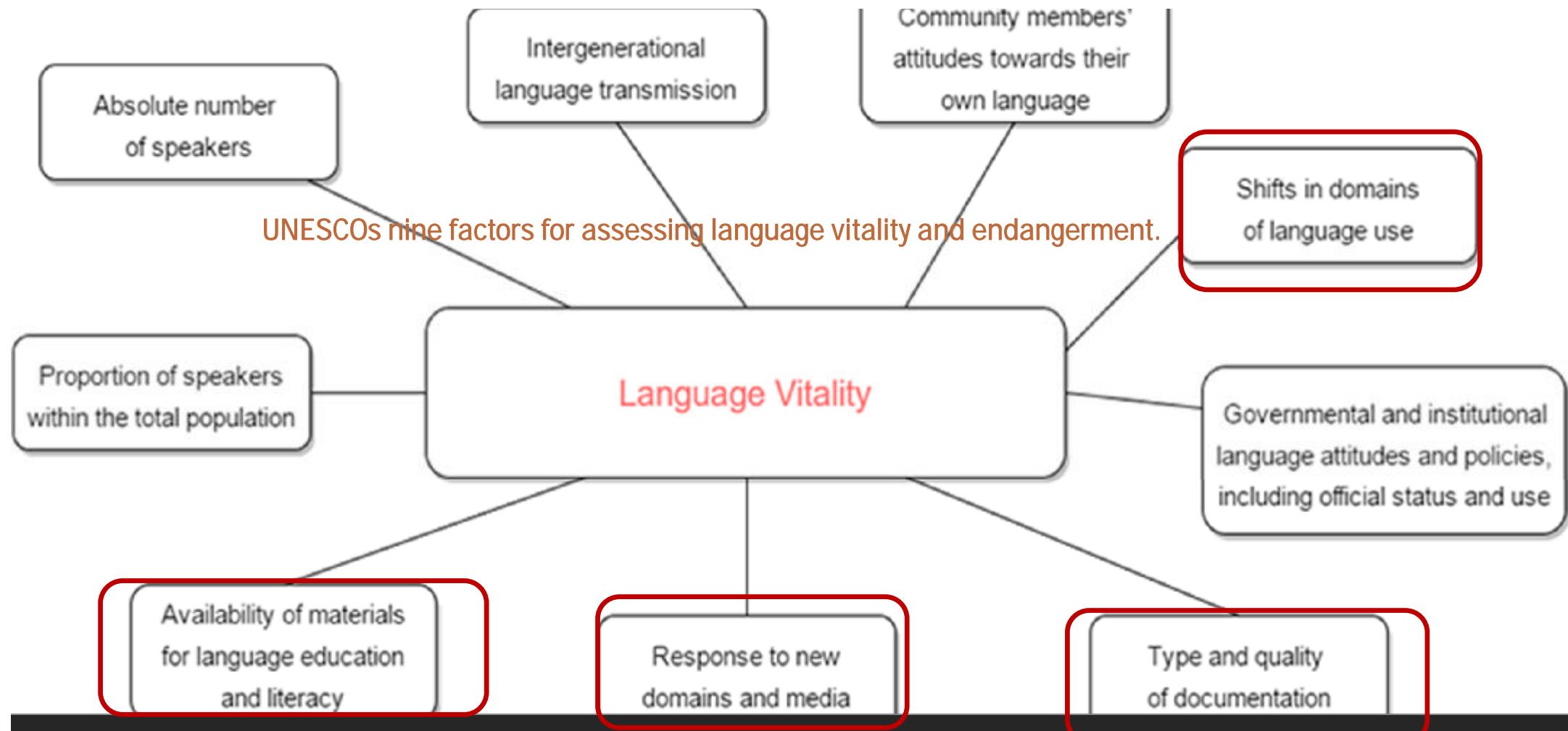


<https://ltgnui.com/2018/08/19/rakstibas-normas-un-rakstibas-briviba/>



RTA dorba grupa sevišķi dasagrīzs šķārsteiklā daīmītena stuosta: „Pirmais latgaliskūs tekstu korpuss pasataise āi itys korpuss ar pasauku „MuLa-01” irā zeimeigs resurss licātūs tekstus i apjam 1 miljonu vuordformu. Vys tik pa mēizraksteibys i morfoloģejis variantu raibums. Myusu tātlicātim piec 2012. gods, i likt izsvoru iz tim dorbum, kā eiga, raistūt tuoluokūs tehnologiskūs volūdys reikus – īim, kluot pi raksteitūs tekstu korpusa planavojam sāļuos sovpateibys. Poši īraksteisim i transkribēsim jaunu mīta A. Kūceņa iniciatīvai runys korpusā īslēgsm Reitu diaspo

UNESCO's nine factors for assessing language vitality and endangerment.



Nobeiguma vietā

PALDIES PAR UZMANĪBU!