



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

Mūsdienu latgaliešu tekstu korpuss: paveiktais un nākotnes ieceres

PROF.SANITA MARTENA, DR.PHILOL.ANTRA KĻAVINSKA
(RĒZEKNES TEHNOLOĢIJU AKADEMIJA)

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

10.-11.03.2021.

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

PRIEKŠLASĪJUMA JAUTĀJUMI

HipiLatLit

 Lietuvių

 Latgaliešu

 Angļu

[SĀKUMS](#) [PRODUKTI](#) [AKTIVITĀTES](#) [GALERIJA](#)

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

Projekta „HipiLatLit” rezultāti:

- [1. Mūsdienu latgaliešu tekstu korpuss](#) (1 miljons vārdlietojumu);
- [2. Lietuvių-latviešu-lietuvių paralēlo tekstu korpuss](#) (8 miljoni vārdlietojumu);
- [3. Lietuvių-latviešu-latgaliešu vārdnīca](#) (10 tūkstoši šķirķļu);
- [4. Starpdisciplināra doktora studiju programma baltistikā.](#)
- Datorlingvistikas modulis filoloģijas maģistrantiem Rēzeknes Augstskolā.

- specializēts korpuss — 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).

Query **volūd.*** 3,605 > GDEX 3,605 (2,730.3 per million)

Page 1 of 19 [Next](#) | [Last](#)

1	diena.lv	dzeivi i sovys zemis viesturi navarātu runuot sovā	volūdā	. Dūmuot , bez sovu sakņu i ar nasauordeitu idēntitati
2	diena.lv	stypri i kai muoti , i mōti . Ir labi , ka es jei sovā	volūdā	varu pīraksteit , cik cīši es juos lūbeju . Bet vyspuor
3	bezpvn.lv	Vysys dzīsmis skaidrā , saprūtamā latgalīšu volūdā -	volūdā	, kurā grupa dzīd i dūmoj . Albumā var atrast arī
4	bezpvn.lv	laikā . Vysys dzīsmis skaidrā , saprūtamā latgalīšu	volūdā	- volūdā , kurā grupa dzīd i dūmoj . Albumā var atrast
5	jureits.wordpress.com	paruoda , ka Latgolys latvīšu volūda jeb latgalīšu	volūda	ir sanuoka par pošu vaļsti . Maņ juopīkreit dokumentā
6	diena.lv	asu suokuse runuot latgaliski , lai jis nacer , ka	volūdu	byus īspiejams maineit . Tys ir par daudz sarežģeitai
7	raibiis.wordpress.com	tū pusi) , cytu fakultati , kur juodazaregistrej	volūdys	kursim . Tai kai breivs - veromīs kartī , ka juodaīt
8	geoviita.wordpress.com	latgaliski ! Kuri ir nūdrūšnuojūši tū , ka latgalīšu	volūda	ari niule ir dzeiva . Maņ ir lela prīca par vysim
9	jureits.wordpress.com	izaveiduot . Tys faktiski ari paruoda , ka Latgolys latvīšu	volūda	jeb latgalīšu volūda ir sanuoka par pošu vaļsti .
10	diena.lv	tikai slapanai , ar sovejim - kai taidā konspirativajā	volūdā	. Koč Latgola jai ir svātuo zeme . Kur zeme ar dabasim
11	diena.lv	krīvim es ni nu šuo , ni nu tuo nasuokšu runuot cytā	volūdā	. Maņ tys ir par daudz suopeigai . Saīt taidys psihologisks
12	diena.lv	Rēzeknē nūtiks informativs seminars par latgalīšu rokstu	volūdu	i juos nuusacejumim . Vēļ es varātu izstuosteit ,
13	geoviita.wordpress.com	Puordeviejs ir jauns puiss , bet narunoj ni angļu , ni krīvu	volūdā	. Labi , ka ir rūkys . Nūpierkam gurķus i tynamīs
14	jureits.wordpress.com	tymā nav ni vuoda par Latgolys latvīšu jeb latgalīšu	volūdu	. Gon latgalīšim , gon Saeimys deputatim maņ juoatguodynoj

15	geoviita.wordpress.com
16	geoviita.wordpress.com
17	diena.lv
18	geoviita.wordpress.com
19	naktineica.lv
20	naktineica.lv
21	edeite.wordpress.com
22	edeite.wordpress.com
23	geoviita.wordpress.com
24	diena.lv
25	cyxob.lv

doc.Number	130
doc.Author	saprgē (Ilze Sperga)
doc.TextTitle	Viltuota dzeive i zogti uobeļeiši
doc.Source	diena.lv
doc.PublicationOfDate	14-Okt-2008
doc.Publisher	SIA Dianas Mediji
doc.Place	http://www.diena.lv/blogi/viltuota-dzeive-i-zogti-uobelesi-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 korpus **līdz šim** ir izmantots?

Korpusā balstīti pētījumi, piemēram:

- Anna Briška (2013). *Mūsdienu latgaliešu rakstu valodas korpus: izveide un izmantošanas iespējas*. Maģistra 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 *lkaunīks* (one token).

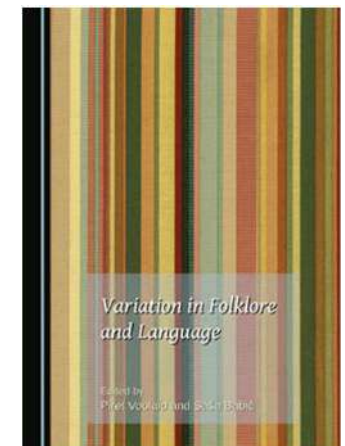
Ethnic nickname *čuhna* (one token).

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

Krīvi (staroveri, pareizticeigī), latgalīši, pūli i igauni Mekši - vysī skolā dzīdēja "Боже, царя храни" i vuicejās 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	Sīkāns Ilmārs
doc.TextTitle	Taidi beja skolotāji
doc.Source	ru.lv
doc.PublicationOfDate	10-Dec-2009
doc.Publisher	SIA Rēzeknes Vēstis
doc.Place	http://lat.ru.lv/index.php?newsid=2757
doc.TextType	elektroniskie



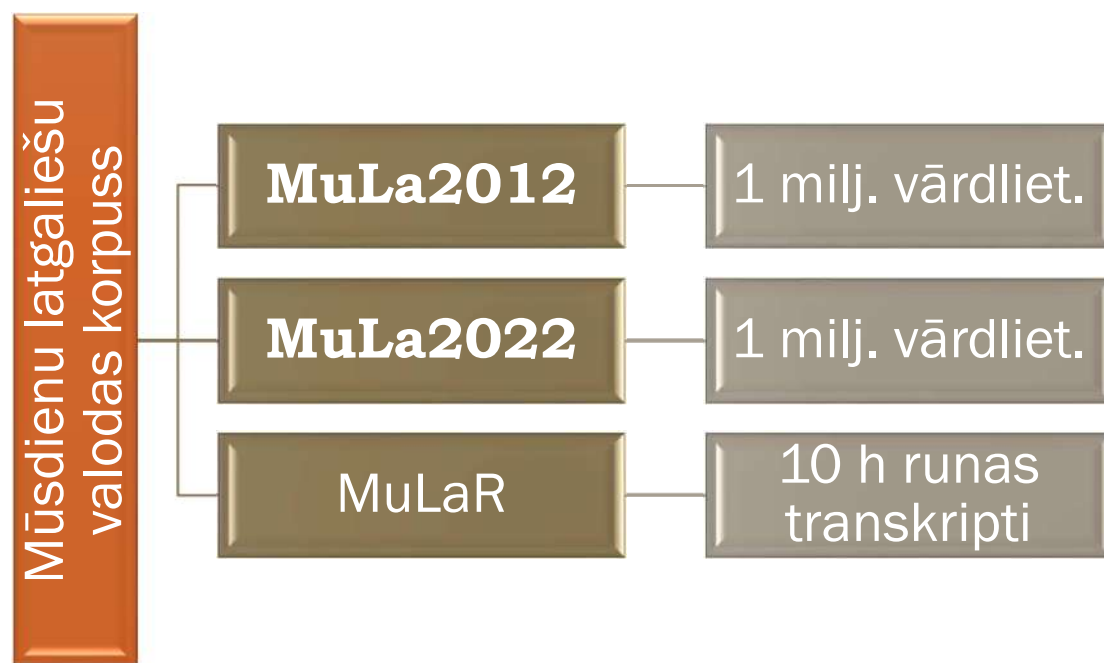


RTA zinātnisko grantu projekts «**Austrumu diasporas latgalisko izlokšņu ierakstu datubāzes izveide**»

<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 korpus?



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

RUNAS IERAKSTU ATŠIFRĒŠANA

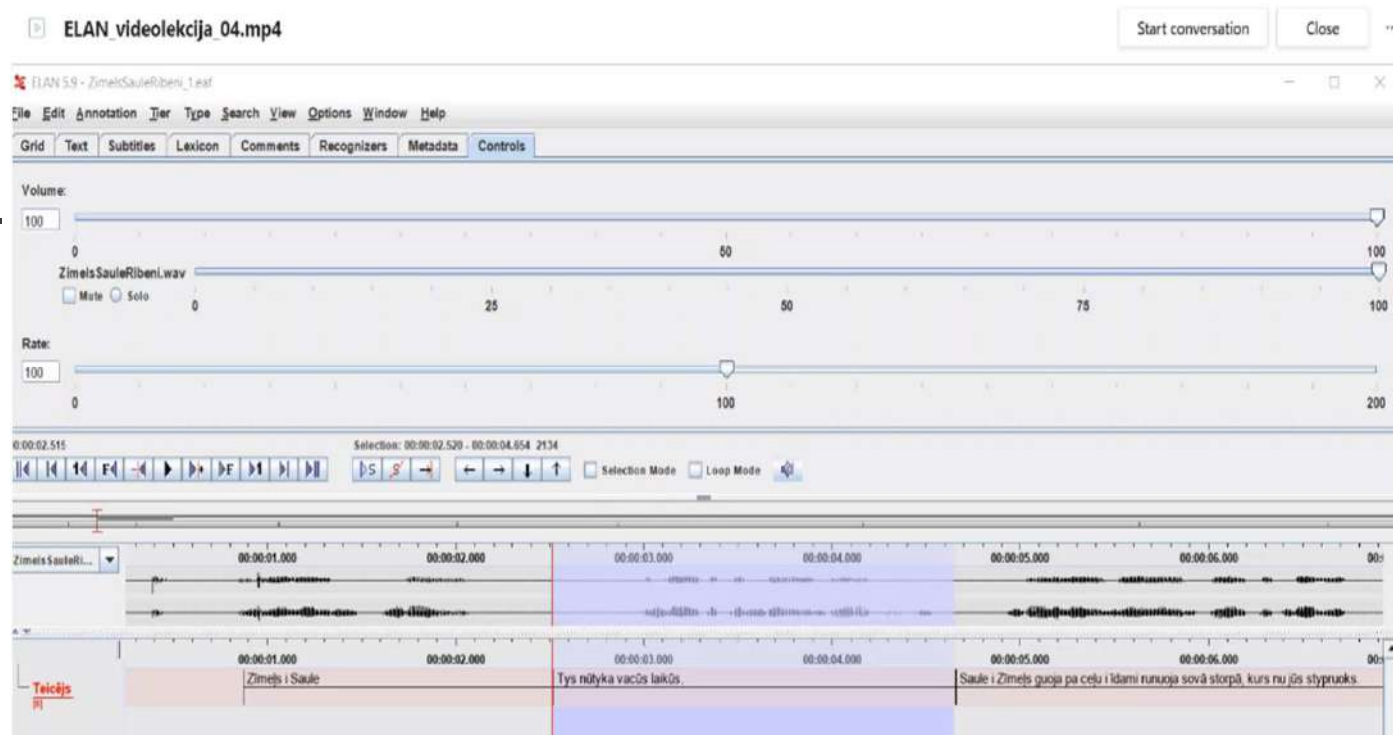
Folkloras un etnolingvistiskās ekspedīcijas (pēc 2012. gada).

Sibīrijas latgaliešu diasporas ekspedīcijas (2017-2018).

Radio un TV raidījumi.

Lauka pētījums 2021.

TriMCo korpusa dati.



4. Vīzija par korpusu izmantošanu plašāk

Uzdevuma komplekts: Rakstības brīvība un rakstības normas

1. Ir vietas un situācijas, kur katrs drīkst rakstīt tā, kā grib, un citas, kur noteikti jāievēro pareizrakstības normas. Diskutējiet, kurā grupā iedert dotās situācijas un teksta tipus, ierakstiet tabulā un papildiniet ar citiem piemēriem:

avīze, sludinājums, privāta vēstule, likumi, komentārs interneta vietnē, drukāts daijliteratūras darbs (piemēram, dzejoļu krājums), grāmatas nosaukums uz vāka, ielas nosaukumi, izziņa, eksāmens skolā, ziņas interneta portālā, dzejolis kādā interneta mājas lapā

A var rakstīt, kā grib	B grūti pateikt, atkarīgi no ... (nosauciet faktorus)	C jāievēro normas

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

No latgališu rakstu valodas viedokļa, šie vārdi ir rakstīti kļūdaini.

Bet kā būtu pareizi?

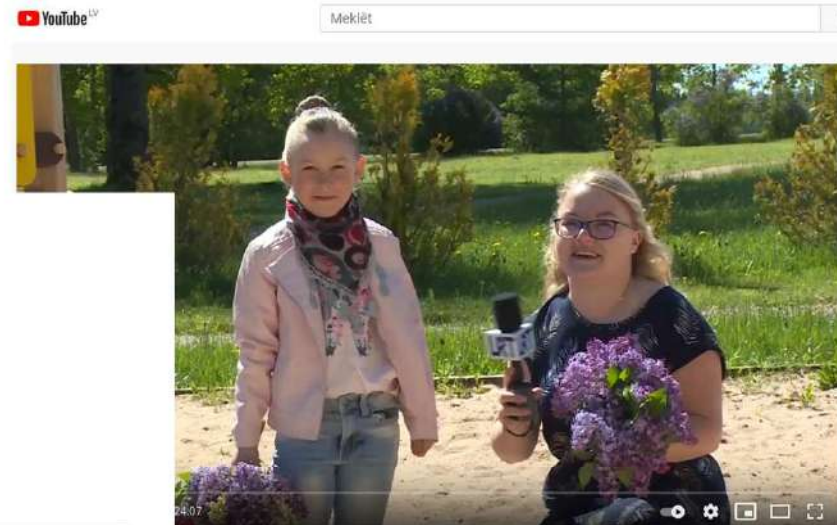


Merķauditorija: skolēni 9. klasē un vidusskolā

Materiāls: Grupas "Latgališu Reps" 2018. gada dziesma

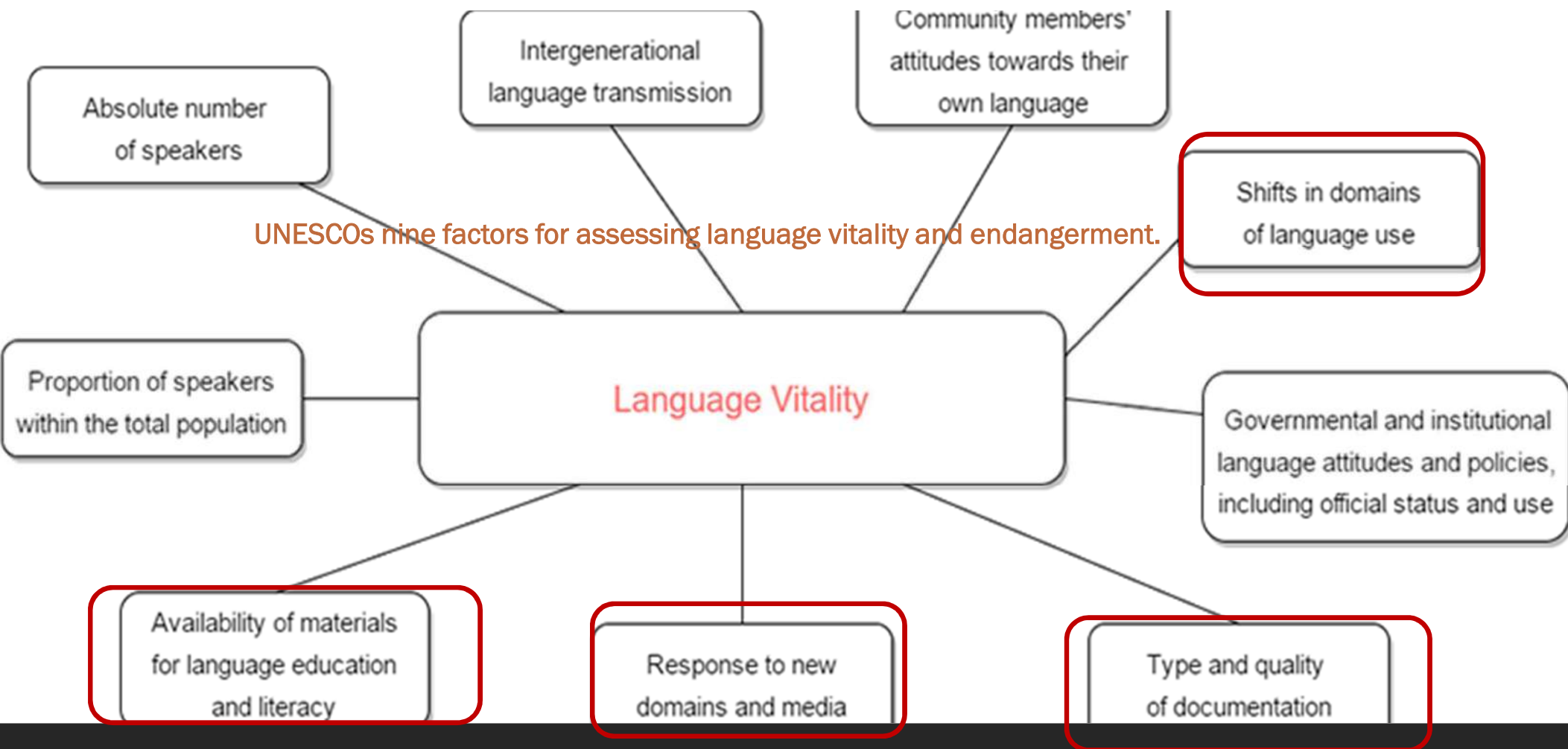
Uzdevums

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



RTA dorba grupa sevišķi dasagrīzs šķārsteiklā daīme tena stuosta: „Pyрмаis latgaliskūs tekstu korpuss pasataise īi itys korpuss ar pasauku „MuLa-01” irā zeimeigs resurss licātūs tekstus i apjam 1 miljonu vuordformu. Vys tik pa m ēizraksteibys i morfologejis variantu raibums. Myusu tār licātim piec 2012. gods, i likt izsvoru iz tim dorbim, ki eiga, raistūt tuoluokūs tehnologiskūs volūdys reikus – l m, kluot pi raksteitūs tekstu korpusa planavojam sas iluos sovpatēibys. Poši īraksteisim i transkribēsīm jaunū mī ita A. Kūceņa iniciatīvai runys korpūsā īslēgsim Reitu diaspo

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



Nobeiguma vietā

PALDIES PAR UZMANĪBU!