



Latvijas Zinātnes padome

Latgaliešu rakstu valodas pareizrakstības pārbaude digitālajā vidē: “MuLa” korpusa iespējas

Antra Kļavinska

Akadēmiķa Jāņa Endzelīna 149. dzimšanas dienas atceres starptautiskā zinātniskā
konference “Valodas reģionālie varianti”, Rīga, 22.- 23.02.2022.

Pētījuma jautājumi

- Kādas ir dažādu valodu (tostarp reģionālo un mazāk lietoto) **iespējas un pieredze pareizrakstības pārbaudes rīku izveidē?**
- Kādus **resursus** (tostarp digitālos) **latgaliešu rakstu valodas pareizrakstības pārbaudei izmanto** latgaliešu tekstu autori, korektori, skolotāji?
- Kādā veidā latgaliešu valodas pareizrakstības pārbaudei var izmantot **Mūsdienu latgaliešu tekstu korpusu (“MuLa”)?**
- Kādas ir **“MuLa” korpusa priekšrocības un trūkumi?**

Metodoloģija

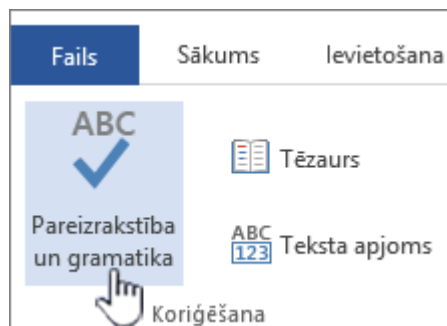


- Teorētiskās literatūras, datu par pareizrakstības rīkiem analīze.
- Daļēji strukturētās intervijas (latgaliešu tekstu autori, korektori, skolotāji).
- Digitālo resursu iespēju salīdzinājums latgaliešu rakstu valodas pareizrakstības pārbaudei.

Iespējas un pieredze pareizrakstības pārbaudes rīku izveidē

- Automatizēta pareizrakstības kļūdu noteikšana tiek izmantota datorizētās valodu apguves lietojumprogrammās un automātiskās vērtēšanas sistēmās (Dikli 2006; Warschauer, Ware 2006).
- Pieejas: vārdnīcā balstīta pieeja (*dictionary based approach*) (Morris, Cherry 1975), attāluma rediģēšanas pieeja (*edit distance approach*) (Levenshtein 1966); statistiskā pieeja (*N-Gram approach*)(Figueroa, Atkinson 2012) u. c.

Pareizrakstības un gramatikas pārbaudes rīki



Atrastas 3 iespējamās kļūdas

- turpmuokais → turpmākais
- juļs sūlei izadavuu → juļs sūlei izadavuu
nepareiza pareizrakstība, paskaidrojums, kas te ir slikti.
vai → juļs sūlei izadavu
vai → ja nu ir vairāki varianti
- Juoņa dīnu), tod – nepareiza interpunkcija

Latgaliešu valodas pareizrakstība

Latgaliešu valodas pareizrakstības rīks atrod kļūdas un piedāvā to iespējamās labojumus latgaliski rakstītā tekstā.

Kādi resursi tiek izmantoti latgaliešu rakstu valodas pareizrakstības pārbaudei?

- (..) pazaveru **pareizraksteibys nūteikumus**. Vysbīžuok - **zvons draugam- volūdnīkam** (..)
- (..) Nūsacejumi ir cīš konspektivi. Daudzys lītys vyspuor nav atrūnamys. Partū izmontoju **gramatiku "Vasals!"**
- (..) Ka verūs kaida vuorda, pasaveru daudzuos **vuordineicuos, ari ME**.



LR Tieslietu ministrijas
Valsts valodas centrs

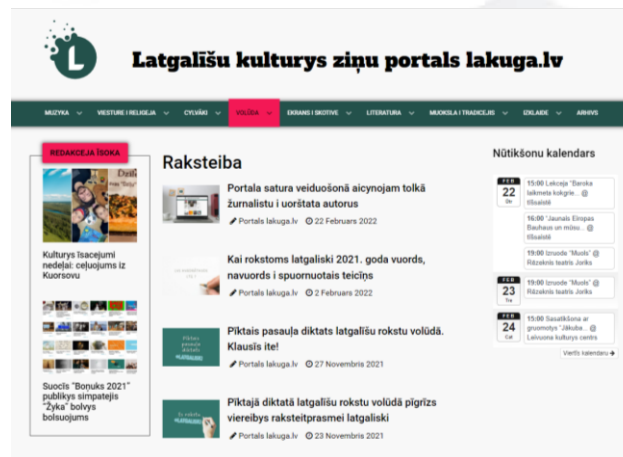
**Latgaliešu
pareizrakstības noteikumi**

LR Tislietu ministrijas
Valsts volūdys centrs

**Latgalīšu
pareizraksteibys nūsacejumi**

Kādi resursi tiek izmantoti latgaliešu rakstu valodas pareizrakstības pārbaudei?

- *Kod juopasaver kaidi vītvuordi latgaliski, ir lakugā volūdys sadaļa "Raksteiba" (..) Dzerauņu nūsaukumus pasaveru Vietvārdu datubāzē.*
- *"Pyрмаis nūteikti ir HipLatLit projekta latgalīšu-latvīšu-lītuviešu vuordineica. Tod ira vuordineica.lv, kas ira taidys breivpruoteigū projekts, un ir ari tei poša Stroda vuordineica (..)." (Husare, "Kolnasāta" 04.04.2020.)*



Kādi resursi tiek izmantoti latgaliešu rakstu valodas pareizrakstības pārbaudei?

- *MuLa naasu lītuojuse, bet tagad ir interese pasaskatēt!*



- *(..) hugo pyrma loba laika nazkas interesants beja izliecs, par kū pasasmieju, i vaira nasavieru.*



S. Martena, N. Naua, A. Briška “Latgaliešu valodas korpuss citu Eiropas mazāk lietoto valodu kontekstā: korpusa raksturojums, lietojums un potenciālā iespējošana” (iesniegts publicēšanai žurnālā “Letonica”)

Aptauja par korpusa izmantošanu (2021), 214 respondenti

		Izmanto (varētu izmantot) pareizrakstības pārbaudei
Zina, lieto	15	6
Zina, nelieto	43	22
Nezina, nelieto	156	71

Digitālie resursi latgaliešu rakstu valodai

Valodu rīki ▼ Par Hugo Palīdzība

HUGO.LV

Latviešu

Sākums / Valodu rīki / Latgaliešu rakstu valodas pareizrakstības rīks

Starptautiskuo zynuotniskuo konference
"Regionaluos volūdys izgleiteibā: nu teksta
leidz tekstprateibai"2022. gods 16. – 17. junī
Rēzeknē (Latvijā)

Atrastas 5 iespējamās kļūdas

- Starptautiskuo → Storptautyskyuo ✓ ✕
- konference → konference ✓ ✕
- Regionaluos → Nacionaluos ✓ ✕
- gods → gūds ✓ ✕
- Latvijā → Latvejā ✓ ✕

Valodu rīki
Par Hugo
Palīdzība

HUGO.lv

Latvian

Sākums / Valodu rīki / Latgaliešu rakstu valodas pareizrakstības rīks

konference

Vārda analīze

konference

konference

ešu dzimte; vienskaitlis; vokatīvs

ešu dzimte; vienskaitlis; nominatīvs

nas tabulu

Locīšanas tabula

Lietvārds **konference**

	Vienskaitlis	Daudzskaitlis
Nominatīvs	konference	konferencis
Ģenitīvs	konferencis	konferenču
Datīvs	konferencei	konferencem
Akuzatīvs	konferenci	konferencis
Instrumentālis	ar konferenci	ar konferencem
Lokatīvs	konferencē	konferencēs
Vokatīvs	konference!	konferencis!

Lietvārds konference

	Vienskaitlis	Daudzskaitlis
Nominatīvs	konference	konferencis
Ģenitīvs	konferencis	konferenču
Datīvs	konferencei	konferencem
Akuzatīvs	konferenci	konferencis
Instrumentālis	ar konferenci	ar konferencem
Lokatīvs	konferencē	konferencēs
Vokatīvs	konference!	konferencis!

konference vs. konference

Baltīšu - latgališu
Latviešu - latgaliešu  **Vurordineica.lv**

konference *subst.* konference; **preses k.** — presis konference; **zinātniskā k.** — ziniskuo konference; **starptautiskā k.** — škārsregionaluo konference

Piezīmes, avoti, varianti

KONFERENCEJA

- psl.) - AK*
- Vajag saorganizēt (kai pioneru, oktobram, tai i naorganizēitu) bārmu **konferencijas**, sapulces, saskrīšonas un t. t. - ("Jaunais Latgalīts" ("Taisneibys" dalykums), 1928 g. majs, Nr. 3, 1
 - ..apsprides, **konferencijas**.. - ("Taisneiba", 1930 10 25, 3 *psl.*) - AK
 - Pymō bēdeigūs zemdaru **konferencija** Sibīrī. - ("Taisneiba", 1928 03 09, 2 *psl.*) - AK

KONFERENCEJA

- Personas vordus: es, mes tev, sev, naskaidri izrunoj (mias, ņav, ..Vysvineibas lat. izgleit. darbiniku **konferencija** nuspridia raksteit e - es, mes, tev, sev. - (A. Eisuļs, J. Siliniks "Latgališu voluda (Etimologija)", 1928, 21 *psl.*) - AK

KONFERENCEJA

- ..storptautu socijalistu **konferencijā**.. - ("Auseklis", 1907 08 16, 4 *psl.*) - AK

Komentāri [0]

Meklēt vēlreiz:

Vārds:

konference vs. konference



LIETUVIŲ - LATVIŲ - LATGALIEČIŲ KALBŲ ŽODYNAS
LIETUVIEŠU - LATVIEŠU - LATGALIEŠU VĀRDNĪCA
LĪTVIŠU - LATVĪŠU - LATGALĪŠU VUORDNEICA

Latviešu
Lietuviešu
Latgaliešu
Angļu

SĀKUMS VĀRDNĪCA

Lietuviešu-latviešu-latgaliešu vārdnīca

Valoda: ☐ Lietuviešu ☒ Latviešu ☐ Latgaliešu

Meklējamais vārds: meklēt

Meklēšanas rezultāts: 1 vienība(-as).

1	2	LT	3	4	5	LV	6	LTG	7
konferencija	konferēncija		sf.	1	116	konference		konference	

konference vs. konference

CONCORDANCE MuLa

simple **konferenc*** 98 (73.94 per million)

☐ Details Left context KWIC Right context

1	☐	<input type="info"/>	doc#15	janizētājā Ukrainas kristianizēšonas tyukstūšgadei veļteitōs	konferences	sēdē. 1989. gods vēļ īvārōjams ar tū, ka gondreiz pēc 50 go	<input type="file"/>
2	☐	<input type="info"/>	doc#97	u paleidzeibu. </s><s> Rodōs īspēja sasaukt kongresus un	konferences	Reigā, Rēzeknē, Daugavpilī un nūdybynōt Latgaļu kuļtūras l	<input type="file"/>
3	☐	<input type="info"/>	doc#113	ais PSRS VDK generals). </s><s> Pēc sanōksmes, preses	konferencē	cyta storpā Sergejs Ivanovs paziņōja, ka gadejumūs, kod ka	<input type="file"/>
4	☐	<input type="info"/>	doc#152	IS SISTEMĀ (1992-2002) </s><s> Pymō pasauīa latgalīōu	konference	1992. godā ir Latgolas atmūdas kulminācijas punkts. </s><s>	<input type="file"/>

CONCORDANCE MuLa

simple **konferenç*** 130 (98.08 per million)

☐ Details Left context KWIC Right context

1	☐	<input type="info"/>	doc#64	yka izanalizāts i daruodeits referatā 2. </s><s> Latgalistikys	konferençē	2009. godā, Feimaņu izlūksne sovā sovumā tīt latgalīšu liter	<input type="file"/>
2	☐	<input type="info"/>	doc#82	Izinem-folkloristem, apdzīduotuojom i boltajom varažbitkom	konferençis	režimā i īsaceja kolpu Juri ar būra veikš, kū juoīlīkam padus	<input type="file"/>
3	☐	<input type="info"/>	doc#96	a daudz loba — organizēja školuojuojim kursus, seminarus,	konferençis	; izdeve "Latgolas školū" (1921. — 1938.) — vierteigu pedač	<input type="file"/>

konference vs. konference

- “(..) praksē nūstastprynuojuse daudz lītuotuo vuorda *konference* i piec juo parauga cytu vuordu raksteiba ar leidzskanū meikstynuošonu (..), tok “Latgalīšu raksteibys nūsacejumu” (2007) 1.6. punktā saceits, ka “leidzskanū burtus / un *n* nameikstynoj jaunuokūs aizjimtūs vuordūs: *kultura, instituts, intelligence, tolerance* i tml.” Komisejā nūlamts itū vaicuojumu izpieteit dziļuok i izsprīst par jaunu, a pagaidom praksē pīsatureit pi Nūsacejumūs saceituo. ” (Ūdre 2019)

MuLa iespējas: frekvence

- *Parosti korpusā vierūs, cik izplateits vuords.*
- *Ja vajag vuorda pamatformu, verūs vuordineicā, ja nazynu kaidu **formu** – **MuLa korpusā**.*
- Leksēma ‘brāļi’ MuLa korpusā konstatēta četros pieraksta variantos:

*brōli – 79,
bruoli - 67,
bruoļi – 2,
brōļi – 1.*

MuLa iespējas: konkordance

- **Kopā un šķirti rakstāmi vārdi:** *pa prīšku* ‘pa priekšu’ / *paprīšku* ‘vispirms’

Arnolds palaide Rutu pa prīšku, pats nūpaka!
(MuLa)

*Jaunos dzeivis vītys Sibirī izalaseit paprīšku
brauc t. s. vītraudzs (ходок) (MuLA)*

MuLa iespējas: metadati

The screenshot displays the MuLa web interface. At the top, there are logos for Latvia and Lithuania, along with the text 'bringing neighbours closer'. Below this, a search bar contains the query 'konference'. The interface shows a list of search results, each with a document title and a snippet of text. A sidebar on the left contains various navigation options like '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', and 'Visualize'. A table at the bottom of the results list shows metadata for the selected document.

doc.Number	1284
doc.Author	Leikuma Lideja
doc.TextTitle	Gruomota Skoluojujīm (pīzeimīs, aizdavumi)
doc.Source	Gruomota Skoluojujīm
doc.PublicationOfDate	1993.gads
doc.Publisher	Izdevniecība Lielvārds
doc.Place	Lielvārde
doc.TextType	drukātie

- (..) ja arī **izmontoju korpusa datus** i kod ir atšķirīgys formys, tod maņ pošai **svareigs kriterejs ir teksta autors** - kurs tū ir rakstejs **voi izdeviejs**. (..) izavieleju tū tekstu, kura autors pat subjektivi viertejūt ir leluoks lītpratiejs latgalīšu volūdys raksteibā.

MuLa trūkumi

A decorative graphic in the top right corner consisting of several 3D cubes and rectangular prisms in shades of light blue, grey, and white, arranged in a cluster.

- Neliels apjoms (~ 1 miljons vārdlietojumu).
- Divas ortogrāfijas.
- Korpuss nav morfoloģiski marķēts.

Kopsavilkums (1)

- Latgaliešu valodas pareizrakstības normēšanā vēl ir neatrisināti, diskutējami jautājumi, trūkst mūsdienīgas pareizrakstības vārdnīcas.
- Latgaliski rakstošie pareizrakstības pārbaudei salīdzinoši maz izmanto digitālos resursus.
- “MuLa” korpusu izmanto galvenokārt vārdformu pareizrakstības pārbaudei (pamatformas pareizrakstības noskaidrošanai galvenokārt tiek izmantotas vārdnīcas).
- Pareizas vārda formas meklējumos korpusā galvenie kritēriji ir:
1) vārdformas lietojuma biežums; 2) teksta avots (autors).

Kopsavilkums (2)

- Iespēja pareizrakstības pārbaudei izmantot vārda / vārdformas lietojuma biežumu, kontekstu, pārbaudīt teksta avota uzticamību, kā arī ātras meklēšanas iespējas ir “MuLa” korpusa kā pareizrakstības pārbaudes rīka priekšrocības.
- “MuLa” būtiskākais trūkums ir tas, ka korpusā nav morfoloģiski marķēts, turklāt tam ir salīdzinoši neliels apjoms, tajā ievietoti periodā no 1988. līdz 2012. gadam publicēti teksti, proti, publikācijas gan pirms, gan pēc patlaban aktuālo “Latgaliešu rakstības noteikumu” (2007) apstiprināšanas.

Kopsavilkums (3)

- Digitālie resursi vēl ir pilnveidojami, lai tie būtu uzticami un ērti izmantojami.
- “MuLa” korpuss tiek pilnveidots un papildināts ar jaunākajiem tekstiem, plānots, ka korpusa 2022. gada versija tiks papildināta ar aptuveni vienu miljonu vārdlietojumu.
- FAIR datu principi (*findable, accessible, interoperable, reusable*). Datu **savietojamība** – priekšnosacījums kvalitatīva automatizēta latgaliešu rakstu valodas pareizrakstības pārbaudes rīka izveidei.

Izmantotā literatūra

- Dikli, S. (2006). An Overview of Automated Scoring of Essays. *The Journal of Technology, Learning and Assessment*, 5(1). Retrieved from <https://ejournals.bc.edu/index.php/jtla/article/view/1640>
- Figueroa, A., Atkinson, J. (2012). Contextual Language Models For Ranking Answers To Natural Language Definition Questions. *Computational Intelligence*. **28** (4): 528–548.
- Husare, E. (2020). Īspiejis vuiceitīs latgalīšu volūdu škārssteikā (Guntas Igavenas intervija ar Edīti Husari). LR Latgales studija, radio raidījums «Kolnasāta». Pieejams: <https://www.lsm.lv/raksts/kultura/kulturtelpa/latgaliesu-valodas-apgusanai--speles-datora-un-tiessaistes-materiali.a356022/>
- Morris, L. Cherry. (1975). Computer detection of typographical errors. *IEEE Transactions on Professional Communication*, 18, 54–64.
- Warschauer, M., Ware, P. (2006). Automated Writing Evaluation: Defining the Classroom Research Agenda. *Language Teaching Research*, 10(2), 157-180. Retrieved February 23, 2022 from <https://www.learntechlib.org/p/101033/>.
- Ūdre, S. (2019). Baze voi bāze, Modyuna, Modūne voi Modūna? Lakuga.lv Pieejams: <https://www.lakuga.lv/2019/06/17/baze-voi-baze-modyuna-modune-voi-moduna/>
- В. И. Левенштейн (1965). Двоичные коды с исправлением выпадений, вставок и замещений символов [Binary codes capable of correcting deletions, insertions, and reversals]. *Доклады Академии Наук СССР*. **163** (4): 845–848.

Paldies!

antra.klavinska@rta.lv

