Digital Resources of Latgalian in the Context of European Lesser-used Languages

Antra Kļavinska, Sanita Martena

16th Conference on Baltic Studies in Europe CONVERGING PATHS: THE BALTIC BETWEEN EAST AND WEST
24-26 April, 2025; University of Cambridge, UK

Overview

1 Latgalian: a short sociolinguistic insight

2 Latgalian (digital) language vitality

3 The process of developing digital resources and tools

4 (Digital) literacy: digital resources for learning Latgalian in the context of other lesser-used languages

5 Conclusions

The main aim of the study

To identify and analyse issues that are important in the process of creating Latgalian language corpora and an electronic dictionary and

to discuss challenges and opportunities in creating and using digital resources and tools for Latgalian in research and education in comparison with other European lesser-used languages.

1 Latgalian: a short sociolinguistic insight

- A Baltic variety closely related to Latvian, officially recognised as "historical variant of Latvian"

- From a perspective of European languages, it may be classified as a regional (autochthonous minority) language similar to e.g. Kashubian, Scots, Low German

- Closely related to Standard Latvian, but with distinct cultural traditions

- Latgale: ~35% use Latgalian regularly

written language: strongly standardized

⇔

spoken language:
lack of standardization & big diversity

1 Latgalian: Language practices

- **Oral**:

- in private communication but also at work places

- in religious contexts (in Catholic churches or in private communication between priests and individuals)

- in the media and in the cultural sector (music – CDs, on Youtube etc., during cultural events, concerts, festivals and similar, some few national TV and radio programmes).

- **Written Latgalian**:

- in media (magazines, newspapers, brochures), www.lakuga.lv

- in private communication (text messages, e-mails, blogs)

- on public signs

- in tourism

- in culture, religion

- education (summer schools and camps, Latgalian language courses, Latgalian as an optional course in schools in Latgale (since 2013))

- science («Via Latgalica», Language Corpora)
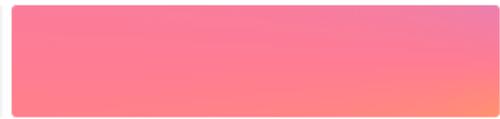
2 Latgalian language vitality
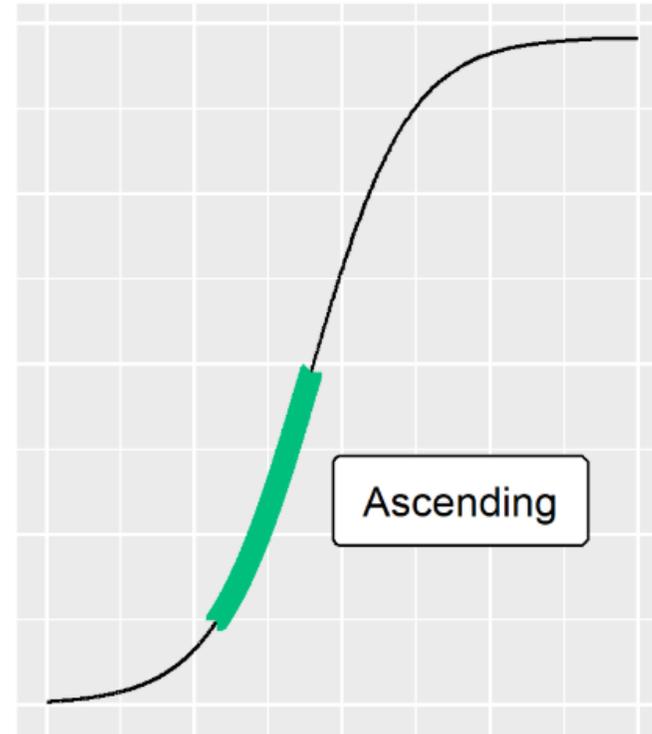
Language Vitality ⭐

- Institutional
- **Stable**
- Endangered
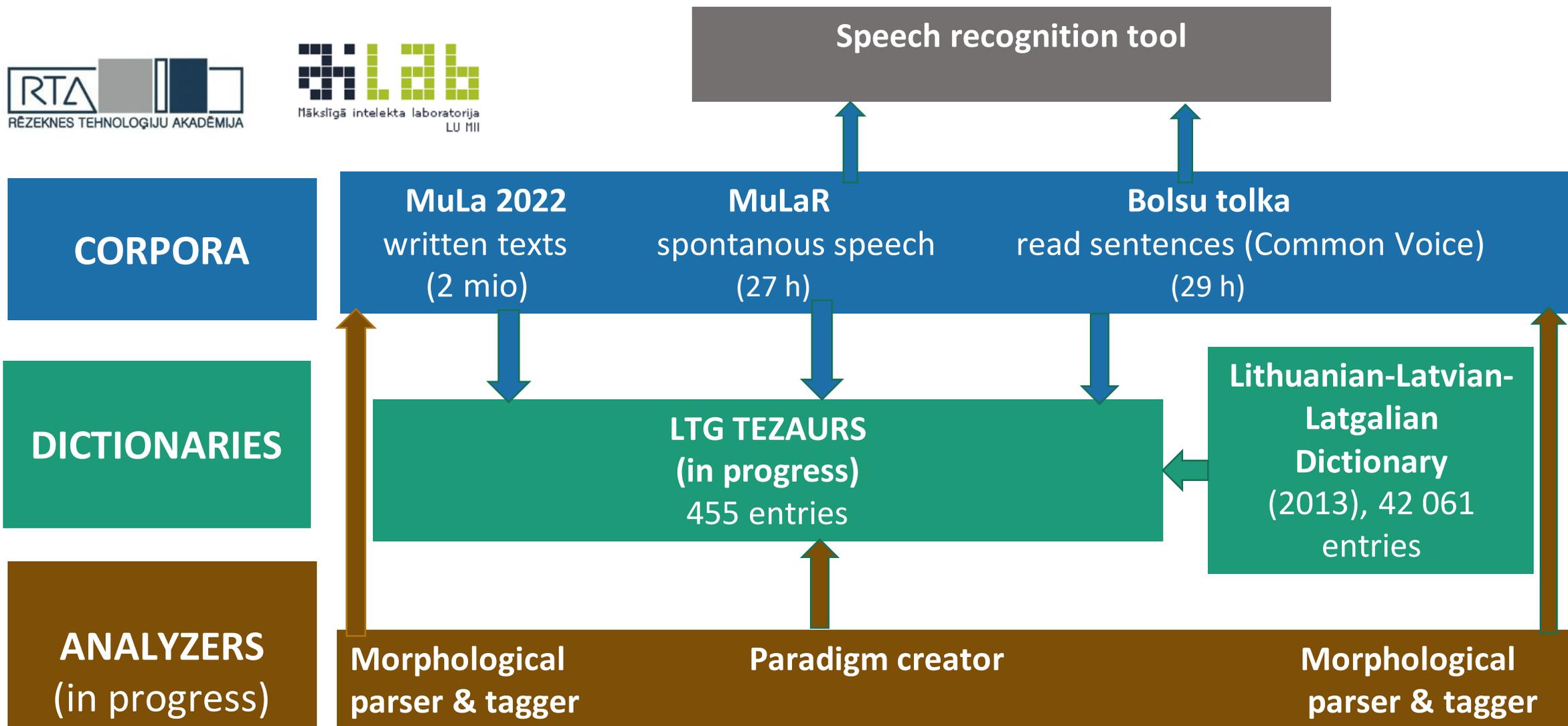- Extinct

▶ Details

Digital Language Support ⭐

Ascending

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2024. Ethnologue: Languages of the World. Twenty-seventh edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

2 Digital language vitality indicators

- **Digital capacity**: evidence of connectivity; digital literacy; internet penetration or digital population size; character/script encoding; <u>availability of language resources</u>.

- **Digital presence and use**: use for e-communication; use on social media; availability of Internet media; Wikipedia.

- **Digital performance**: available Internet services; localised social networks; localised software; machine translation tools/services; dedicated Internet top-level domain.

Siora, C., Quochi, V., Russo, I., Gurruthaga, A., Ceberio, K. 2017. *A Digital Language Vitality Scale and Indicators*. https://www.academia.edu/34966423/A_Digital_Language_Vitality_Scale_and_Indicators

3 Latgalian language digital resources and tools

Speech recognition tool

CORPORA

MuLa 2022
written texts
(2 mio)

MuLaR
spontanous speech
(27 h)

Bolsu tolka
read sentences (Common Voice)
(29 h)

DICTIONARIES

**LTG TEZAURS
(in progress)**
455 entries

Lithuanian-Latvian-Latgalian Dictionary
(2013), 42 061 entries

**ANALYZERS
(in progress)**

**Morphological
parser & tagger**

Paradigm creator

**Morphological
parser & tagger**

Nau, N. (2024) The Latgalian TEZAURS and the challenges of linguistic diversity. Poster presentation.Baltic DH Forum.
https://drive.google.com/file/d/14Rxwj3MOZzuNDgT52MxSBodcnVpHXD-M/view

CONCORDANCE

simple **Kembridž*** • 3
1.07 per million tokens • 0.00011%

	Details	Left context	KWIC	Right context
1	ⓘ doc#150	doeiropīšu saleidzynojūšōs volūdnīceibas doktorands J. Lēlis	**Kembridža**	, Masačūsetsas pavaļstī
2	ⓘ doc#207	:montōta raksteišona ar zaltu, drūši vīn tys ir jaunōks.</s><s>	**Kembridžas**	universitates eksperts Pit
3	ⓘ doc#2139	jōs lingvistikas studijom un 1961. godā Harvarda Universitātē	**Kembridžā**	sajam doktora gradu, aizs

c#150   doeiropīšu saleidzynojūšōs volūdnīceibas doktorands J. Lēlis **Kembridža** , Masačūsetsas pavaļstī un agronoms, valsts univ

Display and count metadata

Select the metadata to be displayed in the concordance. Click 📊 to calculate statistics.

☐ Display above lines ?   ✓ Shorten to 15 ⤢ characters

sitates indoeiropīšu saleidzynojūšōs volūdnīceibas doktorands J. Lēlis Kembridža , Masačūsetsas pavaļstī un ag

type to search 🔍 ⫦ (1)

☐	Token number	269292
✓	**Document number**	**150**
☐	Autors	**Teirumnīks Francis**
☐	Gads	**2001**
☐	Izdevums	**Acta Latgalica 11**
☐	Teksta nosaukums	**Latgaļu sabīdriskō aktivitate trymdā**
☐	Teksta veids	**Literārie teksti-Dokumentālā literatūra;Literārie teksti**
☐	Žanrs	**Atmiņas**

pmāram 1958. godā uzsōka sadarbeibu šaidi četri latgaļu pētnīki: Belarminas koledžas svešvolūdu

</s><s> Leonards Latkovskis, Sr., Luisvila, Kentuki pavaļstī, literaturvēsturnīks un volūdnīks Mikelis Bukšs

rdas Universitates indoeiropīšu saleidzynojūšōs volūdnīceibas doktorands J. Lēlis **Kembridža** ,

un agronoms, valsts universitates bibliotekars Bonifacijs Briška Olbani, Ņujorkas pavaļstī. </s><s> Taidā

kūduls, kas sōka veidōt organizaciju, kuru nūsauce par Latgaļu Pētnīceibas institutu. </s><s> Latgaļu

ciatori beja taidūs īskotūs, ka latvīšu tautas daļai Latgolā, kurus

Language (State)	Status	The name of the Corpus, available from ... (year)	The Size of the Corpus (words)
Basque (Spain)	Official Regional Language	ETC; 2013	355 million
Asturian (Spain)	No Status	ESLEMA; 2008	Not finished
Low Sorbian (Germany)	Minority Language	DTK; 2010	23 mil
Welsh (UK)	Official Regional Language	CorCenCC; 2020	11 mil
Northern Sámi (Norway)	Indigenous (autochthon) Language	SIKOR; 2015	8,9 mil
Silesian (Poland)	No Status	KŚM; 2018	2 mil
Latgalian (Latvia)	Historic Variant of the Latvian Language	MuLa; 2022	2 mil

Martena, S., Nau, N., Briška, A. (2022). Latgaliešu valodas korpuss citu Eiropas mazāk lietoto valodu kontekstā: korpusa raksturojums, lietojums un potenciālā iespējošana, Letonica No 47, 208-224

Latgalīšu volūdys tezaurs
Latgaliešu valodas tēzaurs
455 šķirkļi

[LTG T]:

vārds vai frāze

ltg.tezaurs.lv

Pameklēt plašāk

buļbe

buļbe 🔊 sieviešu dzimtes 5. deklinācijas lietvārds ∧ Locīšana

	Vsk.	Dsk.
Nom.	buļbe	buļbis (ieteicams), buļbes
Ģen.	buļbis (ieteicams), buļbes	buļbu
Dat.	buļbei	buļbem
Akuz.	buļbi	buļbis (ieteicams), buļbes
Lok.	buļbē (ieteicams), buļbie	buļbēs

[LLLD 2013]

buļbeite sieviešu dzimtes 5. deklinācijas lietvārds; deminutīvs ∨ Locīšana

buļba 🔊 sieviešu dzimtes 4. deklinācijas lietvārds ∨ Locīšana [Strods 1933, Bērzkalns 2007]

buļve 🔊 sieviešu dzimtes 5. deklinācijas lietvārds ∨ Locīšana [Let, LLLD 2013]

buļva 🔊 sieviešu dzimtes 4. deklinācijas lietvārds ∨ Locīšana [Bērzkalns 2007]

buļvs 🔊 vīriešu dzimtes 2. deklinācijas lietvārds; Kalupes izloksne ∨ Locīšana [KiV, ADT 1983]

1. Nakteņu sugas kultūraugs; kartupelis.
   ∨ Piemēri   *Apnyka, nūej zuoli papļauņ, buļbis izvogoj, iz mežu moļkys pagrīz.*
   ∨ Saistītās nozīmes   *guļba, kartupeļs, uļbiks.*
   ∨ Stabili vārdu savienojumi   *Trokais buļvs.*

2. Uzturā lietojams šī auga pazemes bumbulis.
   ∨ Piemēri   *Itymūs apstuokļūs tys beja daudz, jau nedeļu jī ar Fraņci naēde nikuo cyta kai buļbis, i to ar māru – pa buļbei obejim vakariņuos, dīniškuos Marei – buļbe, Fraņčam – div.*
   ∨ Saistītās nozīmes   *guļba, kartupeļs, uļbiks.*

NORMATĪVAIS KOMENTĀRS: "Vyspuor buļbu nūsaukšonai Latgolā ir ir ap 15 slaviskūs variantu. Vysus par vīnu pajimt nu nikai nav īspiejams! Raksteisim par tū nasatīpdami - buļbe voi tupiņš! (Leikuma, L. "Gruomota školuotuojim". Lielvārde: Lielvārds, 1993, 59.)

3 Latgalian language digital resources and tools

- Latgalian Verbs https://latvianapps.com

- Latgalian Spelling Tool: https://va.hugo.lv/proofingen

- Pronunciation App: https://fryske-akademy.nl/fa-apps/trainer/

- Google Translate

3 Development of digital resources and tools of lesser-used languages (Latgalian case)

Opportunities and strengths

- long-term, stable cooperation with other European lesser-used language research centers,

- fundraising, new projects, crowdsourcing,

- adherence to FAIR data principles and inter-institutional cooperation in data management,

- raising awareness of the value of Latgalian and promoting it in the local community, especially among young people.

Challenges

- **Linguistic:**
- insufficient data (texts, audio recordings),
- subjectivity, e.g., in manual transcription of audio recordings,
- variation as a challenge for morphological tagging and paradigm creation,
- **extralinguistic:**
- lack of financial and human resources for a planned and sustainable digitization;
- lack of interest / information among potential users,
- lack of digital literacy.

4 (Digital) literacy: Latgalian in Education

4 Developing digital literacy

- **Study courses for Latvian language teachers:** *Language Technology, Language Corpora in Education*.

- **Seminars for Latvian and Latgalian language teachers.**

- **Teaching materials:**

- **Educational computer games:**
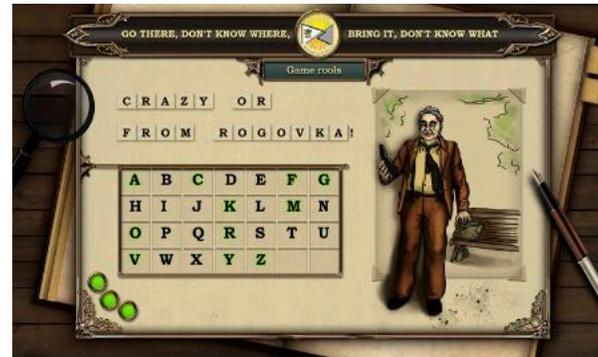
Games | futureofmuseums.eu – travelling through the towns and villages of Latgale, to identify the most notable distinguishing marks of these places, to learn some Latgalian proverbs.

https://iepazisimies.rta.lv/ - a game about important figures in the cultural history of Latgale.

https://www.tandfonline.com/doi/full/10.1080/01434632.2024.2418475

uage & Literature ▶ Journal of Multilingual and Multicultural Development ▶ List of Issues ▶ Volume 46, Issue 3 ▶ Literacy (development) in regio

Journal homepage

≡ ◀) Listen ▶

Literacy (development) in regional collateral languages of Europe – introduction to the special issue of the *Journal of Multilingual and Multicultural Development*

Heiko F. Marten ✉ & Sanita Martena

Developing corpus literacy: A perspective of Latgalian language and cultural studies

Angelika Juško-Štekele & Antra Kļavinska

When the parish turns into a classroom: teaching pluriliteracies in regional collateral languages through place-based education (the case of Latgalian)
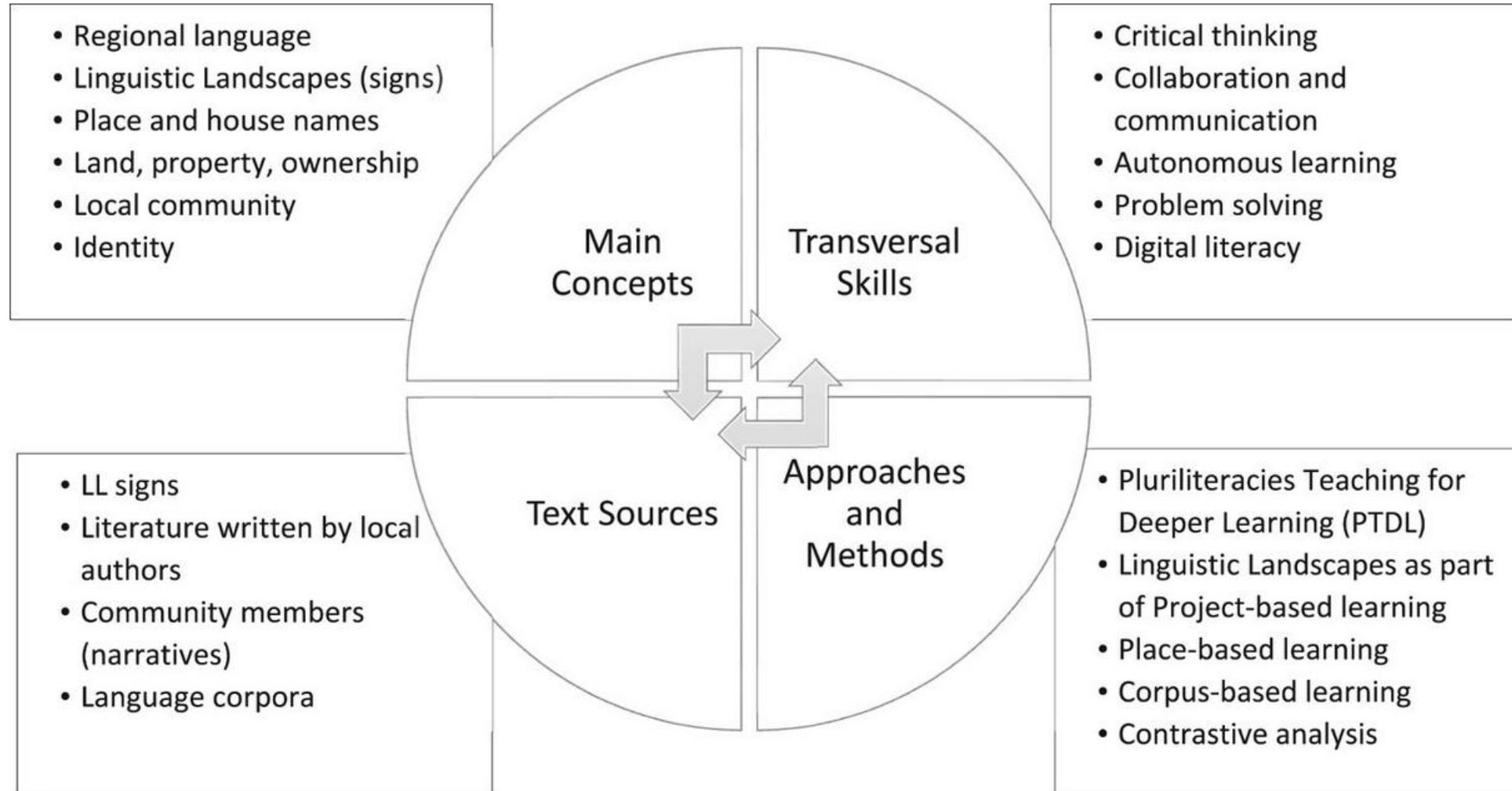
Sanita Martena & Heiko F. Marten

Social values of an added literacy: the case of Latgalian

Nicole Nau & Tomasz Wicherkiewicz

This introduction provides the theoretical context of the Special Issue on Literacy Development in Regional Collateral Languages of Europe and an overview of the papers in the collection. For this end, it first discusses contemporary perceptions of the terms 'literacy' and 'literacies'. These go far beyond traditional understandings, even though expressing one's thoughts through writing texts and meaning-making through reading are still at the core of literacies. Then, readers are introduced to the concept and the ideological contexts of 'regional collateral languages': These are varieties which are, on the one hand, structurally and ideologically closely related, or 'collateral', to major languages, often the main languages of nation states. On the other hand, they are deeply rooted in their regions; the paper therefore argues that they deserve specific attention along the lines of other efforts to maintain and develop autochthonous varieties. Finally, the introduction summarises the papers which focus on Scots in Scotland, Low German in Germany, Csángó in Romania, Silesian, Masurian, and Kashubian in Poland, Võro in Estonian, and Latgalian in Latvia, before identifying common denominators of the case studies and possible topics for future research.

Figure 6 of 6

Figure 6. A set of components in the teaching of regional language literacy through place-based learning.

- Regional language
- Linguistic Landscapes (signs)
- Place and house names
- Land, property, ownership
- Local community
- Identity

Main Concepts

Transversal Skills

- Critical thinking
- Collaboration and communication
- Autonomous learning
- Problem solving
- Digital literacy

- LL signs
- Literature written by local authors
- Community members (narratives)
- Language corpora

Text Sources

Approaches and Methods

- Pluriliteracies Teaching for Deeper Learning (PTDL)
- Linguistic Landscapes as part of Project-based learning
- Place-based learning
- Corpus-based learning
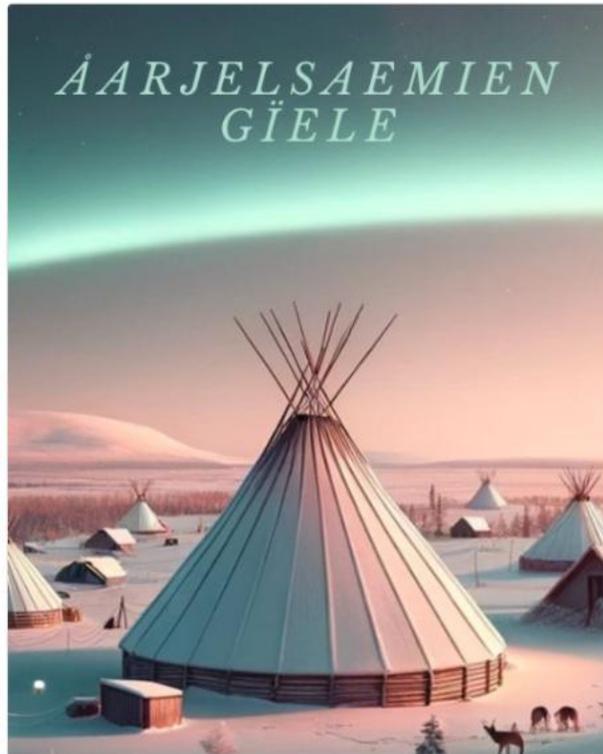- Contrastive analysis

When the parish turns into a classroom: teaching pluriliteracies in regional collateral languages through place-based education (the case of Latgalian)

Sanita Martena & Heiko F. Marten

Education:
Latgalian in the Context of other European Lesser-used Languages

ÅARJELSAEMIEN GÏELE

SOUTH SAAMI

FRYSK

WEST FRISIAN

MIRANDÉS

MIRANDESE

Latgalīšu volūda

LATGALIAN

The Latgalian language, predominantly spoken in Latgale, Latvia, has faced oppression in the past. Over

5 Conclusions (1)

1) In recent years, there has been a greater interest in regional variation and diversity of the Latvian language as a whole (in particular in the field of culture and education)

2) Latvian (lvs) and Latgalian (ltg) are so different that each requires its own set of digital resources and tools, but at the same time linguistically similar enough that tools for Latgalian may be partly based on tools developed for Latvian. In such way collateral languages live side by side and «cooperate» with each other not just in physical environment but also in digital space

3) In the last decade, the Latgalian language has seen progress in all Digital Language Vitality indicators (digital capacity, digital presence and use, digital performance).

4) In the repository of the CLARIN Centre of Latvian Language Resources, 6 resources for Latgalian (corpora and dictionaries) are registered, the development of various tools and analyzers (in cooperation with Latvian and foreign partners) confirms that Latgalian digital capacity is moving from a basic level towards an intermediate one.

5 Conclusions (2)

5) Cooperation with other European colleagues working on lesser-used languages allows to notice a lot of similarities in language (digital) vitality processes, including the role of technologies for documenting, maintaining and developing language

6) Among the younger generation, the number of language users is decreasing, therefore it is crucial to offer modern and creative contents and innovative language teaching approaches to fulfill students' interests and expectations

7) There is a correlation between the literacy of the speakers of a language and the development of language technologies. (Digital) literacy development is therefore particularly significant for the transgenerational transmission and the vitality of lesser-used languages.

Paļdis par viereibu!
Paldies par uzmanību!
Thank you for your attention!

Antra.Klavinska@rtu.lv
Sanita.Martena@rtu.lv