Developing Corpus Literacy: a Perspective of Latgalian Language and Culture Studies

Angelika Juško-Štekele, Antra Kļavinska

Workshop «Documentation for Education of Lesser Used Languages»
Rēzekne, May 2nd, 2023

Overview

- Corpora in the context of documenting lesser used languages.
- Modern Latgalian Language Corpus and Latgalian Speech Corpus - are they used in the language / culture teaching / learning process?
- Teacher education and corpus literacy.
- Some practical examples:
- Getting to know proper nouns using corpus linguistics analysis.
- Understanding and diagnosis of functions of folklore in the national corpus collection.
- Summary.

Corpora of lesser used languages, some examples (Martena, Briška, Naua 2022, 211)

Language (state)	Status	Corpus, available from... (year)	Corpus size (tokens)
Basque (Spain)	official language in the region	ETC; 2013	355 million
Asturian (Spain)	no status	ESLEMA; 2008	not completed
Lower Sorbian (Germany)	minority language	DTK; 2010	23 million
Welsh (UK)	official language in the region	CorCenCC; 2020	11 million
Nordhern Sámi (Norway)	autochthonous language	SIKOR; 2015	8,9 million
Silesian (Poland)	no status	KŚM; 2018	2 million
Latgalian	historic variant of the Latvian language	MuLa; 2012 MuLa; 2022 MuLaR	1 million 2 million not completed

Modern Latgalian Language Corpus (MuLa2012, [MuLa 2022](#)) **and Latgalian Speech Corpus** ([MuLaR](#))

Survey on the use of the MuLa corpus (2021)
(Martena, Briška, Naua 2022)

214 respondents in total

- Know, use
- Know, don't use
- Don't know, don't use

15; 7%
43; 20%
156; 73%
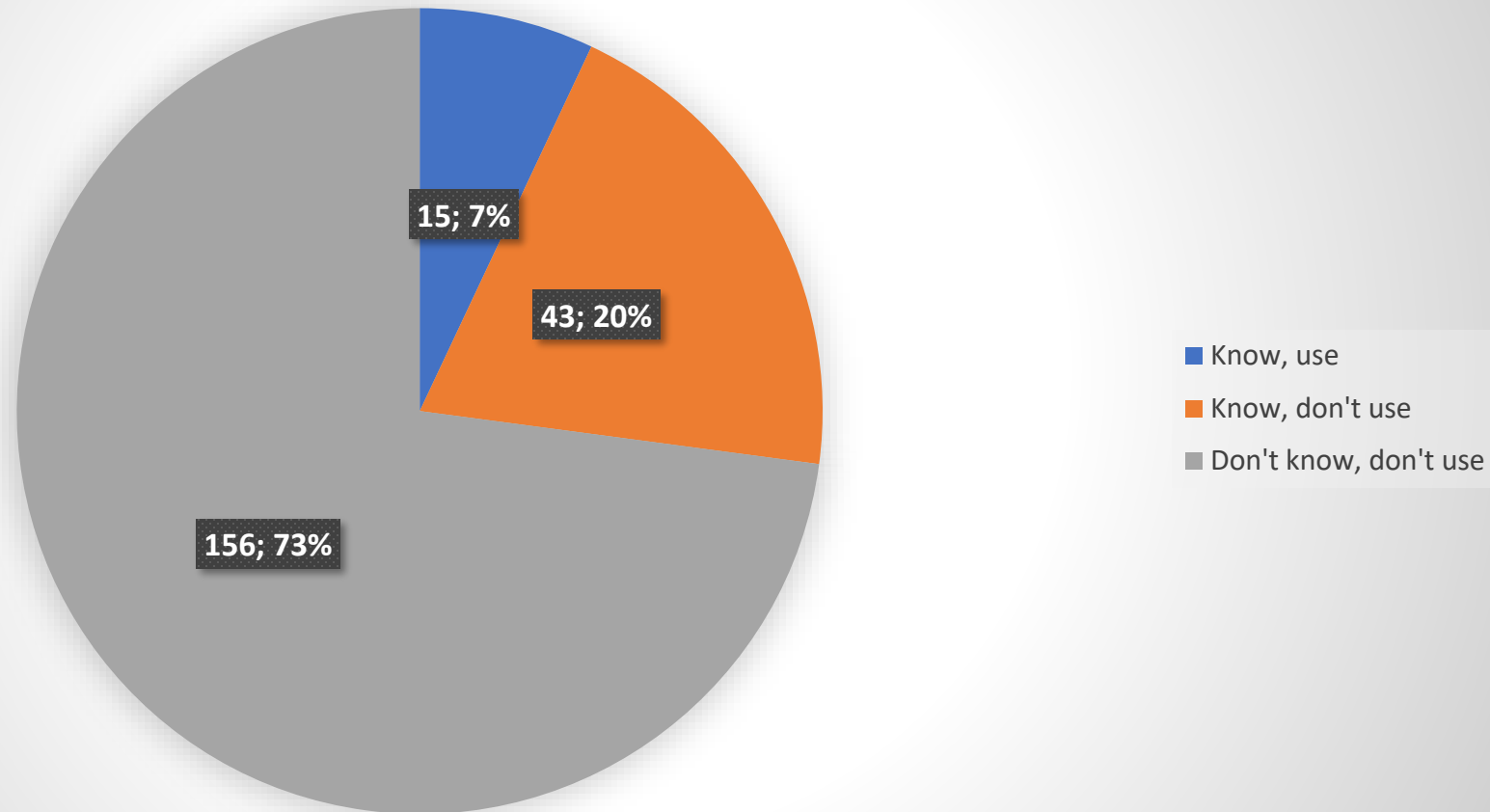
Subcomponents of corpus literacy (Mukherjee 2006, 14)

- understanding basic concepts of corpus linguistics (What is a corpus and what types of corpora are available? What can you do— and cannot do — with a corpus?);

- searching corpora and analysing corpus data by means of corpus software tools, e.g. concordancers (What is corpus software and how can it be used to search a corpus? How can corpus output be analysed?);

- interpreting corpus data (How may general trends in language use/ change be extrapolated from corpus data?);

- using corpus output to generate teaching material and activities.

Educating Language Teachers

- [Video lectures](#) promoting the opportunities offered by the MuLa2022 and MuLaR corpora in the Latgalian language teaching / learning process.

- Seminars for teachers; lectures for students of the RTA study program «Latvian language and literature teacher».

Topic "Our Names" (grades 4 – 6) (Andrejeva *et al.* (2020). The curriculum for the subject "Latgalian written language for grades 1 – 9" 2020, 42)

- **Communicative competence**: Is learning to obtain information from visually and auditory presentation of text and to use it in spoken communication. Discusses the use of personal name variants in everyday life in different communicative situations.

- **Linguistic competence**: Is learning to differentiate between older and modern first names, their meaning, and the full and shortened versions of first names.

- **Sociocultural competence**: Understands that a personal name is a mark of a person's identity and can provide information about the person's belonging to a specific culture.

How do you understand this joke?

- **Vasali! Īpasa**zeisim / ī**sapa**zeisim?

- Do you know people who are called *Pīters, Juoņs, Muora, Klāvs*?

- What names were popular in Latgale before and what are popular now?

- How do you address each other in the family, among friends?

Searching examples by means of corpus software tools, e.g. concordancers, generating teaching materials (LVK 2022 – MuLa 2022)

LVK 2022

MuLa 2022

Other digital tools for generating activities, e.g.,
The 100 most popular newborn names in Latvia
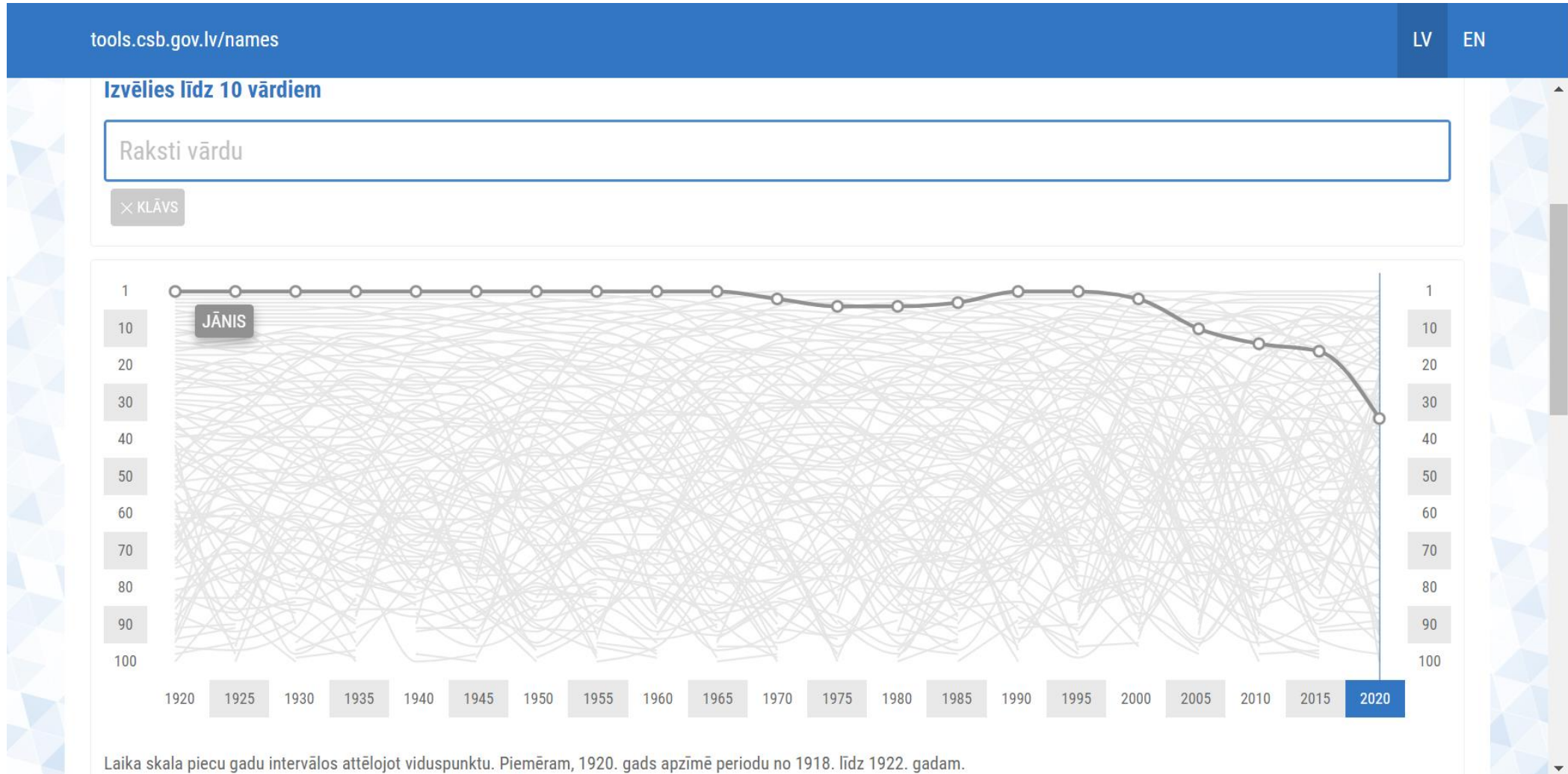
Variation of personal names: MuLaR examples
Query: *Jezup-*

- *nascē ka tur ir **Jāzeps** voi **Jezups** voi sacējom vīnkuorši **Juze*** (M405-1966)

- *i tagad mozais myusu **Jezupeņš** kai mes jū saukojam.*

- *myusu kapitālais seniors kurš aizguoja i jam beja īsauka **Plisa Jezups**.*

- *tegu tegu **Teguškys Jezups** muns tāvs ite jau munu dzedu tai suokuoja.*

- *a tī kolnā atkal dzeivova Kudiņš **Jezups** i jam beja ryži moti. tai i sauce **Ryžais Juze*** (F423-1950)

Topic "Latgalian Cities and Towns and Municipalities" (grades 4 – 6). (Andrejeva et al., 2020, 65)

- **Communicative competence**: Working with a partner or in a group, gets acquainted with one settlement in Latgale and using different information sources prepares a virtual tour for classmates.

- **Linguistic competence**: Knows how to write the names of settlements and their residents in Latgalian.

- **Sociocultural competence**: Is aware that toponyms are an important part of the linguistic landscape and carriers of historical and cultural information, a cultural treasure.

Create the city sign / city slogan in Latgalian!

- How are place names pronounced in Latgalian (Latgalian dialects)?

- How to write a settlement name (place name form) in Latgalian?

- How to name the residents of the settlement?

Karteņa: jaunuo Bolvu nūvoda rūbeža zeime, foto nu Bolvu nūvoda pošvaļdeibys publicitatis materialu

Preiļu nūvoda zeime, foto Amanda Anusāne/portals lakuga.lv

How to pronounce / to write the name of a settlement in Latgalian? (frequency analysis)

- [MuLaR](#) (28.04.2023.)
- **Rēzekn-** 59; **Riezekn-** 44

- *jius sokit Rēzekne taipat kai mes. rēzeknīši soka Riezekne maņ bie puorsteigums. es arī Riezeknī vuiciejūs Riezeknī vo ha ha* (F419-1962, Aglona)

- [MuLa22](#)
- **Rēzekn-** 2,558; **riezekn- /Riezekn-** 39

- *Tautys skaiteišona – **Rēzeknē** dzeivoj 12 tyukstūšu latgaliski runojūšu cylvāku, Reigā –29 tyukstūšys.* (A12, 2013)

- *Lai zineigajim, apsvīdeigajim, uzjiemeigajim i draudzeigajim **riezeknīšim** vēļ daudz jaunu ideju laikim iz prīšku i lai lūbās* (Leikuma 2013)

- Aktivu i radūšu pavasari jyusim vielej LgSC **Riezeknis** biroja koordinatori! (lgsc.lv, 2010)

Check which forms of place name are found in the corpus? Do they comply with contemporary spelling rules?

- MuLa 2022

Latgalian language spelling tool

What are the associations with the place name *Rēzekne?* (collocation analysis) MuLa 2022

Understanding and diagnosis of functions of folklore in the national corpus collection

Types of definitions

- Descriptive: definitions one finds in a dictionary; it specifies what people have meant in the past in using a particular term.

- Stipulative: stipulate what a particular term will mean in a particular context or for a particular purpose.

What is folklore? Some definitions

- «Folklore is the traditions, customs, and superstitions of the **uncultured classes**". (W. Toms) 1846
- "Folklore is the artistic communication in small groups of the experiences, ideals, emotions, and attitudes of a culture." - Alan Dundes
- "Folklore is the traditional art, literature, knowledge, and practices that are disseminated largely through oral communication and behavioral example." - UNESCO
- "Folklore is a conceptual framework, an orientation, and a tool for interpreting and understanding cultural phenomena, from the most ordinary expressions of daily life to the most sophisticated expressions of art." - Regina Bendix
- "Folklore is the artistic expression of a people's culture, transmitted orally from generation to generation." - Richard Dorson
- «**Fakelore** refers to cultural expressions that are presented or marketed as traditional or authentic folklore, but are actually modern creations or adaptations that have little to no basis in actual traditional culture.»- Richard Dorson
- «Folklore is our cultural DNA». - American Folklore Society

Folkloric Competence in student esseys
korpuss.lv – parspriedumi.lv

- Methodology: SOLO (Structure of Observed Learning Outcomes) taxonomy

- Just out of 157 (or 34%) essays containing folklore-related lexemes

- 43% show the lowest level of folklore competence - dominates the naming function without further elaboration

- 29% show the ability not only to name but also to classify into simpler systems 2- 3 realia belonging to the semantic field of folklore

- 6% show the ability to reason, analyse, explain causes, integrate

The field of study of understanding of culture and self-expression in art. Learning outcomes

VSK.K.Li.2. Kultūras procesi un mākslas darbi pauž sabiedrības vērtības, to pārmantojamību un transformāciju, veido intelektuālo, estētisko un emocionālo pieredzi		
Vispārīgais apguves līmenis	**Optimālais apguves līmenis**	**Augstākais apguves līmenis**
K.V.2.7. Nosauc savas personības izaugsmē būtiskas vērtības, balstoties uz savu kā mākslas darba uztvērēja (skatītāja, klausītāja, lasītāja) un kultūras procesos iesaistīta dalībnieka pieredzi.	K.O.2.8. Analizē savu kā mākslas darba uztvērēja (skatītāja, klausītāja, lasītāja), kultūras un sabiedriskos procesos iesaistīta dalībnieka pieredzi personiski nozīmīgu vērtību sistēmas veidošanā. Sarunā ar vienaudžiem prot pamatot savu viedokli par kultūras un mākslas nozīmi katra personībā.	K.A.2.8. Komentē un analizē savu kultūras un mākslas pieredzi, formulē
	K.O.2.9. Saskata savas personības izaugsmei būtiskas multilingvālās un multikulturālās pasaules vērtības, pamato savu izvēli.	
K.V.2.8. Pamato savu kultūras piederību (vienai vai vairākām grupām, kopienām, tautām, valstīm) un identitāti ar piemēriem no ikdienas dzīves, personiskās kultūras pieredzes un piedzīvotā pašizpausmes aktivitātēs.	K.O.2.10. Komentē kultūras un mākslas praksē pieredzēto, piederības izjūtu savai dzīvesvietai un kultūrvēsturiskajam novadam, prezentē savu kultūras piederību un, diskutējot grupā, kritiski reflektē par kultūras nozīmi savas un vienaudžu identitātes veidošanā.	

K.A.2.9. In collaboration with a group of like-minded individuals, research is conducted on **how personal and community identity manifests in the cultural environment (real and digital)** as a result of creative and artistic activities. Reflection is made on the manifestations of one's identity **in the real and digital environment."**

Latvian National Corpora Collection: Korpuss.lv

Counts all the different word forms in KWIC (Key Word in Context)

LV

	Vārdforma	↓ Frequency	Frequency per million	
1	folkloras	321	26.12	•••
2	folkloristikā	34	2.77	•••
3	folklorā	28	2.28	•••
4	Folkloras	24	1.95	•••
5	folkloru	20	1.63	•••
6	folklora	20	1.63	•••
7	folkloristikas	15	1.22	•••
8	folkloristu	13	1.06	•••
9	folkloriste	8	0.65	•••
10	folklorists	6	0.49	•••
11	folklorista	6	0.49	•••
12	folklorai	6	0.49	•••
13	folkloristi	4	0.33	•••
14	folkloristes	4	0.33	•••
15	folkloristiem	3	0.24	•••
16	folkloristam	3	0.24	•••
17	folkloristei	2	0.16	•••
18	Folklorists	2	0.16	•••
19	Folkloriste	2	0.16	•••
20	folklorizētu	1	0.08	•••

LTG

	Vārdforma	↓ Frequency	Frequency per million	
1	folklorys	109	39.00	•••
2	folkloras	23	8.23	•••
3	folkloru	16	5.72	•••
4	folklora	11	3.94	•••
5	folklorā	10	3.58	•••
6	Folklorys	6	2.15	•••
7	folklorai	5	1.79	•••
8	folkloristu	3	1.07	•••
9	folklorists	2	0.72	•••
10	folkloristi	2	0.72	•••
11	FOLKLORAS	2	0.72	•••
12	folkloriziejušuos	1	0.36	•••
13	folkloristikā	1	0.36	•••
14	folkloristikys	1	0.36	•••
15	folkloristika	1	0.36	•••
16	folkloriste	1	0.36	•••
17	folkloristam	1	0.36	•••
18	folklorisks	1	0.36	•••
19	folkloriskais	1	0.36	•••
20	Folklorists	1	0.36	•••

Typology of the different word forms in KWIC

		Folklore	Folkloristics	Folklorist	Folklorizm/ Folklorization	
Lv	count	419	49	53	1	
	%	**80**	**9**	**10**	**~1**	
Ltg	count	73	10	3	3	
	%	**83**	**11**	**3**	**3**	

Folklore (or **traditional and popular culture**) is the totality of tradition-based creations of a cultural community, expressed by a group or individuals and recognized as reflecting the expectations of a community in so far as they reflect its cultural and social identity (Unesco)

An **academic discipline** with a scientific interest about traditional ideas, actions , or consequences.

Someone **trained** in the discipline of folklore (Unesco)

Folklorization or **heritagization** of social practices is an aspect of the infusion of folkloristic/ethnological knowledge, perspectives, and concepts into the public sphere as part of modernity's reflexivity. (V.Tr. Hafstein)

Counts all the different word forms on the first position to the left of KWIC

LV

	Vārdforma	↓ Frequency	Frequency per million	
1	latviešu	26	2.12	···
2	vēstītājas	12	0.98	···
3	mūsdienu	12	0.98	···
4	pasaules	10	0.81	···
5	novada	8	0.65	···
6	tautas	6	0.49	···
7	Latviešu	6	0.49	···
8	patstāvīga	4	0.33	···
9	pamatskolas	4	0.33	···
10	jauniešu	4	0.33	···
11	bērnu	4	0.33	···
12	Liepājas	4	0.33	···
13	Kokneses	4	0.33	···
14	vidusskolas	3	0.24	···
15	uzstāsies	3	0.24	···
16	starptautiskā	3	0.24	···
17	starptautiskais	3	0.24	···
18	somu	3	0.24	···
19	patstāvīgs	3	0.24	···
20	pagasta	3	0.24	···

LTG

	Word	↓ Frequency	Frequency per million	
1	latgalīšu	9	6.79	···
2	latvīšu	6	4.53	···
3	Latgolys	6	4.53	···
4	nūvoda	5	3.77	···
5	jaunīšu	5	3.77	···
6	Latgaļu	5	3.77	···
7	sovu	4	3.02	···
8	Latvīšu	4	3.02	···
9	Latgolas	4	3.02	···
10	Goda	4	3.02	···
11	vītejuos	2	1.51	···
12	tradicionaluos	2	1.51	···
13	tautys	2	1.51	···
14	tautas	2	1.51	···
15	pīsadaleis	2	1.51	···
16	pīraksteitajūs	2	1.51	···
17	noma	2	1.51	···
18	myusu	2	1.51	···
19	latgaļu	2	1.51	···
20	instituta	2	1.51	···

Typology all the different word forms on the first position to the left of KWIC

LV
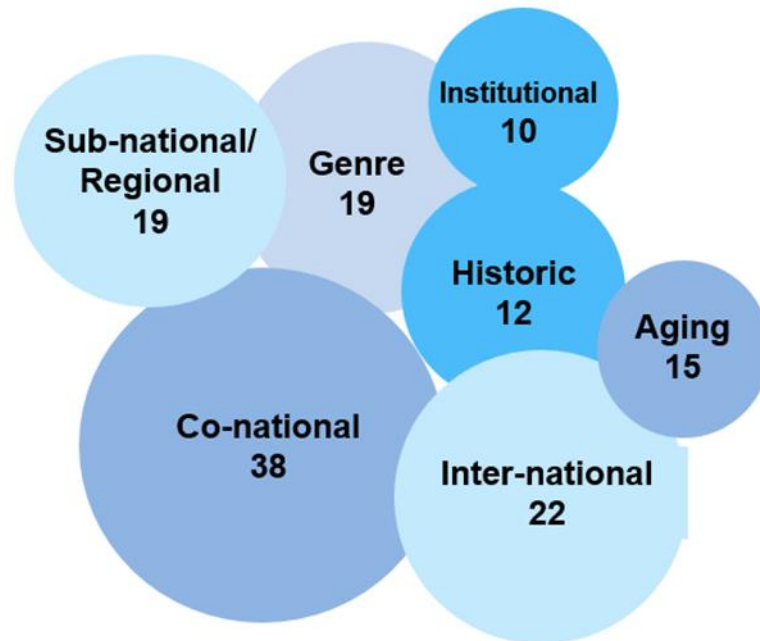
LTG
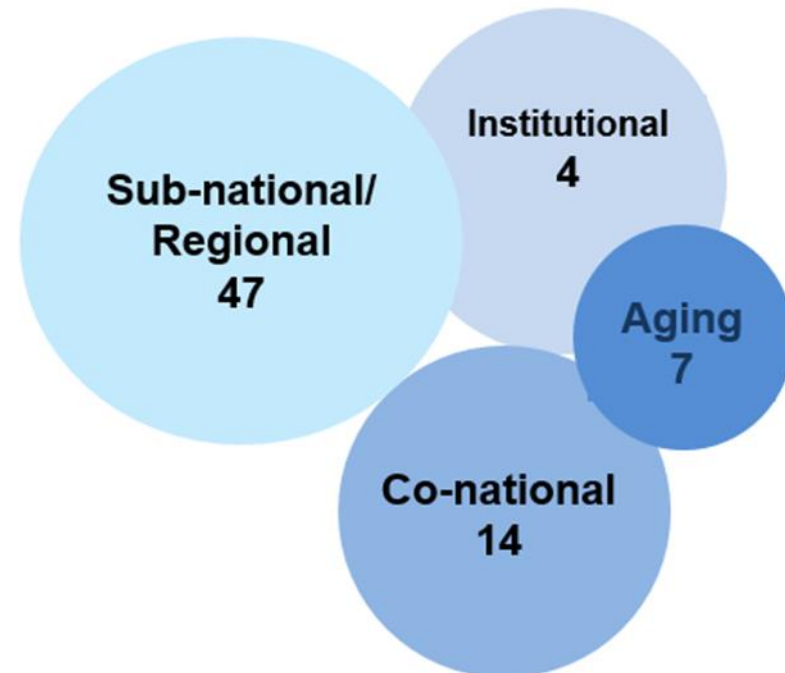
Counts all the different word forms on the first position to the right of KWIC.

LV

	Vārdforma	↓ Frequency	Frequency per million	
1	kopa	44	3.58	•••
2	kopas	40	3.25	•••
3	kopu	16	1.30	•••
4	žanrs	7	0.57	•••
5	festivāls	7	0.57	•••
6	ansamblis	7	0.57	•••
7	izpētes	5	0.41	•••
8	žanri	4	0.33	•••
9	žanra	4	0.33	•••
10	krātuves	4	0.33	•••
11	kopā	4	0.33	•••
12	centra	4	0.33	•••
13	žanru	3	0.24	•••
14	vienības	3	0.24	•••
15	kopām	3	0.24	•••
16	festivāla	3	0.24	•••
17	centrā	3	0.24	•••
18	žanriem	2	0.16	•••
19	svētki	2	0.16	•••
20	pētnieks	2	0.16	•••

LTG

	Vārdforma	Cooccurrences ?	Candidates ?	T-score	MI	↓ LogDice	
1	kūpys	80	126	8.94	11.06	11.42	•••
2	kūpa	77	124	8.77	11.03	11.37	•••
3	kūpu	42	69	6.48	11.00	10.58	•••
4	ekspediceju	11	37	3.31	9.96	8.70	•••
5	draugu	11	190	3.30	7.60	8.46	•••
6	kruotivis	9	11	3.00	11.42	8.45	•••
7	kūpai	8	12	2.83	11.13	8.28	•••
8	kūpas	7	17	2.64	10.43	8.08	•••
9	ansamblis	7	75	2.64	8.29	7.98	•••
10	kūpom	6	10	2.45	10.98	7.87	•••

Typology all the different word forms on the first position to the right of KWIC.
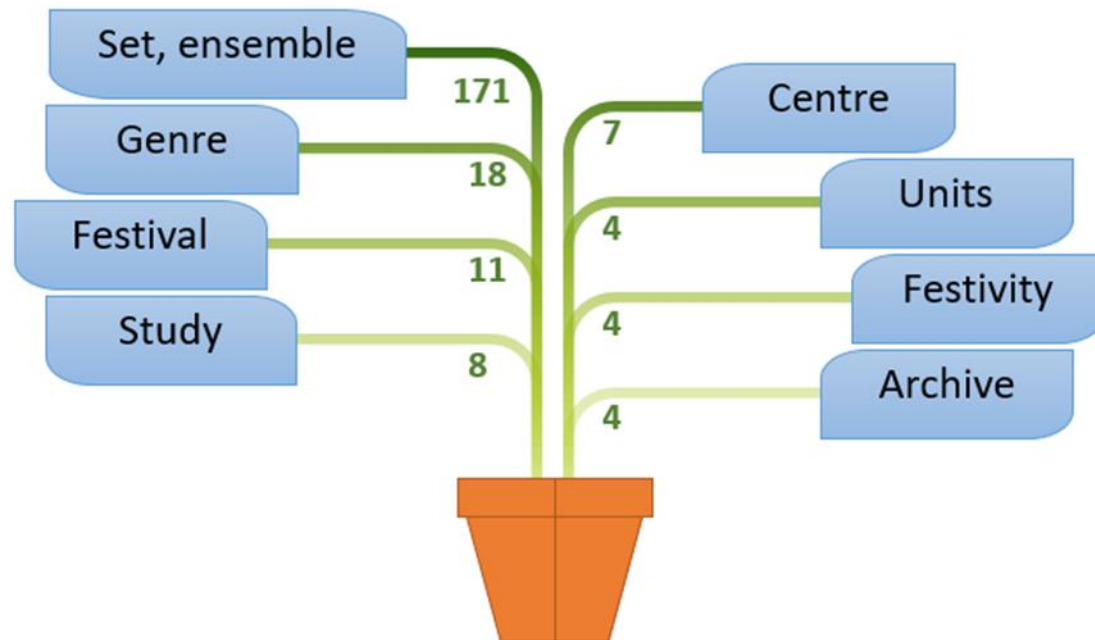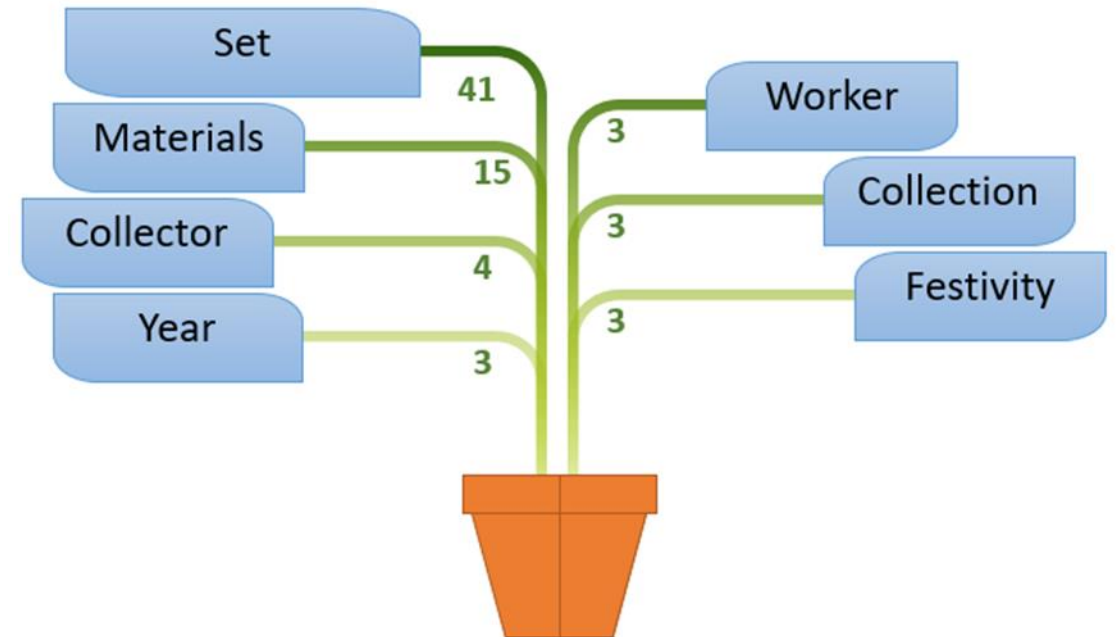
The most popular folklore ansambles in MuLa2022

	Vārdforma	Cooccurrences ?	Candidates ?	T-score	MI	↓ LogDice	
1	Upīte	10	41	3.16	9.82	8.69	•••
2	Ceidari	8	10	2.83	11.54	8.43	•••
3	kūpa	7	124	2.63	7.71	8.03	•••
4	deju	7	163	2.63	7.32	7.97	•••
5	Dyrbyni	5	11	2.23	10.72	7.75	•••
6	Ceiruleits	5	13	2.23	10.48	7.74	•••
7	materialu	6	214	2.43	6.70	7.67	•••
8	„	69	11,386	7.94	4.49	7.54	•••
9	"	56	9,309	7.15	4.48	7.51	•••
10	literatūrā	5	185	2.21	6.65	7.45	•••

Summary

- MuLa corpus shows a high intensity of use of the term of folklore compared to LK2018
- MuLa corpus allows to view not only folklore, but also manifestations of folklore in Latgale culture space
- MuLa corpus convincingly reveals diverse connotations of understanding of Latgalian folklore.
- It is in MuLa's texts that a greater role of personality can be seen in raising awareness of folklore.
- The data of the digital corpora is a grateful material for the diagnosis of folklore comprehension within the framework of the chronology of the texts included in the corps.
- The understanding of folklore in the corpora is mostly linked to the activities of folklore ensembles.
- Spelling and pronunciation of Latgalian proper nouns may be activated by comparing the data of one's native dialect, Latgalian written language and Latvian standard language, thus also activating the comparative aspect of using different corpora materials.
- It is necessary to popularize the digital resources of the Latgalian language and to promote digital literacy (corpus literacy).

References

- Andrejeva, A., Dundure, V., Vulāne, A., & Žulina, V. (2020). *Latgaliešu rakstu valoda 1. – 9. klasei. Mācību priekšmeta programmas paraugs*. Retrieved from : https://www.visc.gov.lv/lv/media/11407/download

- Hague, C., & Payton, S. (2010). Digital literacy across the curriculum. FutureLab. Retrieved from http://www.nfer.ac.uk/publications/FUTL06/FUTL06.pdf

- Heather, Julian & Helt, Marie. 2012. Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education* 20(4): 415–440.

- Martena, S., Briška, A., Naua, N. (2022). Latgaliešu valodas korpuss citu Eiropas mazāk lietoto valodu kontekstā: korpusa raksturojums, lietojums un potenciālā iespējošana. Letonica, (47), 208-2025. Letonica_47_MAKETS_DRUKA_WEB.indd (lulfmi.lv)

- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art – and beyond. In Braun, S., Kohn, K. & Mukherjee J. (eds). *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods,* Frankfurt: Peter Lang, 5–24.

References II

- Olajide, S.B. (2010). Folklore and Culture as Literacy Resources for National Emancipation. International Education Studies, 3, 200-205.

- Ostrak, A., Voolaid, P. (2017). Variation in language, literature, folklore, and music. Programme and abstracts. ELM Scholarly Press.

- Juško-Štekele, A. (2021). Folklore competence in the representation of cultural identity in secondary school students' essay. Education Reform: Education Content Research and Implementation Problems. DOI: https://doi.org/10.17770/er2021.2.6727

- Biggs, J. & Collis, K. (1982). Evaluating the Quality of Learning: The SOLO Taxonomy. New York, Academic Press.

- Coles, M. & Werquin, P. (2006). Education and Training Policy Qualification Systems: Bridges to Lifelong Learning. Paris, OECD.

- Jackendoff, R. (2002). Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford, Oxford University Press.

- Valdimar Tr. Hafstein. Intangible Heritage as a Festival; or, Folklorization Revisited Journal of American Folklore (2018) 131 (520): 127–149. https://doi.org/10.5406/jamerfolk.131.520.0127