

**Annotation, archiving and presentation of Kildin Saami  
lexical data with LEXUS**

Michael Rießler (Berlin) & Jacquelyn Ringersma (Nijmegen)



### **Annotation, archiving and presentation of Kildin Saami lexical data with LEXUS**

Technical and lexicographic issues connected with the creation of a digital electronic dictionary for an endangered Saami language

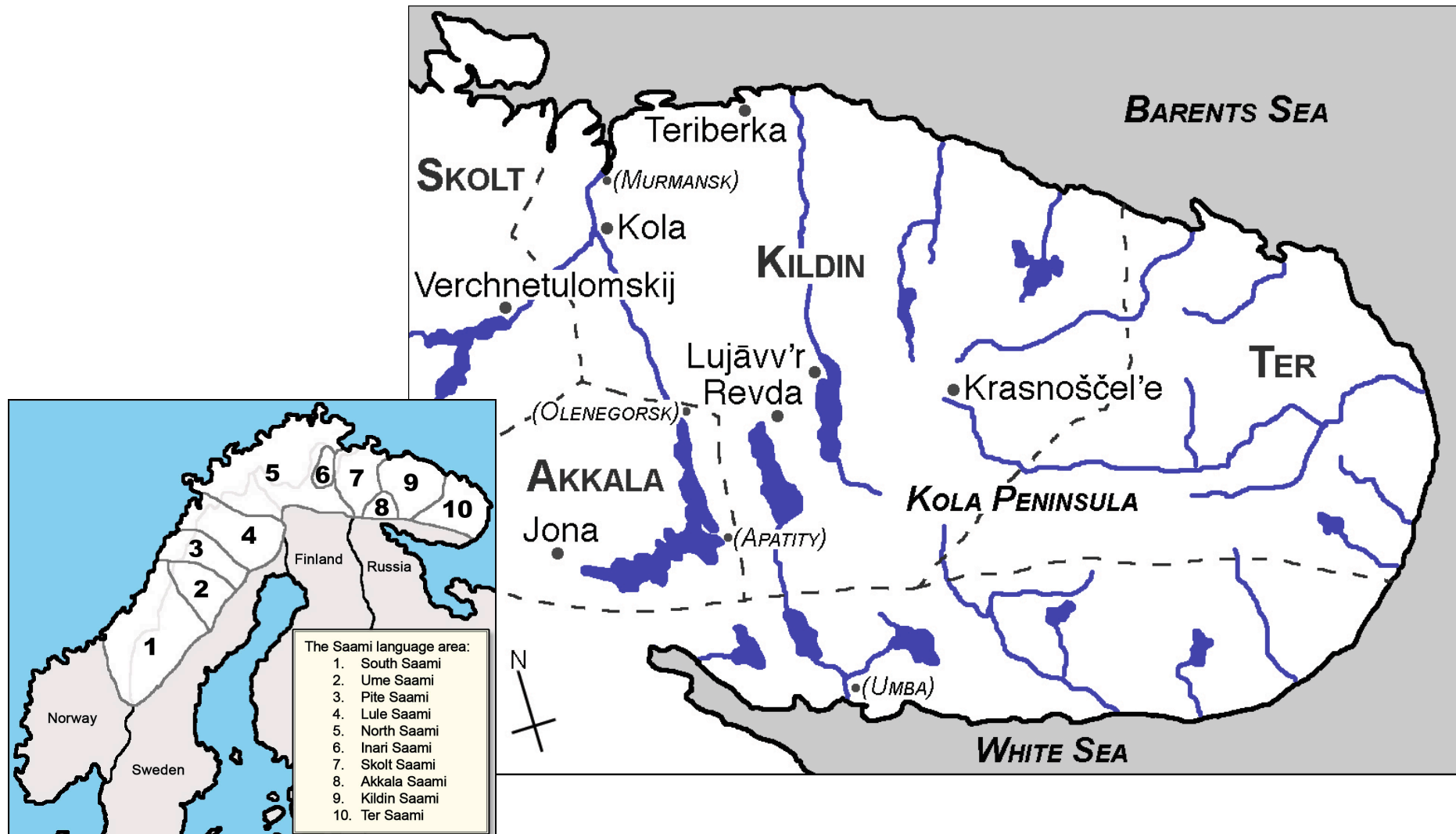
- Kola Saami Documentation Project (DoBeS)
- LEXUS and ViCoS



# Introduction: KSDP



## The Kola Saami languages



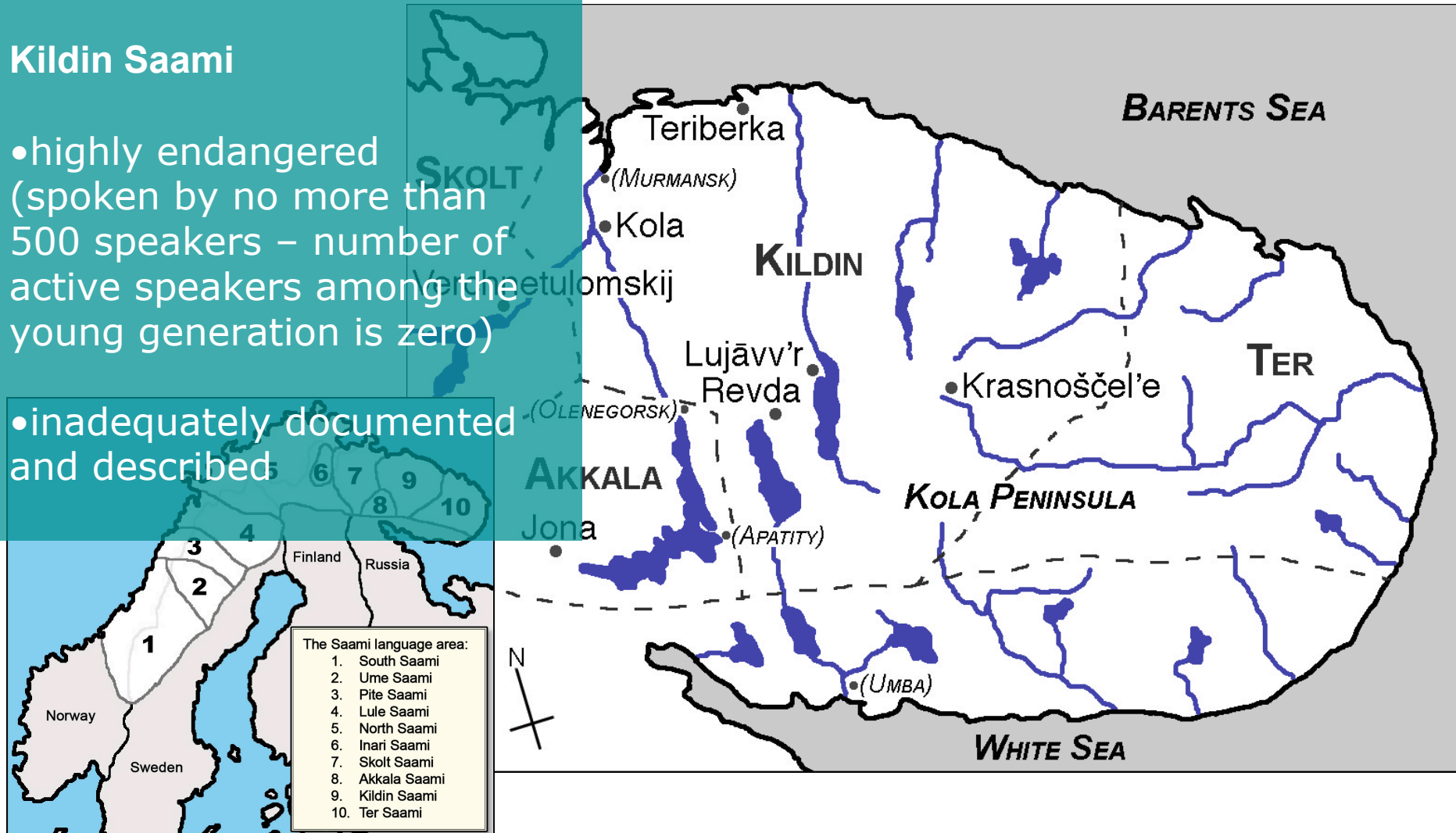


## The Kola Saami languages

### Kildin Saami

- highly endangered (spoken by no more than 500 speakers – number of active speakers among the young generation is zero)

- inadequately documented and described





# Introduction: KSDP



## The Kola Saami Documentation Project (KSDP)

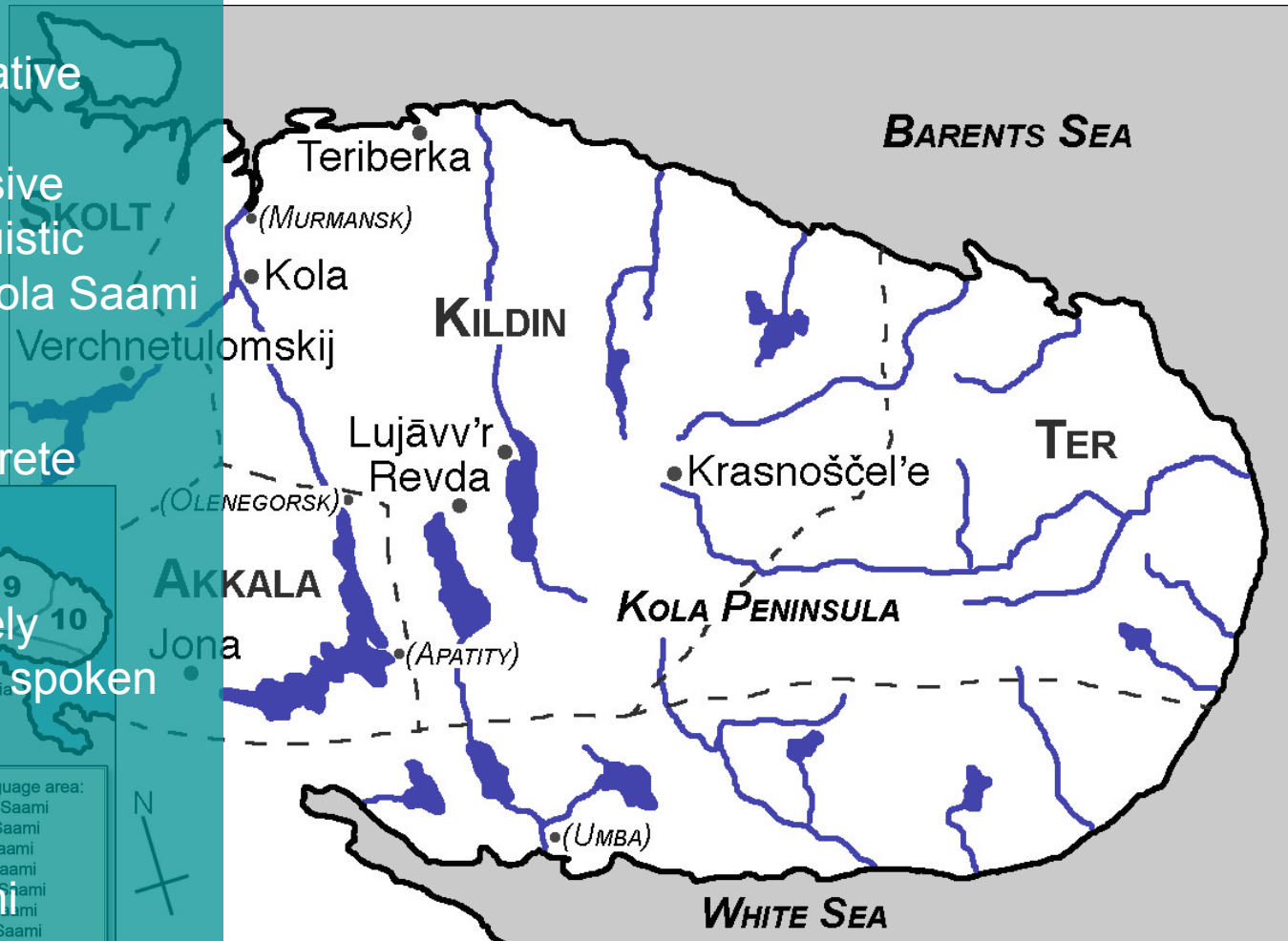
•part of the DoBeS initiative

•creating a comprehensive linguistic and sociolinguistic documentation of the Kola Saami languages

(among other concrete results):

•extensive and freely available corpus of spoken language data with annotations

•digital Kildin Saami dictionary



- The Saami language area:
1. South Saami
  2. Ume Saami
  3. Pite Saami
  4. Lule Saami
  5. North Saami
  6. Inari Saami
  7. Skolt Saami
  8. Akkala Saami
  9. Kildin Saami
  10. Ter Saami



# Introduction: DoBeS



**The DoBeS Initiative (Dokumentation bedrohter Sprachen)**

launched in 2000 by the VolkswagenFoundation (Germany)

Aim: Documentation of languages that are potentially in danger of becoming extinct





## Two pillars

### 1. Documentation (40 teams):

Recordings as primary material

Audio recordings:

present the languages as they are spoken

Video recordings:

information about the environment in which languages are spoken as well as non-verbal communication attitudes

### 2. Archiving

One central online digital archive

Formatting and encoding standards

Long term preservation and persistent accessing

Archiving and data enrichment tools (LAT suite)





## LEXUS

Web based lexicon tool

Based on the ISO recommendations for linguistic resources

**LMF** : Linguistic Markup Framework (lexicon structure)

**DCR**: Data Category Registries (concept naming)

LMF and DCR provide a modular structure for content **interoperability** between (all aspects) of lexical resources





# Why LEXUS



## Heterogeneity of lexica: **Structure**

Structure of lexicon depends on language, linguistic theory, purpose

## Heterogeneity of lexica: **Conceptual**

Large variation in linguistic concepts (attribute) and value naming

## Heterogeneity of lexica: **Format**

Large variation in formats (XML, Toolbox, Chat, Word)

---

## Data **interoperability** problem

cross lexica searches, merging, comparison

## Archive **requirements**

Representation format (XML)

One archive exploitation framework



# Why LEXUS



LEXUS is based on the ISO recommendations

It uses:

**LMF** : default structure is LMF compliant

**DCR**: Data Category Registries

LEXUS advantages:

**Format**: export in standard xml format

**Search**: one or across lexica

**Merge**: Merging of lexica (in development)

**Multi-media lexica**: Lexical encyclopedia

**Linking**: Linking to the archive resources

**Relations**: ViCoS extension



# What can LEXUS do?



- Creation of lexica from scratch, import lexica from other formats
- User defined view of the information in the lexical entries
- Linking multi-media fragments to lexical entries
- Creation of links in images
- Link to resources within the digital archive (or other external web-based resources) – interaction with other archiving tools
- Create conceptual spaces with the ViCoS extension
- Search in and cross lexica
- Export to xml and toolbox format
- Export to IMDI package (in prep.)



Two dictionaries are under creation for KSDP:

**SaRuDic** (Encyclopedic dictionary of Kildin Saami):

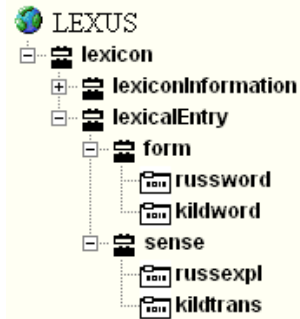
- Kildin Saami – Russian ( – English) dictionary
- Based on data from *Saamsko-russkij slovar'* by Kuruč et al. (1985) / *Electronic Saami Dictionary* by Jur'jev (2003–2007)
- Enriched with data from KSDP's annotated corpus of spoken Kildin Saami (via import from KSDP's Toolbox projects) as well as with data from other existing dictionaries
- Rich structure and content – including derivations and example sentences
- Including multi-media fragments

• **RuSaDic** (Russian-Kildin wordlist):

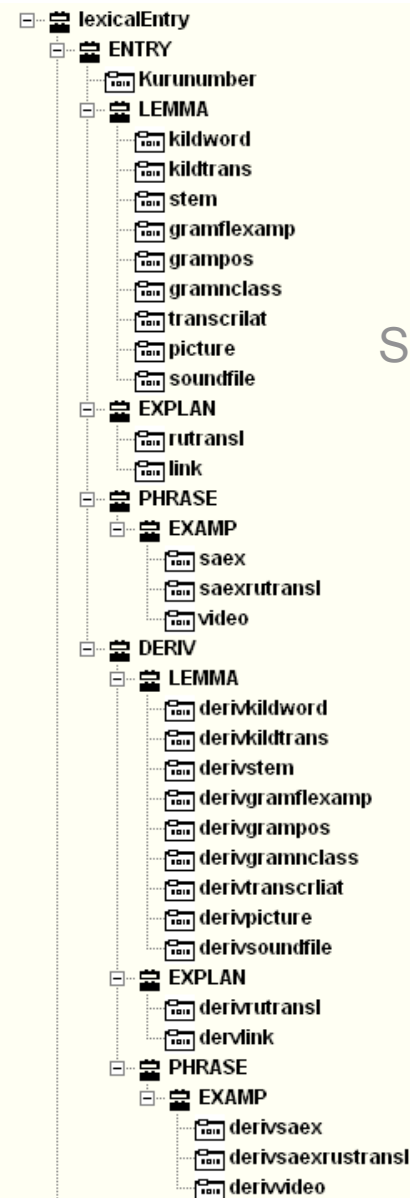
- Preliminary stage (approx. 1000 entries)
- Limited structure
- Limited content (no multi-media)
- Index for SaRuDic



# Kildin Saami dictionaries in LEXUS



RuSaDic structure



SaRuDic structure



# Kildin Saami dictionaries in LEXUS



## RuSaDic structure

- Russian headword
- Additional information about the headword (if necessary)
- Kildin Saami translation (in Cyrillic orthography)
- Kildin Saami translation (in Latin transliteration)

## lexicalEntry

## SaRuDic structure

- Saami headword
- Grammatical information (stem alternation, inflection class)
- Russian translation or explanation
- Saami example phrase(s)/sentence(s)
- Russian translation of example(s)
- Derived word forms (in alphabetical order)
- Grammatical information (stem alternation, inflection class)
- Russian translation or explanation
- Saami example phrase(s)/sentence(s)
- Russian translation of example(s)



## Kildin Saami dictionaries in LEXUS



**САММЬ** (сāмь) |sāmm' (sām')| I, 3 Saami person (*masc., fem. and pl.*); сāммь пӯдә |sāmm' pūd'e| a Saami arrived; сāмь пӯдтәнь |sām' pūdt'en'| Saami (*pl.*) arrived; сāмь кйлл |sām' kīll| Saami language; сōнн лй сāмь олма |sōnn lī sām' olma| s/he is Saami  
сāмас |sāmas| Saami (*adv.*), сārрнэ сāмас |sārre sāmas| to speak Saami  
сāммьленч (сāммья) |sāmm'l'enč (sāmm'l'a)| IV Saami person (*masc., fem. and pl.; originally a diminutive*); сōнн лй сāммьленч |sōnn lī sāmml'enč| s/he is Saami; сый лев сāммья |sijj liev sāmm'l'a| they are Saami  
сāммьлуввэ (сāммьлувэ) |sāmm'luvve (sāmm'lurve)| I to become like a Saami (*esp. according to one's way of life*); сōнн сāммьлувэ |sōnn sāmm'lurve| s/he became like a Saami; сōнн ялл сāмас |sōnn jāll sāmas| s/he lives (like) Saami (*i.e. the Saami way*)



# Kildin Saami dictionaries in LEXUS



**САММЬ** (сāмь) |sāmm' (sām')| I, 3 Saami person (*masc., fem. and pl.*); **сāммь** пӯдә |sāmm' pūd'e| a Saami arrived; **сāмь пӯдтәнь** |sām' pūdt'en'| Saami (*pl.*) arrived; **сāмь кйлл** |sām' kīll| Saami language; **сōнн лй сāмь олма** |sōnn lī sām' olma| s/he is Saami

**сāмас** |sām

Saami (person) **сāммь** |sāmm'|

peak Saami

**сāммьлен'**

Saami (person) **сāмьленч** |sām'l'enč|

n (*masc., fem.*

*and pl.; oriç*

mm'l'enč| s/he

is Saami; **сый лев сāммья** |sij lev samm'l'a| they are Saami

**сāммьлуввэ** (сāммьлувэ) |sāmm'luvve (sāmm'lurve)| I to become like a Saami

(esp. according to one's way of life); **сōнн сāммьлувэ** |sōnn sāmm'lurve| s/he

became like a Saami; **сōнн ялл сāмас** |sōnn jäll sāmas| s/he lives (like)

Saami (i.e. the Saami way)





# Kildin Saami dictionaries in LEXUS



Lexus - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Back Forward Reload Stop Home http://lux07.mpi.nl/mpi/lexus/lexiconOverview.htm Google

Getting Started Latest Headlines

LEXUS Word List View Search Relation types Help sami lexica

## SaRuDic Aug08: Word list


Results 1-25 of 125 for a entries  
<< first < previous Page 1 next > last >>

Lexical Entry	
A	
A-A-A	
аббрай	
аббршз	
аббрье	
аббрьенне	
аббрьювве	
аббрз	
АББРР	
аббрьевь	
абрнз	
абрсз	
абрсзплз	
абрсзпнз	
АВТОБУС	
АВВТА	
АВГУССТ	
автнзплз	
автнзпнз	
авьтвудт	
авьтмусс	
авьтнз	
авьтзм	

**LEMMA**

**АББРР** абър 3 П  
*abb'r*

**MEDIA**



**TRANSLATION**

дождь М

**PHRASES**

абър абр (манн, ёадт)  
дождь идёт

абър оалкувэ  
дождь начался

абър-чалма  
капля (букв. глазок) дождя



## Why ViCoS?

Paper dictionaries' limited usefulness in language maintenance & language revival  
(Manning et al., 2000)

Speech community may prefer *explicit* semantic links, possibly of a wide variety of types

→ introducing *Conceptual Spaces*, where concepts are

- related to others
- anchored in language
- illustrated with multimedia

→ **Vi**sualizing **Co**nceptual **S**paces





Existing lexical resource is starting point

Words offer key access

ViCoS is technology to link words – and the associations they evoke – to other words

Ease-of-use to empower community members to actively describe their language & culture and to learn from such resources  
resources targeted for human consumption

Resulting Conceptual Spaces = sort of informal ontology of fuzzily-defined concepts and relationships  
but where concept nodes are anchored in corresponding formal resources (lexica)



# Connecting RuSaDic and SaRuDic using ViCoS



ViCoS - Visualising Conceptual Spaces - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window

Back Forward Reload Stop Home <http://lux07.mpi.nl/mpi-vicos/vicos.html> Google

Getting Started Latest Headlines

Lexus Workspace Editor ViCoS - Visualising Conceptual Sp...

ViCoS - Max Planck Institute For Psycholinguistics - <http://www.mpi.nl> Lexicon: RuSaDic August 08 Data Categories: russword

а б в г д е ж з и к л м н о п р с т у ф х ц ч ш щ ь э ю я search

- +

Welcome дождь август алфавит

Word list view

- август
- август
- автобус
- автор
- адрес
- аккуратно
- аккуратный
- аккуратный
- аккуратный
- алфавит
- апрель
- апрель
- аркан
- армия
- армия

Lexical entry view for алфавит (RuSaDic August 08)

Lexical Entry	kildword	russword
LE_HANDLE	алфавяйт	алфавит

Save Clear Help Browser RT Mgr Img Mgr Txt Mgr

Source Relation Type Target

Add to Knowledge Space Show in Knowledge Space Edit in Image Mapper Edit in Text Mapper

ViCoS Status Line

Done



# Connecting RuSaDic and SaRuDic using ViCoS



ViCoS - Visualising Conceptual Spaces - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window

Back Forward Reload Stop Home <http://lux07.mpi.nl/mpi-vicos/vicos.html> Google

Getting Started Latest Headlines

Lexus Workspace Editor ViCoS - Visualising Conceptual Sp...

ViCoS - Max Planck Institute For Psycholinguistics - <http://www.mpi.nl> Lexicon: RuSaDic August 08 Data Categories: russword

а б в г д е ж з и к л м н о п р с т у ф х ц ч ш щ ь э ю я search

- +

Welcome дождь август алфавит

Word list view

- август
- август
- автобус
- автор
- адрес
- аккуратно
- аккуратный
- аккуратный
- аккуратный
- алфавит
- апрель
- апрель
- аркан
- армия
- армия

Lexical entry view for алфавит (RuSaDic August 08)

Lexical Entry	kildword	russword
LE_HANDLE	алфавит	алфавит

Save Clear Help Browser RT Mgr Img Mgr Txt Mgr

Source Relation Type Target

Add to Knowledge Space Show in Knowledge Space Edit in Image Mapper Edit in Text Mapper

ViCoS Status Line

Done



# Connecting RuSaDic and SaRuDic using ViCoS



ViCoS - Visualising Conceptual Spaces - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window

Back Forward Reload Stop Home http://lux07.mpi.nl/mpi-vicos/vico5.html Google

Getting Started Latest Headlines

Lexus Workspace Editor ViCoS - Visualising Conceptual Sp...

ViCoS - Max Planck Institute For Psycholinguistics - <http://www.mpi.nl> Lexicon: RuSaDic August 08 Data Categories: russword

а б в г д е ж з и к л м н о п р с т у ф х ц ч ш щ ь э ю я search

Word list view

- август
- август
- автобус
- автор
- адрес
- аккуратно
- аккуратный
- аккуратный
- аккуратный
- алфавит
- апрель
- апрель
- аркан
- армия
- армия

Welcome дождь август алфавит

Lexical entry view for алфавит (RuSaDic August 08)

Lexical Entry	kildword	russword
LE_HANDLE	алфавит	алфавит

Source Relation Type Target

Add to Knowledge Space Show in Knowledge Space Edit in Image Mapper Edit in Text Mapper

ViCoS Status Line

Done



# Connecting RuSaDic and SaRuDic using ViCoS



ViCoS - Visualising Conceptual Spaces - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Back Forward Reload Stop Home http://lux07.mpi.nl/mpi/vicos/viCoS.html

Getting Started Latest Headlines

Lexus Workspace Editor ViCoS - Visualising Conceptual Sp...

ViCoS - Max Planck Institute For Psycholinguistics - http://www.mpi.nl Lexicon: SaRuDic Aug08 Data Categories: kildword

a ä å б в г д е ё ё ж з и й я љ к л ѓ м ж н њ њ о ѓ п р р с т у у ф х ц ч ш  
щ ъ ы ь ь Ъ Э Э Ю ю Я Я search

- +

Welcome дождь август алфавит АББЪР АВГУСТ АЛФАВИИТ

Word list view

alnlzllz  
alnlzlnz  
АЛЛЪТ  
АЛЛЪК  
АЛЛЪКЭ  
АЛЛЪМ 1  
АЛЛЪМ 2  
аллътлэс  
алнэллэ  
алнэлнэ  
алтнэллэ  
алтнэлнэ  
АЛФАВИИТ  
альмсэ  
альмсэллэ  
альмсэлнэ  
АЛЪН  
альнаһтэ  
альнассытэ  
альне

Lexical entry view for АЛФАВИИТ (SaRuDic Aug08)

Lexical Entry	kildword
LE_HANDLE	АЛФАВИИТ

Save Clear Help Browser RT Mgr Img Mgr Txt Mgr

Source Relation Type Target

Add to Knowledge Space Show in Knowledge Space Edit in Image Mapper Edit in Text Mapper

ViCoS Status Line  
Done



# Connecting RuSaDic and SaRuDic using ViCoS



ViCoS - Visualising Conceptual Spaces - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window

Back Forward Reload Stop Home <http://lux07.mpi.nl/mpi/vicos/viCoS.html> Google

Getting Started Latest Headlines

Lexus Workspace Editor ViCoS - Visualising Conceptual Sp...

ViCoS - Max Planck Institute For Psycholinguistics - <http://www.mpi.nl> Lexicon: SaRuDic Aug08 Data Categories: kildword

а б в г д е ё ж з и й я к л м н о п р с т у ф х ц ч ш  
щ ь ы ь ь Ъ Э Э Ю Ю Я Я search

Welcome дождь август алфавит **АББЬР** АВГУССТ АЛФАВИИТ

Word list view

- аллнэллэ
- аллнэлнэ
- АЛЛЪТ
- АЛЛЪК
- АЛЛЪКЭ
- АЛЛЪМ 1
- АЛЛЪМ 2
- алльтлэс
- алнэллэ
- алнэлнэ
- алтнэллэ
- алтнэлнэ
- АЛФАВИИТ**
- альмсэ
- альмсэллэ
- альмсэлнэ
- АЛЪН
- альнайтэ
- альнасьтэ
- альне

Lexical entry view for АЛФАВИИТ (SaRuDic Aug08)

Lexical Entry	kildword
LE_HANDLE	АЛФАВИИТ

Save Clear Help Browser RT Mgr Img Mgr Txt Mgr

Source Relation Type Target

Add to Knowledge Space Show in Knowledge Space Edit in Image Mapper Edit in Text Mapper

ViCoS Status Line  
Done





# Connecting RuSaDic and SaRuDic using ViCoS



ViCoS - Visualising Conceptual Spaces - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window

Back Forward Reload Stop Home http://lux07.mpi.nl/mpi/vicos/ViCoS.html Google

Getting Started Latest Headlines

Lexus Workspace Editor ViCoS - Visualising Conceptual Sp...

ViCoS - Max Planck Institute For Psycholinguistics - http://www.mpi.nl Lexicon: SaRuDic Aug08 Data Categories: kildword

а б в г д е ё е ж з и й я к л л м н н о о п р р с т у у ф х ц ч ш  
щ ь ы ь ь ь ь э э ю ю я я search

Word list view

- аллнэллэ
- аллнэллэ
- АЛЛЪТ
- АЛЛЪК
- АЛЛЪКЭ
- АЛЛЪМ 1
- АЛЛЪМ 2
- аллътлэс
- алнэллэ
- алнэллэ
- алтнэллэ
- алтнэллэ
- АЛФАВИНТ
- альмсэ
- альмсэллэ
- альмсэллэ
- АЛЪН
- альнайтэ
- альнасстьэ
- альне

Lexical entry view for АЛФАВИНТ (SaRuDic Aug08)

Lexical Entry	kildword
LE_HANDLE	АЛФАВИНТ

Save Clear Help Browser RT Mgr Img Mgr Txt Mgr

Source	Relation Type	Target
алфавит август дождь	kildin link	АББЪР АВГУССТ АЛФАВИНТ

Add to Knowledge Space Show in Knowledge Space Edit in Image Mapper Edit in Text Mapper

ViCoS Status Line

Done



# Connecting RuSaDic and SaRuDic using ViCoS



ViCoS - Visualising Conceptual Spaces - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window

Back Forward Reload Stop Home http://lux07.mpi.nl/mpi/vicos/ViCoS.html Google

Getting Started Latest Headlines

ViCoS - Visualising Conceptual Sp... ViCoS Editor and Navigator Lexus

ViCoS - Max Planck Institute For Psycholinguistics - http://www.mpi.nl Lexicon: SaRuDic Aug08 Data Categories: kildword

а б в г д е ж з и й я к л м н о п р с т у ф х ц ч ш  
щ ь ы ь ь ъ ѓ э ю я search

Welcome дождь август алфавит АББЬР

Word list view

A
А.А.А
аббрай
аббршз
аббрье
аббрьенне
аббрьювве
аббрэ
АББЬР
аббрьесь
абрнз
абрсэ
абрсэллэ
абрсэлнэ
АВТОВУС
АВВТА
АВГУССТ
автнэллэ
автнэлнэ
авьтвудт

Lexical entry view for дождь (RuSaDic August 08)

Lexical Entry	kildword	russword
LE_HANDLE	аббрь	дождь

Save Clear Help Browser RT Mgr Img Mgr Txt Mgr

Source Relation Type Target

Add to Knowledge Space Show in Knowledge Space Edit in Image Mapper Edit in Text Mapper

ViCoS Status Line  
Done



# Connecting RuSaDic and SaRuDic using ViCoS



The screenshot displays the ViCoS Editor and Navigator interface in Mozilla Firefox. The main window shows a large empty space with two grey circles: one containing the word 'дождь' (rain) and another containing 'АББЬР' (Abbr). A red arrow points from 'дождь' to 'АББЬР', which is circled in red. To the right, a smaller window titled 'Lexus - Mozilla Firefox' displays a 'SaRuDic Aug08: Lexical entry' for 'АББЬР'. The entry includes a lemma 'АББЬР' with the phonetic transcription 'abbr' and the number '3'. Below the lemma, there is a 'MEDIA' section with a photograph of a field. The entry also includes 'EXPLAN', 'PHRASE', and 'PHRASES' sections. At the bottom of the interface, there are buttons for 'lexus', 'clear', 'save', and 'refetch', and a 'Relation Types:' dropdown menu.



# Connecting RuSaDic and SaRuDic with outside info



Using ViCoS links can be made from ViCoS concepts to external url sources  
(new feature in testing phase)

The screenshot shows two browser windows. The left window is the RuSaDic website (Álgu-tietokanta) with the URL [http://kaino.kotus.fi/algui/index.php?t=sanue&sanue\\_id=7097&kkieli=sa](http://kaino.kotus.fi/algui/index.php?t=sanue&sanue_id=7097&kkieli=sa). The right window is the Russian Wikipedia article for 'Северный олень' (Arctic reindeer) with the URL [http://ru.wikipedia.org/wiki/Олень\\_северный](http://ru.wikipedia.org/wiki/Олень_северный).

The RuSaDic window shows a search interface with a keyboard layout and a list of words: **boazo**, **boazu**, and **boazu** with their respective phonetic transcriptions and links to external sources.

The Wikipedia window shows the article for 'Северный олень' (Arctic reindeer) with a detailed description, a table of contents, and a photograph of a reindeer with large antlers.



# Point`s of discussion



Unicode: LEXUS and ViCoS are online tools

→ sometimes the correct display of characters is problematic

Offline versions of dictionaries

→ LEXUS and ViCoS need database (storage) so CD options are limited.

→ PDF paper output is foreseen (but excludes the video options)

→ Conversion from XML to formatted text is easy

Community based continuation

→ current version: read only (login with username and password)

→ editing: restricted to selected users

→ data curation is an issue



## **Michael Rießler**

Department for Northern European Studies, Humboldt-Universität zu Berlin/Germany

michael.riessler@staff.hu-berlin.de

## **Jacqueline Ringersma**

Max-Planck-Institute for Psycholinguistics, Nijmegen/The Netherlands

jacqueline.ringersma@mpi.nl

DoBeS and KSDP are funded by the Volkswagen Foundation.  
Additional funding for the Kildin Saami dictionary project was provided  
by the Saami Council.

KSDP's Kildin Saami dictionary project also benefits (or benefitted)  
from help provided by: Aleksandr Jur'jev, Aleksandra Antonova,  
Anastasija Nielsen, Anja Behnke, Elena Karvovskaja, Nina Afanas'jeva,  
Jevgen'ja Životova



## References



- Documentation of Endangered Languages (DoBeS) <http://www.mpi.nl/dobes/>
- Jur'jev, A. 2003–2007. *Electronic Saami dictionary 2.0* <http://rusadic.narod.ru/>
- Kola Saami Documentation Project (KSDP) <http://www2.hu-berlin.de/ksdp/>
- Kuruč, R. D., N. E. Afanas'jeva & E. I. Mečkina. 1985. *Saamsko-russkij slovar'* = *Sām'-rūšš soagknehk'* [Saami-Russian dictionary]. Moskva: Russkij jazyk.
- Language Archiving Technology (LAT) <http://www.mpi.nl/lat/>
- Rießler, M. & J. Wilbur. 2007. Documenting the endangered Kola Saami languages. In: T. Bull et al. (eds.) *Språk og språkforhold i Sápmi* (=Berliner Beiträge zur Skandinavistik 11). Berlin: Humboldt-Universität. 39–82.