

LingvoDoc: Extralinguistic Information and Maps

Natalia Koshelyuk

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

October 5, 2021

LingvoDoc: extralinguistic information and maps¹

Kosheliuk Natalia ORCID 0000-0002-5833-7971 Ivannikov Institute for System Programming of the RAS, Moscow (Russia) NKoshelyuk@yandex.ru

Abstract. The paper is a continuation of a series of articles about work and user experience on the LingvoDoc linguistic platform. The methods of adding extralinguistic information to dictionaries and corpora are presented step by step, and the possibilities of working with dialectological maps on the platform are also considered.

Keywords. LingvoDoc, data mining, linguistics, maps, extralinguistic information, maps

1 INTRODUCTION

Extralinguistics is a field of linguistics that studies the total combination of social, historical, geographical and cultural factors associated with the development of language. According to F. de Saussure, one of the founders of structuralism in linguistics, the cultural environment in which a nation develops, its history and customs may influence the development of language. Extralinguistic information, which can now be added to LingvoDoc, in turn, serves as a source of additional information about a language or dialect. On the basis of this information it is possible to judge the conditions of functioning and development of these languages and also this information often helps linguists to ascertain the dialectal affiliation of ancient writing monuments, when this cannot be done by means of classical linguistic analysis. In addition, knowing the area of distribution of a particular linguistic phenomenon, one can build the dialectological isoglosses² that are necessary for drawing up detailed linguistic maps.

¹ Supported by Russian Science Foundation, project no. 20- 18-00403 'Digital Description of Uralic Languages on the Basis of Big Data'

This article presents a sequential description of the steps for working with extralinguistic information and isogloss maps in Lingvo-Doc.

2 HOW TO ADD EXTRALINGUISTIC INFORMATION INTO SOURCE

Work with sources on the LingvoDoc platform always begins with registration in the system or login to your account. After this step, the user will be able to work with the full linguistic base of the platform (in some cases with the prior permission of the author of a dictionary or corpus) and use all of its options.

Expanding a dictionary or corpus with the necessary information is carried out according to a single pattern, the only difference is the following:

1) filling in the appropriate fields of LingvoDoc when creating the resource² (Pic. 1);

2) entering extralinguistic information by editing the available dictionaries and corpora in the personal account (Pic. 2).

¹ Isogloss is a line on the map connecting places with the same linguistic features.

¹ Qv. preprint [Koshelyuk 2021].

Pic. 1. Adding extralinguistic information when creating a resource

Lingvodoc 3.0		Natalia Koshelyuk 🔹	Tasks 💶	Languages (English)
Parent language Select language	Dictionary names and metadata Set dictionary name, translations and metadata	Perspectives Create one or more perspectives		Finish
Please, select the parent language	Crasta Crasta Crasta			

Pic. 2. Adding extralinguistic information by editing



Lingvodoo	3.0						Natalia Koshel	yuk 🔻	Tasks 💶	Languages (E	inglish)
My dictionaries	Available o	fictionaries									
Th	e gospel of Mati	thew and Ma	k in the V	ogul language	(1847-1848) -					Hidden 🔻	
8	Lexical Entries	• Viev	Edit	Publish	Contributions					Published 💌	
Die	tionary of Man	si language, c	ompiled by	y archpriest S	imeon Cherkalov (So	olikamsk city, 1783) 🔹				Published 👻	
	Paradigms 🔻	View	Edit	Publish	Contributions					Published 🔻	
8	Lexical Entries	• Viev	Edit	Publish	Contributions					Published 👻	

After the creation or selection of a desired resource, you need to go to the Properties tab, after which a field will open for filling in or correcting additional information about the source (Pic. 3-5): title, authors, field or archive data, year of publication/collection of the material, genre, subject, localization, etc.

Pic. 3. Switching to fields with extralinguistic data

Lingvodoc 3.0		Natalia Koshelyuk 🔻	Tasks 💶	Languages (English)
My dictionaries Available dictionaries				
The gospel of Matthew and Mark in the Vogul language (1847-1848)				Hidden 👻
Alla Roles				Published 💌
Organizations olikar	msk city, 1783) 🔹			Published 👻
% Statistics B Save dictionary				Published 🔻
K Remove dictionary				Published 👻
Reles Properties Organizations Save dictionary Remove dictionary				

Pic. 4. Extralinguistic information. Title, author, type of source.

Properties		×
Created by: Natalia Koshelyuk Created at: 02.07.2021, 20:03:15 Last modified at: 18.09.2021, 16:34:34		
Translations The pospel of Matthew and Mark in the Vogul language (1947-1948)	English	- Update
Евантелие от Марка и Матфея на мансийском языке (1847-1848 гг.)	Russian (Русский)	• Update
Add		
C Expedition Archive		Save
Authors Interrogator Sove Informant		Save

Pic. 5. Extralinguistic information. Localization



The place of collection of a field material or an archive source can be ascertained exactly by coordinates or conventionally by typing the name of a settlement or region in the search line.

3 HOW TO WORK WITH MAPS

When working with maps in LingvoDoc, one can be offered a wide range of options:

- ✓ Maps
- ✓ Search
- ✓ Storage

In order to go to the desired option, you need to go to the Tools tab, after which all the icons will be displayed on the platform (Pic. 6).

Pic. 6. Tools. Maps, Search, Storage, Languages





Maps is an interactive map of Eurasia, which displays the entire set of loaded dictionaries and corpora for each language and dialect in dotted form (Pic. 7). By selecting the desired source from the suggested list, you can see the location of its collection in detail.





Search allows you to track various linguistic phenomena (grammatical, phonetic, semantic, etc.) both in individual Lingvodoc dictionaries and corpora, and throughout the entire database. Each search can be carried out in a separate tab and the desired parameters can be selected (Pic. 8).

Pic. 8. Search



oc 3.0				Natalia Koshely	uk 🔻	Tasks 1	Languages
Search 1 🗙 Search 2	X Search 3 X	+					
Search							
Search in							
Dictionaries			🗹 Corpora				
Select la	inguages		Grammar			Select tags	
Selected by default:	Show						
Search options							
Ignore adoptions	 Search for adoptions 	O Exclude adoptions	 Ignore e 	etymology 🔘 Has et	ymology	Doesn't have et	tymology
Ignore diacritics							
					_		
	_						
oc 3.0				Natalia Koshely	uk 💌	Tasks 💶	Languages
oc 3.0 Search options				Natalia Koshely	uk 🔻	Tasks 1	Languages
Search options	Search for adoptions	Exclude adoptions	; Ignore	Natalia Koshely etymology 💶 Has e			
Search options	Search for adoptions	Exclude adoptions	; Ignore				
Search options Ignore adoptions Ignore diacritics	Search for adoptions	Exclude adoptions	; Ignore				
Search options Ignore adoptions Ignore diacritics OR/AND mode	Search for adoptions	Exclude adoptions	s Ignore				
Search options Ignore adoptions Ignore diacritics	Search for adoptions	Exclude adoptions	: Ignore				
Search options Ignore adoptions Ignore diacritics OR/AND mode	Search for adoptions	Exclude adoptions	: Ignore				
Search options Ignore adoptions Ignore adoptions Ignore diacritics OR/AND mode AND OR	Search for adoptions Search String	Full string •					tymology
Search options Ignore adoptions Ignore adoptions Ignore diacritics OR/AND mode AND OR		Full string + Full string		etymology 🔲 Has e			tymology
Search options Ignore adoptions Ignore adoptions Ignore diacritics OR/AND mode AND OR		Full string • Full string Sub string	X Add ANI	etymology 🔲 Has e			tymology
Search options Ignore adoptions Ignore adoptions Ignore diacritics OR/AND mode AND © OR AND block	Search String	Full string • Full string Sub string	X Add AND	etymology 🔲 Has e		Doesn't have e	tymology
Search options Ignore adoptions Ignore adoptions Ignore discritics OR/AND mode AND OR	Search String	Full string • Full string Sub string	X Add AND	etymology 🔲 Has e		Doesn't have e	tymology

We display in more details how to carry out a search by individual languages, grammatical indicators and tags (Pic. 9-11).



Pic. 9. Implementation of the Search option by selecting language

Pic. 10. Implementation of the Search option by selecting grammar indicators

Lingvod	oc 3.0	Natalia Koshelyuk 🝷 🛛 Tasks 💶	Languages (English) 🔹
	Search		
	Search in		
	Dictionaries Corr	ora	
	Select languages Grammar	Select tags	
	Selected: Show		



Pic. 11. Implementation of the Search option through tags

Lingvodoc 3.0		Natalia Koshelyuk 🔻	Tasks 💶	Languages (English) 🔻
Search				
Search in				
✓ Dict	tionaries 🗹 Corpo	ora		
	Select languages Grammar		Select tags	
Selecter	d: Show			
				^
Lingvodoc 3.0 Search	Select tags			Languages (English) 🔻
Search in	Radio			
	Clear			
Selecte	Data source -			
	Clear			
Search op	Years			
Igr	secceptors			hology
Igr	Select all Clear all			
OR/AND	maa		Close	
O AN	D OR			

In addition, the Search option makes it possible to combine the data requested in Search into separate areas. A special color palette can

be selected for the borders of each of them (Pic. 12) and the received results can be converted into exe.

Pic. 12. Search. Geographical ranges

		Add OR block	
Search Export to XLSX			
Search 1 Search 2 Search 3			
 Выбрать группы для ареалов 			Режим ареалов
0 🌑 — — О Более О	пересечений		
Search 1 Search 3 Search 5			Реким ареалов
Sand 1 Sand 3 Send 5 • Budgata pyneu gra speance Search 1 Search 3			
Search 1 Search 3 Search 5			Режим ареалов
Search Se			



Storage is a collection of ready-made maps of various language features in various languages. At the moment, the collection consists of 46 maps in Russian (Pic. 13).









CONCLUSION

As can be seen from the overview, the search algorithms built into LingvoDoc can be used to analyze linguistic phenomena in terms of their geographical location, supplementing them with extralinguistic information, which in turn makes it possible to conduct research on the study of a linguistic material on a more meaningful and all-embracing level.

REFERENCES

1. LingvoDoc Homepage, http://lingvodoc.ispras.ru/. Last accessed 09/20/2021

2. Great Russian Encyclopedia, https://bigenc.ru/linguistics/text/2002556. Last accessed 09/20/2021