

# LingvoDoc: Working with Cognate Analysis

Natalia Kosheliuk

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

February 11, 2022

## LingvoDoc: working with Cognate Analysis<sup>1</sup>

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

**Abstract.** This article offers an overview of one of the main LingvoDoc programs – the Cognate analysis option intended for modern etymological and dialect distance research. It describes in detail how this option works and how to implement it. The advantages of the Cognate Analysis algorithm are also highlighted.

Keywords. LingvoDoc, data mining, linguistics, cognate analysis

## **1 INTRODUCTION**

Cognate identification is an important thing for identifying the genetic affinity of languages and their dialects. This method allows linguists to draw conclusions about the development of languages over time and obtain new phonetic and etymological data. Today only a small part of languages has been analyzed in terms of their genetic connections. The main reason is that comparative research in linguistics is still based on the personal work of scientists which is very time-consuming. However, in recent years there has been an active development of methods of computer work on the implementation of these tasks.

In this paper, using the example of the Mansi corpus, we will show how the Cognate Analysis option is implemented in the LingvoDoc and what its advantage is.

#### 2 HOW TO CONDUCT COGNATE ANALYSIS

Initially, work on the LingvoDoc platform begins with standard authorization, downloading the necessary dictionary, or opening an existing one in the database (Pic. 1).

<sup>&</sup>lt;sup>1</sup> Supported by Russian Science Foundation, project no. 20- 18-00403 'Digital Description of Uralic Languages on the Basis of Big Data'

Pic 1. Getting started on LingvoDoc.



To obtain the most accurate and reliable result of etymological analysis, if possible, it is necessary to choose the most qualitative and maximally filled dictionaries (Pic. 2).

Currently, 23 Mansi dictionaries are connected by etymologies on the Linguistics platform. Most of them are archival texts. The decision to involve them in the analysis is due to the rather high accuracy of the reflection of phonetic oppositions in them. Also, these data are important as one of the first written sources on the Mansi dialects and according to them, you can trace how they have changed over the past 100-250 years.

Pic 2. <i>I</i>	Linking	Cognates.
-----------------	---------	-----------

Lingvodoc 3.0				Natalia Koshelyuk 🔻	💭 📾 Eng 💌
Uralic > Mansi > Northern dialect > Dictionary of Ma	nsi language (collected by PS.Pallas XVIII ce	pt. from archive of M.A.Sjögres	n) = > Leftical Enftries =	• •	license: Proprietary
Edit (553) Publish (553) View published	(553) View contributions (14) M	erge suggestions Tools	••••••	<u></u>	Search Q
└ Word ▲ ▼	- Phonemic transcription 🔺 🗸	- Meaning A 🗸	└ Sound ▲ ▼	L Cognates 🔺 🗸	└ Paradigm and contexts ▲ ▼
	∟ шуншъ	∟ блоха		Cognates	
∟? вит "вода"	∟ ви́ту вой	∟ бобр		🥠 Cognates	
∟тёрум	<sup>L</sup> то́ромъ	L Bor		Cognates	
_ пўцыц	∟ пуніть	∟ богатый		Cognates	
<sup>L</sup> ня́р "болото (с морошкой)"	∟ няръ	∟ болото		Cognates	
<sup>L</sup> агмын	<sup>L</sup> а́умы	∟ больной		Cognates	
1⇔ < 1 2 3 4 5 6 7	> → 28				• • •

After selecting the necessary linguistic material for Cognate Analysis, you need to go to the "Tools" tab and start the process (Pic. 3).

Pic 3. Implementation of the Cognate Analysis.

Lingvodoc 3.0	1		Nat	alia Koshelyuk 🔻	🛋 🗰 Eng 💌
Uralic > Mansi > Northern dialect > Dictionary e	of Mansi language (collected by P.S.Pallas XVIII cept	t. from archive of M.	A.Sjögren) + → LeRical Entries +	•••	license: Proprietary
Edit (553) Publish (553) View publi	shed (553) View contributions (14) Me	rge suggestions	Tools •	<u> </u>	Search Q
└ Word ▲ ▼	Phonemic transcription	L Meaning 4	Cognate acoustic analysis	Cognates 🔺 👻	□ Paradigm and contexts ▲ ▼
			Cognate analysis		
	∟шуншъ	∟ блоха	Cognate multi-language analysis Cognate multi-language reconstruction	♦ Cognates	
∟? вит "вода"	∟ ви́ту вой	∟ бобр	Cognate multi-language suggestions Cognate reconstruction	♦ Cognates	
∟ тōрум	∟ то́ромъ	∟ <sub>Bor</sub>	Cognate suggestions Phonemic analysis	♦ Cognates	
∟ пунын	∟пунть	∟ богатый	Phonology Phonological statistical distance	Cognates	
∟няр "болото (с морошкой)"	⊥ <sub>дкн</sub> ⊥	∟ болото	Sound and markup Valency	Cognates	
<sup>L</sup> āгмың	∟ а́умы	∟ больной		♦ Cognates	
1⇔ < 1 2 3 4 5 6	7 > →28				

The LingvoDoc platform user also can choose the necessary dictionaries (Pic. 4): you can stop at all available data or limit yourself to at least two dictionaries. Based on our experience, the optimal number of dictionaries subject to processing should not exceed 20. This will be explained by the need to make some editorial changes to combine the rows: manually transferring to reliable – doubtful rows, and to doubtful – single ones. Editorial work with lists of words represented by a large number of columns resulting from the implementation of the Cognate Analysis option is a time-consuming task and can lead to missing important matches.

Pic 4. Selection of dictionaries necessary for research.



After starting the process of Cognate Analysis, which takes from 1 to 5 minutes, the user receives a result that reflects the genealogical distance between the sources involved in the analysis, as well as fully painted rows of reliable and unreliable correspondences of vowels and consonants (Pic. 5-6).



Briefly outline the principles of the etymological analysis algorithm on the Linguistics platform: at the first stage, for each character from the transcription proposed in each particular dictionary, the algorithm calculates its correspondences in words from other dialects of the same language associated with etymologies with this dictionary. The main calculation is carried out by the roots of words, based on the fact that in them the first vowel corresponds to the first vowel, the first consonant corresponds to the first consonant, the second consonant corresponds to the second. Combinations of vowels and vowels, consonants and consonants are possible. At the output, we get a list of correspondences for each pair of idioms.



этимологический	анализ								
2: The gospel o 3: Concordance 4: Dictionary o	of Matthew and Mark in of glossed corpus of E of Mansi language, comp	the Vogul language vangel of Matthew iled by archpries	as XVIII cent. from and 2 (1847-1848) - Lexical 1878 in Konda dialect t Simeon Cherkalov (Sol 1800 priest K.Slovcov,	l Entries (144 фор t of Mansi languag Likamsk city, 1783	м = 51% от числа соот e - Lexical Entries (: ) - Lexical Entries (:	в.) 156 форм = 55% от ч 151 форм = 54% от ч	нисла соотв.) нисла соотв.)		
Надёжные ряды: [а]-[а]-[а]-[а]-[а] Сомнительные ря ? - ? - ? -[4]; ? -[е]-[е]-[?] [0]-[0]-[0]-[0]-[0] [а]-? - ? -[е]. Материал — надё	//////////////////////////////////////								
1: Dictionary o									
1: Dictionary o	if M.	2: The gospel of	of M.,	3: Concordance	of _	4: Dictionary	of M.	5: Dictionar	y Uppe
[a]	f M.,	2: The gospel o	of M.,	3: Concordance [a]	of _	4: Dictionary [a]	of M.	5: Dictionar	y Uppe
	f м	0.11	оf М., "мороз" "д" "умирать, погибат	[a] atšerem am	of	,	оf М., *Стужа*		у Uppe Чморо: 'Я., с

The author of the dictionary has the opportunity to download the results of the analysis in Excel format (Pic. 7), analyze, verify the correctness of transcriptions and etymologies that led to non-standard lines of correspondence, and make adjustments to transcription and etymology. Further, the above algorithm can be restarted again already on the material verified by the author.

Pic 7. The results of etymological analysis in Excel.

										Nikita Koshelyi	A NC 1	⊡ –	
Лайл Главная													Поделиться
ставить 🤯 🖌 🕺	К <i>К</i> <u>Ч</u> -   Ш -   Шрмфт	<u>∆</u> • <u>A</u> • = ≈	= = €∃ €⊒ Объе Выравни	43.244940	пъ в центре -	upsi vieto (5	форматирование -	Форматировать Стили как таблицу - ячеек стили	Вставить Удалит Ячайог	ь Формат	и филь	овка Найти пр выделил ирование	
•	$\times \checkmark f_t$	Соответствия по	э начальному гласному:										
A	В	C	D	E	F	G	н	1	1	K	L	M	N
Соответствия п	о начальному глас	ному:											
Надёжные ряды													
[a]-[a]-[a]-[a]	a]—[a]												
Сомнительные													
? - ? - ? -[i]-													
[0]-[0]-[0]-[	o]-[o]												
[o]-[o]-[o]-[ [a]-?-?-[e]	o]-[o]												
[o]-[o]-[o]-[ [a]-?-?-[e]	o]—[o]  — ?												
[0]—[0]—[0]—[ [а]— ? — ? —[е] Материал — на,	o]—[o]  — ?												
[о]—[о]—[о]—[ [а]—?—?—[е] Материал— на,	о]—[о]  — ? дёжные ряды:												
[0]—[0]—[0]—[ [а]—?—?—[е] Материал— на, 1: Dictionary of f	о]—[о]  — ? дёжные ряды:		f Matthew and Mark in th		of glossed corpus of I		Mansi language, com		r Pelym dialect of M	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]—[0]—[0]—[ [а]—?—?—[е] Материал— на, 1: Dictionary of f	о]—[о]  — ? дёжные ряды:	[ə]		[ə]		[a]		[ə]		ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]—[0]—[0]—[ [а]— ? — ? —[е] Материал — на 1: Dictionary of 1 [а]	о]—[о]  — ? дёжные ряды:	[а] ачерымъ	'mopos'	[a] atšerem	'mopos'		Mansi language, com 'Crywa'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]—[0]—[0]—[ [а]—?—?—[е] Материал— на, 1: Dictionary of f	о]—[о]  — ? дёжные ряды:	[ə]		[ə]	'мороз' 'я	[a]		[ə]		ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]—[0]—[0]—[ [а]— ? — ? —[е] Материал — на 1: Dictionary of 1 [а]	o]—[o]  — ? дёжные ряды: Mansi language (coll	[ə] ачерымъ амъ	'wopos' 's'	[a] atšerem am	'mopos' 's R'	[a]		[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]—[0]—[0]—[[ [а]—?—?—?—[е] Материал— на, 1: Dictionary of 1 [а] а́лямотъ	о]—[о]  — ? дёжные ряды:	[ə] ачерымъ амъ ал-	'мороз' 'я' 'ужирать, погибать	[a] atšerem am al-uchv	'мороз' 'я	[a]		[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]—[0]—[0]—[0]—[ [а]—?—?—[е] Материал— на, 1: Dictionary of 1 [а] а́лямотъ	o]—{o] — ? дёжные ряды: Mansi language (coll	[ə] ачерымъ амъ	'wopos' 's'	[a] atšerem am al-uchv	'mopos' 's R'	[а] Асерме	'Стужа'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Słovcov,	1905 - Le
[0]-(0]-(0]-( [а]-?-?-[е] Материал — на, 1: Dictionary of 1 [а] а́лямоть а́тлань	o]—[0] — ? дёжные ряды: Mansi language (coll Убить' 'лятьдесят	[ə] ачерымъ амъ ал-	'мороз' 'я' 'ужирать, погибать	[a] atšerem am al-uchv	'mopos' 's R'	[a]		[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]-(0]-(0]-( [а]-?-?-[е] Материал — на, 1: Dictionary of 1 [а] а́лямотъ а́тпанъ атть	о]—{о] — ? дёжные ряды: Малзі Іапдиаде (coll Уубить' 'пятьдесят лять'	[ə] ачерымъ амъ ал-	'мороз' 'я' 'ужирать, погибать	[a] atšerem am al-uchv	'мороз' 'я Я' 'погубить-INF'	[а] Асерме	'Стужа'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]-[0]-[0]-[(]- [а]-?-?-[е] Материал — на, 1: Dictionary of 1 [а] а́лямоть а́тпань атть а́тимь	о]—{о] — ? дёхные ряды: Мапзі language (coll Уубить' Унятьудесят вять' Унят	[ə] ачерымъ амъ ал-	'мороз' 'я' 'ужирать, погибать	[a] atšerem am al-uchv ati	'мороз' 'я Я' 'погубить-INF'	[а] Асерме	'Стужа'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]-[0]-[0]-[(]- [а]-?-?-[е] Материал — на, 1: Dictionary of 1 [а] а́лямоть а́тпань атть а́тимь	о]—{о] — ? дёжные ряды: Малзі Іапдиаде (coll Уубить' 'пятьдесят лять'	[а] ачерымъ амъ ал- очелне	'мороз' 'я' 'умирать, погибать погибать, умирать'	[a] atšerem am al-uchv ati at	'мороз' 'я Я' 'полубить-INP' 'нет не'	[а] Асерме	'Стужа'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]-[0]-[0]-[ [а]-?-?-[е] Материал — на, 1: Dictionary of 1 [а] а́лламоть а́тланъ атть а́тланъ атть а́тланъ	о]—{о] — ? дёхные ряды: Мапзі language (coll Уубить' Унятьудесят вять' Унят	[а] ачерымъ амъ ал- очелне аркып	'мороз' 'я' умирать, погибать погибать, умирать' священник'	[a] atšerem am al-uchv ati at at	'мороз' 'я Я' 'погубить-INF' 'нет не' 'священник'	[а] Асерме	'Стужа'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[0]-[0]-[0]-[ [а]-?-?-[0] Материал — на 1: Dictionary of I [а] алямоть а́тланъ атть атть атть атть	о]—{о] — ? дёхные ряды: Мапзі language (coll Уубить' Унятьудесят вять' Унят	[а] ачерымъ амъ ал- очелне арнып ах-	'мороз' 'я' 'умирать, погибать, погибать, умирать' 'священния' 'собирать'	[a] atšerem am al-uchv ati at at arkep- ach-	'мороз' 'я Я' 'полубить-INF' ' 'нет не' 'сбиденник' 'собирать'	[а] Асерме	'Стужа'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
[ə] а́лямоть а́тпань атть а́тимь атва́тимь	о]—{о] — ? дёхные ряды: Мапзі language (coll Уубить' Унятьудесят вять' Унят	[а] ачерымъ амъ ал- очелне арныл ах- алхлын	'мороз' 'я' погибать, погибать погибать, умирать' 'священнии' 'собирать' 'нести'	(a) atšerem am al-uchv ati at arkep- ach- alment-uchv	'мороз' 'я Я' 'полубить-INF' 'нет не' 'священник' 'священник' 'сбярать' 'нест-INF'	[а] Асерме	'Стужа'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Le
(0)—(0)—(0)—( [а]—?—?—?[е] Материал—на 1: Dictionary of 1 [а] а́лямотъ а́тпанъ атта-ка атта-ка атта-ка	о]—{о] — ? дёхные ряды: Мапзі language (coll Уубить' Унятьудесят вять' Унят	[а] ачерымъ амъ ал- очелне арнып ах-	'мороз' 'я' 'умирать, погибать, погибать, умирать' 'священния' 'собирать'	[a] atšerem am al-uchv ati at at arkep- ach-	'мороз' 'я Я' 'полубить-INF' ' 'нет не' 'сбиденник' 'собирать'	[а] Асерме	'Стужа'	[ə] а́черомъ	'Mopos, c. 17'	ansi language,	author pries	st K.Slovcov,	1905 - Lec

The tool, comparing dictionary materials by a number of parameters (initial vowel, initial consonant, vowel of the first syllable, consonant after the vowel of the first syllable, vowel of the second syllable), finds reliable, doubtful and single series of correlations. The Cognate Analysis option presents the results of comparing dialect materials in the form of a table with detailed lists of correspondences (Pic. 8-9).

Pic	8. Lists of corr	espondences by	vowels/conson	ants, reliable/ui	nreliable lines.	
- E •	<b>.</b>	· · ·	Mansi cognate analysis 2021.08.30 (3) (1) - Ex	cel	🛕 Nikita Koshelyuk 🐽 📼 —	g ×
Файл		Формулы Данные Рецензирование Вид	Справка 🛛 Что вы хотите сделать?		A	Поделиться
Вставить		= = 😸 🗞 - 🕴 Сереносить текст = = = = 💀 💀 🔛 Объединить и поместит	ъ в центре - 🧐 - % 000 % осора	Условное Форматировать Стили аатирование - кактаблицу - ячеек -	В Х А К А К А К А К А К А К А К А К А К А	
Eyşep oder		Выравнивание кя по начальному гласному:	rs Viecno rs	Стали	Вчейог Редактирование	_
AL .				-	-	
	A	В	С	D	E	
1	Соответствия по на	чальному гласному	/:			
2						
3	Надёжные ряды:					
4	[a]—[a]—[a]—[a]—	[a]				
5						
6	Сомнительные ряд	<b>цы:</b>				
7	? — ? — ? —[i]— ?	1. 				
8	? -[e]-[e]- ? -	?				
9	[o]-[o]-[o]-[o]-	-[o]				
10	[a]— ? — ? —[e]—	?				
11						
12	Материал — надёх	кные ряды:				
	Results			1		Þ

D' O Itata of oom ndancas by your ls/consonan raliable/uproliable lin



Pic 9. Lists of correspondences by vowels/consonants, reliable/unreliable lines.

Similar to the analysis presented on the LingvoDoc platform, the Excel table displays the names of dictionaries involved in the study, specific examples of words with their translation, as well as identified consonants and vocalisms (Pic. 10-11). According to the fullness of a particular dictionary, it is likely that the file will not be filled with etymologies for all lines.

Материал —	надёжные ряды:								
татериал —	паделение ряды.								
<u> </u>	of Mansi language (co		of Matthew and Mark		of glossed corpus of	Ev 4: Dictionary o	of Mansi language, co		Jpper Pelym dialect of N
5 [a]		[a]		[a]		[a]		[a]	
5		ачерымъ	'Mopo3'	atšerem	'мороз'	Асерме	'Стужа'	а́черомъ	'Мороз, с. 17'
7		амъ	`я'	am	'n			Амъ	'Я, с. 27'
5					Я'				
алямотъ	убить'	ал-	умирать, поги	δaτε al-uchv	'погубить-INF'				
		очелне	погибать, умир	ать'					
і ітпанъ	`пятьдесят					А́ть таль	'Пятилетие'		
2 аттъ	пять'								
атимъ	`нет			ati	'нет				
атва̀гмъ	не знаю'			at	не'				
5		аркып	'священник'	arkep-	'священник'				
5		ax-	'собирать'	ach-	'собирать'				
7		алхлын	'нести'	alment-uchv	'нести-INF'				
3		алислахъ	'ловить'	alisl-achv	'ловить-INF'				
9		алым	'другой'	alem	'другой'				
D		асрай	'дьявол'	asrai	'дьявол'				

Pic 10. Results of the analysis: dictionary names and identified etymologies.

Pic 11. Results of the analysis: identified phoneme series.

≂ ⊷ <b>- د+ ⊟</b>		<i>.</i>	÷	Mansi co	ognate analysis 2021.08	L30 (3) (1) - Excel			🔪 Nikita Koshelyuk 🛛 📧	m	- a	×
Файл Главная В											<b>Д</b> . Подел	
Вставить буфер обмена	й ∨ 11 ∨ А К. Ц   Ш   <u>Ф</u> , Щрефт			ь и поместить в центре	Общий - % 000 -	Условное форматирован	Форматировать	Стили неек- Янейс	ить Формат 🧳 Со	Ат ртировка Н фильтр - вы чдактировании	делить «	,
B146 - I	X 🗸 fr											,
А 121 Материал — на, 122 123, 1: Dictionary of M		C	D atthew and Mark	E	F F	G	H lansi language, con	I	J er Pelym dialect of N	K lansi langu	L age, autho	or pris
124 (m)		[m]		[m]		[m]		(m)				
125 магъ	'земля'			ma	страна	Mé	'Земля'	Мы	'Земля, с. 12'			
126 127					земля				10 10			
		машъне	`одежда'	mašne	`одежда'			Мы́ушъ	'Одежда, с. 19'			
128 минёнъ 129	'үйди'	мененъ/мынынъ	`идти′	menu-chv men-uchv	'идти-INF'							
				men-uchv				Морохъ пулъ	'Морошка ягода.	4.07		
130 морхъ 131 матимету	`морошка' `старуха					Метмъ аньчукъ	'Престарелый	морохь пуль	морошка ягода,	C. 18		
131 матимегу 132 матимъ	старуха					Метымьеку	Старуха'					
L32 матимайкъ	старыя старик'					метымьеку	старуха					
133 матимаикъ 134 ма́гылъ	'грудь'					Márns	'Грудь'					
134 манылы 135 ма́ить	'печень'					Меть	'Печень'					
136 муй	ัลดันล'					Móy	'Яйцо'					
137 ма́гы	'мёд'			mag	'mēn'	moy	ringo					
138	in the second se	мошъ	`немошь'	moš	'немошь'							
139		мошкатымъ	'одетый'	maškatem	'одетый'							
140		MAKER	'nagath'	michy	'nagath'							
141 [n]		[n]		[n]		[n]		[0]				
142 небакъ	бумага	непек	ъумага	nepek	ъумага			Нэпхъ-оксэ	Ъумажные дены	и, с. 10		
143 хонжёймъ	пишу/	ханш-	писать'	chanš-	писать'			Нэпхъ канштахтэ	м Грамоте учусь, с.			
144								Непхсуа́съ	Бумага, с. 4			
145								Нэ́пэктъ	Газета, с. 7			

Based on the analysis carried out to clarify phonetic and phonological transcription and to calculate regular correlations in words of related languages related by etymologies, at the next stage a program is carried out to determine the proximity of dialects: first, the series of multiple correspondences in the selected group of dictionaries are identified by the vowels of the first syllable and consonants of the first syllable (Pic. 12).

≂ ⊳ۍ دوا ∎									A •	likita Koshelyuk 🛛 📧	- 1	
ійл Главная Вс												А. Поделин
санон занть 💰 - Санон ж К	-> 11-> 			еносить текст идинить и помиссить і икальте		660049	Условное Ф форматирование - п Ст		Еставить Удалить Вставить Удалить Янейся		Ат / ортировка Най фильтр - выде Редистровение	
328 · I )	<ul> <li>√ fx</li> </ul>											
Δ	8	С	D	E	F	G	н	1	1	K L	м	N
Соответствия по гл	вскому первого сл	юга (после согласн	oro) (sec: 1):									
	Dictionary of Mans	The gospel of Matt	th Concordance of g	to: Dictionary of Man	si Dictionary U	oper Pelym dialect of Ma	nsi language, author	priest K.Slovcov, 1905	- Lexical Entries			
Dictionary of Mansi		12	12	19	12							
The gospel of Matti	12	0	2	5	6							
Concordance of glo	12	2	0	11	4							
Dictionary of Mansi	19	2	11	0	19							
Dictionary Upper Pe	12	6	4	19	6							
1												
Соответствия по н	чальному согласн	owy (sec: 1):										
	Dictionary of Mans	The gospel of Matt	h Concordance of a	do: Dictionary of Man	si Dictionary U	oper Perym dialect of Ma	nsi language, author	priest K.Slovcov, 1905	- Lexical Entries			
Dictionary of Mansi		0	0	7	3							
The gospel of Matth	0	6	0	7	3							
Concordance of glo		0	0	7	3							
Dictionary of Mansi	7	7	7	0	6							
Dictionary Upper Pe	3	3	3	0	0							
Суммарная матри	a:											
	Dictionary of Mans	The gospel of Matt	h Concordance of a	do: Dictionary of Man	si Dictionary U	oper Pelym dialect of Ma	isi language, author	priest K.Slovcov, 1905	- Lexical Entries			
Dictionary of Mansi		12	12	26	15							
The gospel of Matth		0	2	14	9							
Concordance of glo		2	6	18	7							
Dictionary of Mansi		14	18	0	19							
Dictionary Upper Pe		9	7	19	0							

Pic 12. Matrix of proximity of dialects by vowels and consonants of the first syllable.

According to Cognate Analysis, based on the identified series of correspondences, all dialects are compared in pairs, the number of transitions between them is calculated, i.e. a matrix of the analyzed phenomenon is formed: 1) vowels of the first syllable, 2) consonants of the first syllable, 3) sum (Pic. 13).

Pic 13. The final matrix of dialect proximity.

1819									
1820 Суммарная матр	ица:								
1821									
1822	Dictionary of Mansi	The gospel of Matth	Concordance of glo	Dictionary of Mansi	Dictionary Upper Pe	ym dialect of Mansi	language, author pri-	est K.Slovcov, 1905 -	Lexical Entries
1823 Dictionary of Man		12	12	26	15				
1824 The gospel of Mat	th 12	0	2	14	9				
1825 Concordance of g	0:12	2	0	18	7				
1826 Dictionary of Man	si 26	14	18	0	19				
1827 Dictionary Upper	Pe 15	9	7	19	0				
1828									

The correlation of reliable series of the initial position of a word allows the program to calculate the distance between the dialects represented in dictionaries and distribute them into groups. As a result, an etymological and phonetic 3D-model of the distance between the dialects of the selected dictionaries is constructed. The resulting distance represents the number of similar and different phonetic transitions between languages and dialects (Pic. 14).

Pic 14. 3D-model of the distance between dialects.



As a result, we get a semblance of a family tree, where the dictionaries involved in the analysis are marked with dots with the appropriate color.

### CONCLUSION

As can be seen from the proposed review of the implementation of the Cognate Analysis, this program allows you to process large data arrays in semi-automatic mode to establish etymological and phonological systems of dictionaries presented on LingoDoc. The advantage of working with this option is the ability to quickly and fairly reliably refine and recheck the boundaries of many phonetic phenomena, identify special features and processes that have occurred in languages over a given period of time, and so on.

#### REFERENCES

1. LingvoDoc Homepage, http://lingvodoc.ispras.ru/. Last accessed 02/06/2022