EasyChair Preprint

№ 7137

LingvoDoc: Working with Phonology

Natalia Koshelyuk

December 3, 2021

LingvoDoc: working with Phonology[1]

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 Phonology option intended for modern experimental phonological research and serving for linguists as an accurate tool for verification and compilation of reliable phonological systems of the world's languages and dialects.
Keywords. LingvoDoc, data mining, linguistics, phonology

1 INTRODUCTION

Phonology is a branch of linguistics that studies the structure of the sound pattern of a language and the functioning of sounds in the linguistic system. The basic unit of phonology is the phoneme.[2] One of the main tasks of phonology is to describe the systems of vowel and consonant sounds. Phonological analysis is the way of implementation of this task, as a result of which distinctive features are established between sounds, lists of phonemes, their classification and, subsequently whole linguistic systems are created. Traditionally, phonological analysis includes three stages: segmentation, identification and classification of phonemes. In addition, segmentation implies the division of the material (text) under study into the smallest segments, considered as phoneme realizations. Identification includes identification of some segments as the realizations

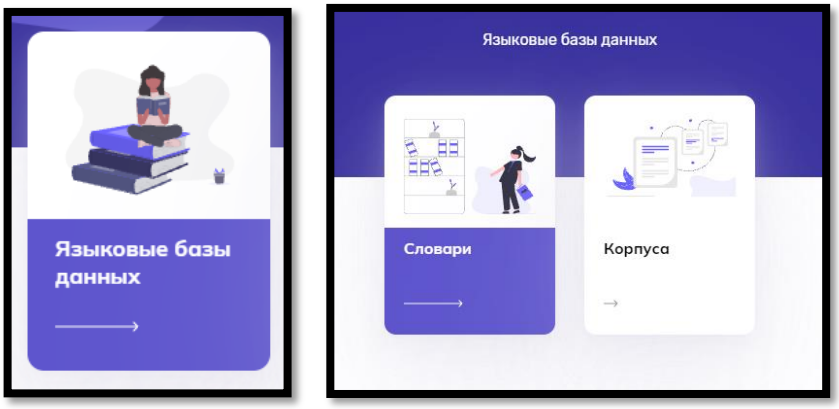
[1] Supported by Russian Science Foundation, project no. 20- 18-00403 'Digital Description of Uralic Languages on the Basis of Big Data'
[2] Phoneme – multifunctional meaning distinguishing unit of a language.

of the same phoneme. Classification includes combining phonemes into certain classes. Conduct of a full cycle of phonological analysis is often quite time-consuming and requires high accuracy.
Below we will demonstrate how it is possible to significantly reduce the resources spent by the researcher, using the built-in and automated Phonology function to perform such an analysis on the LingvoDoc platform, and what else is the value of this program for modern research.

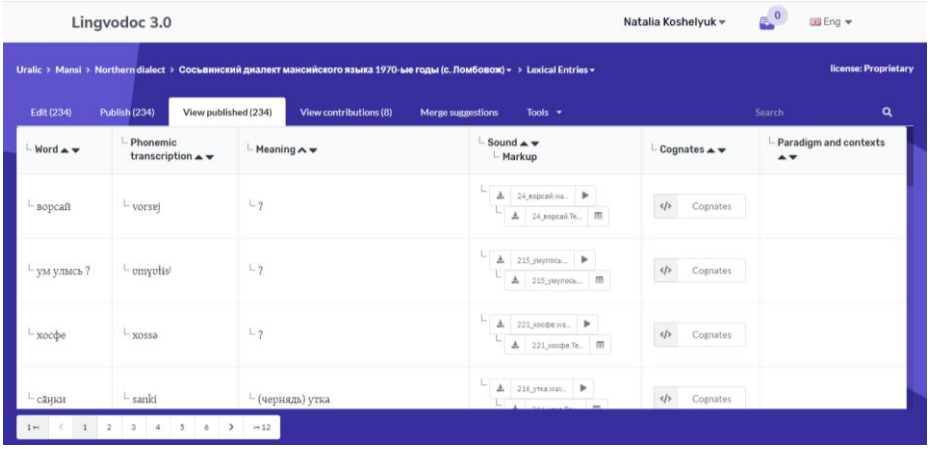
2 HOW TO CONDUCT PHONOLOGY ANALYSIS

To run the program, you need to log in to the http://lingvodoc.ispras.ru/ website and find the required source of field data by searching through all the dictionaries of the database or your own materials (Pic 1-2):

Pic 1. Transition to LingvoDoc linguistic base

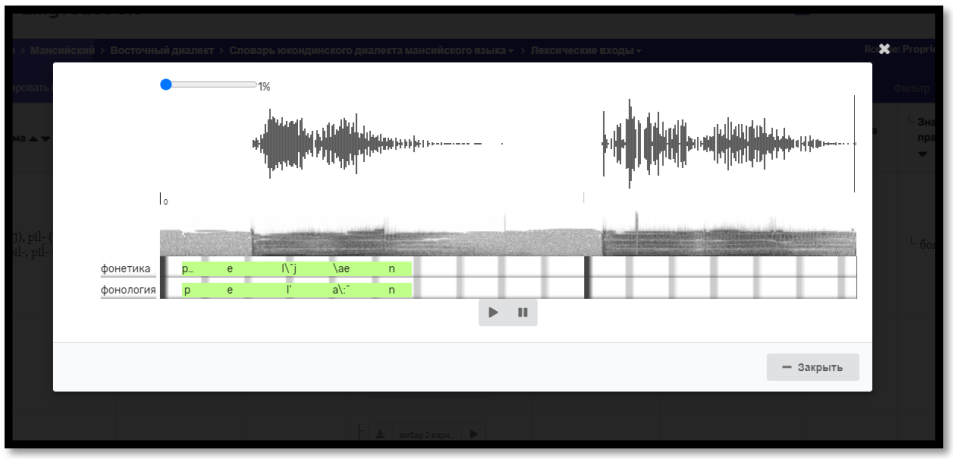


Pic 2. Opening the dictionary for which you want to do Phonological analysis



The Phonology option is designed to verify the correctness of the transcription and analyze phonetic data. Therefore, in order to work with the program, spectrograms, marked up in the special experimental phonetic software Praat (Pic 3), must be uploaded into LingvoDoc, or the audio files of field material must be automatically cut (this option has recently become available on the platform).

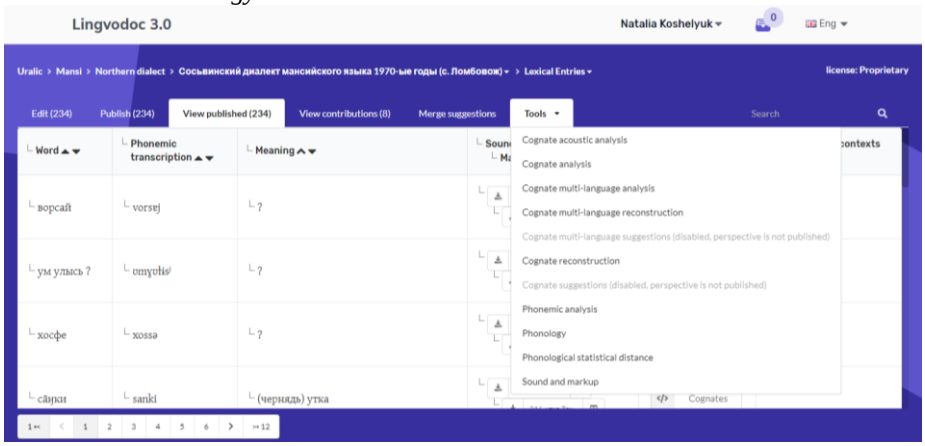
Pic 3. An example of audio segmentation



After placing all the necessary elements of the field material: a word, translation, transcription, sound marking, you can proceed to

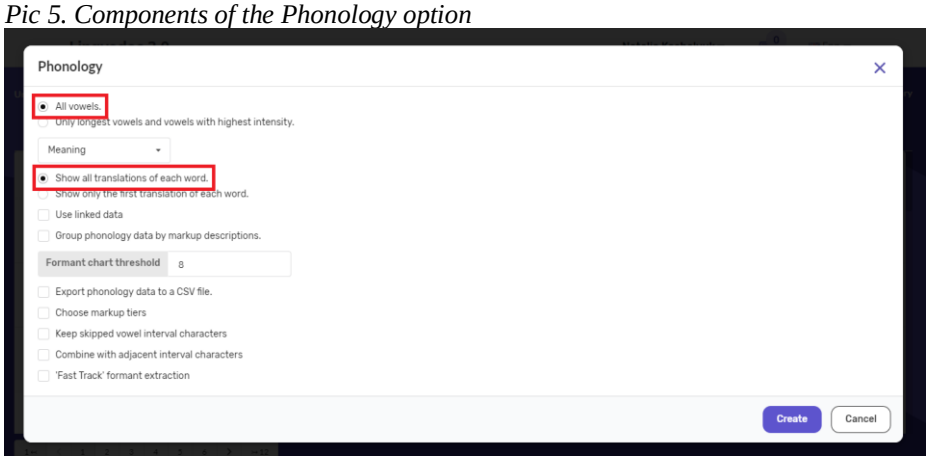
the phonological analysis. To do this, it is necessary to go to the Tools tab on the page of the dictionary under study, and select the Phonology option (Pic 4):

Pic 4. Tools. Phonology



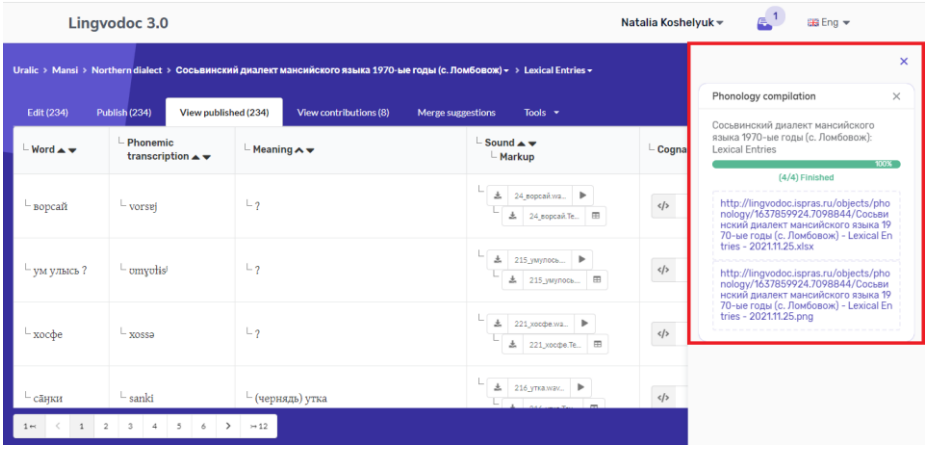
In a special pop-up window, the user can select the parameters he needs, which will be displayed in the final file (Pic 5): all vowels, only long vowels with the highest intensity, all possible translations of a word, interval combinations, etc.

Pic 5. Components of the Phonology option



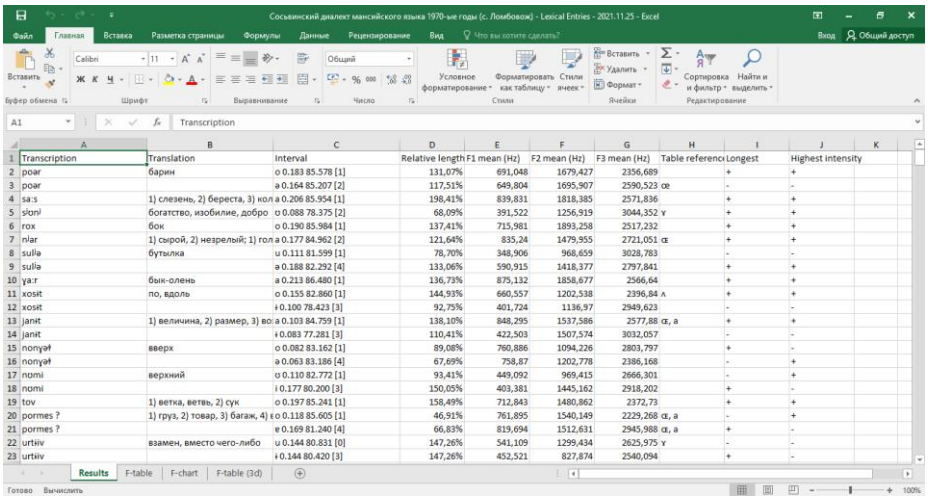
After generating the request, the file in exe. format will be uploaded to your personal account, from where it can easily be downloaded by clicking on the link (Pic 6):

Pic 6. Formation of the result



In the first tab of the resulting analysis of the source phonology, the following elements are displayed: transcription (ex. gr. *poər*, *ɣaːr*, *nʊmi*), translation (ex. gr. *gentleman*, *bull-deer*, *upper*), interval (ex. gr. *ə 0.164 85.207*, *a 0.213 86.480*, *ʊ 0.110 82.772*), relative length (ex. gr. *131,07%*, *136,73%*, *93,41%*), frequency etc. *(Pic 7)*.

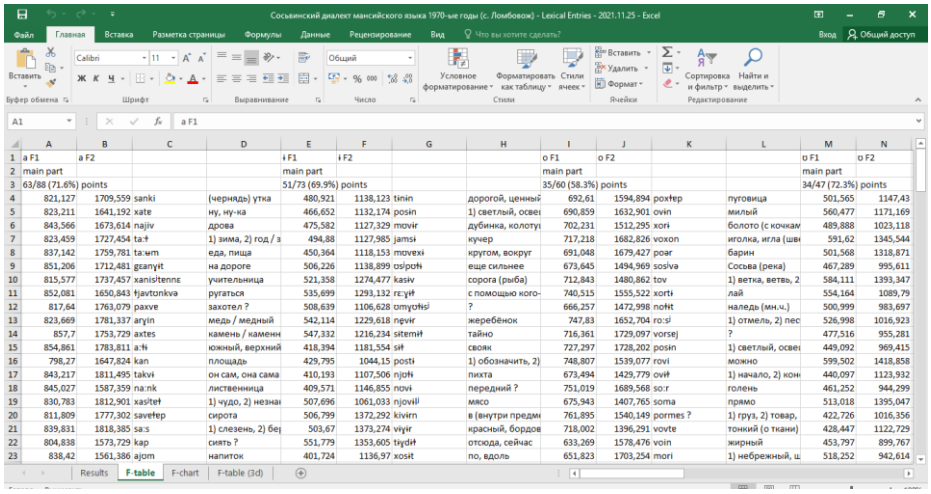
Pic 7. Results



Transcription	Translation	Interval	Relative length	F1 mean (Hz)	F2 mean (Hz)	F3 mean (Hz)	Table reference	Longest	Highest intensity
poar	барин	o 0.183 85.578 [1]	131,07%	691,048	1679,427	2356,689		+	+
poar		a 0.164 85.207 [2]	117,51%	649,804	1695,907	2590,523	œ	-	-
sa:s	1) слезень, 2) береста, 3) кол	a 0.206 85.954 [1]	198,41%	839,831	1818,385	2571,836		+	+
sʃonʲ	богатство, изобилие, добро	o 0.088 78.375 [2]	68,09%	391,522	1256,919	3044,352	ɤ	+	+
rox	бок	o 0.190 85.984 [1]	137,41%	715,981	1893,258	2517,232		+	+
nʲar	1) сырой, 2) незрелый; 1) гол	a 0.177 84.962 [2]	121,64%	835,24	1479,955	2721,051	œ	+	+
sulʲə	бутылка	u 0.111 81.599 [1]	78,70%	348,906	968,659	3028,783		-	-
sulʲə		a 0.188 82.292 [4]	133,06%	590,915	1418,377	2797,841		+	+
yə:r	бык-олень	a 0.213 86.480 [1]	136,73%	875,132	1858,677	2566,64		+	+
xosit	по, вдоль	o 0.155 82.860 [1]	144,93%	660,557	1202,538	2396,84	ʌ	+	+
xosit		ɨ 0.100 78.423 [3]	92,75%	401,724	1136,97	2949,623		-	-
janit	1) величина, 2) размер, 3) во	a 0.103 84.759 [1]	138,10%	848,295	1537,586	2577,88	œ, a	+	+
janit		ɨ 0.083 77.281 [3]	110,41%	422,503	1507,574	3032,057		-	-
nonyəł	вверх	o 0.082 83.162 [1]	89,08%	760,886	1094,226	2803,797		+	-
nonyəł		ə 0.063 83.186 [4]	67,69%	758,87	1202,778	2386,168		-	+
nomi	верхний	o 0.110 82.772 [1]	93,41%	449,092	969,415	2666,301		-	+
nomi		i 0.177 80.200 [3]	150,05%	403,381	1445,162	2918,202		+	-
tov	1) ветка, ветвь, 2) сук	o 0.197 85.241 [1]	158,49%	712,843	1480,862	2372,73		+	+
pormes ?	1) груз, 2) товар, 3) баган, 4)	o 0.118 85.605 [1]	46,91%	761,895	1540,149	2229,268	œ, a	-	+
pormes ?		ɐ 0.169 81.240 [4]	66,83%	819,694	1512,631	2945,988	œ, a	+	-
urtiv	взамен, вместо чего-либо	u 0.144 80.831 [0]	147,26%	541,109	1299,434	2625,975	ɤ	-	-
urtiv		ɨ 0.144 80.420 [3]	147,26%	452,521	827,874	2540,094		+	-

The second tab provides information about specific indicators of vowel sound formants for each of the selected phonemes and makes it possible to subsequently compare the resulting values with the IPA data: F1 shows the maximum phoneme value in Hz, F2 – the minimum (Pic 8).

Pic 8. F-Table

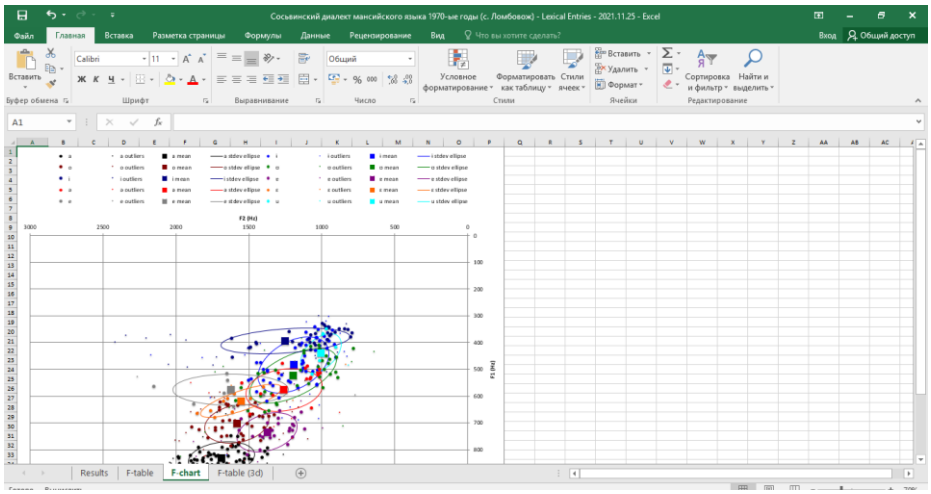


a F1	a F2			ɨ F1	ɨ F2			o F1	o F2			ʊ F1	ʊ F2
main part				main part				main part				main part	
63/88 (71.6%) points				51/73 (69.9%) points				35/60 (58.3%) points				34/47 (72.3%) points	
821,127	1709,559	sanki	(чернядь) утка	480,921	1138,123	tinin	дорогой, ценный	692,61	1594,894	poxʲep	пуговица	501,565	1147,43
823,211	1641,192	xate	ну, ну-ка	466,652	1132,174	posin	1) светлый, освеі	690,859	1632,901	ovin	милый	560,477	1171,169
843,566	1673,614	najiv	дрова	475,582	1127,329	movir	дубинка, колоту	702,231	1512,295	xorɨ	болото (с кочкам	489,888	1023,118
823,459	1727,454	ta:ł	1) зима, 2) год / з	494,88	1127,985	jamsɨ	кучер	717,218	1682,826	voxon	иголка, игла (шв	591,62	1345,544
837,142	1759,781	ta:ɯm	еда, пища	450,364	1118,153	movexɨ	кругом, вокруг	691,048	1679,427	poar	барин	501,568	1318,871
851,206	1712,481	gəanyit	на дороге	506,226	1138,899	osʲpəɲ	еще сильнее	673,645	1494,969	sosʲva	Сосьва (река)	467,289	995,611
815,577	1737,457	xanisʲtennɐ	учительница	521,358	1274,477	kasiv	сорога (рыба)	712,843	1480,862	tov	1) ветка, ветвь, 2	584,111	1393,347
852,081	1650,843	ɬjavtonkva	ругаться	535,699	1293,132	rɛ:yɨł	с помощью кого-	740,515	1555,522	xorti	лай	554,164	1089,79
817,64	1763,079	paxve	захотел ?	508,639	1106,628	omyʊ̈nʲsi	?	666,257	1472,998	noʲkt	наледь (мн.ч.)	500,999	983,697
823,669	1781,337	aryin	медь / медный	542,114	1229,618	ŋpvir	жеребёнок	747,83	1652,704	ro:sʲ	1) отмель, 2) пес	526,998	1016,923
857,7	1753,729	axtes	камень / каменн	547,332	1216,234	sitemʲł	тайно	716,361	1729,097	vorseɲ	?	477,516	955,281
854,861	1783,811	a:ɬ	южный, верхний	418,394	1181,554	sił	своян	727,297	1728,202	posin	1) светлый, освеі	449,092	969,415
798,27	1647,824	kan	площадь	429,795	1044,15	postɨ	1) обозначить, 2)	748,807	1539,077	rovi	можно	599,502	1418,858
843,217	1811,495	takvɨ	он сам, она сама	410,193	1107,506	ŋjoɲ	пихта	673,494	1429,779	ovɬ	1) начало, 2) кон	440,097	1123,932
845,027	1587,359	na:nk	лиственница	409,571	1146,855	novi	передний ?	751,019	1689,568	so:r	голень	461,252	944,299
830,783	1812,901	xasʲtəł	1) чудо, 2) незна	507,696	1061,033	njovilʲ	мясо	675,943	1407,765	soma	прямо	513,018	1395,047
811,809	1777,302	savɛtɐp	сирота	506,799	1372,292	kivirn	в (внутри предм	761,895	1540,149	pormes ?	1) груз, 2) товар,	422,726	1016,356
839,831	1818,385	sa:s	1) слезень, 2) бе	503,67	1373,274	viyir	красный, бордов	718,002	1396,291	vovte	тонкий (о ткани)	428,447	1122,729
804,838	1573,729	kap	сиять ?	551,779	1353,605	tiydʲł	отсюда, сейчас	633,269	1578,476	voin	жирный	453,797	899,767
838,42	1561,386	ajom	напиток	401,724	1136,97	xosit	по, вдоль	651,823	1703,254	mori	1) небрежный, ш	518,252	942,614

The F-Chart tab (Pic 9) contains the generalizing results of the performed analysis and demonstrates the resulting phonetic inventory

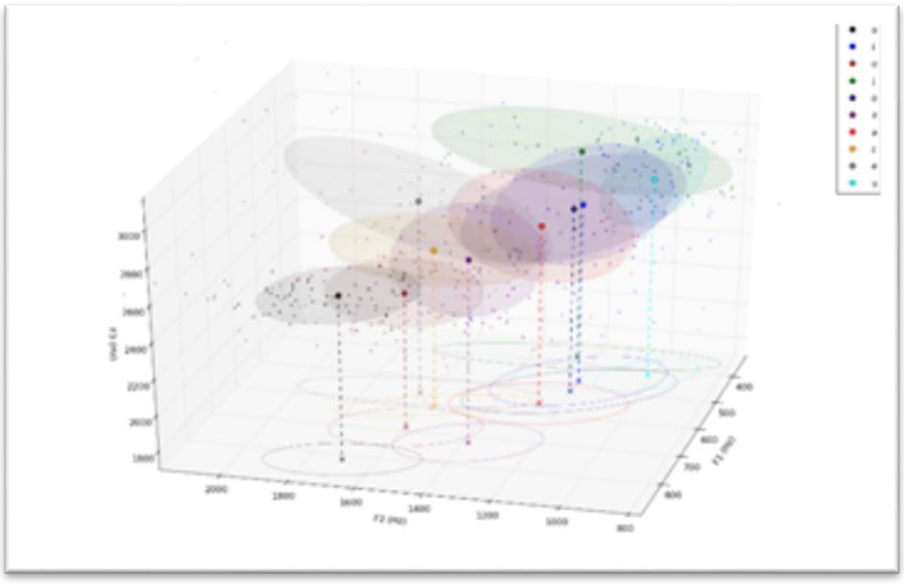
of the dictionary under consideration. The resulting 3D model makes it possible to verify the correctness of the selection of a particular formant generated by the LingvoDoc program at the previous stages of the study.

Pic 9. F-Chart



It is possible to speak about the legitimacy of distinguishing formants of a similar nature of formation (row, rise, position) and sounding on the basis of the overlapping degree of their clouds. As an illustration, we give an example of the correct transcription, when the formants of clouds of different vowels overlap by less than 30%, and incorrect one, when the formants overlap by more than 30% (Pic 10-11):

Pic 10. Correct transcription



Pic 11. Incorrect transcription

CONCLUSION

Thus, the use of the Phonology program allows linguists to conduct experimental phonological research on the basis of LingvoDoc in order to check and verify the phonetic systems of various languages, as well as to clarify and reverify the results of previous years' research on the basis of modern field material.

This function is very important for the study of dialects and accents of insufficiently studied as well as endangered languages because, as the results of already conducted research show, there are often discrepancies between the phonology of such languages and the IPA data used for comparison as the most studied and complete, which testifies the uniqueness of these systems.

REFERENCES

1. Russian Phonetics and Phonology, https://www.philol.msu.ru/~fonetica/index1.htm . Last accessed 10/29/2021
2. LingvoDoc Homepage, http://lingvodoc.ispras.ru/. Last accessed 11/22/2021