

A never-published atlas of Udmurt dialects

Fejes, László

Hungarian Language Center for Linguistics

fejes.laszlo@gmail.com

Abstract

In the first decade of the 21th century, an atlas of Udmurt dialects was prepared for publication. Although hundreds of maps and legends were completed, due to no hope for publication, the project was never finished. The paper describes the material the atlas was based on, how the collection of exercise books was digitized and prepared for the purpose of a dialectal atlas, and how the atlas was generated from the data. The paper also presents some decisions that had to be made during the preparation of the atlas. Finally, the never-published atlas is compared to the published atlas of Udmurt dialects. Despite that the history of the atlas is far from a success story, it shows that, if data are available, a linguistic atlas can be produced even using low-budget tools, in a do-it-yourself way.

Пуштроез

Кызь одйгетй даурлэн нырысетй аръёсаз удмурт диалектъёсья атлас поттыны дасямын вылэм. Кӧня ке сю карта но солы символъёсын валэк-тонъёс лэсьтэмын вылэм но, сое поттыны осконлык ышем бере, уезз пумозяз вуттиллямтэ. Та статья маде, кыӵе материал-тодэтъёс вылэ пыкъясь-кыса атлас лэсьтэмын вылэм, кызы но кыӵе тетрадьёс та атласлы шуы-са дасямын но дигитализировать каремын вал. Статьялэн пумаз поттымтэ атлас мукетыныз, удмурт диалектъёсын поттэм атласэн ӵошатэмын. Атласлэн историез азинэс ой вал ке но, со возматэны быгатэ: тодэтъёс вань дыръя кылъя атласэз, дунтэм тйрлыкъёсты уже кутыса но, “киуж амалэн” дасяны луэ.

1 Introduction

Usually, IWCLUL papers present current achievements in computational approaches to Uralic languages. This paper is exceptional in the sense that it presents a more-than-a-decade-old project, which got stuck in its final phase, although it could have produced an (almost) unprecedented result: an atlas of the Udmurt dialects (based on its working title, *Удмурт вераськетъёсья атлас*, henceforward УВА). The word “almost” indicates that the first volume of another atlas of Udmurt dialects (Насибуллин et al. (2009), henceforward ДАУЯ) was published approximately at the same time when the discussed atlas should/could have been published. Interestingly, the two atlases are so different in their aims and methods, that they cannot even be considered competitors. As ДАУЯ was the first atlas of the Uralic languages of the Russian

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

Federation, it could have been an interesting situation that Udmurt, the only Uralic language with a dialect atlas, could have immediately two of them.

The reasons for the project got stuck are complex. First of all, there was no hope to get financial support for publication. Online publication in PDF format was out of question for several reasons. The main reason is that if the atlas is available online, it is even more hopeless to get financial support to publish it in print. The prestige of an online publication is much lower even today than the prestige of a publication in print, and the difference was even more considerable more than a decade ago. It seemed reasonable to wait for better circumstances. Moreover, the author had permission from the Department of General and Finno-Ugric Linguistics of the Udmurt State University to use the data collected by them for the purpose of publishing a printed atlas. For the same reason, the publication of the bare database was also out of the question. In addition, the author had to leave academia in 2010 and worked outside academia for a living, without time and force to work on the atlas, including search for financial support for publication. When the author could return to research in 2016, he had very different tasks and could find time at least to document the former project only recently.

Section 2 presents where the idea for YBA came from. In Section 3, it is described how a digitally processable data set was produced from the available material. Section 4 discusses the way of generating an atlas from these data. Section 5 outlines the differences between YBA and ДАЯЛ. Section 6 contains some thoughts on the possible future of the YBA project.

2 Background

Ever since the middle of the 1980s, the students of Udmurt philology at the Udmurt State University have had to collect dialect materials from their home village in the second year of their study, and almost every year, they have gone to an expedition together at the end of the year to collect similar materials. The collected material, hand-written into exercise books, consists of two types: texts and the answers to a questionnaire, which will be shortly presented in 2.1. In 2004, the author learnt that a large number of questionnaires were stocked in the rooms of the Department of General and Finno-Ugric Linguistics of the Udmurt State University, not used for any linguistic purposes. Despite that the reliability of the material can be questioned (see 2.2), the author thought this collection was too valuable to be left untouched. The most straightforward idea was to make a dialect atlas based on the material. To make an atlas, survey sites have to be chosen — in 2.3, the applied method will be presented.

2.1 The 400 word program

The questionnaire mentioned above was put together by Valentin Kelmakovich Kelmakov, a (if not the) leading specialist of Udmurt dialectology. It is difficult to determine when the survey was assembled or first published, but the earliest exercise book with the questionnaire is dated to 1983. For the atlas, Кельмаков (2002) was used as a reference.

The questionnaire consists of 400 questions (the name *the 400 word program* — Udmurt 400 кылъем программа — comes from this). In most of the cases, the fieldworker says a Russian word or phrase, which the consultant has to translate into Udmurt. The fieldworker has to try to find a form relevant to the phenomenon the

question serves to observe. E.g. question 221 should observe the use of affricates in the given dialect, and asks for the word ‘good’ (‘добрый, хороший’): /dʒetʃ/ ~ /dʒecʃ/. However, in some dialects, this word is absent or used in a very restricted way, in some greeting forms. If the consultant answers with another word meaning ‘good’, the fieldworker has to record the given form but also to try to ask for synonyms, or ask for the greeting forms containing the searched word. Of course, these rules are applicable for the questions on phonological and morphological phenomena, but not for lexical questions. In addition, there are also semantic questions, when the consultants are given an Udmurt word and they have to translate it into Russian.

The questionnaire consist of 309 questions on phonology,¹ 69 questions on morphology,² 18 questions on the lexicon (vocabulary) and 4 questions on semantics. In addition to the 400 questions, the fieldworker has to record 12 paradigm forms (present tense, positive and negative 1SG ...3PL forms) of two verbs (*тодыны* ‘to know’, *кутскыны* ‘to begin’), i.e. there are 24 additional questions in the questionnaire.

2.2 The material collected

Between 1983 and 2004, more than 3000 exercise books were filled with answers on the questionnaire. The material is geographically unbalanced: since most of the students come from Southern Udmurtia, especially from the environs of Izhevsk, it is not rare that there are more than five, sometimes more than a dozen surveys from the same settlement. Northern Udmurtia is much less documented, while data from the dialects outside Udmurtia are rather sporadic.

In addition, the quality of the data is sometimes questionable. Data were collected by students, not professional fieldworkers. Theoretically, they are checked by the teachers, but, on the one hand, some exercise books seem to be unchecked (or checked but not corrected); on the other hand, for lack of sound recordings, the teachers cannot check whether the written data correspond to the answers given by the consultants. In some cases, it is clear that the student did not understand the task (the recorded answers are irrelevant to the studied phenomenon), or could not consistently record the data. A typical case is when in the answers to the first questions, which aim to reveal whether the dialect has *ɨ* /ɨ/, *ɨ̃* /ɨ̃/ or *ɨ̄* /ɨ̄/, *ɨ* or *ɨ̃* is recorded in all the cases, but later, in answers to other questions, only *ɨ* occurs. In addition, some exercise books are clearly copied from others (self-evidently, these were not used for the project), and it is possible that in some other cases, copying is not so conspicuous. Nonetheless, basically the material seems to be reliable. Data from the same settlements usually show more differences than one would expect if students simply copied the exercise books from each other; however, they are quite consistent to be done at random. Since the students usually document their own dialect, in a certain sense, they are more competent fieldworkers than well-trained but outsider linguists.

¹In fact, in the Udmurt and Russian texts, they are called *phonetic* issues, but it seems that Udmurt (Russian?) linguistic tradition does not always make such a strict distinction between phonetics and phonology as the western one. In any case, most of the problems observed by these questions should be classified as phonological in the western tradition.

²In many cases, these are rather (morpho)phonological questions related to certain suffix morphemes.

2.3 Preparing for an atlas: the choice of survey sites

Theoretically, all of the documented settlements could have been survey sites of the atlas. This choice could have had two disadvantages. First of all, all the available data should have been digitized, although many of these are redundant, because they come from neighbouring settlements without considerable linguistic differences. In addition, too dense survey sites make the map less readable. Moreover, since different areas are documented at a different level, in some areas survey sites could have been dense, while in other areas sparse. Even worse, the density differences would have reflected the number of the students from the area, not the number of the Udmurt settlements (or speakers).

Therefore, a rectangular grid was formed on the map with squares about 15×15 kms. Each square got a two-character code: the first character (a–v) showed its latitude (a is the southernmost, the latitude of Naberezhnye Chelny, while v the northernmost border of Udmurtia), the second one (A–N) showed its longitude (A is the westernmost, N is the easternmost border of Udmurtia). For each square, one representative settlement was chosen, usually the one which was documented by the most surveys. Minimally two filled questionnaires were needed to appoint a survey site for the atlas. Unfortunately, in some cases, the chosen villages, although belonging to different squares of the grid, are quite close to each other, while some territories seem to be uncovered. Finally, 81 survey sites were chosen in the territory of Udmurtia. Later, two survey points were added from the Kirov Oblast (since here there were no settlements documented by two questionnaires, the data from two different but close villages were contracted in both cases) and two from Tatarstan, represented on the same map. In addition, nine survey points were added from Tatarstan, five from Bashkortostan and one-one from the Mari El and the Perm Oblast (nowdays Krai), respectively, which were represented outside (under) the map .

Every survey site had a four-character location code consisting of two letters and two digits. The first two characters showed which grid square it belonged to. In the case of the sites represented outside the map, their first character was x, the second corresponded to their relative position as they are represented under the map, which more-or-less reflected their relative longitudinal position, but ignored the actual distances. The last two characters were digits, and they reflected the relative position of the site in the grid square. Every square was divided into nine equal numbered (5×5 km) squares: 5 was the central square, 1 is the northwestern and 9 is the south-eastern corner. The third character reflected in which ninth the site lies in. In a similar way, every 5×5 km square was divided into nine squares, and the position of the site was specified further by the fourth character. This way, every site could be located with 1–2 km accuracy (see Table 1).

3 From exercise books to data

After the survey sites had been chosen, the material of the exercise books had to be digitized. Each exercise book was represented by one text file, containing exclusively ASCII characters. The data (and the metadata) were simply typed in by the author of the current article. The data were usually written in a well-readable hand, in addition, as the possible answers to the questions formed an almost closed set, it was usually relatively easy to find out what had been intended by the fieldworker. On the contrary, metadata were sometimes written in a hardly readable cursive, and it was difficult to

11	12	13	21	22	23	31	32	33
14	15	16	24	25	26	34	35	36
17	18	19	27	28	29	37	38	39
41	42	43	51	52	53	61	62	63
44	45	46	54	55	56	64	65	66
47	48	49	57	58	59	67	68	69
71	72	73	81	82	83	91	92	93
74	75	76	84	85	86	94	95	96
77	78	79	87	88	89	97	98	99

Table 1: The place of the survey site further specified by two numbers inside the territory specified by two letters

<i>a</i>	a	<i>e</i>	e	<i>u</i>	i	<i>o</i>	o	<i>y</i>	u
<i>я</i>	ja	<i>э</i>	\e	<i>ы</i>	y	<i>ё</i>	jo	<i>ю</i>	ju
<i>б</i>	b	<i>в</i>	v	<i>г</i>	g	<i>д</i>	d	<i>ж</i>	zh
<i>з</i>	z	<i>й</i>	j	<i>к</i>	k	<i>м</i>	m	<i>н</i>	n
<i>п</i>	p	<i>р</i>	r	<i>с</i>	s	<i>т</i>	t	<i>ф</i>	f
<i>х</i>	kh	<i>ц</i>	c	<i>ч</i>	ch	<i>ш</i>	sh	<i>щ</i>	sch
<i>ь</i>	'	<i>ъ</i>	"						

Table 2: The transcription applied in the text files for metadata and meaning given in Russian

find out what is intended to be written (especially with personal names).

The files began with the metadata: every line contained one piece of data, beginning with the data identifier (field name), followed by a colon and the data. The identifiers were abbreviations based on Udmurt phrases, e.g. *gunim*: the name of the village in Udmurt (*гуртлэн удмурт нимыз*), *infvar*: the year of birth of the consultant (*информантлэн вордскем арез*), *ljuk*: the collector (fieldworker) (*люкась*) etc. For the transcription for the metadata, see Table 2.

The linguistic data followed the metadata. Every line contained a three digit code of the question and the answer, separated by a space. The paradigm forms for *тодыны* ‘to know’ and *кутскыны* ‘to begin’ were numbered 400–424.

The Cyrillic-based transcription used in the exercise books was transliterated to a specific code inspired by the Prószyński code. The Prószyński (named after its inventor, Gábor Prószyński)³ is an ASCII-based code developed originally for Old Hungarian texts. The basic idea is that every letter missing from the English alphabet is encoded with a combination of a letter and one or two digits, e.g. *á*: a1, *ö*: o2, *ó*: o3, *č*: c12, *æ*: a36, *f*: s43, *δ*: d50, etc. In the transcription applied (see Table 3), a Roman letter or a Roman letter and a digit corresponds to the original Cyrillic letter. However, there are also some exceptions, e.g. some digits (8, 9) correspond to letters themselves, some punctuation marks (", .) are also applied (since the data are words or, rarely, phrases, these are not needed otherwise), and some other marks are also used (% , ').

If the lack of a form was indicated in the exercise book in any way, a mark hyphen (–) was typed into the place of the data. If the form occurred just in a given phrase

³The first description of the transcription can be found in the unpublished manuscript Prószyński (1985). The earliest use of the term *Prószyński code* (*Prószyński-kód*) is attested in Kornai (1985).

<i>a</i>	a	<i>a</i> ^o	a0	<i>ä</i>	a2	<i>u</i>	i	<i>û</i>	i6
<i>u</i>	i3	<i>o</i>	o	<i>o</i> _λ	o1	<i>ó</i>	o6	<i>ö^b</i>	o3
<i>ö</i>	o2	<i>ö</i> ^o	o4	<i>ö</i>	o5	<i>y</i>	u	<i>ÿ</i>	u2
<i>y</i>	u3	<i>ÿ</i>	u5	<i>yl</i>	y	<i>yl^o</i>	y0	<i>ÿi</i>	y2
<i>yl</i>	y3	<i>ÿi</i>	y7	<i>z</i>	9	<i>z^b</i>	93	<i>z</i>	8
<i>z</i>	e	<i>z</i> _λ	e1	<i>z</i>	e3	<i>z</i>	e3	<i>z</i>	%
<i>β</i>	b	<i>β</i>	v	<i>β</i>	W	<i>w</i>	w	<i>ÿ</i>	u7
<i>z</i>	g	<i>ð</i>	d	<i>Д</i>	D	<i>ð'</i>	d1	<i>ð^β</i>	d5
<i>ж</i>	zh	<i>ж</i>	xh	<i>ж'</i>	x4	<i>z</i>	z	<i>z'</i>	z1
<i>z''</i>	z''	<i>z</i>	x	<i>z''</i>	x''	<i>ÿ</i>	j	<i>ÿ</i>	j7
<i>κ</i>	k	<i>κ</i>	k3	<i>л</i>	l	<i>л'</i>	l1	<i>l</i>	lh
<i>м</i>	m	<i>м</i>	m7	<i>н</i>	n	<i>н'</i>	n1	<i>η</i>	q
<i>н</i>	n.	<i>n</i>	p	<i>p</i>	r	<i>c</i>	s	<i>c'</i>	s1
<i>c''</i>	s''	<i>c'</i>	s6	<i>m</i>	t	<i>m'</i>	t1	<i>mⁿ</i>	t5
<i>m</i>	t7	<i>φ</i>	f	<i>x</i>	X	<i>ç</i>	C	<i>ç'</i>	C1
<i>ç</i>	c	<i>ç'</i>	c1	<i>ç</i>	c7	<i>ÿ</i>	ch	<i>ÿ'</i>	c4
<i>ÿ</i>	c6	<i>u</i>	sh	<i>u</i>	s1s1	<i>'</i>	'		

Table 3: The transcription applied in the text files for Udmurt dialect data

(as /d̥z̥et̥/ ‘good’ in /d̥z̥et̥ lu/ ‘good bye’), the phrase was presented after a backslash (\).

If a synonym was given instead of the expected form, the hyphen was followed by an equals sign (=) and then came the synonym. If the answer was missing (but the lack of the asked item was not indicated), a question mark (?) was written. Any evidently wrong data were written following a question mark as well.

If there were more variants given to the question, they were separated by a comma (,). If the meaning of the word was given in Russian in the exercise book, it was encoded following a hashmark (#) in the transcription similar to the one used for metadata. If the verbal paradigm forms contained a personal pronoun as well, they were written after the verb form separated by an at sign (@).

4 From data to atlas

The idea was to generate an atlas from the text files as automatically as possible. It is important, because this way a new version of the atlas can be done any time (after correcting mistakes, adding new files or even survey sites, changing the way of data representation, the structure of the atlas, etc.). Therefore, a modular process was designed, in which a Unix shell script managed the whole process (all the work was done in Linux), calling Perl scripts and using simple shell commands (such as `uniq` and `sort`).

In principle, the basic task was rearrangement. While the source text files contained the answers given at one survey site at one occasion, in the atlas, answers given to the different questions had to be represented on a different map each; on each map, data for each survey had to be presented, grouped due to the survey sites, even similar data for the same survey site must be grouped together (symbolized by the same sign). In addition, for every map, each type of data must be associated with a map sign (manually, at least for the first time), and for each map, a legend must be

generated, which must enlist all the used signs and all the data they are associated with.

The result was a \LaTeX source file, which had to be compiled by \LaTeX , and the DVI file could be converted to PostScript or PDF, which was ready for printing. An example of an atlas map is presented in Figure 1.

It must be stressed that flexibility is an essential property of the whole approach to the atlas. This means that most of the things done in a particular way could have been done in a different way. However, the description of the decisions made can also demonstrate the possibilities.

The *processed material* was restricted to the first 396 questions of the questionnaire. These ask for an Udmurt equivalent of a Russian word or phrase, i.e. the answer is an Udmurt word (or phrase). Questions 397–400 ask for the meaning of a given Udmurt word, that is, the answer is a Russian word (or phrase). Therefore, a different code is needed to process these answers, the coding of which was delayed, and later, seeing no hope for publication, the needed script was never written. The maps for the paradigm forms of two verbs (*тодыны* ‘to know’, *кутскыны* ‘to begin’) were omitted for a different reason. While all other questions are targeted to explore a given dialectal phenomenon, in these cases, there is no explicit problem the data should answer to. The maps could have been done from several standpoints, but asking novel research questions was out of the scope of the project; therefore, these maps were not prepared. (Representing all variants on the map had no sense, see below.)

The atlas basically consisted of the maps and the legends belonging to them, there were no accompanying comments. Despite that, the atlas had a title and contained some texts; therefore, the *language* of it had to be chosen. It was decided that the atlas will be bilingual: Udmurt for the sake of the language community and English for the international public.

Since there were 396 maps derived and all of them had to have a separate legend (although theoretically the legend could have been placed on the map, for the sake of readability and for aesthetic reasons, this solution was rejected), the length of the whole atlas was about 800 pages. Moreover, the legend sometimes was much shorter than a page, sometimes it exceeded a page length. In addition, it had to be prepared for the addition of explanations to each map. Since every map begins a new page, and each map should be presented on the same side, it could take up very much place. Therefore, it was decided that the map will be presented in two volumes: the first contains the maps, the second one contains the legends (and, desirably, the explanations in the future).

As it was mentioned above, the *maps showed the territory* of Udmurtia, and the Periferic Southern Dialects (PSDs), spoken farther from Udmurtia, were represented under the maps. This solution was chosen because if the PSDs had been represented on the map, the territory of Udmurtia would have been overly compressed. Moreover, PSDs are relatively dispersed, and their representation on their exact place does not add much to our understanding of the dialectal distribution of the given phenomena. In addition, PSDs are poorly represented in our material. Nonetheless, it would have been possible to represent every survey site on their exact place. Similarly, it would have been possible to “magnify” any territory on the map and examine the isoglosses more closely (especially where more survey sites could be added).

The survey sites are not represented on an exact geographical map as a background. As orientation points, six significant settlements of Udmurtia (Izhevsk, Glazov, Votkinsk, Sarapul, Mozhga, Igra) and Agryz (which belongs to Tatarstan, but whose area protrudes into the territory of Udmurtia) are indicated. In addition, the north-

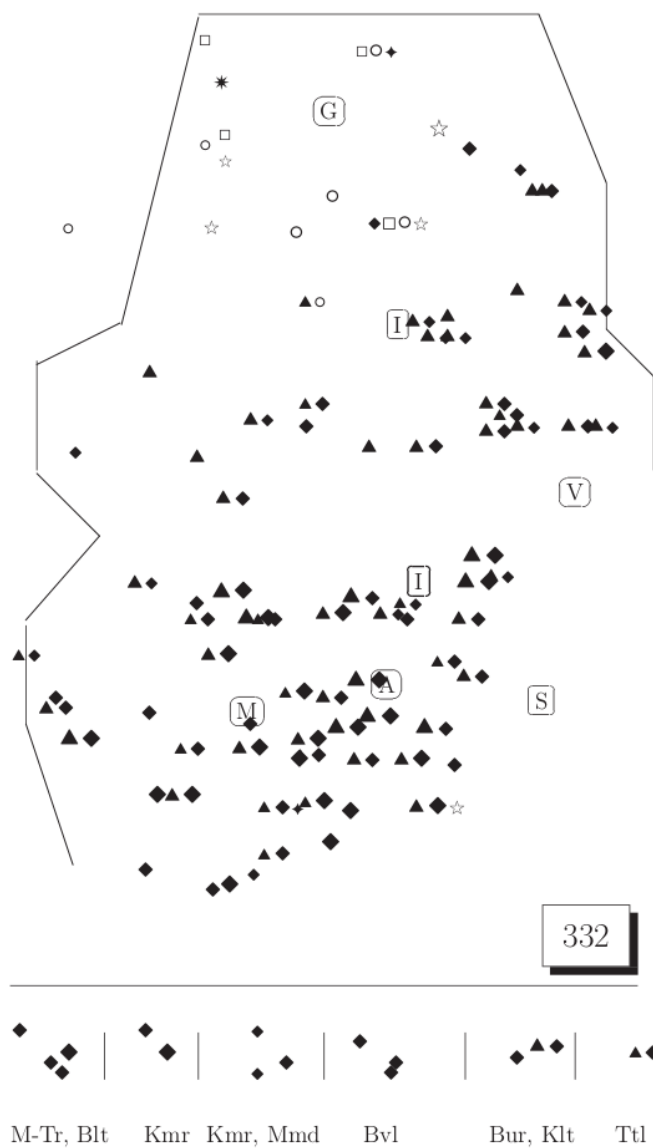


Figure 1: A map from the atlas (332. 'по обрары' – 'along/through the ravine', for the legend, see Figure 2)

eastern, northern and western borders of the Udmurt Republic is also schematically represented.

Many dialect atlases tend to show just one form (meaning, etc.) for one survey site. However, our everyday experience shows that dialects and even individual speakers exhibit variability. Variability can be attested in the material of the atlas as well. In many cases, data collected by different fieldworkers and from different consultants differ; the exercise books sometimes contain more possible answers for the question. It was decided that the atlas should reflect the *local diversity* of dialects; therefore, all data must be represented. However, since the dialect of some sites are documented in more than a dozen exercise books, it makes no sense to put all data on the map. On the other hand, if a form is documented seven times at a site, and another only once, it would be misleading to represent them in the same way. Therefore, if a piece of data on a site occurred just once, it was smaller, if more than four times, bigger than the sign for two to four pieces of data.

However, only *relevant diversity* was reflected on each map. That is, if the question of the questionnaire asked for the quality of a consonant, the differences of vowels were not reflected by the signs. On the one hand, this is practical for the sake of readability; on the other hand, it helps to eliminate the errors of the fieldworkers similar to the one mentioned in Section 2.2.

For the sake of keeping the printing cost low, *no colours* were used in the atlas. The signs chosen to represent the data were taken from the `MnSymbol` package of \LaTeX : triangles (turned into different directions, filled and unfilled), squares, diamonds, circles (all filled and unfilled, containing different patterns), stars (asterisks, different forms and number of points). The signs were chosen in a way that their similarities could reflect the similarities of the linguistic data (e.g., data represented by filled triangles and filled circles resemble one another in a way; while data represented by filled triangles and unfilled triangles are similar in another way.).

The linguistic data are presented in three *transcriptions*: in Cyrillic based Udmurt dialect transcription – see Кельмаков (1998, 44–50) or Кельмаков (2002, 49–56) – for the language community, IPA for the international audience, and Finno-Ugric transcription for western traditional Finno-Ugrists. An example is represented on Figure 2.

5 The differences between the two dialect atlases of Udmurt

An important difference is that while ДАУЯ aims to present a full and balanced picture of the Udmurt dialects, the purpose of УВА is to make use of an incomplete and unbalanced, but already existing collection. Moreover, this collection is constantly growing, and hopefully will grow until Udmurt is spoken or Udmurt philology is taught at the Udmurt University.

By digitizing new exercise books, new survey sites can be added, and the maps can be completed with data for the territories undocumented up to this point relatively easily. On well-documented areas, the survey sites can be made more dense, and more detailed maps of these territories can be produced by relatively small modifications of the scripts. The data for УВА have been collected during a long period (at the time of the preparation, about twenty years, but since then more than thirty years), that is different survey sites can be represented by data from different times. However, on

332-тӱ карта / map 332

- ▲ *н'укэти, н'уктэти*
пукети, пуктети
н'укети, н'уктети
- ◆ *н'ӱти, н'укти, н'ӱти*
пӱ?ти, пукти, пу?ти
н'ӱти, н'укти, н'ӱти
- *н'укэм'у*
пукечи
н'укет'и
- *н'укэсм'у, н'укӱсм'у, н'укым'у*
пукӱси, пукӱси, пукиси
н'укэ'и, н'укӱ'и, н'укис'и
- ☆ *н'укм'у*
пукци
н'укт'и
- * *н'укки*
пукки
н'укки

Figure 2: An example of the legend (332. ‘по овргары’ – ‘along/through the ravine’, for the map, see Figure 1

well-documented territories, it might be possible to do longitudinal analysis and to reveal linguistic change.

Unfortunately, some territories, especially outside Udmurtia and Northern Udmurtia, are poorly represented by УВА. Although the number of the blank spots can be decreased, especially by organizing “expeditions” to these territories, it is a cost-sensitive and time-consuming issue. Moreover, even if special attention is paid to the less-documented areas, the documentation level of different territories will never be balanced. On the contrary, there are no similar problems with ДАУЯ.

From an aesthetic point of view, УВА falls short of ДАУЯ, and even the exact identification of the survey sites on the map is challenging. However, the main purpose is not documenting the survey sites, but to give a general impression on the distribution of certain forms.

The most important difference is that while УВА presents the distribution of phonological and morphological features, and only minimally considers lexical differences, ДАУЯ deals exclusively with lexical issues. As a consequence, the two atlases complement each other, and together they can provide a more complete picture of the Udmurt dialects.

Finally, УВА was prepared based on an existing material and using low-budget tools in a do-it-yourself way. Evidently, the circumstances have changed since the atlas was made, and many things should be done in a very different way today (and even could have been done better at that time). But as this case study shows, making an atlas is not an unachievable purpose even for individual researchers if the linguistic material is available.

6 The future

The simplest way to finish the project would be to find financial support for a printout version, generate the missing maps, possibly improve the appearance, and publish the atlas. However, knowing the circumstances, this scenario seems to be unrealistic.

If we think about online publication, publication in a PDF format is not expedient. It would be much more reasonable to take advantage of the opportunities offered by technology, and to publish maps in an interactive format (e.g. based on OpenStreetMaps), when the user can zoom in and out depending whether they are interested in a specific territory or the general view. However, this would need a completely new way of generating maps from the data, although based on the same principles.

Nonetheless, such a decision is quite risky because of the fast change in technology. For example, in 2006 it could seem a good idea to publish an atlas on CD ROM, which, depending on the technological details, could be completely unusable today. While the preservation of printout books has established standards, web sites easily perish and vanish from the Internet, especially when nobody is involved in them in the hosting institute, if not as an author, at least as a user. As a consequence, online-only publication is not always a completely responsible decision even today.

Acknowledgements

The development of the atlas was financed by the project *NKFI K48309 Permian linguistic databases* led by Sándor Csúcs. I thank the colleagues at the Department of General and Finno-Ugric Linguistics of the Udmurt State University, especially Valentin Kelmakovich Kelmakov, for their support. I am grateful to my Izhevsk family of that time for the ideal circumstances I could work in. I am also thankful to Nóra Wenszky for her assistance in preparing this paper and her useful advice.

References

- András Kornai. 1985. Szótári adatbázis az akadémiai nagyszámítógépen. In *Tanulmányok a nyelvtudomány és társtudományai köréből*, MTA Nyelvtudományi Intézet, Budapest, pages 65–79.
- Gábor Prószéky. 1985. Automatizált morfológiai elemzés a nagyszótári munkálatokban. Kézirat, MTA Nyelvtudományi Intézet.
- В. К. Кельмаков. 1998. *Краткий курс удмуртской диалектологии*. Издательство Удмуртского Университета, Ижевск.
- В. К. Кельмаков. 2002. *Удмурт дуалектология. Студенттэслы но аспиранттэслы юрттос*. 5-тйез, тупатъяса, выльдыса но ватсаса поттэмез. «Удмурт университет» книга поттон корка, Ижкар.
- Р. Ф. Насибуллин, С. А. Максимов, В. Г. Семёнов, and Г. В. Отстановна. 2009. *Диалектологический атлас удмуртского языка. Вып. I*. Научно-издательский центр «Регулярная и хаотическая динамика», Ижевск.