Nomina loci in Beserman Udmurt: elicitation, corpus study, and experiments

Maria Usacheva
Russian Academy of Sciences/Pushkin State Russian Language Institute, Russia

This paper discusses the advantages and disadvantages of elicitation, corpus study, and experimental study in field research. -on’n’ig-forms in Beserman Udmurt which function both as nomina loci (place names) and as converbs are described with the focus on methodology of gathering the data. It is demonstrated that in the case of productive derivation hypotheses formulated on the base of corpus study can more reliably be checked during an experiment with non-verbal stimuli than via elicitation. As for morphosyntactic properties of regular inflected forms, it is easier to study them on the base of corpora data and elicitation. However, experiments provide both natural examples and interesting data for future research.

Keywords: field linguistics, elicitation, experiments, corpus study, Udmurt

The most popular and widespread methods of linguistic fieldwork are elicitation and recording spontaneous texts (typically narratives). Transcribed, glossed and translated texts can be organized into a corpus. On the base of natural examples taken from the corpus, a linguist can formulate a hypothesis and later check it during the fieldwork via elicitation. For investigations of grammar it seems to be sufficient. Nevertheless, in recent decades field linguistics and typology have begun to gather the data with the help of non-verbal stimuli. A wide series of projects based on experiments has been realized in Max Planck Institute of Psycholinguistics. Experiments were conducted in order to gather comparable data from different languages. The projects were devoted to space (Bowerman & Pederson 1992; Bowerman et al. 2004), to time in space (Boroditsky et al. 2008), to expressives (Tufvesson 2007), etc. The idea to study encoding of spatial relations with the help of non-verbal stimuli is very fruitful. It was realized in typological projects (Fortis et al. 2009) and in descriptions of endangered languages (Arjava 2016). Experiments are conducted during field studies of syntax (Polinsky: preprint) and semantics (Arunachalam & Kothari 2011; Majid 2012). Apart from spatiality, non-verbal stimuli are relatively frequently used in studies of possessives (Klumpp 2017; Eisenbeiss 2017a), NP structure (Kozlov et al. 2016; Arkhangel'skij & Usacheva to appear) and ergativity (Longenbaugh & Polinsky 2017). There are special linguistic courses in universities devoted to experiments in the field (Eisenbeiss 2017b).

Why do experimental methods attract more and more attention in field linguistics and typological studies? The main reason seems to be obvious: every method has its limitations, and both elicitation and corpus study are not exceptions in this respect. To demonstrate it, we will describe -on’n’ig-forms in Beserman Udmurt.²

Beserman is an unwritten dialect of Udmurt. It is spoken by ethnic Besermans who live in North-Western Udmurtia (Russian Federation). According to the 2010 Census, there are ca.

---

1 The work was supported by RSF grant 18-18-00462 Communicative-syntactical interface: typology and grammar realized in Pushkin State Russian Language Institute.

2 The present work is devoted to methodological account of gathering data in the field. More information about particular morphological and syntactic properties of the -on’n’ig-forms can be found in Usacheva & Serdobolskaya (2015).
2200 people identifying themselves as Beserman. There are 10 predominantly Beserman villages and 41 villages with a mixed population. Most of Besermans are bilingual in Russian and are familiar with Standard Udmurt. The dialect differs from the literary language in phonetics, morphosyntax and lexicon. In particular, the form described in the present article is attested only in Beserman and is used in neither of the other Udmurt dialects (Kelmakov 1998: 297).

All the examples cited in the present article were collected in the village of Shamardan (Yukamenskoye district, Udmurtia) in 2003–2018. Most of the data were taken from the Beserman corpus of oral speech (ca. 67000 tokens), the Multimedia Beserman corpus (a corpus with sound and video; ca. 38000 tokens) and from the Corpus of examples from the on-line Beserman-Russian dictionary (ca. 83000 tokens). The Beserman corpus of oral speech and the Multimedia Beserman corpus contain examples of spontaneous and quasi-spontaneous speech. Spontaneous texts include recordings of every-day communication (dialogues and polilogues), narratives, receipts, interviews, etc. Quasi-spontaneous speech was recorded during experiments of different kinds: referential communication tasks, retellings of cartoons’ plots or of texts, word games, and serial reproduction. The corpus of examples from the on-line Beserman-Russian dictionary is predominantly composed of sentences got during lexicographic work. These sentences were mostly provided by Beserman speakers when we asked them to create an example where a concrete lexeme from the dictionary had to be used. Some of them are the result of a translation from Russian into Beserman, but these translations were made with a focus on lemmata from the dictionary, not on grammar.

When we searched for -on’n’ig-forms, we found 19 sentences in the oral corpora and 74 sentences in the corpus of examples. These examples gave us much information about the semantics and syntax of the forms in question. As for semantics, corpora examples show that the -on’n’ig-forms can denote place of an action (1) and a process which is parallel to the action denoted by the main verb (2):

(1) Korka-z-ə̑ kwaška, ul-on’-n’ig-z-ə̑ evəł. 
house-P.3-PL be.ruined.PST live-VN-NLOC-POSS.3-PL be.NEG 
‘Their house has been ruined, they have no place to live.’

(2) Mon n’umal kušman-ez bud-șit-on’-n’ig-șn 
I.NOM sweet beet-ACC.3SG grow-CAUS-VN-NLOC-LOC 
baš ’t-i pervoj mesta. 
take-PST.1SG first place 
‘I have won the competition of growing sugar beet.’

The forms in question are also used in when/while-clauses if the places of the two actions coincide:

(3) So dugd-on’-n’ig-a-z mon-e l’oga-m. 
that stop-VN-CVB-LOC-POSS.3 I-ACC trample.down-PST2 
‘He drove over me while he was stopping.’

---

3 All the examples cited in the article belong to Beserman Udmurt.
4 PST2 is a label used for a past tense which also expresses a range of meanings like perfectivity or evidentiality.
Corpora examples also show that the -on’n’ig-forms demonstrate both nominal and verbal properties. It is very common for converbs, but quite surprisingly for place names. One of verbal properties is the ability to attach spatial adjuncts:

(5) Ul’č’a-jə̑n so-os-len traktor-z-ə̑, door-PROL go.out-VN-NLOC-LOC fence door close-RES
Ul’č’ə̑n-so-os-len traktor-zə̑, street-LOC that-PL-GEN1 tractor-POSS.3-PL
es-t’i pot-on-n’ig-ə̑n, ken’er es uš’t-emə̑n.
doors-PROL go.out-VN-NLOC-LOC fence door close-RES
‘Their tractor is on the street, on the exit through the door, the fence’s door is closed.’

Next, -on’n’ig-forms conserve direct object which can be unmarked or bear an accusative marker depending on the referential status of the object. Such split is typical for direct objects of verbs in Beserman (as in other Finno-Ugric idioms). Thus, objects in (6) and (7) are non-specific (generic) and therefore unmarked, whereas a specific definite object in (8) bears the accusative marker:

(6) N’an’ vož’-on’n’ig-a-d n’an’ kema
bread store-VN-NLOC-P.LOC-POSS.2 bread long
vož’-i-d=ke, pə̑k zə̑n lu-e.
store-PST-2.SG=if musty smell become-PRS.3.SG
‘If you store bread in the bread basket too long, musty smell occurs.’

(7) Piroški pə̑ž-on dər-ja ul’ n’an’
patty bake-VN time-ADV unleavened dough
leš’t-on’n’ig-e pun-iš’ko-d vu, piž’, slal.
make-VN-NLOC-ILL put-DETR-2.SG water flour salt
‘While baking pattys, while making unleavened dough put water, flour, salt.’

(8) Č’orog-ez pot-on’n’ig-a-m žilka-ez
fish-ACC go.out-VN-NLOC-P.ILL-POSS.1.SG line-POSS.3.SG
č’ig-i-z.
rip-PST-3.SG
‘While I was pulling out the fish (mentioned before), the fish line ripped.’

The forms which head when/while-clauses agree with nominal subject in possessive:

---

5 In Permic languages case marking of nominal dependents in noun phrases is defined by syntactic position of these phrases. Dependents of nouns in direct object position attach markers of case labeled here as GEN2. Dependents of nouns in other syntactic positions are marked by GEN1. Genitive 2 is a former spatial case which now performs different other functions like marking material, the point of comparison etc.
A Gypsy woman stole their money while telling fortunes.

As one can see from (4) above, there is no agreement with generic zero subject. Thus, in (4) there is no possessive marker on the -on’-ig-form:

‘A tap is used for drawing ile making a blockhouse.’

Examples (3) and (9) above show that -on’-ig-forms which head when/while-clauses attach both nominal and pronominal subjects in nominative. In phrases headed by nouns a pronominal dependent must be in genitive.

As for nominal properties, one can see that -on’-ig-forms which denote places of actions (10) and processes (11)–(13) can be used as adnominal dependents or dependents of relational nouns:

‘The wall between the bedroom and the kitchen in our house is very thin.’

‘In Forest’s day people used to go to the forest with food and drinks, they tied a rag of cloth in a sacred forest (lit. in the forest where one throws sacred objects).’

‘Behind the big road there is a cattlegrave.’

Corpora data also show that -on’-ig-forms denoting place names attach nominal subjects in nominative or in genitive. Case marking of the subjects depends on their referential status. The phrase kurge ‘hen’ in (13) is non-specific and hence unmarked. Kørban ‘a Beserman fest devoted to the end of sowing’ in (14) is marked by genitive because it is opposed to other Beserman fests in the context, and this opposition makes it specific:

‘A Gypsy woman stole their money while telling fortunes.’
There are three henroosts (lit. places where hen sit) in my coop.

And just before the beginning of that Kyrban [fest] the priest came.

The same condition of referential status determines grammatical marking of dependents in phrases headed by nouns.

While studying particular grammatical phenomena we usually also take into account the data from corpora and grammars of genetically related idioms, of languages of the area and (working with dialects) of the literary variety. In the case of the -on’n’ig-forms, searching the corpora of other Udmurt varieties gave no results since such forms are attested neither in other dialects nor in Standard Udmurt (Kelmakov 1998: 297). There is a suffix of nomina loci -on’n’i-, but it does not have process meaning. When/while-clauses in other Udmurt varieties are formed by forms on -ku, -onja- on’n’a- (Georgieva 2017; the -on’n’a-forms also exist in Beserman), in Sharkan Udmurt also by forms on -onna- (Timofey Arkhangelskiy, p.c.). We also have not found the strict analogues of the -on’n’ig-forms neither in genetically related Komi varieties nor in Turkic languages (Tatar and Chuvash) spoken by people who live close to or together with Besermans and Udmurts. In Zyrian when/while-clauses can be formed by conversbs with suffix -ig-(en) which gives the base for suggestions about the etymology of the -on’n’ig- complex (Usacheva & Serdobolskaya 2015).

However, the corpus data do not contain all the information about morphological and syntactic properties of the -on’n’ig-forms. It is a typical situation, especially for the corpora of endangered unwritten idioms because such corpora are very small in comparison with the corpora of idioms with a literary tradition: usually they contain no more that several hundred thousands of tokens. For example, we can say nothing about ability of the -on’n’ig-forms to syntactically behave like nouns when used as conversbs on the base of corpora. We also do not know if the forms in question can attach the nominal plural marker, if they can be modified by adverbs and/or by adjectives, etc. For gathering this information we either have to wait while the constructions we need will occur in the corpus or have to use the elicitation method.

There are also questions that can be answered neither by means of corpora analysis nor by elicitation. They are:

1. How productive are the forms in their function of denoting place names? Do they denote only fixed locations where the process usually takes place (i.e. are they lexicalized or not necessarily)?
2. Are the -on’n’ig-forms which head when/while-clauses preferred to other conversbs also used in this function (to -on’n’a-forms, for instance) when the attention of the speaker is focused both on the location and on the process?

In order to answer these questions, I conducted an experiment with non-verbal stimuli. I formulated the following hypotheses based on the corpus data:
(i) -on ‘n’ig-forms functioning as nomina loci are productive. The complex -on ‘n’ig- can attach to a verbal stem denoting any process.

(ii) -on ‘n’ig-forms which head when/while-clauses are preferred when the place of the process is focused.

When a field linguist is planning an experiment, it is very useful to take into account the genre of the texts he wants to get as a result. It is known that texts of different genres may contain different grammatical forms and constructions. Historical present tense in English, for example, is used in narratives (both oral and written) and not in dialogues (see Wolfson (1989) and the literature cited there); Russian imperfective aspect is used to encode general validity primarily in genres other than narratives (Sitchinava 2011), etc. Before we have shown that -on ‘n’ig-forms can denote places where something happens or processes; they can also head when/while-clauses. Place names occur in texts of different genres whereas names of processes and especially when/while-clauses are more likely to be found in narratives. Therefore, we decided to use one of the most popular techniques of stimulating narratives – a series of pictures which represent a story. In order to avoid problems with items from a culture the speakers are not familiar with we took a plot of a Russian folktale which Beserman speakers definitely know – Ivan tsarevitch, the firebird, and the grey wolf. In order to check the hypothesis 1, we included both images of locations which are usually described by -on ‘n’ig-forms (the stick hens are sitting on, the place in the yard where horses are usually tied, etc.) and images with actions which are not associated with fixed locations (spinning wool, coloring fence, etc.). We also tried to attract the attention to locations where the actions take place by drawing the roads leading to people or animals performing different activities. Here are some of the pictures we used during the experiment:
Figure 1: The pictures used in the experiment with cards
We put the cards in front of Beserman speakers and asked them to tell a story represented on the pictures. If hypothesis 1 was correct, we would get -on’-n’ig-forms of verbs denoting activities not associated with concrete locations. If hypothesis 2 was correct, we would get much more -on’-n’ig-forms denoting processes than in corpora.

Actually, the typical problem with experiments in linguistics is the following. If the experiment is successful (and well-designed), it proves or declines the initial hypothesis which was intended to test. But if the experiment fails, it tells nothing about the hypothesis because there can be many different reasons for this particular experiment to be unsuccessful. For the experiment we conducted with cards there were at least three such reasons. First of all, the pictures were drawn badly. Some of the probationers could not even recognize the story and the heroes. It would be much better if the pictures were drawn by a professional artist. Secondly, the pictures with roads were too complex. There were too many actions on them, and the probationers were confused. Thirdly, speakers of Permic languages I work with have problems with telling stories based on a sequence of cards representing events. Every card is taken as referring to a single event independent from others, and it is very difficult for my probationers to apprehend a sequence of cards as representation of a coherent story. Therefore, as the experiment failed and I could neither prove nor reject my hypotheses, I decided to try a different design. During the second experiment I asked Beserman speakers (five women and one man, all under forty years old) to play a board game designed by me and drawn by the professional artist Tatyana Panova. As one can see from Figure 2, the game represented traditional Beserman activities, ceremonies, fests and frequent problems of life.

![Figure 2: The board-game designed for the experiment](image)
The game was organized as follows. There were two speakers playing. They were asked to comment what they were doing. During the game the two figures moved along different paths. One of the probationers took the figure of a girl, the other took the figure of a boy. The probationers were throwing a dice by turns and then moved their figures according to the number that appeared on the dice. When a figure stood on a circle with a red or a dark-blue number, the probationers had to put a figure on a picture with a special activity (gathering butter mushrooms, catching pigeons to heal stuttering, etc.). Then the figure was returned to the number it stayed before.

During the experiment we recorded four texts (two texts were recorded from the same pair of speakers in different days). The texts in sum last ca. 4 hours; the sub-corpus we made on their base includes approximately 21000 tokens. This experiment cannot be treated as fully successful because instead of descriptions of what the speakers were doing while playing we got narratives about traditional Beserman activities. Nevertheless, for forms functioning as nomina loci the experiment was successful. It has proved that hypothesis 1 is correct because experimental texts contain examples like (15):

\[(15) \quad \text{Odig} \quad \text{pol} \quad \text{mōnam} \quad \text{podруга-ię} \quad \text{lōkt-i-z.} \quad \text{ʒ’eč’era-nō} \]
\[
\text{one} \quad \text{time} \quad \text{I. GEN1} \quad \text{friend-POSS.1.SG} \quad \text{come-PST.3.SG} \quad \text{teeter-INF}
\]
\[
\text{pukš’-i-mō.} \quad \text{ʒ’eč’era-š’ko-mō} \quad \text{tare} \quad \text{olo-malō} \quad \text{mi}
\text{sit-PST-1PL} \quad \text{teeter-DETR-1PL} \quad \text{then} \quad \text{INDF-why} \quad \text{we.NOM}
\]
\[
\text{kōk-na-mō} \quad \text{ʒ’eč’era-n’-n’ig-iś’t ą-mō} \quad \text{uš’-i-mō.}
\text{two-COLL-1PL} \quad \text{teeter-VN-NLOC-P.EL-1PL} \quad \text{fall.down-PST-1PL}
\]

‘Once my friend came to me. We began teetering, and then we both fell down from the place where we were doing it.’

---

6 Picture 3 was taken by Nickolay Philippov.
The speaker is talking about Easter. During the celebration of this fest, Besermans of Shamardan (the village where the experiment was conducted) make big teeters and are teetering on them half of the day. The teeters are made every year in different locations and then removed, so the -on’i’g-form used in (15) definitely refers to the place of the action described. It cannot refer to the action itself (to the process of teetering) because in this function the form can attach only illative marker – not the elative, as in (15). For the teeters itself there is a special word different from the -on’i’g-form.

In addition, Table 1 demonstrates that the share of the -on’i’g-forms denoting place names in experimental texts is about 4.5 times more than the share of these forms in the spoken corpora. These forms are more frequent in the corpus of dictionary examples than in the experimental texts, but the reason is that the lexicon of the dictionary contains a good deal of lexicalized place names with several examples for each such name.

Table 1: The share of -on’i’g-forms in the existing Beserman corpora and in the corpus recorded during the experiment.

<table>
<thead>
<tr>
<th>Form</th>
<th>Spoken corpora</th>
<th>Corpus of examples</th>
<th>Experimental corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>-on’i’g- NLoc</td>
<td>9 of 105000 (0.86 for every 10000 tokens)</td>
<td>57 of 83000 (6.87 for every 10000 tokens)</td>
<td>8 of 21000 (3.81 for every 10000 tokens)</td>
</tr>
<tr>
<td>-on’i’g- process</td>
<td>10 of 105000 (0.95 for every 10000 tokens)</td>
<td>17 of 83000 (2.05 for every 10000 tokens)</td>
<td>2 of 21000 (0.95 for every 10000 tokens)</td>
</tr>
</tbody>
</table>

Therefore, we treat the experiment as successful for the -on’i’g-forms denoting place names. As for these forms functioning as converbs, in experimental texts there are only two examples like (16) containing them:

(16)  Mon žad’-i   isk’il’ta-n’i’g-a-m
      I.NOM get.tired-PST.1SG  sledge-VN-NLOC-P.ILL-POSS.1.SG.
pôr-i   muč’o-je  šădeč’ik-šnô  šăşt otôn
enter-PST.1.SG  bath-ILL  have.rest-INF  for.a.while  there
puk-o.
sit-FUT.1.SG

‘I got tired while sledging and entered the bath. I’ll sit there for a while and have a rest.’

But data from all the corpora show that another strategy used in when/while-clauses – constructions formed by verbal nouns with suffix -on and the postposition dŏrja ‘while’ – is much more frequent than that of using constructions with -on’i’g- (see Table 2).
Table 2: Absolute and relative frequency of -on’n’ig-converbs and alternative strategies in the existing Beserman corpora and in the corpus recorded during the experiment

<table>
<thead>
<tr>
<th></th>
<th>-on’n’ig ‘while’</th>
<th>-on + dərja ‘while’ (the same subject)</th>
<th>-on’a- ‘while’</th>
<th>ku ‘while’ (Standard Udmurt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental corpus</td>
<td>2 (20%)</td>
<td>6 (60%)</td>
<td>1 (10%)</td>
<td>1 (10%)</td>
</tr>
<tr>
<td>Spoken corpora</td>
<td>10 (30%)</td>
<td>18 (55%)</td>
<td>4 (12%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td>Corpus of examples</td>
<td>17 (9%)</td>
<td>139 (77%)</td>
<td>24 (14%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Three possible explanations can be offered for this situation: either hypothesis 2 is wrong; or the -on’n’ig-forms are pushed out by alternative constructions and are therefore too rare to be captured during experiments; or the design of the experiment is unsuitable for stimulating when/while-clauses. In order to choose the right one we computed the frequency of the occurrences of -on + dərja ‘while’ in all the corpora and in the experimental texts. We took into account only the cases when the subjects of the two actions coincide because it is obligatory for the -on’n’ig-forms in converbal function. One can see from Table 2 that the share of construction -on + dərja in the experimental corpus is higher than in spoken corpora, but not very much: 2.85 vs 1.7 for every 10000 tokens. It means that the design was bad, and to test the hypothesis 2 one should make another experiment to get descriptions of parallel actions.

In addition, we would like to say that even unsuccessful experiments are very good for a field linguist because, as a rule, they result in long vivid texts which are not always simple to record in the field. The narratives we got during the experiment with a board game are also very interesting from the ethnographic point of view. Besides, experimental texts often give interesting data concerning different grammatical phenomena. For example, just one 45-minutes dialogue recorded during the experiment described above contains 11 examples like (16) – with combinations of iterative and resultative suffixes in a verb:

(17) - A vot mil’am abi-len... vera-l’Ta-z
and here we,GEN1 grandmother-GEN1 tell-PST2-3.SG
tod-iš’ ko-d=a mugor-a-z so-len lu-žl-i-z
know-PRS.2.SG=Q body-LOC-POS3.SG that-GEN1 become-ITER-PST-3.SG
marõm-eš’ p’aţno-os lûz-eš’.
HES-PL.ADJ spot-PL blue-PL
- Č’epel’-ţlâ-môn kad’.
pinch-ITER-RES like
- Ben, tin’ ŝ’uš’etka=pe Č’epel’-em uj-ţn.
yes here brownie=QUOT pinch-PTCP.PST night-LOC
‘- You know, our granny’s… [she] said that blue spots appeared on her body. - Like pinched several times. - Yes, the brownie supposedly pinched her at night.’

There are only 20 examples of that kind in the oral Beserman corpora, and they do not include classes of verbs represented in texts recorded during the experiment described. Elicitation also did not give reliable results for possibility of such verbs to attach combinations of resultative
and iterative. Therefore, the experiment with the board game gave very important data concerning another, a fully unrelated grammatical theme.

In the present article we tried to show that all three methods used in field linguistics – elicitation, corpora and experiments – have their limitations. It seems rational to combine them. The contribution of each method depends on particular language phenomena. Experiments are useful even if they are unsuccessful, but they must be designed accurately, and the genre of intended texts must be taken into account since different techniques should be used to stimulate texts of different genres. I propose the following workflow for field linguists. First, one can search the corpus of the idiom in question and the corpora of idioms which are genetically or geographically close to it. On the base of the data taken from the corpora, one can formulate a hypothesis and try to test it by experiments. After it, the gap in the data can be filled with the help of elicitation.

References


**Abbreviations**  
ADD – a clitic ‘and’  
COLL – collective numeral  
DETR – detransitive  
EL – elative  
HES – hesitation  
ITER – iterative  
ILL - illative  
NLOC – verbal noun which denotes location of the action  
P.EL – allomorph of the elative case marker which appears before possessive suffixes  
P.ILL – allomorph of the illative case marker which appears before possessive suffixes  
P.LOC – allomorph of the locative case marker which appears before possessive suffixes  
PROL – prolate  
VN – verbal noun

*Maria Usacheva*  
143003,  
18-247 Odintsovo  
Russia  
mashastroeva@gmail.com