



NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Timofey Arkhangelskiy

CLITICS IN THE BESERMAN DIALECT OF UDMURT

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: LINGUISTICS
WP BRP 10/LNG/2014

This Working Paper is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE). Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

*Timofey Arkhangelskiy*¹

CLITICS IN THE BESERMAN DIALECT OF UDMURT²

The Beserman dialect of the Udmurt language has a large number of clitics, mainly discourse markers, which have never been described in detail. These clitics occur very frequently in texts and often appear in clusters. The rules which govern the ordering of clitics within a cluster are especially interesting because the binary relation “clitic A normally precedes clitic B in a cluster” proves to be intransitive, thus making it impossible to use a clitic template to describe the ordering. In this paper, the rules which govern placement of individual clitics in a sentence and their mutual order in clusters are described.

The research is based on a corpus of approximately 64,000 tokens and the author’s own field data collected in the village of Shamardan, Yukamensk region, Udmurtia.

JEL Classification: Z19.

Keywords: corpus data, clitics, clitic clusters, Udmurt, Beserman dialect

1 National Research University Higher School of Economics. Department of Philology. E-mail: tarkhangelskiy@hse.ru

2 The study was implemented in the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) in 2014

Introduction

This research provides a detailed description of the Beserman clitic system that would be useful to both those who are interested in the peculiarities of Udmurt grammar and those who consider the Udmurt data from the broader perspective of linguistic typology or areal linguistics. The paper is primarily focused on clitic clusters, which appear in Udmurt texts very frequently. The mutual positions of clitics inside clusters conform to ordering rules, and these are explicitly stated based on the corpus data and fieldwork. The main point of interest in this description is the intransitivity of the ordering, that is, the fact that knowing the rules for two-clitic clusters, one cannot predict the order in 3-clitic clusters, since in several cases different applicable rules contradict each other. Such unusual configuration makes it impossible to use post-syntactic clitic templates, a model proposed by Perlmutter (1971) and often employed when describing clitic ordering inside clusters.

The article is organized as follows. First, the definition of clitic used throughout the article is given. Then, after a brief description of the characteristics of the dialect and the sources used, a list of the items which were studied is given and I show that they indeed fall under the definition of clitic. The rest of the article will be dedicated to the properties of clitics in Beserman, namely, their position in sentence and their ordering inside clitic clusters.

What counts as a clitic?

Although the notion of clitic has been widely used in linguistic literature at least since Wackernagel's work on Greek second-position clitics (Wackernagel, 1892), there is no consensus as to what exactly counts as a clitic. Clitics have been defined in purely phonological terms (Plungian 2003:29), in terms of a bundle of properties (Zwicky, 1977), as two overlapping classes of different kind, one in terms of phonology and the other in terms of syntax (Anderson, 2005), or even as a separate part of speech (Schachter, Shopen, 2007:52). The general idea behind all possible definitions is that a clitic is a linguistic item which is neither a full-fledged word nor an affix. Because of their intermediate status, clitics may exhibit special phonological and syntactic properties, for example, they often cannot bear stress and are governed by a different set of rules than "ordinary" words. Zwicky (1977) was one of the first to propose a set of criteria for distinguishing between affixes, clitics, and words. Zwicky defines three kinds of clitics, depending on what rules that distinguish words from affixes they violate. These are

special clitics, simple clitics, and bound words. The feature which separates bound words from the first two kinds of clitics is whether there is a “free”, non-clitic variant of the same item. If a clitic has no such variant, it is called a bound word. The other two classes are separated on the basis of their morphosyntax: while simple clitics do not exhibit “special” syntactic behavior which makes them different from the rest of the lexicon, such as second position placement, special clitics exhibit such properties.

As Anderson (2005) points out, the distinction between bound words and the other two kinds of clitics is irrelevant, and each bound word under Zwicky’s classification can be classified as either a clitic with special morphosyntax, or a clitic with “ordinary” behavior. Anderson comes up with two classes for which he provides concise definitions which include one characteristic property each, instead of bundles of properties. He calls the first class “Prosodic clitics”, and the second, “Morphosyntactic clitics”. The two classes are defined independently and have significant overlap, i. e. many clitics are both prosodic and morphosyntactic simultaneously. Nevertheless, there exist clitics which are either prosodic but not morphosyntactic, or vice versa.

In this article, I will adopt the definition given in Anderson (2005:23) which says that a phonological clitic is “[a] linguistic element whose phonological form is deficient in that it lacks prosodic structure at the level of the (Prosodic) Word”. All Beserman items under consideration will be phonological clitics in that sense, while some of them also satisfy the definition of a morphosyntactic clitic.

The language and the sources

Udmurt is a Finno-Ugric language spoken by more than 300,000 in Russia according to the 2010 census. There are a number of grammars of literary Udmurt (e. g. Alatyrev (1970), Winkler (2001)), the first grammars dating back to the 18th century. Udmurt has a number of dialects which can be classified into several groups (e. g. Winkler (2001:6) divides them into five areas). The Beserman dialect, spoken in the northern part of Udmurtia, is probably the most distinct of them. The origins of the dialect and its speakers are not entirely clear, although the most widely accepted hypothesis is that its speakers used to speak one of the neighboring Turkic languages, but shifted to Udmurt at some point (Napol’skikh, 1997; Grosheva et al., 2013). Unlike literary Udmurt, Beserman is an endangered dialect. The number of Besermans decreased significantly in the 20th century, falling to a level of about 2000, according to the 2010 census. Most speakers of Beserman are now elderly people; they all are bilingual in Russian (and some

also are familiar with literary Udmurt). Children in most Beserman families speak to their parents and to each other in Russian rather than in Beserman. There are no publications in Beserman, except for a small number of transcribed texts which were included in descriptions of the language, as there is no official orthography (different set of phonemes makes it difficult to use standard Udmurt orthography for writing in Beserman). The dialect is underresourced: although there exist some publications which investigate Beserman material, for example some of the chapters in Kuznetsova (2012), and a dictionary (Kuznetsova et al., 2013) was published recently, there are no detailed descriptions of its syntax or morphology.

The article is based on the data collected in Shamardan village, Yukamensk region, Udmurtia, during a series of linguistic expeditions starting from 2003. The main sources of the data are a corpus of spoken Beserman around 64,000 tokens in size, which was collected with my active participation, and my own field data obtained during four field trips in 2012–2014.

The clitics under consideration

The Udmurt language, and the Beserman dialect in particular, has a number of clitics. What is interesting about these clitics is that they are very frequent in speech and are often encountered in clitic clusters. Unlike in most European languages which allow clitic clustering, there are no pronominal clitics in Beserman, most clitics being discourse markers. The clitics we will be concerned with in this paper are the following: *no* “and”, *n’i* “already”, *uk* “after all; focus (predicates)”, *na* “else; still; yet”, *ke* “if”, *ik* “focus particle / the same”, *a/wa* “question marker”, *pe* “quotative marker”, *pi* “autocitation marker”, and *val* “retrospective shift” (a fossilized form of BE.PST.3SG which can be used with virtually every verbal form).³

There are other clitics in Beserman which will be not covered in this article. First, all the items in focus are enclitics. I chose to restrict myself to enclitics because the system of enclitics in Beserman is significantly richer than that of proclitics, both in their total number and in the number of combinations they can form. The proclitics (such as *ben* “yes” and *ja* “well”) do not form clusters and occur much less frequently in texts. Second, I excluded enclitics borrowed from Russian such as *ved’* “after all” because of their low frequency and marginal role in the clitic system. Finally, I do not consider two two-syllable words *leš’a* “it seems” and *gine* “only”. Unlike the rest of enclitics, these two have independent stress and most of the time are stressed.

³ Of course, the translations presented here are just labels which shouldn’t be expected to cover every facet of these clitics’ meaning. Since in the article we will be concerned with distributional properties of the clitics, we will not go further in exploring their semantics, although this would undoubtedly make a good topic for research. For the details on the semantics of *ik* and *uk*, see Zubova (to appear).

Nevertheless, it seems that they can constitute parts of longer phonological words and sometimes exhibit special morphosyntactic behavior, e. g. they can appear inside clusters with other clitics. These two items require further investigation, as data on them are scarce and insufficient to make any conclusions.

Clitic placement and clustering in literary Udmurt may be slightly different from those in the Beserman dialect. For instance, in literary Udmurt the order *n'i=uk* is at least as acceptable as *uk=n'i*, while in Beserman the latter is strongly preferred. Besides, the overall frequency of clitics and clitic clusters seems to be much higher in our Beserman corpus than in standard Udmurt texts (but this may be due to the fact that there are no written texts in our corpus).

Let us now prove that the items in question are indeed clitics under our definition. These items are prosodically deficient and thus are not independent words. All these items are always unstressed and cannot appear in a sentence without a host word on their left, i. e. they don't form prosodic words alone. None of these items, except for *val* "retrospective shift", has alternative, "free" or "strong", variants which alone could form a sentence or bear stress (cf. English "don't" or "strong" French pronouns). *Val* is different from other enclitics in that it can occur as an independent word with the meaning "be.PST.3SG". However, it would be better to count stressed and unstressed *vals* as two different lexical items, since the former retains its original lexical meaning, while the latter attaches to other verbal forms, adding the meaning of retrospective shift, and is unchangeable.

On the other hand, the items under consideration are easily distinguished from morphemes. First, they can be differentiated on the basis of stress position. Beserman words always bear stress on their last syllable, except for the imperative and negated forms of certain classes of verbs. When a word is accompanied by one or more enclitics, the position of the stress remains the same, all clitics being unstressed. Another property of Beserman enclitics which helps distinguish them from morphemes is their relative order when several clitics happen to share a host word. Although in most cases, as I am going to show, there is a preferred, or even fixed order, there still is much more variation in their relative positions than there would be if they were morphemes.

General characteristic of the clitic system

Most of these clitics are used very often in speech. In the corpus of spoken Beserman, these clitics are at the top of the frequency list, the most frequent clitic, *no* "and", ranking

second, just below the 3SG pronoun *so*, and *n'i* “already” is third. Table 1 shows raw frequencies and approximate words per million (wpm) rates for the 6 most frequent enclitics:

Table 1. Clitics and their frequencies in the Beserman corpus

clitic	gloss	raw frequency	wpm
<i>no</i>	and	1325	20700
<i>n'i</i>	already	1034	16200
<i>uk</i>	after all; focus (predicates)	419	6500
<i>na</i>	else, still	395	6000
<i>ke</i>	if	385	6200
<i>ik</i>	focus particle / the same	292	4500

It seems that most sentences in our texts, especially in dialogs, have at least one clitic, and often more than one, as in the following example:

- (1) “ben, mar=pe kar-od=na vøldə=wa, mar=a,
ok what=CIT make-FUT.2SG=else then=Q what=Q
pukš’-od=ke=pe, pukš’-ə=n’i”
sit-FUT.2SG=if=CIT sit-IMP.2SG=already

“Well, what can I do, if you’re going to sit here, then sit down already.”

It is not surprising that often two or more clitics attach to one host. When two clitics combine this way, they are said to form a clitic cluster (see e. g. Spencer and Luís (2012:47)). In Beserman, clusters of two clitics are encountered often (such clustering can be seen in the previous example: *pukš’od=ke=pe*); clusters of three clitics appear much less frequently in texts, but are still perfectly grammatical (if they conform to certain rules). A cluster of four clitics occurred only once in our corpus:

- (2) sære so opet’ soš’ed-e šu-e:
then he/this again neighbor-POSS1SG say-PRS.3SG
ju-od=na=wa=pe=no?

drink-FUT.2SG=else=Q=CIT=ADD

“Then this neighbor of mine says again: will you drink more [vodka]?”

Although taken from a real Beserman text, this sentence was rejected as ungrammatical by several informants, so the question of whether clusters of four clitics are possible in Beserman remains open.

Placement of individual clitics

The rules which define the place of individual clitics in a sentence are of two different kinds. A clitic either conforms to the rules of syntax which are applied in the same fashion they are applied to independent words, or it is subject to special rules restricted to clitics, such as the rule of the second position. Clitics from the latter class are called morphosyntactic clitics in Anderson (2005). In Beserman, there are both kinds of clitics. We will begin by looking at the clitics whose position is fixed and located immediately after the phrase they modify, then we will move on to the clause-level clitics whose preferred position is following the predicate, and finally we will look at the clitics which prefer the second position.

Class 1: Clitics with limited scope

The clitics *ik* “focus particle / the same”, *val* “retrospective shift”, *no* “and”, and *a/wa* “question marker” have limited scope and are placed immediately after the constituent they modify. The clitic *uk* can be said to belong to both this class and the class 2, since it may modify only the predicate and not the whole clause, marking contrastive focus.

Ik is different from the rest of the clitics in that it cannot be placed after a predicate. Most of the time it modifies an adverb or a PP:

(3) otč’ə=ik šatt-ez val, kartoška pɣl-ən
 here=FOC weed-POSS3SG be.PST.3SG potato inside-LOC

“Weed grew in this very place, among the potatoes.”

(4) [t’exn’ikum-ez anal-t-em bere]=ik
 college-POSS3SG run.out-CAUS-PTCP after=FOC

“[The records in my employment record book begin] just after I graduated from college.”

If the predicate is in focus, the clitic *uk* should be used; *uk* cannot be used in the cases *ik* can, since, as shown in (Zubova, to appear), these two clitics are in complementary distribution.

Val is a marker with a broad meaning of “retrospective shift” or “discontinuous past” which can be used as a part of an analytic construction with any verbal form (including debitative and imperative), shifting the time reference into the past and/or marking that the desired result was not achieved (see Plungian and van der Auwera, 2006). This clitic always follows another verbal form and can be separated from it only by other clitics. Although it is placed immediately after the verb it modifies by syntactic rules, just as for other Class 1 clitics, the rules that govern the position of *val* in a cluster of clitics are different from those that govern the ordering of the rest of the clitics. Its position is generally freer than for other clitics, but these rules can be partly explained by the fact that unlike other clitics, the clitic *val* originates from the verb form *val* which is an independent word and can act as a host for clitics itself. The same can be said of other analytic verbal constructions in Beserman, the negative and the optative. Therefore, when speaking of rules that define the linear order of clitics, I will not take *val* into account; instead, I will describe its properties in the section dedicated to the analytic verbal forms.

The clitics *no* “and” and *a/wa* “question marker” can appear after any constituent. The former follows one of the coordinands, and the latter follows the information being questioned. The question marker has two phonetic variants: after consonants, *a* is used, and after vowels, both forms are equally possible. This kind of variability is idiosyncratic to this particular clitic and is not a consequence of general phonological restrictions in Beserman. Also, it does not impose any restrictions on the position of this clitic.

(5) mil'am=no muč'o est-ono
we.DAT=ADD bathhouse heat-DEB

“We also have to heat our bathhouse.”

(6) tatč'ə=a tər-od=n'i?
here=Q put-FUT.2SG=already

“Will you put [it] here [and not some other place]?”

These two clitics can occur in pairs with the meaning “both X and Y” ($X=no$ $Y=no$) and “either X or Y” or rather “it may be X or Y or I don't know what else” ($X=a$ $Y=a$). The latter is

most frequently used in two constructions: $X=a\ mar=a$ “X or something” and $X=a\ ug=a$ “either X or not X”. The $X=no\ Y=no$ construction, on the other hand, is one of the main strategies of coordination and does not impose special restrictions on what objects can fill the slots.

- (7) a [kureg=no ž'až'eg=no]=a=n'i tənad?
 so [hen=ADD goose=ADD]=Q=already you.SG.DAT

“So you already have both a hen and a goose, right?”

- (8) otən so gožja=a mar=a kin'ke?
 there this write.PRS.3SG=Q what=Q someone

“Is someone recording this there or something?”

Class 2: Clause-level clitics following the predicate

The second class of Beserman clitics are those which have the whole clause as their scope and prefer to attach to its head, i. e. to the predicate. These clitics differ in to what extent their position is fixed: while some are almost never moved away from the predicate, others are less restricted and can move to other hosts inside the clause. The clitics which prefer to attach to the predicate, in the (approximate) order of increasing freedom of their placement rules, are the following: *uk* “after all; focus (predicates)”, *n'i* “already”, *na* “else; still; yet”, *ke* “if”.

Uk almost always follows the predicate. When presented with a sentence where *uk* was moved one or two words away from the predicate, speakers sometimes do not reject it as ungrammatical altogether, but notice that it is “somewhat ill-formed” compared to the original sentence. Whenever possible, the sentence with *uk* moved to another position acquires new default interpretation. Consider the following two examples:

- (9) so parš' gid'-e pər-iz=uk
 this pig pigsty-ILL enter-PST.3SG=after.all

“This pig did enter the pigsty”.

- (10) so parš'=uk gid'-e pər-iz
 this pig=after.all pigsty-ILL enter-PST.3SG

“This is a pig after all, [that’s why] it entered the pigsty”.

The latter of the examples was reinterpreted as having two clauses because the reinterpretation of *parš'* “pig” as a predicate was possible (no copula is used in the present tense

in Beserman).

The clitic *n'i* “already” also strongly prefers the position after the predicate. Nevertheless, the restrictions for this clitic seem to be weaker than those for *uk*. There is a handful of examples in our corpus where *n'i* is not in its preferred position:

- (11) kartoška puk-t-ənə təb-ət-im abi-n'-əš'=n'i
potato sit-CAUS-INF go.down-CAUS-PST.1PL grandmother-DMS-EL=already
“We already brought the potatoes to be planted from the grandmother’s house”.

The clitic *na* “else; still; yet” behaves in roughly the same way as *n'i*. When this clitic is moved away from the predicate, the opinions of speakers may vary: some say the sentence is well-formed, while others say it is ungrammatical. One interesting case when *na* may not attach to the predicate is when it is used together with its equivalent borrowed from Russian, *eššo / eš's'o*. Such doubling happens occasionally with other clitics and words as well, e. g. Beserman complementizer *šusa* is often used simultaneously with its Russian equivalent *što*, the former appearing at the right border of the subordinate clause, and the latter, at its left border. When *na* is used simultaneously with *eššo / eš's'o*, the two items may go together and precede the predicate.

Finally, *ke* “if” immediately follows the predicate most of the time, but may move around the clause even more freely than *na* “else; still; yet”. When not attached to the predicate, *ke* usually attaches to the word immediately preceding the predicate:

- (12) mone=pi š'az gai=ke kut-iz, mar mon kar-o?
I.ACC=AUTOCIT now road.police=IF catch-PST.3SG
what I.NOM do-FUT.1SG
“And if the road police catches me, what will I do then?”

Class 3: Second position clitics

There are only two clitics which prefer the second position, i. e. which attach to the first prosodic word in the clause: *pe* “quotation marker” and *pi* “autocitation marker”. *Pi* does not occur frequently in our corpus, but on the basis of the existing examples and the data collected through questionnaires, I infer that its distribution is identical or nearly identical to that of *pe*. Therefore, I will use the data for *pe* as representative of the distributions of both clitics.

The rule of second position is really a tendency and not a strict rule, in Beserman, *pe* can appear at the edge of the first constituent, or even further; moreover, it can be repeated in long clauses:

(13) klub-e=pe mama mən-iš'ko
 club-ILL=CIT mom go-PRS.1SG
 “I’m going to the club, mom”. [reported speech]

(14) [ken'er vəlt'i]=pe oš-əl-il'l'am š'ul'-l'os-se
 fence along=CIT hang-ITER-PST.EVID intestine-PL-ACC.POSS3SG
 “His intestines were hung all over the fence [they say]”.

(15) veral-o [jegit' kalək]=pe ərod-eš'
 say-PRS.3PL young people=CIT bad-PL
 “They say, young people are bad”.

Clitics in analytic verbal forms

Analytic verbal forms in Beserman include “retrospective” forms marked with the enclitic *val*, negative forms marked with negative particles (or forms of a negative verb) *ug/um/ud/uz* (present and future) and *ej/em/ed/ez* (past), optative marked with a particle *med*, and prohibitive marked with the particle *medam* (occasionally *medaz* in the third person). All these markers either can be treated as verbs synchronically or originate from verbs (the optative and the prohibitive particles are cognates of the verb *medənə* “want”).

The optative and the prohibitive particles usually precede the main verbal form, but allow other words to separate them from that verbal form. The negative particles always immediately precede the negated verbal form: only clitics (but not independent words) can occur between the two components of an analytic negative form. The same is true for the retrospective shift particle, except that it follows the main verbal form instead of preceding it. As we have seen, clitics often choose verbal forms as their hosts. When the host is a single verb form, such a clitic would appear at its right border (probably separated from it by other clitics), but if a host is a complex form, there are more possibilities: for example, in the negated verb form the clitic can be placed either at the right border of the construction, or between the negative verb and the main verb. When two constructions are combined, e. g. if the main verb is accompanied by both a negative verb and the retrospective shift marker, there are three positions where a clitic can

appear, as shown in the following outline (the possible positions are numbered 1 to 3):

(16) [NEG/OPT] (1) [VERB] (2) val (3)

The optative and prohibitive markers do not have fixed position with respect to the main verb, and when they happen to immediately precede it, they normally are not separated by any clitics. Negation and retrospective shift are more interesting in this respect. When the verb is negated and accompanied by one clitic, the clitic can be placed between the two parts:

(17) kolxoz-e uže um=n'i vel't-iš'ke
 kolkhoz-ILL already NEG.PRS.1PL=already go-NEG.PRS.PL

“[That time,] we already didn’t go to the kolkhoz (collective farm)”⁴.

Every clitic can occupy the position occupied by *n'i* in the previous example (although it is quite rarely occupied by *uk* “after all; focus (predicate)”). On the other hand, it is almost always possible to put the clitic after the main verb without any change in meaning; although there is a preference for attaching it to the negative word, the latter variant can be encountered quite frequently in the corpus. Occasionally the clitic is doubled, i. e. two copies of the clitic occupy both positions at once (recall that the same sometimes happens to the second position clitics, *pe* and *pi*, although the distance between the hosts is usually greater in the case of *pe* and *pi*):

(18) moga, kal' uz=na vu=na marəm-ez?
 wait.IMP.2SG now NEG.FUT.3SG=yet come.FUT.3SG=yet HES-POSS3SG

“Wait, isn’t that thing ready now?”

If several clitics are attached to the complex form, they also can occur between the negation and the main verb:

(19) ja až'ə=ka=n'i, ez=a=n'i=ke vu-ə?
 well look.IMP.2SG=PREC⁵=already NEG.PST.3SG=Q=already=if come-NEG.PST.SG

“Well, go look already, isn’t it ready yet?”

4 Notice that the native Beserman *n'i* “already” is doubled here with its Russian equivalent, *uže*, in a similar way to *na* “else” and *eššo / eš's'o*.

5 *ka* is a relatively infrequent clitic with precative meaning borrowed from Russian and used with imperatives.

It seems that the behavior of those clitics which can appear between the negation and the main verb can be generalized as “if a predicate is a construction involving several consequent verbs, place the clitic after the first of them”. This generalization is corroborated by the fact that in constructions with modal verbs like *bəgatənə* “can” or *luənə* “become; be possible” preceded by the infinitive of a content verb the clitics are often found next to the content verb, followed by either a modal verb alone, or by a negation and a modal verb.

When several clitics are attached to a negated verb, they can usually be placed both between the parts and at the right border, their ordering with respect to each other being governed by the clustering rules dealt with in the next section. Nevertheless, it seems that there are peculiar special cases when each of the two possible places is occupied by one of the clitics. For example, if one of the two clitics is *no* “and”, and the other is *na* “else; still; yet” or *n’i* “already”, the former seems to prefer to attach to the negation, while the latter is expected to occupy the other possible place. Consider the following example and note that the cluster *na=no* would be absolutely grammatical in other contexts:

- | | | | | | |
|-------|----------------------------------|---------------------|-------|----------------------------------|---------------------|
| (20a) | ^{ok} ej=na | až’-əl | (20b) | ^{ok} ej | až’-əl=na |
| | NEG.PST.3SG=yet | see-ITER.PST.SG | | NEG.PST.3SG | see-ITER.PST.SG=yet |
| | “[I] haven’t seen [it] yet” | | | “[I] haven’t seen [it] yet” | |
| (20c) | ^{ok} ej=no | až’-əl=na | (20d) | [?] ej=na | až’-əl=no |
| | NEG.PST.3SG=ADD | see-ITER.PST.SG=yet | | NEG.PST.3SG=yet | see-ITER.PST.SG=ADD |
| | “[I] haven’t even seen [it] yet” | | | “[I] haven’t even seen [it] yet” | |
| (20e) | [?] ej=na=no | až’-əl | | | |
| | NEG.PST.3SG=ADD=yet | | | | see-ITER.PST.SG |
| | “[I] haven’t even seen [it] yet” | | | | |

The situation with the retrospective forms is slightly different. As with the negative forms, *uk* “after all; focus (predicates)” usually attaches to the right border of the complex form, but in this case another clitic, *n’i* “already”, exhibits the same pattern. The other clitics still prefer the position after the first part of the construction. When negation and *val* are used together, these clitics tend to appear after the negation.

Clusters of clitics

The ability to form chains, or clusters, is a well-known property of clitics observed in many languages that has been a subject of numerous investigations (clitic clusters that have attracted most attention are those of Romance, Slavic, and Greek languages). The reason why clitic clusters are an interesting object of study is the rules which govern their ordering: as it is put in Spencer and Luís (2012:9), “clusters of clitics often have properties that we associate more with strings of affixes: the clusters usually assume a fixed, partly idiosyncratic order which often goes against what the rest of the syntax of the language dictates”. Because such an idiosyncratic order (although it often allows variation to some extent) is a property which is characteristic of affixes, not autonomous words, clitic clusters can be said to belong to the domain of morphology, at least to a certain extent.

The rigid ordering rules (particularly those found in most European languages) have led linguists to the idea of templates (Perlmutter (1971) and Bonet (1991); see e. g. Zalizniak (2008) for an elaborate description of the Old Russian clitic system featuring a template with 8 slots). According to the template model, the clitics can be divided into several sets such that no clitics from the same set can appear in the same cluster, and these sets can be ordered, forming a clitic template. The set of clusters allowed by such template can be further narrowed by a set of additional constraints which disallow certain combinations of clitics, such as the “*me-lui constraint” in Romance languages.

The idea of rigid ordering, and the idea of templates in particular, although supported by much evidence from various languages, can be potentially misleading. It is easy to start believing that any clitic sequences can be described in terms of templates and constraints, if they are subject to any ordering rules at all. The rules of clitic ordering in Beserman will show us that there are at least two reasons why this is not necessarily the case.

The first fact about Beserman clitic clusters which makes the Beserman clitic system different from the well-known European pronominal clitic systems is that ordering rules exist, but are in fact not that rigid. The ordering rules seem to be independent of both syntax and phonology and state that for every set of clitics, there is a preferred order in which these clitics appear within a cluster. Nevertheless, in nearly all cases, at least in clusters of size two, the order is not absolutely fixed, and clitics may exchange their places without rendering the sentence ungrammatical and without any change in meaning. When asked about the grammaticality of such alternatively ordered clusters, in most cases the speakers say they are definitely correct,

although sometimes “sound a bit worse” than the clusters with the preferred order. In the corpus, too, we see that alternative orders are visibly less frequent than the preferred ones, but still occur occasionally. It can be inferred from the idea of clitic templates that every clitic normally has its place in a cluster and cannot be put in another place; as noted in Heap (2005:86), “[t]he fixed-order generalization is so close to being categorically true that it is hardly surprising so many linguists have [...] dealt with only invariably fixed-order clitic sequences without taking into account variable sequences”. As we can see from the Beserman data, this is not necessarily true: even if there are ordering rules, the order is flexible to some extent.

Another, even more peculiar, property of the Beserman clitic system is the intransitivity of clitic ordering. Transitivity of a binary relation “ $<$ ” is a property whereby for any a , b , and c , the statements “ $a < b$ ” and “ $b < c$ ” imply “ $a < c$ ”. The clitic ordering rules can be written in the form of such relation if by “ $A < B$ ” we mean “when the clitics A and B stand next to each other in a cluster with no other clitics between them, A should be followed by B ”. It may seem only natural that the default order of clitics must have the property of transitivity. In fact, it has long been known that there are cases where such a property does not hold. Consider the “*me-lui constraint” in French: while, in terms of the ordering relation, $me < le$ and $le < lui$, one cannot say that $me < lui$ since the two clitics do not combine. One possible solution to such issue is using a set of typological or language-specific constraints which would disallow certain combinations of clitics. In this case, it is still possible to employ templates for describing clitic clustering.

Restrictions of such kind are sometimes considered the only obstacle to transitivity (as in e. g. Heggie, Ordóñez, 2005:15–16). Nevertheless, there is another scenario: if any two of the three clitics A , B , and C can be combined in a cluster, it might happen that $A < B$, $B < C$, and $C < A$. Unlike with the “*me-lui”-kind constraints, such a situation would render the use of templates impossible, since a template, even when accompanied by a set of constraints, assumes that whenever any two of the three clitics are compatible within a cluster, they are ordered transitively. This scenario is exactly what happens in Beserman. Consider the clitic ordering graph (fig. 1). An arc between two clitics denotes their default order when they stand next to each other in a cluster with no other clitics separating them. The two numbers over an arc mean the number of instances in the corpus where the two clitics are in the default order, and the number of instances with the opposite order (showing that the ordering rules are flexible to some degree).

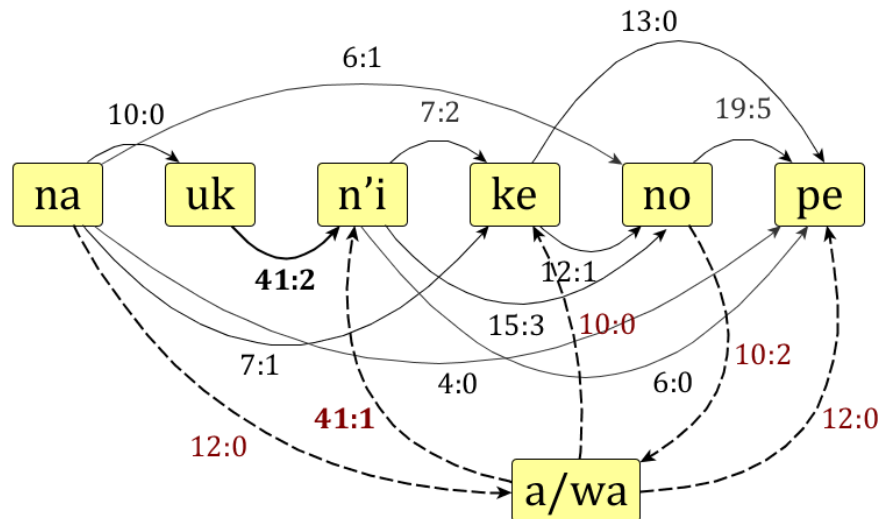


Fig. 1. The graph of Beserman clitics ordering relation.

As it can be seen from the figure, the 6 clitics in the upper row⁶ can be ordered by a template, since the ordering relation between them is transitive (e. g. *n'i* precedes *ke*, *ke* precedes *no*, and *n'i* precedes *no*). The second-position clitic *pe* is always located after all other clitics, which is cross-linguistically common (Luraghi, 2013:170).

The whole system, however, cannot be ordered by any template. The reason for that is that there are two cycles in this graph. The first is $no < a, a < n'i, n'i < no$, and the second, $no < a, a < ke, ke < no$. It is interesting that in both cycles, there is no restriction on co-occurrence of all three clitics in one cluster, although their order in such clusters cannot be predicted on the basis of this binary ordering. In both cases, there is only one idiosyncratic ordering possible: $no=a=n'i$ in the first cycle (cf. (7) and the example below) and $no=a=ke$ in the second; other orders are prohibited.

- (21) kureg=no=a=n'i puza? (*kureg=a=n'i=no, *kureg=n'i=no=a)
 hen=ADD=Q=already lay.eggs.PRS.3SG
 “Is this hen also laying eggs already?”

We see that the rules which govern the behavior of two-clitic clusters do not generalize to three-clitic clusters. It seems that the three-clitic clusters are not built from templates; instead, they exhibit idiosyncratic order which is probably memorized, like a string of affixes where such

⁶ The clitic *ik* “focus particle / the same” was not included in the graph because it does not usually combine with other clitics. As was said earlier, my data on *pi* “autocitation marker” is insufficient, but it seems to occupy the same slot as *pe* (with which it is, of course, incompatible).

behavior is more common (consider, for example, variable nominal affix order in Mari described e. g. in Luutonen, 1997).

Conclusion

The Beserman dialect has a rich system of clitics which provides some insight for the cross-linguistic study of clitics. Individual clitics in Beserman can have narrow or wide scope; in the latter case they may prefer to attach to the predicate or to be put in the second position. In analytic verbal constructions most of the clitics which prefer to appear following the predicate usually are placed after the first prosodic word of the verbal construction. Clitic clusters always have one preferred, or default, ordering. Nevertheless, their ordering poses a problem for the widely used template approach to clitic ordering. The first problem is that their order in most cases is not absolutely fixed and there is room for variation. Another problem which is more fundamental is that the clitic order is intransitive and therefore cannot possibly be described in terms of templates and co-occurrence restrictions. Every clitic cluster has its default order, but it seems that there are no general rules that govern the whole system.

References

- Alatyrev, V. I., ed. (1970). *Grammatika sovremennogo udmurtskogo yazyka* [A grammar of Modern Udmurt]. Izhevsk: Udmurtiya.
- Anderson, Stephen R. (2005). *Aspects of the Theory of Clitics*. Oxford University Press.
- Bonet, Eulalia (1991). *Morphology after syntax: Pronominal clitics in Romance*. MIT Press.
- Grosheva, A. N., Yu. V. Shneider, I. Yu. Morozova, O. V. Zhukova, S. Yu. Rychkov (2013). Genetic diversity of Besermyans inferred from mitochondrial DNA polymorphism. // *Russian Journal of Genetics*, Vol. 49, Issue 11, P. 1168–1174.
- David Heap (2005). Constraining optimality. Clitic sequences and feature geometry. // Heggie, Lorie, and Francisco Ordóñez, eds. (2005). *Clitic and affix combinations: theoretical perspectives*. Vol. 74. John Benjamins Publishing, P. 81–102.
- Heggie, Lorie, and Francisco Ordóñez, eds. (2005). Clitic ordering phenomena. The path to generalizations. // Heggie, Lorie, and Francisco Ordóñez, eds. (2005). *Clitic and affix combinations: theoretical perspectives*. Vol. 74. John Benjamins Publishing, P. 1–29.

Kuznetsova, Ariadna I., ed. (2012). *Finno-ugorskie yazyki: fragmenty grammaticheskogo opisaniya* [Finno-Ugric languages: Fragments of grammatical description]. Moscow: Rukopisnye pamyatniki Drevnej Rusi.

Kuznetsova, Ariadna I. et al., eds. (2013). *Slovar' besermyanskogo dialekta udmurtskogo yazyka* [Dictionary of the Beserman dialect of Udmurt]. Moscow: Tezaurus.

Luraghi, Silvia (2013). Clitics. // Luraghi, Silvia, and Claudia Parodi, eds. (2013). *The Bloomsbury Companion to Syntax*. London, New York: Bloomsbury, P. 165–193.

Luutonen, Jorma (1997). The variation of morpheme order in Mari declension. *Suomalais-Ugrilaisen Seuran toimituksia* 226. Helsinki: Suomalais-Ugrilainen Seura.

Napol'skikh, V. V. (1997). "Biserminy" ["Bisermins"] // *O besermyanakh* [About the Besermans], Shklyayev, G. K. (ed.), Izhevsk, P. 50–54.

Perlmutter, David (1971). *Deep and surface constraints in generative grammar*. New York: Holt, Rinehart and Winston.

Plungian, Vladimir A. (2003). *Obschaya morfologiya: vvedenie v problematiku* [General morphology: An introduction to the field]. Moscow: URSS.

Plungian, Vladimir A., and Johan van der Auwera (2006). Towards a typology of discontinuous past marking. // *Sprachtypologie und Universalienforschung* 59.4, P. 317–349.

Schachter, Paul, and Timothy Shopen (2007). Part-of-speech systems. // Shopen, T. (Ed.). (2007). *Language typology and syntactic description* (Vol. 1). Cambridge University Press, P. 1–60

Spencer, Andrew, and Ana R. Luís (2012). *Clitics: An Introduction*. Cambridge University Press.

Wackernagel, Jakob (1892). Über ein Gesetz der indogermanischen Wortstellung. // *Indogermanische Forschungen* 1, P. 333–436.

Winkler, Eberhard (2001). *Udmurt*. München: Lincom Europa.

Zalizniak, Andrej A. (2008). *Drevnerusskie enklitiki* [Old Russian Enclitics]. Moscow: Yazyki Slavyanskikh Kul'tur.

Zubova, Yulia N. (to appear). *Sintaksis i semantika fokusnykh chastits v besermyanskom dialekte udmurtskogo yazyka (na primere chastits ik i uk)* [Syntax and semantics of focus particles in the Beserman dialect of Udmurt, by the example of the particles *ik* and *uk*]. // *Acta Linguistica Petropolitana*, Vol. XI, part 2. St Petersburg: Nauka

Zwicky, Arnold M. (1977). *On clitics*. Bloomington: Indiana University Linguistics Club.

Timofey Arkhangelskiy
National Research University Higher School of Economics. Department of Philology.
E-mail: timarkh@gmail.com

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

© Arkhangelskiy, 2014