

TIMOFEY ARKHANGELSKIY (Hamburg)

## DIALECTAL VARIATION OF UDMURT DISCOURSE CLITICS

**Abstract.** Udmurt has numerous discourse clitics. Often multiple clitics attach to the same host word, forming clusters. Clitic clusters in European languages have been successfully modeled with clitic templates. In this paper, I am considering ordering of discourse clitics across Udmurt dialects. I demonstrate that there is significant variation between dialects in this respect, as well as some free variation within varieties. While the template approach, with certain caveats, can be applied to each individual Udmurt variety, it cannot be used to model the entire language. I show that the order in clitic clusters is mostly idiosyncratic. In particular, it does not depend on the semantics of the particles or the order of their appearance in the language. Apart from that, I touch upon the peculiar situation whereby certain discourse functions are expressed by particles of completely different origin in different Udmurt dialects. I argue that this is due to a Jespersen's cycle-style transformation, when an existing particle is gradually replaced by a new one through an intermediate stage, when both are used simultaneously.

**Keywords:** Udmurt, discourse particles, clitics, grammaticalization, dialectal variation.

### 1. Introduction

Although my investigation has one subject, it has a variety of objectives. In terms of its subject, this paper is dedicated to certain enclitics and their combinations (clusters) in the dialects of Udmurt. An example of such a cluster can be seen in (1).

(1) Udmurt (Standard; Udmurt duńne, 15.08.2014)

*Žakjapaj, Ud'dad'i-jš pešanaj, koncert-jš bgrd-jša šamen*  
 D. U.-EL grandmother concert-EL cry-CVB in.manner  
*košk-i-z, mezm-o=uk=ni=pe tabere.*  
 leave-PST-3SG miss-FUT.1SG=ENIM=already=REP now

'Džakyapaj, the old lady from Uddjadi, left the concert with tears in her eyes, [saying] she is going to miss [the visitors from Finland she met] already now.'

There are several caveats that have to be made right away. First, as you will see from the list in Table 1, the clitics in question are rather heterogeneous,

Received 24 November 2022, accepted 11 May 2023, available online 10 June 2023.

© 2023 the Author. This is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International Licence CC BY 4.0 (<http://creativecommons.org/licenses/by/4.0>).

and not all of them can be unambiguously called discourse (or modal, or pragmatic) particles in the usual sense (Zimmermann 2011). Nevertheless, each of them has some clearly pragmatic functions in at least one of its uses. Uniting them under that umbrella term is thus not outright incorrect and has an advantage of making it clear right from the start that the combinations I am referring to are not pronominal clitic clusters. While the latter are common in European languages and have been extensively researched (see e.g. Bonet 1991; Cardinaletti 1999; Hellan, Platzack 1999; Bošković 2004), far less is known regarding the former. The only item on my list that probably cannot be described as a discourse clitic at all is the question marker *a*. I am nevertheless including it because it interacts with discourse clitic in clusters much in the same way those clitics interact with each other.

Another caveat is that I remain agnostic as to the part-of-speech characterization of the words in focus, which have been traditionally treated as adverbs, conjunctions and/or particles. When calling them discourse particles, I do not imply that “particle” is their part of speech, or indeed that there are good reasons to have “particles” on the part-of-speech list at all (cf. Zwicky 1985). Neither do I imply, by calling them simply “clitics”, that clitics might be treated as a category of its own (Schachter, Shopen 2007 : 52–53) or considered acategorial (Zwicky 1985).

Finally, there has been a lot of discussion in the theoretical literature, both formal and functional, as to what should count as a clitic and whether “clitic” is a good cross-linguistic category or comparative concept (Zwicky 1977; 1985; Anderson 2005; Spencer, Luís 2012; Haspelmath 2015; 2022). However, Udmurt elements seem to conform to any definition of clitic ever given, including the most recent proposal by Haspelmath (2022) (“a clitic is a bound morph that is neither an affix nor a root”). On the one hand, they cannot be used in isolation and always form one phonological word with their host. On the other hand, they do not have selectional preferences like affixes do, and can attach to words of any part of speech. Therefore, calling them clitics is uncontroversial, regardless of the definition one adopts. While I have no preference for any specific definition, what is important to me in their clitic status is that the way they are positioned inside a cluster can be a priori expected to differ from both the way autonomous words are ordered in a sentence and the way affixes are ordered inside a word.

As to the objectives, there are several. First, I am going to find out which enclitic particles and particle clusters are available across Udmurt dialects. In a sense, this study is aligned in its goals and methods with the recently emerged field of variational pragmatics (Schneider, Barron 2008; Schneider 2010; 2021; Aijmer 2013). The objective of this approach is, according to Schneider (2010 : 239), to add “the pragmatic level to the other language levels overwhelmingly analyzed in dialectology”. Specifically, the first objective of my investigation is subsumed under the study of dialectal variation on the “formal” level of pragmatic analysis (Schneider 2008 : 20). The target variables in my case are clitics (or absence thereof) that are employed in certain pragmatic functions. Out of the main macro-social factors whose influence on the target variable is usually studied in variational pragmatics, I am focusing on the region (or dialect) of the speaker. Although I did not study the other usual factors (social class, ethnicity, gender, age) systematically, they seem to play a much smaller role, if any at all. Inter-speaker variability in my data is only witnessed in the settle-

ments located on borders between contiguous dialectal areas. As to the methods, my investigation satisfies the principles of empiricity (I study language use in corpora and through fieldwork), contrastivity (I compare multiple varieties of the same language) and, although to a lesser extent, comparability (Schneider 2021 : 672–673).

However, my primary focus is not on the semantics of discourse particles, and thus not on the pragmatic level in the strict sense. Instead, I am mainly interested in the syntactic properties of the clitics, namely their position in the sentence and their position relative to each other when multiple clitics attach to the same host word. The second objective of my research is to establish the factors that may or may not affect clitic placement, namely, influence of contact languages, order of their appearance in the language and their meaning. I will only discuss the semantic side when it is relevant for this objective.

Finally, I am going to discuss why different dialects have clitics of different origin that apparently have the same non-trivial meaning. I will demonstrate how dialectal data can shed light on the historical developments that might have resulted in the observed distribution.

Udmurt is a language spoken by around 300,000 people who mostly live in the Russian region of Udmurtia and certain (mostly adjacent) parts of the neighboring regions. Udmurt was only standardized in the 1930s, and the traditional dialects are alive and spoken in Udmurt villages on a daily basis. Traditionally, Udmurt has been divided into four supradialects: Beserman, Northern, Central, Southern. The latter presents a rather sharp internal divide between Southern proper, spoken mainly inside Udmurtia (further referred to as Southern) and Southern Peripheral, spoken mainly in Tatarstan and Bashkortostan (Кельмаков 1998). Although Beserman has been recently recognized as a separate language in Russia,<sup>1</sup> it is actually rather close to, and easily mutually intelligible with, the Northern varieties, which is why treating Beserman on a par with Udmurt dialects as parts of an overarching languoid (Cysouw, Good 2013) makes sense. Dialectometric analysis (Архангельский 2021) shows that most other supradialectal borders, unlike the ones between smaller units, are indeed rather strong and obvious. Although standard registers (standard written Udmurt and the common urban vernacular) have exerted some influence on the dialects through education, books and mass media, most Udmurt speakers never fully master them (Edygarova 2014). Udmurt has long been in contact with Tatar (Turkic > Kipchak), which has had a significant influence on Udmurt (Agyagási 2012), with southern varieties affected much more than the northern ones. Recently, Russian has replaced Tatar as the dominant language.

My investigation is based on the data I collected during two field trips in 2021–2022, as well as on several corpora and published dialectal texts. The field data consists of questionnaire sentences orally translated from Russian into dialectal Udmurt, as well as some monologue and dialogue recordings. During the 2021 field trip, I collected data in 10 settlements in Udmurtia, Tatarstan and Bashkortostan, working with multiple speakers in each place. In 2022, I worked with the members of the Estonian Udmurt community in Tartu and Tallinn, who represented 13 local varieties in total (mostly one speaker per settlement). This data was supplemented with relevant examples from the

<sup>1</sup> According to the “Languages of Russia” project (<http://jazykirf.iling-ran.ru/groups/Permic.shtml>) hosted by the Institute of Linguistics, whose language list is used for official purposes.

transcribed texts published by Keľmakov (Кельмаков 1981) and Nasibullin (Насибуллин 1981), as well as from the Corpus of Standard Udmurt,<sup>2</sup> the Beserman Multimedia Corpus<sup>3</sup> and the Corpus of Tatyshly Udmurt.<sup>4</sup>

The paper is structured as follows. In Section 2, I introduce the subset of Udmurt discourse particles that are the subject of my investigation. In Section 3, I demonstrate how the factors of scope and position inside the sentence can affect clitic placement inside a cluster. In Section 4, the dialectal distribution of clitic templates and variation is covered. Section 5 discusses historical processes that might have led to the observed distribution of discourse particles in Udmurt dialects. This is followed by a conclusion.

## 2. Udmurt discourse particles and clitic clusters

### 2.1. Clitics to be investigated

Udmurt counts a couple dozen clitics that can be subsumed under a broad notion of discourse particles in at least some of their uses. In this paper, I am initially focusing on the eleven particles listed in Table 1, although only a subset of them will be discussed in detail. All of them are enclitics, i.e. they phonologically attach to the right of their host word. In what follows, I am going to use the terms *discourse particles* and *discourse clitics* interchangeably when referring to the eleven items in question. With the exception of *na* 'yet/still' (see details in Section 4), they do not affect the stress placement in the phonological word.<sup>5</sup> As is often the case with pragmatic elements, the translations provided are loose and do not cover the whole range of meanings the particles can have. In order to be more precise, I will translate some of the particles with labels under which their rough equivalents have been described in the typological literature. Some explanations regarding the ranges of meanings available for each particle will be offered next to the examples (2)–(9) that introduce them.

A large number of items in question have been borrowed from Turkic languages throughout the long history of their coexistence with Udmurt in the Volga-Kama area. Therefore, it would be reasonable to investigate to which extent their syntactic properties in Udmurt could be explained by their properties in the donor language. The "Source" column contains the borrowing status. "Native" in this context means 'apparently not borrowed through a relatively recent contact, i.e. from any Turkic language or Russian'.

Before moving further, a couple of short comments has to be made as to the etymology of some of the clitics.

First, it has been claimed by Tarakanov (Тараканов 1993 : 162), who, in turn, was referring to the opinion of Munkácsi, that Udmurt (*i)ni* 'already' is a borrowing of the Tatar *inde* 'id.'. Csúcs (1990 : 180) marks this etymology as dubious. Its areal distribution (it exists and is very frequent everywhere, including Northern areas where many other Tatar borrowings did not spread) and the existence of a very similar item in Komi, which is closely related to Udmurt, also suggest that it was not borrowed from Tatar.

<sup>2</sup> [http://udmurt.web-corpora.net/udmurt\\_corpus/search](http://udmurt.web-corpora.net/udmurt_corpus/search).

<sup>3</sup> <http://multimedia-corpus.beserman.ru/search>.

<sup>4</sup> <http://udmurt.web-corpora.net/tatyshly/>.

<sup>5</sup> This has been mentioned in the descriptive literature (e.g. Winkler 2011 : 31) and is supported by my own observations. However, it has to be mentioned that the claims about the stress placement — both those found in the literature and those made by the author — have never been tested instrumentally, to the best of my knowledge.

List of discourse particles to be investigated

Clitic (standard / most frequent variant)	Dialectal variants	Translation/ label	Source
<i>no</i>		'and'; additive focus particle	probably native (Майтинская 1982 : 126; Федюнева 2008 : 217–218)
<i>ik</i>		emphatic identity (König 1986)	< Chuvash (Wichmann 1987 : 65)
<i>ni</i>	<i>ni, in, ni, ini, in</i>	'already/anymore'	native: cognate <i>nin/ni/nin/ni</i> in Komi (КЭСК 192)
<i>inde</i>	<i>inde</i>	'already/anymore'	< Tatar (Csúcs 1990 : 180)
<i>na</i>		'yet/still'	native: cognate <i>na</i> in Komi (КЭСК 185)
<i>aj / al'i</i>	<i>jaj, ej, jej, al', aji, ajko, al'i</i>	'yet/still'	< Tatar (Тараканов 1982 : 165)
<i>a</i>	<i>wa, ja</i>	question marker	native: cognate <i>e</i> in Komi (КЭСК 209)
<i>pe</i>		reportative marker	native: cognate <i>pe/pe</i> in Komi (КЭСК 227)
<i>d̄ir</i>	<i>d̄ar, d̄är</i>	'probably'	< Tatar (Csúcs 1990 : 162)
<i>uk</i>	<i>ug</i>	enimitive (Panov 2020)	native (Arkhangelskiy 2023a)
<i>eś</i>	<i>jeś</i>	enimitive (Panov 2020)	< Tatar (Csúcs 1990 : 172)

Second, *uk* 'enimitive' has been called either a Chuvash (Федюнева 2008 : 215) or a Tatar (Csúcs 1990 : 298; Bartens 2000 : 322–323) borrowing, without any substantial justification.

However, both theories have serious flaws, as there are significant mismatches in phonology, syntax and semantics between Udmurt *uk* and both of its proposed sources. I argue (Arkhangelskiy 2023a) that *uk* is actually a native item, which has grammaticalized from a reverse-polarity tag question construction.

## 2.2. Functions of discourse clitics

In this section, I explain the primary functions of each of the discourse clitics under consideration, without going into much detail. Whenever semantics of a clitic has been described in the literature, I provide a reference; claims not supported by references are based on my own observations. Examples (2)–(9) below and the corresponding comments only illustrate typical usages of the clitics.

The first clitic, *no*, is the standard coordinating conjunction. Apart from that, it has a range of other uses. All uses of *no* fall under the scope of my investigation. For example, in (2), it is used as an additive particle. Its main function here is to introduce the additive presupposition (Karttunen, Peters 1979), i.e. to signal that there is at least one other focus alternative.

(2) Udmurt (Standard; M̄inam Udmurtfije, 11.12.2012)

*Tue so-os-len rad-a-z̄i Alnaś-jos=no pot-i-z̄i.*  
 this.year that-PL-GEN series-ILL-3PL.POSS Alnash-PL=ADD come.out-PST-3PL

[Volumes of cattle farming have been decreasing in several districts for years.]  
This year, the Alnash district also joined their numbers.'

The clitic *ik* can be primarily described as a particle of emphatic identity, in terms of König (1986). It marks an argument in a proposition as co-referent with an argument of another proposition present in the discourse. Further information on its properties can be found in (Закирова 2019; Zakirova, Kozlov 2022; Зубова 2016).

- (3) Udmurt (Standard; Закирова 2019 : 45)

*Bakč-a-ŋn Maša uža, nŋlpi-os-ŋn=no Maša=ik puk-e.*  
garden-LOC M. work:PRS.3SG child-PL-INS=ADD M.=ID sit-PRS.3SG

'Masha works in the garden, and she is also the one who does the babysitting.'

The clitics *na* 'still/yet' and *ni* 'already/anymore' ((4)–(5), (8)–(9)) can be described as time adverbials in their primary sense. The same concerns *aj/al'i* (7) and *inde*, which are synonymous with *na* and *ni*, respectively, in most (although not all) contexts. However, they have other important senses, which cannot be directly reduced to their primary functions. For example, *na* and *aj/al'i* can mean 'apart from that' and 'else' in certain contexts, just like German *noch* and its other European counterparts. *aj/al'i* (but not *na*) also functions as a hortative particle.

- (4) Udmurt (Bagrash-Bigra, Central/Southern)

*Ta-iz pomidor-jos vož-eš=na, gord-jos-se*  
this-3SG.POSS tomato-PL green-PL=**still** red-PL-ACC.3SG.POSS

*mi ši-i-mj=ni.*

we.EXCL:NOM eat-PST-1PL=**already**

'These tomatoes are still green, we have already eaten the red ones.'

*A*, which marks question focus (5), and *pe*, which marks reported speech (6), are rather standard representatives of the respective cross-linguistic categories. *A* attaches to the head of the focused phrase. *Pe* tends to occupy the second position inside the reported clause, but can be placed elsewhere as well. Although *a* and *pe* are not discourse particles in their main function, *pe* can also be used without any reported speech in a pragmatic sense roughly equivalent to the English 'kind of' (Arkhangelskiy 2023b).

- (5) Udmurt (Standard; Udmurt duńne, 17.07.2009)

*Ma-ke vošt-išk-i-z=a?*  
what-INDEF change-DETR-PST-3SG=Q

'Did something change?'

- (6) Udmurt (Standard; Udmurt duńne, 02.11.2012)

*Ta-iz-lj=pe 10 mill'ion mańet pala kule.*  
this-3SG.POSS-DAT=REP 10 million rouble around is.required

'That one needs around 10 million roubles, he says / they say.'

*Dŋr* (7) implies that the speaker is inclined to believe the proposition to be true, but is not completely sure. It can be used both in declarative sentences and in questions.

- (7) Udmurt (Tatyshly district, Southern Peripheral)

*So-os-len nŋlpi-zŋ u? ve?l-ŋ=dŋr=aj?*  
that-PL-GEN child-3PL.POSS NEG.PRS walk-CNG.SG=**probably=yet**

'Their child probably doesn't walk yet?'

*Uk* (8) and *eś* (9) are representatives of a cross-linguistic category called enimitive by Panov (2020). This means that they mark the proposition as a piece of information that is presumably uncontroversial for both the speaker and the addressee, and can therefore be used as an argument in the discussion. Examples of European enimitives include German *ja* and Russian *ведь*; English does not have a grammaticalized enimitive. Although *uk* has not been called enimitive before, such an analysis is corroborated by the fact that it is typically translated by European enimitives in the dictionaries, such as Russian *ведь*, German *ja* or Finnish *-han/-hän*. Zubova (Зубова 2016 : 448–449) provides a somewhat more detailed analysis of the subsenses of the Russian *ведь* available for *uk*.

(8) Udmurt (Vavozh, Central)

*Mon tɨnɨd vera-j=uk=ini, oʒi kar-ɨnɨ*  
 I:NOM you.SG:DAT tell-PST.1SG=ENIM=**already** so do-INF  
*ug jara!*  
 NEG.PRS be.allowed:CNG.SG  
 'I've already told you you cannot do that, haven't I?'

(9) Udmurt (Tatyshly district, Southern Peripheral)

*Mon tɨnɨd vera-j=eś=ini, oʒɨ kar-ɨnɨ*  
 I:NOM you.SG:DAT tell-PST.1SG=ENIM=**already** so do-INF  
*u? ʒara!*  
 NEG.PRS be.allowed:CNG.SG  
 'I've already told you you cannot do that, haven't I?'

### 2.3. Frequency properties of clitics

As a further step in getting acquainted with the object of my investigation, consider the frequencies of the clitics provided in Table 2. The figures in the three columns represent the ipm (items per million) values in three corpora mentioned above: Corpus of Standard Udmurt (contemporary written texts, mostly newspapers; 9.57M words), Beserman multimedia corpus (spoken; 114K words at the time of writing) and Corpus of Tatyshly Udmurt (spoken; 10K words at the time of writing). In a sense, the two spoken corpora represent the geographical extreme points of the Udmurt-speaking area: Beserman is close to its North-Western edge, while Tatyshly (Southern Peripheral supradialect) is at its South-Eastern edge.

One thing which is immediately clear from Table 2 is that the clitics in question are extremely frequent (some only in certain dialects though). Their frequency is comparable to, or higher than, that of their most frequent European counterparts. For reference, *noch* 'yet/still' and *schon* 'already; indeed' in written German have the ipm values of 3153 and 1544, respectively (DeReWo 2014). One reason for this is that the Udmurt particles tend to be available in a wider range of contexts; see the discussion of the examples (36)–(38) in Section 5, as well as the description of meanings available to *ni* by Zubova (Зубова 2016). However, according to my preliminary observations, they also have a higher probability of appearance than their German or Russian equivalents in contexts where the latter would be felicitous as well. Further investigation is needed to find out why this is happening.

Another obvious conclusion is that there are clitics that exist only in some of the varieties. Out of the three varieties listed, Tatar borrowings *inde* 'already' and *eś* 'enimitive' only exist in the Tatyshly dialect, *uk* is absent there, and *aj/alli* 'yet/still' does not exist in Beserman. Also, out of the pair of quasi-synonyms

Table 2

Relative frequencies of discourse particles in three corpora<sup>6</sup>

Clitic	ipm (Standard)	ipm (Beserman)	ipm (Tatyshly)
<i>no</i>	35,982	18,235	33,961
<i>ik</i>	8810	3891	4207
<i>ni</i>	5931	17,043	10,418
<i>inde</i>	0.8	0	3606
<i>na</i>	2656	5337	1703
<i>aj/al'i</i>	656 [484, 867]	0	5324
<i>a</i>	1781 [1562, 1982]	14,998 [12,142, 17,969]	4756
<i>pe</i>	1530	5775	4809
<i>dj̄r</i>	764 [654, 869]	29	2688
<i>uk</i>	716	11,120	0
<i>eś</i>	0	0	3807

*na* and *aj/al'i* 'yet/still', the former, i.e. the native one, is preferred in the standard written language, but dispreferred in Tatyshly. Although *dj̄r* 'probably' exists in Beserman, it is far less common there than in the South (where it was borrowed from Tatar).

The analysis of a subset of these clitics by Zubova (Зубова 2019) helps make sense of the rest of the discrepancies. Her research reveals that genre and register may dramatically affect their frequencies. While *ik* 'emphatic identity marker' is much more widespread in written texts, *na* 'yet/still', *ni* 'already' and *uk* 'enimitive' exhibit the opposite trend, being more common in less codified texts. The differences in the genre composition of the corpora (especially monologues/dialogues ratio) must also explain the diverging figures for *a* 'question marker' and *pe* 'reportative marker'. To the best of my knowledge, these two are used rather uniformly across Udmurt dialects, being obligatory (*a*) or nearly obligatory (*pe*) in the same sets of contexts.

### 3. Clitic clusters and templates

#### 3.1. Clitic ordering in a cluster

Given the abundance and high overall frequency of discourse clitics in Udmurt, it comes as no surprise that many sentences contain more than one. When multiple clitics choose the same host, all of them form one phonological word, as in (1) or (10).

<sup>6</sup> The occurrences of the particle *uś̄ya* 'isn't it' in Tatyshly, which originates from a combination of a negative verb with a question marker, were excluded from the count for *uk*. Since it is often impossible to tell apart *no* as a coordinating conjunction and as an additive focus particle, all occurrences were counted. In other cases, homonymous entities were disambiguated through random sampling. *Al'i* 'yet/else' is homonymous (and cognate) with *al'i* 'now'; *a* 'question particle' is homonymous with the conjunction *a* 'and, but' borrowed from Russian; *dj̄r* 'probably' is homonymous with *dj̄r* 'time'. This poses a problem since the corpora I used contain ambiguous analyses. When calculating the ipm values for these items in large corpora, I manually disambiguated a sample of at least 100 search hits. The total number of hits was then multiplied by the proportion of the item in question in this sample to obtain an estimate. Since the search hits were randomized, the samples were representative of the respective corpora. For the figures obtained through sampling, Wilson's 90% confidence intervals are indicated in brackets.



- (10) Beserman (Shamardan)  
*Kureg=no=a=ńi puza?*  
 hen=ADD=Q=**already** lay.eggs:PRS.3SG  
 'Is the hen laying eggs already as well?'

The ability to form chains, or clusters, is a well-known property of clitics observed in many languages. Voluminous research on (mostly pronominal) clitic clusters in Romance, Slavic and Greek languages allowed linguists to arrive at certain generalizations. It turns out that if several clitics meet within one phonological word, they tend to be rigidly ordered (Spencer, Luís 2012 : 47), i.e. out of several theoretically possible orderings only one is usually acceptable.

European clitic clusters have been successfully modeled with the help of clitic templates (Perlmutter 1971; Bonet 1991). Templatic organization means that all clitics in a language are split into a number of sets, and the sets can be linearly ordered from left to right, forming a template sequence. Clitics within a set cannot co-occur, and whenever several clitics from different sets form a cluster, their mutual order is the same as the order of their respective sets in the template. The set of clusters accepted by a template can be further narrowed down by additional constraints which disallow certain combinations of clitics. The most widely known example of such a constraint is the *"\*me-lui constraint"*, or Person Case Constraint in Romance languages. According to it, direct object clitics of any person other than 3 (e.g. French *me* 'me') cannot co-occur with dative clitics (e.g. French *lui* 'to him/her'). This constraint has been later shown by Sheehan (2020) not to be restricted to clitic clusters.

As we will see below, the template approach mostly works for each individual variety of Udmurt. Even if my insufficient data does not allow for complete reconstruction of a template in each of the varieties surveyed, it is in principle compatible with the idea of a template. Two potential contradictions to templatic organization can be resolved by stipulating two different items for *no* 'and / additive particle' and taking the syntactic domain into account. However, before we get into that, two general features of Udmurt clitic clusters have to be mentioned, which make Udmurt somewhat different from the well-known Romance and Slavic cases.

First, although the ordering of clitics in a given variety is often rigid, sometimes it is not completely strict. An example of an entirely rigid ordering is (10), where all alternative orders of the three clitics are considered unacceptable by Beserman speakers. However, consider two further examples from the same variety (11)–(12).

- (11) Beserman (Shamardan)  
*Tžnđd so-je vera-z=a=ńi?*  
 you.SG:DAT that-ACC tell-PST.3SG=Q=**already**  
 'Did s/he tell you that already?'

- (12) Beserman (Shamardan)  
*Kal' muket=uk=ńi.*  
 now different=ENIM=**already**  
 'It (the life) is different now, isn't it.'

These orders, *a=ńi* and *uk=ńi*, are clearly considered default by the speakers. However, opposite orders (*ńi=a* and *ńi=uk*) are judged as acceptable by most of them, even if most respondents accept them with reservations, adding that "they sound somewhat worse" or "that's how they would speak in another village". There are also four examples of *ńi=a* and three examples of *uk=ńi* in the Beserman corpus,

produced by different speakers. Although the "default order" is two orders of magnitude more frequent (133 and 182 occurrences in the corpus, respectively), it is nevertheless not absolutely strict.

Moreover, there are varieties where certain pairs of clitics do not have a preferred order, exhibiting both interpersonal and intrapersonal variation; these cases will be discussed in Section 4.

As noted by 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". Absence of variation, inferred from the strict application of the template model, is indeed sometimes presupposed: e.g. Zimmerling and Kosta (2013 : 179) define clitic clusters as "strings of clitics arranged in a rigid order according to language-specific rules". Udmurt data demonstrates that variation is actually part and parcel of clitic clustering.

Second, different varieties of Udmurt have different ordering templates. Just as phonological features or lexical items, the templates clearly demonstrate areal distribution, i.e. settlements that are close to each other are likely to have the same order of clitics in a cluster. Free variation within a template is observed at the border between areas with different templates. Interestingly, some parts of the template are invariable across the whole Udmurt-speaking area, while others fluctuate a lot. This means that a template model — even if one allows for some variation within a template — is only applicable to individual varieties of Udmurt, but not to the language as a whole.

In what follows, I will mostly deal with this dialectal distribution of templates.

## 3.2. Dealing with apparent counterexamples

As mentioned earlier, there are two potential obstacles for the applicability of the template model to Udmurt data — apart from the ubiquitous variation. The first one concerns an apparent violation of transitivity by the *no* 'and / additive particle' clitic and manifests itself uniformly across dialects. The second one involves an apparently semantically motivated variation in position of the reportative clitic *pe*. In both cases, there is an interplay between the scope of the clitic and its position.

### 3.2.1. Violation of transitivity by *no*: narrow vs. wide scope

A binary relation " $<$ " is called transitive if for any  $a$ ,  $b$ , and  $c$ , the statements " $a < b$ " and " $b < c$ " imply " $a < c$ ". This is exactly the sort of relation imposed by a clitic template, since it arranges clitics or sets of clitics along a single axis. In fact, it has long been known that there are cases where such a property does not hold in the strong sense. Consider the "*\*me-lui* constraint" in French. While, in terms of the ordering relation for French pronominal clitics proposed by Perlmutter (1971 : 57), *me* '1<sup>st</sup> person clitic (object/dative)'  $<$  *le* '3<sup>rd</sup> person clitic (object)' and *le* '3<sup>rd</sup> person clitic (object)'  $<$  *lui* '3<sup>rd</sup> person clitic (dative)', one cannot deduce that *me*  $<$  *lui* since the two clitics do not combine and the relation between them is therefore undefined. This is, however, a minor issue that can be remedied by introducing constraints which would disallow certain combinations of clitics. In that 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 e.g. by Heggie and Ordóñez (2005 : 15–16).

Nevertheless, there is another potential 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 be irreconcilable with the idea of a template. However, at the first glance, this is exactly what happens in Udmurt. We will consider Beserman variety in this subsection, however relevant properties of *no* hold in every Udmurt variety I examined.

Consider Beserman examples (13)–(15), which include three clitics, *no* ‘and / scalar focus particle’, *a* ‘question marker’ and *ni* ‘already’.

(13) Beserman (Shamardan)

*Mon tñnñd pis-te vand-i=ni,*  
 I:NOM you.SG:DAT wood-ACC.2SG.POSS cut-PST.1SG=already  
*pñl'-ñl-ono=no=a?*  
 chop-ITER-DEB=ADD=Q

‘I’ve already cut the wood for you, do I have to chop it as well?’

(14) Beserman (Shamardan)

*Ton šñd-de kñs-i-d=a=ni?*  
 you.SG:NOM soup-ACC.2SG.POSS turn.off-PST-2SG=Q=already

‘Have you turned off [the stove under] the soup already?’

(15) Beserman (Shamardan)

*Šž-ez zeg=kad'=ni=no, malñ-ke*  
 sheep-3SG.POSS big=like=already=ADD why-INDEF  
*odig pol=no ez=na pija.*  
 one time=ADD NEG.PST.3=yet give.birth:CNG.SG

‘The ewe is kind of big already, but it has not yet given birth even once for some reason.’

Assuming the template model is correct, how would the three clitics be ordered in the template? The examples suggest that in clusters *no* is followed by *a* (13), *a* is followed by *ni* (14), and *ni* is followed by *no* (15). However, this creates a non-transitive, circular pattern ( $no < a$ ,  $a < ni$  and  $ni < no$ ), exactly like in the hypothetical scenario described above. In addition, all three clitics can appear at once, as in (9), so the situation cannot be remedied with the help of co-occurrence constraints. Importantly, the order of clitics in each of the examples is rather rigid, with the alternatives judged as unacceptable or, in the case of (14), at least degraded by the Beserman speakers. Therefore, this conundrum could not be written off as a mere consequence of free variation.

There is, however, a solution to this apparent contradiction. It turns out that the position of *no* in clitic clusters is strictly determined by its semantics and scope. In (10) and (13), where *no* precedes the other two clitics, it acts as an additive particle, translated as ‘as well’ or ‘even’; the same happens if it is used as a normal coordinating conjunction. In this case, *no* can have variable scope, e.g. over a NP in (10). The meaning of the first *no* in (15) is rather different. There, it has a scope over a clause, attaches to its head and roughly means ‘although the speaker commits to the truthfulness of the proposition expressed in the clause, some natural implications do not hold or are not guaranteed to hold by the speaker’. Among its other functions, this latter *no* is used in concessive clauses (usually in combination with *ke* ‘if’). However, its area of applicability is broader than just concessive contexts. For example, it is often used in monoclausal sentences, where the implications that do not hold are not spelled out explicitly in a sepa-

rate clause. The development of this sense could have been caused or reinforced by the etymologically unrelated Russian conjunction *HO* 'but' (see Kaysina 2015 : 225).<sup>7</sup>

By splitting *no* into two distinct lexical items, 'additive' and 'contrastive', we save the template approach. The two items receive distinct slots in the templates: the additive one, close to their left border, the contrastive one, to the right one. They also have different compatibility constraints. While 'additive' *no* is compatible with pretty much any other clitic, the 'contrastive' one is apparently incompatible with those that contradict the speaker's commitment to the information conveyed in the clause, such as *a* 'question marker' or *dj̄r* 'probably'. This generalization holds across dialects, so that mutual orders in examples like (15) and (16) are the same in all varieties I examined.

(16) Beserman (Shamardan)

*So-ja-z ar-e so ug vel't-δ val,*  
that-ILL-3SG.POSS year-ILL that NEG.PRS walk-CNG.SG be:PST

*a ta-ja-z ar-e vel't-e=no*  
and this-ILL-3SG.POSS year-ILL walk-PRS.3SG=ADD

*biž-δl-e=no=ni.*

run-ITER-PRS.3SG=ADD=**already**

'He (their child) did not walk last year, but this year he already walks and runs.'

With this apparent contradiction resolved, Beserman clitics can be aligned in a template, at least in terms of their default order (Table 3). Only those clitics that exist in Beserman and for which I have enough data are listed here. The additive *no* is labeled as *no<sub>a</sub>*, and the contrastive *no* is labeled as *no<sub>c</sub>*.

Table 3

Beserman clitic template  
(based on corpus data additionally checked with elicitation)

<i>na</i>	<i>no<sub>a</sub></i>	<i>a</i>	<i>d̄δr</i>	<i>uk</i>	<i>ni</i>	<i>no<sub>c</sub></i>	<i>pe</i>
'yet/still'		'question marker'	'probably'	'enimitive'	'already'		'reportative'

Looking at the rest of the clitics, we observe the same trend: obligatorily narrow scope puts the clitic close to the host word, while obligatorily clausal scope places it at the end of the cluster.

One example of the narrow-scope clitic is *ik* 'emphatic identity marker'. When it is used outside of a handful of lexicalised combinations, it always has a narrow scope, marking an argument, an adjunct or (rarely) a predicate as focal and identical to the one in the preceding discourse. It rarely combines with other particles, but the one it easily co-occurs with, *a* 'question marker', is placed to the right of it in every dialect (17).

(17) Udmurt (Staraya Monya, Southern)

*Ta sumka-jez tatčj̄=ik=a pon-o-m?*  
this bag-ACC here:ILL=ID=Q put-FUT-1SG.DELIB

'Shall I put this bag here as well?'

<sup>7</sup> Nevertheless, we are clearly dealing with a native item in the examples above, rather than with a borrowed conjunction. Unlike Russian proclitic *HO*, Udmurt 'contrastive' *no* is an enclitic and can be used in a monoclausal sentence; it also has a wider range of acceptable contexts than the Russian *HO*.

### 3.2.2. Variable position of the reportative *pe*

The reportative clitic *pe*<sup>8</sup> provides an opposite example. Its core function is to mark reported speech or thought, so it is normally encountered in sentences that contain a reported clause and, possibly (although not obligatorily), a framing/main clause, such as *šâr šue* 'says the mouse' in (19).

(18) Beserman (Shamardan)

*Zarńi koko-de=pe mânâm vaj.*  
gold egg-ACC.2SG.POSS=REP I:DAT bring:IMP.2SG  
'Bring your golden egg to me [s/he says].'

(19) Beserman (Shamardan)

*"Mone=no=a=pe možet lež-o-dâ," šâr šu-e.*  
I:ACC=ADD=Q=REP maybe let-FUT-2PL mouse say-PRS.3SG  
'"Maybe you will let me in as well?" says the mouse.'

Corpus data that I examined and elicited examples with *pe* that I collected in the field reveal that, in all dialects, *pe* is in principle grammatical in any position inside the whole sentence, including the framing clause. It seems that its only positional constraint is that it cannot appear inside PostPs or NPs.<sup>9</sup> However, normally *pe* appears inside the reported clause. Its exact position within the clause can differ. In most varieties, including Standard Udmurt and Beserman, *pe* prefers to attach to the first phrase (e.g. NP or PostP) of the reported utterance (18), which makes it a second-position clitic (Bošković 2016, Haspelmath 2022 : 12). This position is however non-rigid, since even in these varieties *pe* occasionally appears in other positions or several times within the reported clause. In some varieties, *pe* may prefer other positions. For example, it seems that in Tatyshly Udmurt, it is more often encountered at the end of the reported clause; further research is needed to clarify its properties across dialects. In any case, *pe* always has a scope over an entire clause, marking it as reported. In clitic clusters, *pe* cross-dialectally almost invariably occupies the rightmost slot (19).

There are however problematic examples like (20), where *pe* is followed by another clitic.

(20) Udmurt (Tatyshly district, Southern Peripheral)

*Moskva-jân dâšet-šk-iškom=pe=ješ tiñi.*  
MOSCOW-LOC teach-DETR-PRS.1PL=REP=ENIM here  
'Look, [she says,] we are studying in Moscow [, doesn't she].'

Although examples like (20) are rare, I have attested them in different dialects, as well as in the standard language. What's more, this does not happen due to limited free variation, as was the case with some Beserman clitics in Subsection 3.1. Comparing such examples with those where the enimitive clitic (*uk* or *eš*, depending on the dialect) precedes the reportative *pe*, we can conclude that the 'outer' enimitive belongs to another syntactic domain than the 'inner' one. In (20), the speaker uses *ješ* to mark the proposition 'she says they are studying in Moscow'

<sup>8</sup> The self-reportative clitic *pej/pâj/pj/pi* (which, unlike *pe*, does not exist in some dialects) seems to follow the same pattern, but I did not include it in my investigation due to its much lower frequency. See Teptiuk 2019; 2021 for more information on Udmurt reportative markers.

<sup>9</sup> To be precise, this constraint is also not absolute, as *pe* can occur e.g. between the nominal head and its adjectival modifier if one of them is in focus, while the other is part of the sentence topic.

as an uncontroversial piece of information that the addressee should accept as an argument in their discussion. The same happens in (21) with *uk*. The 'outer' particles thus describe the point of view of the current speaker and could be said to belong to the 'matrix' of the reported utterance (in terms of Spronck, Nikitina 2019) and thus to a different syntactic domain than *pe*. In (22), on the other hand, the enimitive, probably used here to signal surprise on the part of the reported speaker, who did not expect to see potatoes that big, belongs to the reported clause.

(21) Udmurt (Standard, Udmurt duńne, 01.09.2009)  
*Škola-ın gožja-šk-ınj=no čota-šk-ınj djšet-o=pe=uk!*  
 school-LOC write-DETR-INF=ADD count-DETR-INF teach-PRS.3PL=REP=ENIM  
 'But don't they say that the school is for learning how to write and count?'

(22) Udmurt (Standard, udmurto4ka.blogspot.ru)  
*Kjk mjžjk bjdža kartoška-os-tj=uk=pe.*  
 two fist the.size.of potato-PL-2PL.POSS=ENIM=REP  
 '[S/he says,] "Your potatoes are double the size of a fist, aren't they?"'

In order to account for this scope sensitivity of *pe*, the template model must be further restricted to the clitics that belong to the same syntactic domain (e.g. originate from the same clause).<sup>10</sup> With this additional constraint, *pe* is always the rightmost clitic in the cluster.

### 3.3. Position of clause-level clitics in clusters

The fact that, barring the situation described above, *pe* in Udmurt is always cluster-final, is probably connected to the fact that it is the only clitic that can freely move around the clause and, in most dialects, prefers to occupy the second position. It is generally assumed in syntactic theories that the placement of clause-level second-position clitics is a result of movement to a higher position in the syntactic tree. If that is correct, we could assume that the clitics that undergo such movement are placed further from the host word in clitic clusters than clause-level clitics that do not undergo movement, let alone narrow-scope clitics. This hypothesis is corroborated by the data from standard Komi-Zyrian, where not only *pe* 'reportative marker', but also *ńin* 'already' (cognates of the Udmurt *pe* and *ńi*, respectively) can move to the second position — unlike in Udmurt, where clause-level *ńi* is normally attached to the predicate. In Komi, these two clitics are attested in the order opposite to the one Udmurt prescribes (23). Note that it is clear from the context of that sentence that *ńin* 'already' originates in the reported clause, in contrast to the Udmurt examples with clitics that follow the reportative *pe*.

(23) Komi-Zyrian (Standard; Komi mu, 14.06.2016)  
*Ki-ńjm=pg=ńin uś-i-s.*  
 hand-1PL.POSS=REP=already fall-PST-3SG  
 'They sank into despair already (lit. their hands fell down), they say.'

Since both clitics undergo movement in Komi-Zyrian, the aforementioned ordering rule does not apply to them, which is why *ńin* 'already' can follow the reportative clitic in the cluster.

<sup>10</sup> Impressionistically, I would say that the 'outer' *uk* moves to the reported clause from the main clause after the clitic clusters in the latter have been formed. This, however, would require stipulating elliptic main clauses in examples like (21)–(22), as well as adopting one of the syntactic formalisms and conducting rigorous testing. As this goes beyond both the scope of this paper and my qualification, I use an intentionally vague term 'syntactic domain' here and leave the rest to further research.

The rest of the clitics under consideration, in terms of their scope and placement, occupy an intermediate position between the narrow-scope *ik* 'emphatic identity marker' and the clausal-scope *no* 'contrastive' and *pe* 'reportative marker'. Semantically, they can introduce focus alternatives for constituents of various size. In terms of their placement, *a* 'question marker' always attaches to the focused phrase of an alternative question, while the rest of the clitics usually attach to the predicate, even if they actually modify a particular focused argument or adjunct in the clause. The scope and position, which were demonstrated above to be an important factor in clitic placement, do not play a role when ordering those. So in a sense, the contest over who gets a place closer to the host word does not have obvious winner candidates. Indeed, their placement is subject to a lot of variation, as we will see below. In what follows, I will be only concerned with the clitics other than *no* (both 'additive' and 'contrastive'), *pe* 'reportative' and *ik* 'emphatic identity'.

#### 4. Dialectal variation: constants and variables

##### 4.1. Dialectal templates

If we assume that the problems outlined in Subsection 3.2 are indeed resolved, and free variation can be allowed, then the data I have at hand allows me to come up with a clitic template for each individual variety. However, it turns out that clitics are ordered differently in different varieties, so each variety has its own separate template. Templates reconstructed for different varieties of Udmurt can be so different that it is impossible to produce a general template that would describe the entire language.

The observed variation is, first and foremost, geographic (and, therefore, dialectal) in nature. Settlements located next to each other tend to have same templates, and settlements with same templates form contiguous areas. Whenever I worked with multiple speakers who were born and raised in the same settlement inside one of such areas, no other macrosocial factor (such as gender, age or level of education) played any visible role. However, varieties of settlements located at the borders between different areas sometimes allow for variation. Unfortunately, I do not have enough data to determine if factors like age affect the preference for one of the possible orders in those settlements. One other parameter that could theoretically confound the data is the influence of the standard written register, which most respondents learned in the school and some actively use. While I cannot reliably exclude this possibility, I nevertheless have serious doubts that this would be an important factor. Unlike heavily regulated vocabulary and morphology, things like clitic ordering are seemingly considered "ephemeral" by both Udmurt speakers and language authorities. As a consequence, the latter do not try to standardize them, while the former do not feel the urge to comply with the (apparently non-existent) norm in this respect.

If we compare templates for individual varieties, we will see some stable elements common to all of them, as well as some that are subject to significant variation. Clitics with both stable and variable positions can be seen in Table 4, which presents a unified template schema for all varieties.

The structure of this schema is the following. Each cell corresponds to one clitic slot in a cluster. For adjacent cells, the left slot precedes the right one. Items without parentheses correspond to clitics that can be analyzed as holding invariable positions across varieties. Items in parentheses indicate clitics that appear in the respective slot in at least one dialect, but not in all of them.

Table 4

## Unified template schema for Udmurt clitics

<i>na</i>	( <i>ńi</i> )	<i>a</i>		( <i>ńi</i> )	<i>aj/al'i</i>
	( <i>ńi</i>   <i>dj̄r</i> )	( <i>ńi</i>   <i>dj̄r</i> )	<i>uk</i> / <i>eś</i>	( <i>ńi</i>   <i>dj̄r</i> )	

As the enimitive particles are incompatible with the question marker, I represent the part of the schema between the extreme points *na* 'yet/still' and *aj/al'i* 'yet/still' as consisting of two branches, rendered as two non-adjacent table rows, for convenience. The upper one contains *a* 'question marker', and the lower one, *uk/eś* 'enimitive'. Nevertheless, it can be linearized in each of the dialects, provided certain co-occurrence constraints are added. *Dj̄r* 'probably' is supposedly incompatible with *a* 'question marker', which is why it only appears in the lower row. Items that are altogether absent in a minority of the varieties examined are not taken into parentheses. The item *inde* 'already/indeed', which exists in only one of the varieties I have a reasonable amount of data on (Tatyshly Udmurt), also always occurs at the rightmost end of the cluster. However, I have no data regarding its compatibility and potential mutual order with *aj/al'i* 'yet/still' and did not include it in the template.

The stable part of the template consists of the clitic *na* 'yet/still', which always comes first, its quasi-synonym *aj/al'i*, which always comes last, and particles *a* 'question marker' and *uk* or *eś* 'enimitive', which always come in between and do not combine with one another. On the other hand, *ńi* 'already' and *dj̄r* 'probably' can occupy any place between *na* and *aj/al'i*. Moreover, the position of *ńi* in the upper row does not predict its position in the lower row: attested dialect templates include those where it precedes *a* in the upper line, but follows *uk* in the lower one, as well as vice versa. How variable the position of *ńi* can be even within one settlement, can be seen from the template in Table 5 based on the data from Bagrash-Bigra variety (Central/Southern; a point just north of Agryz on the maps).

Table 5

## Clitic template for Bagrash-Bigra (Central/Southern)

*na*    (*ńi*)    *a*    (*ńi*)    *dj̄r*    (*ńi*)    *uk*    *aj/al'i*

## 4.2. Dialectal distribution of pairwise orders

The pairwise orders are clearly subject to areal distribution. In the following maps (see Fig. 1–3), the distributions of three pairs, *ńi/a*, *ńi/uk* and *ńi/dj̄r*, are plotted.<sup>11</sup> Grey background represents the area where Udmurt is spoken<sup>12</sup> (based on the 2010 census data). Each circle corresponds to a settlement for which the corresponding order has been attested. White circles (○) denote orders with *ńi* 'already' on the left, black circles (●) denote opposite orders, and mixed circles denote varieties for which both orders have been attested.

The distributions mapped have to be taken with a grain of salt for two reasons. First, the amount of data behind each data point ranges from one example taken from a published speech sample to dozens of examples produced by and discussed with multiple speakers. In the former case (which, however, subsumes a minority

<sup>11</sup> The order in the *dj̄r/uk* pair cannot always be deduced from the rest, but my data on it is too scarce.

<sup>12</sup> After Jurij Korjakov (<https://minlang.site/lang/udmurtskiy-yazyk#&gid=1&pid=1>).



of data points), it is possible that the variety is actually a mixed one, i.e. allows both orders, one of which has not been attested because of the insufficient data. Second, there is question of comparability of data. During elicitation, which was the primary means of data collection for varieties representing a total of 23 settlements, I used the same questionnaire to control for factors like semantics and scope. The examples I found in corpora or published collections, on the other hand, are rather heterogeneous. Nevertheless, I believe them to be by and large comparable to the data I collected in the field. Comparison of corpus data for Beserman and Tatyshly Udmurt with the respective elicitation responses shows that the two types of datasets yield the same results, and pairwise orders of clitics tend to be the same regardless of the context — with the exception of two cases addressed in Subsection 3.2.

#### 4.3. Factors that may have influenced the dialectal distribution

Even if the sample the maps are based upon is not nearly as geographically dense as the one in the Dialectological atlas (Насибуллин, Максимов, Семенов, Отставнова 2009), and aforementioned precautions apply, this data still provides a good first approximation of the dialectal distribution of Udmurt clitic templates. What conclusions can we draw from it?

First, let's look at the data from a purely dialectological perspective. The isoglosses one may draw based on these maps do not coincide with the supra-dialect borders in traditional dialectal classifications (Кельмаков 1998; Максимов 1999 : 171). This is hardly surprising for several reasons. First, it is known that isoglosses of different kinds do not necessarily correlate with each other. Specifically, syntactic isoglosses can exhibit low correlation with phonological, lexical and

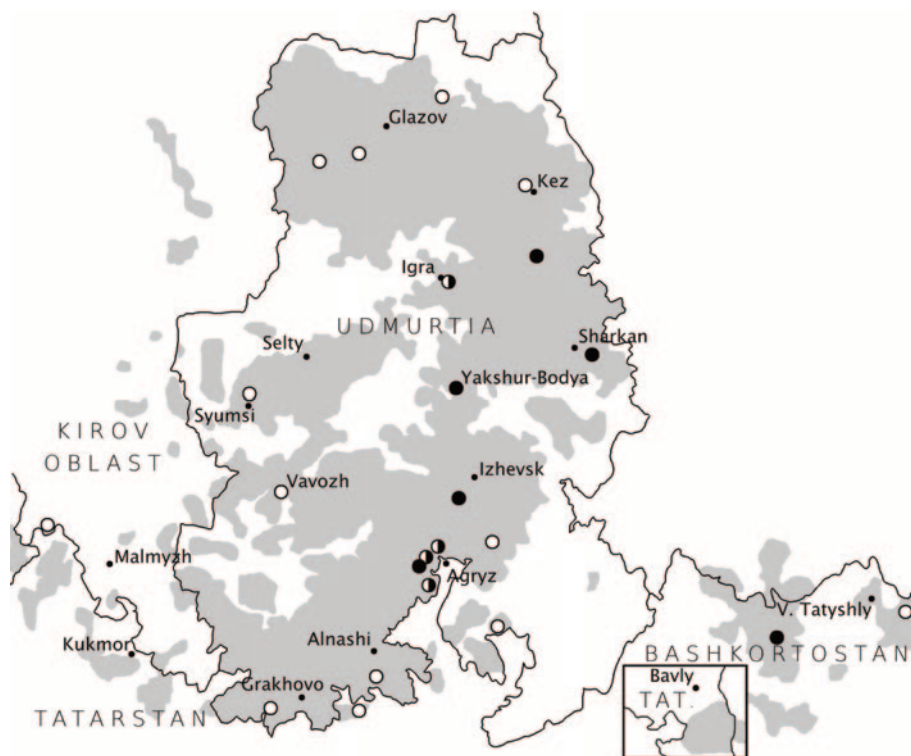


Fig. 1. Areal distribution of the *a ni* (○) vs. *ni a* (●) orders.

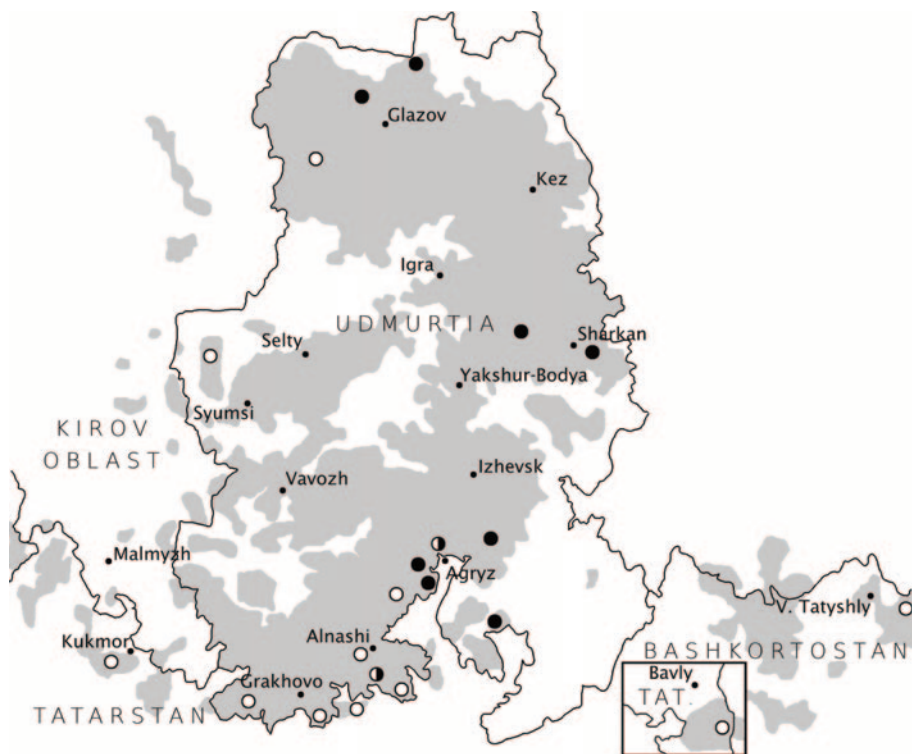


Fig. 2. Areal distribution of the *dij ni* (○) vs. *ni dij* (●) orders.

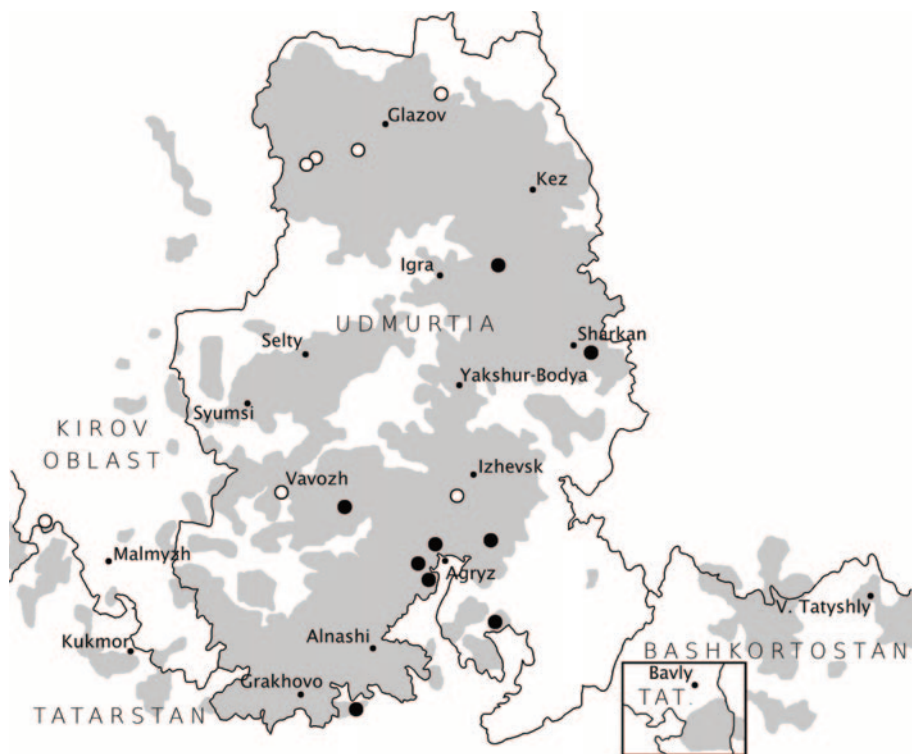


Fig. 3. Areal distribution of the *uk ni* (○) vs. *ni uk* (●) orders.

morphological ones, as demonstrated e.g. by Scherrer and Stoeckle (2016 : 110) for Swiss German. Since traditional Udmurt classifications are based on phonological isoglosses and, to a lesser extent, on vocabulary and morphology, one should not a priori expect clitic ordering isoglosses to align with them well. Second, a dialectometric analysis by Arkhangelskiy (Архангельский 2021) shows that although most of the supradialect borders are rather strong, some are fuzzy, even if one only looks at lexical data. This especially concerns the Eastern part of the border between Northern and Central supradialects.

Although the position of *ñi* 'already' with respect to any of the three other clitics does not automatically define its position with respect to the others, there is visible correlation. There is a large area where *ñi* 'already' occupies the right slot in all three pairs. In terms of traditional classifications, it consists of the western parts of the Northern, Central and Southern Peripheral supradialects, as well as, apparently, the southern and eastern outskirts of the latter. Similarly, in the central and eastern parts of the Central supradialect (the triangle between Igra, Sharkan and Izhevsk) and the central part of the Southern dialect (immediately to the West of Agryz), *ñi* 'already' tends to occupy the left slot in all pairs. The borders between the two areas do not coincide completely for the three clitics, but, nevertheless, seem to lie not very far from one another. Such a distribution would be compatible with a diachronic process whereby *ñi* 'already' started moving either from the inside outwards or from the outside inwards in clitic clusters, outflanking one other clitic at a time (different ones in different varieties). Also, the borders are more vague than the isoglosses one normally sees in dialectal maps. This indicates, apparently, that the transition from one template to another proceeded gradually rather than abruptly. However, further research, preferably involving diachronic data, is needed to support or refute this hypothesis.

Second, the maps allow us to rule out language contact effects. Russian could hardly influence Udmurt clitic templates anyway because Russian equivalents (or particles perceived as equivalents by Udmurt speakers) follow rather different prosodic and syntactic patterns. Unlike in Udmurt, Russian discourse particles are often used as proclitics and are more free to move around the sentence. Tatar, on the other hand, has a more similar system of discourse particles, many of which have actually been borrowed into Udmurt and participate in its clitic clusters. However, comparing the areal distributions with the information on the intensity of language contact, we can conclude that the variation in the mutual order of *ñi* 'already', *a* 'question marker', *dj̄r* 'probably' and *uk* 'enimitive' is unlikely to have resulted from Tatar influence. The intensity of Tatar influence decreases if one moves from the South (many speakers of the Southern Peripheral dialect know at least some Tatar, and some are bilingual in it) to the North. The position of *ñi* 'already', on the other hand, produces more of an East-West divide and involves both areas moderately influenced by Tatar and those where the influence has been minimal. The only way in which Tatar probably has affected Udmurt clitic ordering is the mutual order of the clitics borrowed from Tatar. In the Tatyshly variety, where the largest number of Tatar-borrowed discourse particle has been attested, their order maps that of their Tatar sources (24).

- (24) Tatar:  $-d\hat{a}r < i\check{s} < \ddot{a}le$ ; *inde*  
 Udmurt:  $dj̄r < e\acute{s} < aj/al'i$ ; *inde*

Finally, one takeaway point relevant for the cross-linguistic study of clitic clusters is that semantics clearly does not play any role in ordering discourse clitics.

First, consider the two quasi-synonyms *na* and *aj/al'i* 'yet/still'. Although the sets of contexts where they are used can be slightly different, they have a significant overlap in each dialect, and are actually used interchangeably or simultaneously in most dialects (see next section). Nevertheless, *na* invariably occupies the leftmost slot, and *aj/al'i*, the rightmost one.

Second, *na* 'yet/still' and *ni* 'already' in their core uses can be seen as "semantic twins", just like German *noch* and *schon*, French *encore* and *déjà*, or Russian *еще* and *уже*. One member of the pair normally replaces the other if the sentence containing it is negated. Nevertheless, both Udmurt items for 'yet/still' have cross-dialectally fixed positions in the template, while the position of *ni* 'already' varies both between dialects and within some dialects.

So basically, the positions of discourse clitics within a template, provided there are no scope precedence relations between them, can be better modeled by stochastic diachronic processes rather than explained by their semantics, contact influence or the order of their appearance in the language. But if that is true, then why are the positions of *na* and *aj/al'i* 'yet/still' fixed across dialects? Although I do not have a definitive answer to this, a couple of observations can be made.

In some dialects, *na* seems to be more integrated in the morphology than the rest of the clitics. This tight integration manifests itself in the finite negative verbal complex, which in Udmurt — like in many other Uralic languages — consists of a negative verb followed by a connegative form of the main verb. In principle, all clitics in question (probably with the exception of *inde* 'already') can appear in the position between the negative and the main verbs (Vilkuna 1998 : 212; see Georgieva, Salzman, Weisser 2021 for an analysis).<sup>13</sup> However, in certain dialects *na* has a number of additional properties when placed inside the negative verbal complex than the rest.

First, *na* can appear twice inside the complex, attaching to both the negative verb and the connegative (25). In my data, such doubling is attested mostly in the parts of the Southern (Loloshur-Vozzhi, Kadikovo) and Central (Sjumsi, Ludorvaj, Buranovo) supradialects.

(25) Udmurt (Loloshur-Vozzhi, Southern)

*Ta šž pič'i evəl=ni=no, šžpi ez=na vaj=na.*  
 this sheep small NEG.EXIST=already=ADD lamb NEG.PST.3=**yet** bring:CNG.SG=**yet**  
 'Although this ewe is not young already, it has not yet given birth.'

Second, in combinations like that, at least in some varieties, *na* draws the stress that otherwise would fall on the first syllable of the connegative and/or on the negative verb.<sup>14</sup> No other discourse particle affects stress placement in any way, as far as I know.

Third, certain Southern varieties spoken in the Grakhovo district (the aforementioned village of Loloshur-Vozzhi is situated there) feature a similar construction where the future-tense negative verbal complex interspersed with the double *na* is used with past reference, so (26) is interpreted in the same way as (25).

(26) Udmurt (Grakhovo district, Southern; Атаманов 1981 : 52)

*uz=na iž=na*  
 NEG.FUT.3=**yet** sleep:CNG.SG=**yet**  
 'S/he has not slept yet.'

<sup>13</sup> How often that happens varies a lot between clitics, dialects and registers (Карпова 2006 : 109; Зубова 2019 : 79–80).

<sup>14</sup> See Borise, Georgieva 2021 for an acoustic analysis of the Udmurt stress.

The exact semantic contribution of *na* in (26), beyond its usual meaning 'yet', is unclear. However, most other Udmurt varieties would use the past-tense copula here.

We may conclude therefore, that there are dialects in which *na* 'yet/still' in the negative construction is gradually turning into a morpheme. The fact that no other clitic can appear between *na* and its host word aligns well with this hypothesis.

*Aj/al'i*, on the other hand, has more freedom of movement than the rest of the particles discussed in this section — although, apparently, also only in a subset of the varieties where it exists. In the varieties of Porez (Southern Peripheral), Verkhnjaja Igra, Jenaberdino, Kadikovo (Southern), Buranovo (Central), Stengurt (Northern), *aj/al'i* can move to the sentence-initial position (ceasing to be a clitic) or to the second position (27), (28).

(27) Udmurt (Porez, Southern Peripheral)

*So* *χot'i* *badzjn*, *no* *įžpi=al'i* *ęz* *vaj* *so*.  
 that although big but lamb=**yet** NEG.PST.3 bring:CNG.SG that  
 'Although it (the ewe) is big, it has not yet given birth.'

(28) Udmurt (Stengurt, Northern; Кельмаков 1981 : 91)

*Soku=ved'=al'* *mašina* *ęj* *val*.  
 then=ENIM=**yet** car NEG.PST be:PST  
 'There were no cars yet back then, you know.'

This property can probably be explained by the origin of *al'i*. Just like its Tatar source, *äle*, the Udmurt word is polysemous. Apart from being polyfunctional discourse particles, both *al'i* (in most Udmurt dialects and in standard Udmurt) and *äle* are used as a time adverbial meaning 'now' (which, apparently, is its original sense that gave rise to the discourse uses in Tatar). In this sense, it can be placed anywhere in the sentence, and is often found close to its beginning. Many varieties have both kinds of *al'i*. In all such dialects that I have surveyed, 'lexical' and 'discourse' *al'i* can clearly be described as two separate lexical items. For example, the 'discourse' one is compatible with the past tense, as in (27)–(28), the 'lexical' one is compatible with *ni* 'already' (29), and both can be used in the same sentence simultaneously, as in (30).

(29) Udmurt (Tatyshly district, Southern Peripheral)

*Al'i* *so-os* *jevđl=ni=inde*.  
**now** that-PL NEG.EXIST=already=already  
 'Now, they do not exist anymore.'

(30) Udmurt (Tatyshly district, Southern Peripheral)

*Al'i=no* *ug* *bert* *šusa* *kđll'-e=dđr=äl'i*.<sup>15</sup>  
**now**=ADD NEG.PRS return:CNG.SG COMP stay-PRS.3SG=probably=**yet**  
 '[He says that] even now, she is probably not coming back yet.'

Still, the simultaneous existence of a 'lexical' *al'i*, which has another placement pattern, influences the way the 'discourse' one can be placed inside a sentence. As has been demonstrated for *pe* 'reportative' (as well as Komi-Zyrian *nin* 'already'), second-position requirement or free placement within a sentence correlates with the clitic's preference for the rightmost slots in clitic templates.

<sup>15</sup> The first vowel in the 'discourse' *al'i* is subject to vowel harmony in Tatyshly Udmurt, which provides yet another way of telling the two apart.

## 5. Multiple particles with the same meanings

### 5.1. Clitic replacement through Jespersen's cycle

Let's shift our focus to individual particles now. Data from Udmurt dialects reveal a peculiar situation. Almost without exception, they all have approximately the same set of meanings expressed by discourse particles. However, the particles that express these identical meaning can be different (and of completely different origin) in different dialects.

Consider pairs of examples like (8)–(9) or (31)–(32), which feature translations of the same questionnaire sentences. In Tatyshly (Southern Peripheral), the standard Udmurt enimitive *uk* apparently does not exist, but the same meaning is expressed by *eś*, absent in most other dialects. Similarly, *aj/al'i* is absent in Beserman and the Northern dialects, while *na* only occurs very rarely in some Southern Peripheral varieties.

(31) Beserman (Shamardan)

*Ar ta-leś wal'lo ug=na vel't-δ val.*  
 year this-ABL earlier NEG.PRS=**yet** walk-NEG.PRS.3SG be:PST  
 'Last year, he could not walk yet.'

(32) Udmurt (Loloshur-Vozzhi, Southern)

*So kδl'-em ar-δn ug vetl-δ val=al'i.*  
 that leave-PTCP.PST year-LOC NEG.PRS go-CNG.SG be:PST=**yet**  
 'Last year, he could not walk yet.'

It happens rather often that the same lexical meaning is present everywhere, but expressed by distinct words in different dialects. For example, a lot of vegetables and fruits split the Udmurt-speaking area in two parts, where the Northern part has a native or a Russian name for it, e.g. *jablok* 'apple', and the Southern part has a Tatar one, e.g. *ulmo* 'id.' (Насибуллин, Максимов, Семенов, Отставнова 2009 : 216). There is nothing mysterious in such a situation: since apple trees are cultivated in every settlement where the Udmurts live, there will obviously be a word for 'apple' in each dialect.

Discourse particles are a very different matter though. Unlike apples, they do not refer to objects or concepts that the Udmurt speakers deal with for extralinguistic reasons anyway. A language or a dialect could do without an enimitive or a word with the polysemy of the Udmurt *na* or *ni* just fine. Many languages, like English, do not have grammaticalized enimitives. And while most languages do have temporal/aspectual items equivalent to English *yet*, *still*, *anymore* and *already*, they can group them differently. For example, most languages in the 41-language sample by Vandeweghe (2006 : 181–185) have separate items for *yet* and *still*, in contrast to Udmurt.<sup>16</sup> So if a certain discourse particle does not exist in a certain dialect, there is no reason for its meaning to be represented by another one or for its polysemy to be replicated elsewhere. Indeed, the enimitive apparently does not exist at all in the Bavly variety (Southern Peripheral). Yet in almost all other varieties all meanings listed in Table 1 can be expressed by one particle or another. How did that come to be?

The areal distribution of the pair *na* and *aj/al'i* provides a likely answer. While there are areas where the latter does not exist (most Northern dialects) or the

<sup>16</sup> Regrettably, the Udmurt row in Vandeweghe's table wrongly claims that Udmurt uses *evel* rather than *ni* for 'anymore'.

former is very infrequent and probably limited to a small set of contexts (parts of Southern Peripheral dialects), there is a large area where both are employed either interchangeably (33)–(34) or simultaneously (35). In the contexts where both *na* and *aj/al'i* are acceptable, the speakers claim there is no difference in meaning whatsoever. Their claim is supported by the data: in dialects where both clitics exist, each of them, as well as their combination, regularly appears in translations of same questionnaire sentences.

(33) Udmurt (Bagrash-Bigra, Central/Southern)

*Mon nulesk-i važ vu-i, otjn gubi-os gvel=na.*  
 I:NOM forest-ILL early come-PST.1SG there mushroom-PL NEG.EXIST=**yet**  
 'I came to the forest too early, there are no mushrooms there yet.'

(34) Udmurt (Bagrash-Bigra, Central/Southern)

*Mon nulesk-i važ vu-i, otjn gubi-os gvel=aj.*  
 I:NOM forest-ILL early come-PST.1SG there mushroom-PL NEG.EXIST=**yet**  
 'id.'

(35) Udmurt (Bagrash-Bigra, Central/Southern)

*Mon nulesk-i važ vu-i, otjn gubi-os gvel=na=aj.*  
 I:NOM forest-ILL early come-PST.1SG there mushroom-PL NEG.EXIST=**yet=yet**  
 'id.'

Since *aj/al'i* appeared in Udmurt later than *na* and, like other Tatar borrowings, gradually spread northwards, different latitudinal sections of the Udmurt dialectal continuum presumably represent different stages of the diachronic process whereby *na* is gradually replaced with *aj/al'i*. At the intermediate stages of this process, both clitics are conceptualized as full equivalents and are used simultaneously. In my data, respondents who predominantly used both clitics at once come exclusively from the Central dialectal zone (Sjumsi, Vavozh, Ludorvaj, Buranovo, Iljinskoe). Speakers from the North did not use *aj/al'i* at all. Speakers of the Southern and Southern Peripheral areas, on the other hand, tend to use the two synonyms interchangeably, but not simultaneously (and clearly prefer *aj/al'i* in Southern Peripheral varieties). Even if the newly borrowed particle used to be only a partial semantic match for the old one and originally did not cover all senses expressed by the latter, it probably aligned with it at the stage where both were used at once. As a result, the new particle becomes available in all contexts where the old one was used and eventually replaces it.

The process thus resembles the Jespersen's cycle whereby one construction is gradually replacing another without apparent change in meaning. It has to be noted however, that I am referring to the formal stages of a process that Jespersen (1917) illustrated with gradual diachronic replacement of negative elements, such as French *ne V > ne V pas > (V pas)*. The functional explanation Jespersen offered for the initial stage of the process was that the speakers find the original item, such as the French negative particle *ne*, (semantically) weakened and insufficient and therefore "strengthen" it by adding another element (in that case, the post-verbal element *pas*). While I find this explanation unconvincing even when applied to French (what does a "weakened and insufficient negation" even mean?), it definitely does not work for Udmurt. Udmurt speakers start doubling their discourse particles not out of necessity, but because these particles are available in other languages they speak, and discourse particles are very easily borrowed (Matras 2010 : 81).

The same kind of semantic alignment can be seen in the Russian L2 speech of the Udmurt and Beserman speakers. Certain Russian words, which are perceived as translation equivalents of native discourse particles, acquire new meanings and are used in contexts in which they are hardly ever used by monolingual Russian speakers. For example, Russian *уже* (*uže*) 'already', which is considered equivalent to Beserman *ni*, is almost as frequent in the corpus of L2 Russian produced by native Beserman speakers (15,061 ipm) as *ni* is in the Beserman corpus (17,043 ipm). This presents a stark contrast to standard Russian usage of *уже*: its frequency in the subcorpus of non-public spoken conversations of the Russian National Corpus (which contains texts of approximately the same type and register as the Beserman ones) is four times lower at 4175 ipm. This staggering discrepancy is in part due to usages like (36), where Russian *уже* is used by the Beserman speakers in contexts where standard Russian *уже* would be infelicitous, but Beserman *ni* is appropriate (cf. (37)).

(36) Beserman Russian (Shamardan)

*"Pärtose" vot etot nazjva-jet-sa=uže.*

p. here this.M.SG.NOM call-PRS.3SG-REFL=**already**

'That (a man moving in to his wife's house after the wedding) is called *pärtose* [in Beserman].'

(37) Beserman (Shamardan)

*A ta-iz teń šort=a mar=a šu-išk-e=uk=ni.*

and this-3SG.POSS here yarn=Q what=Q say-DETR-PRS.3SG=ENIM=**already**

'And I guess this thing (yarn) is called *short* [in Beserman].'

I am not able to explain why *ni/uže* is used in these examples, and its meaning here is clearly not temporal or aspectual. However, sentences that mean 'X is called Y' are one of the contexts where *ni* frequently occurs.

Although borrowed *uže* is not nearly as frequent in Beserman as *ni*, it co-occurs with *ni* in about half of all its occurrences in Beserman (38) (see also Kaysina 2014). Given that the equivalence relation has been established, this might prove to be an incipient stage of a shift from *ni* to *uže*.

(38) Beserman (Shamardan)

*Kal=uže šer lešt-o=ni.*

now=**already** rarely make-PRS.3PL=**already**

'These days, they only make [felt boots] rarely.'

There are parallel pairs of items in Udmurt and Komi, which are suspiciously similar in their meaning, and not cross-linguistically widespread, but clearly distinct from the etymological point of view. Such are, for example, the attenuative/comparative marker (*-ges/-gem* in Udmurt, Черемисинова 2019; *-žjk* in Komi, Todesk 2015) or the self-reportative clitic (*pej* in Udmurt and *miša* in Komi, Teptiuk 2021). In light of the situation with Udmurt discourse clitics, it would make sense to investigate if a similar process could have led to the parallel existence of such items.

## 5.2. Possible outcomes of Jespersen's cycle

One possible outcome of the Jespersen's style is the disappearance of the older element because it becomes redundant. There is, however, another option. If the two elements always occur next to each other (unlike French *ne* and *pas*,



which are separated by a verb), they can be lexicalized as a single item. For example, *na* and *aj* could stick together at the intermediate stage, producing a joint clitic *naj*.

Indeed, *na=aj* is a frequent combination in many varieties, and sometimes it shows signs of phonetic reduction, being pronounced as *naj*. Nevertheless, different preferred positions of *na* and *aj/al'i* in the negative verbal complex and their diametrically opposed slots in the clitic template ensure that they are frequently interrupted by other material, as in (39)–(41). This precludes them from collapsing into a single item.

- (39) Udmurt (Kadikovo, Southern)  
*Pojezd ez=na likt-i=aj=no, likt-o-z=djir=ini.*  
 train NEG.PST.3=**yet** come-CNG.SG=**yet**=ADD come-FUT-3SG=probably=already  
 'The train has not arrived yet, but it will probably come [soon].'
- (40) Udmurt (Ludorvaj, Central)  
*Dak mil'am van'=na=ug=aj.*  
 well we.EXCL:GEN EXIST=**yet**=ENIM=**yet**  
 'But we still have some [sacks of pig feed].'
- (41) Udmurt (Buranovo, Central)  
*So-os-len nilpi-zj ug vetl-i=na=djir=aj?*  
 that-PL-GEN child-3PL.POSS NEG.PRS walk-CNG.SG=**yet**=probably=**yet**  
 'Their child probably doesn't walk yet, does he?'

Otherwise, however, lexicalization of a clitic combination would be a realistic scenario. Consider *ajko*, which is one of the regional variants of *aj/al'i*. Unlike all other forms of this particle, *ajko* could not have been derived from the borrowed *al'i* through regular phonological processes. In all likelihood, it is historically a combination of *aj* 'yet/still' and the hortative Russian clitic *ko*.<sup>17</sup> Udmurt *aj/al'i* can also be used as a hortative clitic, attaching to imperatives (42), a function also available to its Tatar source, *äle*. When *ko* was borrowed into the Central dialects that had already borrowed *aj/al'i*, the two were apparently used at once in this context. Since there were no intervening elements in this case, this resulted in a single clitic *ajko*. The equivalence relation between *aj* and *ajko* spread to other senses as well, so that nowadays *ajko* can be used in the rest of the contexts available for *aj/al'i*, in varieties where it exists<sup>18</sup> (43)–(44). The original meaning of *ko* thus ceased to play any role.

- (42) Udmurt (Dubrovskij, Southern)  
*Tani pomidor-jos, salat-e jud=aj/ajko.*  
 here tomato-PL salad-ILL cut=**HORT**  
 'Here are some tomatoes, cut them for the salad.'
- (43) Udmurt (Sjumsi, Central)  
*Ta jž badžjm=ke=no=ini, no ez=na=ajko pija=na.*  
 this sheep big=if=ADD=already but NEG.PST.3=**yet**=**yet** give.birth:CNG.SG=**yet**  
 'Although this ewe is big already, it has not yet given birth.'

<sup>17</sup> Its standard Russian form is *ка* (*ko*); the geographical range of the form *ko* overlaps with the Udmurt-speaking areas (CPHT 1977 : 278).

<sup>18</sup> I have attested *ajko* in the varieties of 8 settlements, which can be roughly covered by a triangle between Sjumsi, Jakshur-Bodja and the South-Eastern angle of the Udmurt-speaking area inside the region of Udmurtia on the maps above.

(44) Udmurt (Staraja Monja, Southern)

*Ta pomidor-jos vož-eś=ajko, vu-em-ze*

this tomato-PL green-PL=**still** come-PTCP.PST-ACC.3SG.POSS

*mi śi-i-m=ni.*

we:EXCL.NOM eat-PST-1PL=**already**

'These tomatoes are still green, we have already eaten the ripe ones.'

## 6. Conclusion

Studying the dialectal distribution of Udmurt discourse particles and clitic clusters allows us to arrive at cross-linguistically relevant conclusions.

First, clitic placement in clusters can be determined by hierarchies of scope (*ik* 'emphatic identity', two flavors of *no* 'and; additive particle') and syntactic constraints (*pe* 'reportative'). Still, these factors are only relevant for those three Udmurt clitics, while the rest can be ordered by templates that can apparently be arbitrary. Semantics, influence of contact languages and the order of appearance of clitics in a language do not have any straightforward effect on their place inside a template.

Second, clitic ordering templates can be subject to variation within a single language. That variation is dialectal in nature: geographically close varieties tend to have same templates. Additionally, there is considerable free variation, both interpersonal and intrapersonal, at the borders between areas that have different templates. That means that the perception of clitic chains as always conforming to a rigid order is not entirely correct. What that also means is that when researching clitic placement in a language with a weak norm and live dialects, one should not expect the data collected from speakers of different varieties to instantiate a single underlying grammatical structure.

Finally, looking at combination of discourse particles provides a potential solution for the 'one meaning, different particles' riddle. The fact that different Udmurt dialects have completely unrelated particles that apparently have same non-trivial semantics could be the result of a gradual, Jespersen's cycle-style replacement. In the course of this process, the meanings of two items are aligned while they are used simultaneously.

The properties of Udmurt discourse clitics and dialectal distribution of clitic templates raises a number of questions. Are same particles always available in the same contexts across dialects? How does the preferred position of clause-level reportative marker *pe* differ across dialects and what can affect its placement? Can we model the diachronic processes that led to the observed distribution of templates, perhaps using diachronic data? Can we find additional evidence in the diachronic data that would shed light on how exactly Jespersen's cycle progresses in the case of discourse particles? Further research is needed to answer these questions.

**Acknowledgments.** This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project ID 428175960.

I am immensely grateful to Nikolai Anisimov (Estonian Literary Museum) for his support when working with the Udmurt community in Estonia.

The publication costs of this article were covered by the Estonian Academy of Sciences and the Estonian Research Council grant PUT 1290.

## Address

Timofey Arkhangelskiy  
Universität Hamburg  
timarkh@gmail.com

## Abbreviations

ABL — ablative, ACC — accusative case, ADD — additive particle, CNG — connegative converb, COMP — comparative, CVB — converb, DAT — dative, DEB — debitive mood, DELIB — deliberative future form, DETR — detransitive derivation, EL — elative, ENIM — enimitive particle, EVID — evidential past, EXCL — exclusive pronoun, EXIST — existential verb, FUT — future, GEN — genitive, HORT — hortative particle, ID — particle of emphatic identity, ILL — illative case, IMP — imperative, INDEF — indefinite marker, INF — infinitive, INS — instrumental, ITER — iterative derivation, LOC — locative, M — masculine gender, NEG — negation, NOM — nominative, PL — plural, POSS — possessive, PRS — present, PST — past, PTCP — participle, Q — question marker, REFL — reflexive, REP — reported speech marker, SG — singular.

L2 — second language, NP — noun phrase, PostP — postpositional phrase, V — verb.

**DeReWo 2014** — Korpusbasierte Wortformenliste DeReWo, DeReKo-2014-II-MainArchive-STT.100000. Institut für Deutsche Sprache, Programmbereich Korpuslinguistik, Mannheim, Deutschland. <https://www.ids-mannheim.de/>; **СПНГ** — Словарь русских народных говоров. Выпуск 12. Зубреха-Калумаги, Москва 1977.

## REFERENCES

- A g y a g á s i, Klára 2012, Language Contact in the Volga-Kama Area. — The Szeged Conference. Proceedings of the 15th International Conference on Turkish Linguistics Held on August 20–22, 2010, Szeged (Studia Uralo-Altica 49), 21–37.
- A n d e r s o n, Stephen R. 2005, Aspects of the Theory of Clitics, New York.
- A r k h a n g e l s k i y, Timofey 2023a [to appear], Etymology of the Udmurt Enimitive *uk* and Grammaticalization of Discourse Particles. — JSFOu 99.
- 2023b [to appear], Reported Speech and Its Extensions in Beserman. — Journal of Uralic Linguistics.
- A i j m e r, Karin 2013, Understanding Pragmatic Markers. A Variational Pragmatic Approach, Edinburgh. <https://doi.org/10.1515/9780748635511>.
- B a r t e n s, Raija 2000, Permiläisten kielten rakenne ja kehitys, Helsinki (MSFOu 238).
- B o n e t, Eulalia 1991, Morphology after Syntax: Pronominal Clitics in Romance, Cambridge, MA.
- B o r i s e, Lena, G e o r g i e v a, Ekaterina 2021, Acoustic Correlates of Stress in Udmurt Verbs: Variation and Theoretical Implications. <https://glowlinguistics.org/44/wp-content/uploads/sites/8/2021/04/slides-borise-georgieva.pdf>.
- B o š k o v i ć, Željko 2004, Clitic Placement in South Slavic. — Journal of Slavic Linguistics 12 (1/2), 37–90.
- 2016, Second-Position Clitics Cross-Linguistically. — Formal Studies in Slovenian Syntax. In Honor of Janez Orešnik, Amsterdam, 23–53.
- C a r d i n a l e t t i, Anna 1999, Pronouns in Germanic and Romance Languages: An Overview. — Clitics in the Languages of Europe, Berlin—New York (Eurotyp 20. Vol. 5), 63–82.
- 2015, Syntactic Effects of Cliticization. — Syntax — Theory and Analysis. An International Handbook. Volume 1, Berlin—München—Boston, 595–653. <https://doi.org/10.1515/9783110377408.595>.
- C s ú c s, Sándor 1990, Die tatarischen Lehnwörter des Wotjakischen, Budapest (Bibliotheca Uralica 10).
- C y s o u w, Michael, G o o d, Jeff 2013, Languoid, Doculect and Glossonym: Formalizing the Notion 'Language.' — Language Documentation and Conservation 7, 331–359.

- E d y g a r o v a, Svetlana 2014, The Varieties of the Modern Udmurt Language. — FUF 62, 376–398.
- G e o r g i e v a, Ekaterina, S a l z m a n n, Martin, W e i s s e r, Philip 2021, Negative Verb Clusters in Mari and Udmurt and Why They Require Postsyntactic Top-Down Word-Formation. — *Natural Language & Linguistic Theory* 39, 457–503. <https://doi.org/10.1007/s11049-020-09484-w>.
- H a s p e l m a t h, Martin 2015, Defining vs. Diagnosing Linguistic Categories: A Case Study of Clitic Phenomena. — *How Categorical are Categories?*, Boston—Berlin, 273–304. <https://doi.org/10.1515/9781614514510-009>.
- — 2022, Types of Clitics in the World’s Languages. <https://ling.auf.net/lingbuzz/007071>.
- H e a p, David 2005, Constraining Optimality. Clitic Sequences and Feature Geometry. — *Clitic and Affix Combinations. Theoretical Perspectives*, Amsterdam—Philadelphia, 81–102.
- H e g g i e, Lorie, O r d ó ñ e z, Francisco 2005, Clitic Ordering Phenomena. The Path to Generalizations. — *Clitic and Affix Combinations. Theoretical Perspectives*, Amsterdam—Philadelphia, 1–29.
- H e l l a n, Lars, P l a t z a c k, Christer 1999, Pronouns in Scandinavian Languages: An Overview. — *Clitics in the Languages of Europe*, Berlin—New York (Eurotyp 20. Vol. 5), 123–142. <https://doi.org/10.1515/9783110804010.123>.
- J e s p e r s e n, Otto 1917, Negation in English and Other Languages, Copenhagen.
- K a r t t u n e n, Lauri, P e t e r s, Stanley 1979, Conventional Implicature. — *Presupposition*, Leiden (Syntax and Semantics. Vol. 11), 1–56.
- K a y s i n a, Inna 2014, The Emergence of an Udmurt-Russian Mixed Code: Evidence from Discourse Markers. — *ESUKA* 5 (2), 9–27. <https://doi.org/10.12697/jeful.2014.5.2.01>.
- — 2015, Grammatical Effects of Russian-Udmurt Language Contact. — *Language Empires in Comparative Perspective*, Berlin, 219–236.
- K ö n i g, Ekkehard 1986, Identical Values in Conflicting Roles: The Use of German *ausgerechnet*, *eben*, *genau* and *gerade* as Focus Particles. — *Groninger Arbeiten zur germanistischen Linguistik* 27, 45–74.
- M a t r a s, Yaron 2010, Contact, Convergence, and Typology. — *The Handbook of Language Contact*, Oxford, 66–85.
- P a n o v, Vladimir 2020, The Marking of Uncontroverted Information in Europe: Presenting the Enimitive. — *Acta Linguistica Hafniensia* 52 (1), 1–44. <https://doi.org/10.1080/03740463.2020.1745618>.
- P e r l m u t t e r, David 1971, Deep and Surface Constraints in Generative Grammar, New York.
- S c h a c h t e r, Paul, S h o p e n, Timothy 2007, Parts-of-Speech Systems. — *Language Typology and Syntactic Description*, Cambridge, 1–60. <https://doi.org/10.1017/CBO9780511619427.001>.
- S c h e r r e r, Yves, S t o e c k l e, Philipp 2016, A Quantitative Approach to Swiss German — Dialectometric Analyses and Comparisons of Linguistic Levels. — *Dialectologia et Geolinguistica* 24 (1), 92–125. <https://doi.org/10.1515/dialect-2016-0006>.
- S c h n e i d e r, Klaus P. 2010, Variational Pragmatics. — *Variation and Change. Pragmatic Perspectives*, Amsterdam (Handbook of Pragmatics Highlights 6), 239–267. <https://doi.org/10.1075/hoph.6.18sch>.
- — 2021, Variational Pragmatics. — *The Cambridge Handbook of Sociopragmatics*, Cambridge, 663–686. <https://doi.org/10.1017/9781108954105.034>.
- S c h n e i d e r, Klaus P., B a r r o n, Anne 2008, Where Pragmatics and Dialectology Meet: Introducing Variational Pragmatics. — *Variational Pragmatics. A Focus on Regional Varieties in Pluricentric Languages*, Amsterdam (Pragmatics & Beyond New Series 178), 1–32. <https://doi.org/10.1075/pbns.178.02sch>.
- S h e e h a n, Michelle 2020, The Romance Person Case Constraint Is Not about Clitic Clusters. — *Dative Constructions in Romance and Beyond*, Berlin (Open Generative Syntax 7), 143–171. <https://doi.org/10.5281/ZENODO.3776543>.
- S p e n c e r, Andrew, L u í s, Ana R. 2012, Clitics: An Introduction, Cambridge.
- S p r o n c k, Stef, N i k i t i n a, Tatiana 2019, Reported Speech Forms a Dedicated Syntactic Domain. — *Linguistic Typology* 23 (1), 119–159. <https://doi.org/10.1515/lingty-2019-0005>.

- Teptiuk, Denys 2019, Quotative Indexes in Finno-Ugric (Komi, Udmurt, Hungarian, Finnish and Estonian), Tartu (Dissertationes Philologiae Uralicae Universitatis Tartuens 21). <http://hdl.handle.net/10062/66597>.
- 2021, Self-Quotative Markers in Permic and Hungarian. — LU LVII, 213—232. <https://doi.org/10.3176/lu.2021.3.04>.
- Todesk, Triin 2015, The Verbal Augmentative and the Inherent Properties of Verbs in Komi. — LU LI, 28—43.
- Vandeweghe, Willy 2006, Aspectual Particles in Some European Languages. — Pragmatic Markers in Contrast, Oxford (Studies in Pragmatics 2), 173—187.
- Vilkuna, Maria 1998, Word Order in European Uralic. — Constituent Order in the Languages of Europe, Berlin—New York, 173—234. <https://doi.org/10.1515/9783110812206.173>.
- Wichmann, Yrjö, Korhonen, Mikko, Uotila, Toivo Emil 1987, Wotjakischer Wortschatz, Helsinki (LSFU XXI).
- Zakirova, Aygul, Kozlov, Alexey 2022, The Emphatic Identity Particle =OK in the Volga-Kama Sprachbund. Presented at Congressus XIII Internationalis Fenno-Ugristarum, Vienna.
- Zimmerling, Anton, Kosta, Peter 2013, Slavic Clitics: A Typology. — STUF — Language Typology and Universals 66 (2). <https://doi.org/10.1524/stuf.2013.0009>.
- Zimmermann, Malte 2011, Discourse Particles. — Semantics, Berlin—Boston (Handbücher zur Sprach- und Kommunikationswissenschaft. Handbooks of Linguistics and Communication Science 33.2), 2012—2038. <https://doi.org/10.1515/9783110255072.2012>.
- Zwicky, Arnold M. 1977, On Clitics, Bloomington. <https://doi.org/10.5281/ZENODO.7436775>.
- 1985, Clitics and Particles. — Language 61 (2), 283—305. <https://doi.org/10.2307/414146>.
- Архангельский Т. А. 2021, Применение диалектометрического метода к классификации удмуртских диалектов. — Урало-алтайские исследования 2 (41), 7—20.
- Атаманов М. Г. 1981, Граховские говоры южноудмуртского наречия. — Материалы по удмуртской диалектологии. Образцы речи, Ижевск, 45—96.
- Закирова А. Н. 2019, Частицы эмфатической идентичности в поволжском языковом союзе. Магистерская диссертация, НИУ ВШЭ.
- Зубова Ю. Н. 2016, Синтаксис и семантика фокусных частиц в бесермянском диалекте удмуртского языка (на примере частиц *ik* и *uk*). — Acta Linguistica Petropolitana XII (1), 442—450.
- 2019, Вариации в использовании модальных частиц в литературном удмуртском языке и в бесермянском диалекте. — Урало-алтайские исследования, № 4 (35), 71—90.
- Карпова Л. Л. 2006, Особенности служебных частей речи среднечепецких говоров (в контексте северноудмуртских диалектов). — Диалекты и история пермских языков во взаимодействии с другими языками, Пермь (Пермистика XI), 104—111.
- Кельмаков В. К. 1981, Образцы удмуртской речи. Северное наречие и срединные говоры, Ижевск.
- 1998, Краткий курс удмуртской диалектологии. Введение. Фонетика. Морфология. Диалектные тексты. Библиография, Ижевск.
- Майтинская К. Е. Служебные слова в финно-угорских языках, Москва.
- Максимов, С. А. 1999, Северноудмуртско-коми ареальные лексико-семантические параллели. Диссертация на соискание ученой степени кандидата филологических наук, Ижевск.
- Насибуллин Р. Ш. 1981, Материалы по удмуртской диалектологии. Образцы речи, Ижевск.
- Насибуллин Р. Ш., Максимов С. А., Семенов В. Г., Отставнов Г. В. 2009, Диалектологический атлас удмуртского языка. Карты и комментарии. Выпуск I, Ижевск.
- Тараканов И. В. 1982, Заимствованная лексика в удмуртском языке (удмуртско-тюркские языковые контакты), Ижевск.

- — 1993, Удмуртско-тюркские языковые взаимосвязи (теория и словарь), Ижевск.
- Ф е д ю н е в а Г. В. 2008, Первичные местоимения в пермских языках, Екатеринбург.
- Ч е р е м и с и н о в а М. О. 2019, Компаративно-аттенуативная полисемия в бесермянском удмуртском. — Типология морфосинтаксических параметров. Том 2, вып. 2, Москва, 128—150.

*ТИМОФЕЙ АРХАНГЕЛЬСКИЙ* (Гамбург)

### ДИАЛЕКТНАЯ ВАРИАТИВНОСТЬ УДМУРТСКИХ ДИСКУРСИВНЫХ КЛИТИК

Удмуртский язык обладает множеством дискурсивных клитик. Нередко несколько клитик входят в одно фонетическое слово, образуя цепочку. В европейских языках цепочки клитик давно и успешно моделируются с помощью так называемых шаблонов. В этой статье я рассматриваю взаимный порядок клитик внутри цепочек в разных удмуртских диалектах. Я показываю, что в этой области диалекты демонстрируют существенные различия, а некоторые диалекты к тому же допускают свободное варьирование. Хотя цепочки клитик и можно, с некоторыми оговорками, моделировать с помощью шаблонов в каждом конкретном диалекте, для удмуртского языка в целом такой подход оказывается неприменим. Порядок клитик в цепочке в общем случае может быть произвольным. В частности, он не зависит от семантики частиц или порядка их появления в языке. Кроме того, я рассматриваю одно примечательное свойство удмуртского языка: существует ряд нетривиальных дискурсивных функций, которые доступны во всех или почти во всех его диалектах, но одна и та же функция может выражаться частицами совершенно разного происхождения в разных диалектах. Я объясняю это явление циклом Есперсена, в ходе которого одна частица постепенно заменяется другой, проходя через промежуточную стадию, когда обе частицы используются одновременно.

*TIMOFEI ARHANGELSKI* (Hamburg)

### UDMURDI DISKURSUSEKLIITIKUTE MURDELINE VARIEERUVUS

Udmurdi keeles on palju diskursusekliitikuid. Sageli on ühes foneetilises sõnas mitu kliitikut, mis moodustavad ahela. Euroopa keelte puhul on mallide abil kliitikute ahelaid edukalt modelleeritud. Oma artiklis käsitlen udmurdi murrete kliitikute järjestust nendes ahelates. Näitan, et murdeti on selles osas olulisi erinevusi ning mõnes murdes ka vaba varieerumist. Kuigi kliitikute ahelaid saab teatud reservatsioonidega modelleerida udmurdi murrete puhul, ei ole selline lähenemine rakendatav udmurdi keele kui terviku suhtes. Kliitikute järjestus võib ahelas olla enamasti meelevaldne. See ei sõltu partiklite semantikast ega nende päritolust. Lisaks käsitlen udmurdi keele tähelepanuväärset omadust, et teatud diskursusefunktsioonid väljendavad eri murretes täiesti erineva päritoluga partiklid. Seletan seda nähtust Jesperseni tsükliga, kus üks keelend asendub järk-järgult teisega, läbides vahepealse etapi, kui mõlemat kasutatakse samaaegselt.