UNIVERSITY OF TARTU
Faculty of Arts and Humanities
Institute of Estonian and General Linguistics
Department of Finno-Ugric Studies
Estonian and Finno-Ugric Languages

Przemysław Podleśny

A comparative study of clause-level constituent order in Erzya and Moksha

Master thesis

Supervisors: Petar Kehayov, PhD
Gerson Klumpp, PhD

Tartu 2024

Affirmation of authorship

I confirm that I have written this thesis myself and have correctly cited the contributions of other authors. The work was written based on the thesis requirements of the Institute of Estonian and General Linguistics of the University of Tartu and is in line with good academic practices.

Abstract

A comparative study of clause-level constituent order in Erzya and Moksha

The thesis describes a comparative investigation with regard to the order of constituents in literary varieties of two languages, Erzya and Moksha, which together constitute the Mordvin branch of the Uralic language family. The research was based on a small, manually annotated corpus of four works of narrative prose penned by two renowned authors, Kuz'ma Abramov and Piotr Levchaev, each contributing two texts to the set. The writers were chosen on the principle of comparability to allow for tentative conclusions concerning differences in the use of the two literary Mordvin languages in the early 2nd half of the 20th century.

The results of the research suggest the existence of differences between Erzya and Moksha in their distribution of the main constituents of transitive clauses and in structuring of clauses with infinitive complements. In general, orderings typical for OV languages seem to be slightly more frequent in Levchaev's Moksha than Abramov's Erzya, although further research will be necessary to confirm most of these tendencies.

Table of Contents

Glosses, abbreviations and acronyms **6**

1 Introduction **9**
1.1 Research goals . . . 9
1.2 The structure of the thesis . . . 10
1.3 Final notes . . . 12
1.4 Acknowledgements . . . 12

2 The Mordvin languages **13**
2.1 Genetic and areal connections . . . 13
2.2 Morphological categories relevant in the study . . . 15
2.2.1 Verbal complex . . . 15
2.2.2 Nominal complex . . . 18

3 Clause-level constituent order in Uralic and Mordvin languages **22**
3.1 Word order in Uralic languages . . . 22
3.2 Word order in Mordvin . . . 25
3.2.1 Statistical data on order of main constituents in clauses . . . 29
3.2.2 Auxiliaries and modifying verbs . . . 33
3.2.3 Interrogative clauses . . . 34
3.2.4 Non-verbal clauses . . . 35

4 Texts used in the study **38**
4.1 Kuz'ma Abramov . . . 40
4.1.1 Эрьке чиресэ кудыне (ECK) . . . 41
4.1.2 Тона пельде (TP) . . . 41
4.2 Pyotr Levchaev . . . 42
4.2.1 Пичетне пичедихть (PP) . . . 43
4.2.2 Кафонц кудат (KK) . . . 43
4.3 Conclusion . . . 44

5 Preparation of the corpus **45**
5.1 Preparation of files for annotation . . . 45
5.1.1 Digitisation of the texts . . . 45
5.1.2 Preparation of the files for annotation in a spreadsheet editor 50
5.2 Annotation . . . 50
5.3 Clause types . . . 51
5.4 Classification of constituents . . . 58
5.5 Morphological annotation . . . 65
5.6 Setting up a searchable corpus . . . 67

5.7 Conclusion 69

6 The corpus study **72**
6.1 Transitive clauses 73
6.2 Dependent clauses 76
6.3 Interrogative clauses 80
6.4 Verbal conjugation paradigm and constituent order 83
6.5 Object type, definiteness and constituent order 88
6.6 Polarity and constituent order 91
6.7 Position of adverbial in finite verbal clauses 93
6.8 Constituent order in clauses with infinitive complements 94
6.9 Order of constituents in nominal predication 98

7 Conclusion **104**

8 Kokkuvõte **106**

Licence **113**

Glosses, abbreviations and acronyms

Morphological glosses:

1/2/3	first/second/third person
PL	plural
SG	singular
ABL	ablative
ADV	adverb
ALL	allative (or dative-allative)
APRT	active (present) participle
CAUS	causative (case/postposition or verbal derivative)
CMP	comparative (case)
CNG	connegative
COM	comitative
COMP	comparative (degree)
COND	conditional mood
CONJ	conjunctive mood
CVB	converb
DAT	dative
DEF	definite
DEM	demonstrative
DES	desiderative mood
DIM	diminutive
ELA	elative
EMPH	emphatic
EX	existential
GEN	genitive
IMP	imperative
INE	inessive
INF	infinitive
INF1	1st infinitive (-мс)
INF2	2nd infinitive (-ма or -мо/-ме)
INF3	3rd infinitive (-мда or -мадо/м(о)до/м(е)де)
FREQ	frequentative
LAT	lative
NEG	negation
NMLZ	nominalization
NOM	nominative

OPT	optative mood
ORD	ordinal
PASS	passive
POSS	possessive pronoun
PROL	prolative
PRS	present (non-past) tense
PST	past tense (bound copula only)
PST1	first (simple) past tense
PTCP	passive participle
PX	possessive declension
QUOT	quotation marker
REFL	reflexive
TRA	translative case
VOC	vocative

Syntactic tags:

A	adverbial of means or degree
ARG_V	complement infinitive
CAUSE	cause
G	goal
INS	instrument
LOC	place
NPR	non-verbal predicate
O	direct object
O'	indirect object
POSS	possessor
PURP	purpose
Q	question word
REL	subordinator
S	subject
SRC	source
T	time
V	verb

Clause types:

DEP	dependent
DIAL	dialogue
EX	existential
IMPERS	impersonal
Q	content interrogative
EMBED_Q	embedded interrogative
INF	infinitive as predicate
MEL	desiderative with мель/мяль
MOZNA	modal possibility with можна
REP	reporting after dialogue

Texts:

ECK	Эрьке чиресе кудыне
KK	Кафонц кудат
PP	Пичетне пичедыхть
TP	Тона пельде

Languages:

E	Erzya
M	Moksha

1 Introduction

The Uralic language family boasts a relatively long history of research, spanning roughly two hundred years of intensive study. There is a long tradition of historical Uralic linguistics and a good number of grammars available. However, one's attention should not be distracted from the fact that much material is a product of its time and from today's point of view could be considered outdated or incomplete. The research tradition is uneven, most attention having been devoted to the large languages of independent Uralic nations: Finnish, Hungarian, Estonian, as well as to diachronic issues, phonology and morphology of the researched languages (see Klumpp et al. 2018: 10-11). As a result, there are still numerous underresearched areas in Uralic studies, particularly in the less prominent languages. Such seems to be the case of constituent order, especially outside the "great three" mentioned above (Vilkuna 2022: 950). The treatment of the issue in most grammars seems to boil down to reporting the preferred, most unmarked ordering, with little treatment of marked orders and factors influencing the ordering. The descriptions tend also to be qualitative and impressionistic.

The Mordvin branch of the Uralic family, consisting of the languages Erzya and Moskha, is one of those afflicted by the aforementioned issue. While some quantitative data does exist (Vilkuna 1998: 186; Kholodilova and Toldova 2018: 610-11), no comparative research of the two languages has been conducted so far in this area — this could also be interpreted as a part of the larger tendency to treat the two languages as interchangeable, leaving branch-internal differences unexplored. Much data is also lacking with regard to different genres of texts, time periods, authors, dialects, etc. (Vilkuna 2022: 960), and the present work is meant to contribute to the mending of this gap.

1.1 Research goals

The aim of this thesis is to investigate clause-level constituent order and its possible correlates in the literary varieties of Erzya and Moksha, the two languages constituting the Mordvin branch of the Uralic family. The study presented here is based on a small, personally annotated corpus of some works by Kuz'ma Abramov and Piotr Levchaev originating from the approximate period of 1960s-1980s and selected as a representation of their language use as manifested in their writing style.

The contribution of this work is meant to be twofold.

On the one hand, it is a study of a small, highly specialised sample. It provides insight into the use of literary language by two well-known authors with a more or less clear dialectal and educational background in a specific genre of writing and time period. Further research can expand on this foundation through addition of

other authors, genres, etc. into the mix.

On the other hand, it might be one of the first, if not the first comparative study of word order within the Mordvin branch. The nature of this work is investigative and its virtue lies not in providing final answers but in presenting new areas of research and leading to questions, inviting comparative exploration of the two Mordvin languages in their literary and oral varieties alike. The main focus will be on finding possible factors and variables affecting word order in Moksha and Erzya, with a particular interest in those whose influence differs between the two languages.

The research questions in this study could be formulated as follows:

1. How are different orderings of clause-level constituents in Moksha and Erzya represented statistically?
2. Which factors may correlate with different patterns of constituent order in Mordvin clauses?
3. In what ways do the two Mordvin literary languages (as used by Kuz'ma Abramov and Piotr Levchaev) differ from one another in terms of the first two questions?

The results presented here are based on a manually annotated corpus consisting of four pieces of literary prose by the aforementioned authors, amounting to a total of 19 539 tokens (8773 Erzya, 10 766 Moksha) with roughly 4500 clauses, of which 3329 (1623 Erzya, 1706 Moksha) contained more than one clause-level constituent (including subjects, verbs, copulas, objects, but also adjuncts, modal particles, as well as negation particles and verbs).

The idea behind the corpus was for it to be only large enough that certain regularities may be registered and compared, and to be somewhat representative for its content as a small subset of Erzya and Moksha. If contrasted with other studies and datasets, it could allow for a fuller picture of Erzya and Moksha as a whole. Although a full and balanced comparison between Erzya and Moksha is not the purpose here, I did not abstain from referring to extant corpora in certain cases where I wanted to check whether an interesting pattern discovered in my data could support a broader generalisation — the relative order of substantive and non-finite constituents in modal verb clauses is a case in point.

1.2 The structure of the thesis

The thesis consists of this introduction and six chapters following it. The first two serve as a theoretical background; the remaining part is a description of the dataset, the proceedings and the outcomes of the study.

In the first chapter, I provide a brief summary of the morphological and morphosyntactic features of Mordvin languages which play a role in the present

study. Several of these features are tested directly for their correlation with clause-level constituent order in Moksha and Erzya, the rest are alluded to in less detail and/or referenced during the discussion of methods used during annotation or patterns found in the data without being in the centre of the research's focus. To the first group belong features such as definiteness marking on substantives and object indexing on verbs; the full variety of possibilities for case marking on direct object NPs is one of those belonging to the latter group. The main focus is on function and on interactions between different aspects of morphology and syntax, as well as on their relevance in this work.

The subject of the second chapter is constituent order in Mordvin, with other Uralic languages serving as a background. The main point of this part is a summary of the available data regarding word order in Erzya and Moksha (on various levels) against which the outcomes of the present study are compared, as well as presentation of this topic in a broader comparative light. The comparative part largely draws on Maria Vilkuna's summary of word order in Uralic languages in Siewierska 1998 and Bakró-Nagy et al. 2022.

The third chapter is a description of the dataset, providing meta-information and summaries which might be important during assessment and classification of the outcomes of the study conducted here, whether in this thesis or in future research. Its contents include a short presentation of the authors whose works were included in the present corpus, synopses along with a short characterisation of the latter as well as a description of the criteria which conditioned this specific choice of texts. The corpus being relatively small and concentrated (not only in the terms of its absolute size but also the narrowness of the choice of its source material), addition of such a chapter was deemed necessary so that an explanation for any possible oddities could be easier to find, even if access to the original texts themselves (which have not yet entered the public domain) is not possible for the reader. Summarising the information in this way enables one also to better understand my own way of looking at the material and the line of reasoning used at certain points of the analysis performed in the later part of the thesis.

The following chapter describes the process through which the corpus used in the study was prepared. Its point is to ensure reproducibility of the present research through a description of the necessary details of the process of data preparation as well as detailed specification of the categories used during annotation, particularly where the categorisation might be less obvious. Examples are provided to illustrate especially the more borderline, less obvious situations. This is also where the outcomes of the choices which had to be made in such cases, some of them indeed arbitrary, are described and, where possible, justified.

Finally, the fifth chapter presents the statistical data produced from the corpus whose way of preparation was described in the previous chapter. The data are

shown from multiple angles, trying to account for the plurality of variables involved in the process of ordering of the constituents. The large amount and diversity of the presented results is an attempt to mitigate the fact that the corpus itself cannot be shared for copyright reasons by providing an extensive overview of the information contained within it. Each set of data is accompanied by discussion, where I bring the most significant findings to the reader's attention and attempt to identify significant or possibly significant tendencies or account for the presence of outliers.

The thesis ends with a summary, presenting the work in a holistic light, recounting the most important outcomes of the study, both in terms of the results themselves and the issues encountered on the way to obtain them.

1.3 Final notes

Throughout this work, "word order" and "constituent order" are used as synonyms (Dryer 2007: 61). The typology of non-finite clauses employed in this work follows Haspelmath (2022). Examples used in this work were glossed according to the Leipzig glossing rules. Examples produced from texts used in the study are signed with an abbreviation representing the text they originate from, while examples from other sources were cited accordingly. All examples are cited verbatim using the original orthography of their source, with two exceptions:

1. Square brackets were added in certain examples to provide context, e.g. where a cited example is a dependent clause and would seem awkward otherwise; the part within the brackets is not glossed but it is translated and likewise indicated by brackets in the translation line.
2. Omissions from the original texts occur at certain points; these are indicated by ellipsis (...). (Original ellipses have been kept intact).

1.4 Acknowledgements

Ending this introduction, I would like to express my gratitude to my supervisors for their guidance, starting with the choice of an appropriate research topic and ending with numerous comments and suggestions for the improvement of this very work.

Additionally, I would like to thank Anna Venchakova and Natalia Abrosimova for their help with analysis and interpretation of some less transparent structures in Moksha and Erzya respectively, which allowed me to better understand the information structure in these languages.

If, however, any errors are still present in this work, I am the only person to be blamed for them.

2 The Mordvin languages

This chapter provides a brief description of Moksha and Erzya, with a particular focus on morphology and those features of their syntax and morphosyntax which will be relevant in the study. These include a brief overview of declension and conjugation paradigms (including ways of expressing nominal definiteness), argument marking and indexing strategies in transitive clauses and strategies used in nominal predication. Non-finite clauses will be touched upon, to a degree necessary in order to understand what distinguishes them from finite clauses. A discussion of known features of constituent order in Mordvin is deferred to the following chapter.

The differences between the two languages, besides certain items of vocabulary, are found mostly in the exact surface-level manifestation of particular features, the particular set of available forms, degree of redundancy within the paradigm etc., so in this chapter, as well as in this work, they will be treated collectively, as long as contrasts do not arise that would warrant a more detailed discussion. For more complete descriptions of the languages, the reader is advised to refer to works cited as sources in this chapter, especially (Bartens 1999), (Hamari and Ajanki 2022) and (Rueter 2023).

2.1 Genetic and areal connections

Erzya and Moksha are two closely related languages, together forming the Mordvin group (branch) within the Uralic language family (Luobbal Sámmol Sámmol Ánte 2022; Saarikivi 2022). They are both (alongside Russian) official languages of the Republic of Mordovia within the Russian Federation, but, at least traditionally, widely spoken also beyond its boundaries (see e.g. Rueter 2013 for a socio-geographical overview). The number of their users cannot be assessed precisely due to methodological issues associated with national censuses in Russia — the tendency to classify Moksha and Erzya both as "Mordvin" without further subdivisions, the ambiguity of questions pertaining to language use and, not unlikely, general dishonesty surrounding the process of data collection. According to the last such census, conducted in 2020, there were altogether 19 911 speakers of "Moksha-Mordvin", 40 045 speakers of "Erzya-Mordvin" and 220 090 speakers of simply "Mordvin" in all of the Russian Federation, a bit more than half of this number in Mordovia itself (Rosstat 2022). Traditionally, based on older data, it has been assumed that there are around twice as many speakers of Erzya as of Moksha (Hamari and Ajanki 2022), giving us a rough estimate of around 185 000 and 90 000 speakers respectively.

Traditionally, Erzya and Moksha used to be grouped together with Mari languages within the so-called "Volgaic" branch, mostly due to their geographic proximity and the presence of a small group of unique lexical items, although the

common innovations are few and limited to a shared portion of vocabulary, and the structural differences are striking (Saarinen 1991). More recently, Nichols (2021) and Grünthal et al. (2022a) (see also Supplement 1 to the latter, Grünthal et al. 2022b) describe a grouping of Mordvin languages together with Finnic (containing Finnish, Estonian, Votic, etc.) and Sámi into "Western Uralic" as quite "clearly genealogical", yet with a rather unclear internal structure. In the flat, "rake"- or "bush"-like model of the Uralic genealogic tree, which has gained popularity in the recent decades, Mordvin remains the closest neighbour of Finnic in a continuum of sorts, generally spanning from the Norwegian Sea in the west to the Yenisei as its eastern border (Luobbal Sámmol Sámmol Ánte 2022: 4).

The literary languages of Erzya and Moksha are, despite some early tradition going over a century back, by and large an achievement of the 20th century, particularly notable is their functional expansion during the interwar period. Despite an ongoing discussion around creating a unified Mordvin standard language, the two traditions managed to establish themselves firmly before any such attempt could come into fruition (Bartens 1999: 22-23; Rueter 2013: 6-11; 2023: 483-4). They represent two language systems, discernible in all continuous written "Mordvin" texts and easily distinguished by their speakers, even if not completely uniform. Dialect variation exists in both languages, so far described mostly in terms of phonological and morphological isoglosses (Rueter 2010: 8, cf. Rueter 2016). A transitional variety used by the ethnic group of Shoksha is sometimes distinguished besides the two but does not have its own literary standard and tends to be classified as a Moskha-influenced dialect of Erzya (Rueter 2016: 111-3; Hamari and Ajanki 2022: 392).

While considered a peripheral member of the so-called Volga-Kama sprachbund — together with the related Mari and Udmurt, and unrelated Turkic languages such as Chuvash, Tatar and Bashkir (Helimski 2003: 159-160) — Mordvin languages differ from the others in this group (other than the other peripheral member, Komi). Their dominant word order is an outlier in the group (VO against more or less flexible OV — cf. Siewierska 1998), a less pronounced influence of Turkic languages visible mostly in vocabulary (more so in Moksha than in Erzya) but not syntax, their tendency towards synthetic rather than analytic constructions, lack of a special evidential past tense, etc. (Saarinen 1991). Nowadays, Russian exerts the greatest influence on Erzya and Moksha, a circumstance shared with numerous other languages, including all members of the aforementioned Volga-Kama sprachbund and most members of the Uralic language family; this is considered one of the factors contributing to a recent shift in preferred word order (cf. Zaicz 1998: 206; Kholodilova and Toldova 2018: 11-12; for an argument about a similar phenomenon ongoing in Udmurt see e.g. Asztalos 2021).

2.2 Morphological categories relevant in the study

Mordvin languages are characterised by their extensive morphology. In the following sections features marked in Mordvin morphology and relevant to syntax, accordingly, important to this work, will be presented, with an emphasis on their function rather than form.

2.2.1 Verbal complex

Mordvin verbs are conjugated for tense, mood, subject person and number, and optionally object person and number. Marking subject and object at the same time requires use of portmanteau suffixes, collectively referred to as objective, or definite, conjugation. Usage of this conjugation is to a degree optional and dependent on the semantics of the verb, coding characteristics of a pronominal object, denoting definiteness of a nominal object or totality of the action connected with it, often resulting in perfective readings. The property of denoting definiteness leads to correlation and interplay with the nominal definite marking. (Hamari and Ajanki 2022: 408)

Extensive subject and object marking allows for easy dropping of both subject and object personal pronouns (Bartens 1999: 112). This has the somewhat unfortunate result of a low number of available clauses with both subject and object appearing overtly; although due to syncreticisms in the objective conjugation paradigm, in some instances additional data (context, an overt pronoun) may be necessary for disambiguation.

(1) Moksha (Rueter 2023: 507)

a. *kel'k-t'ä*
love-1SG>2SG
'I love you (sg.)'

b. *kel'g-samaź*
love-3PL>1
'they love me/us'

(2) Erzya (Bartens 1999: 133)

but'i saj-iks-at — saje-mak!
COND take-DES-2SG take-IMP.2SG>1SG
'if you want to take — take me!'

Among the Uralic languages, Erzya and Moksha belong to those richest in moods, with an unmarked indicative and synthetically marked imperative, conjunctive, conditional, conjunctive-conditional, desiderative and optative (Kehayov 2020; Hamari and Ajanki 2022: 411). Several of these moods exhibit very specific

patterns of usage which might interact with the order of constituents: the (synthetic) imperative is limited to 2nd person and excludes, or almost excludes, presence of the subject, while conditional and conjunctive-conditional verbs (as will be seen later) appear exclusively in dependent (conditional) clauses, mostly (although not exclusively) clause-initially (3). The remaining moods show a more or less typical behaviour and in this study all moods other than imperative are treated equally with the indicative. There is also an analytic 3rd person imperative/optative formed with the particle катк (4), which in this study was not discriminated against, either. (Hamari and Ajanki 2022: 412; Rueter 2023: 509)

(3) Moksha (Kehayov 2020)

Atu mon ušədə-ńd'äŕ-än ava-t'ńə-ń eŕaf-snə-n
or.else 1SG begin-COND-1SG woman-PL-GEN life-PX.3PL-GEN
azəŋkšńə-ma, sat-i kafta ši-t.
chatter-INF2, be.enough-3SG 2 day-PL

'Otherwise, if I begin chattering about the lives of the women, it will take two days.'

(4) Moksha (Rueter 2023: 509)

... katk ńi ańćək miń soda-saśk t'ä kal'akənda-ma-t'
OPT already only 1PL know-1PL>3 DEM babble-NMLZ-GEN.DEF

'Let it just be us who know this conversation'

Phasal and modal auxiliaries take other verbs as arguments. The verb expressing the modal meaning is conjugated and the semantic verb appears as an infinitive; the type of the infinitive is selected by the modal verb and may depend on its semantics, e.g. the Moksha verb ёрамс 'to want', typically taking the -мс infinitive as an argument (5a), takes the ablative infinitive -мда (5b) when indicating a "progressive focused activity" (Rueter 2023: 511).

(5) a. Moksha (Rueter 2023: 522)
a mon jor-an l'ijəńd'-əms
but 1SG want-1SG fly.around-INF1
'but I want to fly around.'

b. Moksha (Rueter 2023: 511)
t'äńi jor-an t'ijə-mda oću saməl'ot, ańćək kodaŋga af
now want-1SG do-INF3 big airplane only somehow NEG
l'iśəńd'-i
work.out-3SG
'Now I'm trying to make a big airplane but it just isn't working out.'

The underlying argument structure of a clause can be preserved but in certain cases the initial object is promoted to an apparent subject, as evidenced by agreement. The verb эрявомс 'to be needed' is such a raising verb in both Erzya and Moksha: example (6a) shows that the verb does not agree in number with the plural agent of ютамс, while example (6b) demonstrates that this lack of agreement is not due to an impersonal nature of the verb but rather to its agreement with the patient of the infinitive.

(6) a. Erzya (TP)
Эряв-сь юта-мс ансяк Угра лее-сь.
need-PST.3SG cross-INF1 only Ugra river-DEF
'[They] only needed to cross the Ugra river.'

b. Erzya (TP)
Эряв-ить муе-мс лушмо-сто-нть след-эст,
need-3PL find-INF1 gully-ELA-DEF trace-PX.3PL
'[We] need to find their traces in the gully.'

Infinitives can also stand on their own in a predicative function, expressing e.g. necessity or possibility. (Hamari and Ajanki 2022: 421)

(7) Moksha (Hamari and Ajanki 2022: 421 with a correction: *езда* → *эзда*)

Мезь-са тага лезд-омс стир-ть, кода ванфто-мс
what-INE again help-INF1 girl-GEN.DEF how protect-INF1
кулома-ть эзда...
death-GEN.DEF ABL

'By what, again, can the girl be helped, how to protect her from death?'

Standard negation in Mordvinic is expressed by means of a negation particle (E) а / (M) аф or a negative auxiliary verb (E) эзь / (M) изь, ашезь with a suppletive paradigm. The usage is conditioned by the tense/mood of the original verb, most importantly, the verb is used in past tenses and the particle is used in the non-past tense. With the negation particle verbs appear unchanged; the negation verb, however, takes the agreement and TAM markings of the original verb, which in turn appears in a form called connegative (Bartens 1999: 141-4; Hamari and Ajanki 2022: 410). Although in e.g. Finnish the negation verb and the semantic verb in connegative are independent to a degree, in Mordvin they are hardly separable. The negation almost always precedes the verb it refers to immediately (no exceptions to this rule exist in the present corpus, although a few interesting constellations can be found with contrastive focus (8a) or ellipsis of the negated verb (8b)).

(8) a. Moksha (KK1)
...сяс и мярьг-ихть аф Самсон (...), а — Самоша-Горюша.
hence and call-3PL NEG Samson but Samosha-Goriusha

'...that's, too, why they call him not Samson, but rather — Samosha-Goriusha.'

b. Moksha (KK2)

Фёдор Ефимыч макс-сы, а тон аф?
Fiodor Efimovich give-3SG>3SG but 2SG NEG

'Fiodor Efimovich gives her, and you do not?'

Different kinds of negators are used as a negative copula ((E) аволь / (M) афоль) and existential copula ((E) арась / (M) аш, аяш) (Bartens 1999: 144; Hamari and Ajanki 2022: 410).

(9) (Bartens 1999: 144)

a. Moksha and Erzya
son af stud'ent
son avol' stud'ent
3SG NEG student
'he is not a student'

b. Erzya
kudo-m araś
house-PX.1SG NEG
'I don't have a house'

c. Moksha
kud uštəms ajaš toń pengə-ńä-t'ńä
house warm NEG 2SG.GEN firewood-DIM-PL.PX.2SG
'you don't have wood to kindle the house'

2.2.2 Nominal complex

The most agreed upon analyses describe Erzya and Moksha as having 12 and 13 cases, the difference being a relatively recently grammaticalised causative case, absent in Erzya. On top of this system is overlaid that of definite and possessive marking, usually described as a system of three declensions: indefinite, definite and possessive, with all of the cases in each language present and distinguished only in the singular forms of indefinite declension and a degree of ambiguity and deficiency elsewhere. A comparable version of the paradigm, although with certain differences, can be found also in pronouns (Hamari and Ajanki 2022: 399-406). The suffixal

complex is further extended by the existence of enclitic copula in opposition to the free copula, traditionally described as nominal conjugation (Rueter 2023: 482).

Four cases of the Mordvin system are particularly important in main participant marking: **nominative**, **genitive** (in certain contexts referred to as genitive-accusative, cf. Bartens 1999: 92, though this convention will not be followed here), **inessive** (almost always in the form of a postpositional construction, Hamari and Ajanki 2022: 427) and **ablative** (also called partitive based on its main types of usage, see Rueter 2023: 493). Below follows a short summary of their functions in this regard.

Nominative is the prototypical subject case in intransitive and transitive sentences. However, an indefinite object of a transitive verb can also appear in the nominative; such an object is considered unmarked. This is opposed to **genitive** marking, used for definite, possessed or otherwise total or individualised, also pronominal objects of finite verbs. Similar split object marking strategies are found in all languages of the Volga-Kama sprachbund, and well beyond it (see e.g Havas 2008: 2008 for a discussion of the phenomenon in Uralic). Vilkuna (1998: 187-8) notes based on Baker (1985: 36, 38-9) that the strategy appears to interact with the word order at least in Permic, suggesting the presence of an incorporational strategy; compare (section on position vs. case) in this work. Genitive is additionally used for indefinite objects of non-finite verbs. Note that in certain personal categories of the Erzyan possessive declension nominative and genitive are usually not distinguished (Hamari and Ajanki 2022: 426-7).

Inessive marking (of the definite declension) is used in the written language almost exclusively in the form of postpositional constructions, though in speech generally also on substantives (note that the latter is only possible in the declension paradigm of Erzya, Moksha being limited to the former alternative). Such object marking indicates an irresultative or imperfective action (loosely adapted from Alhoniemi 1991: 27). A pronominal possessor in such a construction is usually dropped, leaving a form resembling of a specialised personal pronoun:

(10) Moksha (Hamari and Ajanki 2022: 427)

...кивок веле-ста вешенд-и эсо-т.
somebody village-ELA look.for-3SG INE-2SG

'...someone from the village is looking for you.'

Certain verbs of fearing or otherwise negative semantics select objects in **ablative**. With verbs of eating and drinking, ablative object marking is used to express partial quantity. Ablative marking can also appear in partial subjects (11). In Moksha, ablative is used in this latter function more extensively, with various quantifiers. There, the case can be expressed synthetically or, in the definite declension, analytically, with the postposition эзда (Rueter 2023: 493-4).

(11) (Bartens 1999: 94)

a. Moksha
tumə̑-da kolmə̑=l'
oak-ABL three=PST.3SG
'there were **three** oaks'

b. Erzya
tumo-t'ńe-d'e kolmo=l'
oak-PL-ABL three=PST.3SG
'there were three (of the) oaks/the oaks were three'

The information provided above is be summarised in the following table:

Case	Subject	Object
Nominative	+	indefinite only
Genitive	—	+
Inessive	—	+ (postpositional phrase)
Ablative	+ (with additional quantifiers)	+ (with certain verbs and with quantifiers)

In addition to the above, **dative** or (in definite conjugation) **dative-allative** designates an indirect object (receiver).

Generally, subject marking strategies present in verbal clauses are also used in nominal predication. There, however, the translative case can be used in duonominal constructions (Haspelmath 2022) instead of nominative to mark predicate (the behaviour of this kind of marking in Erzya was analysed e.g. by Ajanki (2011)).

(12) Erzya (Ajanki 2011: 398)

Ošo-ńt' ńevt'-ića jalga-ks uĺńe-ś eŕźa-ń
city-GEN.DEF show-PTCP.PRS friend-TRA be-PST1.3SG Erzya-GEN
sodaviks poet.
famous poet

'A famous Erzya poet was used as a guide to show the town.'

Definiteness is another feature possibly relevant for order of constituents, not negligible is its correlation with the case usage in object marking (see above), but it can also play a role in the differentiation of components in duonominal constructions. As indicated in the introduction to this section, this property can be expressed overtly in Mordvin by means of definite declension (nouns marked for possession are also understood as definite; note that possession can also be expressed with a genitive construction without possessor marking on the object, as in 13). Such marking denotes a specific referent, a referent identifiable for the

listener, etc. The associated definiteness can be based on contextual reference or personal/encyclopedic knowledge. Proper names generally do not receive definite marking in Erzya, while they are typically marked for definiteness in Moksha (e.g. not marking 'Moscow' for definiteness is ungrammatical in 14). The usage of definite marking can be limited by deficiency within the paradigm. Unlike in other Uralic languages with definite conjugation, such as Hungarian, conjugation in Mordvin is not directly tied to the definiteness marking of objects (e.g. indefinite, definite and possessive object marking are all possible with the non-definite verb form морась in example 15), but there is a certain degree of correlation, and it is required in the presence of certain other elements, such as superlative markers and universal quantifiers. (Bernhardt 2021: 27-8, 37-9)

(13) Erzya (Bartens 1999: 105)
t'e šabra-n't' t'ejt'eŕ-eś
DEM neighbour-GEN.DEF girl-NOM.DEF
'this is the neighbour's daughter'

(14) Moksha (Kholodilova and Toldova 2018: 128)
*(*mosku) / mosku-s' oc'u oš*
(Moscow) Moscow-NOM.DEF big city
'Moscow is a large city.'

(15) Erzya (Aasmäe 2015: 76)
Сон састынестэ мора-сь моро/моро-нтъ/моро-нзо.
3SG quietly sing-PST1.3SG song/song-GEN.DEF/song-GEN.PX.3SG
'(S)he quietly sang a/the/his(her) song.'

In addition, there exists the bound copula, in some sources called nominal conjugation (Rueter 2023: 482), inflected for person, number and tense (present or past), that is, with a limited paradigm compared to that of the copula verb улемс (see below). This copula can attach not only to nouns, but also to adjectives and postpositions (16). Its presence in a clause is not even limited to the head of the predicate (see the section on non-verbal predication in the next chapter).

(16) Moksha (Kholodilova 2016: 105)
mon lavka-sə=l'-əń i kino-sə=ngə=l'-əń
1SG shop-INE=PST-1SG and cinema-INE=too=PST-1SG
'I've been in the shop, and in the cinema, too.'

Besides the enclitic forms, the copula in Mordvin languages exists as the separate verb улемс, which is suppressed in non-past tense clauses with present reading (it appears elsewhere and in non-past tense clauses with future reading). In the past tense, Erzya uses a paradigm of the frequentative form ульнемс rather than улемс (Turunen 2009: 268-9). Reportedly, only the free copula is allowed in existential and possessive clauses, differentiating them from locative predication (Turunen 2009: 253; Hamari and Ajanki 2022: 424).

3 Clause-level constituent order in Uralic and Mordvin languages

In this chapter, the available information on constituent order in Mordvin is presented, as a starting point for the present study.

The topic of constituent order has received a different degree of attention in various languages and the Uralic family belongs to the less researched — this holds particularly for the 'smaller' languages of language minorities; nevertheless, the general tendencies are quite clear and an outline is provided in the second section. The third, final part is a summary of data regarding word order in Erzya and Moksha based on the sources I managed to reach, which will be the basis for comparison and discussion in the later part of this work. In general, the available data is fractured and incomplete, although based on dialect descriptions (e.g. Ermuškin 2004) and studies in other areas of syntax (in Turunen 2009, stark contrast between literary Erzya and its dialects in terms of usage of different predication strategies is highlighted on many occasions) it might well be the case that literary Erzya and Moksha are more similar to one another (and perhaps to Russian) than to other variations of their respective languages.

3.1 Word order in Uralic languages

Despite a long tradition of research, description of syntax in Uralic languages is found lacking by Klumpp et al. (2018). In the introduction and summary to the chapter dedicated to word order in the Oxford guide to Uralic Languages (2022), one of the most recent comparative summaries of the family, Vilkuna (2022: 550, 560) stresses the necessity of further investigation into the workings of word order in various varieties and genres of the language family, especially in languages other than Estonian, Finnish or Hungarian, echoing her own conclusions from over a quarter of a century ago (Vilkuna 1998: 175, 222). Below is a short summary of the most important generalisations with regard to the language family, which will serve as the starting point and a backdrop for a more detailed discussion of Mordvin in the following section.

In terms of the order of the main constituents in a transitive clause, all Uralic languages are classifiable as either SVO or SOV, displaying a variable degree of flexibility (by which presence, but not equality, of different word orders different from the 'standard' is understood). Although in certain languages of the family both orders are found to be unmarked (even if not equally frequent), usually one of them is used to convey certain pragmatic information (Vilkuna 1998: 175; 2022: 950).

According to Vilkuna (1998: 178; 2022: 950), four groups can be distinguished

with regard to the dominating order and its rigidity, partially reflecting the "contact situation of the languages":

1. Rigid SOV order languages.
2. Eastern SVO, i.e. languages under a strong influence of Russian, with "extensive ordering variation including unmarked, verb or object focusing SOV, variable WH position" (Erzya and Moskha both fall into this category).
3. Western SVO, where the order is defined by "strong verb-second tendencies", with an "initial WH position".
4. Hungarian, with its strong topic-focus-verb structure overriding the syntactic function of the constituents.

Despite the ambiguous position between the two "big" clause-level orders, at a phrase-level structure Uralic languages generally behave like typical head-final languages of the Eurasian OV type (Dryer 1988; 1992), themselves being an important contributor to this category (Vilkuna 2022: 950-51), and the same holds also for Erzya and Moksha:

- adjectives and possessive modifiers, as well as numerals, precede their heads (but with at least some exceptions in Mordvin: see (17) and the explanation below);
- demonstratives precede the head noun (the Mordvin definite declension endings, historically deriving from a postpositional article, from the diachronic point of view could be cited as an exception);
- typical are premodifying modifiers of adjectives and adverbs (e.g. (E) пек паро 'very good', (M) пяк лац 'very well');
- the standard of comparison precedes the adjective/adverb (though the opposite is also possible at least in Moksha, see (18));
- postposition is the preferred type of adposition (no prepositions in Mordvin (Bartens 1999: 163));
- non-finite relative clauses precede their head similarly to adjectival modifiers, finite relative clauses are formed using relative pronouns and follow their heads;
- subordinating elements are verb-patterners (they precede finite subordinate relative clauses in Mordvin);

- subjects of existential sentences and predicate nominals precede the copula in OV languages (and Hungarian) and follow it in VO languages;
- adverbs "are to some degree object patterners" but otherwise appear freely within the clause according to their pragmatic function; "adverbs of manner, intensity, and frequency, such as 'well', 'often', and 'quickly', tend to follow the verb in western Finnic and Saamic and precede it elsewhere" (Vilkuna 2022: 951), with a tendency to appear immediately before the verb "in all the languages that allow unmarked SOV order", Erzya being given as one example of such a language. According to Hamari and Ajanki (2022: 421), in Mordvin "adverbial phrases can occur in any position in a sentence", and this is largely true, although preposed adverbs appear to be slightly more common (total ratios of occurrences of pre- and postverbal adverbials of manner — including converbs — of 101:75 and 157:107 were found in the present corpus in Erzyan and Mokshan texts respectively).

Despite following the general Uralic type, some exceptions to the rules can be found in Mordvin, as in (17a), with an adjective following a noun and receiving an agreement marker in a dialect of Moksha, or cases such as (17b) and (17c), categorised by Vilkuna (1998: 223-6) as possessor extraction, in which the entire construction could be interpreted as crossing the line between attributive and possessive predication.

(17) a. Moksha (Kholodilova and Toldova 2018: 286)

mon rama-jn'ə t'ɛ kn'iga-t' in't'ir'esnaj-t'
1SG buy-PST1.1SG>3 DEM book-PL interesting-PL

'I bought these interesting books.'

b. Moksha (KK1)

Народ-оньке минь простой
people-NOM.PX.1PL POSS.1PL simple

'Ours is a simple people.'

c. Erzya (Vilkuna 1998: 226)

Sudo-zo Olda-en' viškin'e
nose-NOM.3SG Olda-GEN small

'Olda's nose is small' or 'Olda has a small nose.'

In a comparative construction in Moksha, where the pronominal standard of comparison сяда 'this.ABL' is treated as a grammaticalised comparative marker (see Hamari and Ajanki 2022: 422) and the real standard is marked with the postposition коряс 'according to', the adjective can precede said standard:

(18) Moksha (O. E. Polyakov 1995: 28)

Тумо-сь сяда сери келу-ть коряс.
oak-NOM.DEF COMP tall birch-GEN.DEF according.to

'The oak is taller than the birch.'

Certain widespread, easily borrowed formulaic constructions characteristic of specific pragmatic circumstances appear also in Uralic languages, despite apparently breaking the usual conventions of constituent order. Vilkuna (2022: 956-7) mentions the presence of narrative-initial VS ordering 'was man and woman' even in the typically rigidly word-final Khanty and a narrative 'turning point' function of a similar structure in Estonian. Reporting clauses of a similar type following direct speech are yet another example of such a construction.

(19) Moksha (KK1)

(Вдь вов мезе мялец тиемс мархтонк,)

кяжи маци-кс кжназев-сь сон
angry goose-TRA rasp-PST1.3SG 3SG

'(See, this is exactly what he wants to do to us,) rasped he like a mad gander'

While researching, one has to be mindful of such constructions.

Interrogative pronouns (a.k.a. question words, WH-words) tend to appear in situ in OV languages and (mostly) clause-initially in the VO part of the family tree. The same holds for Mordvin, where content interrogative clauses most often start with a question word, although there are rare instances where this does not hold, as in 20 where the apparently topical subject pronoun сон precedes the question word зярдо:

(20) Erzya (ECK)

Сон зярдо ту-сь?
3SG when leave-PST1.3SG

'When did she leave?'

3.2 Word order in Mordvin

Accordingly with the discussion in the previous section, Hamari and Ajanki (2022: 419) describe constituent order in Mordvin in the following way:

"*The Mordvin languages display head-final* NPs*, postpositional phrases, and head-final relative clauses, though relative clauses can also be placed after their*

heads. Auxiliaries are usually preverbal. It is typical of transitive, intransitive, and non-verbal predicate clauses that the subject precedes the predicate but the word order depends on information structuring and the clause type. In transitive clauses the basic word order is SVO*; nevertheless,* SOV *is also common. Variation of word order should be a target of further research.*"

All authors seem to agree in their classification of Mordvin languages as (S)VO with a flexible order of constituents, meaning that all combinations are grammatically allowed, even though not necessarily pragmatically equal (Salamon 1989: 89; Zaicz 1998: 206; Bartens 1999: 168; Rueter 2010: 26; Keresztes 2011: 104; Kholodilova and Toldova 2018: 549-550; Hamari and Ajanki 2022: 419; Rueter 2023: 514). The "default" classification as SVO refers to the pragmatically unmarked, most popular constituent order in transitive clauses with two overt arguments; however, whenever topicality comes into play, and depending on constituent type (including presence or absence of morphological indexing, part-of-speech or zero-occurrence) the actual ordering can differ drastically from the prototype — and certain orderings, while allowed, in reality appear quite rarely (Kholodilova and Toldova 2018: 609). A brief synthesis of the available data is presented below.

The second most popular word order after SVO in Mordvin is SOV. Salamon (1989) describes both most prominent orders as unmarked, devoid of focus; and Vilkuna (1998: 181) claims that SOV is "an archaic variant with no particular function" in Mordvin. Bartens (1999: 168) describes the languages as SVO, mentioning the existence of numerous traits indicating an earlier dominance of SOV. Saarinen (1991: 8) is cited by Vilkuna (1998; 2022) for her claim that 19th century folklore texts appear to indicate that "SOV was once common too". But this should not be taken at face value as an immediate proof of the shift having rapidly taken place during the last two centuries and the SOV ordering being merely a vestige of the very recent shift. For instance, the closely related Finnic languages, treated by typologists as languages which shifted to VO "relatively recently" (Siewierska 1998: 8-9,307) yet appearing as SVX "all through their written history" starting as early as the 16th century (Hamari and Ajanki 2022: 952), apparently still show signs of OV dominance in the oldest written texts and even more recent folklore — V. N. Polyakov et al. (2018) cite studies of Estonian literary texts from 16th century and folklore, among others, folk sayings, which reveal high rates of OV ordering; as well as alleged dominance of SOV in E. Lönnrot's "Kalevala", a literary text based on or at least heavily inspired by genuine folklore material.

Despite the aforementioned claims by, among others, Salamon and Vilkuna of SOV being no different from SVO, Rueter (2010: 26-7) suggests (in his discussion of Erzya syntax) that the former indicates contextuality, contrasting the SVO ordering in a context-free sentence (21a), answering the question 'what happened

while I was away?' with (21b), being rather an answer to the question 'why are you building a fence?' (see also Vilkuna 2022: 953 for comparison).

(21) a. Erzya (Ermuškin 2004: 155)

veŕgize-ś *sala-ś* *ŕeve*
wolf-NOM.DEF steal-PST1.3SG sheep

'The wolf stole a sheep.'

b. Erzya (Rueter 2010: 27)

veŕgize-ś *ŕeve* *sala-ś*
wolf-NOM.DEF sheep steal-PST1.3SG

'The wolf stole a sheep.'

A complicating circumstance is that of the different distribution of word order in presence and absence of an overt subject, and perhaps even depending on its quality.

At least in the case of Moksha, SOV is "the most common order when both S and O are instantiated by pronouns", with 2SG subject being the only to trigger a "strong SVO ordering, such that a 1SG object is most likely to follow the finite verb, and a 3SG object has a 50% chance of [either] preceding [or] following the finite verb" (Rueter 2023: 514).

Dialectal differences seem to add even more confusion to the issue of SVO vs. SOV or VO vs. OV. For instance, definite objects are said to appear preverbally and indefinite postverbally in the Srednetyoshski dialect of Erzya. There, the OV order is generally preferable to VO in the absence of a *pronominal* subject, however, in presence of such a subject, OSV is the most likely outcome (Ermuškin 2004: 155-6).

Vilkuna (1998: 187-9) analyses the occurrences of preverbal unmarked objects, suggesting that such objects, "at least in Permic", with a tendency to precede their verbs without interruption, might be similar to incorporated pre-verbal objects in Hungarian. Note that this behaviour harmonises with OV languages' preference for suffixation (Greenberg 1990; Dryer 1992): the presence of suffixes would be an obstacle to postverbal incorporation. However, it seems not to apply to Mordvin languages, where non-definite nouns appear preverbally only very rarely (see also the note at Vilkuna 1998: 229, according to which "in the present corpus, the number of SXV clauses with an object is only 5, and not a single SOV with a nominative object was found. In the 36 OV clauses, the number of nominative objects was 16 (44.4%), while in VO and O-clauses, the proportion of nominative objects was 31.8% and 26.3%, respectively").

According to Ermuškin (2004: 157), indefinite arguments in the dialect are supposedly most likely to appear to the right of their definite counterparts, resulting

also in VS or OVS presentational-type orderings for indefinite subjects. OVS can be employed also in "out-of-the-blue" clauses in numerous Uralic languages (taking on the form of OVS in VO languages and OSV in OV languages), including Erzya (Vilkuna 1998: 190). In these clauses, such as 22, the subject is not topical and the event expressed by the verb is mentioned seemingly without any previous context (note also the indefiniteness of the subject *guj*).

(22) (Vilkuna 1998: 190)

Mon' *susk-imim* *guj*
1SG.GEN bite-PST1.3SG>1SG snake

'A snake bit me'

In Salamon's assessment (1989: 89), object-initial clauses highlight definiteness (topicality) of the object, at the same time highlighting the subject as contrastive (the example given there is 23, which suggests that only OSV was meant here), in OVS, a secondary stress can be felt on the verb.

(23) Moksha (Salamon 1989: 89)

śe *t́äjxí́r-ińt́* *f́ed́a* *ńäĺǵ-iźä* *rozbojńik-iń*
DEM girl-GEN.DEF Fedya seize-PST1.3SG>3SG highwayman-GEN
ḱecta
hand.ELA

'ezt a lányt Fegya ragadta el a rablótól'
'as for this girl, it was Fedya who snatched her away from the kidnapper'

The activity is stressed when the verb is used in the initial position, the secondary stress is placed on the final element: the subject in VOS, the object in VSO (Salamon 1989: 89).

(24) Moksha (Salamon 1989: 89)

a. *polavt-i̯ź* *zi̯ń* *si̯ńś* *e̯st* *nurdo-st*
exchange-PST1.3PL>3 3PL 3PL.GEN.EMPH own sledge-PX.3PL
'They traded their own sledge.'

b. *kiĺ-ś* *ĺišme* *vardo* *i* *śej-ś* *uńaša-ń*
harness-PST1.3SG horse Vardo and tear-PST1.3SG Unyasha-GEN
usko-mo
drag-INF2
'Vardo harnessed a horse and left, taking Unyasha away.'

In Ermuškin's data adjuncts of time, place or manner as a rule appear preverbally and the immediately preverbal position bears focus (Ermuškin 2004: 157). Also (Vilkuna 1998: 202-3) highlights the tendency for the verb in Erzya to appear in the immediately preverbal position. Examples of different pragmatic meanings in a transitive clause with variously ordered elements, including a pre- and postverbal adverbial, have been provided by (Zaicz 1998: 206).

(25) Erzya (Ermuškin 2004: 157)
baba-j, *taw–kodamo* *ser'ej* *ašo* *pr'a* *od* *c'ora* *min'ek*
grandmother-VOC some.sort.of tall white head young man POSS.1PL
ambar *vaks* *lotka-s'*
granary next.to.ALL stop-PST1.3SG
'Grandma, some tall, blond young man has stopped next to our granary!'

3.2.1 Statistical data on order of main constituents in clauses

There are a few datasets pertaining to the distribution of different word orders; I managed to find three. One of them originates from Vilkuna's older summary of word order in Uralic languages (1998: 186):

Table 1. Word order in Erzyan clauses according to Vilkuna (1998). Note that the definitions of S, O and X are more relaxed for languages other than Finnish, therefore the results may not be fully comparable.)

Order	Erzya	Komi	Udmurt	Finnish (X=O; S,O=NP)
SVX	65	49	37	423
SXV	11	19	218	10
XSV	13	8	83	27
XVS	16	12	53	74
VSX	2	10	10	7
VXS	1	0	4	0
Total	108	98	405	541

In this study, X represents "a complement — an object, a predicate complement, or an oblique complement of some sort, and an NP or a PP". There were a total of 920 Erzya, 963 Komi and 5500 Udmurt clauses in the respective corpora. Erzya and Komi are represented here by complete texts, Udmurt — a subcorpus where "each individual text of the complete corpus is represented by approximately 200 clauses of running text". The Finnish data provided for comparison comes from the "so-called HKV corpus" with about "10 000 clauses of expository Finnish prose, mainly newspaper texts". (Vilkuna 1998: 185-6)

In Vilkuna's opinion (1998: 181), the degree of relationship of the SVO/SOV order alternation in Mordvin with that found in Russian requires further investigation. More recently, she writes: "the literary norms and the written texts are influenced by Russian normative grammar, while everyday discourse may reflect the extreme variability of spoken Russian word order" (Vilkuna 2022: 951). This seems accurate: the statistical distribution of word order found in Erzya corpora, rather than to its close relative Finnish, seems more similar to the areally connected Komi placed under a similarly strong Russian influence, or even to Polish where, according to Siewierska and Uhlířová (1998: 121), the SVO order decidedly dominates with an over 69.1% share of all clauses with both overt arguments, three other orders (VSO, VOS, OVS) contribute 7-11% each and the rate of occurrence of the two remaining orders is negligible.

Another, more specialised study was conducted by Kangastus during her research of participant coding in Erzya. The study was based on the Erzya-language magazine *Сятко* and concerned exclusively clauses with objective conjugation. (Kangastus 2012: 53-54)

Table 2. Word order in clauses with objective conjugation, according to Kangastus (2012). Out of 525 examined clauses, 100 (19% of the sample) contained neither an overt subject nor an object)

Order	Occurrences	Share
SVO	104	19.8%
SOV	26	5.0%
OVS	19	3.6%
OSV	17	3.2%
VSO	3	0.6%
VOS	2	0.4%
Total	171	33.3%

Order	Occurrences	Share
SV	36	6.9%
VS	3	0.6%
Total	39	7.4%

Order	Occurrences	Share
VO	135	25.7%
OV	80	15.2%
Total	215	41.0%

Kangastus notes that the internal structure is rather loose in that there may

be intervening elements between the main constituents, such as adverbials of time and manner. VO order is prevalent both with and without an overt subject, the null-subject being the most common (Kangastus 2012: 54-55). The frequency of OV relative to VO is much higher than that of SOV in relation to SVO — in light of the information provided above it is indeed an expected outcome.

Additionally, she provides also results of cross-examination of the arguments' position with their type (pronominal or otherwise) (Kangastus 2012: 56):

Table 3. Order of different types of objects

Argument type	SV	VS	OV	VO
Pronoun (incl. pers. pronouns)	87 (68)	4 (3)	48 (24)	22 (13)
NP	96	23	94	223
Total	183	27	142	245

Citing Bartens, she notes that a pronominal subject is more likely to appear in the topical, initial position due to its typically anaphoric function; an indefinite subject is more likely to trigger reversal of the word order, especially if it is semantically agentive (Bartens 1999: 173). A pronominal object, even if limited to personal pronouns, precedes the verb more often than follows (Kangastus 2012: 57); it seems not unlikely that many of the OV cases were triggered by the presence of a pronominal subject in the same clause as in such a case they tend to appear right next to one another (as in Moksha according to Rueter, mentioned above), although exceptions also occur:

(26) Erzya (Kangastus 2012: 57, with spelling corrections)

a. [--] *śeste iśt'a ladja-sińek: tiń mińek kavańa-samiź,*
then this.way arrange-1PL>3 2PL 1PL.GEN indulge-3PL>1
miń — tink...
1PL 2PL.GEN
'then we arrange it like this: you indulge us — we indulge you'

b. *Kodak sońś Sardanapal pačkod'-i t'e tarka-ńt'eń,*
as.soon.as 3SG.EMPH Sardanapal arrive-3SG DEM place-ALL.DEF
mon śeske vana t'e vař-ińe-va-ńt' sonze
1SG immediately lo! DEM hole-DIM-PROL-DEF 3SG.GEN
l'ed'-sa nal-so-ńt'
shoot-3SG>3SG arrow-INE-DEF
'As soon as Sardanapal himself arrives here, I will immediately shoot him with an arrow through this small hole here.'

For Moksha, a statistical overview can be found in Kholodilova and Toldova (2018), based on the variety of the villages Lesnoe Cibaevo, Lesnoe Ardaševo and

Lesnye Siyaly (Лесное Цибаево, Лесное Ардашево, Лесные Сиялы respectively), located in the sphere of influence of the northern section of the central dialect group of Moksha. The data was collected separately for intransitive and transitive clauses (Kholodilova and Toldova 2018: 610-11) based on a subcorpus of KMT (A Corpus of Moksha Texts, collected from native speakers during the fieldwork) with 1160 clauses (847 sentences), all clauses containing a verb in indicative or conjunctive mood, with omission of clauses where all predicate arguments are missing:

Table 4. Relative position of the verb and the subject in intransitive clauses with respect to the type of the subject phrase, based on Kholodilova and Toldova (2018: 610).

	Personal pronoun	Definite/possessive NP	Non-definite NP	Total
SV	97	188	11	296
VS	4	59	51	114
Total	101	247	62	410

It is noted that definite NPs tend to appear clause-initially, though their number of postverbal occurrences is not negligible. VS is typical for introductory clauses, e.g. with movement verbs (*sams*/*sašəndəms* 'to come') or stative/locational verbs (*er'ams* 'to live', *aščəms* 'to be located') — compare this to Bartens' assertion that non-definite subjects are likely to trigger an inverse ordering of the constituents.

Table 5. Distribution of constituent order in Moksha depending on the conjugation type on the predicate based on Kholodilova and Toldova (2018: 611)

Order	Definite (objective)		Indefinite (subjective)		Total	
SOV	17	89	2	51	19	140
OV	66		49		115	
OSV	1		0		1	
OVS	5		0		5	
SVO	28	87	24	54	52	141
VO	54		28		82	
VSO	3		2		5	
VOS	2		0		2	
Total	176		105		281	

Here, with subjective conjugation (and, respectively, zero-marking on the object), SOV hardly ever occurs.

In transitive clauses, the subject is rarely situated after the verb (or object), unlike in non-transitive clauses (Kholodilova and Toldova 2018: 551):

Clause type	SV	VS
Intransitive	296	114
Transitive	72	12
Total	368 (74%)	126 (26%)

Object-initial clauses are rare, typically with the object acting as the topic:

(27) Moksha (Kholodilova and Toldova 2018: 612)
ava-n' *panar̥-n'ə-n'* *maša* *kuvac* *stakšn'ə-sin'ə, a*
woman-GEN dress-PL.DEF-GEN Masha long.time sew-3SG>3PL and
al'ɛ-n' *panar̥-n'ə-n'* *kurək-stə*
man-GEN dress-PL.DEF-GEN quick-ADV
'Masha sews dresses slowly, but shirts — quickly.'

VS and OS-types are the least common, together constituting 20 out of 5000 clauses in the entire KMT corpus (Kholodilova and Toldova 2018: 609).

The subject influences the position of the object: SOV is less common than SOV but in absence of the overt subject (which is the majority of all cases), OV is preferred over VO (Kholodilova and Toldova 2018: 611). The reverse is also true: the majority of OV clauses in the corpus are clauses with their subject absent.

The relative frequencies of different orderings of constituents follow a similar hierarchy with both types of conjugation, although the exact percentages differ.

Subjective: OV (47.57%) > VO (27.18%) > SVO (23.30%) > SOV (10.30%).

Objective: OV (40.00%) > VO (32.73%) > SVO (16.97%) > SOV (1.94%).

3.2.2 Auxiliaries and modifying verbs

Vilkuna (1998: 212) reports no auxiliaries in Erzya, although the verb кармамс 'to begin' does have a secondary auxiliary function in denoting future tense in both Erzya and Moksha (Hamari and Ajanki 2022: 411). This verb precedes its complements as do other verbs with this ability (Hamari and Ajanki 2022: 419), although historically auxiliaries used to follow their verbs, as can be seen in a form of leftovers fossilised as morphology (Bartens 1999: 168-9). The order within an infinitival complement is typically VO in Erzya, although dependents of infinitives, including objects, are easily extractable and can end up even on the other side of the finite verb than the infinitive (Vilkuna 1998: 217-8).

In the dialect of Moksha described by Kholodilova and Toldova (2018), the order of infinitive and its object depends on the type of the infinitive: VO is ungrammatical with the *-(ə)ma* infinitive and allowed with the *-(ə)ms* infinitive, although only if the object bears plural nominative marking, rather than the singular genitive (an object of an infinitive cannot be unmarked). A sample from

the newspaper “Мокшень правда”, representing the literary language, showed preference for VO with both the aforementioned types of infinitive (although somewhat less with the –ма infinitive), and a noticeable preference for OV with the -мда infinitive. (Kholodilova and Toldova 2018: 676-7)

Table 6. Position of object in relation to its head in infinitival complements, based on a sample from "Мокшень правда". The relative frequencies of particular types of infinitive in the table are not representative of their relative frequencies within the corpus. (Adapted from Kholodilova and Toldova 2018: 677).

Infinitive	VO	OV	proportion of VO
-мс	55	9	0.9
-ма	45	23	0.7
-мда	3	13	0.2

3.2.3 Interrogative clauses

In Erzya, including the Srednetyoshski dialect, polar questions are marked by intonation only (stressing the preverbal focus), although question particles can also be added (e.g. ли, паряк) (Vilkuna 1998: 208; Ermuškin 2004: 158). In Moksha, the situation is similar, with optional question particles such as ли, што ли, деряй (with certain phonetic and orthographic variation) and others. The stressed element within the question (if existing) can be moved to a position of focus: clause-intitial or preverbal (Erina 1997: 121-126; Kholodilova and Toldova 2018: 548). Different possible orders, showing different pragmatic focus, are presented in (28).

(28) Moksha (Kholodilova and Toldova 2018: 548)

a. *maša t’ɛči pid’ə-s’ (l’i) jam?*
Masha today cook-PST1.3SG Q porridge
‘Did Masha cook porridge today?’ (unmarked)

b. ***maša*** *(l’i) jam-t’ t’ɛči pid’ə-z’ə?*
Masha Q porridge-GEN.DEF today cook-PST1.3SG>3SG
‘Was it **Masha** who cooked the porridge today?’

c. *maša (štol’)* ***jam*** *pid’ə-s’? /*
Masha or.what porridge cook-PST1.3SG
jam *(l’i) pid’ə-s’ maša?*
porridge Q cook-PST1.3SG Masha
‘Is it **porridge** that Masha cooked today?’

d. *maša* ***t’ɛči*** *jam-t’ pid’ə-z’ə? /*
Masha today porridge-GEN.DEF cook-PST1.3SG>3SG

t'ɛči *(l'i) maša jam-t' pid'ə-z'ə?*
today Q Masha porridge-GEN.DEF cook-PST1.3SG>3SG
'Was it **today** that Masha cooked the porridge?'

Content questions in Erzya and Moksha typically start with a WH-word, but when the word functions as the predicate, it appears clause-finally (29). Vilkuna (1998: 208) states that "frequent but not obligatory sentence-initial position [of the WH-word] is reported for Erzya and Komi and is supported by the present corpora", with 3 non-WH-initial and 37 WH-initial occurrences in her Erzya corpus.

(29) Erzya (Ermuškin 2004: 158)
n'et' jarmak-tne ko-sto?
DEM.PL money-NOM.PL.DEF where-ELA
'Where is this money from?'

3.2.4 Non-verbal clauses

Turunen (2009: 261) distinguishes between three types of non-verbal predication in Erzya (example 30), and the same categories can be applied also to Moksha:

1. "Predicative suffix" construction (in this work considered as bound-copula construction);
2. Zero-copula construction;
3. Copula construction.

(30) Erzya (Turunen 2009: 261)
a. *Kodamo kežej=at...*
what.like angry=2SG
'How angry you are...'
b. *Ton laborant.*
2SG technician
'You are a technician.'
c. *Ška-t'ńe ul'ńe-śt' śed'e vešola-t.*
time-PL.DEF be-PST1.3PL CMP cheerful-PL
'Those times were happier.'

Although the subject preceding the predicate is considered unmarked, as is typical in Uralic languages (Vilkuna 1998: 198), other orders are also possible. The

flexibility may depend on the type of predication (bound copula/free copula/zero-copula) and the class of the predicate (indefinite/definite noun, adjective, locative phrase. . .).

The first and last type of predication are characterised by a fairly free word order (Turunen 2009: 266-7) and no necessity for an overt subject (30a). The zero-copula construction is simultaneously described as one with a "rigid" word order, but compare (31), where adjectives functioning as predicates precede their pronominal subjects (obligatorily agreeing with them in number; see Bartens 1999: 108 and Rueter 2023: 517).

(31) a. Erzya (Turunen 2009: 284)

Vaj, čaula-t miń!
oh stupid-PL 1PL

'Oh, we are stupid!'

b. Moksha (Rueter 2023: 517)

kel'mə t'alə-ś
cold winter-NOM.DEF

'it is one cold winter'

Although the bound copula used in the first type of predication named by Turunen typically attaches to the head of the predicate, it can also appear on other constituents, in particular, if they receive focus through a modification in the word order — this applies to both Mordvin languages. The most typical case is that of the copula appearing on an adjective attribute of the predicate noun when their order is reversed (32), with the head noun receiving obligatory definite marking). (Turunen 2006: 183-5; Kholodilova 2016: 251-2)

(32) a. Erzya (Turunen 2006: 184)

mon lomań-eś b'eŕań=an, mon lomań-eś plohojń=an
1SG person-NOM.DEF bad=1SG 1SG person-NOM.DEF bad=1SG

'I am a bad man, I am a wretched man'

b. Moksha (Kholodilova 2016: 251)

ťiń saśet-tńə plox=tadə
2PL neighbour-PL.DEF bad=2PL

'you are bad neighbours'

Sometimes, at least in Moksha, the copula can appear on the subject, rather than the predicate, when it is in focus. Also in this case, a shift in word order is possible, although not necessarily required. (Kholodilova 2016: 249-250)

(33) Moksha (Kholodilova 2016: 249-250)

a. *kucə* *mon=əĺ-əń*
house.INE 1SG=PST-1SG

b. *mon=əĺ-əń* *kucə*
1SG=PST-1SG house.INE
'It is me who was at home'

4 Texts used in the study

This chapter presents two well-established Mordvin authors: Kuz'ma Abramov known for his works in Erzya and Piotr Levchaev writing in Moksha, as well as synopses of their stories used in the present corpus. Presentation of this meta-information of sorts has a double purpose: it demonstrates the way in which the two parts of the dataset are comparable with one another, while highlighting certain differences which may affect the statistical results. As has been mentioned in the introduction, it is not possible in a thesis of this format to even attempt to make a complete overview of Mordvin-language literature. A good understanding of the smaller set of data should therefore compensate for a possible lack of the representativeness generally expected of a good corpus (Stefanowitsch 2020: 28-36).

The choice of authors and works was narrowed down by several principles, listed below:

- The texts should be available in their original form, whole and unabridged, to eliminate the risk of errors caused by the process of digitisation or arbitrariness of the corpus' author, as well as later difficulties possibly caused by a lack of the complete context. In the best case, such texts should be available in the form of a physical copy or a good-quality scan. Investigating whole texts instead of snippets provides a sense of balance of the data, with introductions and endings included along with the content; it is also easier to follow the texts, knowing them from the start to the end.

- The texts selected for the study should be comparable: created in a roughly similar time period (e.g. works from the 1920s would not be quite comparable with those written in the 1950s), of a similar genre.

  - Ideally, the texts should represent simple narrative or expository prose: word order in e.g. poetry is a much more difficult issue to tackle — this being said, theatre plays might be an interesting type of data, which might, however, require a specific approach.
  - I wanted this study to be based on a stable, developed form of the relatively ew literary languages (a partially arbitrary decision but also a way to avoid excessive difficulties with the material), which could not be guaranteed in the case of the texts published in the interwar period which have entered public domain and are available online on sites such as Wikisource[1] or Mordovskaya Biblioteka[2].

[1] https://wikisource.org/wiki/Main_Page/Эрзянь, https://wikisource.org/wiki/Main_Page/Мокшень (both last accessed 16.05.2024)

[2] http://mordovlib.ru/ (access 03.04.2024)

- The authorship should be known and the authors would at best both be respected literary figures within their respective traditions.

This essentially limited my choice to the physical collection of the library of the Institute of Estonian and General Linguistics of the University of Tartu, and under the aforementioned constraints, the selection of works used in this study appeared satisfactory. The texts are four pieces of narrative prose fiction, all written/published approximately in the 1960s-70s; three of them were included in the corpus in their entirety, one is represented by its first two chapters (it could be said that each chapter constitutes an entity in its own regard), each contribution constitutes roughly a quarter of the corpus. As for the genre, the Erzya texts are classified in their sources as ёвтнема 'short story', those written in Moksha as повесть 'novella', the main difference between the two types apparently being their usual size and scope (повесть is typically longer than ёвтнема but shorter than a novel; the longest text, "Кафонц кудат" 'Two pairs of houses', of which only a part was used in this study, is also a повесть); I considered them similar enough in terms of their content to not affect the comparability negatively (though see commentary later in this chapter). Altogether, they form a corpus of 19 539 tokens: 8773 Erzya and 10 766 Moksha, with around 3329 clauses (E 1623, M 1706) containing more than one constituent and thus being useful in the present study.

Regarding the authors: although it cannot be said that their biographies, characteristics and reception fully mirror each other (this is indeed very rarely the case), both were born and grew up in a similar time and conditions, and both are respected and considered important creators within their literatures in their own right. Both are featured in a curriculum of the national literature classes in Mordovia (Katorova 2022: 19-21, 49-51) and both are regularly named among the most important writers in their respective languages. For instance, in an article[3] dedicated to the 95th anniversary of the Mokshan literary journal Мокша, Levchaev is mentioned among the contributors during the interwar period, and his post-war collection of short stories "Стирнят-якстернят" 'Pretty girls' is also named. Fragments of his works are frequently cited as examples by Serebrennikov et al. (1998) in their Moksha-Russian dictionary. Abramov, on the other hand, is one of the only three people mentioned under the headword "Мордовские писатели" 'Mordvin writers' in the Great Russian Encyclopedia[4]. For both of them, there exists a Wikipedia article in at least their native language[5]. The biographies presented below are mostly based on the respective entries in Bryzhinskiĭ et al. (2001), available verbatim also in the alphabetic index of the website "Библиотека

[3]https://natlibraryrm.ru/moksha-zhurnalti-95/ (access 03.04.2024)
[4]https://bigenc.ru/l/mordovskie-pisateli-26907b (access 03.04.2024)
[5]Levchaev: https://mdf.wikipedia.org/wiki/Пята_Левчаев;
Abramov: https://myv.wikipedia.org/wiki/Абрамов,_Кузьма_Григорьевич

мордовской литературы" 'The Library of Mordovian Literature' dedicated to literary publications in Mordovia[6].

4.1 Kuz'ma Abramov

Kuz'ma Abramov (Кузьма Григорьевич Абрамов), was born in the year 1914 to a peasant family in the Starye Naymany village in the South-Eastern part of present-day Mordovia, in 1920 his family moved to the settlement of Molnya not far away (Abramov 2009: 56). His place of birth and upbringing would suggest the strongest influence of the South-Eastern (or Sura) dialect of Erzya (Rantanen et al. 2022: Mordvin). After graduating from school in 1933 he attended a preparatory course at the pedagogical university of Mordovia but dropped out of the studies and obtained employment as a teacher. Before being drafted into the army in WW2 he worked in the Mordovian publishing house as an editor. After the war, in the early 1950s, he completed his education at the Pedagogical University.

Although Abramov's first publication (a collection of poetry) appeared before WW2, his literary output derives almost entirely from the post-war years, up until his death in 2008 in Saransk. Abramov entered the Writer Union of USSR in 1949, started publishing on a larger scale in the early 1950s with theatre plays and short stories, their number increasing towards the end of the decade. In 1957 he published his first novel, and that is the genre which became his specialty during the later part of his career, and which won him appreciation and respect in the Erzyan-speaking literary circles.

Kuz'ma Abramov is probably one of the most important representatives of Erzyan literature and his texts should likewise represent an 'ideal' writing style in the Erzya literary language. A 7-volume collection of his works (in Russian translation) appeared towards the end of the 20th century, and a 12-volume complete collection of his works was published after his death.

The two short stories (ёвтнема) used in this study were taken from the 7th volume of the posthumous collection, which, in addition to a set of short stories contains also the novel "Эсеть канстось а маряви" 'One does not feel their own burden'. Despite Abramov's general fame, his short stories do not seem to enjoy as much popularity as his novels, and mentions of them are not easy to find e.g. during internet search. It is, however, assumed that they are representative of the author's style regardless of their own popularity. Even though they are not explicitly dated, and I could not find any information in this regard — perhaps an outline of the contents of each volume of the "Collected works" is provided in the

[6]Accessed through a snapshot on the Internet Archive; the original website is currently inaccessible, not unlikely due to geographical blockade of certain websites hosted in the Russian Federation: https://web.archive.org/web/20200128081743/http://www.lib.e-mordovia.ru/Library/?Char=%CB&id=2

preface to the first one, but that was inaccessible to me — one can assume that they were written around the 1970s. This is based on two reasons: one external, i.e. their grouping in the collective edition of his works separately from both the earliest short stories and the unreleased texts and together with "Эсеть канстось а маряви" 'One does not feel their own burden' first published in 1967, and one internal, i.e. the action time of "Эрьке чиресэ кудыне" 'A little house at the lakeside' identifiable as close to 1980.

4.1.1 Эрьке чиресэ кудыне (ЕСК)

The story, with its title translatable as 'A little house at the lakeside', appears to be set around 1980, and nothing seems to indicate that this is not contemporary to its time of writing. The main character, Nikolai Chebretsov, returns from summer holidays to discover his youngest son having recently returned from the army draft and about to marry. Amidst preparations to the ceremony, a letter from Nadezhda, a mysterious woman from Chebretsov's past arrives, awakening his long suppressed memory from the summer of 1945.

Structurally, the story is divided into six chapters, with alternating segments set in the present and in the past usually coinciding with chapter divisions. Roughly 35 years separate the two time planes: this difference can be inferred from the fact that Chebretsov, now 59, barely managed to work a year in his profession obtained after a seven years' education at school and a six-month professional course before "the war" started, and at its end served in the rank of lieutenant — suggesting early 1920s as his birth date. The remark about him "not hurrying with marriage" before the war, which appears in the first paragraph of the story is another reason to assume that Chebretsov was not underage at the time of its outbreak (presumably understood here as 1941).

The anterior plane explores themes such as youth, attraction, but also self-dedication and loyalty. The issue of marital relations dominates in the contemporary plane, where now submissive and withdrawn Chebretsov suffers under the tight control of his dominating wife. The language of this story is rather ordinary, without any recognisable stylistic features. There is a fair share of both narration and dialogue.

4.1.2 Тона пельде (ТР)

'From the other side', with a subtitle "Войнань шкадо ёвтнема" 'A war-time story' is, in contrast to the slice-of-life type represented by ЕСК, more of an action story. In April of 1942, as a result of an unsuccessful breakthrough attempt, Syres'kin, a roughly 40-years-old *starshina* from the Soviet Army, is left stranded along with his wounded superior, the 28-years-old lieutenant Pichagov, behind the enemy

lines. They take up a dangerous journey across the occupied territory in order to secretly return to the friendly camp across a gap in the German entrenchments. On their way, they encounter Sima, a young nurse whose company has similarly been eradicated and who joins them in their attempt to escape the hostile land.

Similarly to ECK, here too the language is nothing out of the ordinary. Descriptions of the unfriendly environment at the end of the long winter of 1942 are important contributors to the general "survival story" mood. Dialogues appear, though they seem less frequent than in ECK, in line with the fact that it is the action, rather than the interaction, that drives the story.

4.2 Pyotr Levchaev

Pyotr Levchaev (Пётр Иванович Левчаев), was born in the year 1913 in the village of Promzino in the south-western part of present-day Mordovia (i.e. at the northern edge of the area of the Western dialect of Moksha (Rantanen et al. 2022: Mordvin)) to a peasant family and died in 1985 in Saransk. After graduating from a school for peasant youth he studied at the Workers' Faculty of Mordovia, whereupon he entered the Teachers' Institute of Mordovia where he studied in the years 1935-1937. In 1938 he was arrested and after a sham trial sent to a labour camp on Kolyma, in 1949 released and sent for permanent settlement to Krasnoyarskiy Krai ("Память" 2024). Only after Stalin's death, during the thawing period in 1956, was he rehabilitated and allowed to return to Saransk.

The literary activity of Levchaev started already in the interwar period. Initially, he was more active in poetry than prose: his first publication was that of the poem "Тонафнема" 'Study' in the newspaper "Од веле" 'New village' in 1930. In 1934, he became a member of the Union of Writers. During this period, Levchaev's works were selected into collections of Mokshan literature such as "Валда эряфти" 'Towards the bright life' (1934), featuring the most endorsed (and ideologically suitable) works in the nascent literary language.

The period of exile and estrangement created a large gap in Levchaev's literary career: his two consecutive book publications, "Даволхт" 'Storms' (1936) and "Вирсь увнай" 'The forest is murmuring' (1959) are separated by more than 20 years. Although in his later period Levchaev became known rather for his literary sketches and essays, full of wonder and fascination with the severe beauty of the cold North, some works of the later period also include short stories and novels, both in Mokshan and Russian.

Levchaev's texts were drawn from an anthology of his works published under the title Кярьмаз 'Bunch' (published in 1972). The two chosen for this study were originally published in 1962 in a collection called simply "Повестть. Расскаст" 'Novellas. Short stories', both belong to the former category. They should reflect the mature phase of his output but without an excessive artistic tint.

Events of both Mokshan texts take place before the revolutions which brought upon significant changes in the social relations and structure and both texts treat the oppression of Mokshans by the local aristocracy as their point of focus. Still, they manage to be distinct due to the differences in their attitude to realism, character building and style of narration. One thing to be wary about: the texts' setting in the pre-war period may be a cause for stylisation (archaising), especially within the dialogues.

4.2.1 Пичетне пичедихть (PP)

'The pines are mourning' is a story almost gothic in its tone, reminiscent of a dark fairytale where bad luck follows the main characters at every step and death is a frequent occurrence. The events take place in a forest inhabited by Mokshans living in dugouts, or pit-houses (Moksh., Russ. землянка) and earning their living by woodcutting, and at the nearby manor of the forest's owner, boyar Orest L'vovich Modzalevskiy. The latter uses his wealth and influence to seduce a young Moksha girl, Natasha, whom he subsequently sends away having heard the news of her pregnancy. Later, he takes a similar approach with Natasha's younger sister, Ulyasha, but this time his attempts are met with scorn. Ultimately, Modzalevskiy's dark deeds are not left unavenged, although an ambiguous end is the best the remaining characters can hope for.

The story is told from the point of view of a passer-by, visiting the forest and recollecting (somewhat wistfully) its events decades (maybe more than half a century) after they took place. Descriptions of nature are used to depict the characters' inner state of mind and to underscore the grim mood of the tale; however, they do not dominate. There is a certain balance between dialogue and simple narration, although the latter takes the leading role, leading the reader along a timeline of events spanning several years.

4.2.2 Кафонц кудат (KK)

This story, whose title can be translated as 'Two pairs of houses', is in some ways the opposite of PP, not shying away from humour despite the harshness of the reality it describes. It tells of the strife between the Mokshan folk and the local nouveau-riche aristocrat Makar Antonovich Tulai, whose ambitious projects spell economic doom for the local inhabitants and who abuses his power and connections to maintain full control of the local populace. Currently, as his son is coming of age, it is high time to secure a bride for him...

The story is longer and split into 12 chapters, of which only the first two were included in this study as the entire story is rather large in volume. The first chapter provides exposition to the story in the form of a dialogue: the local folk discuss

recent events by the fireplace, explaining their situation to Kirei Mikhailovich, a globetrotter currently employed by Tulai as a seasonal worker. The second chapter depicts negotiations between Tulai, the local owner, and Fyodor, son of Chil'ka (Чилькань Фёдор), whose daughter the former projects as his son's future wife. The large number of named characters comes to attention: this is likely due to the length of the story and hypothetically could have an impact on the structure of the narration, e.g. through a higher rate of occurrence of non-anaphoric subjects (although this is a feature of dynamics of the narration rather than of the number of actors).

The language of this story is lively, energetic, full to the brim with dialogue rich in idiomatic expressions, exclamations and discourse particles, etc. (This is also true of PP, though there it is less noticeable due to the lower importance of its dialogue). The dialogue is intended to provide characterisation, for instance, Tulai is depicted continuously repeating "мон шутендамазень аф кельгса" 'I don't like tricksters!'.

In citations, the two chapters are referred to as KK1 and KK2 respectively. This distinction is made due to the functional differences of the two chapters, the first being of an expository character, the other serving as a more standard episode of a story.

4.3 Conclusion

Kuz'ma Abramov and Pyotr Levchaev, born in a mere year's distance from one another, come from similar backgrounds: originating from peasant families, they obtained a 10-year school education, whereupon both directed their efforts towards obtaining higher education (with various rates of success). Both worked as teachers and made their first literary attempts before World War 2, their more 'mature' phase falling on the post-war years. There is one significant difference in their biographies: Levchaev spent almost two decades in imprisonment and exile while Abramov's career was 'only' interrupted by a few years of military service. In a sense, the former's new start after his return to Saransk could be analogous to Abramov's first serious steps on the literary arena around the same period, albeit with more experience to back it.

Levchaev's stories, due to their setting, are more likely than Abramov's to contain archaic elements of style, perhaps even consciously introduced by the author. Even if they do, they represent a specific type of storytelling, which is in itself a worthy object of research.

5 Preparation of the corpus

In this chapter, the procedure of preparation and annotation of the corpus used in the present study is described in detail. On the one hand, all key steps of the process are presented to give the reader an impression of the practical aspect of the study. On the other hand, the categories employed in annotation as well as in the discussion of the results in the following chapter are presented, with a particular emphasis on the less obvious examples.

5.1 Preparation of files for annotation

5.1.1 Digitisation of the texts

The texts selected for use in the study were digitised manually by typing the text in a text editor. Digitisation through optical text recognition of scanned pages of the books was considered but no ready algorithm was available for Mordvin languages and the algorithms recognising Russian and Bulgarian to which I had the access struggled with the text, requiring multiple manual adjustments after scanning and providing no advantage of time, consistency or precision over manual typing[7]. Additionally, the line-breaking along clause lines (see below) still had to be done manually.

The texts, when typed down, were split into lines following these rules (**NOTE**: from now on, unless otherwise specified, by "clauses" I will mean clauses *relevant for this study*, i.e. with at least two constituents qualifying for annotation — simple clauses, such as lone-standing verbs, do not contribute any useful information in the present context and generally were omitted during annotation):

1. No line should contain more than one clause (verbal or non-verbal).
2. Breaking of clauses into two or more lines is not allowed.
3. Repetition of material is not allowed (each token of the original text must appear exactly once in the digitised version).

[7]Characters such as л and д; н, и and к were confused too often, and due to their similar appearance on a computer screen this kind of errors seemed easier to miss than it is to mistype a character. One of perhaps the worst cases of misrecognition was identification of г as Latin r, which was very easy to miss, but utterly confusing to the automatic analysers used in later steps. In the end, it turned out not to be too much of a problem, as such cases could sometimes be identified later (e.g. by the inability of an analyser to deal with a specific form), the raw text level was mostly meant for the human eye and the automated parsing anyway turned out to be only useful to increase readability — but not as a tool for corpus annotation. However, at that stage it was not yet fully known how the data would be processed and imperfections of this kind were extremely undesirable.

Additionally, to allow for an estimation of the size of the dataset on the go, it was preferable that as many lines as possible do contain a "useful" clause (possible with an addition of parentheticals or neighbouring simple clauses with less than two constituents of the types considered in this study).

The original order of appearance was preserved to maintain integrity and readability of the texts, the only exception being clauses enclosed on both sides by another clause: when one clause was enclosed by another relevant clause from both sides, the internal clause was split off to a separate, usually following line, and a 'replacement string' "(_)" was inserted into the external clause in the place of each word thus deleted in order to represent the cutout segment. This facilitated readability and preserved the original distances between constituents in case they would be considered as a factor in this study (in the end, distances were not considered here except for some qualitative remarks). The division into clauses was further refined during the annotation process.

For instance, (34) is a valid clause, with a subject, a verb and an adverbial. So is (35), with only two constituents (a verb and its verbal argument), whose order is investigated in this study, making the clause eligible for separation. (36) is a pair of nested clauses. The clause containing the relative pronoun кить 'who.PL' would be tagged as a dependent clause at a later stage. In (37), a minimal finite phrase панжезень 'I opened it' follows a short non-phrasal utterance consisting of a conjunction and a particle (и афкукс 'and indeed'), a commonplace occurrence in dialogues. Neither of these is interesting from the point of view of this study, consequently, they did not need to be kept separated; keeping them together increases readability and reduces the number of 'noise' clauses. Note: the tagging of the verb панжезень in (37) was fully optional.

(34) Erzya (ECK)

Весемесь	*ютась*	*вадрясто.*
everything.DEF	pass.PST1.3SG	well
S	V	O

'Everything went well.'

(35) Erzya (ECK)

карми	*кевкстнеме,*
start.3SG	ask.INF2
V	V_ARG

'she'll start inquiring'

(36) Erzya (ECK)

Кавто	*солдаттнэ,*	*(_)*	*(_)*	*(_),*	*умок*	*ёматотсть.*
two	soldier.PL.DEF				long.ago	disappear.PST1.3PL
	S				T	V

кить *састь* *мартост,*
who.PL come.PST1.3PL with.PX.3PL
REL, S V COM

'The two soldiers, who arrived with them, had long since disappeared.'

(37) Moksha (KK1)

И *афкукс,* *панжезень...,*
and really open.PST1.3SG>3PL
(V)

'And he opened them, indeed...'

According to rule 3, if two verbs in two consecutive clauses shared some material, e.g. a single subject, it was only present within the clause containing the verb appearing closer to it. For instance, sentences of a frequence structure shown in example (38) were split into two clauses each, reducing the count of possible "full" clauses, but with the benefit of consistency. Compound verbs consisting of two heads and marked orthographically with a hyphen (Hamari and Ajanki 2022: 429-30) were treated as single verbs (39).

(38) Erzya (ECK)

Лейтенантось *озась* *лавсинентень,*
lieutenant.DEF.NOM sit.PST1.3SG shelf.DIM.ALL.DEF
S V

кирвазтсь *папироса.*
light.up.CAUS.PST1.3SG cigarette
V O

'The lieutenant sat on the bunk and lit a cigarette.'

(39) Moksha (KK2)

вадерязень *-удерязень* *штыкокс* *ащи*
stroke.PST1.3SG>3PL comb.PST1.3SG>3PL bayonet.TRA stand.PTCP
V

усанзон
moustache.GEN.PX.3SG
O

'he stroked his moustache, protruding like two bayonets'

A distinction was made from the start between narration and dialogue. I assumed that the text would be organised differently in these two styles of storytelling, not only due to the presence of subjects and objects in 1st and 2nd person, but also different topicality and speech style: dialogue is more prone to dialectal and

idiolectal stylisation for the purposes of characterisation. A notable example of this is the frequent repetition by Makar Tulai in KK of the phrase "мон шутендамазень аф кельгса" 'I don't like tricksters' and its variations[8]. Accordingly, it was important to track all occurrences, not only of dialogues, but also monologues, internal monologue (except that of the narrator), etc. Usually, they were clearly designated by punctuation and text formatting: em-dashes and indentations ("audible" speech) or quotation marks (internal monologue), but some cases proved more difficult. Examples include quotations of characters' thought process without any marks inside narration, cases of text between em-dashes within a spoken paragraph that could be interpreted as either an intrusion within the dialogue or an explanatory remark of the narrator (but narrator's remarks were typically additionally preceded by an additional comma) and editing errors — e.g. in TP on page 397 of the edition I used, an entire paragraph of narration is erroneously marked with an em-dash.

In the files, the distinction was made by placing a special unique symbol at the beginning of every clause classified as dialogue — a symbol distinct from the dash typically used to denote dialogues, as, of course, not every clause within dialogue starts with one, and clauses outside dialogue can also be singled out with dashes; I chose simply a single dot.

Besides dialogue, another type of clause was distinguished explicitly: reporting clauses with VS ordering used after direct speech (from now on referred to as REP), a frequent and formulaic type widespread among European languages (Vilkuna 2022: 956-7). To mark it, I employed a similar strategy here as with dialogue, just using a different symbol.

These clauses were identified by their special orthographic marking: the dialogue is interrupted by a dash and a comma (or any punctuation other than a full stop), and followed by non-dialogue starting with a non-capital letter. Only the first main clause after the dash received the REP marking, as other clauses never followed the VS pattern (they almost never contained a subject at all, and even then, the subject would not be the speaker). Below is an example of dialogue (40a)[9] followed by a reporting clause (40b) and its unmarked continuation (40c):

— *(...) Вов кодама сон ломань тонь азорце, Кирей Михалыч! — аделазе алясь баснянц и ушедсь шава трубканц нозома.* (KK1)

[8]It should be noted here that dialogue is not to be mistaken with oral language: it is merely a different storytelling tool, with its own rules and perks, arbitrarily interwoven into the narration. As a result it is not representative of any real discourse, and they represent, in any case, a single author's vision of what, and how, the personae dramatis might say in a given situation; a narrator can reveal himself, too, which happens e.g. at the end of Levchaev's PP: "*Мон аф весть яканъ-шяянь ся сире пичетненъ алга...*" 'Not once and not twice have I walked under those old pine trees...'.

[9]Note that this part is marked as a dependent clause due to the presence of the relative pronoun кодама 'what.like'.

(40) a. (DIAL, DEP)
Вов кодама сон ломаниць тонь азорце
lo! what.like 3SG person.NOM.DEF POSS.2SG master.PX.2SG
'Now you see what kind of person your master is'

b. (REP)
аделазе алясь баснянц
finish old.man.NOM.DEF fable.GEN.PX.3SG
'finished his tale the old man '

c. (*unmarked*)
и ушедсь шава трубканц нозома.
and begin.PST1.3SG empty pipe.GEN.PX.3SG suck.INF2
'and started to suck on his empty pipe.'

The only exception to the above rule was 41, where two parallel clauses were clearly connected and together followed the typical VS pattern of REP: бта 'as.if', depending on the context, was treated as either a relativiser or a particle, and in this case the presence of a dash suggests the latter interpretation, although by analysing бта биркас шапа керсь 'as if he were carving a notch on a tally stick' as a dependent and мярьгсь (...) Ачурсь 'said Ačur' as the main clause one could remove the 'exceptionality'.

(41) Moksha (KK1)

a. (REP)
мярьгсь —
say.PST1.3SG
'said'

b. (REP)
бта биркас шапа керсь Ачурсь
as.if tally.stick.ALL notch cut.PST1.3SG Ačur.NOM.DEF
'Ačur as though cut a notch in a tally stick.'

The rule of not marking dependent clauses as REP was applied in (42), where it can be seen that the VS pattern appears in the second, main clause.

(42) Moksha (KK2)
бта и ашезе няй Аннать,
as.if and NEG.PST1.3SG>3SG see.CNG Anna.GEN.DEF
N V O
'as if he didn't even see Anna,' (DEP)
кизефтсь Тулаюсь
ask.PST1.3SG Tulai.NOM.DEF
V S
'asked Tulai' (REP)

5.1.2 Preparation of the files for annotation in a spreadsheet editor

When each clause in the dataset (in some cases, with additional elements) was represented by a single row of a text file, each text file representing one of the texts used for the corpus (two text files were used to correspond to the two chapters of KK), the files were transformed with a script for its later annotation in a spreadsheet editor (easy to control interlinearity and allows flexibility with the format):

1. Word separators were changed from spaces to tabulations, so that the data would be easily transferable between a text editor and a spreadsheet editor[10].
2. An additional, reserved "0th" column was added (represented by a line-initial tabulation) to easily allow for clause-level annotations, such as whether the clause comes from a dialogue or to distinguish main and dependent clauses (see below); special symbols used for DIAL and REP were placed in this column, so that the textual content of all clauses would start from the first.
3. 5 additional lines were added between each pair of subsequent clauses:
   - The first three of the newly added lines under each entry were filled word-by-word with analyses, glosses and translations provided by Moksha and Erzya morphological analysers[11], coalesced for legibility — this was meant to assist annotation by provision of quick reference and was discarded at the later stage and in no way interfered with the rest of the data.
   - The fourth line was meant for tags; column 0 was automatically fitted with appropriate DIAL or REP marking based on the symbol in the 0th column of the entry, otherwise the line was left empty.
   - The fifth line was left empty to provide easy optical separation of clauses.

5.2 Annotation

The files created as the output of the previous step could easily be copy-pasted into a spreadsheet editor (Figure 5.1). For some time, rather than a spreadsheet, I considered using a dedicated tool, such as FieldWorks by SIL[12], but decided that it would be likely to unnecessarily complicate the process, especially due to the fact that I was much more interested in syntactic than morphological glossing.

[10] Em-dashes (—) left inside the text ended up being treated as separate words due to their separation on both sides, but as I was not counting distances, the only downside of this is a very slightly overestimated token count in the corpus.

[11] Created by Timofey Arkhangelskiy and available as open-source Python packages: https://github.com/timarkh/uniparser-grammar-erzya; https://github.com/timarkh/uniparser-grammar-moksha

[12] https://software.sil.org/fieldworks/release-notes/

2959	.	Мекс	кадовить	ськамот?
2960		мекс	кадо-в-ить|кадо-в-и-ть	ськамо-т
2961		ADVPRO	V,pass,imp,2.s,sg.s,3.o,pl.o|V,pas	PRO,nom,sg,2sg|PRO,2sg
2962		зачем, отчего, почему	оставить	один, в одиночку
2963	DIAL			
2964				
2965		Тейтересь	мизолдозевсь.	
2966		тейтер-есь	мизолдо-зев-сь	
2967		N,anim,hum,sg,nom,def	V,inch,non_obj,3.s,sg.s,pst	
2968		девушка, дочь	улыбаться	
2969				
2970				

Figure 5.1. Two rows of the TP file seen in a spreadsheet editor.

The clauses in the files opened as spreadsheets were annotated by entering appropriate labels ('tags') in the dedicated 'tag' row, under the constituents they referred to. Appropriate tags were also added, where necessary, in 'column 0' of the tag rows, to mark different types of clauses. The two types of annotation, although appearing next to each other, used distinct sets of categories and corresponding markers. The respective strategies are described in the following subsections.

The tagging produced in this phase was not final and during the later stages received some revisions and refinement (as well as corrections); the description provided below refers to the final variant of the system. Note that a single clause or constituent could receive more than one label — for instance, a clause could be at the same time dependent, a part of a dialogue and a modal statement. This allowed flexibility and reduced the number of categories overall, but in the case of constituents encouraged overly creative (and not necessarily justified) mixing for the sake of generalisation, e.g. only later during the process did I realise that infinitive clauses should in most instances be treated separately, not as instances of nominal predication.

5.3 Clause types

The main distinction between narration (unmarked), dialogue and direct-speech-reporting clauses has been described in the section dedicated to the file preparation.

The second most important distinction, made already during the annotation phase, was that between main (unmarked) and dependent (DEP) clauses. Main clauses of content questions (Q) were additionally distinguished within the former type: dependent clauses are characterised by a relativiser or subordinator appearing

in a clause-initial position, and in content (WH-word) questions the main clause contains a question word typically occupying an initial or focused position.

Although in the presence of a dependent clause main clauses might be structured differently than otherwise to accommodate its presence, this has been assumed to be a less important factor than the clause itself being dependent. Hence, the position, or even presence, of dependent clauses was not marked in their main clauses. Finite dependent clauses in the function of a constituent within a matrix clause were also ignored; only if a correlative pronoun introducing the dependent clause was present, such as сянь in 43, this demonstrative was considered a constituent of the matrix clause and tagged appropriately.

(43) Moksha (KK1)
А сянь юкстаськ,
and DEM.GEN forget.PST1.1PL>3
O V
[*што минь попоньке сими-ярхцай сонь жа мархтонза!*]
'And this we forgot [that our priest feasts with no-one else but him!]'

Polar (yes/no) questions were not marked, with the exception of embedded polar questions (EMBED_Q) marked with a combination of the predicate and its negation ('polar question complements' in Klumpp 2016: 255-6), requiring the verb to appear clause-initially. There were three such clauses in the dataset, all in Erzya (44).

(44) a. Erzya (ECK)
[*Лейтенантось эзь кевкстне,*]
вечксы а вечксы цёрантъ.
love.3SG>3SG NEG love.3SG>3SG boy.GEN.DEF
'[The lieutenant did not ask] if she loved the boy.'

b. Erzya (TP)
[*лоткась ды кувать кунсолось*]
марязевить а марязевить кортамо вайгельть.
hear.PASS.3PL NEG hear.PASS.3PL speak.NMLZ voice.PL
'[he stopped and listened for a long time] if no sounds of talking could be heard.'

c. Erzya (TP)
[*Старшинась эзь карма ванномонзо,*]
маштовсь эли эзь
kill.PASS.PST1.3SG or NEG.PST1.3SG
'[The starshina did not start checking] if he had been killed'

The third distinction was based on the type of construction, in parallel to the categories listed above. Most of them, though not all, were marked explicitly with a 'clause-level' tag (details below). This was to ensure a reasonably 'pure' meaning of syntactic categories in the researched data. Mixing could lead to inaccurate results, see for instance how differently various types of subjects in Estonian interact with word order in Metslang's 2013 research. Although averaged results would not be completely invalid, investigating different types of constructions separately provides more fine-grained data and gives a more accurate overview of the structure of a language.

The clauses which were considered 'default' and not marked in any way included finite-verb (transitive and intransitive) clauses and clauses with nominal predication (without a finite verb other than the copula, including all types described by Haspelmath (2022) as predicational — also appertentive (Haspelmath 2022: 7-8) — as well as all types of duonominal clauses; more details later in this chapter, where the identification of the predicate is discussed).

The special types of clauses distinguished by annotation and omitted from all statistics and discussion presented in the next chapter (except for clauses with infinitive complements investigated as a distinct type of structure in section 6.8), were:

Impersonal clauses (IMPERS) — usually experiencer clauses or weather predicates without a semantically identifiable subject (45a). Sometimes, such clauses had simultaneously certain properties of other non-standard types, such as (45b), where a special non-finite form of the verb кирдемс 'to ask' is used in reference to the experiencer's state of mind (but not modal possibility or necessity, as with the standard infinitives in -мс or -мо/е/а), or (45c), a time-predicating clause which could be analysed as a subtype of either a classificational or existential (see Ajanki 2014 for discussion). Clauses like (45d), with an element classified by Kholodilova and Toldova (2018: 706) as an 'emotive modal predicate' were for the sake of brevity also classified as impersonal, rather due to their experiencer semantics than their syntactic properties.

(45) a. Erzya (ECK)
кармась тензэ марявомо, мерят весе рунгозо
start.PST1.3SG DAT.3SG feel.PASS.INF2 say.2SG all body.PX.3SG
ливти та-ков алов.
fly.3SG somewhere.LAT down.LAT
'he started to feel as if his whole body was falling somewhere down'

b. Moksha (PP)
а Орест Львовиченди аф кирдемшка ни
but Orest L'vovich NEG ask.INF.COMP already
'And Orest L'vovich does not feel like asking anymore'

c. Erzya (ECK)

Ульнесь сёксень чокшне.
be.PST.3SG autumn.GEN evening

'It was an autumn evening.'

d. Moksha (KK2)

Улема, визькс ули тяфтама инжить эзда!
seemingly shame be.3SG such guest.PX.2SG ABL

'One could think you're ashamed of such a guest of yours!'

Infinitive clauses (INF) — could be understood a subtype of IMPERS. Clauses with an -MC or -MV infinitive similar to nominal predication but semantically rather modal, expressing necessity (Bartens 1999: 144-152). Further subdivisions, probably along the infinitive type (as the two infinitives appear to have different semantics and a different set of applications) could be proposed but as these clauses were not an object of this study anyway, splitting them off from the rest was enough.

(46) a. Erzya (TP)

Миненек муемс костояк ютамо тарка
1PL.DAT find.INF1 somewhere.ELA cross.NMLZ place

'We need to find a crossing point somewhere'

b. Erzya (ECK)

Тонеть, паряк, лездама?
2SG.DAT apparently help.INF2

'I guess you could use some help?'

c. Moksha (PP)

И тейть сянкса аш месть пичечнемс=ка.
and DAT.2SG DEM.CAUS EX.NEG what.PL be.agitated.INF1=EMPH

'There is no reason to worry because of you, either.'

Existential clauses (EX) — usually marked by the presence of the existential copula or its negative variant (47a). In some cases it was not obvious whether a clause was existential or predlocative (Haspelmath 2022: 5-7), especially in absence of an existential copula. These clauses, not unlike 47b, were judged individually based on contextual and pragmatic clues, such as the 'reversed' word order and the indefiniteness of "немецть" in the given example. There seems to be a certain grey zone between impersonal, existential and infinitive clauses as the latter appear to use the existential negator, rather than the ones found in verbal in nominal predication (47c, compare 46c above) (Bartens 1999: 144).

(47) a. Moksha (KK1)
Улихть, брат, стама орёлхт=ка!
be.3PL brother such eagle.PL=EMPH
'There are even such eagles, my brother'

b. Erzya (TP)
тосо, нама, немецть
there.INE of.course German.PL
'There are Germans there, after all.'

c. Moksha (KK1)
Кенярткшнемс аш мезень лангс!
be.happy.INF1 EX.NEG what.GEN on.ALL
'There is nothing to be happy about!'

Possessive clauses (POSS) — a subtype of existential clauses, with a similar structure but expressing possession. The possessor number/person is marked on the possessum with the possessive declension and can optionally be present also as a free genitive NP. Some instances could be interpreted as semantically more existential than strictly possessive, and in some cases it was not easy to decide that a clause was not, in fact, predlocative (48c is an example of such ambiguity).

(48) a. Erzya (ECK)
Венченк ули?
boat.PX.2PL be.3SG
'Do you have a boat?'

b. Erzya (TP)
Сонзе кияк ули: авазо, сазорозо,
POSS.3SG someone be.3SG mother.PX.3SG younger.sister.PX.3SG
ялаксонзо?
younger.brother.PX.3SG
'Does he have anyone: a mother, a little sister, a little brother?'

c. Moksha (KK1)
Маряк, сонь седиенц вастса кельме кев.
hear.IMP.2SG POSS.3SG heart.GEN.PX.3SG place.INE cold stone
'Look, he has a cold rock instead of a heart.'

Clauses with infinitive complements — featuring verbs such as капшамс 'hurry', эрявомс 'be needed', савомс 'be necessary', (E) ушодомс / (M) ушедомс 'begin'. This category was not marked explicitly at the clause level but rather by tagging their complement infinitives as ARG_V. While in most cases objects in infinitive complements retain their status and connection to their head (nominative objects of finite verbs become genitive objects of the corresponding infinitives, see

Bartens 1999: 91), morphologically passive modal verbs эрявомс and савомс raise objects of their infinitive arguments to subjects. The latter behaviour was the main reason for me to avoid too early assumptions about the role of the verbal infinitive and treat is simply as a “complement". Complements with adverbial reading, usually expressing purpose (50), were not included in this category.

(49) a. Moksha (PP)
аф кармат нужань няендема
NEG start.2SG poverty.GEN see.FREQ.INF2
‘You will not be seeing poverty!’

b. Erzya (TP)
Ютамс эрявсь ансяк те юткось.
cross need.PASS.PST1.3SG only DEM distance.DEF.NOM
‘Only this stretch was left to cross.’

(50) Moksha (PP)
и сон озась малазост курендама
and 3SG sit.down.3SG near.ALL.PX.3PL smoke.INF2
‘and he sat next to them to smoke’

Causative clauses with subject-to-object raising (CAUS) — similar to the previous type but due to their nature allowing the presence of two different objects: one of the infinitive, and one of the main verb. Although there were several clauses of this type, most of them only had one explicitly expressed direct object, whether due to intransitivity of the verb in the complement (51a), to use of object marking on the finite verb making no need for an overt pronoun or due to raising of the subject of the infinitive to an indirect, rather than direct object (51b). Example (51c) was the only one in the corpus with two overt direct objects. They were not included in statistics on clauses with infinitive complements due to the structural differences between this type and the simpler one mentioned above.

(51) a. Erzya (ECK)
Николай Петрович тердинзе инжетнень
Nikolai Petrovich invite.PST1.3SG>3PL guest.GEN.PL.DEF
ойсеме столь экшка.
take.seat.INF2 table at.PROL

‘Nikolai Petrovich invited the guests to take seat at the table.’

b. Erzya (ECK)
Генералось кармавтсь весененъ срадомс
General.NOM.DEF start.CAUS.PST1.3SG all.DAT disperse.INF1
аламонь-аламонь.
few.DISTR

'The general made everyone disperse into small groups.'

c. Moksha (PP)

Орест Львович ве лангс кучезе Ареф
Orest L'vovich night on.ALL send.PST1.3SG>3SG Aref
атянь велетненъ ланга вирень керыень
old.man.GEN village.GEN.PL.DEF on forest.GEN cut.PTCP.GEN
кочкама
gather.INF2

'At night, Orest L'vovich sent the old Aref to gather woodcutters from the villages.'

Desiderative clauses with the word мель (MEL) — characterised by the presence of the word (E) мель / (M) мяль 'will' as a constituent, with an -мс infinitive which often developed into an entire non-finite clause. There is a certain degree of similarity to possessive clauses (52a) here, although the ability of the word мель to accept the clitic copula in this construction (52b) raises questions about the relatedness of these structures. Constructions with certain verbs other than the copula ул(ьн)емс 'be' following the same general pattern were also included here, such as 52c with the verb самс 'come' expressing a sudden emergence of the wish. The word *mɛl'* itself, as well as its combinations with verbs were classified by Kholodilova and Toldova (2018: 704) as types of emotive predicates, making them similar to the можна-type clauses mentioned below.

(52) a. Moksha (PP)
Андрейть мялец самс кудалама!
Andrei.GEN.DEF will.PX.3SG come.INF1 court.INF2
'Andrei wants to come courting!'

b. Moksha (KK2)
мялец=оль шнафтомс пря Фёдортъ инголе
will.PX.3SG=PST.3SG praise.INF1 head Fyodor.GEN.DEF in.front.of
'He was keen on boasting in front of Fyodor'

c. Erzya (ECK)
Сась тензэ мель аламос аштемс съкамонзо,
some.PST1.3SG DAT.3SG will little.ALL be.located.INF1 alone.3SG
сэтьместэ, оймазь.
quiet.ADV calm.down.CVB
'He felt like being alone for a while, in peace, calming down.'

Modal clauses with the predicative modal word можна 'one may' (MOZNA) — an interesting blend of infinitival and modal verb clauses, where the

uninflected (except for accepting bound copula, as seen in 53) word можна, rather than a verb, serves as a predicate governing the -мс infinitive of the main verb. These were few in number, and were only found in Moksha texts, nevertheless they were considered specific enough not to group them with any other type of clause (a 'modal predicate' in Kholodilova and Toldova (2018: 703), one of its most important differences from the one encountered in the MEL type is that it raises the object of the infinitive to its subject, rather than the subject of the infinitive to its possessor; therefore, they would not be compatible if I wanted to collect data on their order of constituents).

(53) Moksha (PP)

синь можна=т мимс и рамамс танцти суском лангс,
3PL may.be-PL sell.INF1 and buy.INF1 tasty bite on.ALL
ярмакта!
money.ABL

'one can sell and buy them with a tasty treat, with money! '

5.4 Classification of constituents

This section describes the types of constituents annotated and investigated in this study: verbs, subjects, objects, indirect objects/experiencers, nominal predicates, negators, adverbials, dependent infinitives, etc.

Not every word was tagged as not all of them fall into a category relevant in this study — adjectives and demonstratives within noun phrases, dependent nouns in postpositional constructions, loose particles and interjections, etc. were ignored (even though presence of such constituents may have an influence on constituent ordering, this was not researched in this study). Within a phrasal constituent, typically only the head (though see for instance the paragraph on nominal predication) was tagged. When a phrase was discontinuous, the head was its only tagged element and all cases of other relevant constituents intervening were appropriately marked. Unlike Vilkuna (1998), I do not consider cases of possessors appearing separate from the possessees as discontinuous: rather, the possessor in these phrases seems to function as an indirect object and was accordingly marked as both POSS and O'; see also the paragraph on indirect objects later in this section.

The exact position of the tags was not really important as distances were not measured; only their order mattered, indicating the relative position of the constituent phrases they represented. Parentheticals, usually separated with punctuation, were ignored. Sometimes, when two constituents of the same type (other than a finite verb, e.g. objects or non-finite verbal arguments) appeared in the same clause, and it was necessary that only one constituent of a given type is tagged (this was true for all constituents other than the various types of adjuncts),

the more developed one was chosen — for instance, a non-finite purpose expression with an object instead of one without an object — but such occurrences were rather rare.

Categories annotated and used in this study (provisionally annotated but later unreferenced categories are not described here) included:

Verb (V): a finite verb, transitive 54a or intransitive 54b. The tense or mood did not matter at this point. In clauses with a negator (whether particle or auxiliary), the semantic part (connegative) was labelled as the verb (see example 59 below). As written earlier, each clause can contain at most one such verb. The free copula улемс was treated as a verb (54c) but its negations ((E) аволь, арась / (M) аф, аш and variants) were not. Only free constituents were annotated, so the bound copula was not eligible for marking as a verb.

(54) a. Moksha (PP)

Орест Львович тянь приметазе
Orest L'vovich DEM.GEN notice.PST1.3SG>3SG
S O V
'Orest L'vovich noticed this'

b. Erzya (ECK)

Николай Петрович удось вельть берянькстэ.
Nikolai Petrovich sleep.PST1.3SG quite bad.ADV
S V A
'Nikolai Petrovich slept quite badly'

c. Erzya (ECK)

Сёрмась ульнесь ава ломаненъ пельде.
letter.NOM.DEF be.PST1.3SG woman person.GEN side.ABL
S V SRC, NPR
'The letter was from a woman.'

Subject (S): the subject, defined according to the type of sentence it appears in. Understood as typically a nominative, usually definite NP with which a finite verb agrees, or the semantic subject, the actor of the verb (see 54 for typical examples). In absence of a finite verb, the subject is identified on the basis of such features as being a noun-headed NP in nominative, more rarely a quantitative phrase with an ablative, high on the scale of topicality and definiteness — as opposed to the predicate, which is additionally often indicated by the bound copula. In clauses predicating quantity, such as (55a), I treated the quantifier as the predicate and the quantified entity as the subject; if not for the presence of the bound copula, this could be, however, considered an instance of an existential clause (the availability of bound copula usage is a diagnostic for predicate constructions, as opposed to existential; see Turunen 2009: 253).

A non-finite verb, or even a clause, could be considered a subject if it appeared in a construction, and in a (usually gerund-like) role which appropriately strongly supported such a qualification. For instance, in (55b) the presence of bound copula and a complete lack of necessitative meaning typical of infinitive clauses warrants treatment of меленцяв молемс 'mill.LAT go.INF1' as a nominalised phrase rather than as a free infinitive. Similarly, a word typically functioning as an attribute could stand alone as a subject if it was used in a nominalised meaning, particularly with definite declension marking — such as тнярось 'this.much.NOM.DEF' in (55c).

(55) a. Moksha (PP)

Аф кржа=ль ни эряфсонза тяфтама сускомда.
NEG few=PST1.3SG already life.INE.PX.3SG such bite.ABL
N NPR LOC S
'There have already been not too few such bites in his life.'

b. Moksha (KK2)

Меленцяв молемс тякос=ель прясонза.
mill.LAT go.INF1 only.NOM.DEF=PST1.3SG head.LOC.PX.3SG
S NPR LOC
'Going to the mill was the only thing on his mind.'

c. Moksha (KK2)

И кодама прорвас тнярось тяльги?
and what.like abyss.ALL this.much.NOM.DEF fit.3SG
Q, G S V
'And by what kind of abyss could this much be contained?'

Object (O): a direct object of a transitive verb. It can be unmarked or marked with genitive, optionally ablative, inessive or a postpositional phrase expressing one of these cases (56). Its person/number can be indexed on the verb by means of objective (definite) conjugation. With the exception of ablative and inessive (when clearly used for object marking, i.e. next to a transitive verb and expressed with a postpositional phrase), constituents marked with spatial or semantic cases were considered indirect objects or adverbials, rather than simply objects. As opposed to the subject treatment above, infinitives, along with their clauses, were treated as special arguments of the finite verbs with which they appear, rather than simply as clausal objects.

(56) a. Moksha (KK1)

Ладна, норада, панчсайне ортатненъ!..
okay wait.IMP.2PL open.1SG>3PL gate.GEN.PL.DEF
V O
'Alright, you just wait, I'll open the gate!'

b. Moksha (PP)
Алёна баба эвондсь тянь=ге эзда.
Alyona old.woman take.fright.PST1 DEM.GEN=EMPH ABL
S V O
'That, too, frightened the old Alyona.'

c. Erzya (ECK)
Кудосо эйсэнзэ учость тевть, мелявкст.
home.INE INE.3SG wait.PST1.3PL matter.PL worry.PL
LOC O V S
'Back home, all kinds of issues and worries were waiting for him.'

Note that intransitive verbs of e.g. movement may appear together with objects such as expressions of time or distance — these verbs are in those instances considered transitive, which can be shown by their ability to take on objective conjugation (57a, 57b). Expressions of time or distance marked similarly to objects can also be used with transitive verbs (57c), in such situations the actual objects would take precedence in marking.

(57) a. Erzya (TP)
Те ютконтъ сынь молизь ашолгадомс.
DEM stretch.GEN 3PL go.PST1.3PL>3 dawn.INF.ALL
O S V T
'They passed (lit. went) that stretch by dawn.'

b. Erzya (ECK)
Николай Петрович удось пель чи ды ве.
Nikolai Petrovich sleep.PST1.3SG half day and night
S V O
'Nikolai Petrovich spent half a day and a night asleep.'

c. Moksha (Kholodilova and Toldova 2018: 558)
vas'ɛ vet'ə minutat mar't' esə anas'
Vasya five minute.PL apple.GEN.DEF INE ask.for.PST1.3SG
S T O V
'Vasya kept asking for the apple for five minutes'

Indirect object (O'): usually expressed with dative or dative-allative (the latter in the definite declension); beneficients/maleficients expressed e.g. with postpositions пачка 'for' or инкса 'for the sake of' were also treated as members of this category (58a). External possessors, usually in clauses expressing uncontrolled events (58b), though not exclusively (58c) were treated as a kind of indirect object, but additionally marked with a special POSS tag.

(58) a. Moksha (KK1)

А кинь инкса мон тяфта старандан?
but who.GEN for 1SG like.this make.effort.1SG
O' S A V
'And for whose sake am I exerting myself so much?'

b. Moksha (KK1)

Кинь-бди мчкоргодсть пейнза.
someone.GEN be.gnashing.PST1.3PL tooth.NOM.PL.PX.3SG
O', POSS V S
'Somebody's teeth gnashed'

c. Erzya (ECK)

Минек ранязентъ керизъ омбоце
POSS.1PL wound.PTCP.GEN.DEF cut.PST1.3PL>3 second
O', POSS V
пильгензэ...
leg.GEN.PX.3SG
O
'They cut the other leg of our wounded...'

Verbal argument (ARG_V): a non-finite verb appearing as an argument of another verb (modal or controlling). Theoretically, these could be chained so this marking is limited to arguments of finite verbs.

Negator (N): negation verb (in the present dataset always immediately preceding the main verb) or negation particle, tagged rather to denote a clause as a whole as negative than to track the position of the negator itself. Negations were not tracked in clauses with the NPR marking.

Both of the last categories are represented in example (59).

(59) Moksha (PP)

аф кармат нужань няендема!
NEG start.2SG poverty.GEN see.FREQ.INF2
N V O ARG_V
'you will not suffer poverty!'

Nominal predicate (NPR): in a generalised sense, applied both to non-verbal predicates and to duonominal clauses treated by Haspelmath (2022) as non-predicational. It could be an adjective, a locative phrase, an indefinite noun (sometimes marked with the translative case) in classificational and definite noun in identificational clauses. Any of these could bear the predicative conjugation marking. In absence of any case/copula marking, the (non-topical) carrier of new information was marked as NPR. A further distinction of different kinds of clauses

could be made but this would likely require a larger dataset and was considered as beyond the scope of this work.

Clauses of the mixed classificational-attributive type, where a definite noun broadly categorising the subject is followed (not necessarily immediately) by its seemingly dependent adjective which conveys the new information, were common occurrence (example 60; see also the section on nominal predication in the theoretical chapter), appearing in both languages.

(60) a. Erzya (ECK)

Сонсь *Николай* *Петрович* *ломанесь* *аволь*
3SG.EMPH Nikolai Petrovich person.NOM.DEF NEG
S

покшкавтниця
be.pretentious.APRT
NPR

'Himself, Nikolai Petrovich is not a haughty person'

b. Moksha (KK2)

стирсь *сон* *цебярь*
girl.NOM.DEF 3SG pretty/good
S NPR

'she's a pretty girl'

Adverbial (T, INS, LOC, SRC, G, CAUSE, PURP, A): different types of adverbials (time, instrument, location, source, goal, cause, purpose, means) were distinguished in case that the differences between them would become an object of study at a later point.

Interrogative pronoun (Q): WH-words in content interrogative clauses.

The interrogative pronoun месть/мезть 'what.PL' is the most intriguing member of this category: it often appears to be a (nominative plural) object, but usually of a transitive zero verb. It typically represents the reasoning or purpose of an action, and perhaps a shade of irritation on the speaker's side (reminds one of Russian чего 'what.GEN', Polish co/czego 'what/what.GEN', Estonian mis 'what' in similar contexts). It was considered a nominal predicate in clauses without a verb, such as (61a) and (61b), although one could perhaps argue that they are a distinct type of clauses, or even not clauses at all. With a verb, месть/мезть was marked as an adverbial or an object. Мезть is frequently used in collocation with the verb кортамс in Erzya serving as its apparent object, and in (61c) the pronoun was accordingly tagged as the object of кортат. On the other hand, in (61d) мезть is clearly not connected syntactically with the intransitive verb аштетядо and so it was marked as an adverbial of purpose.

Note also that at least Moksha has the verb месендемс ‘do.what’ (Serebrennikov et al. 1998), which can be used as an interrogative or relative verb (see example 62d below). As its usage precludes the presence of a direct object in the same clause, its occurrences were excluded from all statistics where the position of the object is relevant (the same holds for both interrogative and dependent clauses featuring this verb).

(61) a. Moksha (PP)

Уляша, тон месть?!
Ulyasha 2SG what.PL
S NPR

‘Ulyasha, what on Earth are you doing?!’

b. Erzya (ECK)

Тон мезть?
2SG what.PL
S NPR

‘What’s up with you? What’s going on?’

c. Erzya (ECK)

Мезть кортат?
Ulyasha 2SG what.PL
O V

‘What are you talking about?’

d. Erzya (ECK)

Тынь мезть тесэ аштетядо?
2PL what.PL here.INE stand.2PL
S PURP LOC V

‘What are you standing here for?’

Subordinator (REL): in this work a generalised category including complementisers, adverbialisers and relative pronouns, in most cases located at the very beginning, or within the first constituent, of a dependent clause (62a). If the subordinator was a dependent in an NP, the tag was deferred to the last constituent (usually the head) within the phrase (62b). Often, it appeared accompanied by other tags, such as S, O or NPR. In very few cases more than one subordinator appeared within a single clause (62c). An interesting member of this class is the verb месендемс ‘do.what’ mentioned in the previous paragraph (62d).

(62) a. Moksha (KK1)

[*И озай Кирей Михайлович,*] ***коза*** *максыхть вастоня*

where.ALL give.3PL place.DIM

REL, G V O

'[And so Kirei Mikhailovich sits down] where they give him a place.'

b. Erzya (ECK)

[*Неят,*] ***зняро*** *кавалерт састь.*

how.many suitor.PL come.PST1.3PL

REL, S V

'[You see,] how many suitors have come!'

c. Erzya (ECK)

[*Эрьвась нуртнесь сонсь ды симсь,*]

кинень *зняро эряви.*

who.DAT how.much be.needed.3SG

REL, O' REL, S V

'[Everyone poured [the drinks] [for] themselves and drank] to their hearts' content '

d. Moksha (PP)

Вона ***месендеть*** *Наташать мархта*

lo! do.what.PST1.2SG Natasha.GEN.DEF with

REL, V O'

'Look what you did to Natasha!'

5.5 Morphological annotation

When the syntactic annotation was complete, the extra lines used for human readability (three lines of automatic glosses and empty lines between clauses) were dropped with the help of another script, leaving only the clauses and their respective sets of tags. Then came time for the additional annotation of morphology of verbs and objects, which was necessary for some elements of the research (morphological parameters important in this work have been described in section 2.2).

Unfortunately, while the analysers were undoubtedly helpful in the task of syntactic annotation, they are far from perfect and could not be relied upon. As one reason for this, the ambiguities within the paradigms (and possibly some cases of ambiguity between derivational and declensional/conjugational suffixes) can be cited. Knowledge of the context is therefore necessary for disambiguation. What's worse, the parsing mechanism itself was unreliable, often failing to recognise a word altogether (this seemed to be especially the case with the Moksha analyser, but it

can also be a result of Levchaev's use of uncommon vocabulary or a higher number of morphophonological alternations in Moksha than in Erzya). As a consequence, not only was glossing of at least a part of the verbs by hand unavoidable, but in addition I could not have full certainty that the analysers would not miss certain analyses in those verbs that they did find a match for.

For these annotations, a second 'tag' line was created. In the case of verbs, tags were added to the constituents earlier tagged as V, even if it was the negation verb that bore the TAM and agreement markers within the clause (this would make searching for relevant tokens in the corpus at the later stage easier). For instance, two clauses ("Эзить кенере кудов само, мельгат ливтить сёрмат" 'You did not hurry back home, letters are flying after you!') previously represented in the following way:

.	Эзить	кенере	кудов	само,
DIAL	N	V	ARG_V	
.	мельгат	ливтить	сёрмат.	
DIAL	G	V	S	

...looked like this after the annotation:

.	Эзить	кенере	кудов	само,
DIAL	N	V	ARG_V	
		PST, SUBJ, S2SG		
.	мельгат	ливтить	сёрмат.	
DIAL	G	V	S	
		PRES, SUBJ, S3PL		

Notice that some suffixes, such as (E) -ить PST.2SG and PRES.3PL, had to be disambiguated based on contextual clues; although this particular pair did not have much influence on the results of this research as both belong to the subjective paradigm and this is all we need to know here, one should be aware that the same suffix is also used for PST.2SG>3PL (for the notation, see e.g. example 1). Although the aforementioned example only holds for Erzya, there are still instances of similar ambiguity in Moksha, too, only with different groups of suffixes.

Objects underwent a similar process of morphological annotaiton, being divided into pronouns and nouns (postpositional phrases were divided based on the type of their dependent NP), and further into indefinite (unmarked), definite and possessed. Additionally, the case form was marked for all objects. Possessed objects were understood to be in genitive, even if this cannot be marked with morphology for most combinations of possessors and possessees in Erzya (Hamari and Ajanki 2022: 402-3). It should also be noted that some forms were found in the present

which did not follow the standard paradigms, such as турвам in (63a), which would have an expected singular nominative interpretation, rather than genitive of any number (perhaps it was simply a typo). Object tags use the same line as those assigned to verbs, so the structure of the files did not change with this step. The reflexive pronoun пря, homonymous with the word for 'head', was tagged as a noun or a pronoun, depending on its meaning in the given clause (e.g. in the context of example 63b it is known that the presence of Kirei Mikhailovich was not limited to his head, and a literal 'head' would be unlikely to appear as an object without a possessive suffix).

(63) a. Erzya (TP)

Монь кондямо виев шайтаносъ, истя пек
POSS.1SG similar strong devil.NOM.DEF so much
лепштинзе турвам, натой керявсть
press.PST1.3SG>3PL lips.NOM.SG?.PX.1SG even cut.PASS.PST1.3PL
пейтненень.
tooth.ALL.PL.DEF

'The fiend was as strong as me, pressed my lips so much they even got cut against the teeth.'

b. Moksha (KK2)

Кирей Михайлович няфтсь пря и тага
Kirei Mikhailovich show.PST1.3SG head and again
юмафтсь.
disappear.PST1.3SG

'Kirei Mikhailovich showed up and disappeared again.'

5.6 Setting up a searchable corpus

With all the annotations in place, it was time to start browsing data for different patterns and looking for similarities and differences. This could be done with manually written script queries, however, it seemed more convenient to use an existing browning mechanism in the form of tsakorpus[13] developed by Timofey Arkhangelskiy and supporting the "Corpora of Uralic Volga-Kama languages"[14].

The files obtained by the process described above were translated into the corpus format described in Tsakorpus' documentation[15] with a custom script, and indexed on a local installation of the engine. Each clause was indexed as a separate sentence.

[13]https://github.com/timarkh/tsakorpus
[14]http://volgakama.web-corpora.net/index_en.html
[15]https://tsakorpus.readthedocs.io/en/latest/

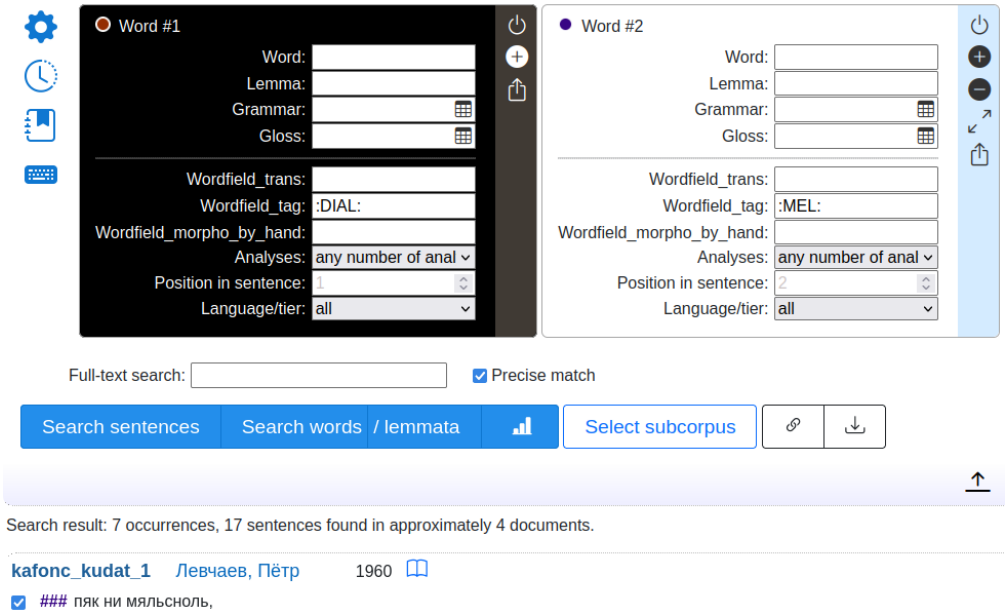

Figure 5.2. Results of search for dialogue clauses (DIAL) with a desiderative (MEL) construction. Note the erratic counting of occurrences.

I configured the corpus to allow for syntactic and morphological annotations as additional word fields. As it turned out that the queries based on additional wordfields work slightly unexpectedly (looking for fitting substrings, not necessarily full words as is the case with the main fields), all tags added through annotation were supplied with special guard characters to avoid accidental collisions (e.g. verbs and verbal arguments were represented in the corpus as :V: and :ARG_V: respectively, rather than simply V and ARG_V, so that instances of ARG_V would not be returned when searching for V).

The clause-level tags, stored in the zero-column, were represented in the corpus as tags on a unique clause-starter lexeme, appearing at the beginning of every corpus entry. This is not a clean solution from the design point of view, but it was easy, effective and served its purpose.

Besides the peculiar treatment of extra word fields, the query mechanism sometimes worked also in other unexpected ways. Negative query items behave especially strangely, e.g. a negative query forbidding that the first item within a clause (in this case, it is the special clause-starter token) have a particular type of tag cannot be combined with a positive query referring explicitly to the same element (Figure 5.2–5.4). Instead, a positive query had to be made. Also, the occurrence counting is sometimes bugged with negative query items, however, the "sentence" count is the relevant number of hits for us — the structures we are after

appear at most once in any clause, and if a clause is returned by the engine, it must contain the relevant structure.

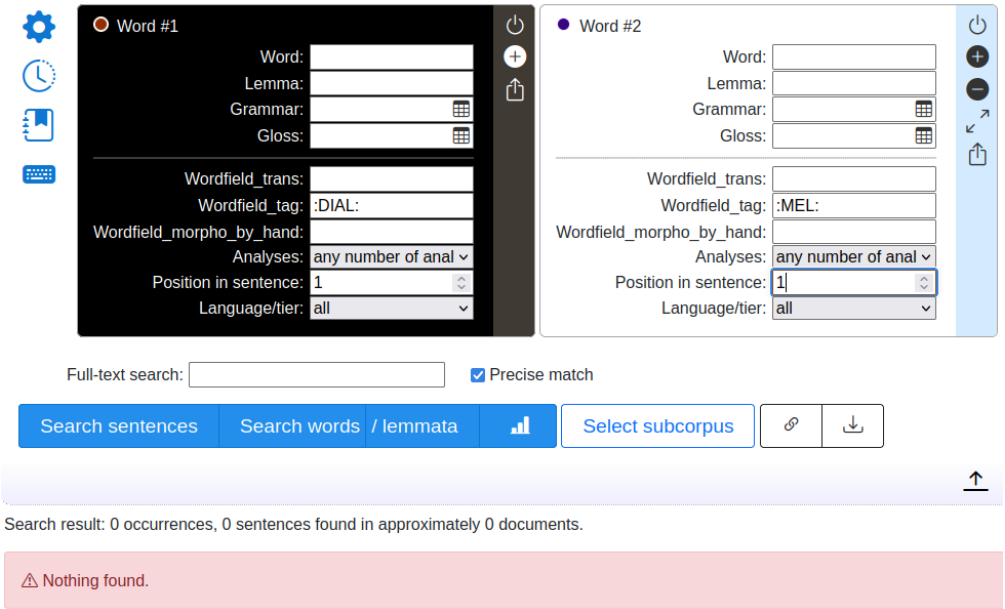

Figure 5.3. The same query as in Figure 5.2, but now looking for the MEL indicator only on the first token of the clause. Technically it should work exactly the same as in Figure 5.2, as MEL is a clause-level tag and can only appear on the clause-starter token, but instead it returns no results.

When searching for occurrences of a particular word order, the "distance" parameter was used, so that each consecutive constituent would be at a positive distance from its predecessor, usually set to "from 1 to 100" to catch all viable instances — the number 100, much greater than the length of any clause in the corpus, functions as an "infinity" (Figure 5.5). The "Select subcorpus" tool was used to refer in each query to an appropriate part of the data, usually a single text or a set of texts in a single language. Over the course of the research, results of the queries were reviewed and the source documents appropriately corrected and reindexed for maximum correctness and accuracy.

5.7 Conclusion

As can be seen based on this chapter, the preparation of the present corpus, along with establishing a working strategy for its use was a complex and demanding process, not the least due to my prior lack of experience in this area. Not all of the choices made on the way to its completion were optimal but for all there is at least some sort of justification.

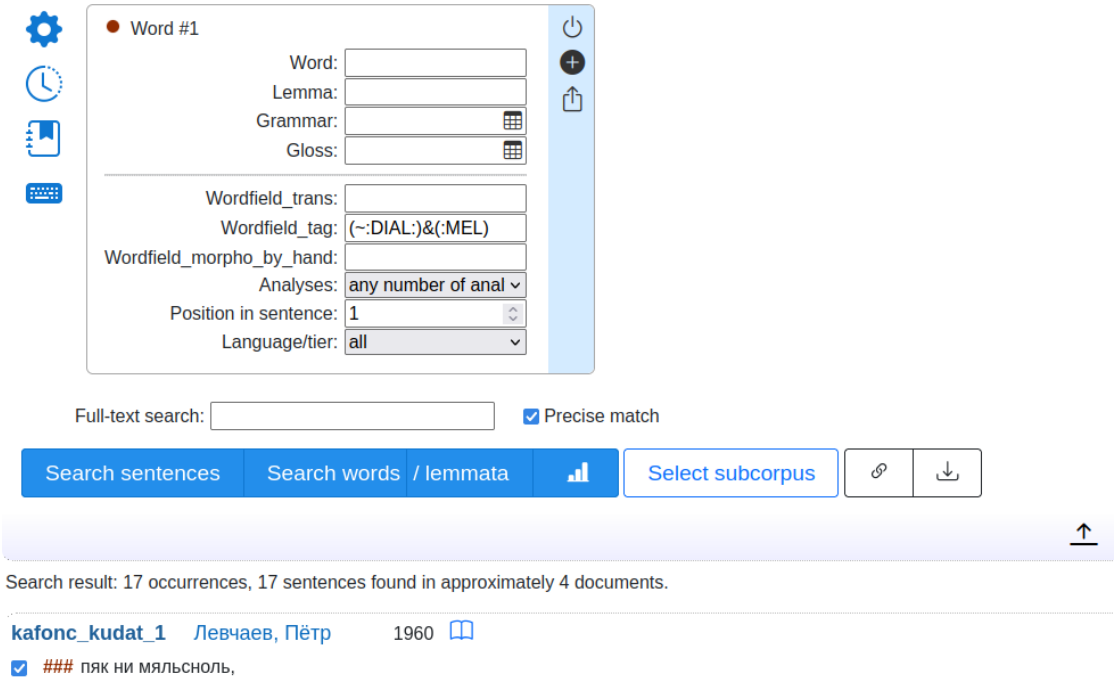

Figure 5.4. The same query as in Figure 5.2 and 5.3, but reformulated as a single-item query.

The inability to use automatic analysers for tagging was underwhelming but they still fulfilled well the purpose they were most likely created for: they did allow me to parse the text quicker and with greater ease than if I had to consult a dictionary every time I found several unknown words in a row. This is the function they also fulfil in Timofey Arkhangelskiy's corpora of Erzya and Moksha (Arkhangelskiy 2019: 132-4); and just like their advantages, their imperfections impact those corpora, too — making them effective tools in qualitative research or when simply searching for examples of language use, but not immediately useful for collection of statistics.

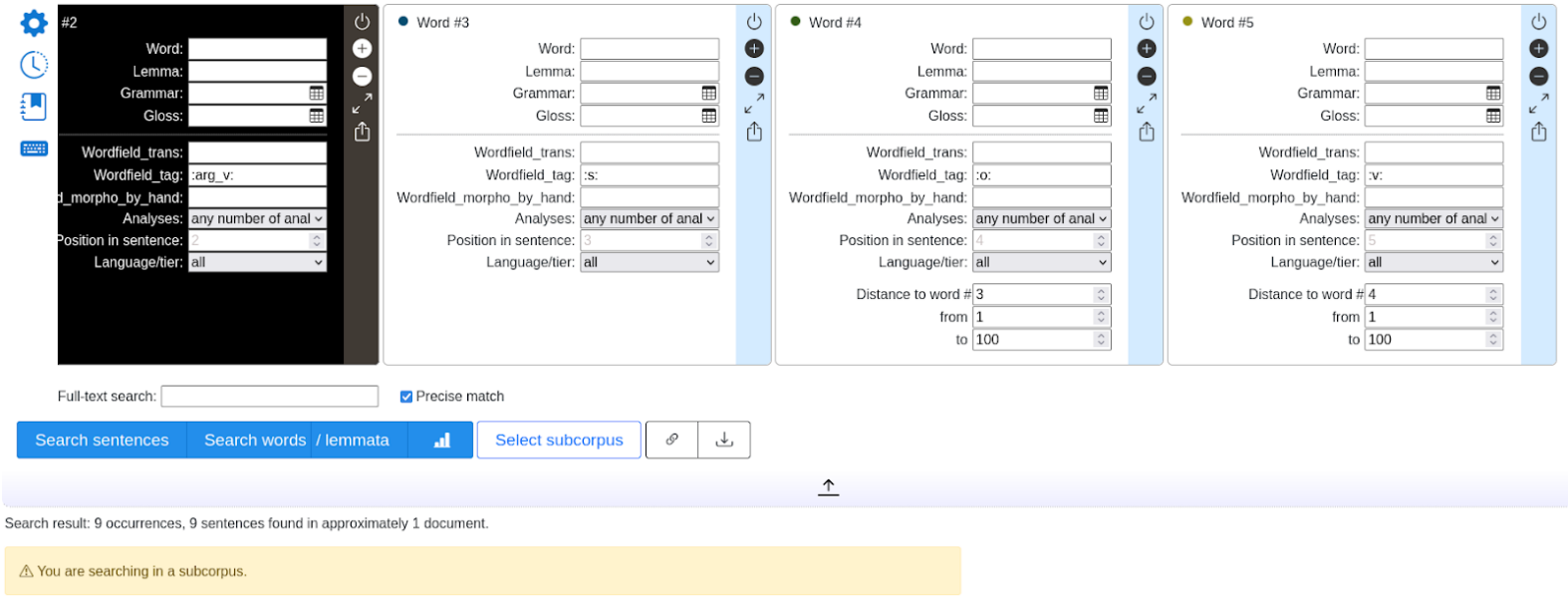

Figure 5.5. A query looking for instances of SOV word order, excluding clauses with infinitival complements. Actual queries used during the study contained some additional elements excluding undesirable types of clauses from search results.

6 The corpus study

The main objective of this study is the collection of quantitative data regarding ordering of constituents in some common syntactic structures in Mordvin languages, identification of likely factors affecting the ordering, and identification of possible areas where the ordering strategies may differ between the languages. This chapter describes in detail the outcomes of this research, comparing the results with earlier authors where possible. The strategies used in category assignment and tools used during the process of preparation of corpus and its subsequent browsing have been described in the previous chapter.

In this study, additionally to the statistics of ordering of overt S/A and O arguments in relation to the verb, some likely contributing or at least correlating factors have been investigated. In some cases there was a precedent in the form of earlier quantitative research (Vilkuna 1998; Kangastus 2012; Kholodilova and Toldova 2018). These results are compared with the new ones from the present corpus.

Other features, such as the influence of verb polarity on the order of constituents in a transitive clause, the position of an indirect object in a ditransitive clause or the order of constituents in non-verbal predicative/identificational clauses have not been approached quantitatively before, at least not in the available literature. They are a contribution to the research of constituent order in Mordvin, and their advantage is in having data collected simultaneously from both Moksha and Erzya side by side, allowing for their comparison.

Care was put into separation or elimination of possible sources of data noise. This was achieved by two means:

1. Exclusion of special clause types (dependent clauses, questions, modal clauses, clauses with modal/aspectual auxiliaries; for more details see the previous chapter) in the majority of study cases discussed here.

2. Separation of different storytelling styles (narration and dialogue), which, as can be seen below, produced two subsets.

Although the exclusions reduce the size of the available dataset, they allow for a more accurate observation of the trends appearing in Erzya and Moksha than e.g. the across the board approach taken by Vilkuna (1998).

Where the data size allowed for it, each of the four texts was considered separately, to better demonstrate the variety within each language (or, to be more exact, within each author's language usage) and compare it to the differences between Erzya and Moksha themselves. Indeed, in several instances the difference between two texts written by one author turned out to be greater than that between the two authors, exemplifying the difficulties inherent to this kind of research and

justifying the caution in drawing definite conclusions. Where possible, I tried to account for the differences on the basis of the thematic and stylistic peculiarities of their source texts.

All percentages given here were rounded to one decimal point for a balance of accuracy and readability — given that no sample examined here is larger than 500 entries, and most are much smaller, any differences beyond that point are irrelevant in comparison with the error margin generated by possible inaccuracies in categorisation or annotation. Where not said otherwise, the results are based on narration (not dialogue or explanatory phrases of the VS type) and exclude all the special subtypes of clauses listed in the previous chapter (content questions, dependent clauses, modal clauses, clauses with modal/aspectual verbs, etc.) and clauses with imperative verbs, including idiomatic phrases and expressions such as секень вант 'I tell you it is going to happen' (lit. 'watch this', cf. Russian того и гляди 'id.;).

6.1 Transitive clauses

The first result presented here is the distribution of constituents in non-dependent transitive clauses in the narration part of the texts in the corpus. As the total number of such clauses was well over 100 in three of the four texts, and not much below that number in the remaining one (TP), the statistics for each text are shown separately.

Table 7. Order of constituents in transitive clauses.

	Erzya						Moksha					
	ECK		TP		Total		KK		PP		Total	
VO	**101**	**77.7%**	**69**	**81.2%**	**170**	**79.1%**	**131**	**79.9%**	**128**	**73.6%**	**259**	**76.6%**
pure VO	50	38.5%	39	45.9%	89	41.4%	87	53.0%	82	47.1%	169	50.0%
SVO	49	37.7%	30	35.3%	79	36.7%	40	24.4%	45	25.9%	85	25.1%
VSO	1	0.8%	0	0.0%	1	0.5%	2	1.2%	1	0.6%	3	0.9%
VOS	1	0.8%	0	0.0%	1	0.5%	2	1.2%	0	0.0%	2	0.6%
OV	**29**	**22.3%**	**16**	**18.8%**	**45**	**20.9%**	**33**	**20.1%**	**46**	**26.4%**	**79**	**23.4%**
pure OV	14	10.8%	10	11.8%	24	11.2%	16	9.8%	32	18.4%	48	14.2%
SOV	10	7.7%	3	3.5%	13	6.0%	13	7.9%	9	5.2%	22	6.5%
OSV	1	0.8%	2	2.4%	3	1.4%	2	1.2%	1	0.6%	3	0.9%
OVS	4	3.1%	1	1.2%	5	2.3%	2	1.2%	4	2.3%	6	1.8%
Total	**130**		**85**		**215**		**164**		**174**		**338**	

Based on the data presented above, it appears that non-initial subject placement is a marginal phenomenon in both Erzya and Moksha. The total average percentage of occurrences of the VO order is 79.1% in Erzya and 76.7% in Moksha texts — the difference seems not too significant in comparison with the variation between the texts themselves, although it is interesting that both Erzya texts feature a

higher VO occurrence rate than the average in Levchaev's Moksha ones. Still, at least within the realm of narration, a much larger corpus would be necessary to confirm whether there is any meaningful difference in distribution between the two languages.

A very similar result is obtained when only clauses without an overt subject are considered (Table 8).

Table 8. Order of constituents in transitive clauses without an overt subject.

	Erzya					
	ECK		TP		Total	
pure VO	50	78.1%	39	79.6%	89	78.8%
pure OV	14	21.9%	10	20.4%	24	21.2%
Total	**64**		**49**		**113**	

	Moksha					
	KK		PP		Total	
pure VO	87	84.5%	82	71.9%	169	77.9%
pure OV	16	15.5%	32	28.1%	48	22.1%
Total	**103**		**114**		**217**	

Now, let us see the same type of data based on dialogue for comparison (Table 9).

Table 9. Order of constituents in transitive clauses within dialogue.

	Erzya						Moksha					
	ECK		TP		Total		KK		PP		Total	
VO	**28**	**66.7%**	**24**	**72.7%**	**52**	**69.3%**	**38**	**45.8%**	**32**	**60.4%**	**70**	**51.5%**
pure VO	21	50.0%	15	45.5%	36	48.0%	32	38.6%	20	37.7%	52	38.2%
SVO	7	16.7%	9	27.3%	16	21.3%	6	7.2%	12	22.6%	18	13.2%
VSO	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
VOS	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
OV	**14**	**33.3%**	**9**	**27.3%**	**23**	**30.7%**	**45**	**54.2%**	**21**	**39.6%**	**66**	**48.5%**
pure OV	9	21.4%	7	21.2%	16	21.3%	30	36.1%	10	18.9%	40	29.4%
SOV	2	4.8%	2	6.1%	4	5.3%	13	15.7%	9	17.0%	22	16.2%
OSV	1	2.4%	0	0.0%	1	1.3%	1	1.2%	1	1.9%	2	1.5%
OVS	2	4.8%	0	0.0%	2	2.7%	1	1.2%	1	1.9%	2	1.5%
Total	**42**		**33**		**75**		**83**		**53**		**136**	

The sample is, unfortunately, smaller than the previous one, but still large enough for some observations, especially if combined data is taken into account for each language.

First of all, there is a marked difference in the frequency of usage of VO versus OV order between Erzyan and Mokshan texts: while the Erzyan dialogue differs

in this regard from the narrative only slightly (even if noticeably), in Mokshan texts we see an almost equally frequent usage of both orders, and indeed in KK the frequency of OV is *higher* than that of VO. Also the average frequency of SOV is in Moksha higher than that of SVO, although in Erzya SOV is a far second to SVO. This could be explained in several ways: perhaps a higher expressiveness of the characters' speech, a stylistic device used consciously by the author to evoke the feeling of a rustic, or archaic, type of language (if perceived archaicity was the goal, then it might fall in line with Saarinen's (1991: 50) claim that 19th century's folklore texts indicate an earlier dominance of VO); finally, it could be an individual preference of the author — in any case, with a p-value of 0.021 it seems to be more than a mere accident, even though a larger data sample would be desirable. For instance, there are as many as 5 appearances of the phrase мон шутендамазень аф кельгса 'I am not fond of jokers/pranksters' in the second chapter of KK, repeated in various contexts by Tulai, which almost artificially increases the occurrence rate of SOV.

Some word orders (namely, verb-initial) are unattested at all, which might be due to the smaller number of clauses overall or some pragmatic qualities that make them more acceptable in storytelling than in a representation of a human conversation. It is also interesting that the number of OVS clauses is in both languages higher than that of OSV clauses, 11:6 in non-dialogue and 4:3 in dialogue — although the minuscule data size unfortunately does not allow us to draw any conclusions here.

In any case, these results do not exactly match those obtained by Vilkuna (1998) in her comparative study of Erzya, Komi, Udmurt and Finnish (reviewed in more detail in Chapter 3). There, each of SXV, XSV and XVS types in Erzya was roughly 4-6 times less frequent than SVX, here — only the subject-initial types (SVO, SOV) roughly follow this proportion, with object-initial clauses being much less frequent than Vilkuna's generalised X-initial type. This indicates that inclusion of different types of clauses really distorts the image. X, believed to be a generalisation of object-patterners, in Vilkuna's data appears preverbally in 36.8% of all clauses with an overt subject, whereas in the present corpus, in main clauses, this percentage is 20.4% in Erzya narration (23.1% overall) and 25.8% in Moksha narration (34.7% overall), that is, with a much lesser frequency.

A tendency visible in both datasets is the marginal character of verb-initial clauses, especially in Erzya, and perhaps also a general preference to separate S and X/O with a verb (based on comparison of the frequencies of SVX and XVS with SXV and XSV respectively).

Table 10. Order of constituents in data analysed by Vilkuna (1998) and those presented in this study (for which X=O).

	Erzya (Vilkuna 1998)		Erzya (narration)		Erzya (narration + dialogue)		Moksha (narration)		Moksha (narration + dialogue)	
SVX	65	60.2%	79	77.5%	95	76.0%	85	70.2%	103	62.4%
SXV	11	10.2%	13	12.7%	17	13.6%	22	18.2%	44	26.7%
XSV	13	12.0%	3	2.9%	4	3.2%	3	2.5%	5	3.0%
XVS	16	14.8%	5	4.9%	7	5.6%	6	5.0%	8	4.8%
VSX	2	1.9%	1	1.0%	1	0.8%	3	2.5%	3	1.8%
VXS	1	0.9%	1	1.0%	1	0.8%	2	1.7%	2	1.2%
Total	**108**		**102**		**125**		**121**		**165**	

6.2 Dependent clauses

The number of dependent clauses relevant for the research turned out to be underwhelmingly low, which meant an accordingly little amount of data to work with. These results are nevertheless presented here for documentation purposes, and to allow for a short commentary.

Table 11. Order of constituents in dependent clauses in Erzya and Moksha narration. The numbers in brackets in the OV row and its subtypes indicate the number of occurrences where the object is not a subordinator or an NP containing a subordinator.

	Erzya		Moksha	
	ECK	TP	KK	PP
VO	**7**	**2**	**8**	**4**
incl. SVO	0	2	2	2
OV	**4 (1)**	**3 (0)**	**9 (5)**	**4 (1)**
incl. OVS	3 (1)	2 (0)	4 (3)	1 (0)
incl. OSV	0	0	1	0

It should come as little surprise that in my data, in VO clauses, the subject, if present, always precedes the verb (the other VO types are anyway very rare). On the other hand, perhaps it is a subordinator in the role of a subject that triggers the VO order of constituents; unfortunately, there is only a single example of such a clause in the data (64a), other SVO dependent clauses having standard subjects (64b).

(64) a. Moksha (KK1)

[*Гаройть аляц аф мяляфты сянь колга баснятка,*]

кие	*тяза*	*путозе*	*васенце*	*кудть*
who	to.here	build.PST1.3SG>3SG	first	house.GEN.DEF
REL, S G		V		O

'[Garoi's father would not be able to recall even any legend telling] who built here the first house'

b. Erzya (TP)

Зярдо *немецтне* *нолдыльть* *ракета,*
when German.PL.DEF let.out.PST2.3PL rocket
REL, T S V O
[*сынь прыльть сэрей ловонть потс.*]

'Whenever the Germans fired a rocket, [they would dive into the deep snow.]'

It might be more interesting that the subject typically comes as the last constituent in dependent OV clauses, which is evident even if this small sample. If so, it should be justification enough for treating dependent clauses separately, as it is a clear pattern that could otherwise be lost, or even worse, misinterpreted (seeing that OVS is in general a rare type, without much advantage in frequency over the other object-initial order, OSV).

The only example of an OSV clause in this section of data is (65), where the verb is marked by its late position as conveying more important information than the subject.

(65) Moksha (KK1)

[*книжканзон да листовканзон шири,*]

конатненъ *аф* *весть* *сторожсъ,* *фкя сельме* *Миронць,*
which.PL.GEN.DEF NEG once custodian.DEF one eye Miron.DEF
REL, O T S
мушендозенъ *кошкастонза.*
find.FREQ.PST1.3SG>3PL hovel.ELA.PX.3SG
V SRC

'[to his books and leaflets,] which the custodian, one-eyed Miron, not once found in his hovel'

There seems to be no perceivable difference between dependent clauses in Erzya and Moksha, although the fact that out of all clauses where the object does not occupy the initial position as a subordinator (and which would therefore trivially belong to the OV type), 6/18 take the OV order in Moksha, and only 1/10 in Erzya, seems not without significance. Again, this alone might be an accident, or a result of factors left unnoticed her, but taken together with other observations, it seems that Levchaev is somewhat consistently more OV-oriented in comparison to Abramov, whatever the reason.

There were very few dependent clauses in dialogue (Table 12). The relatively larger number of 'unforced' OV orderings, at least in Erzya, catches attention,

Table 12. Order of constituents in dependent Erzya and Moksha clauses in dialogue. The numbers in brackets have the same meaning as in Table 11

	Erzya		Moksha	
	ECK	TP	KK	PP
VO	**1**	**1**	**3**	**1**
incl. SVO	0	0	0	1
OV	**9 (2)**	**2 (1)**	**3 (3)**	**4 (0)**
incl. OVS	1 (0)	0	0	0
incl. OSV	1 (0)	0	0	0
incl. SOV	0	0	2	0

Table 13. Order of constituents in dependent Erzya and Moksha clauses in dialogue and narration together. The numbers in brackets have the same meaning as in Table 11

	Erzya			Moksha		
	ECK	TP	Total	KK	PP	Total
VO	**8**	**3**	**11**	**11**	**5**	**16**
incl. SVO	0	2	2	2	3	5
OV	**13 (3)**	**5 (1)**	**18 (4)**	**12 (8)**	**8 (1)**	**20 (9)**
incl. OVS	3 (1)	2 (0)	5 (1)	4 (3)	1 (0)	5 (3)
incl. OSV	1 (0)	0	1 (0)	1	0	1
incl. SOV	0	0	0	2	0	2

although with such a small data size no conclusions can be drawn from this fact. If one were to assume equivalency between dependent clauses in the narration and in dialogue, a slightly higher rate of OV would be found in Moksha (36%) than in Erzya (27%), however, this is far from a meaningful result.

In dialogue, we have a second appearance of the OSV order (66), this time in Erzya (ECK).

(66) Erzya (ECK) (a lieutenant to his two subordinates)
[*Монень а эряви содамс,*]
мезе тынь арсиде!
what 2PL think.PST1.2PL
REL, O S V
'[I do not need to know,] what you thought!'

This ordering in (66) is clearly for focus reasons: putting тынь at the end of the sentence would result in undesirable focusing of the subject pronoun, while the emphasis of the sentence is rather on the act – and the object – of thinking of the

addressees expressed by the verb арсиде. The role of the (most likely unstressed) pronoun here is for the speaker to distance himself from the listeners.

Note how different it is from the case of clauses with nominal subjects, where the clause-final position is the preferred, apparently unmarked one. This seems to be the cause in the only OVS clause appearing this part of the dataset (67), where Надя is clearly topical as the clause concludes Nadya's backstory related by her mother, (but so are Russian partisans referred to with the pronoun тенст).

(67) Erzya (ECK)

[*Мон а содан,*]

мезень	*лезэ*	*тенст*	*макссь*	*Надя.*
what.GEN	advantage	DAT.3PL	give.PST1.3SG	Nadya
	REL, O	O'	V	S

'[I do not know] what kind of advantage Nadya gave them.'

In this part of the dataset, there is also a unique double appearance of the SOV order in KK1 (68).

(68) a. Moksha (KK1)

[*азода спасиба ся ломанти,*]

кие	*инксонт*	*пакарензон*	*тапай*
who	CAUS.2PL	bone.GEN.PX.3SG	break.3SG
REL, S	O'	O	V

'[say thanks to the person], who gives it all for your sake'

b. Moksha (KK1)

[*Тон кулендить (...),*]

кода	*тя*	*Тулаюсь*	*аф*	*кунара*	*минь*	*эшелямазь*
how	DEM	Tulai.DEF	NEG	long.ago	1PL.GEN	bathe.PST1.3>1PL
REL		S		T	O	V

апак	*уштт*	*баняса?*
NEG.CVB	kindle.CNG	bath.INE
		LOC

'[Have you heard (...),], how that Tulai not so long ago bathed us in an unheated bath?'

The subject-initial position in (68a) is forced due to its appearance as a subordinator кие, and further by the strong emphasis on пакарензон 'bone.GEN.SG.PX.3SG' (highlighting the devotion with which the person being referred to allegedly serves the villagers' well-being). The order in (68b) is likely the effect of the object being expressed with a pronoun, минь — as will be seen later, pronouns favour the preverbal position, as indicated by Rueter (2023: 514).

6.3 Interrogative clauses

It seems that all observations valid for dependent clauses should also follow for interrogative (content) clauses (embedded or not), and perhaps also for embedded polar interrogative clauses due to the common property of a single constituent of the sentence (interrogative pronoun and verb along with its negation, respectively) being selected and fronted, the rest of the clause being organised around this movement.

All examples of embedded polar interrogative clauses found in the texts have already been cited in 3.2.3. Unfortunately, their simple structure leaves nothing to be discussed — as expected, all of them feature their focused element (verb) as the initial constituent, with any possible arguments and complements following it.

As for content interrogative clauses, they would normally be expected to appear only in dialogues, but rhetorical and embedded questions are found also elsewhere. Tables 14 and 15 depict their statistics of word order.

Table 14. Constituent order in content interrogative clauses (including embedded questions) in narration. The numbers in the brackets relate the number of clauses where the initial element is not the focus of the question.

	Erzya		Moksha	
	ECK	TP	KK	PP
VO	**1**	**0**	**0**	**0**
OV	**0**	**2 (0)**	**5 (0)**	**1 (1)**
incl. OSV	0	1 (0)	1 (0)	0
incl. OVS	0	0	1 (0)	0
incl. SOV	0	0	0	1 (1)

Table 15. Constituent order in content interrogative clauses (including embedded questions) in dialogue. For the meaning of the numbers in the brackets, see above.

	Erzya		Moksha	
	ECK	TP	KK	PP
VO	**2**	**1**	**4**	**1**
incl. SVO	0	0	1	0
OV	**6 (0)**	**1 (0)**	**2 (2)**	**2 (0)**
incl. OSV	1 (0)	0	1 (1)	0
incl. OVS	1 (0)	0	0	0

Most of the questions found in the narration have their object as focus, consequently, in most of them the object is clause-initial (69a) due to the usual fronting of WH-words regardless of their syntactic role. There are only two exceptions,

where the object is not the focus: one VO clause in Erzya (69b) and one OV in Moksha (69c).

(69) a. Moksha (KK2)

— а мезе сон азы?
but what 3SG say.3SG
Q, O S V

'— and what will she say?'

b. Erzya (ECK)

мейс сезнизе сёрмантъ...
for.what tear.PST1.3SG>3SG letter.GEN.DEF
Q, PURP V O

'why did he tear up the letter...'

c. Moksha (PP)

Да кие сонь эздонза Фана атянь вастса
and who POSS.3SG ABL.3SG Fana old.man.GEN place.INE
Q, S O LOC
атказаль.
decline.CONJ.3SG
V

'And who would decline it in the place of the old Fana?'

In the dialogue, there are more interrogative clauses with different foci; those with OV ordering are almost all object-focused, with two exceptions in Moksha: one of them having a stressed topical object before the subject interrogative pronoun (70a), the other (70b) being a very emotional negative (as shown later in this chapter, negative clauses tend to favour preverbal objects).

(70) a. Moksha (KK2)

Сянь кие аф содасы, Макар Антоныч!
DEM.GEN who NEG know.3SG>3SG Makar Antonovich
O S N V

'Who doesn't know that, Makar Antonovich!'

b. Moksha (PP)

Мес верда шокши педлитъ ашитъ
why blood.ABL suck.APRT leech.GEN.DEF NEG.PST1.2SG>3
Q, CAUSE O N

шарфта *кевкс?!*
turn.into.CAUSE.CNG stone.TRA
V

(speaking to God) 'Why haven't you turned the blood-sucking leech into stone?'

VO clauses seem to be the default whenever anything other than the object is the focus of the question:

(71) a. Erzya (ECK)
Мекс *венчесь* *а* *кунсолы* *эйсэнь?*
why boat.DEF NEG listen.3SG INE.1SG
Q, CAUSE S N V O
'Why is the boat not obeying me?'

b. Moksha (KK2)
Ой, кормилескяй, *да* *коса* *тяфта валондыть*
oh, breadwinner.DIM.VOC and where like.this drench.PST1.2SG>3
Q, LOC A V
шалхка варятнень?
nose hole.GEN.PL.DEF
O
'O, dear breadwinner, and where have you got so drunk?'

All over subjects appearing in OV clauses (dialogue or not) are pronouns. Personal (i.e. non-interrogative) pronouns are, as was the case with dependent clauses, found preverbally when not focused, and postverbally otherwise (example 72, focus marked in bold).

(72) a. Erzya (ECK)
[*Мон экшелявлинь — те мезеяк арась, кудосан.*]
Вана мезе теевлить ***тон?***
lo! what do.CONJ.2SG 2SG
Q, O V S
'If I were to take an [unvoluntary] bath — tha's not a problem, I am at home. But, look, what would **you** do?'

b. Erzya (TP)
Зняроце ***ве*** *сынь* *а* *удыть* *паро ладсо.*
how.many.ORD night 3PL NEG sleep.3PL good order.INE
Q, O, T S N V A
'Which night (in a row) are they not sleeping properly?'

c. Moksha (KK1)

[*сембе ванцть Кирей Михайловичень лангс,*]

мезе азы ***сон***

what say.3SG 3SG

Q, O V S

'[everybody looked at Kirey Mikhaylovich, wondering] what he is going to say'

d. Moksha (KK2)

[*чипордань сельмот ванць Ага бабань лангс*]

— а ***мезе*** *сон азы?*

but what 3SG say.3SG

Q, O S V

'[his eyes blinking, he looked at the old Aga] — and what is she going to say?'

From this part of the review it should be evident that transitive interrogative clauses and dependent clauses share many traits which distinguish them from main clauses. Despite their relatively small number, their inclusion into the main dataset would obfuscate the results; but it is also their small number which does not allow me to analyse them in more detail.

6.4 Verbal conjugation paradigm and constituent order

Tables 16 and 17 contain a cross-comparison of the constituent order in Moksha and Erzya with the type of conjugational marking on the verb (subjective or objective). Tables 18 and 19 recapitulate the statistics available in extant sources, mentioned in 3.2.1.

Even a brief comparison reveals huge disparities in the data.

In Kangastus' corpus, SVO is not even four times more common than SOV, whereas my Erzya dataset has with verbs in objective conjugation an overwhelming ratio of 44:1. In addition, in her results both object-initial orders, OSV and OVS, appear with a non-negligible frequency, which was also the case in Vilkuna's 1998 sample but not in the present data (regardless of the verbal conjugation type). It could be due to the nature of the material used by her, namely, 2003's issues of the journal "Сятко", containing texts of various styles and genres, poetry, prose and drama alike (Kangastus 2012: 8), that our results differ. It is likely that the later date of these texts and their variety is also the cause of the significantly lower relative frequency of VO ordering than in both the halves of the new datasets presented here, whether it be simply due to greater flexibility of ordering or a

Table 16. Constituent order in Erzya, divided into subjective and objective conjugation.

		Subjective		Objective		Total	
VO	pure VO	86	50	84	39	170	89
	SVO		35		44		79
	VSO		0		1		1
	VOS		1		0		1
OV	pure OV	27	10	18	14	45	24
	SOV		12		1		13
	OSV		1		2		3
	OVS		4		1		5
%VO	**Total**	**76.1%**	**113**	**82.4%**	**102**	**79.1%**	**215**

Table 17. Constituent order in Moksha, divided into subjective and objective conjugation.

		Subjective		Objective		Total	
VO	pure VO	82	51	177	118	259	169
	SVO		28		57		85
	VSO		2		1		3
	VOS		1		1		2
OV	pure OV	39	27	40	21	79	48
	SOV		8		14		22
	OSV		1		2		3
	OVS		3		3		6
%VO	**Total**	**67.8%**	**121**	**81.6%**	**217**	**76.6%**	**338**

Table 18. Constituent order in Erzya, based on Kangastus (2012) and reformatted for convenience.

		Objective	
VO	pure VO	244	135
	SVO		104
	VSO		3
	VOS		2
OV	pure OV	142	80
	SOV		26
	OSV		19
	OVS		17
%VO	**Total**	**63.2%**	**386**

Table 19. Constituent order in Moksha, based on Kholodilova and Toldova (2018: 611).

		Subjective		Objective		Total	
VO	pure VO	54	28	87	54	141	82
	SVO		24		28		52
	VSO		2		3		5
	VOS		0		2		2
OV	pure OV	51	49	89	66	140	115
	SOV		2		17		19
	OSV		0		1		1
	OVS		0		5		5
%VO	**Total**	**51.4%**	**105**	**49.4%**	**176**	**50.2%**	**281**

higher degree of emotionality in genres other than the one presented here, i.e. short story/novella.

On the other hand, the data presented in Kholodilova and Toldova (2018) (based on a non-literary dialect of Moksha) generally confirm the pattern of non-relevance of VSO, VOS, OSV and OVS types of clauses. Somewhat greater flexibility seems to be found in objective than subjective clauses, although the higher occurrence rate of the less popular types in this category could be due to the larger size of this part of the sample.

Kholodilova and Toldova (2018) present Moksha as a language with an essentially even distribution of VO and OV word order, independently of the conjugation. My data, on the other hand, clearly demonstrate both languages as VO, though with a quantitative disproportion between the two conjugational types. Namely, verbs with object indexing in both languages precede their objects roughly four times as often as they follow them (the VO pattern is present in approximately 82% of all objective clauses in both languages). The situation of verbs marked only for their subject is different: they follow their objects more often, with only 76.1% of VO occurrences in Erzya and an even lower rate of 67.2% in Moksha; although the difference is not statistically significant due to the small sample size, the matter might be worth further study.

The data in Kholodilova and Toldova (2018) suggest a preference for SVO against SOV (overwhelmingly so for subjective clauses and slightly in objective) but for OV over VO in absence of an overt subject (more so in the case of subjective than objective sentences, so that the two tendencies even out), mirroring Ermuškin's conclusions about the Srednetyoshski dialect of Erzya mentioned earlier in this work, as well as Kangastus' results.

The present data confirm the preference for SVO over SOV, but on a different scale in Erzya and Moksha, which in turn both also differ from those in Kholodilova

and Toldova (2018): in Moksha, the tendencies are similar in both types of conjugation, with a consistent ratio of roughly 4 to 1 (77-80% of the sum of all SOV and SVO occurrences); but in the Erzyan part of the data, SOV clauses are essentially absent with objective conjugation verbs (so a reversed case of the Mokshan data presented by Kholodilova and Toldova, where this happens in the subjective conjugation instead), and SVO is less than 3 times as common as SOV (making its advantage lower than that in my Mokshan data but higher than that of the objective conjugation in Kholodilova and Toldova (2018)).

The only example of an Erzyan SOV objective clause is the first sentence of ECK (73), which could perhaps be interpreted as a variation of an "out-of-the-blue" statement (Vilkuna 1998: 190, see also 3.2 in this work), although with Чебрецов Николай Петрович appearing clause-initially, rather than finally, as a definite subject (in spite of their usual lack of overt marking, personal names are functionally definite in Erzya (Bernhardt 2021: 40-41, 48-52)).

(73) Erzya (ECK)

Чебрецов Николай Петрович отпусконзо те иестэ
Chebretsov Nikolai Petrovich vacation.GEN.PX.3SG DEM year.ELAT
S O

ютавтызе оймсема кудосо, Кавказсо,
pass.CAUS.PST1.3SG>3SG rest.NMLZ house.INE Caucasus.INE
V LOC

Эльбрусонтъ маласо.
Elbrus.GEN.DEF near.INE

'Nikolai Petrovich Chebretsov spent his summer vacation at a holiday house in Caucasus, near the Elbrus.'

The 'pure' VO/OV clause type presented here reflects the general trend of a preferred postverbal object position, but a closer examination reveals something interesting: contrary to Ermuškin's (2004: 155-6) observation, supported by Rueter (2010: 26-7), to the effect that in clauses without an overt subject the object could be more likely to appear preverbally, the frequency of OV in relation to VO is in fact lower than that of SOV in relation to SVO, curiously enough, with verbs of opposite conjugations: subjective in Erzya and objective in Moksha.

The difference in Erzya could be due to the generally low acceptance of preverbal indefinite objects (see the following section on object types), and the real reason for a higher OV rate in the presence of a subject is that most of these clauses in the corpus are of negative polarity (it might be the case that negation of a verb pragmatically requires presence of a nearby overt subject). Without the negative sentences, there are 4 occurrences of pure OV and 3 of SOV (against 46 occurrences

of ‘pure’ VO and 33 of SVO), suggesting a rather equal distribution independent of the presence of the subject. The impact of negation on word order is described further below.

As for the difference in Moksha, note that the occurrence ratio of SOV and SVO appears independent of verbal conjugation type, indicating that it is the lack of the subject that causes the difference in distribution of the ‘pure’ VO and OV orders. It might be the case that definite objects require presence of a nearby subject (in the same clause or a neighbouring one) to precede their verbs, whereas indefinite objects are more liable to the strategy compared by Vilkuna 1998 to incorporation in Hungarian and (potentially) Permic languages (postverbal objects indeed make an impression of being more ‘tangible’, ‘countable’ or contrastive, perhaps intended to become the topic of the following discourse; actions associated with them seem more concrete, comp. e.g. (74a, 74b) and (74c, 74d)).

(74) a. Moksha (KK2)

[*шинек-венек увнасть-кенордасть Тулаютъ заведениянза:*]

яжасть	*почфт,*	| *тисть*	*ямкст,*	| *лифтсть*	
grind.PST1.3PL	flour.PL	make.PST1.3PL	porridge.PL	pull.out	
V	O	V	O	V	
канцёр вай.					
hemp	oil				
O					

‘Tulai’s enterprises operated day and night in full swing: they ground flour, made porridge, extracted hemp oil.’

b. Moksha (PP)

Цёратне,	*кле,*	*вирь*	*керыхть*
boy.PL.DEF	QUOT	forest	cut.3PL
S		O	V

— *ярмак*	*кргайхть*	*тойнь*	*каямс,*
money	rake.3PL	bride.price.GEN	pour.INF1
O	V		PURP

а	*стирнятне*	*куйгорхт-керхт*	*ваткайхть*
but	girl.PL.DEF	birch.bark-tree.bark	clean.up.3PL
	S	O	V

— *ярмак*	*пуроптыхть*	*казненъ*	*рамсемс.*
money	summon-3PL	present.GEN	buy.INF1
O	V		PURP

‘Boys, they say, chop the woods — they scrape money to pay bride price,

and the girls clean up birch bark and such — they summon money to buy presents.'

c. Moksha (KK1)

сон=га *сюволясь* *кулуфнень* *потмоса,*
3SG=EMPH rummage.PST1.3SG ash.GEN.PL.DEF inside.INE
S V LOC

вешендсь *модамарь.*
find.PST1.3SG potato
V O

'he, too, rummaged in the ashes, found a potato.'

d. Moksha (PP)

Маласа *чикоргоды* *сязъганць* | — *пара* *куля* *кандсь!*
near.INE screech.3SG jackdaw.DEF good news bring.PST1.3SG
LOC V S O V

'The jackdaw screeches nearby — it has brought (some) good news!'

As a side note, it is an intriguing detail that in Erzya subjective and objective clauses appear to have a roughly equal frequency, with subjective conjugation slightly prevailing, while in Moksha, both in my data and in Kholodilova and Toldova (2018), objective conjugation is used on around 63% of all verbs (63.9% here and 62.6% there); this could perhaps be related to the higher rate of occurrence of definite-marked objects in Moksha, seen in Table 20 in the next section.

6.5 Object type, definiteness and constituent order

Table 20 presents a comparison of the distribution of VO and OV orders with different types of objects (recall that in the present study the label "pronoun" encompasses all kinds of pronouns, not only personal). The differences in the distribution are minor, with one exception: indefinite nominal objects are much more likely to precede their verbs in Moksha (30.9% chance) than in Erzya (10.7%). Note that this mirrors the higher rate of occurrences of preverbal objects with subjective conjugation (where one expects to find indefinite objects) in Moksha when compared to the Erzyan portion of the corpus, and the similarity of the distribution in objective-conjugation clauses (where definite objects are expected).

In Erzya, the tendency to appear preverbally rises in the order: NDEF < DEF/POSS < PRON, with indefinite noun objects preceding their verbs only once in 10 occurrences. In Moksha, the first two elements are reversed, giving the ordering DEF/POSS < NDEF < PRON.

Table 20. Relative position of the object to the verb, divided by object type.

		Erzya		Moksha	
Pronoun	VO	27	60.0%	20	57.1%
	OV	18	40.0%	15	42.9%
	Total	45		35	
Definite or possessed NP	VO	93	81.6%	174	83.3%
	OV	21	18.4%	35	16.7%
	Total	114		209	
Indefinite NP	VO	50	89.3%	65	69.1%
	OV	6	10.7%	29	30.9%
	Total	56		94	
Total		**215**		**338**	

Review of dialogue data shows a slightly different pattern: here, both pronouns and indefinite nouns lean towards the OV ordering visibly more in Moksha than in Erzya. All classes of objects, except Erzya indefinite nouns, have a higher likelihood of appearing preverbally in dialogues than in narration.

Table 21. Relative position of the object to the verb, divided by bject type, in dialogues.

		Erzya		Moksha	
Pronoun	VO	9	47.4%	6	25.0%
	OV	10	52.6%	18	75.0%
	Total	19		24	
Definite or possessed NP	VO	22	68.8%	42	70.0%
	OV	10	31.3%	18	30.0%
	Total	32		60	
Indefinite NP	VO	22	91.7%	22	41.5%
	OV	2	8.3%	31	58.5%
	Total	24		53	
Total		**75**		**137**	

In both languages, pronouns have the highest likelihood of appearing preverbally, in 40-43% of all cases in the narration part, and up to three quarters of all occurrences in dialogues. This could be somehow related to the fact that preverbal pronoun position is unmarked in dependent clauses (see above); perhaps pronouns have a higher tendency to become unstressed, highly thematic expressions (preferring preverbal positions), used for contrast, connection or specification in cases of ambiguous indexing rather than as fully fledged constituents.

Recall Rueter (2023: 514) stating that SOV is especially popular if both S and O are pronouns: this is not verifiable for the narration due to lack of relevant data but appears true for dialogue: there, one finds 10 affirmative clauses (E — 2, M — 8) with SOV order, 2 (1 each) with SVO; within this set, all Moksha clauses feature personal pronouns in the roles of S and O. The one with SVO order is (75), whose ordering might be owed to euphonic reasons; the only Erzyan clause within this set where both S and O are instantiated by pronouns is SOV.

(75) Moksha (PP)

и	*синь*	*пчкафтсазь*	*сонь*
and	3PL	heal.3PL>3	3SG.GEN
	S	V	O

'and they will heal her'

Still, even in the absence of an overt subject, there are 6 OV and only 2 VO affirmative clauses with object pronouns in Moksha dialogue (and 2 to 7 in the narration, for a total of 8 to 9), and the proportion is 3 OV and 5 VO (4 to 16 including the narration) in Erzya. It seems that in Moksha indefinite pronouns (мезе-бди, мезе кельк 'anything') favour the preverbal position, and that inessive postpositions in Erzya (counted as pronouns due to their zero-possessor being classified as a pronoun) appear mostly postverbally.

In negative sentences, there is a tendency for a preverbal pronominal object mainly in the form of a negative pronoun. There are too few data points to establish any pattern for other pronouns appearing as objects in negative clause. Although the Erzyan data might suggest that lightweight, monosyllabic pronouns (many of them, in fact, referring to participants of the conversation, who are naturally very topical, which might be an alternative explanation to be proven or disproven with a larger dataset) could favour the preverbal position (comp. 76a), more complex ones appearing postverbally, this might not extend to Moksha. There, I found three instances of VO order where O is a short pronoun, though notably all of them involve some sort of particular focus, on the object (76c) or the verb (76d).

(76) a. Erzya (TP)

немецтненень	*прям*	*аволия*	*максо*
German.ALL.PL.DEF	head.PX.1SG	NEG.CONJ.1SG>3SG	give.CNG
O'	O	N	V

'I wouldn't let the Germans take me on!'

b. Erzya (ECK)

секс	*эзинек*	*содаво*	*вейкенек-вейкенек,*
for.this.reason	NEG.PST1.1PL	know.PASS.CNG	one.1PL~REFL
CAUSE	N	V	S

‘this is why we weren’t recognised by one another’

c. Moksha (PP)

Маряк, *сон=га* *аф* *кельгсамазь* ***минь,*** *ашу*
hear.IMP.2SG 3SG=EMPH NEG love.3SG>1PL 1PL.GEN poor
S N V O
ломаттненъ. . .
person.GEN.PL.DEF

‘See, he, too, does not love **us**, poor people.’

d. Moksha (KK1)

Аф ***тисы*** *сянь,* | *совестец* *аф* *саты*
NEG do.3SG>3SG DEM.GEN conscience.PX.3SG NEG be.enough.3SG
N V O S N V
грабамс *сембе* *велеть.*
rob.INF1 all village.GEN.DEF
PURP
‘He **won’t do** it, he does not have what it takes to rob the entire village.’

6.6 Polarity and constituent order

Based on the data shown in Table 22, it appears that polarity is a fairly good predictor of constituent order: the majority (even if not particularly overwhelming) of negative clauses follow the OV order in most data points, whereas affirmative clauses tend to be VO (seemingly more often in Erzya than in Moksha), and due to their domination this is also the general trend.

Table 22. Relative position of V and O in positive and negative transitive clauses.

	Erzya						Moksha					
	ECK		TP		Total		KK		PP		Total	
Affirmative	**110**		**76**		**186**		**152**		**159**		**311**	
VO	95	86.4%	65	85.5%	160	86.0%	125	82.2%	123	77.4%	248	79.7%
OV	15	13.6%	11	14.5%	26	14.0%	27	17.8%	36	22.6%	63	20.3%
Negative	**20**		**9**		**29**		**12**		**15**		**27**	
VO	6	30.0%	4	44.4%	10	34.5%	6	50.0%	5	33.3%	11	40.7%
OV	14	70.0%	5	55.6%	19	65.5%	6	50.0%	10	66.7%	16	59.3%
Total	**130**		**85**		**215**		**164**		**174**		**338**	

In transitive clauses expressing full negation, the “empty” object slot of a transitive verb can be filled in with a negative pronoun preferring the preverbal

position (e.g. (E) мезеяк, мезтькак ‘nothing’; see 77a, 77b). Sometimes, an NP marked with the intensifying clitic (E) -як (and variants) (M) -вок, translatable as ‘not even a’, appears as an object in such sentences and is located at the preverbal focus position (77c).

(77) a. Erzya (TP)

Лейтенантось тень каршо=як мезеяк эзь
lieutenant.DEF DEM.GEN against=EMPH nothing NEG.PST1.3SG
S O’ O N
мере.
say.CNG
V

‘To that, the lieutenant did not say anything, either.’

b. Erzya (ECK)

Сон ней содась весемедентъ,
3SG now know.PST1.3SG everything.ABL.DEF
S T V O
секс мезтькак эзь кевкстне.
for.this.reason nothing.PL NEG.PST1.3SG ask.CNG
PURP O N V

‘Now he knew everything, so he did not ask anything.’

c. Moksha (PP)

но землянкав пря=вок изь няфне
but earthhouse.LAT head=NEG.EMPH NEG.PST1.3SG show.CNG
G O N V

‘But he did not show himself at all to the earthhouse.’

Table 23. Relative position of V and O in negative clauses where O is not instantiated by a negative pronoun.

	Erzya			Moksha		
	Narration	Dialogue	Total	Narration	Dialogue	Total
VO	9	3	12	11	4	15
OV	8	6	14	12	16	28
Total	**17**	**9**	**26**	**23**	**20**	**43**

Negative pronouns, despite heavily contributing to the disproportion between statistics of object position in positive and negative transitive clauses, are certainly

not the only reason for the difference. Even after exclusion of all appearances of pronouns such as (E) мезеяк 'nothing', киньгак 'nobody', мездеяк 'nothing.ABL', (M) мезевок 'nothing', the statistic still indicates a much higher rate of OV occurrence than in positive clauses, especially in dialogue. In the narration, objects of negative clauses seem to have a roughly 50% chance of appearing either pre- or postverbally in both Erzya and Moksha, which is strikingly different from the 86% rate of VO occurrence in Erzya and 80% in Moksha.

6.7 Position of adverbial in finite verbal clauses

Table 24 presents the statistics of position of an adverbial of manner and degree relative to the verb and object in transitive clauses. In general, a preverbal position of such adverbs is favoured in both Erzya and Moksha (slightly more in Moksha), although postverbal position is not all too rare in either language. Adverbials rarely occur between verbs and their objects.

Table 24. Position of adverbial of means, degree, etc in transitive clauses in narration and dialogue.

	Erzya				Moksha			
	ECK	TP	Total		KK	PP	Total	
AVO	7	7	14	50.0%	9	14	23	63.9%
VOA	5	2	7	25.0%	4	2	6	16.7%
OVA	1	3	4	14.3%	1	0	1	2.8%
OAV	0	1	1	3.6%	1	1	2	5.6%
VAO	2	0	2	7.1%	0	2	2	5.6%
AOV	0	0	0	0.0%	1	1	2	5.6%
Total	**15**	**13**	**28**	**100.0%**	**16**	**20**	**36**	**100.0%**

Table 25 presents statistics on the position of adverbial in intransitive clauses. Here, the preverbal tendency is more visible in Erzya than in Moksha, especially in clauses without an overt subject. In clauses containing a non-zero subject, however, the adverb prefers to appear at the opposite side of the verb, so that both verb-initial and verb-final orderings are relatively infrequent. The function of the third most frequent ordering, AVS, seems to be putting stress on both the adverb and the subject (78).

(78) a. Moksha (PP)
Шобда, пелькс: аф лац увнайхть сире пичетне
dark danger: NEG well shimmer.3PL old pine.PL.DEF
A V S

землянкатъ *вельхксса.*
earthhouse.GEN.DEF above-INE
LOC

'It's dark and full of danger: the way the old pine trees above the earthhouse are shimmering is not a pleasant one.'

b. Erzya (ECK)
шождынестэ *чаръкодьсы* *грамотас* *содыця*
light.DIM.ADV understand.3SG>3SG alphabetism.ALL know.APRT
A V
эрьва *ломанесь*
every person.DEF
S

'every person with elementary education understands it easily'

Table 25. Position of adverbial of means, degree, etc in intransitive clauses in narration and dialogue.

	Erzya				Moksha			
	ECK	TP	Total		KK	PP	Total	
VA	**21**	**27**	**48**	**48.0%**	**32**	**45**	**77**	**52.4%**
pure VA	8	11	19	19.0%	21	30	51	34.7%
SVA	12	16	28	28.0%	10	12	22	15.0%
VSA	1	0	1	1.0%	1	2	3	2.0%
VAS	0	0	0	0.0%	0	1	1	0.7%
AV	**29**	**23**	**52**	**52.0%**	**34**	**36**	**70**	**47.6%**
pure AV	18	9	27	27.0%	23	21	44	29.9%
SAV	5	11	16	16.0%	10	8	18	12.2%
ASV	1	1	2	2.0%	0	1	1	0.7%
AVS	5	2	7	7.0%	1	6	7	4.8%
Total	**50**	**50**	**100**	**100.0%**	**66**	**81**	**147**	**100.0%**

6.8 Constituent order in clauses with infinitive complements

Tables 26 and 27 present the data on the relative position of a modal/auxiliary verb and its infinitive argument. It appears that the VO-patterning V+ARG_V order is overwhelmingly dominant in both Erzya and Moksha, regardless of the storytelling style or the type of the infinitive. The slight percentage differences in Moksha seem as more of an artefact of the small sample than an indicator of

a significant difference. Neither does the type of the infinitive seem to affect the ordering in any significant way. It might be worth noting that in negative clauses the relative frequency of the "reversed" ordering seems higher than otherwise: out of 30 negative clauses (dialogue and non-dialogue, 7 (23.3%) have the ARG_V+V ordering, e.g. (79)).

Table 26. Ordering of a finite verb and its infinitive argument depending on style (narration/dialogue).

	Erzya					Moksha				
	V+ARG_V		ARG_V+V		Total	V+ARG_V		ARG_V+V		Total
Narration	93	90.3%	10	9.7%	103	43	95.6%	2	4.4%	45
Dialogue	41	93.2%	3	6.8%	44	25	83.3%	5	16.7%	30
Total	**134**	**91.2%**	**13**	**8.8%**	**147**	**68**	**90.7%**	**7**	**9.3%**	**75**

Table 27. Ordering of a finite verb and its infinitive argument depending on the argument infinitive type.

	Erzya					Moksha				
	V+ARG_V		ARG_V+V		Total	V+ARG_V		ARG_V+V		Total
-мс	55	93.2%	4	6.8%	59	17	85.0%	3	15.0%	20
-мV	75	90.4%	8	9.6%	83	42	91.3%	4	8.7%	46
-мVдV/-мда	4	80.0%	1	20.0%	5	9	100.0%	0	0.0%	9
Total	**134**	**91.2%**	**13**	**8.8%**	**147**	**68**	**90.7%**	**7**	**9.3%**	**75**

(79) Erzya (TP)
Ловось соламо а снарнти=як.
snow.DEF melt.INF2 NEG try.CNG=NEG.EMPH
S ARG_V N V

'The snow is not even trying to melt.'

A different situation from the aforementioned can be found in the ordering of the infinitive argument of a modifying verb and its object, as shown in tables 28 and 29.

In Erzya, the rates of objects preceding and following their verbs are quite similar to those found with finite verbs. In Levchaev's Moksha, however, a preference for the object to precede its head infinitive is found with -ма infinitives, a tendency similar to that mentioned in Kholodilova and Toldova (2018: 676-7). 80a (Erzya, with VO order) and 80b (Moksha, with OV order) are typical examples of transitive clauses with the verb кармамс in their respective languages.

Table 28. Relative ordering of an infinitive verb and its object in narration and dialogue.

	Erzya					Moksha				
	V_ARG+O		O+V_ARG		Total	V_ARG+O		O+V_ARG		Total
Narration	20	76.9%	6	23.1%	26	3	33.3%	6	66.7%	9
Dialogue	4	80.0%	1	20.0%	5	4	40.0%	6	60.0%	10
Total	**24**	**77.4%**	**7**	**22.6%**	**31**	**7**	**36.8%**	**12**	**63.2%**	**19**

Table 29. Relative ordering of different types of infinitives and their objects. The statistic for complements of the -мV verb кармамс is shown separately in the final row.

	Erzya					Moksha				
	V_ARG+O		O+V_ARG		Total	V_ARG+O		O+V_ARG		Total
-мс	10	83.3%	2	16.7%	12	5	62.5%	3	37.5%	8
-мV	13	72.2%	5	27.8%	18	2	18.2%	9	81.8%	11
-мVдV/-мда	1	100.0%	0	0.0%	0	0	—	0	—	0
Total	**24**	**77.4%**	**7**	**22.6%**	**31**	**7**	**36.8%**	**12**	**63.2%**	**19**
'кармамс'	6	85.7%	1	14.3%	7	1	14.3%	6	85.7%	7

(80) a. Erzya (ECK)

Авась *кармась* *кайсеме* *муськематненъ* *пикс* *ланга*
woman.DEF start.3SG pour.INF2 laundry.GEN.PL.DEF rope on
S V ARG_V O G

'The woman started spreading out the laundry on a rope '

b. Moksha (PP)

сиренясь *арась* *шкайняниц* *инголи* / *и*
old.DIM.DEF stand.PST1.3SG icon.GEN.PX.3SG in.front.of.LAT | and
S V G

кармась *молитванц* *митертьфтема*
start.PST1.3SG prayer.GEN.PX.3SG whisper.INF2
V O ARG_V

'the old woman stood in front of her icon and started whispering her prayer '

Although this tendency is an interesting finding, Levchaev's prose appears to be more of an exception than the rule for Moksha: in addition to the data based on a sample from "Мокшень правда" analysed by Kholodilova and Toldova (2018), I also tried myself to estimate the frequency of similar usage elsewhere in Moksha, in artistic prose rather than in journalism. For this purpose, I searched for all usages of the lexeme кармамс in works classified as fiction within the Moksha Corpus

(Arkhangelskiy 2018–2021), obtaining 39 hits from mostly pre-WW2 prose. This sample contained ten examples of ARG_V+O and only one of O+ARG_V ordering, confirming the hypothesis that Levchaev's usage is rather exceptional.

It is worth mentioning that in the type of sentences discussed here, the object is connected to its infinitive rather loosely and can be moved e.g. to a position before the finite verb for pragmatic reasons, or for seemingly no reason at all: the object in (81a) appears clause-initially as a highly topical constituent; in (81b) the early position might indicate the speaker's emotional ; as 81c shows, it does not have to be verb-adjacent, either — this behaviour is similar to other Uralic languages (Vilkuna 2022: 959).

(81) a. Erzya (TP)

Атаканть	*снартокшность*	*тееме*	*зоря*	*ланга*
attack.GEN.DEF	intend.PST1.3PL	do.INF2	dawn	on
O	V	ARG_V		T

'They intended to do the attack at dawn'

b. Moksha (KK1)

Корхтайхть,	*Тулаюсь*	*нинге*	*кафта*	*заведениям*	*кармай*
talk.3PL	Tulai.DEF	even.more	two	enterprise.PL	start.3SG
	S			O	V

строяма?..
build.INF2
ARG_V

'They say, Tulai is going to build two more enterprises?'

c. Erzya (TP)

Старшинась	*снартнесь*	*сювордамс*	*конёв*	*потс*	*чувто*
starshina.DEF	attempt.PST1.3SG	wrap.INF1	paper	inside.ALL	tree
S	V	ARG_V		G	

нупонть.
moss.DEF.GEN
O

'The starshina tried to wrap tree moss with paper. '

The complex formed by the verb and its infinitive complement also allows interruptions other than the indirect object, especially, although not exclusively, in Levchaev's texts (82). Usually, an adverb or a secondary complement of the infinitive appears in this position.

(82) a. Moksha (KK1)

а *сядо=нга* *пяк* *кельгсы* *кундаф* *калдонза*
but COMP=EMPH very love.3SG>3SG catch.PTCP fish.ABL.PX.3SG
A V INS

ломаненъ *каваняманц*
person.GEN treat.to.INF2.3SG
O ARG_V

'but he loves treating people to his catch even more'

b. Erzya (ECK)

(...) *лейтенантось* *снартокшнось* *весть* *аволдазь* *керямс*
lieutenant.NOM.DEF want.PST1.3SG once shake.CVB cut.INF1
S V A ARG_V

вечкеманъ *сюлмонтъ*
love.NMLZ.GEN knot.GEN.DEF
O

'the lieutenant wanted to cut the knot of love in one go'

6.9 Order of constituents in nominal predication

Tables 30 and 31, 32 and 33 present the data on distribution of constituent orderings in clauses with nominal predication and identificational clauses in the present corpus. The negative copulae (E) аволь / (M) аф, and (E) арась / (M) аш (which in addition to existential clauses appear in constructions such as (E) тя аш мезевок 'it's nothing') are not considered here. Different types of predication (classificational, attributional, predlocative...) are not distinguished here.

Table 30. Order of subject and predicate in Erzya narration and dialogue.

	Narration				Dialogue				Total	
	ECK		TP		ECK		TP			
S+NPR	41	89.1%	30	90.9%	24	100.0%	21	95.5%	116	92.8%
NPR+S	5	10.9%	3	9.1%	0	0.0%	1	4.5%	9	7.2%
Total	**46**	**100.0%**	**33**	**100.0%**	**24**	**100.0%**	**22**	**100.0%**	**125**	**100.0%**

The preferred ordering is subject-predicate, more commonly in dialogues than elsewhere, with slightly more variation in Moksha than in Erzya; and this does not seem to stem from a higher rate or range of usage of the bound copula in Moksha than in Erzya as only 4 such instances can be unambiguously identified among all 21 NPR+S clauses in Moksha (and there is also one clause-initial predicate with bound copula among the 9 NPR+S clause in Erzya).

Table 31. Order of subject and predicate in Moksha narration and dialogue.

	Narration				Dialogue				Total	
	KK		PP		KK		PP			
S+NPR	19	79.2%	25	83.3%	48	85.7%	32	91.4%	124	85.5%
NPR+S	5	20.8%	5	16.7%	8	14.3%	3	8.6%	21	14.5%
Total	**24**	**100.0%**	**30**	**100.0%**	**56**	**100.0%**	**35**	**100.0%**	**145**	**100.0%**

Table 32. Position of the free copula ул(ьн)емс in clauses with nominal predication (narration and dialogue considered together) in Erzya.

	Erzya					
	ECK		TP		Total	
COP+NPR	**17**	**89.5%**	**14**	**87.5%**	**31**	**88.6%**
S+COP+NPR	13	68.4%	13	81.3%	26	74.3%
NPR+COP	**2**	**10.5%**	**2**	**12.5%**	**4**	**11.4%**
NPR+COP+S	1	5.3%	2	12.5%	3	8.6%
S+NPR+COP	1	5.3%	0	0.0%	1	2.9%
Total	**19**		**16**		**35**	

Table 33. Position of the free copula ул(ьн)емс in clauses with nominal predication (narration and dialogue considered together) in Moksha.

	Moksha					
	KK		PP		Total	
COP+NPR	**9**	**90.0%**	**3**	**75.0%**	**12**	**85.7%**
S+COP+NPR	4	40.0%	2	50.0%	6	42.9%
NPR+COP	**1**	**10.0%**	**1**	**25.0%**	**2**	**14.3%**
NPR+COP+S	0	0.0%	0	0.0%	0	0.0%
S+NPR+COP	0	0.0%	0	0.0%	0	0.0%
Total	**10**		**4**		**14**	

Despite the small data size, it seems that the nominal predicate is a somewhat stronger object-patterner than the object itself, with both subject and full copula rarely following it (whether alone or together). The less common "reversed order" can be found e.g. in (83a), a specificational clause with what I interpreted as a topical predicate marked with translative case. (83b) has a focused predicate and the ordering in (83c) seems to be conditioned if not by focus then by the kind of contextuality Rueter (2010: 26-7) suggested as a possible reason for OV ordering in Erzya — perhaps a certain degree of personal relevance for the speaker plays an important role here. In clause (83d), the intent could have been to put emphasis on

стака; note also, however, that here стака 'difficult, heavy' (classified as a special 'emotive' predicator by Kholodilova and Toldova (2018: 706)) appears with a verbal infinitive as its subject, making the clause somewhat distinct from the rest of the examples provided here.

(83) a. Erzya (ECK)

Истямо сяткокс Николай Петровичень туртов ульнесь
such spark.TRA Nikolai Petrovich for be.PST1.3SG
NPR O' V
те сёрмась.
DEM letter.DEF
S

'For Nikolai Petrovich, this letter was just such a spark [as the ones that have just been described].'

b. Erzya (TP)

Кавонест туртов ульнесь те винтовкась ды
two.COM.3PL for be.PST1.3SG DEM rifle and
NPR, O' V
лейтенантонь пистолетэсь.
lieutenant.GEN gun.DEF
S

'This rifle and the lieutenant's gun had to be shared by the two of them.'

c. Moksha (KK1)

тячи-ванды орёлоньке лийкстай, | кудса ули!
today-tomorrow eagle.PX.1PL fly.3SG home.INE be.3SG
T S V LOC, NPR V

'today-tomorrow our eagle will fly, he will be home! '

d. Erzya (TP)

нинтень съкамонзо стака ули кастомост
wife.ALL.DEF alone.3SG heavy be.3SG grow.CAUS.NMLZ.3PL
O' NPR V S

'bringing them up alone will be hard for the wife '

Orderings in which an intervening element would appear between the copula and the predicate were almost unattested in main clauses in the present corpus, example 85d presented a bit further below being the only such occurrence. On the

other hand, a unique ordering of the main constituents, with the subject placed between the (apparently fronted) copula and the predicate, was found three times in counterfactual conditional clauses in Erzya (84).

(84) a. Erzya (ECK)
Улевель=гак *Надя* *велув* *немецтнень* *марто,*
be.CONG=EMPH Nadya together German.GEN.PL.DEF with
V S A COM
[*(...) сонзэ кадовлизь кудосо аштеме?*]
'If Nadya had been together with the Germans, [would they have let her stay at home?]'

b. Erzya (TP)
Улевель=гак *весь* *валдо*
be.CONJ=EMPH night.DEF bright
V S NPR
[*нама, седе шождынель чарькодемс, мезть неть истятнэ.*]
'If the night had been bright [it would probably have been easier to understand what these were.]'

A thing which immediately comes to attention is that usage of an overt verbal copula in nominal predication (which in a decisive majority of cases precedes its complement in both languages) is relatively low. The latter ratio is probably due to the different semantics of the usage of the verb улемс in the past tense leading to a preference of the clitic copula usage; it is not too surprising, knowing that bound copula in Moksha is generally found to have a wider range of application than in Erzya (see Kholodilova 2016: 256 for a comparison of various tendencies).

Among usages of an overt copula in Moksha, there are two expressions referring to specific points in time, of which one (85a) contains an adverb of time as its predicate and the other classifies/characterises a (semantically rather weak) subject instantiated with the word 'time' (85b). Others could be described as perfective, describing "having been" somewhere as a background for other events (85c), or referring to a temporary state, additionally marked with the translative case (85d). All remaining uses in Moksha are in present tense with future reference or in moods other than indicative (these require usage of a full, non-deficient copula, as in 85e and 85f).

(85) a. Moksha (PP)
Тевсь *ульсь* *илять*
deed.DEF be.PST1.3SG evening.DEF.GEN
S V T, NPR
'It happened at evening (lit. the event was at the evening)'

b. Moksha (PP)

Пингсь	*ульсь*	*вчёк*	*шада*	*ведь*	*ланга,*	*кельме,*
time.DEF	be.PST1.3SG	just	seasonal.flood	water	on	cold
S	V				NPR	
начка,	*рдаз.*					
wet	dirt					

'The time was just in the season of floods, cold, wet, dirty.'

c. Moksha (KK1)

Корхтайхть,	*ульсь*	*Питерьсо=нга*
talk.3PL	be.PST1.3SG	Petersburg.INE=EMPH
	V	NPR

'They say he has even been to Petersburg.'

d. Moksha (KK1)

эста	*сон*	*ульсь*	*Тулаюти*	*работникокс*
then	3SG	be.PST1.3SG	Tulai.DAT.DEF	worker.TRA
T	S	V	O'	NPR

'at that time he used to be a labourer for Tulai'

e. Moksha (PP)

вечна	*катк*	*улихть*	*сюдофто=нга*	*сюдофт*
forever	OPT	be.3PL	curse.PTCP.ABL=EMPH	curse.PTCP.PL
T		V		NPR

'May they be the most accursed forever'

f. Moksha (PP)

аньцек	*ульхть*	*ломань,* |	*тят*	*переченда!*
only	be.IMP.2SG	person	NEG.IMP.2SG	resist.CNG
	V	NPR	N	CNG

'just be a person, do not resist!'

Perhaps the difference in usage is related to the availability of the derived verb ульнемс 'be.FREQ.INF1' next to the basic улемс 'be.INF1' in past-tense usage in Erzya, which allows a wider range of usage than the underived verb. Although the frequentative form is much more common in Erzya, an example of past-tense use of its base verb in Erzya can also be found in the present data (86). It is perhaps worth noting here that while a dynamic interpretation ('to become') is sometimes possible with this verb, in this example, as I have been told by a native speaker, only the static reading is possible.

(86) Erzya (ECK)

Авась	*ульсь*	*вельть*	*кортыця.*
woman.DEF	be.PST1.3SG	quite	talk.APRT
S	V		NPR

'The woman was quite talkative.'

As a closing remark, it should be noted that the clitic copula always appears in postposition to either the head of the predicate or the focused element in the clause, so the two mechanisms could be seen as complementary. It could perhaps even be hypothesised that the somewhat larger preference for head-final ordering in transitive clauses in Moksha is in a way reflected in the language's relatively higher frequency of copula usage.

7 Conclusion

The results of the study presented in this work indicate various types of tendencies in constituent ordering as well as differences between the two Mordvin literary languages. The following are the most important conclusions:

1. The distribution of constituent ordering in clauses is highly dependent on text genre and language variety.

2. Verb-initial and object-initial clauses are extremely rare in narration.

3. SVO is more common than SOV in both languages.

4. In narration, the VO and OV types (ignoring the position of an optional subject) appear with a similar distribution of 3-4 to 1 in both languages. In dialogue, the ratio changes somewhat in favour of OV, to a greater degree in Moksha than in Erzya: in Levchaev's dialogues, the frequency of OV clauses rivals that of VO — although this might be a feature of conscious stylisation, given the nature of the texts used in the corpus.

5. The overall tendency for OV–patterning usage and constructions appears to be somewhat stronger in Moksha.

6. Main indicative clauses differ from dependent and content interrogative clauses in their properties of word order. The two latter seem to be more or less similar to one another, for instance, in both these types the position of a topical subject is preverbal when instantiated by a personal pronoun and postverbal as an NP.

7. Pronouns, especially negative pronouns, have a higher tendency than other types of objects to precede their verb.

8. Nouns not marked for definiteness rarely precede their verbs in Erzya but relatively often in Moksha. In the latter case, preverbal and postverbal usage could depend on the referentiality of the object, with less concrete, more background objects occurring preverbally and specific, referential objects following their verbs.

9. The position of an object can be restricted (usually to a postverbal position) depending on the presence or absence of an overt subject: for example, when the subject is present, objects of verbs with objective conjugation marking almost always appear postverbally in Erzya; and contrary to some expectations, objects of subjective verbs in Erzya and objective verbs in Moksha are less likely to precede their verbs with a null-subject.

10. Verbs are preceded by their objects noticeably more often in negative than in positive clauses.

11. Adverbials of manner and degree appear preverbally and postverbally with a roughly equal frequency in intransitive clauses and more often preverbally in transitive clauses.

12. Infinitive complements of modal and auxiliary verbs almost always (roughly 9 of out every 10 cases) follow their heads in both languages, independent of the type of the infinitive.

13. In constructions with modal and auxiliary verbs, objects of the infinitive verb follow it in Erzya and precede it in Moksha, and the result seems statistically significant; however, the latter tendency could be a property of Levchaev's style or idiolect as data from other sources shows more similarity to the "Erzya"-style patterning.

14. The full copula улемс/ульнемс consistently, although not exclusively, follows its subject and precedes its complement. Its usage is restricted in Moksha past tense, the bound copula being preferred instead.

These results are based on a small, very specific data sample; a larger, more representative corpus would be necessary to allow for generalisation of the tendencies observed here. Although a tentative explanation for some of the outcomes has been provided in this thesis, further work as well as higher degree of participation of native-speaker expertise would be required to fully account for all of the various tendencies spotted in the data.

In general, the assertions of earlier authors (Vilkuna 1998; 2022; Rueter 2010) concerning the role of context and focus on word order in Mordvin hold well, while those of an equivalent unmarked role of OV and VO do not: every, or almost every case of OV can be accounted for in one way or another, making it far from unmarked. To this, I add a hypothesis that in some instances the difference between the two orders can depend on the degree of referentiality or contrastiveness of an indefinite object.

Many of the regularities observed here call for further research and hopefully the data provided here can be of use to those who will undertake it.

8 Kokkuvõte

Käesolevas lõputöös uurin ja võrdlen kahe mordva kirjakeele — ersa ja mokša — sõnajärge osalause ja lihtlause tasandil. Siin kirjeldatud uurimus põhineb väikesel korpusel, mis koosneb kahe tunnustatud mordvalasest autori, Kuzma Abramovi (ersa) ja Piotr Levtšajevi (mokša) neljast kirjanduslikust proosateosest. Kummagi autori loomingust võtsin kaks teost, mille märgendasin käsitsi osalausete ja nende lauseliikmete tüübi suhtes. Tulemusena valmis 3329 analüüsitava liht- ja osalausega (1623 ersa keeles, 1706 mokša keeles) korpus, kusjuures analüüsitavaks pean lauseid, mis sisaldavad vähemalt kaht selles töös uuritavat lauseliiget.

Korpuse suurus võimaldas valida uurimiseks ainult puhtaimat tüüpi liht- ja osalaused, kus iga lauseliikme roll on selge ning vaevata tuvastatav. Väljajäetud kõrvalaused ning eriküsimusi ja erikonstruktsioone sisaldavad osalaused jätsid andmestiku piisavalt suureks. Lisaks sellele eristasin jutustamist ja dialoogi.

Ülal kirjeldatud täpse lähenemisega sain mõnevõrra erinevaid tulemusi kui leebemaid kriteeriume kasutanud teadlased. Need tulemused viitavad erinevat tüüpi tendentsidele mordva kirjakeelte sõnajärjes ning erinevustele kahe mordva kirjakeele vahel. Kõige olulisemad järeldused on järgmised:

1. Tendentsid mordva keelte lauseliikmete järjekorras sõltuvad oluliselt tekstižanrist ja keele varieeruvusest (dialektist, murdest, jne).
2. Verbi ja objektiga algavaid (osa)lauseid esineb jutustamisel üliharva.
3. SVO on mõlemas keeles levinum kui SOV, v.a. mokša dialoogid, kus on ülekaalus SOV.
4. Jutustamises esinevad mõlemas keeles VO ja OV tüübid (subjekti asukoht ja olemasolu ei ole oluline) sarnase jaotusega: 3-4 VO juhtu ühe OV kohta. Dialoogis muutub suhe mõnevõrra OV kasuks, mokša keeles rohkem kui ersa keeles. Levtšajevi dialoogides konkureerib OV sagedus VO omaga — arvestades uuringus vaadeldud tekstide ainest võib see olla teadlik stiliseerimise tunnus.
5. Üldine kalduvus OV-keeltele omaste mustrite kasutamisele näib mokša keeles olevat mõnevõrra tugevam kui ersa keeles.
6. Deklaratiivsed pealaused erinevad kõrvallausetest ja eriküsimustest oma sõnajärje omaduste poolest. Kaks viimast tunduvad sarnased: näiteks on mõlemas tüübis asesõnaline subjekt tavaliselt enne verbi ning nimisõnafraasiline subjekt pärast verbi.
7. Asesõnadel, eriti eitavatel asesõnadel, on muud tüüpi objektidest suurem kalduvus eelneda oma tegusõnale.
8. Määramata nimisõnast objektid esinevad ersa keeles harva verbide ees, kuid mokša keeles suhteliselt sageli. Mokša keeles võib sellise objekti verbieelne- või -järgne kasutus sõltuda selle referentsiaalsusest, kusjuures verbile eelnevad

vähem konkreetsed, rohkem topikaalsed objektid ja järgnevad spetsiifilisemad objektid.

9. Objekti asukoha mobiilsus lauses võib olla piiratud (tavaliselt verbijärgsele positsioonile) olenevalt sellest, kas subjekt on lauses väljendatud või välja jäetud. Näiteks esinevad subjekti olemasolul määrava (objekt-) pöördtüübi verbide objektid ersa keeles peaaegu alati verbijärgselt. Määramata (subjekt-) pöördtüübi verbide objektid ersa keeles ja määrava pöördtüübi verbide objektid mokša keeles eelnevad oma verbile väiksema tõenäosusega, kui lausest on subjekt välja jäetud.
10. Eitavates lausetes on objekt märgatavalt sagedamini tegusõna ees kui jaatavates lausetes.
11. Viisi- ja hulgamäärused esinevad intransitiivsetes osalausetes verbieelselt ja -järgselt ligikaudu võrdse sagedusega ja transitiivsetes osalausetes sagedamini verbieelselt.
12. Infinitiivitäiendid järgnevad verbidele mõlemas keeles peaaegu alati (umbes 9 korda 10-st), sõltumata infinitiivi tüübist.
13. Infinitiivitäiendiga konstruktsioonides järgneb objekt infinitiivile ersa keeles ja eelneb mokša keeles; viimane tendents võib aga olla Levtšajevi stiili või idiolekti omadus, kuna muudest allikatest pärinevad andmed näitavad rohkem sarnasust „ersa-stiilis“ mustriga.
14. Täiskoopula улемс/ульнемс (eesti keeles *olema*) järgneb järjekindlalt, kuigi mitte ainult, oma subjektile ja eelneb oma täiendile.

Ülal toodud tulemused põhinevad väikesel, väga spetsiifilisel andmevalimil ning siin täheldatud tendentside tõestamiseks oleks vaja suuremat, esinduslikumat korpust. Kuigi selles lõputöös on mõnede tulemuste kohta esialgne selgitus, oleks kõigi korpusest leitud tendentside täielikuks arvessevõtmiseks vaja edasist tööd ja emakeelena kõnelejate suuremat osalust.

Üldiselt jäävad pidama varasemate autorite väited (Vilkuna 1998; 2022; Rueter 2010) konteksti ja fookuse rolli kohta mordva sõnajärjes. Mõnikord esinev väide OV ja VO järjestuste samaväärse markeerimatuse kohta ei näi aga olevat tõsi: peaaegu iga OV järjestust saab ühel või teisel viisil põhjendada, see pole kaugeltki markeerimata. Enda poolt panen ette hüpoteesi, et (vähemalt mokša keeles) võib mõnel juhul kahe järjekorra erinevus sõltuda määramatu objekti referentsiaalsuse või kontrastiivsuse astmest.

Paljud siin täheldatud seaduspärasused nõuavad edasist uurimist ja loodetavasti on siin esitatud andmed kasulikud neile, kes seda ette võtavad.

References

Aasmäe, Nina (2015). *Кортатано эрзякс. Поговорим по-эрзянски.* Ed. by Tõnu Seilenthal. Transl. by Denys Teptiuk and Tony Keller. Vol. 8. Tartu Ülikooli Paul Ariste Soome-Ugri Põlisrahvaste Keskuse Üllitised. Tartu. ISBN: 978-9985-4-0904-6.

Abramov, Vladimir (2009). "Слово о писателе. К 95-летию народного писателя К. Г. Абрамова". In: *Финно–угорский мир* 4, pp. 56–64. ISSN: 2541-982X. URL: https://csfu.mrsu.ru/arh/2009/4/56-64.pdf (visited on 04/03/2024).

Ajanki, Rigina (2011). "Double encoding of nominal and adjectival predicates: A study of the nominative–translative switch in Erzya". In: *Suomalais-Ugrilaisen Seuran Aikakauskirja* 2011.93. ISSN: 1798-2987, 0355-0214. DOI: 10.33340/susa.82503.

— (2014). "Zero transitive clauses in Erzya and Finnish". In: *Juuret marin murteissa, latvus yltää Uraliin.* Ed. by Sirkka Saarinen et al. Suomalais-Ugrilaisen Seuran Toimituksia 270. Helsinki: Suomalais-Ugrilainen Seura, pp. 3–17. ISBN: 978-952-5667-63-9.

Alhoniemi, Alho (1991). "Zur Kasuskennzeichnung des Objekts im Mordwinischen". In: *Suomen kielitieteellisen yhdistyksen vuosikirja 1991.* Helsinki: Suomen kielitieteellinen yhdistys, pp. 18–30. ISBN: 978-951-95264-5-4.

Arkhangelskiy, Timofey (2018–2021). *Moksha corpus.* URL: moskha.web-corpora.net.

— (2019). "Corpora of social media in minority Uralic languages". In: *Proceedings of the Fifth International Workshop on Computational Linguistics for Uralic Languages.* Proceedings of the Fifth International Workshop on Computational Linguistics for Uralic Languages. Tartu, Estonia: Association for Computational Linguistics, pp. 125–140. DOI: 10.18653/v1/W19-0311.

Asztalos, Erika (2021). "From head-final towards head-initial grammar: Generational and areal differences concerning word order usage and judgement among Udmurt speakers". In: *IMPACT: Studies in Language, Culture and Society.* Ed. by Diana Forker and Lenore A. Grenoble. Vol. 50. Amsterdam: John Benjamins Publishing Company. ISBN: 978-90-272-6001-7. DOI: 10.1075/impact.50.06asz.

Baker, Robin (1985). *The development of the Komi case system: a dialectological investigation.* Suomalais-Ugrilaisen Seuran toimituksia 189. Helsinki: Suomalais-Ugrilainen Seura. 266 pp. ISBN: 978-951-9019-91-8.

Bakró-Nagy, Marianne Sz, Johanna Laakso, and E. K. Skribnik, eds. (2022). *The Oxford guide to the Uralic languages.* Oxford, United Kingdom: Oxford University Press. ISBN: 978-0-19-182151-6.

Bartens, Raija (1999). *Mordvalaiskielten rakenne ja kehitys.* Suomalais-Ugrilaisen Seuran Toimituksia 232. Helsinki: Suomalais-ugrilainen seura. ISBN: 978-952-5150-22-3.

Bernhardt, Mariann (2021). *Määräisyys ja objektin merkitseminen mordvalaiskielissä*. Turku: University of Turku. ISBN: 978-951-29-8639-2. URL: https://www.utupub.fi/handle/10024/152777 (visited on 05/16/2024).

Bryzhinskiĭ, A. I., O. V. Pashutina, and E. I. Chernov (2001). *Pisateli Mordovii: biobibliograficheskiĭ spravochnik*. Saransk: Mordovskoe knizhnoe izd-vo. 591 pp. ISBN: 978-5-7595-1386-5.

Dryer, Matthew S. (1988). "Object-verb order and adjective-noun order: Dispelling a myth". In: *Lingua* 74.2-3, pp. 185–217. ISSN: 00243841. DOI: 10.1016/0024-3841(88)90059-9.

— (1992). "The Greenbergian word order correlations". In: *Language* 68.1, pp. 81–138. ISSN: 1535-0665. DOI: 10.1353/lan.1992.0028.

— (2007). "Word order". In: *Language typology and syntactic description*. Ed. by Timothy Shopen. 2nd ed. Cambridge, UK ; New York: Cambridge University Press, pp. 61–131. ISBN: 978-0-521-58857-7.

Erina, Ol'ga (1997). *Частицы в мордовских языках*. Dissertationes philologiae uralicae Universitatis Tartuensis 1. Tartu: Tartu Ülikooli Kirjastus. 149 pp. ISBN: 978-9985-56-268-0.

Ermuškin, Grigorij I. (2004). *Einheitliche Beschreibung der Dialekte uralischer Sprachen. 9, Srednetëškij dialekt ėrzja-mordovskogo jazyka*. Mitteilungen der Societas Uralo-Altaica 24. Hamburg: Finnisch-Ugrisches Seminar der Universität Hamburg.

Greenberg, Joseph H. (1990). "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements". In: *On Language*. Ed. by Keith Denning and Suzanne Kemmer. Stanford University Press, pp. 40–70. ISBN: 978-1-5036-2321-7. DOI: 10.1515/9781503623217-005.

Grünthal, Riho et al. (2022a). "Drastic demographic events triggered the Uralic spread". In: *Diachronica* 39.4, pp. 490–524. ISSN: 0176-4225, 1569-9714. DOI: 10.1075/dia.20038.gru.

— (2022b). "Drastic demographic events triggered the Uralic spread: Supplements". Version 1.0. In: *Diachronica* 39.4. DOI: 10.5281/ZENODO.6345559.

Hamari, Arja and Rigina Ajanki (2022). "Mordvin (Erzya and Moksha)". In: *The Oxford Guide to the Uralic Languages*. Ed. by Marianne Bakró-Nagy, Johanna Laakso, and Elena Skribnik. 1st ed. Oxford University PressOxford, pp. 392–431. ISBN: 978-0-19-182151-6. DOI: 10.1093/oso/9780198767664.003.0023.

Haspelmath, Martin (2022). "Nonverbal clause constructions". URL: https://www.academia.edu/71539642/Nonverbal_clause_constructions (visited on 05/16/2024).

Havas, Ferenc (2008). "Unmarked Object in the Uralic Languages. A Diachronic Typological Approach". In: *Linguistica Uralica* 43.1, p. 1. ISSN: 0868-4731. DOI: 10.3176/lu.2008.1.01.

Helimski, Eugene (2003). "Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: A survey". In: *Nyelvtudományi Közlemények* 100, pp. 156–167. URL: http://epa.niif.hu/04100/04182/00233/pdf/EPA04182_nyelvtudomanyi_kozlemenyek_2003_156-167.pdf (visited on 05/16/2024).

Kangastus, Anna-Maria (2012). "Subjektin ja objektin referenssi ersämordvan objektikonjugaatiolauseissa". MA thesis (unpublished). University of Helsinki. 115 pp.

Katorova, A. (2022). *Примерная рабочая программа учебного предмета «Родная (мордовская) литература» для 10–11 классов среднего общего образования (2022).* URL: https://fgosreestr.ru/uploads/files/beb4b9b47ff5223127f2da6fd7ba7fd8.docx (visited on 04/03/2024).

Kehayov, Petar (2020). "Between Facts and Speech Acts: the Conditional and Conditional-Conjunctive in Moksha Mordvin". In: *Linguistica Uralica* 56.1, p. 18. ISSN: 0868-4731. DOI: 10.3176/lu.2020.1.03.

Keresztes, László (2011). *Bevezetés a mordvin nyelvészetbe.* Debrecen: Debrecen University Press. 147 pp. ISBN: 978-963-318-167-6.

Kholodilova, Maria (2016). "Moksha non-verbal predication". In: *Mordvin languages in the field.* Ed. by Ksenia Shagal and Heini Arjava. Uralica Helsingiensia 10. Helsinki: University of Helsinki, Department of Finnish, Finno-Ugrian and Scandinavian Studies, pp. 229–259. ISBN: 978-952-5667-78-3.

Kholodilova, Maria and Svetlana Toldova, eds. (2018). *Элементы мокшанского языка в типологическом освещении.* Moscow: Буки Веди. 1013 pp. ISBN: 978-5-4465-2097-8.

Klumpp, Gerson (2016). "Semantic functions of complementizers in Permic languages". In: *Complementizer Semantics in European Languages.* Ed. by Kasper Boye and Petar Kehayov. De Gruyter, pp. 529–586. ISBN: 978-3-11-041661-9. DOI: 10.1515/9783110416619-016.

Klumpp, Gerson, Lidia Federica Mazzitelli, and Fedor Rozhanskiy (2018). "Typology of Uralic languages: current views and new perspectives. Introduction to the special issue of ESUKA – JEFUL". In: *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 9.1, pp. 9–30. ISSN: 2228-1339, 1736-8987. DOI: 10.12697/jeful.2018.9.1.01.

Luobbal Sámmol Sámmol Ánte (2022). "Proto-Uralic". In: *The Oxford Guide to the Uralic Languages.* Ed. by Marianne Bakró-Nagy, Johanna Laakso, and Elena Skribnik. 1st ed. Oxford University PressOxford, pp. 3–27. ISBN: 978-0-19-182151-6. DOI: 10.1093/oso/9780198767664.003.0001.

Metslang, Helena (2013). "Coding and behaviour of Estonian subjects". In: *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 4.2, pp. 217–293. ISSN: 2228-1339, 1736-8987. DOI: 10.12697/jeful.2013.4.2.12.

Nichols, Johanna (2021). “The Origin and Dispersal of Uralic: Distributional Typological View”. In: *Annual Review of Linguistics* 7.1, pp. 351–369. ISSN: 2333-9683, 2333-9691. DOI: 10.1146/annurev-linguistics-011619-030405.

"Память" (2024). *Левчаев Петр Иванович*. Проект «Бессмертный барак». URL: https://bessmertnybarak.ru/levchaev_petr_ivanovich/ (visited on 05/16/2024).

Polyakov, Osip E. (1995). *Учимся говорить по-мокшански*. Saransk: Mordovskoe knizhnoe izd-vo. 197 pp. ISBN: 978-5-7595-0965-3.

Polyakov, Vladimir N., Elena A. Makarova, and Ekaterina Paramoshko (2018). “High-frequency contrastive grammar features of the Uralic languages”. In: *XLinguae* 11.1, pp. 163–185. ISSN: 13378384, 2453711X. DOI: 10.18355/XL.2018.11.01.15.

Rantanen, Timo et al. (2022). “Best practices for spatial language data harmonization, sharing and map creation—A case study of Uralic”. In: *PLOS ONE* 17.6. Ed. by Søren Wichmann, e0269648. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0269648.

Rosstat (2022). *Итоги ВПН-2020. Том 5 Национальный состав и владение языками. Таблица 4. Владение языками и использование языков населением*. URL: https://rosstat.gov.ru/storage/mediabank/Tom5_tab6_VPN-2020.xlsx (visited on 05/19/2024).

Rueter, Jack (2010). *Adnominal person in the morphological system of Erzya*. Suomalais-Ugrilaisen Seuran toimituksia 261. Helsinki: Société Finno-Ougrienne. 236 pp. ISBN: 978-952-5667-24-0.

— (2013). “The Erzya Language. Where is it spoken?” In: *Études finno-ougriennes* 45. ISSN: 0071-2051, 2275-1947. DOI: 10.4000/efo.1829.

— (2016). “Towards a systematic characterization of dialect variation in the Erzya-speaking world: Isoglosses and their reflexes attested in and around the Dubyonki Raion”. In: *Mordvin languages in the field*. Ed. by Ksenia Shagal and Heini Arjava. Uralica Helsingiensia 10. Helsinki: University of Helsinki, Department of Finnish, Finno-Ugrian and Scandinavian Studies, pp. 109–148. ISBN: 978-952-5667-78-3.

— (2023). “Moksha Mordvin”. In: *The Uralic Languages*. Ed. by Daniel Abondolo and Riitta-Liisa Valijärvi. 2nd ed. London: Routledge, pp. 481–526. ISBN: 978-1-315-62509-6. DOI: 10.4324/9781315625096-11.

Saarikivi, Janne (2022). “The divergence of Proto-Uralic and its offspring: A descendent reconstruction”. In: *The Oxford Guide to the Uralic Languages*. Ed. by Marianne Bakró-Nagy, Johanna Laakso, and Elena Skribnik. 1st ed. Oxford University PressOxford, pp. 28–58. ISBN: 978-0-19-182151-6. DOI: 10.1093/oso/9780198767664.003.0002.

Saarinen, Sirkka (1991). “Typological differences between the Volgaic languages”. In: *Suomen kielitieteellisen yhdistyksen vuosikirja 1991*. Helsinki: Suomen kielitieteellinen yhdistys, pp. 43–52. ISBN: 978-951-95264-5-4.

Salamon, Ágnes (1989). “Tárgyas szerkezetek az erza-mordvinban”. In: *Folia Uralica Debreceniensia* 1, pp. 87–108. ISSN: 0239-1953.

Serebrennikov, B. A., A. P. Feoktistov, and O. E. Poliakov, eds. (1998). *Мокшанско-русский словарь: 41000 слов*. Moscow: Русский язык. 919 pp. ISBN: 5-200-02012-3.

Siewierska, Anna, ed. (1998). *Constituent order in the languages of Europe*. Empirical approaches to language typology EUROTYP 20,1. Berlin: Mouton de Gruyter. 829 pp. ISBN: 978-3-11-015152-7.

Siewierska, Anna and Ludmila Uhlířová (1998). “An overview of word order in Slavic languages”. In: *Constituent order in the languages of Europe*. Ed. by Anna Siewierska. Empirical approaches to language typology EUROTYP 20,1. Berlin: Mouton de Gruyter, pp. 105–149. ISBN: 978-3-11-015152-7.

Stefanowitsch, Anatol (2020). *Corpus linguistics: A guide to the methodology*. Zenodo. DOI: 10.5281/ZENODO.3735822.

Turunen, Rigina (2006). “Complex Morphosyntactic Features of Nominal Predicates in Erzya”. In: *SKY Journal of Linguistics*, pp. 173–187. ISSN: 1456-8438.

— (2009). “A typology of non-verbal predication in Erzya”. In: *Acta Linguistica Hungarica* 56.2-3, pp. 251–313. ISSN: 1216-8076, 1588-2624. DOI: 10.1556/ALing.56.2009.2-3.4.

Vilkuna, Maria (1998). “Word order in European Uralic”. In: *Constituent order in the languages of Europe*. Ed. by Anna Siewierska. Empirical approaches to language typology EUROTYP 20,1. Berlin: Mouton de Gruyter, pp. 173–233. ISBN: 978-3-11-015152-7.

— (2022). “Word order”. In: *The Oxford Guide to the Uralic Languages*. Ed. by Marianne Bakró-Nagy, Johanna Laakso, and Elena Skribnik. 1st ed. Oxford University PressOxford, pp. 950–960. ISBN: 978-0-19-182151-6. DOI: 10.1093/oso/9780198767664.003.0049.

Zaicz, Gábor (1998). “Mordva”. In: *The Uralic languages*. Ed. by Daniel Abondolo. First edition. Routledge language family descriptions. London; New York: Routledge, pp. 184–218. ISBN: 978-1-138-65084-8.

Licence

Non-exclusive licence to reproduce the thesis and make the thesis public

I, **Przemysław Podleśny**,
(*author's name*)

1. grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

   A comparative study of clause-level constituent order in Erzya and Moksha,
   (*title of thesis*)

   supervised by Gerson Klumpp and Petar Kehayov.
   (*supervisor's name*)

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Przemysław Podleśny
27/05/2024