

# **A grammar of the Shughni language**

a doctoral dissertation by

Clinton Parker

DEPARTMENT OF LINGUISTICS, MCGILL UNIVERSITY

MONTREAL, QC, CANADA

DATE



# Acknowledgements

The completion of this dissertation was made possible by the support of many, many people – so many, in fact, that it would be impossible to name them all here. But I would like to acknowledge those who have gone out of their way to support me through the ups and downs of this process.

First and foremost, I would like to thank the Shughni speakers who have generously shared their language with me and without whom this dissertation would have never been possible. Thanks to Mr. Noor Ali Dost, my first Shughni consultant, for his patience as a teacher, and to his entire family for their generosity, hospitality, and lasting friendship. I am very grateful to my two primary language consultants in Tajikistan: to Bakhmal Odibekova for her ever-entertaining Shughni lessons and for connecting me to many members of the Shughni-speaking community, and to Shahlo Shomansurova for her valuable linguistic insights, for her patience as I asked question after question about small grammatical details, and for sharing many moments of nerdy linguistic humor. You have all stuck with me until the bitter end, and I am forever appreciative of that.

My time in Tajikistan would not have been the same without Bakhmal Odibekova's parents, Mr. Davlat Odibekov, his wife Shahlo Shodmonbekova, as well as their son Nekruz. Thanks for opening up your home to me, for feeding me, and for making me feel at home during long stays in Tajikistan. Davlat, I will never get tired of trying to decipher your incessant Shughni puns, and Shahlo, I appreciate the long sessions of *qilaptow*-ing! Spending time with you both helped ease the struggles of dissertation writing and have made this process much more enjoyable!

Many thanks to those who participated in various fieldwork activities and phonetic experiments. Each and every one of you brought unique insight about your language, and it was a pleasure to work with each of you. I would also like to thank the people of Khorog, Tajikistan, for making me feel welcome. To all Shughni speakers, thank

you for always making me feel like learning your language is a worthwhile endeavor, and for always being willing to laugh with me as I inevitably made mistakes during the process.

I received much helpful feedback from members of the *Iranistics* group at the Higher School of Economics in Moscow, particularly from Sasha Sergienko and Prof. Vladimir Plungian. Thanks to you both for taking the time to read sections of this thesis and for your eagerness to share helpful insights.

I am appreciative of those who made my stays in Tajikistan possible and enjoyable. Thanks to the staff of American Councils for International Education, especially Mollima Nisso Rasulova and Jace Livingston; to Dr. Parvonakhon Jamshedov, who was a gracious academic host at the Tajik National Academy of Sciences during my time in Dushanbe; and to Parvina Adylova and Pablo Tutillo Maldonado, who went out of their way to provide support of all kinds. Thanks to Alexander Maier and Nick Seay for their friendship, companionship, and shared love of Tajikistan. I would also like to thank my host mother in Dushanbe, Muyassar Sharipova, as well as her sons, who have always welcomed me to their home with open arms and large plates of *osh*. You all and many others have shaped the positive experiences which have always left me wanting to return to Tajikistan.

I am grateful for the support of my professors at McGill, especially that of my advisor Jessica Coon, who has always been available and who has always been a positive voice even in times when I have doubted myself. Thanks also to the supervisors of my evaluation papers, Nico Baier, Janine Metallic, and Lisa Travis, and to my committee members Martina Martinović and Francisco Torreira. Many thanks to James Crippen and Geoffrey Haig for thorough reviews and helpful comments on this dissertation. To the members of my cohort, Emily Kellison-Linn, Gouming Martens, Yeongwoo Park, and Jia'er Tao: I look fondly upon the nights as Ph.D. plebeians working through problem sets until the early morning hours. Most of all, I appreciate your friendship, as it has been there through the best and worst of times along this journey.

Last but certainly not least, I would like to thank my parents, Fred and Victoria Parker, for their unwavering support in everything I set out to do in life. They have always encouraged me to pursue what I love, and this project certainly would not have been possible without them.

The research for this dissertation was supported by a Title VIII Combined Research & Language Training grant through American Councils for International Education, a FRQSC Doctoral Research Grant, and a Fulbright Student Research Grant.

# Abstract

This dissertation provides a comprehensive grammatical description of Shughni, an understudied Eastern Iranian language spoken in Afghanistan and Tajikistan. Contributing to this goal are three primary objectives: (i) to identify, analyze, and describe areas of the language's grammar which have heretofore not been investigated; (ii) to synthesize and build upon existing research on the language, much of which was published in Russian during the Soviet era and has therefore been hard to access for many scholars outside the former Soviet Union; and (iii) to point out phenomena in Shughni which are not covered in depth here and are ripe for future investigation. Novel contributions of this thesis include, among others, an analysis of definiteness in Shughni nominals, a thorough description of the language's system of triple deixis, particularly as it pertains to demonstratives and adverbs, and a description of tense and aspect situated in a Neo-Reichenbachian framework. It provides a thorough description of morphosyntactic alignment in Shughni from synchronic, diachronic, and typological perspectives, and it builds upon the work of Soviet-era scholars in providing a historical account of several irregularities found in the modern Shughni verbal system, notably the suppletion in infinitive, present, and past stems, as well as the development of syncretism in feminine and plural past and perfect forms. Ultimately, it is the author's hope that this thesis will both contribute to the documentation of Shughni and spur future research into un(der)-investigated areas of its grammar.

# Résumé

Cette thèse fournit une description grammaticale exhaustive de la langue shughni, une langue peu étudiée du groupe iranien oriental qui est parlée en Afghanistan et au Tadjikistan. Trois objectifs principaux contribuent à la réalisation de ce but: (i) identifier, analyser et décrire les domaines de la grammaire qui n'ont pas été étudiés jusqu'à présent; (ii) synthétiser et développer les recherches existantes sur la langue, dont une grande partie a été publiée en russe à l'époque soviétique et, par conséquent, a rendu difficile l'accès de nombreux chercheurs en dehors de l'ancienne Union soviétique; (iii) souligner des phénomènes de la langue shughni qui sortent du cadre de cette thèse et méritent d'être étudiés dans le futur. Parmi ses nouvelles contributions figurent l'analyse de la définitude dans les expressions nominales, la description approfondie du système déictique, particulièrement en ce qui concerne les démonstratifs et les adverbes, et l'analyse du temps et de l'aspect grammatical selon le modèle néo-Reichenbachien. Elle donne aussi une description approfondie de la structure d'actance du point de vue synchronique, diachronique et typologique. La thèse s'appuie également sur les résultats des chercheurs de l'ère soviétique afin d'offrir un compte rendu historique de plusieurs irrégularités dans le système verbal du shughni moderne, notamment la supplétion dans les radicaux infinitifs, présents, et passés, ainsi que le développement du syncrétisme dans les formes féminines et au pluriel du passé et du parfait. En fin de compte, l'auteur espère que cette thèse non seulement contribuera à la documentation de la langue shughni, mais aussi stimulera de futures études dans des domaines non explorés et peu explorés de la grammaire.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Geography of the Shughni-speaking region . . . . .	2
1.2	Sociolinguistic information . . . . .	6
1.3	Previous research on Shughni . . . . .	7
1.3.1	Early works on Shughni . . . . .	8
1.3.2	Works on specific aspects of Shughni . . . . .	8
1.3.3	Grammatical sketches and dictionaries . . . . .	10
1.4	Impetus for the present study . . . . .	11
1.5	Methodology . . . . .	13
1.5.1	Data from fieldwork sessions . . . . .	13
1.5.2	Data from natural speech . . . . .	16
1.5.3	Data from previous publications . . . . .	17
1.6	Overview of thesis . . . . .	17
<b>I</b>	<b>Preliminaries</b>	<b>21</b>
<b>2</b>	<b>Historical and genealogical considerations</b>	<b>23</b>
2.1	Broad genetic relations: Indo-Iranian . . . . .	24
2.2	Old, Middle, and Modern Iranian languages . . . . .	25
2.2.1	Old Iranian . . . . .	28
2.2.2	Middle Iranian . . . . .	31
2.2.3	Modern Iranian . . . . .	36
2.3	Eastern vs. Western Iranian languages . . . . .	38
2.4	Pamir languages . . . . .	40
2.5	Shughni-Rushani Subgroup . . . . .	45
2.6	Historical and genealogical considerations: Summary . . . . .	49
<b>3</b>	<b>Sound system and orthography</b>	<b>51</b>
3.1	Phoneme inventory . . . . .	52
3.1.1	Consonants . . . . .	52
3.1.2	Vowels . . . . .	58
3.2	Phonotactics . . . . .	64
3.2.1	Overview and permitted syllable shapes . . . . .	65
3.2.2	Coda clusters . . . . .	65
3.2.3	Resolution of illicit coda clusters in loanwords . . . . .	69
3.2.4	Social aspects of Shughni phonotactics . . . . .	70

3.3	Stress . . . . .	71
3.3.1	Stress and derivational affixes . . . . .	72
3.3.2	Stress and inflectional affixes . . . . .	73
3.3.3	Unstressed elements: Case-like suffixes and morphophonological clitics . . . . .	74
3.3.4	Stress in loanwords . . . . .	75
3.4	Orthography . . . . .	76
3.4.1	Latin script . . . . .	77
3.4.2	Cyrillic script . . . . .	78
3.4.3	Arabic script . . . . .	80
3.4.4	Orthography used in this thesis . . . . .	89
3.4.5	Shughni scripts side-by-side . . . . .	91
<b>4</b>	<b>Grammatical Overview</b>	<b>93</b>
4.1	Typological overview . . . . .	94
4.2	Nominals . . . . .	96
4.2.1	Pronouns . . . . .	96
4.2.2	Nominal inflection . . . . .	101
4.2.3	Nominal derivation . . . . .	102
4.2.4	Order of elements in noun phrases . . . . .	103
4.3	Case-like suffixes & adpositions . . . . .	104
4.3.1	Case-like suffixes: Locative and dative . . . . .	104
4.3.2	Prepositions . . . . .	106
4.3.3	Postpositions . . . . .	107
4.4	Verbs . . . . .	108
4.4.1	Verb stems and verbal inflection . . . . .	108
4.4.2	Transitive, unergative, and unaccusative verbs . . . . .	111
4.4.3	Verbal inflection . . . . .	120
4.4.4	Complex verbs . . . . .	122
4.4.5	Passive-like constructions and causatives . . . . .	123
4.5	Complex clauses and question formation . . . . .	126
4.5.1	Complex clauses . . . . .	127
4.5.2	Question formation . . . . .	128
<b>II</b>	<b>Nominals</b>	<b>131</b>
<b>5</b>	<b>Noun phrases, nominal inflection, and nominal derivation</b>	<b>133</b>
5.1	Content and reference of noun phrases . . . . .	134
5.1.1	Content, size, and order of elements in noun phrases . . . . .	135
5.1.2	Generic reference . . . . .	137
5.1.3	Indefinite reference . . . . .	138
5.1.4	Definite reference . . . . .	141
5.2	Number in nominals . . . . .	147
5.2.1	Additive plural . . . . .	148
5.2.2	Associative plural . . . . .	151
5.2.3	Similative plural . . . . .	152
5.2.4	Collective plural . . . . .	154
5.3	Nominal derivation . . . . .	155
5.3.1	Formation of nouns . . . . .	156
5.3.2	Formation of adjectives . . . . .	168



5.4	Numerals	176
5.4.1	Numerals: Form	176
5.4.2	Usage: Shughni vs. Tajik numerals	179
<b>6</b>	<b>Pronouns and Demonstratives</b>	<b>183</b>
6.1	Pronouns	184
6.1.1	Personal pronouns	184
6.1.2	Interrogative pronouns and <i>wh</i> -phrases	185
6.1.3	Indefinite pronouns	192
6.1.4	Reflexive and emphatic pronouns	194
6.2	Demonstratives	200
6.2.1	Form and syntax of demonstratives	201
6.2.2	Exophoric uses of demonstratives	203
6.2.3	Endophoric uses of demonstratives	212
6.2.4	Deictic presentatives	218
<b>7</b>	<b>Grammatical Gender in Nominals</b>	<b>221</b>
7.1	Overview of grammatical gender in Shughni	222
7.2	Historical considerations in Shughni grammatical gender	224
7.2.1	Old Iranian basics	225
7.2.2	Factors moderating sound change	226
7.2.3	Sound changes affecting the expression of gender	227
7.3	Morphological expression of gender in Shughni nominals	228
7.3.1	Expression of gender in Shughni nouns	230
7.3.2	Expression of gender in Shughni adjectives	232
7.3.3	Gender concord in noun phrases (vs. verb phrases)	234
7.4	Gender assignment in nouns	237
7.4.1	Overview of gender assignment: Three primary factors	238
7.4.2	Thematic classes and grammatical gender	240
7.4.3	Other semantic factors in gender assignment	247
<b>III</b>	<b>Verbs</b>	<b>253</b>
<b>8</b>	<b>Verb Stems, Inflection, and Modification</b>	<b>255</b>
8.1	Verb stems and usage	256
8.1.1	Verb-stem paradigms	256
8.1.2	Basic uses of verb stems	258
8.2	Verbal inflection	261
8.2.1	Person and number agreement	261
8.2.2	Verbal negation	263
8.3	Copular constructions	265
8.3.1	Second-position copular clitics	266
8.3.2	Copular verb <i>vidow</i>	268
8.3.3	Compatibility of the copula with different lexical classes	271
8.3.4	Existential and possessive constructions	272
8.3.5	Summary: Copular constructions	274
8.4	Verbal and predicate modification	275
8.4.1	Types and order of adverbs	275
8.4.2	Adverbial derivation and morphology	277

8.4.3	Adverbs of degree . . . . .	281
8.4.4	Epistemic adverbs . . . . .	284
8.4.5	Deictic adverbs . . . . .	285
<b>9</b>	<b>Regular and Irregular Verbs</b>	<b>293</b>
9.1	Regular and irregular verbs . . . . .	294
9.1.1	Monosyllabic and polysyllabic stems . . . . .	294
9.1.2	Regular verbs . . . . .	296
9.1.3	Irregular verbs: Consonant alternations . . . . .	299
9.1.4	Irregular verbs: Vowel alternations . . . . .	303
9.1.5	Verb-stem shortening . . . . .	310
9.2	Gender in verbs . . . . .	315
9.2.1	Which verbs distinguish gender? . . . . .	316
9.2.2	Historical considerations in gender-distinguishing verbs . . . . .	316
9.2.3	Vowel and consonant alternations in past and perfect stems . . . . .	317
9.2.4	Gender distinction in onomatopoeic verbs . . . . .	320
9.3	Recent and ongoing changes in Shughni verb stems . . . . .	324
9.3.1	Borrowings in the Shughni verb system . . . . .	324
9.3.2	Leveling in verb-stem paradigms . . . . .	325
<b>10</b>	<b>TAM and Evidentiality</b>	<b>331</b>
10.1	Definitions and framework used for TAM . . . . .	332
10.2	Temporal and aspectual compatibilities of verb stems . . . . .	335
10.2.1	Present stems . . . . .	335
10.2.2	Past stems . . . . .	344
10.2.3	Perfect Stems . . . . .	348
10.2.4	Tense and aspect: Interim summary . . . . .	350
10.3	Infinitive stems and aspectual reference . . . . .	352
10.3.1	Progressive aspect . . . . .	352
10.3.2	Inceptive aspect . . . . .	354
10.3.3	Prospective aspect . . . . .	355
10.4	Grammatical expressions of mood . . . . .	357
10.4.1	Grammatical distinctions between realis and irrealis . . . . .	357
10.4.2	Types of irrealis clauses . . . . .	360
10.4.3	Mood: Summary . . . . .	367
<b>11</b>	<b>Argument Structure and Transitivity</b>	<b>369</b>
11.1	Resultative participles and passive-like constructions . . . . .	370
11.1.1	Verbal resultative participles: Suffix <i>-ak</i> . . . . .	370
11.1.2	Adjectival passive participle: Suffix <i>-in</i> . . . . .	374
11.1.3	Summary: Resultative participles . . . . .	376
11.2	Causative . . . . .	377
11.2.1	Morphological causatives . . . . .	377
11.2.2	Periphrastic causatives . . . . .	384
11.2.3	Semantics of morphological and periphrastic causatives . . . . .	387
11.2.4	Section summary: Causatives . . . . .	389
11.3	Complex verbs . . . . .	390
11.3.1	Structure of complex verbs . . . . .	392
11.3.2	Transitivity in complex verbs . . . . .	393
11.3.3	Other issues in complex verbs . . . . .	397

11.4	Reflexive verbs . . . . .	402
11.5	Oblique-first constructions . . . . .	404
11.5.1	Accusative-first constructions . . . . .	407
11.5.2	Dative-first constructions . . . . .	408
11.5.3	Subjecthood and oblique-first constructions . . . . .	411
11.5.4	Summary: Oblique-first constructions . . . . .	413
<b>IV</b>	<b>Syntax</b>	<b>415</b>
<b>12</b>	<b>Other Syntactic Phenomena</b>	<b>417</b>
12.1	Complex clauses . . . . .	418
12.1.1	Relative clauses . . . . .	418
12.1.2	Temporal adverbial clauses . . . . .	428
12.1.3	Other subordinate clauses . . . . .	431
12.1.4	Complement clauses and dependent speech . . . . .	434
12.1.5	Coordination of clauses . . . . .	439
12.1.6	Complex clauses: Summary . . . . .	440
12.2	Information structure and clause-level word order . . . . .	442
12.2.1	Expression of topic in Shughni . . . . .	443
12.2.2	Expression of focus in Shughni . . . . .	448
12.2.3	Information structure and clause-level word order: Summary . . . . .	454
12.3	Questions . . . . .	455
12.3.1	<i>Wh</i> -questions . . . . .	455
12.3.2	Polar questions . . . . .	458
12.3.3	Echo questions . . . . .	459
12.3.4	Other question particles . . . . .	463
12.3.5	Questions: Summary . . . . .	466
12.4	Morphosyntactic alignment and vestigial ergativity . . . . .	467
12.4.1	Morphosyntactic alignment: Overview and terminology . . . . .	467
12.4.2	History and diversity of morphosyntactic alignment in the Iranian past tense . . . . .	471
12.4.3	Morphosyntactic alignment in Shughni . . . . .	483
12.4.4	Morphosyntactic alignment in Shughni: Summary . . . . .	490
<b>13</b>	<b>Conclusion</b>	<b>493</b>
	<b>Appendices</b>	<b>499</b>
	<b>Appendix A Nouns</b>	<b>501</b>
	<b>Appendix B Verb stems</b>	<b>505</b>
B.1	Consonant irregularities in verb stems . . . . .	507
B.2	Vowel irregularities in verb stems . . . . .	509
	<b>References</b>	<b>517</b>



# List of Tables

2.1	(Attested) Middle Iranian Languages. . . . .	35
2.2	Genetic groupings of Iranian languages. . . . .	38
2.3	Shughni-Rushani language varieties. . . . .	47
3.1	Shughni consonant phonemes. . . . .	53
3.2	Examples of Shughni consonant phonemes in onset (word-initial) and coda (word-final) positions. . . . .	54
3.3	Examples of minimal pairs in Shughni: Consonants. . . . .	55
3.4	Single and geminate /t/ in Shughni. . . . .	58
3.5	Shughni vowels phonemes . . . . .	59
3.6	Examples of minimal pairs in Shughni: Vowels. . . . .	59
3.7	High, middle, and low allophones of short vowels. . . . .	62
3.8	Morphophonological correspondences of short and long vowels: Evidence from verb shortening. . . . .	63
3.9	Permitted syllables shapes in Shughni. . . . .	65
3.10	Consonant clusters in native Shughni words. . . . .	67
3.11	Consonant clusters in Shughni inflected verbs. . . . .	68
3.12	Clusters of three consonants in Shughni. . . . .	68
3.13	Permitted clusters in Arabic borrowings. . . . .	69
3.14	Resolved illicit clusters in Arabic borrowings. . . . .	70
3.15	Stress-attracting affixes in Shughni. . . . .	73
3.16	Cyrillic characters used for Shughni. . . . .	79
3.17	Shughni consonant letters in the Arabic script. . . . .	81
3.18	Shughni consonants represented by multiple Arabic letters. . . . .	83
3.19	Shughni short vowels in Arabic orthography. . . . .	85
3.20	Shughni long vowels in Arabic orthography. . . . .	88
3.21	Orthography used in this thesis: Consonants. . . . .	90
3.22	Orthography used in this thesis: Vowels. . . . .	91
4.1	Direct personal pronouns. . . . .	97
4.2	Oblique personal pronouns. . . . .	97
4.3	Demonstrative pronouns. . . . .	97
4.4	Fundamental wh-words in Shughni. . . . .	98
4.5	Formation of indefinite pronouns in Shughni. . . . .	99
4.6	Formation of plural nouns in Shughni. . . . .	101
4.7	Example stem vowel alternations in grammatical gender distinction. . . . .	102
4.8	Common derivational affixes in Shughni. . . . .	103
4.9	Order of elements in NPs. . . . .	103
4.10	Common prepositions in Shughni. . . . .	107
4.11	Common prepositions in Shughni. . . . .	108

4.12	Distribution of verb stem types in Shughni.	111
4.13	Verb classes and compatibility with the third-singular past-tense clitic =(y)i.	114
4.14	Verb classes and compatibility with the third-singular past-tense clitic =(y)i.	116
4.15	Properties of transitive, unergative, and unaccusative verbs.	120
4.16	Present-tense agreement suffixes.	121
4.17	Past-tense clitics.	121
4.18	Common vowel alternations in gender-distinguishing verb stems.	122
4.19	Common light verbs used in complex verbs.	123
4.20	Intransitive~causative pairs.	126
5.1	Cardinal articles: Shortened forms of Shughni numerals.	140
5.2	Summary: Definite types and demonstrative use in Shughni.	147
5.3	Additive plural suffixes in Shughni.	148
5.4	Vowel alternations in plural forms.	149
5.5	Compatibility of the additive plural with different types of NPs.	150
5.6	Formation of the similative plural via reduplication.	153
5.7	Abstract-noun forming suffix -(y)i with nouns.	156
5.8	Noun-forming suffixes which select for nouns.	158
5.9	Abstract-noun forming suffix -(y)i with adjectives.	159
5.10	Occupational nouns with the suffix -iĵ.	160
5.11	Processual nouns with the suffix -idz.	160
5.12	Nouns containing the once productive agentive suffix -idz.	161
5.13	Gerunds formed with the suffix -ak.	161
5.14	Suffixes used in de-verbal derivation of nouns.	162
5.15	NOUN + NOUN compounds in Shughni.	163
5.16	Genealogical diminutive -bic/-buc.	164
5.17	Verb-based compound nouns with a borrowed (Tajik) present stem.	165
5.18	Occupational compounds with the suffix -iĵ.	166
5.19	NOUN + INF. STEM compounds in Shughni.	167
5.20	Verb-based compound nouns: Summary.	167
5.21	Denominal adjective-forming suffixes.	169
5.22	Denominal adjective-forming prefixes.	170
5.23	Deadjectival adjective-forming suffixes.	172
5.24	Compound adjectives: Examples.	173
5.25	Compound adjectives with gender-distinguishing adjectival components.	174
5.26	Denominal components of compound adjectives with voicing of initial consonant.	175
5.27	Shughni numerals 1-10.	177
5.28	Shughni numerals above 10.	177
5.29	Borrowed Tajik numerals in Shughni.	179
5.30	Contexts/quantities for Shughni numerals (1-10).	180
5.31	Quantities expressed by Tajik numerals.	180
5.32	Contexts requiring Tajik numerals in Shughni.	181
6.1	Personal pronouns.	184
6.2	Fundamental interrogative pronouns.	186
6.3	Complex interrogative pronouns.	191
6.4	Common indefinite pronouns in Shughni.	193
6.5	Demonstrative pronouns and grammatical gender.	201
6.6	Demonstrative pronouns and the precision particle <i>ik</i>	212
6.7	Shughni deictic presentatives.	219

7.1	Proto-Iranian vowels (cf. Cantera 2017: 482).	226
7.2	Historic developments of the Shughni vowel system.	229
7.3	Gender-distinguishing noun pairs.	231
7.4	Gender-distinguishing adjective pairs.	233
7.5	Gender agreement with adjectives and verbs.	237
7.6	Predominantly masculine thematic categories.	242
7.7	Predominantly feminine thematic categories.	243
7.8	Split thematic categories.	244
7.9	Summary: Thematic groups of nouns by gender.	246
7.10	Examples of (masculine) abstract nouns from various etymological sources.	247
7.11	Pairs of FEM~MASC homonyms.	248
8.1	Shughni verb stems: Form, inflection, and usage.	261
8.2	Present suffixes.	262
8.3	Past clitics.	262
8.4	Second-position clitics used in present realis copular constructions.	266
8.5	Verb-stem paradigm of <i>vidow</i> ‘be’.	268
8.6	Present-tense paradigm of <i>vidow</i> ‘be’.	268
8.7	Distribution of the Shughni copula across grammatical environments.	270
8.8	Summary of copula form and distribution.	275
8.9	Types of adverbs and examples.	277
8.10	Suffixed and reduplicated adverbs.	280
8.11	Degree adverbs in Shughni.	282
8.12	Common epistemic adverbs in Shughni.	284
8.13	Deictic adverbs with prepositional locative elements.	287
8.14	Deictic adverbs with suffixal locative elements.	288
8.15	Deictic adverbs of manner.	289
8.16	Deictic temporal adverbs in Shughni: Single lexemes.	290
9.1	Polysyllabic verb stems and their ancient Iranian pre-verb counterparts.	295
9.2	Regular Shughni verbs.	298
9.3	Historical affixes and consonant irregularities in verb stems.	301
9.4	Consonant alternations in verb stems due to regular sound changes.	302
9.5	Common vowels in (non-3SG) present stems.	305
9.6	Common vowels in (non-3SG) present stems.	306
9.7	Common vowels in past and perfect stems.	307
9.8	Common vowels in infinitive stems.	308
9.9	Irregular stem vowels: Summary in four examples.	309
9.10	Stem-shortening in the verb <i>xīdow</i> ‘eat’.	310
9.11	Verbs which undergo present-stem shortening: Type 1.	312
9.12	Verbs which undergo present-stem shortening: Type 2.	312
9.13	Verbs which undergo imperative shortening.	314
9.14	Shortening of past stems.	314
9.15	Shortening of perfect stems.	314
9.16	Gender-distinguishing pairs of past stems.	319
9.17	Gender-distinguishing pairs of perfect stems.	321
9.18	Gender-distinguishing onomatopoeic verbs.	322
9.19	Verbs with parallel irregular and leveled paradigms.	326
9.20	Plural perfect forms.	328

10.1	Summary: Temporal and aspectual references. . . . .	334
10.2	Semantic contributions of <i>=ta</i> . . . . .	344
10.3	Basic grammatical differences between realis and irrealis. . . . .	359
10.4	Imperative verb forms. . . . .	362
10.5	Indicative vs. irrealis verb forms. . . . .	367
11.1	Verbal resultative participles from unaccusative and transitive verbs. . . . .	370
11.2	Adjectival resultative participles in Shughni. . . . .	375
11.3	Intransitive~causative ablaut pairs. . . . .	379
11.4	Intransitive~causative pairs with suffixation. . . . .	382
11.5	Labile infinitive verbs. . . . .	383
11.6	Common verbs used in ‘coercion’ causatives. . . . .	386
11.7	Eventivity and agentivity of causative types. . . . .	389
11.8	Light verbs in Shughni. . . . .	393
11.9	Types of non-verbal components in complex verbs. . . . .	394
11.10	Transitivity alternations in complex verbs. . . . .	395
11.11	Unaccusative complex verbs with <i>čīdow</i> ‘do’. . . . .	397
11.12	Non-present stems of <i>dāktow</i> and <i>lāktow</i> . . . . .	401
11.13	Reflexive verbs in Shughni (examples). . . . .	405
11.14	Accusative-first constructions in Shughni. . . . .	407
11.15	Dative-first constructions with complex verbs. . . . .	409
11.16	Dative-first constructions in Shughni. . . . .	410
12.1	<i>Wh</i> -words used in free-choice free relatives. . . . .	426
12.2	Subordinators <i>ca</i> and <i>di</i> in temporal adverbial clauses. . . . .	430
12.3	Verbs which take complement clauses with the complementizer <i>idi</i> . . . . .	435
12.4	Coordinating conjunctions in Shughni. . . . .	440
12.5	Shughni complementizers. . . . .	441
12.6	Example utterances and <i>wh</i> -echo responses. . . . .	461
12.7	Example utterances and <i>wh</i> -echo responses. . . . .	462
12.8	Focused polar echo questions. . . . .	463
12.9	Development of case in Iranian languages. . . . .	476
12.10	Morphosyntactic alignment in Persian . . . . .	478
12.11	Morphosyntactic alignment in the Pashto past tense. . . . .	480
12.12	Davani verbal suffixes . . . . .	481
12.13	Davani subject clitics . . . . .	481
12.14	Davani past-tense alignment. . . . .	483
12.15	Shughni pronouns. . . . .	484
12.16	Shughni present suffixes . . . . .	486
12.17	Shughni past clitics . . . . .	486
12.18	Shughni past-tense alignment (non-3sg). . . . .	489
12.19	Shughni past-tense alignment (3sg). . . . .	489
12.20	Shughni past-tense subjects and overt clitics. . . . .	490
A.1	Predominantly masculine thematic categories. . . . .	501
A.2	Predominantly feminine thematic categories. . . . .	502
A.3	Split thematic categories. . . . .	503
A.4	Examples of (masculine) abstract nouns. . . . .	504
B.1	Etymology of agreement suffixes and past-tense clitics (cf. Dodykhudoeva 1988: 6-8). . . . .	506



B.2	Consonant alternations in verb stems due to regular sound changes. . . . .	507
B.3	Stem-consonant voicing alternations. . . . .	508
B.4	Present (non-3sg) stem vowels in irregular verbs. . . . .	509
B.5	Irregular third-singular present stems. . . . .	511
B.6	Past stem vowels in irregular verbs. . . . .	512
B.7	Infinitive stems of irregular verbs. . . . .	514



# List of Figures

1.1	The Shughni-speaking Region of Tajikistan and Afghanistan . . . . .	4
1.2	The Shughni-speaking region within Central Asia . . . . .	5
2.1	Extant Indo-European Branches . . . . .	26
2.2	Historical Periods of Iranian Languages (approx.) . . . . .	27
2.3	Map of the Pamir Languages (Kim 2014) . . . . .	42
2.4	Pamir Languages within Eastern Iranian . . . . .	44
2.5	Shughni-Rushani Languages . . . . .	46
3.1	Shughni short and long vowel correspondences. . . . .	61
3.2	Shughni scripts side-by-side . . . . .	92
4.1	Stem formation in three regular verbs . . . . .	109
4.2	Stem formation in three irregular verbs . . . . .	109
6.1	Canonical contexts for the use of proximal demonstratives (Image source: Wilkins 2018) . . . . .	204
6.2	Peripheral contexts for the use of proximal demonstratives. . . . .	205
6.3	Canonical contexts for the use of medial demonstratives. . . . .	206
6.4	Peripheral contexts for the use of medial demonstratives. . . . .	207
6.5	Canonical context for the use of distal demonstratives. . . . .	208
6.6	Decision tree: Non-contrastive uses of exophoric demonstratives . . . . .	210
6.7	Choice of demonstrative degree in pure text deixis. . . . .	218
7.1	Umbrella nouns and thematic categories. . . . .	246
7.2	Part/Whole relations in grammatical gender. . . . .	251
7.3	Part/Whole relations in borrowed Russian lexemes. . . . .	252
8.1	Division of labor in verbal negation prefixes. . . . .	265
8.2	Reversal of degrees deictic in pronouns and adverbs. . . . .	286
8.3	Structure of deictic locative adverbs. . . . .	287
10.1	Relevant time intervals in example (216). . . . .	334
10.2	Present stem encoding relative tense. . . . .	338
10.3	Past stem encoding relative tense. . . . .	346
12.1	Relativization strategies in Shughni externally headed relative clauses. . . . .	423
12.2	Accusative and ergative alignment . . . . .	468
12.3	Split-intransitive alignment . . . . .	470
12.4	(Hypothesized) step-by-step development of ergative alignment in Iranian . . . . .	475



# Glossing abbreviations

-	Affix boundary
1	First Person
2	Second Person
3	Third Person
=	Clitic boundary
A.PTPL	Adjectival Resultative Participle
ABBR	Abbreviated form
ADJ	Adjective
AGT.SFX	Agentive Suffix
ART.SG	Singular Article
ASS.P	Associative Plural
CAUS	Causative
CF	Counterfactual Suffix
CMPR	Comparative
COLL	Collective Plural
COMP	Complementizer
CTR.TOP	Contrastive Topic
DAT	Dative
DEM	Demonstrative
DEO	Deontic modality
DIM	Diminutive
DUB	Dubitative
ECWH	Wh-Echo Question
EEP	Existential Emphatic Particle
EPIS	Epistemic modality
FAC	Factual Enclitic
FOC	Focus
F	Feminine
GER	Gerund
HON	Honorific
IMP	Imperative
INCEP	Inceptive Aspect
INCTAG	Incredulous Tag Question
IND	Indicative
INTR	Intransitive
IPFV	Imperfective Aspect
IRR	Irrealis

LF	Long Form (of Numeral)
LOC	Locative
M	Masculine
NEG	Negation
NMZ	Nominalizer
NUM	Numeral
PEJ	Pejorative
PFV	Perfective
PL	Plural
PREC	Precision Marker
PROG	Progressive Aspect
PRON.EMPH	Emphatic Pronoun
PROSP	Prospective Aspect
PRSV	Presentative
PRS	Present
PST	Past
PTCP	Participle
PURP	Purposive
QUOT	Quotation Marker
REFL	Reflexive Pronoun
REL.ADJ	Relational Adjective
REL	Relativizer
SBJV	Subjunctive
SF	Short Form (of Numeral)
SG	Singular
SIMV	Similative Plural
SUBR	Subordinator
SUP	Superlative
TAG	Tag Question
TOP	Topic
V.PTPL	Verbal Resultative Participle
VOC	Vocative

# Chapter 1

## Introduction

This dissertation provides an overview of several aspects of the Shughni language (ISO 639-3: *sgn*), an Eastern Iranian (Indo-European) language spoken by roughly one hundred thousand people primarily in the mountainous regions of northeastern Afghanistan and southeastern Tajikistan, but also in various diaspora communities throughout the world. Specifically, it examines the variety of Shughni spoken in Khorog, Tajikistan, the capital and largest city of the Gorno-Badakhshan Autonomous Province (GBAO). Where applicable, information is given on other varieties of the language, such as those spoken in Afghanistan and in the regions surrounding Khorog in Tajikistan.

The scholarly value of this dissertation rests on the following three primary objectives: (i) to identify, analyze, and describe areas of the language's grammar which have heretofore not been investigated; (ii) to synthesize and build upon existing research on the language, much of which was published in Russian during the Soviet era and has therefore been hard to access for many scholars outside the former Soviet Union; and (iii) to point out phenomena in Shughni which are not investigated in detail in this thesis and are ripe for future investigation.

Novel contributions of this thesis include, among others, an analysis of definiteness in Shughni nominals, a thorough description of the language's system of triple deixis, particularly as it pertains to demonstratives and adverbs, an analysis of the Shughni tense and aspect system through the lens of a Neo-Reichenbachian framework, and a comprehensive account of morphosyntactic alignment in Shughni from synchronic, diachronic, and typological perspectives. It also builds upon the work of Soviet-era scholars in providing a historical account of several irreg-

ularities found in the modern Shughni verbal system, notably suppletion in infinitive, present, and past stems, as well as the development of syncretism in feminine and plural past and perfect forms.

The coverage of topics is at once broad and thorough, touching upon issues ranging from historical linguistics and effects of language contact to core components of the language's grammar, including phonology, morphology, and syntax. Ultimately, it is my hope that this thesis will at once contribute to the documentation of Shughni and spur future research into un(der)-investigated areas of the language.

The remainder of the Introduction is organized as follows. The first three sections deal with Shughni background. Section 1.1 gives an overview of the **geography of the Shughni-speaking region**; Section 1.2 provides **sociolinguistic and ethnographic information**; and Section 1.3 provides a look at **previous research** on the language. Next, Section 1.4 provides further information on the **impetus for the thesis**, and Section 1.5 lays out the methodology used. Lastly, Section 1.6 provides an outline of the **structure for the remainder of the thesis**.

## 1.1 Geography of the Shughni-speaking region

Shughni is spoken in two neighboring regions: the Badakhshan Province of Afghanistan and the Gorno-Badakhshan Autonomous Province of Tajikistan (Tajik Вилояти Мухтори Кӯҳистони Бадахшон, *Viloyati Mukhtori Kūhistoni Badakhshon*, often referred to as GBAO). Each of these regions has a distinct administrative subregion called 'Shughnon' (Tajikistan) or 'Shughnan' (Afghanistan), in which Shughni is the predominant mother tongue.<sup>1,2</sup> In Tajikistan, Shughni is also the primary mother tongue of those living along the Shahdara River in the Roshtqal'a District to the south of the Shughnon..

In Afghanistan, Badakhshan is one of thirty-four provinces (Dari ولايت *wilāyat*, ultimately from Arabic وِلَايَة *wilāya* 'province; state; governorate'), and within Badakhshan, Shughnan is one of thirty districts (Dari ولسوالي *wuluswālī*). On the Tajik side, the Gorno-Badakhshan Autonomous Province is one of four major administrative regions in

---

<sup>1</sup>The abbreviation 'GBAO' – generally pronounced as an acronym, rather than an initialism – comes from the Russian name for this province: Горно-Бадахшанская автономная область (*Gorno-Badakhshanskaya avtonomnaja oblast*).

<sup>2</sup>The difference in spelling between the Afghan (Shughnan) and Tajik (Shughnon) districts corresponds to the Persian long *ā*, which is rendered as <ā> in Dari but as <o> in Tajik. The same can be said of Badakhshan (Afg.) and Badakhshon (Taj.), Rushan (Afg.) and Rushon (Taj.), etc. Note that the spelling of this district, particularly in Afghanistan, is also occasionally with an *i* rather than an *u*, hence *Shighnan*.



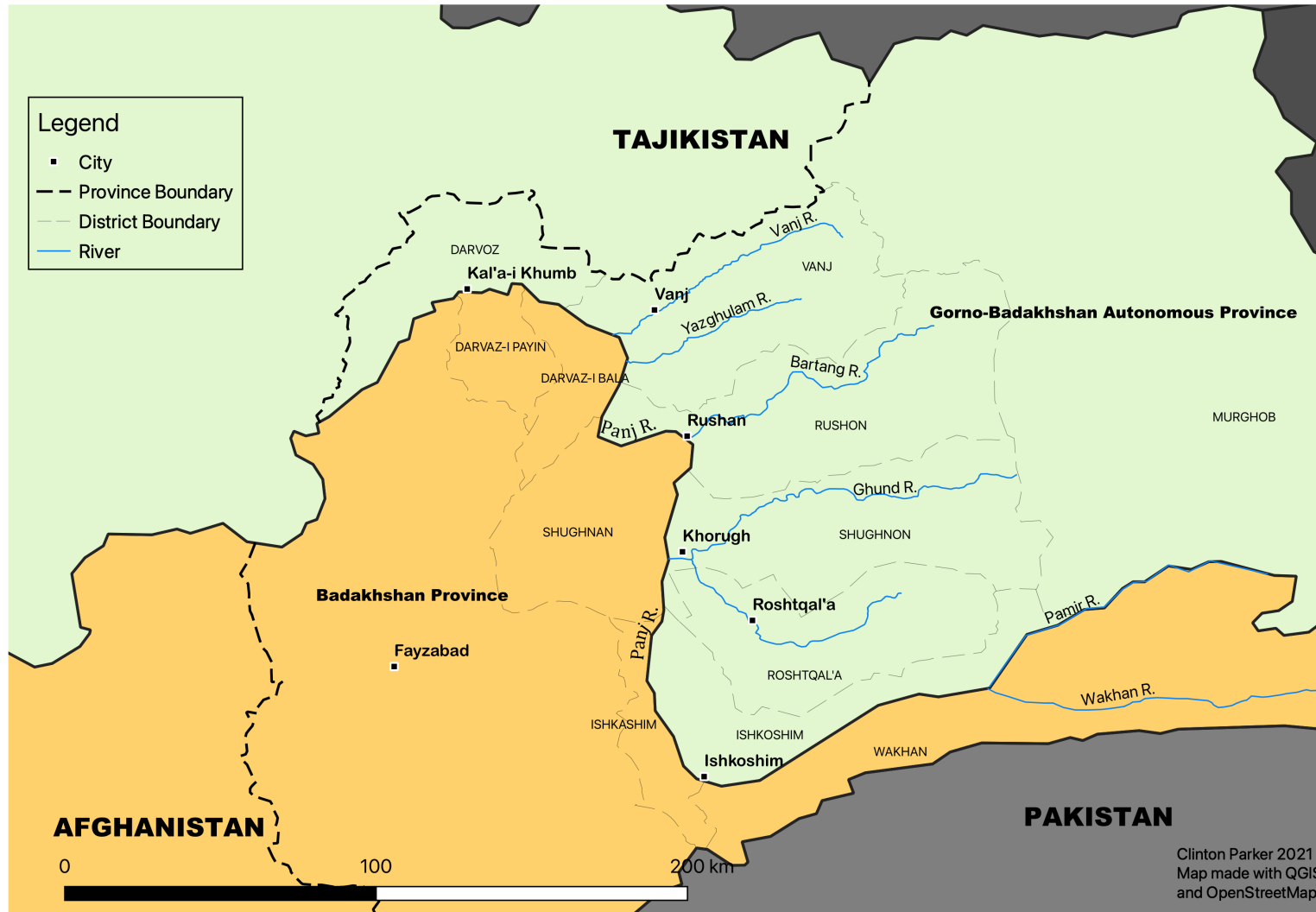
Tajikistan known as *viloyat* (Tajik вилоят, of the same origin as the Dari above). GBAO, like Afghan Badakhshan, is further subdivided into smaller districts, which in Tajikistan are known as *nohiya* (Tajik ноҳия; cf. Farsi ناحیه, ultimately from Arabic نَاحِيَة *nāḥiya* ‘zone; region’). These include six smaller districts distributed roughly north to south and each centered upon the tributaries of the Panj, as well as the larger district of Murghob, which forms the entire easternmost portion of GBAO, and the entire eastern portion of Tajikistan more generally. This region of Tajikistan is characterized by many remote valleys separated from one another by formidable mountain ranges, many of which have peaks reaching over 6,000 meters in height. Each of the smaller districts in GBAO corresponds to one or two such valleys, while Murghob is primarily high-altitude plateau. Thus, the Vanj and Yazghulam Rivers run through the Vanj District; the Bartang River through Rushan; the Ghund River through Shughnan; and so on. A map showing this region of Tajikistan and Afghanistan is given in Figure 1.1.

This region is of noteworthy linguistic diversity, and each valley is generally home to a specific linguistic variety. In some cases, the name of a particular linguistic variety corresponds to the name of the river along which it is spoken; hence, the Bartangi language is spoken along the Bartang River; Yazghulami is spoken along the Yazghulam River; the Old Wanji language, now dormant, was spoken along the Vanj River; and so on.

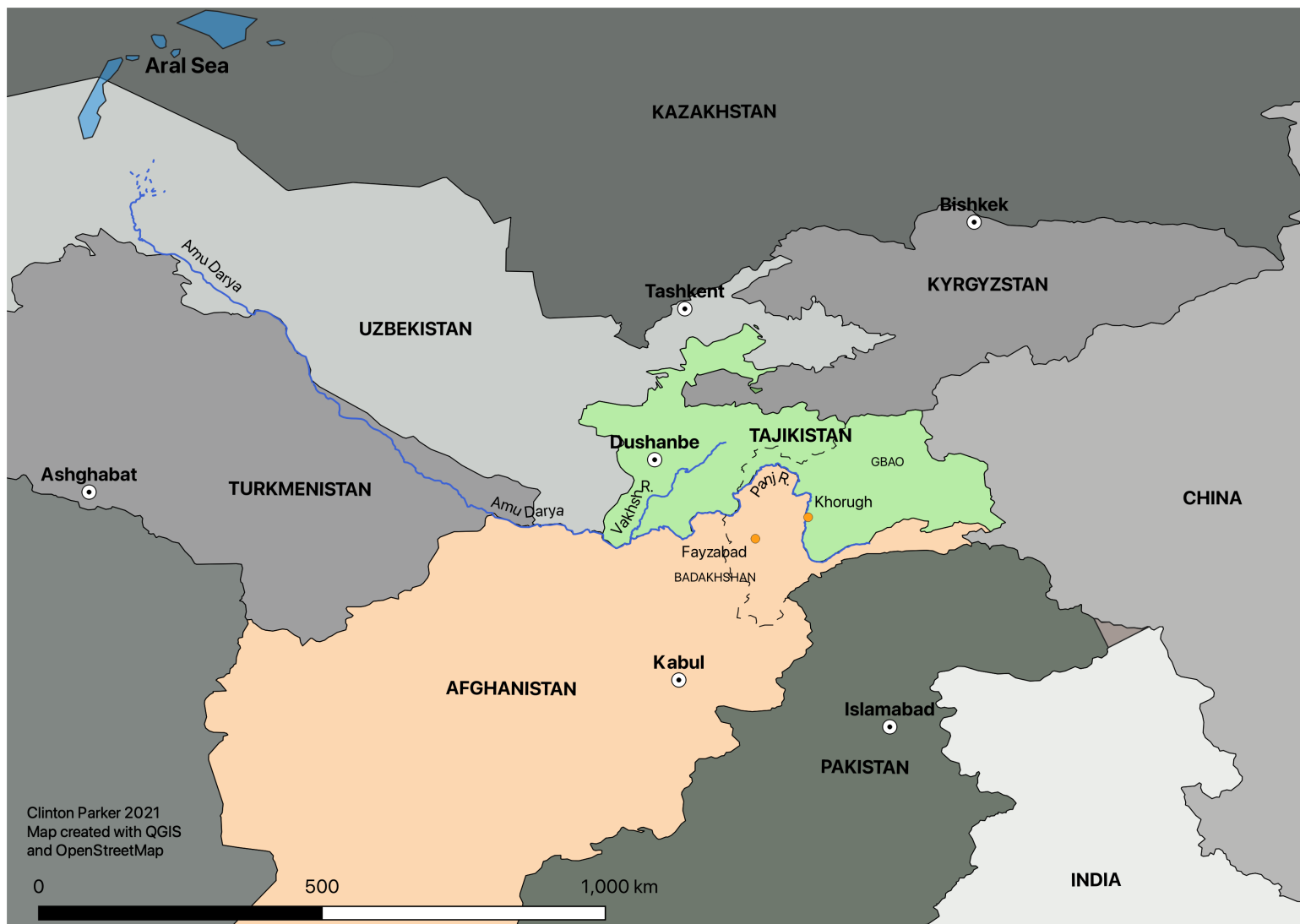
A geographical centerpiece of the Shughni-speaking region is the Panj River, which flows out of the Wakhan Corridor and marks the border between Afghanistan and Tajikistan for its entire length of nearly a thousand kilometers. The Panj first runs a southwesterly course from its source at the confluence of the Pamir and Wakhan Rivers, before bending toward the north near the village of Ishkashim, about a hundred kilometers from its source. Roughly another hundred kilometers past Ishkashim, it passes by the administrative center of GBAO, Khorog, where it is met by two important tributaries: the Shahdara from the southeast and the Ghund from the northeast. From Khorog, the Panj flows north to the town of Rushan, where the Bartang River runs into it from the east. After passing a few more tributaries in GBAO, notably the Yazghulam and Vanj Rivers, whose valleys are home to the Yazghulami and now extinct Old Wanji languages, respectively, the Panj resumes a westerly and southwesterly course for several hundred more kilometers before joining the Vakhsh River near the border with Uzbekistan. This confluence marks the start of the great Amu Darya (Oxus), which flows generally northwest through Turkmenistan and Uzbekistan to the Aral Sea. A birds-eye view of the Shughni-speaking region, showing the course of the Panj and Amu Darya, all as well as the countries surrounding the region, is shown in the map in Figure 1.2.

In Tajikistan, the heart of Shughni-speaking territory is the capital of GBAO, Khorog, a town of approximately

**Figure 1.1:** The Shughni-speaking Region of Tajikistan and Afghanistan



**Figure 1.2:** The Shughni-speaking region within Central Asia



20,000 inhabitants situated at 2,000 meters of elevation. The language is also spoken all along the valley of the Shahdara River, which flows into Khorog from the southeast, as well as along the valley of the Ghund River, which flows into Khorog from the northeast. In addition, the Shughni-speaking region includes a roughly sixty-kilometer stretch of the right bank (i.e. Tajik side) of the Panj River, beginning around the village of Darmorakht, south of (above) Khorog, and ending around the Bajuw Valley, north of (below) Khorog (e.g., Payne 1980: 417). Some of the larger Shughni-speaking villages located on the Tajik bank of the Panj include Pish, to the south of Khorog, and Porshinev and Sokhcharv, to the north. (For a more detailed map of this area than that provided in this section, see Figure 2.5.)

A few distinct Shughni dialects have been distinguished in the literature (e.g., Edelman & Dodykhudoeva 2009b: 788; Payne 1980: 417). The principal Shughni variety is that spoken in and around Khorog, as well as along the Panj River. For the remainder of this thesis, I refer to this dialect as ‘Shughni proper’. Other noteworthy varieties of Shughni include Bajuwī, spoken in the Bajuw Valley north of Khorog, near the northern limit of the Shughni-speaking territory where it borders with that of Rushani, and Barwozi, a severely threatened variety spoken in the highest part of the Shahdara Valley. There is also some variation in Shughni speech throughout each of the Ghund and Shahdara Valleys. Edelman & Dodykhudoeva (2009b: 788) note that these dialects vary somewhat considerably due to the presence of immigrants from other parts of the Pamir, including the valleys of Rushan and the Wakhi-speaking region.

## 1.2 Sociolinguistic information

By most estimates, the number of Shughni speakers is roughly 100,000 (e.g., Edelman & Yusufbekov 2000; Edelman and Dodykhudoeva 2009b), a figure which includes the speakers of Afghan and Tajik Badakhshan, of which about 20,000 are residents of Khorog, as well as diaspora communities elsewhere in these countries and in other parts of the world. Significant communities of Shughni speakers are found in Kabul, Dushanbe, and other areas of Tajikistan and Afghanistan outside Badakhshan, as well as in Russia, Canada, and the United States.

Although it may be said that the vitality of Shughni is threatened to some extent due to external political and economic pressure, most notably from Tajik and Russian, the language is still transmitted orally from generation to generation and is the primary means of communication for ethnically Shughni families both within and outside

Badakhshan. Shughni, which does not have official status in Tajikistan, together with Russian and Tajik, both of which do have official status in the country, is one of the languages used for communication among speakers of different Pamir languages, and can therefore be thought of as a *lingua franca* of the region. Additionally, there are ongoing efforts led by native-speaker academics, community members, and outsider academics to preserve the language and encourage its widespread usage in the community (see Dodykhudoeva & Ivanov 2008 and references therein).

According to Mueller et al. (2010b), Shughni fulfills complementary roles with Tajik and Russian in the lives of Shughni speakers in Tajikistan, though the domains in which each language is used may overlap with one another to some degree. These roles can be summarized as follows. Shughni is generally the sole language of the home, for interactions among Shughni-speaking community members, and among colleagues in certain workplaces. Although the law in Tajikistan requires primary and secondary education to be taught in Tajik, Shughni is commonly used in the classrooms of lower-level primary schools, where most students do not have the necessary level of Tajik to understand concepts taught in the language. Tajik and Russian, on the other hand, are used for interactions with those who do not speak Shughni and in official (governmental) workplaces. They are also the primary languages used in higher education and in the media. It is worth noting, however, that English is a language of increasing importance in the Pamir region, where it is used as the medium of instruction at the University of Central Asia in Khorog and in several of the non-governmental organizations operating in the region. (For a discussion of the attitudes of Shughni speakers toward the various languages spoken in the Pamir region, see Bolander 2016.)

### 1.3 Previous research on Shughni

Although the Pamir languages have received relatively little scholarly attention in comparison with more widely spoken languages of the Iranian group, such as Persian, Kurdish, and Pashto, a rich and diverse body of work has nonetheless developed for these languages over the course of the past century and a half. Among the Pamir languages, Shughni is perhaps the most widely studied, with works on various aspects of the language, from topics of historical linguistics to phonetics, morphology, lexicography, sociolinguistics, and others. In this section, I provide an overview of previous research on Shughni. For ease of presentation, I divide these works into three categories: (i) early works on Shughni, up until approximately 1950 (1.3.1); (ii) studies on specific linguistic

aspects of Shughni (1.3.2); and (iii) grammatical sketches and dictionaries (1.3.3).

### 1.3.1 Early works on Shughni

A comprehensive overview of the earliest research activity on Shughni and other Pamir languages is given in Oranskij (1975). According to him, the Pamir languages are mentioned in the writings of travelers to the region as early as the first half of the nineteenth century, but they do not appear in scholarly works until the 1870s, beginning with Robert Shaw's (1876) description of Wakhi and Sarikoli and then of Shughni (Shaw 1877). Shaw's work on the language, published in the *Journal of the Asiatic Society of Bengal*, includes a Shughni story accompanied by a translation, some commentary on the language's grammar, and a word list. Further material on Shughni was gathered by Ivanov (1884) and then published by Salemann (1895), who later made a report on his own trip to Central Asia (Salemann 1898). Also in the late 19th century, a number of linguists used the available data on Shughni to make historical linguistic and etymological commentary on the language (e.g., Tomaschek 1880; Geiger 1887; and Geiger and eds.).

Research on Shughni continued in the first few decades of the twentieth century in the form of further expeditions to the Pamirs, first in 1914 by the French linguist Robert Gauthiot and Russian linguist Ivan Ivanovich Zarubin (see Gauthiot and Zarubin 1914), and then in 1926 by the Norwegian linguist Georg Morgenstierne (see Morgenstierne 1926). Gauthiot and Zarubin's expedition was cut short by the onset of the First World War, which ultimately claimed the life of Gauthiot. Zarubin, however, resumed his work on the Pamir languages after the war. In addition to producing prolific work on Bartangi and Roshorvi (e.g., Zarubin 1926; 1927; 1930; 1937), he published a collection of texts along with a dictionary of Shughni (Zarubin 1960). Other notable scholars who worked on the Pamir languages during this period include the Irishman George Abraham Grierson (e.g., Grierson 1920; 1921); the German scholar Wolfgang Lentz (e.g., Lentz 1933); and the Swedish linguist Hannes Sköld (e.g., Sköld 1936).

### 1.3.2 Works on specific aspects of Shughni

In the decades following the Second World War, as work on the Pamir languages garnered support within Soviet academic circles, and as a number of native speakers of Pamir languages pursued research on their mother tongues, we begin to see more and more works on specific aspects of Shughni and other Pamir languages.

A description of Shughni phonology, together with a study on vowel length in the language, is given by Valentina S. Sokolova (1953: 85-106; 135-139). In the same study, she offers descriptions of other varieties of the Shughni-Rushani subgroup and of other Eastern Iranian languages, including Yazghulami, Wakhi, Ishkashimi, Yaghnobi, and Ossetic. Work on the history of the Shughni sound system is taken up Rakhim Dodykhudoev (1962; 1963; 1964a; 1964b). More recently, Mingzhen Bao (2013) published a study on the glottalization of Shughni vowels in word-initial position.

Works which look more broadly at the genetic classification of Shughni, the Shughni-Rushani languages, and the Pamir languages more generally, include those by Joy Edelman, Tatiana Pakhalina, and Valentina Sokolova. Sokolova (1967) establishes the genetic relationship between Yazghulami and the Shughni-Rushani group, and Sokolova (1973) examines possible genetic relations between Yazghulami, Shughni-Rushani, and other Pamir languages, particularly Munji. Pakhalina (1983; 1989) looks at the historical phonetics and historical morphology of the Pamir languages, respectively. Edelman (1976) examines the history of Shughni-Rushani and Yazghulami demonstrative pronouns, the history of the consonant systems of Pamir languages (1980a), and the notion of a substrate language in the history of the Central Asian Sprachbund (1980b). Edelman (1986; 1990; 2009) is a three-volume series on comparative phonology, morphosyntax, and lexicon of Eastern Iranian languages, respectively. Leila Dodykhudoeva (1988) provides a look at Shughni verbs from a historical perspective, in particular through the lens of the reconstructed verbal systems of Old Iranian languages. Finally, Antje Wendtland (2009) gives an up-to-date overview of thought on genetic relations among Eastern Iranian languages, including the Pamir languages.

Research on the dialectology of Shughni is found primarily in the works of Dodkhudo Karamshoev, who provides a description of the Bajuwi dialect of Shughni (1963a), as well as a look at some specific features of Bajuwi – nominals (1962a) and vowels (1963b). Karamshoev also looks at the relation of Bajuwi to Shughni proper (1962b) and at the division of Shughni dialects more broadly (1966).

Investigation into specific aspects of Shughni grammar includes a study on the pronominal system of the language (Alamshoev 1994) and a study on grammatical gender in Shughni-Rushani (Karamshoev 1978). In addition, the morphosyntactic alignment system of Shughni and other Pamir languages has received a considerable amount of attention in recent years. Notable works on this topic include those by Payne (1980) and Wendtland (2009), both of whom offer general overviews of alignment in the Pamir languages, with theoretical analyses of vestigial ergativity in Shughni provided by Stump & Hippisley (2011) and Parker (2020). Other works on specific topics in Shughni

grammar include studies on cleft sentences (Barie 2009), deixis (Mueller 2014), a special class of verbs meaning ‘to pour’ (Armand 2022), and variation in the position of second-position (Wackernagel) clitics within the clause (Chistiakova 2022). Note that the latter article is part of an ongoing project at the Higher School of Economics (HSE) in Moscow. A number of other works on various aspects of the language, including pronouns, possessive expressions, morphosyntactic alignment, and grammatical gender, are under way as a part of this project.<sup>3</sup>

### 1.3.3 Grammatical sketches and dictionaries

Since around the middle of the twentieth century, several grammatical descriptions and dictionaries have been published on Shughni. In many cases, grammatical descriptions of the language are quite brief and give only cursory overviews of the topics discussed, hence the impetus for the present study.

Grammatical overviews of Shughni and other Pamir languages often appear in edited volumes constituting collections of grammatical descriptions of languages in a particular region or genetic group. Russian-language descriptions include those on the entire Shughni-Rushani group by Sokolova (1966) in *Языки народов СССР (Languages of the Peoples of the USSR)*; Edelman (1987) in *Основы иранского языкознания (Fundamentals of Iranian Linguistics)*; and Edelman (2000) in *Языки мира (Languages of the World)*. Russian-language descriptions of Shughni itself include Edelman & Yusufbekov (2000) in *Языки мира* and Dodykhudoeva (2003) in *Языки российской федерации и соседних государств (Languages of the Russian Federation and Neighboring States)*. Pakhalina (1969) gives a rather thorough overview of the Pamir languages generally, and Karamshoev (1963a), mentioned above, provides a description of the Bajuwi dialect.

English-language descriptions include Edelman & Dodykhudoeva (2009a) in *The Iranian Languages* and Payne (1989) in *Compendium Linguarum Iranicarum*, both of which examine the Pamir languages generally, as well as Nawata (1979) and Edelman & Dodykhudoeva (2009b) on Shughni specifically.

Tajik-language descriptions of Shughni include Bakhtibekov (1979) and the pedagogical grammar of Alamshoev (2015). In addition, Karamshoev et al. (2019) have published a pedagogical grammar of the language written entirely in Shughni.

---

<sup>3</sup>Details can be found at the following link: <https://ling.hse.ru/pamir/seminar>.



Besides Zarubin's (1960) *Шугнанские тексты и словарь* (*Shughni Texts and Dictionary*), Shughni dictionaries have been compiled by Morgenstierne (1974) (an etymological dictionary of the Shughni-Rushani group), Karamshoev (1988a), and Alamshoev & Alamshoev (2020). A searchable online version of Karamshoev's dictionary has been put together by researchers at HSE as part of the same project mentioned above (see Makarov & Melenchenko 2021 for the website itself and Makarov et al. 2020 for a description of the project).

## 1.4 Impetus for the present study

Although not as widely studied as other Iranian languages such as Persian or Kurdish, Shughni has received a considerable amount of scholarly attention over the past century and a half. It has been a subject of interest for researchers in various parts of the world, including not only Tajikistan and Afghanistan, where it is the mother tongue of a sizable portion of the population, but also Russia, western Europe, and more recently in North America (see Section 1.3 for a more detailed look at previous research on Shughni). Academic work on the language has been galvanized by both native speakers and non-native speakers alike. And although the language is commonly described as unwritten, Shughni speakers in both Afghanistan and Tajikistan have gone to great lengths to develop orthographies and to produce literature in the language.

This dissertation, therefore, does not constitute the first attempt at a description or analysis of Shughni grammar, as this has been undertaken by many scholars before. Nor is it a last-minute effort to document and preserve the language before it becomes dormant, as most signs point toward the ongoing vitality of the language in the communities where it is spoken as a mother tongue. It is, however, the most up-to-date, extensive, and comprehensive description of the language at this time. It is also the first English-language description of Shughni of any length. Existing descriptions of the language – even those in Russian and Tajik – are typically of journal-article length (see, for instance, Edelman & Dodykhudoeva 2009b, for an example of such a recent description of the language). In this dissertation, I expand both the breadth and depth of content in a way that has not been done before.

In many parts of this dissertation, I examine aspects of Shughni grammar which have not yet been described. In others, I look at previously described facets of the language in greater detail than before, or I provide alternative analyses to those provided before. Hence, one of the major contributions of this study is to bring attention to previously undescribed aspects of Shughni, to fill in gaps in existing descriptions, and to identify areas for future

research when they are beyond the scope of the present study. At the same time, the dissertation contributes to the documentation of Shughni insofar as it relies on novel linguistic material (both elicited and heard in natural speech) to draw conclusions about the language's grammar.

Importantly, not all the information presented here is novel. A second, and in my opinion equally important goal of the dissertation is to bring together existing research on the language. A noteworthy source of motivation for this study is the fact that, despite the considerable breadth and value of previous research on the language, much of the extant work on Shughni has been and continues to be carried out by disjointed groups of scholars who, because of barriers of various kinds, not least of which are language barriers, have not been in consistent, fruitful communication with one another. Thus, for instance, academics working on Shughni in Tajikistan and Russia may not collaborate consistently with those in the Americas, due primarily to language barriers and distinct platforms for disseminating research. Moreover, a significant portion of the existing body of work on Shughni is written in Russian or Tajik, which means it is relatively inaccessible to many scholars whose native and/or main professional language is English. With these things in mind, this dissertation seeks not only to bring together previous research on Shughni in such a way that it is accessible to English-language scholars, but also to promote communication and collaboration between researchers based in various parts of the world.

In addition to remarks on this dissertation's scholarly contribution, a couple notes on its limitations are in order here. As a descriptive grammar, this study is meant to provide a snapshot of modern Shughni grammar, a goal which, for various reasons, has proven to be somewhat idealistic.

First and foremost, there is no standardized variety of Shughni. As is the case with many other languages across the world, what we might refer to as a single 'Shughni language' is in fact a multitude of closely related linguistic varieties spoken across neighboring regions of Afghanistan and Tajikistan. Particularly complicating is the fact that speakers of the language are distributed somewhat evenly in number between Afghanistan and Tajikistan, and there is generally little sustained contact between speakers in these regions. While the Shughni varieties in Tajikistan continue to be influenced primarily by Russian and Tajik, those in Afghanistan are more prone to influence from Dari. Moreover, even within these countries, and in Tajikistan in particular, speakers of Shughni have been spread out across remote valleys separated by impassable mountains for quite some time. Hence, there have been multiple distinct dialects of Shughni identified even within Tajikistan. As mentioned above, I focus primarily on the variety of Shughni spoken in Khorog, Tajikistan, but I make reference to the features of other varieties where applicable.

Beyond dialectal variation, Shughni speech differs considerably across generations and across registers. It is here that one notes the influence of Russian and Tajik, which are the languages of mainstream education and media and which have official status in Tajikistan. It is not uncommon to encounter Shughni words in dictionaries and other publications from a few decades ago which have now become antiquated or even forgotten among the younger generations. Importantly, this phenomenon is not restricted to niche terms, such as those for modern technology, but is also found for such commonplace nouns as ‘table’ and such commonplace verbs as ‘decide’. Although I do not treat generational variation in great detail in this dissertation, I point out variation across generations and registers where possible, and simply note that this is a ripe topic for further investigation.

Finally, it is worth mentioning that although much of the linguistic phenomena described here are treated in more detail than ever before, there is often even more to be said about them that goes beyond the scope of the dissertation. Indeed, as I will continue to note, most of the topics discussed here – even down to individual morphemes and grammatical phenomena – could constitute the topic of an entire dissertation. As such, it is my hope that this work provides not only a reference for the current state of our knowledge on the language, but also a basis and source of inspiration for future investigation into a variety of aspects of the language.

## 1.5 Methodology

Data for the present study come from a wide range of sources, of which three basic types can be distinguished:

- (i) data elicited during fieldwork sessions with native speakers;
- (ii) naturally occurring speech (both in recorded stories and in conversations with native speakers); and
- (iii) data taken from previous publications on the language.

Each type of data is described, along with relevant issues, in the following three subsections.

### 1.5.1 Data from fieldwork sessions

The majority of example sentences in the thesis come from data elicited during fieldwork sessions. This subsection first looks at the types of elicitation methods used (Section 1.5.1.1) and then provides information regarding the

consultants whom I worked with (Section 1.5.1.2).

### 1.5.1.1 Elicitation techniques and examples

Data from fieldwork sessions were garnered using elicitation techniques including translation tasks, grammaticality judgments, and acceptability judgments (see, e.g., Payne 1997; Matthewson 2004; Bower 2008, and references within these works on various elicitation methods). Initial work on a particular topic was often carried out with a questionnaire containing a series of *translation tasks*, in which the consultant translated English or Russian utterances targeting a specific linguistic form or phenomenon into Shughni, given a specific context. This was often followed up by tasks involving *grammaticality judgments*, in which the consultant indicated whether a minimally varied version of the initial utterance in Shughni was possible in a specific context. In many cases, particularly when investigating topics such as tense, aspect, and mood (TAM), as well as information structure, it was necessary to determine the kind of contexts in which a grammatical utterance was acceptable. Here, *acceptability judgments* were used to distinguish the kinds of contexts in which a type of utterance was felicitous from those in which it was infelicitous.

A typical line of questioning during elicitation is given in the hypothetical exchange in (1). Utterances by the consultant are labeled with “SC” (for “Shughni Consultant”), and those by me are labeled “CP”. The area of grammar being investigated is information structure and word order, and the questions seek to determine the types of contexts in which particular word orders are possible. I begin here with a translation task (accompanied by a context) to establish a baseline sentence from which to vary word order. I know (from previous publications and my own observations) that this is the predominant word order in the language. Subject, object, and verb are labeled with subscript S, O, and V, respectively.

#### (1) Example elicitation session 1: Information structure and word order

**CP:** Let’s say a group of friends has just finished a meal in a restaurant, and they are splitting the bill by determining who ate what. One person is doing the calculations, and they want to know what Nekruz ate. They ask, ‘What did Nekruz eat?’. How would they say that?

**SC:** They would say, *Nekrūz<sub>s</sub> = i čīz<sub>o</sub> xūd<sub>v</sub>*.

**CP:** OK, and then someone responds and says, ‘Nekruz ate the noodles.’ How would they say that?

**SC:** *Nekrūz<sub>s</sub> = i lapšā<sub>o</sub> xūd<sub>v</sub>*.

**CP:** What if they said it like this: *Lapša<sub>o</sub>=yi Nēkrūz<sub>s</sub> xūd<sub>v</sub>*. Would that work as a response to the question *Nēkrūz=i čīz xūd?*

**SC:** No, it doesn't work there. That would sound strange.

**CP:** Let's say, for example, Maryam is correcting Alik, who is saying "Clinton ate the noodles". Could Maryam correct Alik by saying, *Nay, lapša<sub>o</sub>=yi Nēkrūz<sub>s</sub> xūd<sub>v</sub>?*

**SC:** Yes, it works in that context.

The approach in many elicitation sessions was informed by a publication or group of publications designed specifically for the investigation of a particular topic. Elicitation sessions on information structure, for example, were designed with the help of the discussion in Aissen (to appear). Beyond information structure, the examination of demonstratives was aided by Wilkins' 2018 questionnaire on demonstratives, and the investigation of TAM by the information in Cover 2015 and Cover & Tonhauser 2015.

### 1.5.1.2 Consultants

Over the course of my graduate career and during the preparation of this thesis, I have worked with three primary language consultants, all three of whom are native speakers of Shughni and also speak Persian to a near-native level. Two of these consultants, both of whom are young women in their twenties, were born and raised in Khorog, have remained in Tajikistan for the vast majority of their lives, and also speak Russian to near-native levels. The other consultant, a man aged over 50, was born and raised in the Shughni-speaking region of Afghanistan's Badakhshan Province but has lived in Montreal for over a decade. Because the present study focuses on a description of the variety of Shughni spoken in Khorog, Tajikistan, it was verified that the sentences provided by the consultant from Afghanistan were grammatical in the Shughni dialect of Khorog.

In addition to the three primary consultants, larger groups of speakers were recruited for fieldwork on certain phenomena. For instance, twelve speakers were consulted in the investigation of the semantics of morphological causatives versus syntactic causatives. Other aspects of the grammar, including the deictic system and information structure, were investigated in similar ways. All additional consultants (i.e. beyond the three primary consultants I worked with) were born and raised in Tajik Badakhshan and were between 18 and 55 years of age. At the time of my fieldwork, some were living in Badakhshan, some in Dushanbe, and some outside Tajikistan (e.g., in Europe).

### 1.5.2 Data from natural speech

Data from natural speech used in this thesis are generally of two kinds. First, some data come from interviews which I recorded during the fall of 2019 in Khorog and transcribed and translated with the assistance of language consultants. And second, other data come from conversations with native speakers of Shughni where I was either a participant or a witness. In such cases, I would hear a noteworthy linguistic form, make note of it along with the context in which it was uttered, and later confirm with a consultant that the utterance in question was indeed well-formed and felicitous (and that it was not, for instance, a slip of the tongue on the part of the speaker or misheard by me). If needed, I would then follow up with further questions about its form and use. An example of this type of approach is given in (2), where I ask about a suffix-like element *ik* which I ultimately analyze as an echo-question particle.

#### (2) Example elicitation session 2: Echo question particle *ik*

**Initial utterance and context:** One day, while watching television at home with some Shughni-speaking friends, someone yells something from another room in the house. Their words are hard to make out, and the person next to me yells back *čīz ik*. I haven't heard this particular sequence before, but I know *čīz* means 'what' and I know *ik* to be used as a diminutive morpheme. Over the next few weeks, I begin to notice people saying *čīz ik* more and more, so I decide to inquire about it during an elicitation session.

**CP:** I often hear people say *čīz ik*. Have you heard this before?

**SC:** Yes. People say that often.

**CP:** If someone yells something from another room, and you haven't heard what they said, could you say *čīz ik*?

**SC:** Yes, exactly.

**CP:** So it's like when you're asking for clarification?

**SC:** Yes.

**CP:** Can you think of other words with *ik* besides *čīz*?

**SC:** Yes, you can say, for example, *cawaxt ik*, meaning 'when' or *tar kā yik*, meaning 'to where'.

**CP:** And you would be saying these after you heard a time or a location but want to clarify?

**SC:** Yes, that's correct.

**CP:** For example, if someone says, 'I'm going to school', and they also say when they will go, but you don't hear what they say. Could you say *cawaxt ik* in this context?

**SC:** Yes.

A discussion similar to the one in (2), together with follow-up elicitation sessions, led to the analysis of *ik* as a

*wh*-echo particle. Not all such discussions were fruitful, however. In some cases, my consultants indicated that the form I heard was ungrammatical and that I likely either misheard it or it was a speech error on the part of the speaker. Note that all sentences taken from naturally occurring speech – whether heard in recorded interviews or everyday conversations – were verified with native speakers to ensure that they were indeed grammatical (and do not contain speech errors or marginal constructions).

### 1.5.3 Data from previous publications

Finally, a number of example sentences come from previous publications on Shughni. The majority of these come from Soviet-era publications, especially Dodkhudo Karmshoev's (1988a) three-volume dictionary and two-volume work on gender (Karmshoev 1978; 1986), but also from certain more recent publications, including short descriptions of Shughni such as Edelman & Yusufbekov 2000 and Edelman & Dodykhudoeva 2009b.

Unless otherwise noted, example sentences come from my own fieldwork; those from other sources are cited appropriately. Note that I do not distinguish between examples elicited during fieldwork sessions and those heard in naturally occurring speech. Nonetheless, as mentioned above, all examples which come from naturally occurring speech were verified with my language consultants.

## 1.6 Overview of thesis

This thesis is divided into four sections, that is, groups of chapters which share related topics or a common aim. The first group of chapters, which consists of the following four chapters, deals with preliminaries and sets up the detailed presentation of Shughni grammatical phenomena which follows. First, **Chapter 2** looks at historical and genealogical considerations with the goal of locating Shughni within the Indo-European language family. In doing so, this chapter provides detailed information on both bygone Iranian languages, including attested and unattested Old and Middle Iranian languages, as well as modern Iranian languages from both the Eastern and Western groups. A key goal of this chapter is to clarify the notion that the Pamir languages (and arguably, Eastern Iranian more generally) constitute a Sprachbund or linguistic area, rather than a stand-alone genetic subgroup. The next two chapters deal with the sounds of Shughni. **Chapter 3** presents the fundamentals of Shughni phonology, including

consonant and vowel phonemes, along with important allophones and phonological processes. This chapter also discusses Shughni orthography, examining Arabic-, Cyrillic-, and Latin-based scripts which have been developed for the language and introducing the orthography to be used in this thesis. The last chapter of the preliminary group is **Chapter 4**, which offers a concise presentation of the fundamentals of Shughni grammar, including general typology, constituent order, and basics of nominals and verbs.

The second group of chapters turns to topics in Shughni nominals. **Chapter 5** presents the content and order of elements in noun phrases and discusses nominal inflection, derivation, and (in)definiteness. **Chapter 6** then looks at pronouns and demonstratives and includes a detailed presentation of the language's triple deictic system. The final chapter in the second section is **Chapter 7**, which examines grammatical gender in the language, a complex topic which is linked to the etymology, phonological shape, and meaning of nouns. It will be seen here that Shughni differs markedly from most European languages with grammatical gender in that it exhibits a semantically-based system, where the most important factor in determining the gender of a Shughni noun is not its phonological shape, but rather its meaning.

The third section includes four chapters on various aspects of the verbal system. **Chapter 8** can be thought of as an introduction to verbs. It describes the fundamental building blocks of the verbal system, including the concept of verb stems, including the copula, as well as verbal inflection and modification. **Chapter 9** then puts forth a definition of regular and irregular verbs and deals with a number of issues within this topic. On the one hand, this chapter draws from the work of Soviet-era linguists in discussing the historical linguistic factors which have led to irregularities in the verbal system. On the other, it presents a number of topics such as the shortening of verb stems, which, to the best of my knowledge, have not been discussed in any previous literature. **Chapter 10** turns to issues in tense, aspect, and mood (TAM). It takes a neo-Reichenbachian approach to understanding tense and aspect, and it uses this approach to account for several types of mismatch between a verb's stem and the tense it denotes, such as when a past stem is used to refer to a future event. The last chapter in this group is **Chapter 11** on argument structure and transitivity. Here, two types of resultative participles (used in passive-like constructions) are examined, along with causative constructions, complex verbs, and reflexive verbs. This chapter also includes the description of a topic which has not been examined in the literature in any detail, if at all, namely oblique-first constructions, in which neutral word order involves an oblique argument in first position and in many cases a nominative theme in second position. These constructions are varied in their form and semantics, but are generally similar to dative-first constructions in languages such as Icelandic in that the nominal argument is less agentive



than a prototypical subject.

The final section consists only of **Chapter 12**, which addresses a number of syntactic phenomena which are not addressed explicitly in the preceding chapters. Specifically, this chapter looks at complex clauses, including relative, adverbial, and complement clauses, as well as various means for coordinating clauses. It also examines question formation and provides a description of another novel topic in the literature on Shughni, namely that of echo questions. And finally, this chapter offers a detailed description of morphosyntactic alignment in Shughni. By drawing from what is known about the history of alignment in Iranian, the discussion here couches Shughni alignment, including a non-canonical (split-intransitive) pattern found only in one small corner of its grammar, within the broader context of alignment systems in Iranian more generally.

**Chapter 13** revisits the main points of the thesis, discusses areas in which further research is most needed, and offers concluding remarks.



## **Part I**

# **Preliminaries**



## Chapter 2

# Historical and genealogical considerations

In this chapter, I outline some of the fundamental concepts surrounding the history and genealogical relations of the Iranian languages with the intent of demonstrating, to the extent possible, where Shughni falls within the Iranian branch and, more broadly, the Indo-European family. The discussion here incorporates both bygone Iranian languages and modern languages, and we will see that the genealogical status of Shughni, like that of many Iranian languages, is still not fully understood. These subjects are treated in considerable detail, but a concise summary is given in Section 2.6.

For more broad coverage of the Iranian languages, including an overview of both ancient and modern Iranian languages, the reader is directed to the following works: the *Compendium Linguarum Iranicarum* (Schmitt 1989b), an edited volume of works describing Old, Middle, and Modern Iranian languages, with entries in German, French, and English; *The Iranian Languages* (Windfuhr 2009b), a similar work with entries only in English; and the *Encyclopedia Iranica*, which provides relatively concise overviews by leading scholars on specific topics in Iranian linguistics, in addition to topics in history and archaeology.

An extensive body of work on the Iranian languages also exists in Russian, most notably the six-volume series *Основы иранского языкознания – Fundamentals of Iranian Linguistics*, edited by V.S. Rastorgueva (1979; 1981; 1982; 1987; 1991; 1997); the two-volume *Опыт историко-типологического исследования иранских языков – Essays on the historical-typological study of Iranian languages*, also edited by V.S. Rastorgueva (1975a; 1975b);

and the three-volume *Языки мира: Иранские языки – Languages of the world: Iranian languages*, edited by Andrej Kibrik (1997; 1999; 2000).

The remainder of this chapter is organized as follows. Section 2.1 sets the stage with a look at the formation and **broad genetic structure of the Indo-Iranian languages**. Next, Section 2.2 provides a detailed **overview of Old, Middle, and Modern Iranian languages**. Then, Section 2.3 discusses the most fundamental genetic distinction in Iranian languages: **Eastern and Western Iranian languages**, followed in Section 2.4 by a discussion of the **Pamir languages**, a smaller group of Eastern Iranian languages spoken in the Pamir Mountains and surrounding region. Section 2.5 looks within the Pamir languages at the tight-knit **Shughni-Rushani group** of genetically related languages, of which Shughni is a part. A brief summary of the section is given in Section 2.6.

## 2.1 Broad genetic relations: Indo-Iranian

Shughni is an Indo-European language belonging to the Indo-Iranian branch, one of the eight branches of Indo-European with languages still spoken today, alongside Celtic, Germanic, Italic, Balto-Slavic, Albanian, Greek, and Armenian.

The Indo-Iranian languages are further subdivided into Iranian, which includes Shughni, Persian, Pashto, and many other languages spoken from central Turkey and the Caucasus to Pakistan, Afghanistan, and western China, and Indo-Aryan, which includes many major languages of the Indian subcontinent such as Hindi-Urdu, Punjabi, etc. (See Jain & Cardona 2007 for an overview of the Indo-Aryan languages and Kapović et al. 2017 for an overview of Indo-European.)

Aside from these two clearly defined branches of Indo-Iranian, at least two other language groups are sometimes considered to be branches of the Indo-Iranian family: (i) the Nuristani languages, spoken primarily in northeastern Afghanistan, and (ii) the Dardic languages, spoken in northern India, northern Pakistan, and eastern Afghanistan. The genetic classification of both Nuristani and Dardic remains controversial, however. Degener (2003), for instance, outlines three models for Nuristani, one of which has Nuristani as a distinct branch of Indo-Iranian, one which places it within Iranian, and one which places it within Indo-Aryan (see references in this work for further discussion of Nuristani). Likewise, the Dardic languages, generally considered to be a branch within Indo-Aryan,

have also been argued to be a separate branch of the Indo-Iranian languages (see Kulikov 2017, Bashir 2007, and references in those works).

Regarding the chronology of the development of Indo-Iranian languages, the Proto-Indo-Iranian language is thought to have been spoken in the eastern European steppe sometime between 3500 and 2500 BCE (Windfuhr 2009a: 5), which would place it around a thousand years removed from Proto-Indo-European, generally thought to date to at least 4000 BCE. Iranian and Indo-Aryan are thought to have split by 2000 BCE, shortly after which we begin to see the first attested ancient languages from the group (Sims-Williams 2017: 264; Skjærvø 2009b: 43). Both Vedic Sanskrit, an ancient Indo-Aryan language, and Avestan, an ancient Iranian language with striking similarities to Vedic Sanskrit, date to sometime in the second millennium BCE.

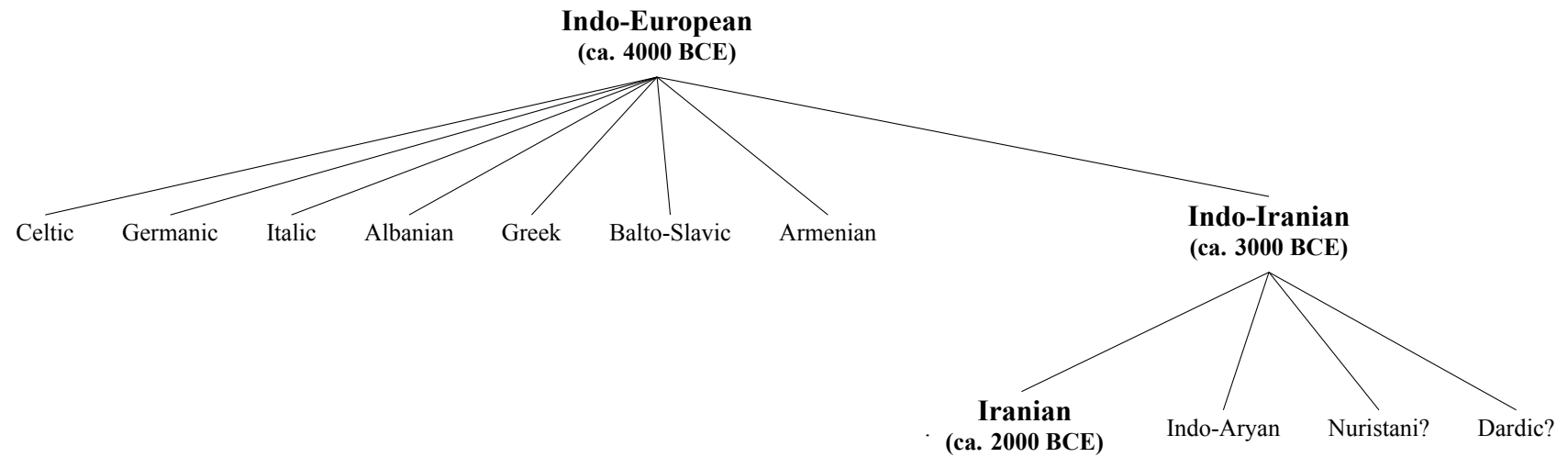
The broad genetic relations discussed here are schematized in Figure 2.1.

## 2.2 Old, Middle, and Modern Iranian languages

In terms of their distribution over time, Iranian languages are traditionally categorized into Old Iranian, Middle Iranian, and Modern Iranian. Old Iranian is associated with the time period after the breakup of Proto-Iranian and up until around the third century BC; Middle Iranian refers to the period from third century BC up until the Arab conquest around 700 AD; and Modern Iranian constitutes the period after the Arab conquest, or after 700 AD. These time periods are schematized in Figure 2.2.

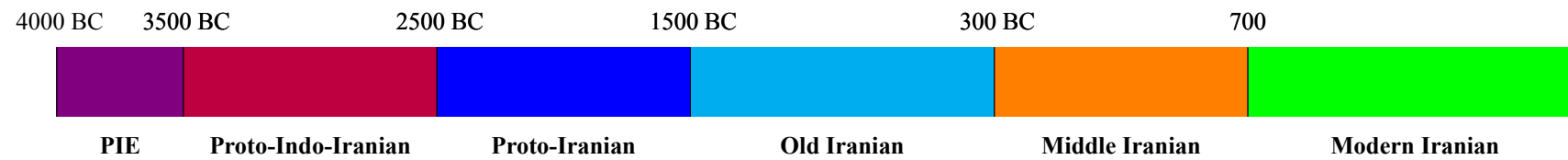
In general, the division of each phase of Iranian coincides with major historical events in the Iranian-speaking world. Thus, Old Persian is associated with the Achaemenid Empire; Middle Persian and Parthian with the Sassanian and Arsacid Empires, respectively; and Modern Persian with the time following the Islamic conquest. To paraphrase Sundermann (1989c), in transitional periods in the history of Iranian peoples, the older language stage that had been undergoing developments was completed, as it were, and the new stage was then given recognition in the newly established political space. However, this is not to say that political or social changes are the sole reason for novel linguistic developments, though in certain cases they may well be an important factor. The profound changes undergone by Persian after the Islamic conquest in the seventh century have been likened, for instance, to the changes undergone by Old English after the Norman conquest in the eleventh century.

**Figure 2.1:** Extant Indo-European Branches





**Figure 2.2:** Historical Periods of Iranian Languages (approx.)



Bygone Iranian languages, attested in various forms including written documents and inscriptions dating from times well before the birth of Christ, form the basis for our understanding of genetic relations among modern Iranian languages and linguistic changes which have shaped them into what they are today. However, for a number of reasons, not least of which is the shortage of sources and the fragmentary nature of many existing sources, attaching a specific time period to a particular Old or Middle Iranian language is an imprecise science. In particular, because many of the texts in which we see ancient languages are religious in nature, it is possible that the language of the texts represents the spoken language of an era before the dating of the texts, much like the English language of the King James Version of the Hebrew Bible, still used in many religious settings today, represents a variety of English spoken several centuries ago. Moreover, although certain extralinguistic influences, such as the Muslim conquest of certain areas or an invasion of other tribes into Iranian-speaking regions, can be dated with more certainty, it is often difficult to know the timeline over which their effects took place on spoken Iranian languages. Therefore, periods of time associated with Old and Middle Iranian languages, particularly with respect to when they were used as vernacular languages, should generally be understood as rough estimates.

### 2.2.1 Old Iranian

At the Old Iranian stage, two languages are attested in textual form: Old Persian, associated with Southwestern Iran, and in particular the Fārs Province, and Avestan, which is associated with Northeastern Iran.<sup>1</sup> Old Persian, on the one hand, is known from several secular inscriptions written in Cuneiform script, the most important of which is the Behistun Inscription in the Kermanshah Province of modern-day Iran. These inscriptions detail the deeds of the Achaemenid king Darius I (522-486 BCE), and the invention of the Old Persian script used in the inscriptions is also generally attributed to him. Avestan, on the other hand, is known from the religious texts which make up the sacred Zoroastrian texts known as the *Avesta*. Avestan is written in a modified form of the Pahlavi (i.e., Middle Persian) script. (The fact that we find an Old Iranian language written in a Middle Iranian script will be elucidated below.)

---

<sup>1</sup>Locating the region(s) in which Avestan was spoken is complicated by a number of factors, including the existence of Old Avestan and Young Avestan, two languages represented in the *Avesta* which were separated not only in time, but also likely in place. It is also complicated by the fact that the *Avesta*, unlike the inscriptions of Old Persian, does not contain clear reference to historical events. However, there is general agreement among scholars of Old Iranian that the homeland of Avestan speakers corresponds to modern-day northeastern Iran. For more on this, see Gnoli (1987) and references therein.

Both the Old Persian and Avestan scripts, for different reasons, present difficulties which make a proper analysis of the languages they represent rather challenging. In general, whereas the Old Persian Cuneiform script, as a syllabary, is ambiguous and leaves much to be desired in the way of phonemic and phonetic representation, the Avestan script, as used in the *Avesta*, is the opposite. That is, the Avestan script is used in a way that reflects the precise phonetic details of the language as it was used in slow chanting, and in turn obscures the phonemic structure of the language to a considerable degree. (See Hoffman 1987; Kellens 1989; Sims-Williams 2017; and Windfuhr 2009a for a more thorough discussion of these scripts and the languages they represent.)

The history of the *Avesta* presents additional challenges for interpreting the Avestan language. In particular, Avestan is not known to have been written down until the late Sassanian period (224-651 CE), and the oldest extant manuscripts are less than a thousand years old. Thus, the content of the texts was passed down orally for several thousand years before being written, and must have come into contact with Zoroastrian clergy who were native speakers of a variety of Iranian dialects separated in both time and space. Indeed, careful linguistic examination reveals that the language in the texts has clearly been influenced by multiple vernacular languages over the course of its transmission (see Kellens 1989: 35-37 for an overview of these influences). A second complication for interpreting the language in the *Avesta* is that it is not a single language, but rather at least two. The first of these, Old Avestan (also called Gathic Avestan), is a variety found in one set of texts, including the set of seventeen hymns known as the *Gāthās*. Old Avestan bears striking resemblance to Vedic Sanskrit, which has led to the notion that Old Avestan and Vedic Sanskrit, though clearly separated by a number of isoglosses, were likely not far removed from one another in the diachronic development of Indo-Iranian, perhaps being spoken languages sometime between 1500 and 1000 BCE. The second language represented in the *Avesta*, known as Young Avestan, shows divergences from Old Avestan which have allowed scholars to conclude that it is not only a language spoken at a later time than Old Avestan, from which it is separated by roughly half a millennium, but it is a different dialect of Iranian and therefore not a direct descendant of Old Avestan. It is generally thought that Old and Young Avestan represent ‘crystallized’ forms of older Iranian dialects which were used to transmit religious texts orally among Zoroastrian clergy. (See Skjærvø 2009b and references therein for more on the languages in the *Avesta*.)

Despite these difficulties, however, the Old Persian and Avestan texts still have much to teach us about ancient Iranian languages. Importantly, the texts reveal two distinct Old Iranian languages with clear isoglosses. Avestan represents a language, or more precisely, multiple languages, which are thought to have been spoken between approximately 1500 BC and 500 BC, and which therefore fall squarely within the Old Iranian period. On the other

hand, Old Persian, as it is seen in the Behistun inscriptions, is a late-stage Old Iranian language on the verge of the Middle Iranian period. Moreover, whereas Old Persian is clearly the ancestor of Southwestern Iranian languages including Modern Persian, it is not the case that Avestan is the direct ancestor of Eastern Iranian languages such as Shughni. Instead, Avestan is better thought of as a Central Iranian (i.e., non-Southwestern) language (e.g. Skjærvø 2009b: 51).

Although Avestan and Old Persian are the only Old Iranian languages with direct textual evidence, it can safely be assumed that they were not the only Iranian languages or dialects to have existed during the Old Iranian period. Convincing indirect evidence exists for at least two further Iranian languages: Median and Scythian. Median was the language of the Medes, who ruled an area in modern-day western Iran in the first half of the first millennium BCE, while Scythian was spoken by nomadic tribes to the north of Iran. In the case of Median, more substantial evidence is provided by the existence of many non-Perside (i.e., non-Old-Persian) Iranian words in Old Persian. These words often exhibit shared isoglosses with the language(s) of the *Avesta*, and, furthermore, many of them are attested in the Behistun inscriptions along with the cognate Old Persian word. Hence, there are a number of doublets in Old Persian wherein the non-Perside counterpart is assumed to be from Median. Scythians, for their part, are linked to tribes with Iranian names that are referenced in a variety of historical sources, and it is through such ancient writings that we can assume that Scythian tribes lived to the north of present-day Iran during the Old Iranian period.

Evidence for other Old Iranian languages comes in multiple forms. According to Schmitt (1989a: 86), the diversity of attested Middle Iranian languages alone is enough to provide a strong case for the existence of a multitude of Old Iranian languages. In addition to Median and Scythian, we can be fairly certain that the Middle Iranian language Parthian was preceded by an ancestor from the Old Iranian period, and that the two attested Old Iranian languages, namely Old Persian and Avestan, were accompanied by multiple dialectal varieties. Substantial support for this latter point is the notion that certain features in Middle Persian, usually understood to be the direct descendant of Old Iranian, indicate that it likely descended from a different dialect of Old Persian than the one we find in the Behistun inscriptions. Likewise, as mentioned above, Young Avestan appears to have descended from a variety other than Old Avestan. Other important evidence comes from onomastics, and in particular the fact that the endonyms of many Middle Iranian languages, modulo relevant sound changes, match the ethnonyms we find in Old Iranian textual sources, such as the Behistun inscriptions. (For more on the diversity of Old Iranian languages, see Schmitt 1989a.)

### 2.2.2 Middle Iranian

By the Middle Iranian period (ca. 300 BCE), we start to get textual evidence for several Iranian languages spoken across a vast geographical expanse from the modern-day Xinjiang Province of northwestern China, to Iran, the shores of the Black Sea, and well into Eastern Europe. In addition, by this time, most attested languages already display typical Eastern Iranian features or Western Iranian features (see the following subsection for more on this distinction). Thus, Middle Iranian languages are generally classified as either Western Iranian or Eastern Iranian, though this classification is not perfect, as we will see below.

Like Old Iranian, Middle Iranian languages are attested in a variety of sources, from religious documents – including Buddhist, Zoroastrian, Christian, and Manichean – to administrative documents, inscriptions on coins, bowls, helmets, and other physical items. However, compared to evidence of Old Iranian languages, that of Middle Iranian languages is generally both richer with regard to the amount of data available, as well as more diverse with regard to the types of sources in which we find the languages and the scripts by which they are represented.

The most well attested Middle Iranian languages are the two Western Iranian languages *Middle Persian* – sometimes also *Pahlavī*, as it is called in Manichean texts – and *Parthian*, both of which were spoken in modern-day Iran. The Eastern Middle Iranian languages for which we have significant direct textual evidence include *Khotanese* and *Tumshuqese*, sometimes together referred to as Saka, spoken around the Tarim Basin in Xinjiang, and *Sogdian*, *Bactrian*, and *Chorasmian*, spoken in different areas of Central Asia. Languages for which we have less evidence are the *Scythian* languages spoken by tribes in the northwestern part of the Iranian linguistic area, from the Caucasus along the north coast of the Black Sea and into the lower reaches of the Danube in Eastern Europe.

Below, I discuss each of these language groups in more detail; these details are summarized in Table 2.1. However, for a more thorough overview of Middle Iranian languages, the reader is directed to the sources referenced at the beginning of this section, namely the *Compendium Linguarum Iranicarum* (Schmitt 1989b); *Iranian Languages* (Windfuhr 2009b); and the *Encyclopedia Iranica*. Note, however, that these sources, although rather comprehensive, are meant to offer summaries and overviews of Iranian languages. Original publications on the decipherment and analysis of Old and Middle Iranian texts are cited within these volumes.

**Middle Persian** (Skjærvø 2009a; Sundermann 1989a) and **Parthian** (Sundermann 1989b) were two closely related Western Iranian languages spoken natively in different areas of modern-day Iran. Middle Persian, the descendant

of Old Persian, is associated with the area of Pārs in southwestern Iran, and was the court language of the Perside kings in this area and of the Sassanian Empire (224 CE – 651 CE). The name ‘Pahlavī’ is sometimes also used to refer to Middle Persian, though it more specifically refers to the Middle Persian language found in Manichean texts. Parthian, on the other hand, was the court language of the Arsacid Empire (247 BCE – 224 CE) and was spoken to the northeast of where Middle Persian was spoken, including the modern provinces of Khorasan and Golestan in Iran, and parts of Turkmenistan. Both languages are attested in religious and secular documents in a number of related scripts derived ultimately from the Aramaic script. Middle Persian is notably found in a fragmentary translation of Psalms (i.e., the *Pahlavi Psalter*), and it is generally thought that other, more complete translations of the Bible must have also existed in the language. Records of the importance of the Parthian language throughout the Near East and Middle East survive in the form of Parthian loanwords in a multitude of languages from the region, including Persian, Aramaic, Sogdian, Armenian, and more. Both languages exhibit features associated with Western Iranian languages, notably a comparatively simpler nominal and verbal inflection system than that of Eastern Iranian languages and the preservation of word-initial voiced stops *\*b*, *\*d*, and *\*g*.

**Sogdian** (Sims-Williams 1989b; Yoshida 2009; 2016) was an Eastern Iranian language spoken in Sogdiana, a region of oasis states which included much of modern-day northern Uzbekistan and western Tajikistan, and at times parts of Kyrgyzstan and southern Kazakhstan. Sogdiana was bounded by the Oxus River (modern Amu Darya) in the south and by the Jaxartes River (modern Syr Darya) in the north. The primary city of Sogdiana was Samarqand, with other major cities Bukhara and Kish (modern Shahrīsabz) in modern-day Uzbekistan, and Panjikent and Khujand in modern-day Tajikistan. The city of Samarqand, and Sogdiana generally, was an important hub along the Silk Road, and many Sogdian traders established diaspora communities along trade routes well into China, where the bulk of Sogdian material has been discovered. As a result of this widespread presence, Sogdian became a kind of *lingua franca* along the Silk Road, with some Turkic-speaking peoples, such as the Uyghurs, even adopting it as an administrative language. A wealth of Sogdian texts has been discovered along the Silk Road in China, especially at Turfan and Dunhuang. Sogdian materials include secular documents such as receipts and legal documents, as well as religious texts, including texts from the Manichean and Christian religions. The language was written in at least four scripts: the Sogdian script was used for secular documents; the Manichean script for Manichean texts; the Syriac script for Christian texts; and the Brahmī script was used for Buddhist documents, though only fragmentary documents have been found with Sogdian represented by the Brahmī script. Regarding its linguistic features, Sogdian displays some of the canonical Eastern Iranian sound developments, such as the

voicing of the clusters *\*xt* and *\*ft*, as well as the spirantization of *\*b*, *\*d*, and *\*g* in initial position; however, *\*č* is preserved as such and does not undergo depalatalization as in many other Eastern Iranian languages.

**Chorasmian** (Durkin-Meisterernst 2009; Humbach 1989; MacKenzie 1991) – also commonly spelled ‘Khwarezmian’ – was an Eastern Iranian language spoken in Chorasmia (also ‘Khwarazm’), an ancient region centered around the lower reaches of the Oxus River in modern-day western Uzbekistan and Turkmenistan. Chorasmia was bounded in the north by the Aral Sea and in the south by the Karakum desert. Several sites in Uzbekistan and Turkmenistan, particularly the city of Khiva, Uzbekistan, have been identified as important Chorasmian settlements. The Chorasmian language is attested in two stages: Old (or ‘Middle’) Chorasmian, written in an Aramaic-derived script and found in inscriptions on vessels, ossuaries, and other media, and Late Chorasmian, written in an Arabic-derived script and found in the interlinear glosses of a few Arabic documents, including a dictionary and a book on case law. Chorasmian shares a number of features with Sogdian and other Middle Eastern Iranian languages, particularly the sound developments typical to Eastern Iranian languages such as the spirantization of *\*b*, *\*d*, and *\*g* in initial position.

**Bactrian** (Sims-Williams 1988; 1989a) was an Eastern Iranian spoken in the region of Bactria, which was centered around modern-day northern Afghanistan, southern Tajikistan, and southern Uzbekistan. It was bounded in the south by the Hindu Kush, and in the north by the Oxus River, though its borders extended beyond the river at times. Bactria, with its capital at Balkh, was ruled by the Greeks for several centuries after the conquest of Alexander the Great in the fourth century BCE, and even after the Greeks were ousted in the second century BCE, the Greek language continued to be used for some time in official capacities. Eventually, the Bactrian language was written down using the Greek script and is attested through coins, seals, and inscriptions dating to the first through third centuries CE, as well as manuscripts from several centuries later. Writings in Bactrian have also been found in the Manichean script. The language shares a number of sound developments typical of Eastern Iranian, but its relatively simple morphological system resembles Western Iranian languages.

**Khotanese and Tumshuqese** (Emmerick 1989; 2009) were two closely related Eastern Iranian languages spoken within and nearby the Kingdom of Khotan, a Buddhist kingdom around the Taklamakan Desert in the modern-day Xinjiang Uyghur Autonomous Region in the People’s Republic of China. Together, Khotanese and Tumshuqese are often referred to as *Saka*, where Tumshuqese is the more archaic variety and Khotanese the more innovative. Geographically, Tumshuqese is associated with the village of Tumshuq (now spelled ‘Tumxuk’), which lies on the

northwestern edge of the Taklamakan Desert, via the discovery of documents in the language there. Khotanese, on the other hand, is associated with an area to the southeast of Tumshuq centered around the town of Khotan (modern-day Hotan), which was the center of the Kingdom of Khotan. The extant materials in Khotanese far outnumber those in Tumshuqese, and the linguistic features of Khotanese have been analyzed to a greater extent than those of Tumshuqese. Virtually all documents in the language are written in variants of the Brahmi script and are Buddhist in content. Khotanese exhibits a number of features associated with the Eastern Iranian languages, but in many cases the developments are not as complete as in other Eastern Iranian languages. For instance, in Khotanese the palatal affricate *\*č* developed into alveolar *ts* only before non-palatals, and in initial position the stops *\*b* and *\*d* are thought to have been spirantized to *v* and *ð*, but *\*g* remains a stop.

**Scythian** (Abaev and Bailey 1985; Bielmeier 1989; Schmitt 2018) is generally conceived of as a group of several related Eastern Iranian languages and dialects spread out across both time and space. Unfortunately, because of a general lack of materials in these languages, very little is known about them or about their relation to other Iranian languages. Somewhat more is known about the Scythians as an ethnic group, however, as they are mentioned in a variety of historical sources and are linked to many archaeological findings. We know, for instance, that the Scythians were powerful nomadic warrior tribes who originally occupied the northern extremes of the Iranian-speaking ancient world, most notably the steppe above the Black Sea and the Caspian Sea. Over time, however, Scythian peoples migrated in multiple directions and ended up as far west as the Iberian peninsula and North Africa. The majority of evidence for Scythian languages comes in the form of onomastics – i.e., personal names, ethnonyms, and toponyms of clear Iranian origin which are identified in a variety of historical sources. Several have been identified in Greek inscriptions along the north shore of the Black Sea as well as in the writings of classical historians. A particularly well known Scythian group is the Alanians, where the name ‘Alan’ can be derived from *\*arya-* ‘Aryan’. The language of the Alans is attested in an inscription on a tombstone in southern Russia and in the work of Byzantine historian Johannes Tzetzes (12th century CE), who offers two lines of what he refers to as the Alanian language. Both instances of Alanian are written in Greek script. Likewise, a short word list on the back of a document from the fifteenth century, found in the 1950’s in a Hungarian library, represents the Jassic variety, brought to Hungary by Alanians fleeing the Mongols in the thirteenth century. The Scythian languages are survived by modern Ossetian, a direct descendant of Alanian, spoken in Turkey, Georgia, and southern Russia.

Finally, it is important to note that, as in the Old Iranian period, the Middle Iranian languages found in texts which have survived to modern times were more than likely only a small portion of the languages and dialects which



**Table 2.1:** (Attested) Middle Iranian Languages.

	LANGUAGE	REGION	ATTESTATION	SCRIPTS	FEATURES
WESTERN	<b>Middle Persian</b>	Pārs Province southwestern Iran	many secular documents & religious documents	primarily Aramaic-derived scripts, including Pahlavī, Manichean, Syriac, etc.	relatively simple morphology, initial <i>*b</i> , <i>*d</i> , <i>*g</i> preserved
	<b>Parthian</b>	<i>Parthia</i> northeastern Iran	religious and secular documents, especially bilingual Pahlavī-Parthian inscriptions	primarily Aramaic-derived scripts, especially Pahlavī	relatively simple morphology, initial <i>*b</i> , <i>*d</i> , <i>*g</i> preserved
EASTERN	<b>Sogdian</b>	<i>Sogdiana</i> Uzbekistan, Tajikistan	secular documents, religious texts	Sogdian, Manichean Syriac, Brahmī	initial <i>*b</i> , <i>*d</i> , <i>*g</i> spirantized, <i>*xt</i> , <i>*ft</i> voiced, but <i>*č</i> preserved
	<b>Chorasmian</b>	<i>Chorasmia</i> western Uzbekistan, northern Turkmenistan,	vessels, ossuaries (Early Ch.), scholarly documents (Late Ch.)	Aramaic (Early Ch.), Arabic (Late Ch.)	initial <i>*b</i> , <i>*d</i> , <i>*g</i> spirantized, <i>*č</i> depalatalized <i>*xt</i> , <i>*ft</i> voiced
	<b>Bactrian</b>	<i>Bactria</i> northern Afghanistan, southern Tajikistan	coins, seals inscriptions	Greek, Manichean	initial <i>*b</i> , <i>g</i> spirantized, <i>*d &gt; l</i> (likely via <i>*δ</i> ), but relatively simple morphology
	<b>Khotanese &amp; Tumshuqese</b>	<i>Kingdom of Khotan</i> Xinjiang, China	Buddhist documents	Brahmī	initial <i>*b</i> , <i>*d</i> spirantized, but <i>*g</i> preserved, and <i>*č</i> depalatalized only in certain positions,
	<b>Scythian</b>	Eurasian steppe Ukraine, Caucasus, eastern Europe	primarily onomastics, one inscription, a couple lines in a historical document	no known widespread writing, inscription in Greek letters	In Ossetic, <i>*č</i> depalatalized, <i>*xt</i> , <i>*ft</i> voiced, <i>*g</i> spirantized (Digor), <i>&gt; q</i> (Iron), but <i>*b</i> , <i>*d</i> preserved

existed at the time. Evidence for other Western Middle Iranian languages comes, for example, in the form of Iranian words borrowed into other languages (e.g. Armenian), but whose form indicates sound developments which differ from those in Middle Persian or Parthian (see Sundermann 1989c: 106-107). The diversity of Eastern Iranian languages is hinted at, for instance, by the fact that Yaghnobi, though generally considered to be a descendant of Sogdian, nonetheless displays a number of divergences from the Sogdian represented in extant texts – divergences which, importantly, cannot be accounted for by changes subsequent to the Middle Iranian period.

### 2.2.3 Modern Iranian

In modern times, the Iranian languages are a diverse group spoken by some 200 million people from Central Turkey in the west to the Xinjiang Province of China in the east. Three countries have a dialect of Modern Persian as an official language: *Farsi* in Iran, *Dari* in Afghanistan, and *Tajiki* in Tajikistan.<sup>2</sup> In Afghanistan, the Eastern Iranian language Pashto is also an official language, and in Iraq, Kurdish is an official language alongside Arabic. Dozens of other Iranian languages and dialects are spoken in these four countries and many others, including Russia and Georgia (Ossetic); Turkey (Ossetic and Kurdish); Syria, Armenia, and Azerbaijan (Kurdish varieties), Oman (Kumzari), Pakistan (Balochi, Pashto and Pamir languages), Turkmenistan (Balochi); Uzbekistan (Persian varieties); and China (Sarikoli and Wakhi).

On the one hand, the typological diversity within the Iranian group belies several millennia of isolated development, during which individual and groups of Iranian languages developed their own unique features and idiosyncrasies. On the other, the striking similarities found across many Iranian languages reflect not only their common origin, but also the prolonged contact which took place, and continues to take place, among speakers of Iranian languages and between speakers of Iranian and non-Iranian languages. In the latter sense, the Iranian languages are part of

---

<sup>2</sup>A few notes on the Persian language are in order here. Throughout this dissertation, I use the term ‘Persian’ to refer to the continuum of generally mutually intelligible Western Iranian language varieties which are spoken from Iran to Afghanistan and Tajikistan, and whose standardized varieties are Farsi (Iran), Dari (Afghanistan), and Tajiki (Tajikistan). In some cases, I use the name of one of the specific standard varieties to distinguish it from the other two.

Varieties of Persian vary rather substantially from one another, both within each country and from country to country, and in terms of phonology, morphology, syntax, and lexicon. In addition, whereas in Iran and Afghanistan, Persian is written with an Arabic-based script, in Tajikistan it is written with a Cyrillic-based developed during the time of the Soviet Union. (For a detailed overview of the variation among different dialects of Persian, see Lazard 1989 and Windfuhr and Perry 2009.)

broadest linguistic area of both genetically related and unrelated languages which share many features.

All Iranian languages, to one degree or another, have received some influence from both Arabic and Turkic, as speakers of these language groups have had a significant presence throughout the Iranian-speaking area. Persian, for its part, has often been the vehicle by which Arabic and Turkic influence has reached minority Iranian languages outside the center of the Iranophone world, and Persian itself has exerted further influence as a superstrate language.

In the outer reaches of the Iranian-speaking region, a number of non-Iranian languages have influenced and continue to influence Modern Iranian languages. In the west, Ossetic speakers in Russia have consistent contact with Russian and Caucasian languages, in Georgia with Kartvelian, and in Turkey with Turkish; speakers of Kurdish are in contact with Turkish and Arabic. In the east, the Persian varieties of Tajikistan and Uzbekistan are spoken alongside Turkic languages Uzbek and Kyrgyz; Pamir languages, Pashto, and Balochi all show evidence of prolonged contact with Indo-Aryan languages and possibly non-Indo-European languages spoken in the region before the arrival of Iranians (a modern descendant of which might be Burushaski); and Sarikoli along with the Wakhi spoken in Xinjiang are influenced by the majority languages of the region, Uyghur and Mandarin Chinese. Of course, in modern times, Russian has crept into many of the languages of the former Soviet Union, including many Iranian languages such as Tajiki and the Pamir languages of Tajikistan.

In some cases, specific modern Iranian languages can be linked to specific Middle Iranian languages, albeit with varying degrees of certainty. For instance, Modern Persian is the direct descendant of Middle Persian, which in turn is the descendant of Old Persian. Modern Ossetic is descended from the Alanian variety attested on a gravestone in southern Russia. Yaghnobi is the descendant of a Sogdian dialect, and modern-day Wakhi bears considerable resemblance to Khotanese Saka. In the vast majority of cases, however, the Old and Middle stages of Modern Iranian languages must be inferred via the reconstruction of bygone language varieties which are not attested in writing. This is precisely the case for Shughni and the Pamir group more generally, as we will see in the following subsection.

## 2.3 Eastern vs. Western Iranian languages

The Iranian languages are traditionally subdivided further into the Western and Eastern groups, which are themselves subdivided further into Northwestern/Southwestern and Northeastern/Southeastern. In reality, however, genetic relations among the Iranian languages are not as clear-cut as these purported subgroups might imply. (For more on genetic relations among Iranian languages, see Windfuhr 2009a and references cited therein.)

Genetic relations among the Western Iranian languages are complex, but it is generally agreed upon that Persian and related dialects make up a Southwestern Iranian supgroup, while Zazaki, Gorani, Kurdish, Balochi, among others, make up a Northwestern Iranian subgroup. Among the Eastern Iranian languages, Ossetic, Yagnobi, Pashto, and the Pamir languages make up the Northeastern group, while Parachi andOrmuri constitute the Southeastern group.

**Table 2.2:** Genetic groupings of Iranian languages.

Genetic Group	Languages
NW Iranian	<i>Zazaki, Gorani, Kurdish, Balochi, . . .</i>
SW Iranian	<i>Farsi, Dari, Tajiki (and dialects)</i>
NE Iranian	<i>Ossetic, Yagnobi, Pashto, Pamir Languages</i>
SE Iranian	<i>Parachi, Ormuri</i>

The notion that the Eastern Iranian languages are a single genetic subgroup descended from a common ancestor has yet to be properly demonstrated, however, as there is no unifying innovation which brings them together (e.g., Sims-Williams 1996; Wendtland 2009). Rather, this grouping is made on certain developments and archaisms which are widespread – but not universal – in Eastern Iranian languages, but rare or unattested in Western Iranian languages. Such developments include certain phonological phenomena, such as the voicing of Old Iranian clusters *\*fɪ* and *\*xɪ* to *vd* and *ɣd*, respectively, as in Shughni (*(w)ūvd* ‘seven’ < *\*hafta* (cf. Persian *haft*); the development of a dental affricate *ts̪* (generally written <c>) from Old Iranian palatal *\*č̪*, as in Shughni *cavor* ‘four’ < *\*čaθwar* (cf. Farsi *čahār*); and the spirantization of initial stops *\*b*, *\*d*, and *\*g* to *v*, *ð* and *ɣ*, as in Shughni *vud* ‘was’ (cf. Persian *bud*). However, none of these phenomena fully unifies the Eastern Iranian languages to the exclusion of Western Iranian languages. For instance, Yagnobi, which shares striking similarities with the Middle Iranian language Sogdian, does not exhibit the voicing of *\*fɪ* or *\*xɪ*; a number of Eastern Iranian languages, including Yazghulami and Munji, preserve Old Iranian *\*č̪* at least in certain positions; and initial stops *\*b*, *\*d*, and *\*g* are preserved as

such in Parachi andOrmuri.

Morphological characteristics which set certain Eastern Iranian languages apart include more conservative (i.e., more complex) inflection systems than those found in Western Iranian languages; a tripartite deictic system inherited from Proto-Indo-European (cf. the deictic system in Shughni); and the development of a second-person pronoun with a prefixed *t-*, which is itself a form of the second-person singular pronoun (see, e.g., Wendtland 2009: 179-180). Like the phonological phenomena discussed above, each of these morphological characteristics unites only a portion of Eastern Iranian languages to the exclusion of others, and for each feature the grouping of languages is different. Although many Eastern Iranian languages preserve a system of grammatical gender (e.g. Shughni, Pashto), this system is lost in other languages (e.g. Sarikoli). In Yazghulami, the tripartite deictic system has been replaced with a two-degree system (proximal and distal). The peculiar formation of second-person plural pronouns is found in the Middle Iranian language Bactrian and modern languages Ishkashimi, the Shughni-Rushani group, and possibly Pashto and Ormuri, but is not found in the Middle Iranian languages Chorasmian and Khotanese, nor in modern languages Yaghnobi, Munji, Wakhi, Parachi, or Ossetian. Compare, for instance, Shughni 2PL *tama*; Ishkashimi *ṭmṭx*; Sar. *tamaš*; but Munji *mof*, Wakhi *sa(y)-iš(t)* (DIR) and *sav* (OBL).

Finally, certain lexical items are common in Eastern Iranian languages but rare or unattested in Western Iranian languages. These include, for instance, *\*kuta* ‘dog’, which has as reflexes Shughni *kud*, Yazghulami *k<sup>o</sup>od*, and Yaghnobi *kut*; as well as *\*drawa-* ‘(individual) hair’, which has reflexes in the modern languages as Shughni *cīw*, Yazghulami *cu*, Yaghnobi *diraw*, and Ormuri *drī* (Sims-Williams 1996). (For an overview of the distinctive properties of Eastern Iranian, see Edelman 1980a; Wendtland 2009.)

Ultimately, the term ‘Eastern Iranian’ is useful insofar as it allows us to discuss a group of languages which are undoubtedly Iranian languages but which fall outside Western Iranian groups such as Persian, Kurdish, and Balochi. However, at least at this point in time, Eastern Iranian languages should not be considered a single genetic grouping. To quote Nicholas Sims-Williams (1988) on the matter, “if one reconstructs ‘proto-Eastern Iranian’ in such a way as to account for all the features of the group, it proves to be identical to the ‘common Iranian’ reconstructible as the ancestor of the whole Iranian family. It is therefore more useful to conceive of Eastern Iranian as a ‘Sprachbund’ or areal grouping of languages”. As we will see in the following subsections, the same can be said for the Pamir languages, which share a number of distinguishing phenomena amongst them, but not in a way that allows us to reconstruct a “proto-Pamir” language.

## 2.4 Pamir languages

The Pamir languages are a group of around a dozen Eastern Iranian languages spoken in the geographical region surrounding the Pamir mountains. They are spoken in an area which includes the westernmost parts of Xinjiang, much of eastern and southeastern Tajikistan, the northeastern part of Afghanistan including the Wakhan Corridor, and northern Pakistan, particularly in Gilgit-Baltistan and Khyber Pakhtunkhwa. The exact number of Pamir languages depends, as it does for language estimates elsewhere, on where one draws the line between a language and a dialect, but most estimates range between ten and fifteen. According to the most recent scholarship (e.g., Edelman & Dodykhudoeva 2009a), the Pamir languages can be divided into four genetic subgroups. These groups are, in the order they are discussed below: (i) Munji-Yidgha; (ii) Wakhi; (iii) Ishkashimi; and (iv) the North Pamir group. Crucially, however, despite their geographical proximity as well as a number of grammatical and phonological similarities, the Pamir languages have thus far not been shown to constitute a single cohesive genetic subgroup set apart from other Eastern Iranian languages.

The **Munji-Yidgha** group includes the two related languages Munji, spoken in the Kuran wa Munjan District of the Badakhshan Province in northeastern Afghanistan, and Yidgha, spoken in the Lotkoh Valley of the nearby Chitral District of Gilgit-Baltistan in Pakistan. Munji has been said to have roughly 5,000 speakers (Beyer and Beck 2011) and two major dialects which correspond to the upper and lower portions of the Munjan Valley (e.g. Williamson 2017). Notable works on Munji and Yidgha include a short overview in the *Compendium Linguarum Iranicarum* (Skjærvø 1989); a couple grammatical sketches (Morgenstierne 1938b; Grjunberg 1972); an examination of genetic relations between Munji-Yidgha and other Pamir languages (Sokolova 1973); a sociolinguistic assessment of Munji (Beyer and Beck 2011); and a phonological sketch of Munji (Williamson 2017).

**Wakhi** is spoken by a total of approximately 37,000 people dispersed among all four countries in which Pamir languages are represented: Pakistan, Afghanistan, Tajikistan, and China. In Pakistan, the language is spoken in Gojal, Ishkoman, and Chitral in the extreme north of the country. In Afghanistan and Tajikistan, Wakhi is spoken in the Wakhan Corridor along the Panj River above Ishkashim. In China the language is spoken in various settlements in the western part of Xinjiang. Scholarly work on the language has been done primarily for the Wakhi spoken along the Wakhan Corridor, which forms a dialect continuum (e.g. Grjunberg and Steblin-Kamenskij 1976; Pakhalina 1975) and for Wakhi spoken in the Upper Hunza in Pakistan (e.g. Lorimer 1958). (See also Elena Bashir's (2009)

chapter in *The Iranian Languages* for an overview of the language.)

The **Ishkashimi** group includes the Ishkashimi language, the closely-related Sanglechi dialect, as well as the now extinct Zebaki dialect. Ishkashimi proper is likely spoken by no more than a couple thousand people on both sides of the Panj River near the village of Ishkashim (i.e., in both Afghanistan and Tajikistan). In Tajikistan, the language is used only by residents of the villages of Ryn and Sumgin. See Mueller et al. (2010a) on the sociolinguistic situation of Ishkashimi. Sanglechi is generally considered to be a dialect of Ishkashimi and is spoken in the Sanglech Valley, which runs south-southwest from the village of Ishkashim. Sanglechi may have an additional thousand speakers or more. Works dedicated specifically to Ishkashimi and Sanglechi are scarce, though one may find information regarding them in some of the general literature on Pamir languages. Works on Ishkashimi include a sketch by Pakhalina (1959) and a short article by Steblin-Kamenskij (1998) in the *Encyclopedia Iranica*. A relatively recent doctoral dissertation was published on Sanglechi by Yusufbekov (2000) and elaborated upon in Yusufbekov and Dodykhudoeva (2008).

The **North Pamir group** (also called the ‘Shughni-Yazghulami’ group in some Russian-language sources) includes, on the one hand, the Yazghulami language and the closely related but now extinct Wanji language, and on the other, the Shughni-Rushani languages, discussed below, to which Shughni belongs. That Yazghulami and the Shughni-Rushani group belong to a single genetic subgroup of Eastern Iranian languages was first demonstrated by Valentina Sokolova (1967) in her work titled *Генетические отношения язгулямского языка и шугнанской языковой группы* – *Genetic relations between Yazghulami and the Shughni Group* – in which she reconstructs the vowel system of the Proto-Shughni-Rushani and Proto-Yazghulami languages and then uses these reconstructed systems to reconstruct the vowel system of the Proto-Shughni-Yazghulami language. Yazghulami is spoken by a few thousand people in the Yazghulam Valley of Tajikistan, a right tributary to the Panj River located roughly a hundred kilometers downstream (north) of Khorugh. The language is noteworthy among Pamir languages in that its speakers are Sunni Muslims, whereas the majority of other Pamir speakers are Ismaili. The primary sources on Yazghulami include Joy I. Edelman’s grammatical sketch (1966) and dictionary (1971). A section on Yazghulami phonology is also included by Sokolova (1953: 176-208) in her collection of phonological sketches on Eastern Iranian languages.

The map in Figure 2.3 (from Kim 2014) shows the geographic distribution of Pamir languages.

Much like the Eastern Iranian languages, discussed in section 2.3, the Pamir languages share a number of dis-

**Figure 2.3:** Map of the Pamir Languages (Kim 2014)

tinctive features among them, but do not show a single innovation shared by all languages. These broadly shared features, for which there are almost always exceptions, can be taken as evidence of prolonged contact among these languages and perhaps also as evidence of contact with common substrate languages once spoken in the region.<sup>3</sup>

<sup>3</sup>The Pamir languages are considered to be part of a larger *Sprachbund*, often referred to as the Central Asian Sprachbund, which includes, in addition to languages spoken in the Pamirs, languages of the Hindu Kush, certain spurs of the Karakoram Range, and parts of the Himalaya. Unlike the Pamir group, languages of this larger linguistic area are not all Eastern Iranian and are in fact strikingly diverse in their genetic classifications. They include other Eastern Iranian languages such as Pashto, Parachi, andOrmuri; the Western Iranian language Balochi; Nuristani languages; Indo-Aryan languages such as Domaaki, Pahari dialects, Punjabi, and Sindhi, as well as the Dardic languages; and several non-Indo-European languages such as the isolate Burushaski, a handful of Dravidian languages including Brahui, and a few Tibeto-Burman languages.

Certain typological features are found throughout these languages which suggest prolonged contact with one another and perhaps effects of substrate languages. Particularly telling are features which are shared among a genetically related portion of these languages, but are uncommon among the languages of the larger family. In the case of Eastern Iranian, such ‘anomalous’ features include the presence of retroflex consonants, notably retroflex fricatives and affricates; a class of semantically intransitive verbs which behave morphosyntactically as transitive; gender systems based not on the inherited Indo-European stem, but rather on semantic divisions along lines of individual vs. general, animate vs. inanimate, concrete vs. abstract, etc.; and certain means of expressing alienable vs. inalienable possession.

Ultimately, these features, when considered together, suggest a certain typology for a substrate language or group of languages which would



Examples of such features can be found in both the phonology and morphosyntax of these languages. Regarding their phonology, most Pamir languages share the development whereby Proto-Iranian \*č becomes depalatalized; however, this has not occurred in Yazghulami or Munji. Likewise, the typical Eastern Iranian voicing of \*xt has occurred in Yazghulami, Ishkashimi, Munji, and Wakhi, but the cluster undergoes further simplification in Shughni and Sarikoli. Regarding their morphosyntax, the majority of Pamir languages, including Shughni, exhibit distinct strategies in the present and past tenses for marking the person and number features of subjects (and sometimes objects). In this system, present-tense subject features are marked with a verbal affix, while past-tense subject features are marked with a second-position clitic which is relatively free with respect to the types of phrases it can attach to. However, this system is absent in Munji, where the person and number features of both present- and past-tense subjects are co-referenced by a verbal affix. (For a summary of such innovations, archaisms, and their relevance in the genealogy of the Pamir languages, see Payne 1989: 420-423 and Wendtland 2009. For a more thorough treatment of these matters, see Edelman 1986; 1990; 2009; and Sokolova 1967; 1973.)

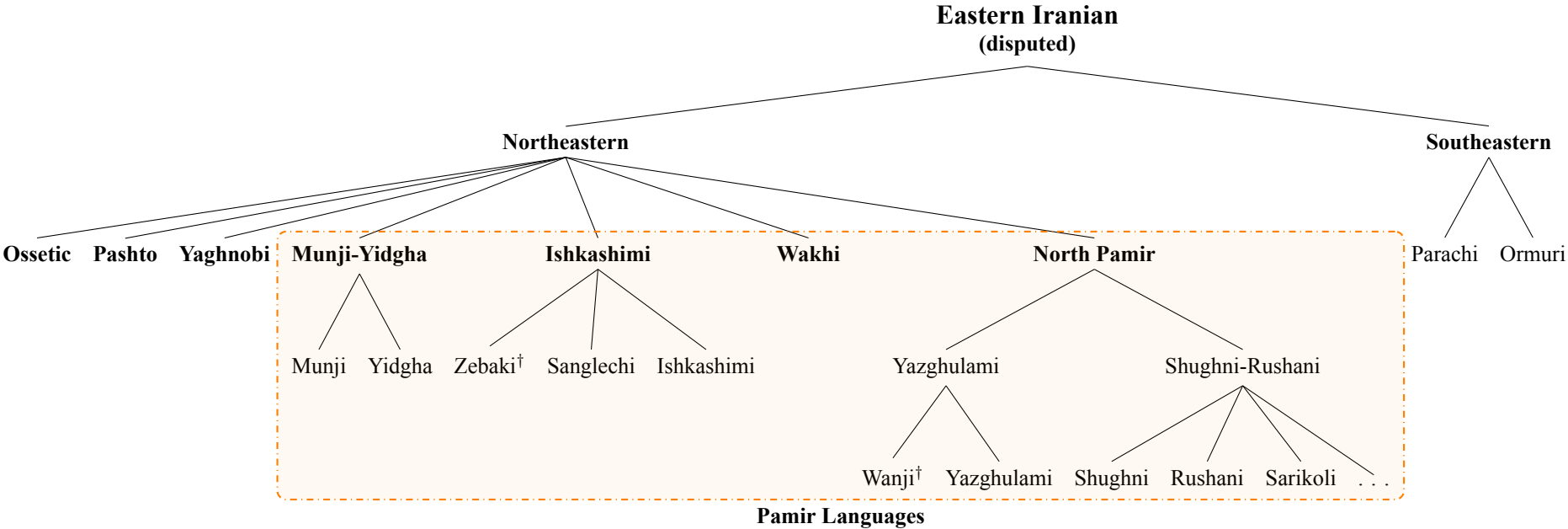
Upon close examination, the distinctive features of Pamir languages are distributed in such a way that precludes their classification as a single genetic subgroup. Instead, the Pamir languages are better thought of as a *Sprachbund* or linguistic area, rather than a genetic grouping. The position of the Pamir languages within the Eastern Iranian group is schematized in Figure 2.4. Note that each of the genetic subgroups belonging to the Pamir group are on the same level as other Northeastern Iranian groups, which include Ossetic, Pashto, and Yaghnobi. Hence, it has not been demonstrated that the Pamir languages are any more closely related to one another than to the other members of the Northeastern branch.

According to Edelman & Dodykhudoeva (2009a), we can think of each the four Pamir language groups as representing a distinct wave of immigration into the region by distinct Iranian-speaking groups. These waves of immigration are also likely to have occurred at distinct periods in the history of Iranian languages, and may be separated by hundreds of years. The ancestors of the modern-day Shughni-Rushani speakers, whose languages are the topic of the next subsection, would have been one such group who immigrated into the Pamir area many

---

have co-existed with the ancestors of the modern languages of the Central Asian Sprachbund and, over time, would have projected certain features onto these languages. In particular, this substrate very likely had active-stative morphosyntactic alignment along with a robust and complex noun classification system based on semantic criteria. The Burushaski language, a language isolate spoken in northern Pakistan, exhibits a number of these phenomena and may well be the modern offspring of the language family – or one of the language families – which formed the substrate for this Sprachbund. For more detail on this linguistic area, see Edelman 1980a and references therein.

Figure 2.4: Pamir Languages within Eastern Iranian



centuries ago.

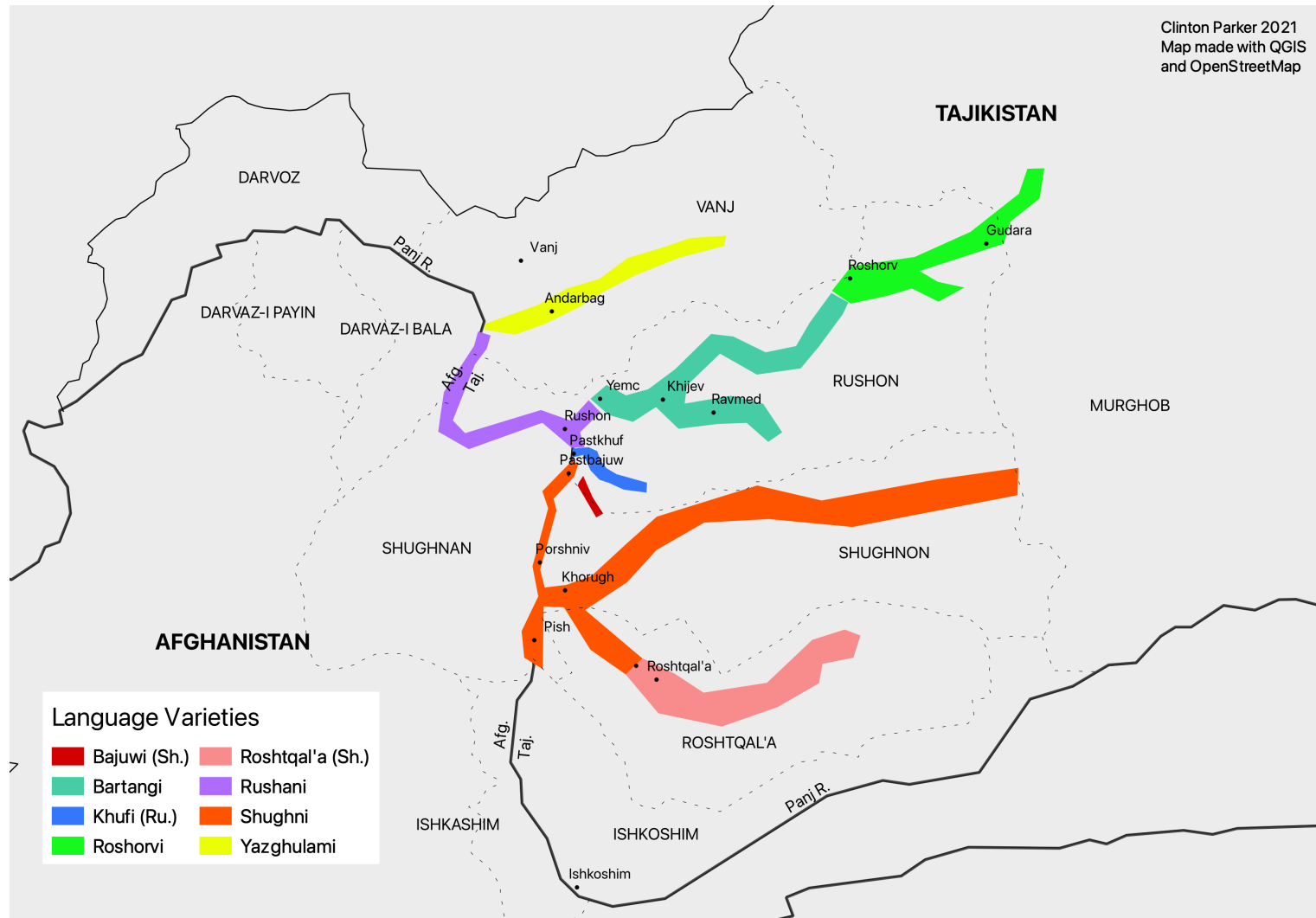
## 2.5 Shughni-Rushani Subgroup

The Shughni-Rushani group consists of several closely related language varieties spoken in the Gorno-Badakhshan region of Tajikistan, the Badakhshan Province of Afghanistan, and the Tashkurgan Tajik Autonomous County in western Xinjiang. Together with Yazghulami, these languages form the North Pamir genetic subgroup of Eastern Iranian languages (see Figure 2.4).

Within the Shughni-Rushani group, four smaller genetic subgroups are distinguished (e.g. Edelman & Dodykhudoeva 2009a; Payne 1989); they are the following: (i) *Shughni*, together with several dialects, including *Bajuwi*, spoken in a small valley toward the northern extreme of the Shughni-speaking area, possibly extinct *Barwozi* in the upper reaches of the Shahdara; and a few microdialects in the Ghund Valley; (ii) *Rushani*, spoken to the north of the Shughni area along the Panj River from Pastkhuf to Shipad, and *Khufi*, spoken in the Khuf Valley; (iii) *Bartangi*, spoken in the Bartang Valley which flows into the Panj at Rushan, and *Roshorvi*, spoken along the Gudara and Murgab Rivers, both of which flow into the Bartang; and (iv) *Sarikoli*, spoken in Tashkurgan, Xinjiang, with at least three distinct dialects. A map of Shughni-Rushani varieties in Tajikistan and Afghanistan – together with Yazghulami but leaving out Sarikoli – is shown in Figure 2.5, and general information on each variety of the group is given in Table 2.3.

The Shughni-Rushani languages share a number of phonological, morphosyntactic, and lexical features, many of which set them apart from other Pamir languages and other Iranian languages more generally. However, it should be noted that in many of these respects, Sarikoli has developed differently than the other languages in the group. For instance, all varieties except Sarikoli exhibit some form of vestigial ergativity – i.e., traces of a formerly robust ergative construction in the past tense – whether it be in the form of oblique past-tense subjects (e.g. Rushani), a special third-person singular agreement marker found only for past-tense transitive and unergative subjects, but not unaccusative subjects (e.g. Shughni), or gender distinction in the past stem of unaccusative verbs, but not unergative or ergative (e.g. Shughni). In fact, Sarikoli is the only variety which has not preserved grammatical gender. With respect to phonology, Sarikoli is unlike other languages of the subgroup in that it lacks phonemic vowel length.

**Figure 2.5:** Shughni-Rushani Lanugages



**Table 2.3:** Shughni-Rushani language varieties.

SUBGROUP	SUBVARIETIES	COUNTRIES	REGION	NO. OF SPEAKERS
<b>Shughni</b>	<i>Shughni proper</i>	Tajikistan, Afghanistan	along the Panj River from Darmorakht (upriver) to Dasht (downriver)	ca. 80,000-100,000 (Edelman & Dodykhudoeva 2009a)
	<i>Barwozi</i>	Tajikistan	upper Shahdara Valley	
	<i>Bajuwi</i>	Tajikistan	Baju(w) Valley, north of Khorog	
<b>Rushani</b>	<i>Rushani proper</i>	Tajikistan, Afghanistan	along the Panj River from Pastkhuf (upriver) to Dasht (downriver)	ca. 15,000 (Payne 1989)
	<i>Khufi</i>	Tajikistan	Khuf Valley. (east of Pastkhuf)	
<b>Bartangi</b>	<i>Bartangi proper</i>	Tajikistan	Bartang Valley (east of Rushan)	ca. 5,000 Payne (1989), citing 1970's estimates
	<i>Roshorvi</i>	Tajikistan	Along Gudara & Murgab rivers (above Bartang River)	
<b>Sarikoli</b>	<i>Central dialect</i>	China	Tashkurgan & surrounding area	ca. 40,000 (Kim 2017)
	<i>Near-Eastern dialect</i>	China	east and southeast of Tashkurgan, e.g. Baldir, Mariong (villages)	
	<i>Far-Eastern dialect</i>	China	east and southeast of Tashkurgan, e.g. Burangsal, Tung, Kichik Tung (villages)	

Nonetheless, the vowels of all Shughni-Rushani varieties, including Sarikoli, correspond systematically. These correspondences are displayed in full by Sokolova (1967) and elaborated upon by Edelman (1986). These authors demonstrate the development of Proto-Iranian vowels, modulo certain phonetic positions, such as *i*- and *a*-umlaut, into the modern Shughni-Rushani languages. An example of one such vowel development is the Proto-Iranian long *\*ā*, which becomes Shughni *ō*, Rushani *ū*, Bartangi *ō*, and Sarikoli *u*. Thus, the Proto-Iranian word *\*bāraka-* – via the additional changes of initial *\*b* to *v*, noted above, as well as *\*k > \*č > j* – becomes Shughni *vorj*, Rushani *vūrj*, Bartangi *vorj*, and Sarikoli *vurj*. On the basis of correspondences like this one, Sokolova and Edelman are able to reconstruct the vowel system not only for Proto-Shughni-Rushani, but also for Proto-Shughni-Rushani-Yazghulami (i.e., Proto-North-Pamir).

Notable morphosyntactic features shared by all languages of the subgroup include the use of second-position clitics as agreement markers in the past tense; the preservation of a triple deictic system distinguishing proximal, medial, and distal degrees; and the development of a system of spatial opposition in three fundamental prepositions: *Sh*. *tar* (horizontal plane), *ar* (down), and *pi* (up). Consider the following examples from Rushani (Sokolova 1967: 117):

- (3) a. **Tar** way qīwt.  
to him called.pst.3sg  
'He called to him (along a horizontal plane).'
- b. **Ar** way qīwt.  
down.to him called.pst.3sg  
'He called (down) to him.'
- c. **Pa** way qīwt.  
up.to him called.pst.3sg  
'He called (up) to him.'

Previous research on the Shughni-Rushani languages has focused primarily on Shughni, for which there are a few grammatical sketches, including Bakhtibekov 1979 and Edelman & Dodykhudoeva 2009b, as well as Karamshoev 1963a on its Bajuwi dialect. (See section 1.3 for a more extensive discussion regarding previous research on Shughni). Aside from Shughni, grammatical sketches exist for Rushani (Faizov 1966); Sarikoli (Gao 1963; 1985; Kim 2017; Pakhalina 1966); Bartangi (Karamkhudoev 1973); and Roshorvi (Kurbanov 1976). Sarikoli has received some considerable scholarly attention in recent years, the output of which includes analyses on Sarikoli

diphthongs (Arlund 2006), subordination (Kim 2014), the reflexive pronoun (Kim 2015), and verbal morphology and grammatical aspect (Palmer 2016). It is safe to say, however, that all Pamir languages, including those of the Shughni-Rushani group, have received relatively little attention in linguistics, particularly in comparison with more widely spoken Iranian languages such as Persian, Kurdish, and Pashto.

Finally, note that throughout the literature on Pamir languages, the language varieties of the Shughni-Rushani group are referred to as both ‘languages’ and ‘dialects’, seemingly without much consideration for the implications of these terms. Sokolova (1953) notes that mutual intelligibility is achieved fairly easily amongst speakers of the Shughni-Rushani varieties, particularly among the varieties in Afghanistan and Tajikistan (i.e., excluding Sarikoli). She further speculates that the reason many Shughni-Rushani varieties are referred to as individual languages, rather than dialects, is the lack of a singled standardized variety of which the others could be called dialects. I take no position on this issue in this dissertation and wish simply to acknowledge its existence. Throughout the rest of the dissertation, as a matter of convention, I will refer to each genetic subgroup of the Shughni-Rushani group as a ‘language’ (i.e. the Shughni, Rushani, Bartangi, and Sarikoli languages), and each subvariety as a ‘dialect’ (e.g. Bajuwi dialect of the Shughni language; Khufi dialect of the Rushani language; etc.). Further, unless otherwise noted, a form or construction referenced for a particular variety can be understood to represent the principal dialect (i.e., ‘X form found in Rushani’ represents Rushani proper and not Khufi; and ‘Y construction found in Sarikoli’ represents the Central dialect of Sarikoli; and so on).

## 2.6 Historical and genealogical considerations: Summary

To summarize the crucial points of this section, Shughni belongs to the Iranian branch of Indo-European languages. Due to a number of features, it is further classified as Eastern Iranian, a group which includes Ossetic, Pashto, Yaghnobi, Parachi,Ormuri, and the other Pamir languages, and which opposes the larger Western Iranian group, which includes Persian, Kurdish, Balochi, and many others. The legitimacy of Eastern Iranian as a distinct genetic subgroup within Iranian remains controversial, however, as there does not exist any innovation shared by all Eastern Iranian languages to the exclusion of all Western Iranian languages. It can be said, nonetheless, that Shughni displays many of the features associated with Eastern Iranian and few of those associated with Western Iranian.

Within Eastern Iranian, Shughni is often further subcategorized as a Pamir language, another group of languages

which share many features but which have not been demonstrated to constitute a distinct genetic subgroup among Eastern Iranian languages. Instead, the Pamir languages should be thought of as a kind of areal grouping or *Sprachbund* – that is, a group of Iranian languages spoken in the same geographic region which share a number of phonological, morphosyntactic, and lexical features, due, at least in part, to the extended contact they have had with one another and with a common set of non-Iranian languages also spoken in the area.

Within the Pamir languages, however, there are four accepted groups of genetically related languages: Wakhi, Ishkashimi, Munji-Yidgha, and the North Pamir group. The latter includes Yazghulami and the Shughni-Rushani languages, to which Shughni belongs. The Shughni-Rushani group consists of several varieties spoken in neighboring regions of Tajikistan and Afghanistan, including Shughni, Rushani, Bartangi, and Roshorvi, as well as the Sarikoli language spoken in the nearby Xinjiang Province of China.



## Chapter 3

# Sound system and orthography

This chapter provides an overview of Shughni phonology, including phoneme inventory, phonotactics, stress, and orthography. Section 3.1 first provides the Shughni **phoneme inventory** and discusses the characteristics of individual phonemes and classes of phonemes. Section 3.2 then examines **phonotactics**, including permitted syllable structures and the way the language deals with foreign coda clusters in borrowed words. Section 3.3 turns to **word-level stress**, and then Section 3.4 concludes with an examination of Shughni **orthography**, including the three scripts with which it is typically written – Arabic, Cyrillic, and Latin. This section also presents the orthography to be used in the remainder of the thesis. For the reader who is looking for a brief overview of Shughni phonemes and the orthography used in this thesis, Tables 3.21 and 3.22 provide a succinct look at both.

To my knowledge, the most in-depth work on Shughni phonology comes from Valentina S. Sokolova's (1953) *Очерки по фонетике иранских языков* (*Sketches of the phonology of Iranian languages*), which details the phonological systems of various (Eastern) Iranian languages. The reader is referred to this work for a comparison of the phonology of the Shughni-Rushani languages (excluding Sarikoli), as well as Ossetian and Yaghnobi. A short summary of the phonology of Pamir languages is given in Edelman & Dodykhudoeva 2009a: 777-780, and a more detailed look at Shughni phonology is provided later in the same volume (Edelman & Dodykhudoeva 2009b: 789-792). A brief overview can also be found in Edelman & Yusufbekov 2000: 227-229. Notably absent from all these works, as well as the present thesis, is a thorough examination of intonation, which remains a topic for future work.

### 3.1 Phoneme inventory

The Shughni phoneme inventory consists of thirty-nine total phonemes, including twenty-nine consonants and ten vowels. Its consonant inventory, discussed in Section 3.1.1, is virtually identical to the consonant inventories of the other Shughni-Rushani languages (e.g., Edelman & Dodykhudoeva 2009a). In its vowel system, discussed in Section 3.1.2, Shughni exhibits fundamental characteristics of phonemic vowel length shared by most Shughni-Rushani languages (e.g., Sokolova 1953).

#### 3.1.1 Consonants

This subsection provides the Shughni consonant inventory (Section 3.1.1.1) and examines three important characteristics of the language's consonant system (Section 3.1.1.2).

##### 3.1.1.1 Consonant inventory

Shughni's twenty-nine consonant phonemes are given in Table 3.1. These are the same consonants given in other sources on the language (e.g., Edelman & Dodykhudoeva 2009b: 790; Sokolova 1953: 136), with the exception of the alveolar flap /ɾ/, which is generally listed in these publications as a trill /r/. The status of the rhotic phoneme as a flap or trill is discussed further in Section 3.1.1.2 below.

Note that in Table 3.1, whenever the IPA symbol for a given phoneme differs from the grapheme used in this thesis, the latter is given in angled brackets. However, for the purpose of the discussion on phonological phenomena here, all word forms in the present chapter are given in IPA. The orthography to be used in the remainder of the thesis is presented in Section 3.4.

All Shughni consonants can appear in onset or coda position. Table 3.2 shows each consonant in word-initial, medial, and final positions. Table 3.3 gives examples of minimal pairs involving contrasts in voicing, place of articulation, and manner of articulation.

**Table 3.1:** Shughni consonant phonemes.

MANNER \ PLACE	Bilabial	Labiodental	Interdental	Alveolar	Alveopalatal	Palatal	Velar	Uvular
<b>Plosive</b>	p    b			t        d			k        g	q
<b>Affricate</b>				ṭṣ <c>    ḍḏ <dz>	ṭʃ <č>    ḍʒ <ǰ>			
<b>Fricative</b>		f        v	θ        ð	s        z	ʃ <š>    ʒ <ž>		x <ḫ>    ɣ <ḡ>	χ <x>    ʁ <y>
<b>Nasal</b>	m			n				
<b>Approximant</b>	w					j <y>		
<b>Lat. approx.</b>				l				
<b>Flap</b>				r <r>				

**Table 3.2:** Examples of Shughni consonant phonemes in onset (word-initial) and coda (word-final) positions.

PHONEME	ONSET (WORD-INITIAL)		CODA (WORD-FINAL)	
	PLOSIVES			
p	<b>p</b> ε:χto:w	<i>cook</i>	lap	<i>very</i>
b	<b>b</b> i:r	<i>bed</i>	xa: <b>b</b>	<i>night</i>
t	<b>t</b> i:r	<i>top</i>	sat	<i>went (M)</i>
d	<b>d</b> u:s	<i>a little</i>	ʈʃi: <b>d</b>	<i>house</i>
k	<b>k</b> al	<i>bald</i>	ʈʃal <b>ak</b>	<i>bucket</i>
g	<b>g</b> ul	<i>flower</i>	sø: <b>g</b>	<i>tale</i>
q	<b>q</b> i:ʈʃ	<i>stomach</i>	cilu <b>q</b>	<i>vertical</i>
	AFFRICATES			
ʈs	<b>ʈ</b> sε:m	<i>eye</i>	xa <b>ʈs</b>	<i>water</i>
ɖz	<b>ɖ</b> zalik	<i>small (F)</i>	zari: <b>ɖz</b>	<i>partridge</i>
ʈʃ	<b>ʈ</b> ʃo:r	<i>husband</i>	ʃi <b>ʈʃ</b>	<i>now</i>
ɖʒ	<b>ɖ</b> ʒiri:b	<i>sock</i>	xi: <b>ɖʒ</b>	<i>bull</i>
	FRICATIVES			
f	<b>f</b> iro:χ	<i>wide</i>	so: <b>f</b>	<i>pure</i>
v	<b>v</b> irø:y	<i>eyebrow</i>	ziv	<i>tongue</i>
θ	<b>θ</b> i:r	<i>ash</i>	me: <b>θ</b>	<i>day</i>
ð	<b>ð</b> ust	<i>hand</i>	tu: <b>ð</b>	<i>mulberry</i>
s	<b>s</b> i:vd	<i>shoulder</i>	wu: <b>s</b>	<i>beam</i>
z	<b>z</b> i:rd	<i>yellow</i>	vaz	<i>goat (F)</i>
ʃ	<b>ʃ</b> i:nto:w	<i>laugh</i>	no: <b>ʃ</b>	<i>apricot</i>
ʒ	<b>ʒ</b> o:w	<i>cow</i>	ta <b>ʒ</b>	<i>pull (IMP)</i>
x	<b>x</b> u:vd	<i>milk</i>	va: <b>x</b>	<i>rope</i>
ɣ	<b>ɣ</b> in	<i>wife</i>	ʁø: <b>ɣ</b>	<i>ear</i>
χ	<b>χ</b> u:ðm	<i>sleep</i>	mε <b>χ</b>	<i>nail</i>
ʁ	<b>ʁ</b> a:ʈs	<i>girl</i>	ʈsu <b>ʁ</b>	<i>ripped</i>
	SONORANTS			
m	<b>m</b> aðo:r	<i>noon</i>	ri: <b>m</b>	<i>poplar</i>
n	<b>n</b> i:wdo:w	<i>cry</i>	bø: <b>n</b>	<i>beard</i>
w	<b>w</b> ø:n	<i>wool</i>	yi: <b>w</b>	<i>one</i>
j	<b>j</b> u:rx	<i>bear</i>	no: <b>j</b>	<i>throat</i>
l	<b>l</b> ø:vdo:w	<i>say</i>	bida: <b>l</b>	<i>change</i>
r	<b>r</b> u:ʃt	<i>red</i>	xa: <b>r</b>	<i>city</i>

**Table 3.3:** Examples of minimal pairs in Shughni: Consonants.

VOICING							
<u>p~b</u>		<u>t~d</u>		<u>k~g</u>		<u>f~v</u>	
/pi:r/	‘old’	/ta:k/	‘trousers’	/kaxt/	‘grain’	/fa:m/	‘know!’
/bi:r/	‘bed’	/da:k/	‘give’	/gaxt/	‘return’	/va:m/	‘I bring’
<u>θ~ð</u>		<u>s~z</u>		<u>ʃ~ʒ</u>		<u>x~ɣ</u>	
/θud/	‘burnt’	/sa:r/	‘morning’	/ʃi:rak/	‘(type of insect)’	/xin/	‘hear!’
/ðud/	‘smoke’	/za:r/	‘poison’	/ʒi:rak/	‘stone’	/yin/	‘wife’
<u>χ~ɣ</u>		<u>ts~dz</u>		<u>tʃ~dʒ</u>			
/χa:m/	‘I eat’	/tsuq/	‘vertical’	/tʃo:j/	‘tea’		
/ɣa:m/	‘mourning’	/dzuq/	‘pinch’	/dʒo:j/	‘place’		
PLACE							
<u>f~θ</u>		<u>θ~s</u>		<u>s~ʃ</u>		<u>ʃ~x</u>	
/to:fʃ/	‘stain’	/θits/	‘burnt (F)’	/sut/	‘went (M)’	/tʃu:fʃ/	‘barley’
/to:θtʃ/	‘bowl’	/sits/	‘gone (F)’	/ʃut/	‘lame (M)’	/tʃu:xtʃ/	‘watched’
<u>x~χ</u>		<u>ts~tʃ</u>		<u>v~ð</u>		<u>ð~z</u>	
/xa:r/	‘city’	/tsurt/	‘rustles’	/vud/	‘was (M)’	/ðar/	‘far’
/χa:r/	‘eat!’	/tʃurt/	‘feeling(s)’	/ðud/	‘smoke’	/zar/	‘gold’
<u>z~ʒ</u>		<u>ʒ~ɣ</u>		<u>ɣ~ɣ</u>		<u>dʒ~dʒ</u>	
/zɛ:xt/	‘take (INF)’	/ʒa:f/	‘complain!’	/ɣatʃ/	‘tree bark’	/vu:ɣdʒ/	‘long’
/ʒɛ:xt/	‘run (INF)’	/ɣa:f/	‘Cirisium’	/ɣatʃ/	‘chalk’	/vu:ɣdʒ/	‘brought (PRF)’
MANNER							
<u>s~ts</u>		<u>ʃ~tʃ</u>		<u>z~dz</u>		<u>ʒ~dʒ</u>	
/sa/	‘go!’	/ʃi:nt/	‘laugh (INF)’	/zi:n/	‘kill!’	/ʒaqt/	‘press (INF)’
/ca/	‘(SUBR)’	/tʃi:nt/	‘dig (INF)’	/dʒi:n/	‘daughter (ABBR)’	/dʒaqt/	‘bark (INF)’

### 3.1.1.2 Selected features of the Shughni consonant system

This subsection points out a few important features of the Shughni consonant system. These are: (i) the lack of a glottal fricative /h/, a common feature within the Pamir group but not among Iranian languages more generally; (ii) the existence of a velar pair /x/ and /ɣ/ alongside a uvular pair /χ/ and /ʁ/, a typologically uncommon feature also found in many Pamir languages; and (iii) the status of the rhotic phoneme as a flap /ɾ/ or trill /r/, which has generally not been discussed in the literature on Shughni.

**3.1.1.2.1 Lack of phonemic *h* in Shughni.** The lack of phonemic *h* is a feature common to all Pamir languages, including Shughni (Edelman & Dodykhudoeva 2009a: 777; Payne 1980: 424-425; Sokolova 1953: 135). Nonetheless, the sound [h] occurs as an allophone in Shughni in at least two environments. First, before a word-initial or syllable-initial /a/, there is an optional [h]-like or ‘rough breathing’ sound (Bao 2013; Edelman & Dodykhudoeva 2009a: 777; 2009b: 791; Sokolova 1953: 98), although further research is needed to determine whether this is a full consonant sound or the devoicing of the word-initial vowel, and whether this phenomenon occurs in the same way across all Pamir languages. In Shughni, for example, the word /araj/ ‘three’ may be realized as [haraj] or [araj].<sup>1</sup>

And second, Shughni contains a number of borrowings from Arabic and Tajik, both of which possess the phoneme /h/. In more formal registers, some Shughni speakers may pronounce loanwords as they are pronounced in their source language, particularly those of Tajik origin.<sup>2</sup> The Shughni word /amkor/ ‘co-worker’, for instance, is a borrowing from Tajik – cf. *хамкор* /hamkor/ – and does in fact contain an initial phoneme /h/ in its original form. With this in mind, another reason one might hear an initial [h] in this word in Shughni is the fact that in more formal registers, some Shughni speakers may use its Tajik pronunciation.

<sup>1</sup>The appearance of a word-initial [h]-like sound is akin to the phenomenon whereby word-initial short vowels /i/ and /u/ are optionally preceded by their corresponding glides, [j] and [w], respectively (see Sokolova 1953 on this phenomenon). For instance, the word /is/ ‘feeling’ is realized either [jis] or [is], and the word /uz/ ‘I’ is realized either [wuz] or [uz]. Importantly, however, whereas approximants /j/ and /w/ are independent Shughni phonemes in their own right, the [h]-like rough-breathing sound found before word-initial vowels does not have a corresponding phoneme.

<sup>2</sup>Arabic also has voiced pharyngeal fricative /ħ/, which contrasts with the voiceless glottal fricative /h/. Words borrowed into Tajik (and other Persian dialects) from Arabic retain both of these Arabic phonemes as /h/, as the pharyngeal fricative /ħ/ does not exist phonemically in Persian. Further, many if not all loanwords of Arabic origin in Shughni have entered the language via Persian (Dari and Tajiki), and Shughni therefore inherits the same neutralization of Arabic /ħ/ and /h/.

Given the allophonic status of [h] in Shughni, determining how to represent this sound orthographically is not a trivial matter. This is discussed further in Section 3.4, but for now note that the [h]-like sound preceding word-initial and syllable-initial vowels tends not to be represented orthographically in any script. Phonemic /h/ in borrowed words, for its part, is generally represented in Arabic-based Shughni scripts, but not in Latin- or Cyrillic-based scripts.

**3.1.1.2.2 Velar fricatives /x/ and /ɣ/. Shughni contains a velar fricative pair /x/ and /ɣ/, which contrasts with both the uvular pair /χ/ and /ʁ/ and the post-alveolar pair /ʃ/ and /ʒ/. The existence of these three pairs of fricatives within a single consonant system is found in all North Pamir languages (i.e. the Shughni-Rushani group and Yazghulami) as well as Wakhi. However, this pair does not exist in all Pamir languages, and does not have the same origin in all Pamir languages in which it does exist.**

In the North Pamir group, the modern velar fricatives /x/ and /ɣ/ came about after an older velar pair – shared with some Western Iranian languages – shifted to uvular position, as it did in the Western Iranian language Persian. Unlike in Persian, however, the voided velar space was subsequently filled via further sound shifts in the North Pamir languages. The velar fricatives of the Pamir languages Munji and Yidgha, on the other hand, have a completely different origin (Edelman & Dodykhudoeva 2009a; Sokolova 1973), and in Ishkashimi and Sangilechi such a velar pair does not exist.

As with the sound [h], there are certain orthographic issues involving the velar fricatives /x/ and /ɣ/. In many Latin and Cyrillic scripts, the velar pair is generally represented with a caron (e.g. <ǰ> and <ǧ> in some Latin scripts), while the uvular pair is represented without a caron (e.g. <x> and <y> in the same Latin scripts). In Arabic-based scripts, the letters <ځ> and <چ>, borrowed from the Pashto alphabet, are used to represent /x/ and /ɣ/, respectively, while the Arabic letters <خ> and <غ> represent /χ/ and /ʁ/, respectively.

**3.1.1.2.3 The phoneme /r/** The last issue regarding Shughni consonants concerns the rhotic sound, which has been described by some authors as an alveolar trill /r/ (e.g. Sokolova 1953: 136, who does not note the variation in the pronunciation of this sound) and by others as an alveolar liquid without reference to its specific manner of articulation (cf. Edelman & Dodykhudoeva 2009b: 790, where this phoneme is nonetheless transcribed as <r>).

In my data this sound is generally produced as a flap in all positions – i.e. word-initially, medially, and finally –

although it may be realized as a trilled [r] under certain conditions. One such scenario is when two sounds [r] follow one another, where result is a single trill [r]. Sequences of two adjacent rhotic sounds may occur word-internally, as in the word /tʃurast/ > [tʃurast] ‘murmuring’, or across word boundaries, as in /tar raqs/ > [taraqs] ‘to (the) dance’. Examples of a single rhotic sound [r], which I analyze as a sequence of a single phoneme /r/, and instances of a trilled [r], which I analyze as a sequence of two phonemes /rr/, are given in Table 3.4.

**Table 3.4:** Single and geminate /r/ in Shughni.

Single /r/						Two phonemes /rr/ > [r]		
INITIAL		MEDIAL		FINAL		MEDIAL (ONLY)		
/raqosa:/	‘dance’	/piro:/	‘before’	/duto:r/	‘dutar’	/tʃurrast/ >	[tʃurast]	‘murmuring’
/ra:m/	‘mercy’	/viro:d/	‘brother’	/par/	‘wing’	/arra:/ >	[ara:]	‘saw (tool)’
/re:d/	‘stay’	/firi:pt/	‘arrive’	/ʁa:ra:m/	‘we eat’	/parra:/ >	[para:]	‘edge’

### 3.1.2 Vowels

This subsection is organized similarly to Section 3.1.1 on the Shughni consonant inventory. Here, Section 3.1.2.1 first gives the language’s vowel inventory together with minimal pairs. This is followed by an examination of three characteristics of the Shughni vowel system: (i) the distinction between short and long vowels; (ii) the monophthongization of the diphthong [aj]; and (iii) a number of issues surrounding the vowel phoneme *û*, which I take to be the front vowel [ø].

#### 3.1.2.1 Vowel inventory

Shughni has a total of ten vowel phonemes, of which three are short and seven are long. These are exhibited in Table 3.5. Similar to the presentation of consonants above, an approximate IPA value of each vowel is followed by the orthographic representation used in the remainder of this thesis, given in parentheses. Minimal pairs exhibiting the phonemic status of individual vowel phonemes are given in Table 3.6, where the first column in each pair shows the vowel phoneme in question using the orthography of this thesis. In the second and third columns, the



minimal pair is given in its orthographic and phonetic representations, respectively.

**Table 3.5:** Shughni vowels phonemes

	Front		Central		Back
<b>High</b>	<b>i:</b> ( <i>ī</i> )				<b>u:</b> ( <i>ū</i> )
<b>High-mid</b>	<b>e:</b> ( <i>e</i> ) <b>ø:</b> ( <i>û</i> )	<b>ɪ</b> ( <i>i</i> )		<b>ʊ</b> ( <i>u</i> )	
<b>Low-mid</b>	<b>ɛ:</b> ( <i>ê</i> )			<b>ɔ:</b> ( <i>o</i> )	
<b>Low</b>			<b>ɐ</b> ( <i>a</i> )	<b>ɑ:</b> ( <i>ā</i> )	

**Short vowels (orthographic):** *a, i, u*

**Long vowels (orthographic):** *ā, ī, e, ê, ū, û, o*

**Table 3.6:** Examples of minimal pairs in Shughni: Vowels.

PAIR	ORTH.	PHON.	GLOSS	PAIR	ORTH.	PHON.	GLOSS
<b>a</b>	<i>yax</i>	[jɐχ]	‘sister’	<b>ī</b>	<i>ðīd</i>	[ði:d]	‘hits’
<b>ā</b>	<i>yāx</i>	[ja:χ]	‘ice’	<b>ê</b>	<i>ðêd</i>	[ðɛ:d]	‘war’
<b>i</b>	<i>pic</i>	[pits̩]	‘rotten.F’	<b>ī</b>	<i>ðīd</i>	[ði:d]	‘hits’
<b>ī</b>	<i>pīc</i>	[pi:ts̩]	‘face’	<b>e</b>	<i>ðêd</i>	[ðɛ:d]	‘falls’
<b>u</b>	<i>nur</i>	[nur]	‘today’	<b>e</b>	<i>ser</i>	[se:r]	‘satisfied’
<b>ū</b>	<i>nūr</i>	[nu:r]	‘light’	<b>ê</b>	<i>sêr</i>	[sɛ:r]	‘threshing’
<b>i</b>	<i>sir</i>	[sɪr]	‘secret’	<b>ū</b>	<i>mūd</i>	[mu:d]	‘died.M’
<b>ê</b>	<i>sêr</i>	[sɛ:r]	‘threshing’	<b>o</b>	<i>mod</i>	[mo:d]	‘died.F’
<b>i</b>	<i>yid</i>	[jɪd]	‘(DEM.MED)’	<b>ū</b>	<i>būm</i>	[bu:m]	‘owl’
<b>e</b>	<i>yed</i>	[je:d]	‘bridge’	<b>û</b>	<i>bûm</i>	[bø:m]	‘roof’
<b>u</b>	<i>ðud</i>	[ðud]	‘smoke’	<b>o</b>	<i>kor</i>	[ko:r]	‘work’
<b>o</b>	<i>ðod</i>	[ðo:d]	‘fell’	<b>û</b>	<i>kûr</i>	[kø:r]	‘blind’
<b>u</b>	<i>gul</i>	[gul]	‘flower’				
<b>û</b>	<i>gûl</i>	[gø:l]	‘mute’				

The most significant difference between the configuration of vowels given here is with the vowel which is represented orthographically as *ũ*. Whereas previous authors have traditionally classified this vowel as a back vowel (e.g., Karamshoev 1963a; Edelman & Yusufbekov 2000; Edelman & Dodykhudoeva 2009b: 789; cf. also Sokolova 1953: 86 who notes that this vowel is “slightly moved forward”), I take this vowel to be the front mid vowel [ø:] (with the length symbol used due to the fact that it belongs to the category of long vowels in Shughni, to be discussed below). This is corroborated by acoustic measurements taken recently during fieldwork by researchers at HSE and laid out in Ivanova 2023 (see also Makarov & Plungian 2023, who likewise indicate that *ũ* is a front vowel).

Additionally, the configuration in Table 3.5 indicates that the low vowels *a* and *ā* differ in both quality and quantity. This observation is contra traditional views of the Shughni vowel system, such as the claim made by Sokolova (1953: 98) that “[short] /a/ is equal in quality to [long] /a:/.” It is, however, in line with the acoustic data presented in the report of Ivanova (2023), who provides evidence that *a* and *ā* differ not only in quantity, but also in quality.

### 3.1.2.2 Selected phenomena in the Shughni vowel system

We now turn to three phenomena involving vowels in Shughni, including: (i) the distinction between short and long vowels (Section 3.1.2.2.1); (ii) the contraction or monophthongization of the diphthong /aj/ (Section 3.1.2.2.3); and (iii) a few issues surrounding the phoneme *ũ* (Section 3.1.2.2.2).

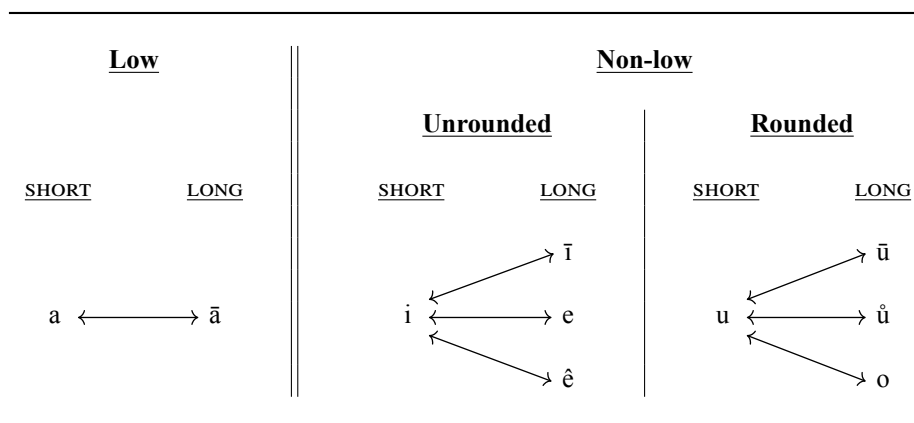
**3.1.2.2.1 On short and long vowels** The ten Shughni vowels have traditionally been categorized into series of short and long, with three short vowels – *a*, *i*, and *u* – and seven long vowels – *ā*, *ī*, *ū*, *e*, *ē*, *o*, and *ũ* (cf., e.g., Edelman & Dodykhudoeva 2009b: 789; Sokolova 1953). However, as noted by Makarov & Plungian (2023), quality alone is sufficient to distinguish Shughni vowel phonemes from one another. Even those vowels generally distinguished orthographically only by a macron – i.e. *a*~*ā*; *i*~*ī*; and *u*~*ū* – in fact differ from one another in quality. In general, the short vowel in each pair is less peripheral than its long counterpart, a notion which is represented in Table 3.5. Note that such differences in quality between long vowels and their short-vowel counterparts are not uncommon in the world’s languages (cf., e.g., Maddieson et al. 2001 who point to similar issues in the Tlingit vowel system, including the idea that long vowels are generally more peripheral than short vowels in this language).

Nonetheless, multiple authors have demonstrated that there are meaningful distinctions in duration between the

group of three vowels traditionally deemed short and the group of seven vowels deemed long. This was first done using phonetic measurements by Sokolova (1953), who notes that the duration of short vowels in monosyllabic words never equal to the duration of long vowels in the same phonetic position. In her study, even the longest variants of short vowels never reached the duration of the shortest variants of long vowels. Ivanova (2023) provides more recent evidence for the presence of short vs. long classes of vowels and demonstrates that distinctions in duration between these two classes are statistically significant.

Beyond the traditional grouping of short and long vowels, each long vowel has been said to correspond in quality to one of the three short vowels. Thus, the low long vowel  $\bar{a}$  corresponds to the low short vowel  $a$ ; each of the three unrounded, front long vowels ( $\bar{i}$ ,  $e$ ,  $\hat{e}$ ) correspond to the unrounded short vowel  $i$ ; and each of the rounded long vowels ( $\bar{u}$ ,  $o$ , and  $\hat{u}$ ) correspond to the rounded short vowel  $u$ . This gives the picture in Figure 3.1. Note that the non-low groups labeled “non-rounded” and “rounded” in Figure 3.1 have traditionally been given the labels “front” and “back”, as the vowel  $\hat{u}$  has generally been analyzed as a back vowel. In light of the recent phonetic evidence provided by Ivanova (2023) indicating that  $\hat{u}$  is in fact a front vowel, I use labels making reference to rounding which more readily distinguish these two groups of vowels.

**Figure 3.1:** Shughni short and long vowel correspondences.



In addition to differences in duration, short and long vowels are also distinguished by their stability in quality. Whereas long vowels are generally stable in quality, short vowels vary considerably depending on phonetic position (cf. Sokolova 1953; see also the short discussion in Edelman & Dodykhudoeva 2009b: 789). The short vowel  $i$

can correspond in quality to the long vowels  $\bar{i}$ ,  $e$ , and  $\bar{e}$ , and the short vowel  $u$  can correspond in quality to the long vowels  $\bar{u}$  and  $o$ . Short  $i$  has its highest (most [i]-like) variant in open syllables word-medially, its middle (most [e]-like) variant in closed syllables, and its lowest (most [ɛ]-like) variant in open syllables word-finally. Likewise, short  $u$  has its highest (most [u]-like) variant in open syllables word-medially, its middle (most [o]-like) variant in closed syllables, and its lowest (most [ɔ]-like) variant word-finally. The allophones of  $i$  and  $u$  are summarized, together with examples, in Table 3.7.<sup>3</sup>

**Table 3.7:** High, middle, and low allophones of short vowels.

<u>Short <math>i</math></u>			
<u>VARIANT</u>	<u>PHONETICALLY</u>	<u>EXAMPLE</u>	<u>GLOSS</u>
<b>High</b>	[i]	<i>siyo</i> > [sijɔ:]	‘black’
<b>Middle</b>	[e]	<i>mis</i> > [mes]	‘also’
<b>Low</b>	[ɛ]	<i>ki</i> > [kɛ]	‘do!’
<u>Short <math>u</math></u>			
<u>VARIANT</u>	<u>PHONETICALLY</u>	<u>EXAMPLE</u>	<u>GLOSS</u>
<b>High</b>	[u]	<i>guno</i> > [gunɔ:]	‘sin’
<b>Middle</b>	[o]	<i>kud</i> > [kod]	‘dog’
<b>Low</b>	[ɔ]	<i>xu</i> > [χɔ]	‘and (then)’

Importantly, although none of the variants of short  $u$  correspond closely in quality to  $\bar{u}$ , there is morphophonological evidence which supports the notion that these vowels nonetheless share a correspondence. Evidence for this comes from a phenomenon whereby certain words shorten via the dropping of their final consonant (as in certain verb forms, such as *sāw* ‘go!’ > *sa*, and denominal locative suffixes, such as *tīr* ‘on’ > *ti* ). When this occurs, the

<sup>3</sup>I follow previous authors in transcribing the short vowels /i/ and /u/ as such. However, as pointed out by Sokolova (1953: 92), the fact that these vowels are written as /i/ and /u/ is merely convention, since, given their range in quality, they could just as easily be written /e/ and /o/.

previous vowel, if long, is also shortened, and the long vowel becomes its short counterpart according to the correspondences given in Figure 3.1. Crucially, we find one example in which the shortening of the imperative form *lūv* ‘say!’, which contains long *ū*, leads to the form *lu*, with short *u*. Further examples of this phenomenon, also providing evidence for the phonological correspondences in Figure 3.1, are given in Table 3.8.

**Table 3.8:** Morphophonological correspondences of short and long vowels: Evidence from verb shortening.

GROUP	CORRESPONDENCE	LONG FORM	SHORT FORM	GLOSS
<b>Low</b>	$\bar{a} \leftrightarrow a$	<i>sāw</i>	<i>sa</i>	‘go!’
<b>Non-low, Unrounded</b>	$\bar{i} \leftrightarrow i$	<i>parjīv</i>	<i>parjī</i>	‘take away!’
	$\bar{e} \leftrightarrow i$	<i>zēz</i>	<i>zi</i>	‘take’
<b>Non-low, Rounded</b>	$\bar{u} \leftrightarrow u$	<i>čūd</i>	<i>ču</i>	‘did’
	$\bar{u} \leftrightarrow u$	<i>lūv</i>	<i>lu</i>	‘say!’

The phenomenon of verb shortening therefore supports the notion that there is not only a phonetic correspondence, but also a morphophonological correspondence between each of the Shughni low vowels and (a group of) long vowels. Although the phenomenon of verb shortening has been pointed out in previous publications, especially the three-volume dictionary of Karamshoev (1988a), it has not been connected to the correspondences of long and short vowels shown in Figure 3.1. A detailed phonological analysis of the Shughni vowel system remains a goal for future investigation, and the morphophonological evidence for vowel correspondences given here may prove a useful component of such an analysis.

**3.1.2.2.2 The vowel *ū*.** The vowel *ū*, which I take to be phonetically the front mid vowel /ø:/, has an eclectic set of sources. It is found not only in native Shughni words, where it occurs as a result of both older and newer sound changes, but also in borrowed words, where it arises due to multiple regular phonological processes in the language.

The oldest source of the vowel *ū* in Shughni is the diphthong *\*āu*, where it arises in words such as *kūr* ‘blind’ and *rūpc* ‘fox’. According to Sokolova (1953: 91), however, instances of *ū* from this source are relatively rare in comparison with newer sources. She gives a list of approximately thirty words which she claims are the only words in which *ū* has this most ancient provenance.

In addition, *û* arises as the result of three regular phonological processes in Shughni. First, it arises from the long vowel *o* before nasals in closed syllables, both in borrowed words such as *tobistûn* ‘summer’ (cf. Tajik *tobiston*), and in native Shughni words such as *pûnd* (etymologically /po:nd/). (For a detailed discussion of this process, including its interaction with processes of derivation, see Makarov & Plungian 2023: 189–190.) Second, *û* arises from the contraction of *aw* and *āw*, a process which we find in both nouns, such as *wûn* ‘wool’ (etymologically /wawn/ – cf. Yazghulami *wawn*, Bartangi *wāwn*), and in verb stems, such as *sûd* ‘goes; becomes’ and *nûd* ‘cries’ (etymologically /sa:wd/ and /na:wd/ – cf. Bartangi *sāwd*, *nāwd*). Finally, Shughni *û* arises in borrowed words from /u/ when adjacent to uvulars and pharyngeal consonants, as in *mûlat* ‘period (of time)’, ultimately from Arabic *muhla*, likely via Tajik *muhlat*, and *tûmat* ‘libel; slander’, also ultimately from Arabic, likely via Persian *tuhmat*.

**3.1.2.2.3 Contraction of the diphthong /aj/.** The final phenomenon to be discussed in this subsection involves the diphthong /aj/ (transcribed in the orthography used here as *ay*), which is found relatively rarely in modern Shughni. According to Sokolova (1953: 88), it is mainly preserved word-initially in borrowings such as *ayb* ‘stain’. Word-medial instances of this diphthong are more rare and apparently represent older borrowings, such as *sayli* ‘walk; stroll’ and *bayt* ‘line (of poetry)’.

Generally, this diphthong undergoes contraction to a long mid-vowel – either *e* or *ê* – in both native and borrowed words. The diphthong *ay* becomes *e* in all instances except when it is next to uvulars, where it becomes *ê*. Thus, from Tajik *najza* ‘spear’, we get Shughni *nezā*, but from Tajik *xayr* ‘OK; alright’, we get Shughni *xêr*. In native words, we can compare, for instance, the Shughni verb *rimed* ‘orders; commands’ with the Bartangi form *rimayd*, which provides evidence that this process targets not only borrowed words, but also native words. Nonetheless, it is not immediately clear why other instances of *ay* persist in Shughni, as they do in the words mentioned above.

## 3.2 Phonotactics

This section examines Shughni phonotactics, including syllable shapes and consonant clusters found in both native Shughni words and borrowed words. Section 3.2.1 first provides an overview of Shughni phonotactics and permitted syllable shapes in the language. Section 3.2.2 then looks at word-final clusters, including those found in monomorphemic words and those which arise only via verb inflection. Section 3.2.3 examines the resolution of

non-permitted word-final clusters in borrowed words, and lastly, Section 3.2.4 discusses social aspects of Shughni phonotactics. Note that in this section and the following one on stress, transcription is mostly done with the orthography used in this thesis. Where applicable, the IPA is used, generally with a form given in slashes.

### 3.2.1 Overview and permitted syllable shapes

Onset clusters in Shughni are prohibited, while coda clusters are allowed and quite prevalent (e.g., Edelman & Yusufbekov 2000: 229; Edelman & Dodykhudoeva 2009b: 791). Any of the twenty-nine consonants in the phoneme inventory of Shughni can appear as the sole consonant in syllable onset position or as the sole consonant in syllable coda position (see Table 3.2). Permitted syllable shapes in Shughni are shown in Table 3.9.

**Table 3.9:** Permitted syllables shapes in Shughni.

SHAPE	EXAMPLE	GLOSS
<b>V</b>	<i>ā</i>	(vocative particle)
<b>VC</b>	<i>is</i>	‘feeling’
<b>CV</b>	<i>tu</i>	‘you (sg)’
<b>VCC</b>	<i>axk</i>	‘speaking (not mute)’
<b>CVC</b>	<i>mot</i>	‘tired’
<b>CVCC</b>	<i>zīrd</i>	‘yellow’
<b>CVCCC</b>	<i>šarθk</i>	‘clay’

Onset clusters found in loanwords are resolved with an epenthetic vowel (e.g. Russ. примерно *primerno* > Sh. *pirimerna* ‘approximately’). Loans containing onset clusters are almost always from Russian, where onset clusters are typical, rather than Persian or Arabic, as onset clusters are prohibited in these languages.

### 3.2.2 Coda clusters

Shughni allows a wide variety of coda clusters in monomorphemic words. Coda clusters in native Shughni words generally have a stop, affricate, fricative, or nasal in final position. Examples of coda clusters permitted word-finally in monomorphemic native Shughni words are shown in Table 3.10. Note that the cluster in question is given in the IPA (with its orthographic form in parentheses when it differs from the IPA form), while example words in this table and subsequent tables are given in the orthography used in this thesis. Some of the words in

this table, such as *jāld* ‘fast’, may be Persian borrowings. To my knowledge, however, none of the words in this table are Arabic borrowings, which may have distinctive coda clusters discussed below.

In addition to clusters found in monomorphemic words, such as those given in Table 3.10, four more types of consonant clusters arise in inflected verbs: STOP-STOP, AFFRICATE-STOP, STOP-AFFRICATE, and AFFRICATE-AFFRICATE. These clusters arise when a verb stem with a final stop or affricate is followed by an inflectional morpheme, which may be the third-person singular present ending or past ending (either /-t/ or /-d/) or a perfect ending (/t͡s/, /t͡ʃ/, /d͡z/, or /d͡ʒ/). These types of clusters, which are not found outside inflected verbs, are given in Table 3.11.

Clusters of three consonants are possible and are found primarily in inflected verb forms. There are nonetheless a few monomorphemic words with word-final clusters containing three consonants – e.g. *šarθk* (/ʃarθk/) ‘mud’. Inflected verb forms which contain three final consonants consist of a stem ending in two consonants with the addition of an inflectional morpheme. Thus, for instance, the verb with infinitive form *niḍv̄dow* ‘to stick (INTR)’ has present stem *niḍaf̄c-* (/niḍaft͡s-/), and when the third-person singular present inflection /-t/ is added to the present stem, the result is a word-final cluster of three consonants: *niḍaf̄ct* (/niḍaft͡st/). Further examples of clusters containing three consonants are given in Table 3.12.



**Table 3.10:** Consonant clusters in native Shughni words.

STOP-FINAL			AFFRICATE-FINAL			FRICATIVE-FINAL			NASAL-FINAL		
<u>Fricative-Stop</u>			<u>Fricative-Affricate</u>			<u>r-Fricative</u>			<u>Fricative-Nasal</u>		
<b>xp</b> ( <i>x̣p</i> )	<i>tu<u>x̣p</u></i>	‘sour’	<b>fts</b> ( <i>f̣c</i> )	<i>sif<u>c</u></i>	‘beads’	<b>rθ</b> ( <i>rθ</i> )	<i>var<u>θ</u></i>	‘both’	<b>xm</b> ( <i>x̣m</i> )	<i>pa<u>x̣m</u></i>	‘wool’
<b>st</b>	<i>ðu<u>st</u></i>	‘hand’	<b>xts</b> ( <i>x̣c</i> )	<i>amba<u>x̣c</u></i>	‘pine tree’	<b>rx</b> ( <i>ṛx̣</i> )	<i>kir<u>x̣</u></i>	‘slippery’	<b>ðm</b>	<i>xū<u>ðm</u></i>	‘sleep’
<b>ft</b> ( <i>št</i> )	<i>rū<u>št</u></i>	‘red’	<b>χts</b> ( <i>xc</i> )	<i>poyta<u>xc</u></i>	‘gift’	<b>rχ</b> ( <i>rx</i> )	<i>čar<u>rx</u></i>	‘cycle’			
<b>xt</b> ( <i>x̣t</i> )	<i>gū<u>x̣t</u></i>	‘meat’	<b>ft̪</b> ( <i>f̣č</i> )	<i>to<u>f̣č</u></i>	‘stain’	<b>rv</b> ( <i>rv</i> )	<i>wū<u>rv</u></i>	‘boiling’			
<b>θk</b>	<i>šar<u>θk</u></i>	‘clay’	<b>θt̪</b> ( <i>θč</i> )	<i>to<u>θč</u></i>	‘bowl’	<b>rð</b> ( <i>rð</i> )	<i>zor<u>ð</u></i>	‘heart’			
<b>sk</b>	<i>divū<u>sk</u></i>	‘snake’	<b>ʃt̪</b> ( <i>šč</i> )	<i>čū<u>šč</u></i>	‘barley’	<b>rɣ</b> ( <i>ṛɣ̣</i> )	<i>wū<u>ṛɣ̣</u></i>	‘wool’	<u>r-Nasal</u>		
<b>xk</b> ( <i>x̣k</i> )	<i>yū<u>x̣k</u></i>	‘tear’	<b>xt̪</b> ( <i>x̣č</i> )	<i>ra<u>x̣č</u></i>	‘nit (louse egg)’				<b>rm</b> ( <i>rm</i> )	<i>čī<u>rm</u></i>	‘worm’
<b>vd</b>	<i>wū<u>vd</u></i>	‘seven’	<b>χt̪</b> ( <i>xč</i> )	<i>mu<u>xč</u></i>	‘crop-eared’				<b>rn</b> ( <i>rn</i> )	<i>xū<u>rn</u></i>	‘crow’
<b>zd</b>	<i>sī<u>zd</u></i>	‘silverberry’	<b>vd̪</b> ( <i>vdz</i> )	<i>sā<u>vdz</u></i>	‘green’						
<b>ʒd</b> ( <i>žd</i> )	<i>yā<u>žd</u></i>	‘dirty’	<b>ɣd̪</b> ( <i>ɣdz</i> )	<i>fīrē<u>ɣdz</u></i>	‘flea’						
<b>zg</b>	<i>pibiz<u>g</u></i>	‘bladder’	<b>ɣd̪ʒ</b> ( <i>ɣ̣j̣</i> )	<i>yo<u>ɣ̣j̣</u></i>	‘flour’						
<u>Nasal-Stop</u>			<u>Nasal-Affricate</u>								
<b>mb</b>	<i>nā<u>mb</u></i>	‘moisture’	<b>nts</b> ( <i>nc</i> )	<i>civī<u>nc</u></i>	‘wasp’						
<b>nd</b>	<i>cū<u>nd</u></i>	‘how much’	<b>nt̪</b> ( <i>nč</i> )	<i>šū<u>nč</u></i>	‘laughter’						
<b>ng</b>	<i>rā<u>ng</u></i>	‘type’	<b>nd̪</b> ( <i>ndz</i> )	<i>pī<u>ndz</u></i>	‘five’						
<u>r-Stop</u>			<b>nd̪ʒ</b> ( <i>ṇj̣</i> )	<i>gā<u>ṇj̣</u></i>	‘dirty’						
<b>rt</b> ( <i>rt</i> )	<i>čur<u>t</u></i>	‘mood’	<u>r-Affricate</u>								
<b>rk</b> ( <i>rk</i> )	<i>pār<u>k</u></i>	‘leaf’	<b>ṛts</b> ( <i>rc</i> )	<i>kur<u>rc</u></i>	‘deep’						
<b>rd</b> ( <i>rd</i> )	<i>zī<u>rd</u></i>	‘yellow’	<b>rt̪</b> ( <i>rč</i> )	<i>xar<u>rč</u></i>	‘food’						
<b>rg</b> ( <i>rg</i> )	<i>pūr<u>g</u></i>	‘mouse’	<b>rd̪</b> ( <i>rdz</i> )	<i>vēr<u>rdz</u></i>	‘mare’						
<u>l-Stop</u>			<b>rd̪ʒ</b> ( <i>ṛj̣</i> )	<i>vor<u>ṛj̣</u></i>	‘steed’						
<b>lt</b>	<i>tul<u>t</u></i>	‘mess’									
<b>ld</b>	<i>ǰāl<u>d</u></i>	‘fast’									

**Table 3.11:** Consonant clusters in Shughni inflected verbs.

<u>Stop-Stop</u>			<u>Stop-Affricate</u>		
<b>pt</b>	<i>fir̄pt</i>	‘arrived’	<b>pt̪</b> ( <i>pč</i> )	<i>fir̄pč</i>	‘arrived (PRF)’
<b>kt</b>	<i>lākt</i>	‘leave; allow (PRS OR PST)’	<b>kt̪</b> ( <i>kč</i> )	<i>lākč</i>	‘left; allowed (PRF)’
<b>qt</b>	<i>lāqt</i>	‘swing; rock (PRS OR PST)’	<b>qt̪</b> ( <i>qč</i> )	<i>lāqč</i>	‘swung; rocked (PRF)’
<u>Affricate-Stop</u>			<u>Affricate-Affricate</u>		
<b>t̪st</b> ( <i>ct</i> )	<i>wuct</i>	‘move; budge (PRS OR PST)’	<b>t̪st̪</b> ( <i>cč</i> )	<i>wucč</i>	‘moved (PRF)’
<b>d̪zd</b> ( <i>dzd</i> )	<i>and̪dzd</i>	‘wake up (PRS OR PST)’	<b>d̪zd̪</b> ( <i>dzj</i> )	<i>and̪dzj</i>	‘woken up (PRF)’

**Table 3.12:** Clusters of three consonants in Shughni.

<u>Monomorphemic</u>		
<b>rθk</b> ( <i>rθk</i> )	<i>šarθk</i>	‘clay’
<u>Polymorphemic</u>		
<b>rθt</b> ( <i>rθt</i> )	<i>q̄rθt</i>	‘snore(d) (SG.PRS OR PST)’
<b>rθt̪</b> ( <i>rθč</i> )	<i>q̄rθč</i>	‘snored (PRF)’
<b>ft̪st</b> ( <i>fct</i> )	<i>wiðafct</i>	‘stick (INTR, 3SG.PRS)’
<b>x̪t̪st</b> ( <i>ǰct</i> )	<i>pēǰct</i>	‘ask(ed) (3SG.PRS OR PST)’
<b>x̪t̪st̪</b> ( <i>ǰcč</i> )	<i>pēǰcč</i>	‘asked (PRF)’
<b>χ̪t̪st</b> ( <i>xct</i> )	<i>angaxct</i>	‘get stuck (3SG.PRS)’
<b>χ̪t̪st̪</b> ( <i>xcč</i> )	<i>angaxcč</i>	‘got stuck (PRF)’

An additional group of clusters are not found in native Shughni words, but appear in Arabic loanwords and are permitted in the language without modification. In some cases, these represent combinations of sounds which are not found in native words, including NASAL-FRICATIVE (e.g. *sinf*), *l*-NASAL (e.g. *ilm*), and *l*-FRICATIVE (e.g. *qulf*). In other cases, borrowed words contain a type of cluster which is found in native words, but a particular combination of sounds which is not found in native words, as is the case for words like *dars* (*r*-FRICATIVE). Clusters of this type are shown in Table 3.13.

**Table 3.13:** Permitted clusters in Arabic borrowings.

CLUSTER	EXAMPLE	GLOSS
<b>nf</b>	<i>sinf</i>	‘class’
<b>lm</b>	<i>ilm</i>	‘science’
<b>lf</b>	<i>qulf</i>	‘lock’
<b>lʁ (lɣ)</b>	<i>talɣ</i>	‘bitter’
<b>rs (rs)</b>	<i>dars</i>	‘class’
<b>sm</b>	<i>qism</i>	‘part’

### 3.2.3 Resolution of illicit coda clusters in loanwords

Whereas some word-final clusters not found in native Shughni words are permitted to remain in their original form, others are not permitted to remain in their original form and are modified to conform to Shughni phonotactics. The vast majority of these clusters are found in words of Arabic origin and contain a final *l* or *r*, a pattern which betrays Shughni’s general rejection of coda clusters with final *l* or *r*.

Most illicit clusters in word-final position are resolved with a word-final release vowel, which permits the final consonant to be resyllabified as an onset. When this happens, stress remains on the penultimate syllable (e.g. *fikr* > *fik.ri*), making words of this type easily distinguishable, as stress otherwise almost always falls on the final syllable of a monomorphemic word in Shughni. In a few cases, rather than a word-final release vowel, an epenthetic vowel is inserted between the consonants of the illicit combination (e.g. *‘qa.bar* ‘grave’ < Ar. *qabr*). Even in these cases, stress remains on the penultimate syllable and not on the epenthetic vowel. Examples of illicit clusters which are resolved are given in Table 3.14.

**Table 3.14:** Resolved illicit clusters in Arabic borrowings.

CLUSTER	ORIGINAL	RESOLVED	GLOSS
<u>FINAL <i>r</i></u>			
<b>tr</b> ( <i>tr</i> )	<i>atr</i>	<i>átri</i>	‘perfume’
<b>kr</b> ( <i>kr</i> )	<i>fīkr</i>	<i>fíkri</i>	‘thought’
<b>mr</b> ( <i>mr</i> )	<i>umr</i>	<i>úmri</i>	‘life’
<b>sr</b> ( <i>sr</i> )	<i>asr</i>	<i>ásri</i>	‘century’
<b>zr</b> ( <i>zr</i> )	<i>uzr</i>	<i>úzri</i>	‘evidence’
<b>ʃr</b> ( <i>ʃr</i> )	<i>našr</i>	<i>našri</i>	‘publication’
<b>χr</b> ( <i>xr</i> )	<i>faxr</i>	<i>fáxri</i>	‘pride’
<b>br</b> ( <i>br</i> )	<i>qabr</i>	<i>qábar</i>	‘grave’
<u>FINAL <i>l</i></u>			
<b>tl</b>	<i>qatl</i>	<i>qátli</i>	‘murder’
<b>ql</b>	<i>naql</i>	<i>náqli</i>	‘narration’
<b>fl</b>	<i>tifl</i>	<i>tifli</i>	‘infant’
<b>sl</b>	<i>nasl</i>	<i>násli</i>	‘generation’
<b>jl</b> ( <i>yl</i> )	<i>sayl</i>	<i>sáyli</i>	‘outing’
<u>OTHER</u>			
<b>tn</b>	<i>matn</i>	<i>mátni</i>	‘text’

### 3.2.4 Social aspects of Shughni phonotactics

A final aspect of Shughni phonotactics to be discussed in this subsection regards the influence of certain social factors on the pronunciation of syllable shapes and consonant clusters which are typically prohibited in the language. In general, it can be said that prestige registers of Shughni are more permissive of foreign coda clusters, both in onset position and in coda position. Thus, in higher registers, for instance, one may be more likely to hear unresolved onset clusters such as *pr* in Russian *primerno* ‘approximately’, or unresolved illicit coda clusters such as *kr* in *fīkr* ‘thought’ or *tn* in *matn* ‘text’.

Although no formal study exists on this topic for Shughni, related sociolinguistic phenomena have been documented for Tajik. Windfuhr & Perry (2009: 428), for instance, have the following to say about Russian loanwords

in Tajik:

*In Tajik, Russian borrowings have added alien initial clusters to the corpus, as zveno ‘team, unit’, Stalinobod (former name of Dushanbe). Such loanwords have been reproduced in their original orthographies (regarded as normative) in successive writing systems; the degree to which speakers assimilate them to native phonotactic and phonetic norms depends partly on non-linguistic factors such as their level of education and cultural predilections.*

These authors note further that in Persian more generally, certain phonological phenomena associated with Arabic borrowings – for instance, geminate consonants (a marginal feature in Persian phonology); the glottal stop (found only in Arabic borrowings such as *ba’d* ‘after’), and post-consonantal final *h* (e.g. Ar. *subh* ‘morning’) – are typically only realized in high registers and slow articulation.

In the same vein, speakers of Shughni who have extensive formal education and are aware of the phonological and phonotactic nuances of Persian, Russian, Arabic, and even English, may allow for some foreign syllable shapes and clusters in their speech. An intriguing point for future research involves the types of speech acts in which this occurs and the degree to which it occurs in Shughni.

### 3.3 Stress

Shughni prosody has not been examined in detail in any work, although brief remarks are made by Edelman & Yusufbekov (2000: 228-229) and Edelman & Dodykhudoeva (2009b). Basic prosodic characteristics include the lack of lexical tone, the fall in pitch at the end of declarative sentences, and the general “distinctive interrogative intonation” associated with interrogative sentences (Edelman & Dodykhudoeva 2009b: 791-792). This section focuses on word stress, an aspect of Shughni prosody which is relatively well documented, although the acoustic properties of stress in the language have yet to be examined.

Stress in Shughni is not lexically contrastive and almost always falls on the final syllable in multisyllabic words.<sup>4</sup>

---

<sup>4</sup>A few exceptions to word-final stress in monomorphemic words are pointed out by Olson (2017), namely *xúbaθ* ‘self (EMPH)’; *áraj* ‘three’, and *nála* (tenseless particle introducing direct speech), each of which have stress on the first syllable. The emphatic pronoun *xúbaθ*

However, word stress is relatively weak and is generally subordinate to phrasal stress (Edelman & Yusufbekov 2000; Edelman & Dodykhudoeva 2009b). In polymorphemic words, stress can end up on non-final syllables, with (i) derivational affixes, (ii) inflectional affixes, and (iii) case-like suffixes and morphophonological clitics all behaving differently with regard to stress. The vast majority of derivational suffixes attract stress, while most if not all derivational prefixes do not attract stress. Inflectional affixes, on the other hand, seem to be lexically specified for stress: plural suffixes and verbal negation prefixes are stressed, while verbal agreement suffixes are unstressed. Morphophonological clitics – i.e. phonologically dependent morphemes which are not selective regarding the category of the word or phrase to which they attach (following Zwicky and Pullum 1983) – are always unstressed. Each of these categories, as they relate to stress, is discussed in further detail in the subsections below: derivational affixes in Section 3.3.1, inflectional affixes in Section 3.3.2, and unstressed elements, including case-like suffixes and morphophonological clitics in Section 3.3.3. Lastly, Section 3.3.4 offers a few remarks regarding stress in loanwords.

### 3.3.1 Stress and derivational affixes

Most derivational suffixes attract stress, including both productive suffixes (e.g., the nominalizing suffix *-i*, the diminutive suffixes *-ak* and *-ik*, the augmentative *-aki*, the adjectival participle *-in*, and the material suffix *-in* meaning ‘made of’; etc.), as well as non-productive suffixes such as the relational adjective suffix *-in̄j* and the suffix *-doj̄*. Further examples are given in Table 3.15.

Shughni does not possess many productive prefixes; nonetheless, a few productive derivational prefixes have been borrowed from Tajik, including */be:-/* ‘without’, *ba-* ‘with’, and *no-* ‘not’. Such prefixes typically appear within loanwords of which they are a part, such as *bemazá* ‘tasteless; not tasty’; *babarakát* ‘blessed’, and *nobalád* ‘unfamiliar’, but they may also appear together with native Shughni words, as in *bexúðm* ‘sleepless’. In any case, these prefixes do not attract stress and therefore follow the general pattern whereby word-final stress is not affected by derivational affixes.

One exception to the rule that derivational morphology conforms to word-final stress is the adverb-forming suffix *-*

---

‘self’ (examined in Section 6.1.4) may have acquired its exceptional stress pattern via analogy with words containing the adverb-forming suffix *-aθ* (e.g., *jáldaθ* ‘quickly’), which does not attract stress (Edelman & Yusufbekov 2000: 228). Etymologically, however, *xúbaθ* does not contain this morpheme (e.g. Sokolova 1967).

**Table 3.15:** Stress-attracting affixes in Shughni.

<b>Productive Suffixes</b>			
<u>SUFFIX</u>	<u>GLOSS</u>	<u>EXAMPLE</u>	<u>GLOSS</u>
-i	NMZ	<i>garmí</i>	‘warmth’
-ak	DIM	<i>qũ ‘lák</i>	‘small lake’
-ik	DIM	<i>kamkamík</i>	‘little by little’
-aki	AUG	<i>yulayakí</i>	‘enormous’
-in	ADJ.PTCP	<i>tícín</i>	‘gone.F’
-in	‘made of’	<i>đorgín</i>	‘wooden’
-dor	‘having’	<i>búndór</i>	‘bearded’
-gũy	‘speaking’	<i>rostgũy</i>	‘one who speaks the truth’
<b>Non-Productive Suffixes</b>			
<u>SUFFIX</u>	<u>GLOSS</u>	<u>EXAMPLE</u>	<u>GLOSS</u>
-ĩnj	relational ADJ	<i>biyorĩnj</i>	‘yesterday’s’
-doj	agentive?	<i>ýewdój</i>	‘hunter’
-eĵ	appointment?	<i>kurtayėĵ</i>	‘cloth used for making clothes’
-ĩdz	agentive?	<i>xābĩdz</i>	‘stick (used for whipping)’

*aθ*, a productive suffix which attaches primarily to nouns to form adverbs of time (e.g. *sārakí* ‘morning’ > *sārakíyaθ* ‘in the morning’) and adjectives to form adverbs of manner (e.g. *beyam* ‘relaxed’ > *beyamaθ* ‘in a relaxed manner’).

### 3.3.2 Stress and inflectional affixes

Inflectional affixes in Shughni are more variable with respect to stress than derivational affixes. On the one hand, final stress is maintained by nominal plural suffixes (including the most productive plural suffix *-en*, as well as non-productive suffixes such as *-orĵ*, *-galā*, and *-jev*) and the comparative suffix *-di*. On the other, word-final stress does not occur with two important series of inflectional morphemes: the verbal negation prefixes *na-* (indicative) and *mā-* (prohibitive/subjunctive), which attract stress, and the person-number present agreement suffixes, which do not attract stress.

Stress in the present-tense verbal complex, in particular, follows a rather different set of rules than in words of other categories, and therefore merits special attention here. Stress in present-tense verbs can be described as follows.

In affirmative present-tense verbs (i.e. in the absence of a negation prefix), stress falls on the final syllable of the verb stem. Thus, in the case of monosyllabic stems such as *sāw-* ‘go; become’, we have present-tense forms such as *sāwī* ‘you (SG) go’ and *sāwām* ‘we go’ (cf. present-tense agreement suffixes *-i* (2SG) and *-ām* (1PL)). In the case of bisyllabic stems such as *firāp-* ‘arrive’, we have forms such as *firāpum* ‘I arrive’ (cf. present-tense agreement suffix *-um*) and *firāpi* ‘you (SG) arrive’.

If a negation prefix is present, whether *na-* (IND) or *mā-* (PROH/SBJV), then primary stress in the verbal complex is on the negation prefix, and the final syllable of the verb stem receives secondary stress (Edelman & Yusufbekov 2000: 228). Thus, to use the same verbs as above as examples, we have forms such as *nāsāwi* ‘you don’t go’ and *māsāwi* ‘(so that) you don’t go’ for a monosyllabic stem, and *naḡfirāpi* ‘you don’t arrive’ and *māḡfirāpi* ‘(so that) you don’t arrive’ for a polysyllabic stem.

### 3.3.3 Unstressed elements: Case-like suffixes and morphophonological clitics

Shughni has at least two types of phonologically dependent morphemes which generally do not receive stress and which attach at the right edge of a phonologically independent word. The first is case-like suffixes, including *-ti* ‘on’, *-and* ‘in’, *-ard* ‘toward; around’, *-va* ‘along’, and *-ec* ‘until’. The latter suffix is commonly found together with the preposition *to* ‘until’, as in *to iyūnec* ‘until June’ and *to xābec* ‘until the evening’. Note that in both cases, stress falls on the nominal object of the adpositional phrase and not on the suffixal element. One possible exception is the suffix *-aḡ*, a locative suffix which expresses movement along or toward a location and is invariably stressed, as in *tar cīdāḡ* ‘(toward) home’.

The second type of morpheme in question is that of second-position morphophonological clitics, of which Shughni has at least two types: (i) a person-number clitic which is also used to index the person-number features of past-tense subjects and present-tense indicative copular subjects (see Section 4.4.3.1 on these morphemes), and (ii) a factual clitic *=ta* used to indicate habitual actions and future temporal reference (see Section 10.2.1.2). Both types target the right edge of the first syntactic (or prosodic) phrase of a clause, excluding any complementizers or conjunctions, and neither type receives stress under any circumstances.

The examples in (4) and (5) illustrate. In each example, the clitic (bolded) attaches to a complex phrase – the subject NP in (4) and a fronted PP in (5) – and in each case stress falls on the final stress-attracting element of the word



(underlined). In both examples, the stress-attracting morpheme is the nominal plural suffix *-en* (homophonous with the 3PL agreement clitic), but in the first example the unstressed possessive/locative suffix *-and* intervenes between the two.

- (4) [Wev nān-en-and]<sub>NP</sub>=**en** lap xušrūy čīd-en.  
 their mother-PL-POSS=COP.3PL very beautiful house-PL

‘Their mothers have very beautiful houses.’

- (5) [Tar māš xār-en]<sub>PP</sub>=**ta** yad-en.  
 to our city-PL=FAC come.PRS-3PL

‘They will come to our cities.’

### 3.3.4 Stress in loanwords

Stress in loanwords sometimes exhibits deviations from the general rule that stress is word-final in monomorphemic words. I briefly describe two such cases here.

First, as discussed in Section 3.2.3, a number of Arabic loanwords contain word-final consonant clusters which are not permitted in Shughni and are resolved in the language through the use of an unstressed epenthetic short *i*. Such words include Ar. *umr* ‘life’ > Sh. *úmri*; Ar. *fikr* ‘thought’ > Sh. *fíkri*; and Ar. *naql* ‘narration’ > Sh. *náqli*.

Second, there are many Russian loanwords in Shughni which do not conform to the pattern of word-final stress. As discussed in Section 3.2.4, a number of sociolinguistic factors are at play in determining not only the extent to which Russian loanwords are used, but also the extent to which they are adapted to Shughni phonotactics. In my observations, it is much less common for Shughni speakers to adapt stress in Russian loans to Shughni stress patterns than to adapt Russian coda clusters to Shughni phonotactic constraints. For instance, a loan like *primérno* ‘approximately’ is much more likely to have its initial consonant cluster resolved with an epenthetic vowel (*pirímerna*) than to have its stress shifted to the final syllable (*p(i)rimerná*). Moreover, in many cases where a Russian word has penultimate stress, the final unstressed vowel is dropped in its Shughni realization. For instance, Russian *pečénie* ‘cookie’ becomes Shughni *pičén*, and Russian *zubnáya pasta* ‘toothpaste’ becomes Shughni *zubnoy pāst*.<sup>5</sup> Ultimately, however, more research is needed to determine precisely how speakers deal with stress patterns in Russian borrowings which do not conform to those of Shughni.

<sup>5</sup>The neutralization of Russian gender endings (cf. *zubnaya* > *zubnój* ‘dental’) is an intriguing point for future research.

## 3.4 Orthography

Shughni is written using alphabets derived from Latin, Cyrillic, and Arabic scripts. Because there is no single accepted governing body which regulates the Shughni language, there does not exist a single standard version for any of the three types of Shughni alphabets, and therefore we find numerous different versions of each script.

The choice of which style of alphabet is used depends, for the most part, on geographical location. Thus, as one might expect, Cyrillic-based scripts are traditionally used in Tajikistan, a former Soviet state whose official language, Tajik, is written using the Cyrillic alphabet, and where Russian, which enjoys official status as the country's language of "interethnic communication", is still commonly used in many settings, including higher education. Likewise, in Afghanistan, where official languages Dari and Pashto are written with Arabic-based scripts, Shughni also tends to be written with an Arabic-based script. Latin-based alphabets, for their part, are often used by professional linguists in scientific publications on Shughni, though they are gaining popularity as a means of everyday written communication among Shughni speakers in Tajikistan, particularly among young people.

In Latin- and Cyrillic-based Shughni scripts, the biggest discrepancies are found in the representation of certain consonant phonemes, while vowel letters remain relatively stable. In Arabic-based scripts, the situation is reversed: consonant letters remain relatively stable, while the most complex issue involves devising a system to represent the ten vowels of Shughni.

Shughni orthography is a complex topic, and this section presents a considerable amount of information on each type of script used to represent Shughni. For the reader who wishes only to see the script used in this thesis, Section 3.4.4 provides a table of the graphemes used and a brief explanation of certain choices. In the remainder of this section, each type of script is presented in turn: Latin (3.4.1), Cyrillic (3.4.2), and Arabic (3.4.3). Section 3.4.4 then lays out the writing system used to represent Shughni in this thesis. Section 3.4.5 closes by displaying a short passage from a novel written in Shughni using each of the three scripts.

### 3.4.1 Latin script

The first Western scholars to produce work on Shughni used Latin-based alphabets for the language – e.g. Shaw (1877) and Morgenstierne (1938a; 1974). Latin-based scripts were also employed by several authors publishing from the Soviet Union, including Dodykhudoeva (1988), Edelman (1986; 1990; 2009), Sokolova (1953; 1967; 1973), and Zarubin (1960). Virtually all recent English-language publications also use some variation of a Latin-based transcription – e.g. Payne (1980), Barie (2009), Edelman & Dodykhudoeva (2009b), Stump & Hippisley (2011), Mueller (2014), and Parker (2020). Many recent Russian-language publications also employ a Latin-based script for Shughni (cf. Armand 2022 and Chistiakova 2022).

#### 3.4.1.1 Notes on Latin consonant letters

Many of the Latin letters used for Shughni consonants correspond to the same phoneme as it is written in English (e.g. <p> for /p/, <g> for /g/, and so on). However, consonant phonemes traditionally written in English as digraphs are generally transcribed in Shughni as single graphemes with diacritics, which helps preserve a one-to-one correspondence between graphemes and phonemes. Examples include <Šš> for /ʃ/, <Žž> for /ʒ/, <Čč> for /tʃ/; and <Ĵĵ> for /dʒ/. The phonemes /θ/ and /ð/, for their part, have traditionally been written in Shughni with their respective IPA characters (with capital letters <Θ> and <Δ>, respectively).

The phoneme /ts/ is represented with <Cc>, a transcription method often used for the same Russian phoneme. The phoneme /dz/, for its part, has been represented in many scholarly works on the language with two distinct graphemes for capital and lowercase letter. The uppercase letter is a version of <Z> modified with a hook, giving <Ž>, and the lowercase letter is the Cyrillic letter *Abkhazian dze* <з>, used to represent the same phoneme in the Northwest Caucasian language Abkhaz.

A noteworthy peculiarity of the Latin script (and Cyrillic, as discussed below), is that the grapheme <Xx> corresponds to the voiceless uvular fricative /χ/, while the voiceless velar fricative /x/ is represented with <Ǧǧ>. Similarly, the grapheme <y> corresponds to the voiced uvular fricative /ʁ/, while the grapheme <ǧ> is used to represent the voiced velar fricative.

Finally, as discussed above, Shughni has a complicated relationship with the sound [h], which does not exist as a native Shughni phoneme, but comes to the language along with numerous Arabic and Persian loanwords. Latin

scripts, including all those mentioned here and the one used in this thesis, opt not to represent /h/ in borrowed words and therefore have no use for a grapheme <h>.

### 3.4.1.2 Notes on Latin vowel letters

Vowel letters in Latin-based Shughni scripts are relatively stable. The only fundamental difference in choice of vowel characters is for the vowel /ɛ:/, which Sokolova represents with the grapheme <ē>. In the majority of subsequent works (e.g. Edelman & Dodykhudoeva 2009b), it is represented as <Ê>.

Another issue lies in whether diacritics, particularly macrons, are used on certain long vowels. Sokolova uses a macron with the letters for all seven vowels which belong to the class of long vowels (as discussed in Section 3.1.2.2.1), even those which are not opposed by an identical short-vowel character. That is, she chooses the graphemes <ē>, <ĕ>, <ō>, and <û> (with macrons), despite the fact that, under her system, there are no counterpart letters <e>, <ε>, <o>, or <ũ> with which they contrast. In the subsequent chapters of this thesis, I do not use macrons over long vowels which do not have a short vowel represented by the same IPA character. That is, I use macrons with the graphemes <ā>, <ī>, and <ū>, which contrast with <a>, <i>, and <u>, respectively. However, I do not use macrons on <e>, <ê>, <o>, or <ũ>, as these graphemes are not opposed by identical short-vowel characters.

For a summary of the Latin letters typically used to represent Shughni, see Tables 3.21 and 3.22. These tables show the orthography used in this thesis and are generally representative of Latin characters used in other works.

## 3.4.2 Cyrillic script

A Cyrillic alphabet was developed for Shughni in the 1930s and subsequently used for several years in the schools of Shughni-speaking regions (Edelman & Dodykhudoeva 2009b: 788). Today, however, all primary and secondary school instruction is officially in Tajik.

Although some Soviet-era scholars chose to use a Latin script for the language in their works, others used a Cyrillic script, perhaps most notably Karamshoev in his three-volume Shughni-Russian dictionary (Karamshoev 1988c; 1988d; 1988b), and Bakhtibekov (1979) in his concise grammar of the language. More recently, Alamshoev (2015)

and Karamshoev et al. (2019) have published pedagogical grammars of Shughni, both of which use a Cyrillic script for the language. The Cyrillic alphabet given in Table 3.16 is the same one found in all of these works.

**Table 3.16:** Cyrillic characters used for Shughni.

IPA	Grapheme	IPA	Grapheme	IPA	Grapheme
a	Аа	i	Ии	u:	Ўў
a:	Ăă	i:	Йй	ø:	Ӣӣ
b	Бб	j	Йй	f	Фф
v	Вв	k	Кк	χ	Хх
w	Ww	l	Лл	x	Ӧӧ
g	Гг	m	Мм	ts	Цц
d	Дд	n	Нн	dz	Зз
ð	Δð	o:	Оо	tʃ	Чч
θ	Θθ	p	Пп	ʃ	Шш
e:	Ее	r	Рр	ʁ	Ғғ
ē	Êê	s	Сс	ʏ	Ӣӣ
ʒ	Жж	t	Тт	q	Ққ
z	Зз	u	Уу	dʒ	Ҷҷ

### 3.4.2.1 Notes on Cyrillic consonant letters

Many Cyrillic consonant letters used to represent Shughni sounds correspond to the exact same phoneme in Russian. Thus, for instance the grapheme <Цц> represents the phoneme /ts/ in both languages. However, the Shughni phoneme inventory does not align perfectly with that of Russian.

Three Shughni consonant phonemes not shared with Russian are found in Tajik, and these languages, in turn, use the same graphemes for these three consonants. These include <Ғғ> for /ʁ/; <Ққ> for /q/; and <Ҷҷ> for /dʒ/. All three of these letters represent letters of the Russian alphabet modified with diacritics. The first letter is the Russian letter <Гг> modified with a stroke, and the second and third are the Russian letters <Чч> and <Кк>, each modified with a descender.

Note that the letters <Ӣӣ> and <Ӧӧ>, like their counterparts in the Latin script given above, somewhat non-intuitively correspond to /ʏ/ and /x/, respectively, while <Ғғ> and <Хх> correspond to /ʁ/ and /χ/, respectively.

One reason for this mismatch is likely that the latter two sounds exist in Tajik, in which they are represented by the same letters, while the sounds /ɣ/ and /x/ do not exist in Tajik.

The letters <Δð> and <Θθ> are generally used to represent /ð/ and /θ/, respectively, in both Latin and Cyrillic scripts for Shughni. However, Edelman & Dodykhudoeva (2009b: 778) give <Џ> and <В> – i.e. Cyrillic letters modified with a caron. The use of these Cyrillic letters to represent the phonemes /θ/ and /ð/, rather than the Greek (and IPA) letters used in most scholarly works, helps to maintain homogeneity with the Cyrillic script used to represent Russian and Tajik, and it was likely for this reason that they were chosen in the development of the early Shughni alphabet.

### 3.4.2.2 Notes on Cyrillic vowel letters

The vowel letters used in Cyrillic scripts are virtually identical across these works. However, there is a key difference regarding the representation of certain vowels preceded by the palatal glide /j/ in Tajik and in Shughni. Where as Tajik uses Russian “soft” vowel letters to represent some sequences of vowels preceded by a /j/, the Shughni alphabets given here use in all cases <Йй> /j/ followed by the vowel letter. Thus, in Tajik, the sequence /ja/ is written as <я>; /jo/ as <ё>; /ju/ as <ю>, while these same sequences in Shughni are written respectively as <йа>; <йо>; and <йу>. For instance, the sequence /jaχ/ (‘ice’ in Tajiki; ‘sister’ in Shughni) is written as <яx> in Tajik and as <йax> in Shughni.<sup>6</sup>

### 3.4.3 Arabic script

Arabic-based scripts have been developed to represent the varieties of Shughni spoken in Afghanistan (cf. Mueller 2014: 5-6 on the teaching of Shughni in primary schools of the Badakhshan Province). In general, the Arabic-based script for Shughni follows the conventions of the alphabet of Farsi in Iran and the alphabet of Dari in Afghanistan. Complications arise, in part, due to the comparatively larger consonant inventory of Shughni, but perhaps the greatest challenge is to devise a way of representing the ten vowels of Shughni, where Farsi, Dari, and Modern Standard Arabic have comparatively fewer. In this subsection, the representation of Shughni consonants and vowels in the Arabic script are examined separately: consonants in Section 3.4.3.1 and vowels in Section 3.4.3.2. The

---

<sup>6</sup>The word for ‘ice’ in Shughni contains long /a:/: <йāx> /ja:χ/.

information here comes primarily from the the work of two scholars, Haidari (2004) and Dost-Mohammad et al. (2011).

### 3.4.3.1 Shughni consonant phonemes in Arabic-based orthography

Table 3.17 shows the letters used to represent native Shughni consonants in Arabic-based scripts. This is not an exhaustive list of consonant letters used in Shughni, however. Another group of consonant letters correspond to Arabic sounds which do not exist in Shughni; these are examined in Section 3.4.3.1.2.

**Table 3.17:** Shughni consonant letters in the Arabic script.

IPA	Grapheme	IPA	Grapheme	IPA	Grapheme
b	ب	ð	ڈ	q	ق
p	پ	r	ر	k	ک
t	ت	z	ز	g	گ
θ	ٹ	ʒ	ژ	l	ل
dʒ	ج	ɣ	ړ	m	م
tʃ	چ	s	س	n	ن
χ	خ	ʃ	ش	w	و
ts	ث	x	ځ	v	ف
dʒ	خ	ʁ	غ	j	ي
d	د	f	ف		

There are six Shughni consonants which do not exist in Arabic. For four of these consonants, their corresponding letters are taken from the Farsi alphabet: /ʒ/ <ژ>; /tʃ/ <چ>; /p/ <پ>; and /g/ <گ>. Letters for two consonant phonemes which do not exist in Farsi or Arabic are taken from the Pashto alphabet: /x/ <ځ> and /ɣ/ <ړ>. (See David 2014: 25-29 on Pashto orthography.)

**3.4.3.1.1 Phonemes /θ/ and /ð/ in Arabic-based scripts.** Two Shughni consonants, the interdental fricative pair /θ/ and /ð/, exist in Arabic but not in Farsi, Dari, Tajik, or Pashto. These sounds present a challenge for developers of Shughni Arabic-based scripts because the letters used to represent /θ/ and /ð/ in Arabic – namely <ٹ> and <ڈ> – appear in Arabic loanwords in Dari and Pashto but are pronounced as /s/ and /z/, respectively. And because Shughni has acquired Arabic borrowings primarily via Dari and Tajik, it shares the same pronunciation

of these letters as /s/ and /z/, even though their original Arabic sounds, /θ/ and /ð/, are Shughni phonemes. For instance, the Arabic word /mi:ra:θ/ <ميراث> ‘heritage’ has been borrowed into Persian as /mi:ra:s/, and from there into Shughni (cf. Sh. /mi:ro:s/). In Shughni, this word retains this modified pronunciation with /s/ despite the fact that Shughni phonology is capable of accommodating the original Arabic segment /θ/.

Thus, the dilemma for developers of Shughni Arabic-based alphabets becomes whether to represent native Shughni /θ/ and /ð/ with the letters to which they correspond in Arabic, or to use new letters. This is a point on which Haidari (2004) and Dost-Mohammad et al. (2011) disagree. The former chooses to use the original Arabic letters <ث> and <ذ> to represent the Shughni phonemes /θ/ and /ð/, respectively. He argues that the Arabic-based Shughni script should limit the use of graphemes which are unfamiliar to the Shughni-speaking population, effectively accepting the ambiguity brought on by this choice – that is, that the letter <ث> represents both /s/ and /θ/, and the letter <ذ> represents both /z/ and /ð/.

Dost-Mohammad et al. (2011), on the other hand, advocate for the use of separate Arabic letters to resolve this situation, arguing that this ambiguity adds unnecessary stress on the reader, who must decide on a word-by-word basis whether these letters belong to an Arabic loanword or a native Shughni word. For this reason, these scholars propose the use of the letter <ث> to represent /θ/ and <ذ> to represent /ð/. These letters maintain a connection with their Arabic counterparts by sharing the same shape, but allow for each Shughni phoneme to be represented by a different letter.<sup>7</sup>

**3.4.3.1.2 Representing Arabic consonants which do not exist in Shughni.** Another issue in Arabic-based Shughni scripts is one which is also found in scripts for the languages Dari, Farsi, and Pashto (among others), namely that many borrowings from Arabic contain consonant sounds which do not exist in these languages. For instance, in Arabic, the letters <ص> and <ض> represent the pharyngeal voiceless alveolar fricative [sʕ] and the pharyngeal voiced alveolar stop [dʕ], respectively, but neither of these sounds exist in Persian or in Shughni.

In Persian and Shughni, these phonemes are realized in borrowed words as /s/ and /z/, respectively. Thus, the

---

<sup>7</sup>In the Cyrillic alphabet developed for Tajik, the same letters are used for /s/ and /z/ regardless of whether they originated as a pharyngeal consonant in an Arabic loanword. Thus, for instance, the /s/ in *cyxōar* (*suhbat* – ‘speech; talk’) and *cā* (*sī* ‘thirty’) are both written with the same letter <c>, despite the fact that the first comes from the emphatic /s/ in Arabic *صحبَة* – *ṣuḥba* ‘friendship’, while the second word is a native Tajik word.



word < صبر > ‘calm’, phonologically /sʰabr/ in Arabic, is rendered /sabr/ in these languages, and the word < ضعیف > ‘weak’, phonologically /dʰaʕi:f/ in Arabic, is rendered /za(j)i:f/. Nonetheless, both the Farsi and Dari varieties of Persian write such words as in the original Arabic, and Arabic-based Shughni scripts tend to follow the same convention.

The pharyngeal alveolar stops are not the only consonants which are realized differently in Persian and Shughni, however. There are two graphemes which correspond to the Shughni phoneme /t/, three which correspond to /s/, and four which correspond to /z/. Moreover, the Arabic graphemes corresponding to the pharyngeal fricative pair /ħ/ and /ʕ/, realized /h/ in Persian, are not pronounced in Shughni. The voiceless glottal fricative /h/ – also found in Persian, where it is represented with < ه > – is generally written as such in Farsi and Dari loanwords containing this phoneme, though it not pronounced in Shughni. Thus, Persian loanwords such as < همکار > – /hamkār/ > ‘co-worker’ and < هر > – /har/ ‘each’ are written the same as in Persian but pronounced /amkor/ and /ar/, respectively.

Table 3.18 shows the groups of graphemes which represent a single phoneme (or none at all) in Shughni. The origin of example Shughni words is Arabic, unless otherwise noted using parentheses.

**Table 3.18:** Shughni consonants represented by multiple Arabic letters.

Letter	Arabic phoneme (IPA)	Letter name	Shughni phoneme	Example usage in Shughni	
ع	ʕ	'ayn	∅	عدالت	/ado:lat/ ‘justice’
ه	h	hā	∅	هنر	/unar/ ‘talent’ (Persian)
ح	ħ	ḥā	∅	حکومت	/uku:mat/ ‘government’
ت	t	tā	t	تیر	/ti:r/ ‘top’ (Shughni)
ط	tʰ	ṭā	t	طاقت	/to:qat/ ‘patience’
س	s	sīn	s	سیزد	/si:zd/ ‘silverberry’ (Shughni)
ص	sʰ	ṣād	s	صبر	/sabri/ ‘calm’
ث	θ	thā	s	میراث	/miro:s/ ‘heritage’
ز	z	zāy	z	زیرد	/zi:rd/ ‘yellow’ (Shughni)
ض	dʰ	ḍād	z	ضعیف	/zayi:f/ ‘weak’
ظ	ðʰ	ẓā	z	ظالم	/zo:lim/ ‘tyrant’
ذ	ð	dhāl	z	ذات	/zo:t/ ‘kind; sort’

### 3.4.3.2 Shughni vowels in Arabic-based orthography

While Arabic consonant letters are rather easily adapted for Shughni, a system for representing Shughni vowels presents more of a challenge, most notably because the Arabic script represents only the long vowels of Modern Standard Arabic with a full grapheme, and their short counterparts are optionally represented with diacritics. This system is largely preserved in the writing systems of Farsi and Dari, which have comparatively fewer vowel phonemes than Shughni.

With this in mind, it appears that in devising an Arabic-based writing system for Shughni, there is a delicate balance to strike between maintaining the orthographic conventions of Farsi and Dari, with which most native Shughni speakers in Afghanistan are familiar, while simultaneously representing the Shughni vowels a way which minimizes ambiguity. This is no easy task, as we will see below.

This subsection first presents the representation of Shughni short vowels in the Arabic script (Section 3.4.3.2.1), followed by the representation of Shughni long vowels in the Arabic script (Section 3.4.3.2.2).

**3.4.3.2.1 Shughni short vowels in the Arabic script.** The three short vowels, namely /a/, /i/, and /u/, coincide with those found in both Arabic and Dari, and thus can be written with the same diacritics used to mark short vowels in these languages. Short /a/ is marked with a *fathah* (a diagonal diacritic above the preceding consonant letter) as in the Shughni words <كَش>, /kaʃ/ ‘hot’, and <غَل>, /ʁal/ ‘still; yet’. Short /i/ is marked with a *kasrah* (a diagonal diacritic below the preceding consonant letter), as in the Shughni words <سِت>, /sit/ ‘dirt’, and <مِس>, /mis/ ‘also’. Short /u/ is written with a *damma* (a rounded diacritic above the preceding consonant), as in the words <سُت>, /sut/ ‘went (m.)’, and <كُد>, /kud/ ‘dog (m.)’.

In word-initial position, an *alif* <ا> is used to represent a short vowel, with the appropriate diacritic mark to indicate its quality. In word-medial position, a short vowel has no corresponding grapheme, and its only indication is the diacritic. In word-final position, however, Shughni Arabic-based orthography follows the conventions of Farsi and Dari in writing short vowels with a full grapheme. The letter *wāw* <و> is used for word-final /u/, as in the word <تو> *tu* ‘you (sg.)’; the letter *hā* <ه> is used to represent word-final short /a/, as in the factual particle <ته> /ta/; and the letter *yeh* <ي> is used to represent word-final /i/, as in <تي> /ti/ ‘go; leave’. In this sense, there is potential ambiguity between the short vowels and their long vowel counterparts in word-final position.

However, in reality, short /a/ only occurs word-finally in a handful of cases, such as /aga/ ‘if’ and /sa/ ‘go!’; long /a:/ is much more common in word-final position. On the other hand, short /i/ is much more common in word-final position than its long counterpart /i:/, which only occurs in a small number of borrowings, such as /si:/ ‘thirty’ (< Tajik). Neither short /u/ nor long /u:/ are commonly found in word-final position, and despite the existence of the minimal pair /ku/ ‘(emotive particle)’ and /ku:/ ‘mountain’, little confusion should occur between these two phonemes in word-final position. Table 3.19 gives examples of each short vowel word-initially, word-medially, and word-finally:

**Table 3.19:** Shughni short vowels in Arabic orthography.

	DIACRITIC	FINAL	MEDIAL	INITIAL
a	<i>fathah</i> َ	سه <i>sa</i> ‘go!’	كش <i>kaš</i> ‘hot’	آسید <i>asīd</i> ‘this year’
i	<i>kasrah</i> ِ	گرمی <i>garmi</i> ‘warmth’	مِس <i>mis</i> ‘also’	اسلاح <i>islo</i> ‘correction’
u	<i>dammah</i> ُ	خو <i>xu</i> ‘ate’	کُد <i>kud</i> ‘dog’	أُتاق <i>utoq</i> ‘room’

**3.4.3.2.2 Shughni long vowels in the Arabic script.** The question of how to represent long vowels in Shughni is more complex than that of how to represent short vowels. Fundamentally, this is because the Arabic script, as originally designed for the Arabic language, is only equipped to represent three long vowels – /a:/, /i:/, and /u:/. This leaves four remaining Shughni long vowel phonemes without counterpart graphemes in Arabic or Persian. Below, I present the method(s) set forth by Haidari (2004) and Dost-Mohammad et al. (2011) to represent each of the Shughni long vowels.

**The vowel /i:/** coincides in quality with the same long vowel in Arabic, Farsi, and Dari, and therefore the same letter <ي> is used in Shughni (e.g. <بیر> – /bi:r/ ‘bed’ and <تید> – /ti:d/ ‘go.INF’).

**The Shughni vowels /e:/ and /ɛ:/** present several potential issues for the Arabic-based Shughni script. First, the long vowels /i:/ and /e:/, though distinct phonemes in many varieties of Persian, are represented with a single grapheme <ي> (cf. Dari *šīr* ‘milk’ and *šer* ‘lion’, each written <شیر>). And second, the phonemic distinction between /e:/ and /ɛ:/ is not found in Persian and is rather unique even among Pamir languages.

To partially resolve these issues, Haidari opts to use the Pashto letter <ې> (used for /e/ in Pashto) to represent both /e:/ and /ɛ:/ in Shughni, thus eliminating confusion with /i:/. In effect, a similar amount of ambiguity arises

in Shughni between words with /e:/ and /ɛ:/ as in Dari between words with /i:/ and /e:/. To illustrate, consider the Shughni minimal triplets /χi:r/ ‘sun’ vs. /χe:r/ ‘cessation’ vs. /χɛ:r/ ‘alright’ and /ði:d/ ‘hits’ vs. /ðe:d/ ‘falls’ vs. /ðɛ:d/ ‘war’. Under Haidari’s system, one word from each triplet is written distinctly from the other two. Specifically, /χi:r/ ‘sun’ is written <خیر>, and /ði:d/ ‘hits’ is written <ځید>, respectively. Both /χe:r/ ‘cessation’ and /χɛ:r/ ‘alright’ are written <خېر>, and both /ðe:d/ ‘falls’ and /ðɛ:d/ ‘war’ are written <ځېد>.

**The vowel /o:/**, which corresponds etymologically to Dari and Farsi /a:/ (Tj. /o:/), is written with the Arabic letter *alif* <ا> as it is in these languages. That is, it is written as the plain letter word-medially and word-finally, but with a *maddah* (a tilde-shaped diacritic) <ā> in its initial form. This letter is found, for instance, in the words <پروا> – /parwo:/ ‘care; concern’; <مال> – /mo:l/ ‘livestock’; and <آت> – /o:t/ ‘visible’.

**Shughni /u:/**, like /i:/, coincides in quality with the same vowel in both Arabic and Persian, and it is therefore written in the same way as in these languages, namely with the letter *wāw* <و>. Note that the same letter is also used to write the phoneme /w/, a redundancy which is also found in Arabic and Persian. In Shughni, the vowel /u:/ is generally accompanied by a preceding alveolar glide /j/ in word-initial position, yielding the sequence <وو>, as in the word <ووفد> – /wu:vd/ ‘seven’. It is found most often word-medially, for instance, in the words <مون> – /mu:n/ ‘apple’; and <چود> – /tʃu:d/ ‘did’. It only rarely occurs word-finally, as in <کو> – /ku:/ ‘mountain’.

**The vowel /ø:/** is not found in Arabic, Pashto, or any of the Persian varieties using the Arabic script, and therefore a different way of writing this vowel must be devised for Shughni. Here, the systems of Haidari and Dost-Mohammad et al. differ, but only slightly. Because of its association with /u/ as a rounded vowel, both have proposed that this vowel be written with a modified version of the letter *wāw* <و>. Haidari, for his part, chooses to use a ring on top of *wāw*, yielding <ؤ>. Dost-Mohammad et al., on the other hand, propose the use of the diacritic *alif kharjīyah*, ‘dagger alif’ or ‘superscript alif’, thus giving the letter <ؤ>. This vowel is found, for instance, in the words <دوند> – /dø:nd/ ‘so much; this much’ and <اون> – /ø:n/ ‘yes’.

**Long /a:/** in Shughni is ideally represented differently than long /a:/ in Persian because, as mentioned above, Persian long /a:/ corresponds etymologically to Shughni /o:/. Persian /a:/ and Shughni /o:/ are written via a plain *alif* <ا> in medial and final positions, and via an *alif* with the *maddah* diacritic <ā> in initial position. For long /a:/ in Shughni, Haidari opts to use the *alif* with the *maddah* diacritic <ā> in word-initial and word-medial positions, and the letter <ا> in word-final position, where it is also used for short /a/. Despite potential issues in distinguishing /a:/ from /o:/ in initial position and /a:/ from /a/ in final position, Haidari’s strategy generally avoids ambiguity.

First, Shughni /a:/ rarely occurs word-initially in native words, where ambiguity could occur with /o:/. Where /a:/ occurs word-initially in borrowings, it is generally the result of compensatory lengthening due to a neighboring phoneme not pronounced in Shughni – often Arabic or Persian /h/, or Arabic /ʕ/. Because Arabic and Persian borrowings are spelled in their original Arabic forms, there is no <Ā> word-initially to represent Shughni /a:/. For instance, the word *āwol* ‘state; condition’, a loanword of Arabic origin, is spelled in its original Arabic form as <أحوال> (rather than <آوال>, which could be read /a:wol/ or /o:wol/).

Long /a:/ is common word-medially in native Shughni words such as /ka:ts/ and /kulla:/, which are written in Haidari’s system as <كاٲ> and <كلا>, respectively. Here, long /a:/ is represented distinctly from both short /a/ and long /o:/. Word-medial short /a/ is not written with a full letter, but rather via a diacritic, and word-medial long /o:/ is an *alif* without the *maddah* diacritic.

Word-finally, long /a:/ is likewise distinguished from /o:/ via the *maddah* over the *alif*. Short /a/, for its part, only occurs word-finally in a handful of function words such as /aga/ ‘if’ and in shortened verb stems such as /sa/ ‘go!’. Short /a/ is not found word-finally in any nouns or adjectives (cf. Sokolova 1953). Ambiguity is thus generally avoided using Haidari’s system.

A summary of Shughni long vowels in the Arabic script is given in Table 3.20. Graphemes in parentheses are used indicate the hypothetical form for a vowel in cases where there are no words for which the vowel is written in the position in question. For instance, an initial /ê/ would be written with <آٲ>, but this vowel does not occur word-initially, so this grapheme appears in parentheses.

**Table 3.20:** Shughni long vowels in Arabic orthography.

Word-Initial					
VOWEL	FORM		EXAMPLE	LATIN	GLOSS
a:	( آ )		—	—	—
e:	اڤ		اڤچه	<i>ečā</i>	floor
ɛ:	( اڤ )		—	—	—
i:	( اي )		—	—	—
o:	آ		آت	<i>ot</i>	visible
u:	( او )		—	—	—
ø:	اۈ		اۈن	<i>ũn</i>	yes

Word-Medial					
VOWEL	FORM		EXAMPLE	LATIN	GLOSS
	NON-CONN.	CONN.			
a:	آ	ل	غاڭ	<i>yāc</i>	girl
e:	ڤ	ڤ	خبر	<i>xer</i>	cessation
ɛ:				<i>xêr</i>	OK; alright
i:	ي	ي	چيد	<i>čīd</i>	house
o:	ا	ا	مال	<i>mol</i>	livestock
u:	و	و	مون	<i>mūn</i>	apple
ø:	ۈ	ۈ	دۈند	<i>dūnd</i>	so much

Word-Final					
VOWEL	FORM		EXAMPLE	LATIN	GLOSS
	NON-CONN.	CONN.			
a:	ه	ه	گنده	<i>gandā</i>	bad
e:	ې	ې	بېمېنې	<i>ǰumne</i>	tomorrow
ɛ:	( ې )	( ې )	—	—	—
i:	ي	ي	سي	<i>sī</i>	thirty
o:	ا	ا	پروا	<i>parwo</i>	care
u:	( و )	( و )	—	—	—
ø:	ۈ	ۈ	—	—	—

### 3.4.4 Orthography used in this thesis

In the remainder of this thesis, I use a Latin-based orthography which incorporates elements of early Latin-based orthographies, such as that of Sokolova (1953; 1967; 1973). The writing system used here is meant to be as efficient as possible while remaining easily readable for academics who are either more or less familiar with Shughni, and for Shughni speakers who are not familiar with technical linguistics topics or the International Phonetic Alphabet.

Among the choices of characters made for this writing system, a few are worth mentioning. I follow the authors of early Shughni publications in using the graphemes <ŷ> and <Ÿ> to represent the phonemes /ɸ/ and /ʏ/, respectively, and the graphemes <x> and <ẋ> to represent the phonemes /χ/ and /x/, respectively. I also follow virtually all authors in using a caron over several letters to avoid the use of digraphs and to avoid certain letters particular to the International Phonetic Alphabet. Thus, <č> represents /tʃ/; <š> represents /ʃ/; <ž> represents /ʒ/. The phoneme /dʒ/ is represented as <ǰ>, a letter which is used by most Soviet scholars. I use it here to avoid any potential confusion with the IPA symbol /j/, which I represent as <y>.

Note further that I opt to use the digraph <dz> to represent the phoneme /d͡z/. I avoid using the characters <Zz> because of the lowercase letter's close resemblance to the IPA character <ʒ>, which is also a phoneme in Shughni. Moreover, I choose to represent this phoneme without a top-tie bar, as no top-tie bar is used for any other phonemes and, in my opinion, this would add unnecessary complexity to the writing system. I am not aware of any word or sequence of graphemes in which this digraph would cause confusion between the single phoneme /d͡z/ and the sequence of individual phonemes /d/ and /z/.

I use a macron over only those long vowels whose grapheme also has a short counterpart. Therefore, <ā> is used for /a:/; <ī> is used for /i:/; and <ū> is used for /u:/. All three short vowels are represented with the plain version of these graphemes. The long vowels /e:/, /o:/, and /ø:/ are all written without a macron, as <e>, <o>, and <û>, respectively. I follow most authors in representing the long vowel /ɛ:/ with a circumflex: <ê>.

The writing system used in this thesis is given in Tables 3.22 (vowels) and 3.21 (consonants). These tables include examples of each grapheme in use.

**Table 3.21:** Orthography used in this thesis: Consonants.

CONSONANT LETTERS					
IPA Value	Grapheme	Examples			
b	b	<i>bīr</i>	‘bed’	<i>ǰāb</i>	‘night’
ts	c	<i>cem</i>	‘eye’	<i>ȳac</i>	‘girl’
tʃ	č	<i>čīd</i>	‘house’	<i>wīdīč</i>	‘bird’
d	d	<i>divi</i>	‘door’	<i>bād</i>	‘after(ward)’
ð	ð	<i>ðīs</i>	‘ten’	<i>tūð</i>	‘mulberry’
f	f	<i>fānd</i>	‘lie’	<i>šāf</i>	‘saliva’
g	g	<i>gul</i>	‘flower’	<i>wārg</i>	‘lamb’
ɣ	ȳ	<i>ȳac</i>	‘girl’	<i>boȳ</i>	‘garden’
ʝ	ŷ	<i>ŷīnik</i>	‘woman’	<i>maŷ</i>	‘sheep’
ɖʒ	ǰ	<i>ǰāld</i>	‘fast’	<i>lunǰ</i>	‘cheek’
k	k	<i>kud</i>	‘dog’	<i>pérnak</i>	‘heel’
l	l	<i>lap</i>	‘very’	<i>kāl</i>	‘head’
m	m	<i>mūn</i>	‘apple’	<i>sām</i>	‘I go’
n	n	<i>nēdz</i>	‘nose’	<i>qapqūn</i>	‘trap’
p	p	<i>pīš</i>	‘cat’	<i>tuxp</i>	‘sour’
q	q	<i>qīč</i>	‘stomach’	<i>tsiluq</i>	‘vertical’
r	r	<i>rūšt</i>	‘red’	<i>qār</i>	‘wrath’
s	s	<i>sīvd</i>	‘shoulder’	<i>mis</i>	‘also’
ʃ	š	<i>šič</i>	‘now’	<i>māš</i>	‘we’
t	t	<i>tēr</i>	‘black’	<i>kut</i>	‘short’
θ	θ	<i>θīr</i>	‘ash’	<i>meθ</i>	‘day’
v	v	<i>vo</i>	‘again’	<i>tagov</i>	‘below’
w	w	<i>wūvd</i>	‘seven’	<i>žow</i>	‘cow’
χ	x	<i>xoŷ</i>	‘six’	<i>rux</i>	‘dawn’
x	ǰ	<i>ǰīrbǰ</i>	‘frog’	<i>woǰ</i>	‘grass’
j	y	<i>yīw</i>	‘one’	<i>noy</i>	‘throat’
ʒ	ž	<i>žīr</i>	‘stone’	<i>étaž</i>	‘floor’
z	z	<i>zimc</i>	‘land’	<i>wīz</i>	‘load’
ɖʒ	dz	<i>dzulik</i>	‘small’	<i>wīxidz</i>	‘key’



**Table 3.22:** Orthography used in this thesis: Vowels.

VOWEL LETTERS					
IPA Value	Grapheme	Examples			
a	a	<i>abri</i>	‘cloud’	<i>va</i>	‘along’
i	i	<i>(y)is</i>	‘feeling’	<i>jūri</i>	‘harmony’
u	u	<i>uǰast</i>	‘sigh’	<i>tu</i>	‘you (sg.)’
a:	ā	<i>āwol</i>	‘condition’	<i>qissā</i>	‘story’
e:	e	<i>bel</i>	‘shovel’	<i>ǰumne</i>	‘tomorrow’
ɛ:	ê	<i>ôêd</i>	‘war’	<i>firêǰdz</i>	‘flea’
i:	ī	<i>pīc</i>	‘face’	<i>sī</i>	‘thirty’
o:	o	<i>osil</i>	‘harvest’	<i>sado</i>	‘voice’
u:	ū	<i>ǰūvd</i>	‘milk’	<i>dūs</i>	‘a bit’
ø:	ũ	<i>ũn</i>	‘yes’	<i>sũg</i>	‘legend’

### 3.4.5 Shughni scripts side-by-side

As a means of bringing together different scripts used to write Shughni, I conclude the section on orthography by presenting a short Shughni text written in three different alphabets, which are found in Figure 3.2. The Latin text uses the writing system which is to be used in the remainder of the thesis; the Cyrillic script is that used by Karamshoev in his (1988) dictionary, and the Arabic script follows that of Haidari (2004), with the exception that the phonemes /θ/ and /ð/ are represented with the graphemes <ث> and <ذ>, respectively (as suggested by Dost-Mohammad et al. 2011). The text itself is an excerpt from the novel *Zindagi az naw ca sūd sar* (*When Life Begins Again*), recommended to me by Ms. Shahlo Shomansurova. A translation, also provided by Ms. Shomansurova, is provided below the texts. At this point, one last issue arises, namely whether to write certain phonologically weak morphemes, such as aspectual and emotive particles, as separate words or as attached to the word which precedes them. I follow here the conventions used in the novel and generally write these morphemes together with the word preceding them.

Figure 3.2: Shughni scripts side-by-side

Йида йи сайлийат му қати ар му фикрийен чӯд. Вузум хойх чӯд жйwчгай қуват  
турд дивийtow. Ту лāk ар ху дарӯн wi вири ху йис wi ки ху фām хурд йид  
Худойчӯник Рухи ар ту дарӯн. Йисат wi ди чӯд, му гāпента мис тайор сен. Дӯнчат  
тута йиси му гāпенен дисаθ yuborak wi тирйн виц.

بيده يي سَيلَيت موقتې آر مو فِکريېن چود. وُزُم خايِش چود ژيوجگېش قُوت تُرد دِښتاو. تو لاک آر خو درون وي ږي  
خو پِس وي کي خو فām خُرد پد ځخايجوئیک روخي آر تو درون. پست وي دي چود، مو گپېنته مِس تيار سېن. دُونجَات  
توته پسي مو گپېنېن دِسْت غبارک وي تيرين څخ.

Yida yi sayliyat mu qati ar mu fikriyen čūd. Wuzum xoiḥ čūd žīwǰgaḥ quwat turd  
diviḫtow. Tu lāk ar xu darūn wi viri xu yis wi ki xu fām xurd yid Xuḍoyjūnik Ruxi ar tu  
darūn. Yisat wi di čūd, mu gāpenta mis tayor sen. Dūnǰat tuta yisi mu gāpenen disaθ  
yuborak wi tirin vic.

*Here you had a walk with me inside my thoughts. I wanted to show you the power of love, that you could find it  
and feel it and then realize that it's the Light of God inside you. When you feel it, this will also be the end of my  
words, because you will realize that all I say about love is just dust in the shadow of what it really is.*

## Chapter 4

# Grammatical Overview

This chapter is designed to consolidate information on the most basic aspects of Shughni morphosyntax into a single part of the thesis. In doing so, it is meant to serve two further purposes. First, the reader wishing only to find basic information on Shughni can do so without having to sift through the more dense chapters which follow. And second, the reader who intends to go beyond this chapter will be equipped with a foundation in the most important building blocks of Shughni grammar. With this in mind, the material in this chapter is covered rather concisely, and in each section the reader is pointed to the areas of the thesis which examine these topics in more detail.

The aspects of Shughni grammar covered in this chapter are the following. The first section of the chapter gives a **typological overview**, outlining key grammatical characteristics of the language and indicating how they fit into the larger areal-linguistic picture. Section 4.2 then looks at **nominals**, and Section 4.3 introduces **adpositions and case-like suffixes**. Next, Section 4.4 provides an overview of the **verbal system**, including the notion of verb stems, as well as basic verbal inflection and formation and use of passive-like constructions and causatives. Finally, Section 4.5 presents a few aspects of Shughni syntax, namely **complex clauses and question formation**.

## 4.1 Typological overview

Shughni shares many typological features not only with genetically related languages such as Persian and Pashto, but also with genetically unrelated languages of the Central Asian Sprachbund, notably the Turkic languages. Most fundamentally, Shughni shares with these languages basic SOV word order. Shughni is pro-drop, and clause-level word order is flexible. Deviations from SOV are common and are generally used to indicate nuances in information structural notions such as topic and focus. Specifically, topicalized and focused elements – especially when contrastive – tend to appear first in the clause regardless of their syntactic relations. When objects or verbs are fronted, particularly when an overt subject is present, the result is non-canonical word order. Information structure and clause-level word order are discussed in more detail in Section 12.2.

Shughni displays a tendency toward right-headedness, but it is not right-headed in every aspect of its grammar. It possesses a mix of prepositions and postpositions, as well as at least one circumfix-like element (*to-X-yec* ‘until X’). It also has a mix of complementizers and subordinators that appear at both the left and right edge of the clause. In this regard, Shughni is like Pashto and many varieties of (vernacular) Tajik Persian, which similarly possess both left-headed and right-headed areas of their grammar.

Shughni uses both a binary case system (direct and oblique) and a binary system of grammatical gender (feminine and masculine), though neither case nor gender show up in all areas of nominal inflection. A formal distinction in grammatical case only shows up in the first-person singular pronoun and a subset of demonstrative pronouns. Direct case is reserved for subjects, and oblique case is used with non-subjects, including direct and indirect objects, objects of adpositions, and possessors.

Grammatical gender, like case, is explicitly marked only on a subset of nouns and agreeing elements. Although all nouns are inherently specified for grammatical gender, formal marking of gender only appears on a subset of demonstrative pronouns (Section 6.2.1), a subset of adjectives and participle-like unaccusative verbs (Section 9.2), and a handful of nouns which show gender distinctions based on natural sex. Both the assignment and morphosyntactic expression of gender in Shughni are examined in detail in Chapter 7.

In its verbal system, Shughni is similar to other Indo-Iranian languages of Central Asia in making widespread use of complex verbs, which consist of an inflecting light verb and a non-verbal element which holds the verb’s

core semantic content. The non-verbal component of a complex verb may be either indigenous or borrowed, but many non-verbal elements have been borrowed from Persian, Arabic, and more recently Russian. At the same time, the language also uses a number of simplex verbs, the majority of which are Indo-European in origin. Most simplex verbs are monomorphemic and do not undergo (productive) derivational operations through affixation or stem-internal processes, with the exception of a few suffixal morphological causatives. Passive-like constructions, for their part, are done periphrastically. The various aspects of the language's verbal system are explored in the four chapters of Part III of the thesis.

Shughni, like a number of other Iranian languages, exhibits elements of both accusative and ergative alignment, although in Shughni the ergative pattern is restricted to a very small corner of the grammar. The language is fully nominative-accusative throughout its case-marking system and nearly fully nominative-accusative throughout its verbal agreement system. The only exception is found in its past-tense agreement system, where in the third-person singular, unaccusative subjects pattern with objects in taking no agreement clitic, while unergative and transitive subjects trigger an overt agreement clitic. This phenomenon has been called *vestigial ergativity* (e.g. Stump & Hippisley 2011) and is generally agreed to be a relic of a formerly robust system of split ergativity which has eroded at different rates in different languages. Details on verbal inflection in the past and present tenses can be found later in the present chapter (in Section 4.4.3), while morphosyntactic alignment and the phenomenon of vestigial ergativity are the topic of Section 12.4.

Finally, although most of Shughni's typological characteristics are shared by Pashto, Persian, and other widely spoken Iranian languages, a few surprising characteristics – often shared with Indo-Aryan, Dravidian, Burushaski, and other languages of South Asia – suggest the influence of a non-Indo-European substrate, perhaps a relative of Burushaski or other languages spoken in the region before the arrival of Indo-Europeans. These characteristics include, for instance, the use of correlative clauses (as in Hindi) in addition to externally headed relative clauses (as in Persian) (see Section 12.1.1.2); a reordering of the elements of certain adpositional phrases involving inalienable possession (*my in hand* rather than *in my hand*) (see Section 4.3.2); the order of elements in numerals eleven to nineteen such that the numeral 'ten' precedes the remainder (e.g. 'ten and five' = 'fifteen'), rather than vice versa, as in Persian (see Section 5.4.1); and a system of gender assignment which is based partly on notions of concrete and abstract (see Section 7.4). The latter phenomenon, in particular, has been linked to a similar semantically based system of noun classification in Burushaski (e.g. Edelman 1980b; Edelman & Dodykhudoeva 2009b). The typological features mentioned here are fleshed out in a bit more detail in the subsections that follow, and in even

greater detail in the remainder of the thesis following this chapter.

## 4.2 Nominals

This section looks at core issues in the form and formation of nominals. It is divided into short subsections on **pronouns** (4.2.1), **nominal inflection** (4.2.2), **nominal derivation** (4.2.3), and the **order of elements in noun phrases** (Section 4.2.4).

The discussion here provides an introduction to aspects of Shughni grammar which are addressed in more detail in Part II of the thesis, which includes Chapter 5 on nominal inflection, derivation, and (in)definiteness, Chapter 6 on pronouns and demonstratives, and Chapter 7 on the assignment and morphological expression of grammatical gender.

### 4.2.1 Pronouns

This subsection provides a brief presentation and discussion of the various types of pronouns in Shughni: personal pronouns (4.2.1.1); demonstrative pronouns 4.2.1.2; interrogative pronouns (4.2.1.3); indefinite pronouns (4.2.1.4); and emphatic and reflexive pronouns (4.2.1.5).

#### 4.2.1.1 Personal pronouns

Shughni has four personal pronouns: first- and second-person singular and plural. Demonstrative pronouns, to be discussed below, are used in the function of third-person pronouns. Personal pronouns, like demonstrative pronouns, exhibit a direct~oblique case distinction, although a formal distinction only occurs in the first-person singular, where we find *wuz* (DIR) and *mu* (OBL). In the other three cells of personal pronouns, the direct and oblique forms are identical. The paradigms of direct and oblique personal pronouns are shown in Tables 4.1 and 4.2, respectively.

**Table 4.1:** Direct personal pronouns.

	SING	PLUR
1	<b>(w)uz</b>	<i>māš</i>
2	<i>tu</i>	<i>tama</i>

**Table 4.2:** Oblique personal pronouns.

	SING	PLUR
1	<b><i>mu</i></b>	<i>māš</i>
2	<i>tu</i>	<i>tama</i>

Note that throughout the thesis, Shughni personal pronouns are glossed with their English equivalent in the corresponding grammatical context. For example, *wuz*, which is always a subject, is glossed as ‘I’, and *mu* is glossed as ‘me’ when it is an object and ‘my’ when it is a possessive adjective.

#### 4.2.1.2 Demonstrative pronouns

Demonstrative pronouns, displayed in Table 4.3, employ a system of triple deixis inherited from the old Proto-Indo-European system (e.g. Edelman & Dodykhudoeva 2009a: 794). A direct~oblique distinction is made in each of the three degrees – proximal, medial, and distal. Moreover, gender is distinguished in the oblique singular forms for all three degrees, as well as in the direct forms of the distal. Plural forms, for their part, show the direct~oblique distinction but no gender distinction.

**Table 4.3:** Demonstrative pronouns.

	Proximal		Medial		Distal	
	DIR	OBL	DIR	OBL	DIR	OBL
SG.F		<i>mam</i>		<i>dam</i>	<i>yā</i>	<i>wam</i>
SG.M	<i>yam</i>	<i>mi</i>	<i>yid</i>	<i>di</i>	<i>yu</i>	<i>wi</i>
PL	<i>māð</i>	<i>mev</i>	<i>dāð</i>	<i>dev</i>	<i>wāð</i>	<i>wev</i>

Demonstrative pronouns may either modify a noun as part of a noun phrase or, as mentioned above, function as a third-person pronoun. This distinction is exhibited in examples (6) and (7). In the former, the distal demonstrative forms part of a phrase with the head noun *čorik* ‘man’, while in the latter it functions as a stand-alone third-person pronoun.

#### (6) Demonstrative modifying noun

[**Yu**      čorik] azam      yat.  
 DEM.DIR.M man   from.there come.PST  
 ‘**That** man came from there.’

#### (7) Stand-alone demonstrative as subject

[**Yu**] azam      yat.  
 he   from.there come.PST  
 ‘**He** came from there.’

Throughout the thesis, I gloss demonstrative pronouns which modify nouns differently from those which are used as standalone pronouns. Specifically, I gloss the former as DEM and include at least the case and gender, where gender is distinguished. In instances where the deictic degree of the demonstrative is relevant to the discussion, I also include this information (i.e. PROX, MED, or DIST). I gloss demonstratives functioning as pronouns similar to how I gloss personal pronouns, namely by using their corresponding English pronoun in the context at hand.

The reader is directed to Section 6.2 for a more detailed discussion of the usage of demonstrative pronouns and the deictic system.

#### 4.2.1.3 Interrogative pronouns

Shughni has both simplex and complex interrogative pronouns and adjectives. Complex interrogative forms consist of a simplex interrogative pronoun as a base, with the addition of a case-like suffix or a derivational suffix. Common interrogative pronouns are given in Table 4.4. The first and second columns contain simplex interrogative pronouns, and the third contains complex interrogative pronouns.

**Table 4.4:** Fundamental wh-words in Shughni.

FORM	GLOSS	FORM	GLOSS	FORM	GLOSS
<i>čidūm</i>	which	<i>čīz</i>	what (thing)	<i>čīzard</i>	why
<i>cūnd</i>	how much/many	<i>čīr</i>	what (activity)	<i>čīzin</i>	what kind (of)
<i>čay (či)</i>	who(m)	<i>kā</i>	where	<i>cawaxt</i>	when

The structure of the complex interrogative pronouns in the third column is as follows. The simplex interrogative pronoun *čīz* ‘what’ combines with the dative suffix *-(a)rd* to form *čīz-ard* ‘why’. This same simplex pronoun takes the derivational (adjective-forming) suffix *-in* to form *čīz-in* ‘what kind (of)’. And *cawaxt* ‘when’ contains an initial bound interrogative pronoun *ca-*, roughly meaning ‘what/which’, followed by *waxt* ‘time’.

Note the existence of two words meaning ‘what’ – *čīz* and *čīr* – where the former questions an object and the latter questions an activity, as in the common greeting *čīrita?* (lit. *čīr-i=ta* ‘what are you doing’, glossed *what-2SG=FAC*). A more thorough discussion on *wh*-words in Shughni is provided in Section 6.1.2.



#### 4.2.1.4 Indefinite pronouns

Indefinite pronouns in Shughni are typically formed on the basis of the *wh*-words in Table 4.4 and can be divided into four types: (i) **assertive existential** (~ ‘something’); (ii) **elective existential** (~ ‘anything’); (iii) **negative** (~ ‘nothing’); (iv) **universal** (~ ‘everything’).

Indefinite pronouns in Shughni have a prefixed element resembling a quantifier – *ar* ‘each’, *yi* ‘one’, or *fuk* ‘all’. Negative and universal indefinite pronouns optionally take the augmentative suffix *-aθ*. The formation of Shughni indefinite pronouns is exhibited in Table 4.5. Examples are included of each type of pronoun based on the *wh*-word *čīz* ‘what’, which results in the meanings ‘something’, ‘anything’, and so on.

**Table 4.5:** Formation of indefinite pronouns in Shughni.

TYPE	FORMATION	EXAMPLE	GLOSS
Assertive Existential	<i>ar</i> + WH + <i>ca</i>	<i>arčīzca</i>	‘something’
Elective Existential	<i>yi</i> + WH	<i>yičīz</i>	‘anything’
Negative	<i>yi</i> + WH (+ <i>aθ</i> )	<i>yičīz(aθ)</i>	‘nothing’
Universal	<i>fuk</i> + WH (+ <i>aθ</i> )	<i>fukčīz(aθ)</i>	‘everything’

The usage of these different types of indefinite pronouns is based primarily on polarity and mood: assertive existential pronouns are used in affirmative declarative environments, elective existential in imperative and interrogative environments, and negative pronouns are used in negative polarity environments. Universal pronouns may be used in any environment. A more complete discussion of indefinite pronouns, including examples of their usage, is given in Section 6.1.3.

#### 4.2.1.5 Reflexive and emphatic pronouns

Shughni has two pronouns which correspond to English ‘-self’ pronouns: a subject-oriented oblique anaphor *xu* (glossed REFL) and an emphatic pronoun *xubaθ* (glossed PRON.EMPH), which may be in either direct or oblique position. The oblique anaphor *xu* is used in all contexts typical of oblique pronouns, including as both direct and

indirect object (in reflexive-like constructions), object of adposition, and possessor. This pronoun has a single form *xu* and is used only when co-referential with the subject nominal – i.e. that which is in direct case and agrees with the finite verb. Personal (and demonstrative) pronouns are used in oblique position when the entity to which they refer is not co-referential with the subject.

Examples of *xu* are given in (8) and (9). Note that in (8a), which contains a second-singular subject, the reflexive pronoun *xu* – here with a possessive function – cannot be replaced by the personal pronoun *tu*. In (8b), the use of *xu* is only felicitous if it is co-referential with the subject noun *yu čorik* ‘that man’; otherwise, a demonstrative pronoun *wi* is required. The example in (9) shows *xu* as a reflexive object with the verb *zār xu čīdow* ‘to injure oneself’.

(8) **Subject-oriented oblique anaphor *xu*: As possessive adjective**

- a. Tu=t **xu** (//\***tu**) mošīn zinod o?  
 I=1SG REFL (\*your) car wash.PST PQ  
 ‘Did you wash your car?’
- b. Yu čorik<sub>i</sub>=i **xu**<sub>i/\*j</sub> (//\***wi**<sub>i\*/j</sub>) tilivīzor werūn čūd.  
 that.M man=3SG REFL his television broken do.PST  
 ‘That man broke his television.’

(9) **Subject-oriented oblique anaphor *xu*: As direct (reflexive) object**

- Yā=yi carāng zār **xu** čūd?  
 she=3SG how injured REFL do.PST  
 ‘How did she injure herself?’

The emphatic pronoun *xubaθ* is used when the identity of an argument is being emphasized against other possibilities. It often follows another nominal with which it is co-referential. An initial example of *xubaθ* is given in (10). Both *xu* and *xubaθ* are discussed in greater detail in Section 6.1.4.

(10) **Emphatic anaphor *xubaθ***

- Wuz=um **xubaθ** awqot taylor čūd.  
 I=1SG PRON.EMPH food prepared do.PST  
 ‘I myself prepared food.’

## 4.2.2 Nominal inflection

Nouns in Shughni may inflect for both number (addressed in Section 4.2.2.1) and gender (addressed in Section 4.2.2.2). However, there are key differences between each type of inflection. Whereas number inflection is done via suffixation, gender inflection is done via stem-internal vowel alternations. Moreover, while virtually all Shughni nouns may inflect for number, overt morphological inflection for gender only occurs with a handful of adjectives and animate nouns with natural sex.

### 4.2.2.1 Number

Inflection for number is mostly regular. The plural for most nouns is formed by adding the suffix *-(y)en* to the singular, with the epenthetic glide occurring after vowels. However, two irregularities in the formation of plurals are of note. First, certain nouns, often those denoting familial relations, may take a suffix other than *-en* (e.g. *-orĵ* > *dũmod-orĵ* ‘sons-in-law’). And second, the stem vowel of a few nouns changes to *-a-* when the plural suffix *-en* is added (e.g. *kud* ‘dog’ > *kad-en* ‘dogs’). Table 4.6 exhibits examples of these three phenomena: regular pluralization via *-en*, irregular pluralization via the use of a suffix other than *-en*, and the stem-internal vowel change in the plural of some nouns.

**Table 4.6:** Formation of plural nouns in Shughni.

SUFFIX <i>-en</i>			OTHER SUFFIX			STEM-VOWEL CHANGE WITH <i>-en</i>		
SING.	PLUR.	GLOSS	SING.	PLUR.	GLOSS	SING.	PLUR.	GLOSS
<i>tāt</i>	<i>tāt-en</i>	‘father(s)’	<i>dũmod</i>	<i>dũmod-orĵ</i>	‘son(s)-in-law’	<i>kud</i>	<i>kad-en</i>	‘dog(s)’
<i>mošĭn</i>	<i>mošĭn-en</i>	‘car(s)’	<i>pitiš</i>	<i>pitiš-ũn</i>	‘cousin(s)’	<i>puc</i>	<i>pac-en</i>	‘son(s)’
<i>pũd</i>	<i>pũd-en</i>	‘road(s)’	<i>yax</i>	<i>yaxdzinen</i>	‘sister(s)’	<i>čĭd</i>	<i>čad-en</i>	‘house(s)’

See Section 5.2 for a more detailed discussion on pluralization, including the grammatical conditions under which plural suffixes are used. This section also contains information on other types of plural besides the fundamental (additive) plural discussed here, including the associative, similitive, and collective plurals.

### 4.2.2.2 Grammatical gender in nouns and adjectives

In adjectives and nouns, gender is marked not via suffixation, as it often is in other Indo-European languages, but rather by stem-internal vowel alternations, and in a few cases, stem-final consonant alternations. Moreover, gender marking is not realized by a single pair of vowels; instead, a number of vowels participate in the alternations in question. In some cases, the same vowel may mark masculine gender in one word but feminine in another. For example, the vowel *o* is used in the masculine form of the gender-distinguishing noun *vorj* ~ *vêrdz* ‘horse (M~F)’, but in the feminine form of the gender-distinguishing adjective *rũšt* ~ *rošt* ‘red (M~F)’.

The reasons for such overlap, and for the phenomenon of gender marking stem-internally, ultimately boil down to the historical development of the Shughni vowel system and the fate of erstwhile gender-marking suffixes in the language (Sokolova 1967; Karamshoev 1978; Karamshoev 1986). These historical considerations, as well as the intersection between grammatical gender and semantic notions of concreteness vs. abstractness are discussed in more detail in Section 7.2.

Table 4.7 shows just a few of the gender-distinguishing vowel alternations in Shughni.

**Table 4.7:** Example stem vowel alternations in grammatical gender distinction.

<u>Nouns</u>			<u>Adjectives</u>		
VOWELS (M~F/PL)	EXAMPLE	GLOSS	VOWELS (M~F/PL)	EXAMPLE	GLOSS
<i>u~i</i>	<i>puš</i> ~ <i>piš</i>	‘cat’	<i>u~a</i>	<i>kurc</i> ~ <i>karc</i>	‘deep’
<i>u~a</i>	<i>bakul</i> ~ <i>bakal</i>	‘calf (of a cow)’	<i>ũ~o</i>	<i>rũšt</i> ~ <i>rošt</i>	‘red’
<i>o~ê</i>	<i>vorj</i> ~ <i>vêrdz</i>	‘horse’	<i>ĩ~ã</i>	<i>cĩx</i> ~ <i>cãx</i>	‘bitter’

### 4.2.3 Nominal derivation

Nominal derivation in Shughni is done via affixation and compounding. However, by far the majority of nominal derivation in Shughni – including the derivation of both nouns and adjectives – is done via affixation, with a strong preference for suffixation. The language employs a large number of both native and borrowed (Tajik) derivational affixes. To my knowledge, the only derivational prefixes used are borrowed from Tajik. Table 4.8 exhibits just a few of the most commonly used derivational affixes in Shughni.

**Table 4.8:** Common derivational affixes in Shughni.

FORM	SELECTS	DERIVES	MEANING	EXAMPLE	
-(y)i	ADJ	N	abstract quality	<i>garmi</i>	‘warmth’
-idz	V	N	process	<i>zidāridz</i>	‘sweeping’
-ik	N	N	(endearing) diminutive	<i>pucik</i>	‘beloved son’
-gār*	N	ADJ	occupation	<i>naqligār</i>	‘talkative’
no-*	ADJ	ADJ	not ADJ	<i>noḃalad</i>	‘unfamiliar’

\*Borrowed from Tajik.

The use of compounding is much more restricted. It is used, for instance, in the formation of occupational nouns by juxtaposing a noun with a (borrowed) present verbal stem, as in *yoc* ‘fire’ + *kun-* ‘do’ > *yockun* ‘fire keeper’. Nominal derivation in Shughni is treated in much more detail in Section 5.3.

#### 4.2.4 Order of elements in noun phrases

Noun phrases in Shughni may consist of a single bare noun (e.g. *žīr* ‘stone’), though most often a noun is combined with at least one other element. The order of elements in (native) noun phrases is rigid, with nouns always in final position. A noun phrase may also contain a demonstrative or determiner, a numeral, an adjective, or some combination thereof, all preceding the noun in that order – hence, (DEM)–(NUM)–(ADJ)–NOUN. Examples are given in Table 4.9.

**Table 4.9:** Order of elements in NPs.

NP ELEMENTS	EXAMPLE	GLOSS
N	<i>žīr</i>	‘stone’
DET N	<i>yā žīr</i>	‘that stone’
DET NUM N	<i>wāḃ cavor žīr</i>	‘those four stones’
DET NUM ADJ N	<i>wāḃ cavor ḡulayaki žīr-en</i>	‘those four humongous stones’

Although the order ADJ–NOUN is found in virtually all noun phrases containing two native Shughni words, the reverse order may occur in the form of *izofat* constructions borrowed from Tajik (e.g. Edelman & Dodykhudoeva

2009b: 803), hence NOUN-*i* ADJ, where the *izofat* *-(y)i* links the noun with its modifier.<sup>1</sup> Thus, we may find either *aʃoib kalīmā* (ADJ-NOUN) ‘strange word’ or *kalīmā-yi aʃoib* (NOUN-ADJ), with the latter typically found in higher registers. Combinations of adjectives and nouns may also be borrowed from Russian, though in this case the order of elements matches that of Shughni (i.e. ADJ-NOUN) – as in *akademīčiski proyekt* ‘academic project’. Types of noun phrases and the order of elements within them are discussed in Section 5.1.

### 4.3 Case-like suffixes & adpositions

Shughni employs a mix of prepositions, postpositions, and case-like suffixes. Neither prepositions nor postpositions clearly outnumber the other, nor does there seem to be a common semantic thread which unites either type. Adpositions and case-like suffixes are not examined explicitly in the latter sections of the thesis. However, they consistently appear in example sentences and as pieces of other constructions. This section therefore provides an overview of each, which the reader may wish to use as a reference throughout the thesis. **Case-like suffixes** are examined in Section 4.3.1, **prepositions** in Section 4.3.2, and **postpositions** in Section 4.3.3.

#### 4.3.1 Case-like suffixes: Locative and dative

In this subsection I distinguish three case-like suffixes which I consider to be distinct from postpositions because of their variability in form and their semantic versatility. First is the **locative/possessive suffix** *-(y)and* (*-nd* with vowel-final pronouns, as in *tu-nd* ‘yours’ and optionally *-ndi* when following non-pronouns ending in a vowel, as in *Dūšanbi-ndi* ‘in Dushanbe’). This suffix is used to express *inessive* meaning – i.e. location or origin inside of – and possession.<sup>2</sup> Examples are given in (11).

#### (11) Locative *-and*: Location (within) and possession

<sup>1</sup> The palatal glide *y* is inserted when the *izofat* directly follows a vowel.

<sup>2</sup> Note that for ease of presentation I gloss this suffix as either LOC or POSS depending on its function in the clause. It is possible, of course, that this is a single morpheme expressing both location and possession. Nonetheless, it is noteworthy that certain variants of this suffix are compatible only with locative or possessive uses. For instance, the shortened form *-nd* is only possible on pronominals in expressing possession, and the form *-(a)ndi* is possible only with a locative sense. I leave a detailed treatment of the possible unification of these forms into a single morpheme for future research.

a. **Locative -and**

Bač-galā=ta daryo-**yand** šinowāri kin-en.  
 child-PL=HAB river-LOC swimming do.PRS-3PL  
 ‘Children swim in the river.’

b. **Possessive -and**

Wev-**and** ačaθ pūl nist!  
 DEM.DIR.PL-POSS no money be.NEG  
 ‘They have absolutely no money.’

Second is the **the dative suffix -(y)ard** (*-rd* with vowel-final pronouns, as in *tu-rd* ‘you-DAT’ and optionally *-ra* when following non-pronouns ending in a vowel, as in *Sayora-ra* ‘Sayora-DAT’), which has canonical dative usages as well as the ability to express approximateness in location and time. Examples of this suffix are shown in (12).

(12) **Dative -ard: Indirect object and approximateness**a. **Indirect-object -ard**

Wuz=ta ukmandaθ arčīzca xu virod-**ard** dāk-um.  
 I=FAC definitely something REFL brother-DAT give.PRS-1SG  
 ‘I’ll definitely give something to my brother.’

b. **Approximate location -ard**

Māš=ām čīd-**ard** wi kud xīkūd=atā wi=yām na-virūd.  
 we=1PL house-DAT DEM.OBL.M dog search.PST=but him=1PL NEG-find.PST  
 ‘We searched for that dog around the house but didn’t find him.’

And third is the **the locative suffix -ti**, which expresses *adessive* and *allative* meaning – i.e. location or movement onto – and can also be used to mark instrumental, causative, and certain participial constructions. An example of a traditional locative meaning with *-ti* is shown in (13a), while an instrumental usage is shown in (13b).

(13) **Locative -ti: Location (on) and causative, instrumental**a. **Locative -ti**

Ik=šič pūnd-**ti** žiniḡ dis lap.  
 PREC=now road-LOC snow so much  
 ‘There is a lot of snow on the road right now.’

b. **Instrumental -ti**

Yu            čorik=i   xu   tāt=at   nān   pūl-ti   yi čīd   xarīd čūd.  
 DEM.DIR.M man=3SG REFL father=and mother money-INS a house buy do.PST  
 ‘That man bought a house using his parents’ money.’

### 4.3.2 Prepositions

Shughni’s inventory of prepositions include the language’s three core directional and locative adpositions: *tar*, *ar*, and *pi*, which encode horizontal, downward, or upward direction or location, respectively. These prepositions encode inherent deixis, as the speaker is forced to indicate whether movement or location is down, up, or horizontal from the point of reference.

Examples with each preposition are given in (14)–(16). Note that the use of one of these prepositions when discussing movement from one geographic location to another, especially regarding movement within the mountainous Badakhshan region where Shughni is spoken, generally makes reference to the change in elevation. For instance, because Khorugh is located at significantly higher elevation than Dushanbe, speakers typically use the preposition *ar* ‘down’ when describing movement from Khorugh to Dushanbe, and *pi* ‘up’ when describing movement the other direction (see Mueller 2014 for more on the use of these prepositions within the mountainous environment of the Shughni-speaking region).

(14) **Preposition *tar*: Upward location or movement**

Māš=ām **tar** boy sat.  
 we=1PL to gardens go.PST  
 ‘We went to the garden. (The garden is not located significantly above or below us.)’

(15) **Preposition *ar*: Downward location or movement**

Māš=ām fukaθ **ar** Dūšanbi sat.  
 we=1PL all down.to Dushanbe go.PST.PL  
 ‘We all went to Dushanbe. (From a place higher up.)’

(16) **Preposition *pi*: Upward location or movement**

Memūn-en=en mis **pi** Xaray sat.  
 guest-PL=3PL also to Khorog go.PST.PL  
 ‘The guests also went to Khorog. (From a place below Khorog.)’

Other common prepositions are given in Table 4.10. Note that two of these prepositions – *mi* and *či* – are restricted



to use with a few body parts, indicating an object's location, in the case of *mi*, and a person's position resting on the body part, in the case of *či*.

**Table 4.10:** Common prepositions in Shughni.

PREPOSITION	GLOSS	EXAMPLE	GLOSS
<i>az</i>	'from'	<i>az Tojikistūn</i>	'from Tajikistan'
<i>be</i>	'without'	<i>be tuyaθ</i>	'without you'
<i>bādi</i>	'after'	<i>bādi kor</i>	'after work'
<i>tūnec</i>	'while'	<i>tūnec tu nān na-yaθč</i>	'while your mom hasn't arrived'
<i>mi</i>	'in (a body part)'	<i>mi θust</i>	'in (one's) hand'
<i>či</i>	'on (a body part)'	<i>či dām</i>	'on (one's) back'

It is also worth noting that in certain cases in which the object of a preposition is a noun phrase with a possessor, particularly in instances of inalienable possession involving body parts, the order of the preposition and possessor is obligatorily inverted. Hence, where we would expect the order PREP–POSSESSOR–POSSESSEE, as in 'in my hand', we get instead POSSESSOR–PREP–POSSESSEE, or the equivalent of 'my in hand'. Compare, for instance, the alienable possessive construction *pi mu čīd*, lit. '(up) to my house' (PREP–POSS–N), with the inalienable possessive construction *mu mi θust*, lit. 'my in hand' (POSS–PREP–N). This syntactic strategy of expressing inalienable possession, where the preposition is moved inside the possessive NP, is also found in Yidgha and the other languages of the Shughni-Rushani group and has been linked, at least in part, to the influence of analogous constructions in neighboring non-Indo-European languages such as Burushaski (e.g., Edelman 1980b).

### 4.3.3 Postpositions

Common postpositions are given in Table 4.11. Note that the postposition *ja* indicates physical location with or on a person, as in 'in one's possession' or 'at one's location'.

The word *xez* exhibits behavior in between a noun and a postposition. On the one hand, it is used only in locative constructions and does not have an independent usage as a noun outside such constructions. On the other, it cannot be used as the sole locative element in a phrase and must occur immediately following a noun. Thus, we find examples such as *tar mu xez* 'to me'; *māš xez-and*; and *universitet xez-ard* 'around the university'.

**Table 4.11:** Common prepositions in Shughni.

<u>PREPOSITION</u>	<u>GLOSS</u>	<u>EXAMPLE</u>	<u>GLOSS</u>
<i>va</i>	‘along’	<i>pūnd va</i>	‘along the road’
<i>ja</i>	‘at; with’	<i>mu ja</i>	‘with me; in my vicinity’
<i>qati</i>	‘with’	<i>namak qati</i>	‘with salt’
<i>jāt</i>	‘for’	<i>māš safar jāt</i>	‘for our trip’
<i>garginuṣṣā</i>	‘across (from)’	<i>tu čīd garginuṣṣā</i>	‘across from your house’

## 4.4 Verbs

The presentation of Shughni verbs in this section is a primer for the four chapters in Part III of the thesis, which includes Chapter 8 on verb stems, inflection, and adverbial modification; Chapter 9 on regular and irregular verbs; Chapter 10 on tense, aspect, and mood (TAM); and Chapter 11 on argument structure and transitivity. The present section provides a look at the fundamentals of the verbal system and is divided into subsections on verb stems (Section 4.4.1), classes of (in)transitive verbs – i.e. transitive, unergative, and unaccusative verbs (Section 4.4.2), verbal inflection (Section 4.4.3), complex verbs (Section 4.4.4), and passive-like and causative constructions (Section 4.4.5).

### 4.4.1 Verb stems and verbal inflection

A fundamental concept in Shughni verbs is the notion that each verb – except for a handful of defective verbs – has four stems: *infinitive*, *present*, *past*, and *perfect*. All combinations of tense, aspect, and mood in the language are formed using one of these verb stems in combination with inflectional elements, auxiliary verbs, and in some cases non-verbal elements such as adpositions.

#### 4.4.1.1 Formation of verb stems

In regular verbs, each of the four stems can be predicted based on the form of any of the other stems. Infinitive stems, whether of a regular or irregular verb, always contain a final *-t* (following voiceless consonants and nasals) or *-d* (following vowels and voiced non-nasal consonants). Given the infinitive of a regular verb, the three other

stems will be as follows: the past stem is identical to the infinitive; the present stem is the infinitive without the final *-t/-d*; and the perfect stem is the infinitive stem with the final *-t/-d* affricated to *-č/-ǰ*. Three sample regular verbs are given in Figure 4.1 to illustrate this pattern.

**Figure 4.1:** Stem formation in three regular verbs

INF	<i>qīwd-</i>	INF	<i>binêst-</i>	INF	<i>nivîšt-</i>
PST	<i>qīwd</i>	PST	<i>binêst</i>	PST	<i>nivîšt</i>
PRF	<i>qīwǰ</i>	PRF	<i>binêšč</i>	PRF	<i>nivîšč</i>
PRS	<i>qīw-</i>	PRS	<i>binêš-</i>	PRS	<i>nivîš-</i>
GLOSS	‘call’	GLOSS	‘lose’	GLOSS	‘write’

However, many Shughni verbs are irregular. In these cases, the knowledge of one stem is not sufficient to determine the form of the other stems. Taking infinitive stems as a base, the remaining verb stems may be irregular for a variety of reasons. For instance, they may be irregular because of a change in stem vowel, a change, deletion, or addition of the stem-final consonant, or, in at least one case, suppletion. In the case of perfect stems, a preceding fricative *-θ/ð* or *-x/ǰ* may be inserted immediately before the final affricate. Three sample irregular verbs are given in Figure 4.2 to illustrate these types of irregularities.

**Figure 4.2:** Stem formation in three irregular verbs

INF	<i>wīnt-</i>	INF	<i>xīd-</i>	INF	<i>čīd-</i>
PST	<i>wīnt</i>	PST	<i>xūd</i>	PST	<i>čūd</i>
PRF	<i>wīnč</i>	PRF	<i>xūǰǰ</i>	PRF	<i>čūǰǰ</i>
PRS	<i>wīn-</i> (3SG <i>wīn-</i> )	PRS	<i>xār-</i>	PRS	<i>kin-</i>
GLOSS	‘see’	GLOSS	‘eat’	GLOSS	‘do’

These verbs are presented from left to right in order of degree of irregularity, as it were. In the first verb *wīnt* ‘see’, the only irregularity is the change of present-stem vowel from long *ī* to short *i*, though the long vowel is preserved in the third-singular form. In the verb *xīd* ‘eat’, the present stem vowel is changed to *ā* rather than *ī* with the addition of a consonant *r*. In the past and perfect stems the stem vowel is *ū*. Also, in the perfect stem of *xīd* we find the insertion of a fricative *ǰ* before the final affricate. The final verb *čīd* ‘do’ is similar to *xīd* in its past and perfect stems, but its present stem exhibits suppletion.

Present stems in Shughni are primarily monomorphemic, with the status of past, perfect, and infinitive stems

depending on one's analysis of the stem-final stop (in the case of the former two) and affricate (in the case of the latter), and in particular whether these portions of the stems are analyzed as distinct morphemes. In any case, many verb stems, especially those with two or more syllables, contain what was once a productive pre-verbal element in Old Iranian (e.g. *fīrīptow* 'arrive', where the underlined portion *fīr-* descends from *\*fra* 'forth; forward'; see Section 9.1.1). However, such elements are no longer productive in the modern language. New verbal lexemes are mostly formed through borrowing, via the use of (native) light verbs together with a non-verbal (borrowed) element.

#### 4.4.1.2 Distribution and usage of verb stems

As mentioned above, all tense, aspect, and mood distinctions in Shughni are expressed using one of a verb's four stems. Each stem, in turn, is used in multiple types of constructions and is compatible with a specific set of inflectional and/or derivational affixes.

**Present stems** are used in present-tense indicative and subjunctive constructions, as well as in future and habitual clauses in combination with the second-position clitic *=ta*. In these constructions, present stems take the present-tense person-number agreement suffixes. Bare present stems are used as second-singular imperative forms and may also attach to certain derivational affixes to create verbal nouns.

**Past stems** are used invariably with past-tense second-position agreement clitics. They appear predominantly in non-perfect constructions with past temporal reference. In some instances, however, a past stem is used in a sentence with non-past temporal reference. This only occurs in certain subordinate clauses whose matrix counterpart has a present stem. It will be argued in Section 10.2.2 that even in these cases, past verb stems still have (relative) past temporal reference.

**Perfect stems** are used in perfect constructions, generally with past temporal reference. As in English, perfect stems differ from strictly past-tense clauses in that they inherently imply that the predicate is relevant to the circumstances at the time of utterance. Perfect stems are also used in past subjunctive clauses, including conditionals, as well as to indicate indirect evidence for the information being relayed (a type of evidential usage). In all of the previously mentioned usages, perfect stems call for past-tense agreement clitics, but they are also used as a base for forming resultative participles in *-ak* and *-in*, in which case the person-number agreement appears on the auxiliary verb.

Finally, **infinitive stems** are used in a variety of contexts. Bare infinitive stems (e.g. *vīd* ‘bring’) are used with certain adpositions and locative suffixes in periphrastic constructions expressing nuances in aspect. For instance, the locative suffix *-ti* combines with bare infinitive stems to form a present active participle, and the prepositional elements *daraw* and *či* combine with bare infinitive stems to show inceptive and prospective aspects, respectively. Infinitive stems are widely used with the suffix *-ow* (e.g. *vīd-ow* ‘to bring’), which I gloss PURP for *purposive*. This suffix is not only used in a verb’s citation form, but is also used with infinitive verbs in argument position, to express a purposive relationship between a non-finite and finite clause, and with infinitives in adverbial clauses.

The distribution of these four types of verb stems is summarized in Table 4.12.

**Table 4.12:** Distribution of verb stem types in Shughni.

<u>Stem</u>	<u>Construction Type</u>	<u>Inflection Type</u>
PRESENT	<ul style="list-style-type: none"> <li>• present indicative &amp; subjunctive</li> <li>• future &amp; habitual (clitic =<i>ta</i>)</li> <li>• imperative</li> </ul>	<ul style="list-style-type: none"> <li>• present agreement suffixes</li> <li>• no gender agreement</li> </ul>
PAST	<ul style="list-style-type: none"> <li>• past (non-perfect)</li> </ul>	<ul style="list-style-type: none"> <li>• past-tense clitics</li> <li>• gender distinction in unaccusatives</li> </ul>
PERFECT	<ul style="list-style-type: none"> <li>• past (perfect)</li> <li>• past conditional</li> <li>• resultative participles</li> </ul>	<ul style="list-style-type: none"> <li>• past-tense clitics</li> <li>• gender distinction in unaccusatives</li> </ul>
INFINITIVE	<ul style="list-style-type: none"> <li>• verbal nouns</li> <li>• aspectual distinctions (with adpositions &amp; locative suffixes)</li> </ul>	<ul style="list-style-type: none"> <li>• none</li> </ul>

#### 4.4.2 Transitive, unergative, and unaccusative verbs

Before moving forward, now is an ideal time to point out the existence of three classes of verbs in Shughni which will be useful in subsequent discussions of various grammatical phenomena throughout the thesis. In Parker 2020, I apply the labels **transitive**, **unergative**, and **unaccusative**, originally given in Perlmutter 1978, to these classes of Shughni verbs. I will retain these same labels throughout this thesis.

The fundamental idea is that, in addition to *transitive* verbs, which require direct objects, there are two classes of

intransitive verbs – unergative and unaccusative. Subjects of *unergative* verbs share certain syntactic characteristics with subjects of transitive verbs, while subjects of unaccusative verbs share certain characteristics with objects of transitive verbs. Note that I follow several authors – including Comrie (1978) and Dixon (1979; 1994) – in using the labels A and O for the subject and object of a transitive verb, respectively, and S<sub>A</sub> and S<sub>O</sub> for the subject of unergative and unaccusative verbs, respectively.

From a semantic perspective, the set of Shughni verbs which I label unaccusative generally correspond to the kinds of verbs identified by Perlmutter (1978) as unaccusative. That is, they are by and large verbs of directed motion such as *sittow* ‘go’, *yattow* ‘come’, and *firiṭtow* ‘arrive’, as well as non-volitional actions such as *mīdow* ‘die’ and *virixṭow* ‘break’. Unergative verbs, for their part, generally denote manner of motion or volitional actions – e.g. *žēṭow* ‘to run’ and *nīwdow* ‘to cry’ (cf. Parker 2020: fn. 12).<sup>3</sup>

In what follows, I discuss four aspects of the syntactic behavior of each class of verbs which may serve as diagnostics to distinguish unergative and unaccusative verbs (with a summary provided in Section 4.4.2.5):

Diagnostic 1: **Compatibility with the third-singular past-tense clitic =(y)i**, found with transitive and unergative verbs, to the exclusion of unaccusatives (Section 4.4.2.1);

Diagnostic 2: **Compatibility with resultative (passive-like) participles (suffixes -ak and -in)**, available with transitive and unaccusative verbs, to the exclusion of unergatives (Section 4.4.2.2);

Diagnostic 3: **Gender distinction in past and perfect stems**, occurring only in (a subset of) unaccusative verbs (Section 4.4.2.3); and

Diagnostic 4: **Morphological causatives**, available only for unaccusative verbs (Section 4.4.2.4).

---

<sup>3</sup>There are a number of Shughni verbs which constitute apparent exceptions to the correlation between the notion of volitionality (or agentivity) and unergativity, on the one hand, and a lack of volitionality and unaccusativity, on the other. For instance, there are some verbs, such as *riwiṭtow* ‘fly away’ and *zibīntow* ‘jump’, which seem to denote agentive actions but nonetheless behave syntactically as unaccusative in Shughni. Similarly, the verbs *nīwdow* ‘cry’ and *šīntow* ‘laugh’, which pattern as unergative verbs in Shughni, may have unaccusative counterparts in other languages. From a pan-Iranian perspective, two classes of intransitive verbs – one which patterns syntactically with transitive verbs and one which forms its own class of intransitive verbs – are found in other Eastern Iranian languages (e.g. Pashto; see David 2014). Nonetheless, the content of each class differs from language to language, and a closer examination of these classes across Iranian languages remains an important goal for future research.

#### 4.4.2.1 Diagnostic 1: Compatibility with third-singular past-tense clitic =(y)i

Both transitive and unergative verbs obligatory trigger a second-position clitic =(y)i with third-person singular past-tense subjects, while unaccusative verbs never trigger this clitic. This characteristic aligns the subjects of transitive verbs with the subjects of unergative verbs ( $A$  and  $S_A$ ), both of which trigger a clitic, and subjects of unaccusative verbs with objects of transitive verbs ( $S_O$  and  $O$ ), which do not trigger a clitic.

The examples in (17)–(18) illustrate this pattern. The examples in (17a) and (17b) show that the third-singular clitic is required for the transitive and unergative verbs *žėydw* ‘read’ and *žėxtow* ‘run’, respectively. Example (18a) shows that the clitic is illicit with the third-singular subject of the unaccusative verb *tīdw* ‘leave’, and example (18b) demonstrates that a third-singular clitic is likewise incompatible with the third-singular object of the transitive verb *žėydw*.

#### (17) Compatibility of transitive and unergative subjects with the 3<sub>SG</sub> past-tense clitic =(y)i

##### a. Transitive ( $A$ ): *žėydw* ‘read’

Yā\*(=yi) yi kitob žėyd.  
she=3<sub>SG</sub> a book read.PST  
‘She **read** a book.’

##### b. Unergative ( $S_A$ ): *žėxtow* ‘run’

Yā\*(=yi) lap žėxt.  
she=3<sub>SG</sub> much run.PST  
‘She **ran** a lot.’

#### (18) Incompatibility of unaccusative subjects and transitive objects with the 3<sub>SG</sub> past-tense clitic =(y)i

##### a. Unaccusative ( $S_O$ ): *tīdw* ‘leave’

Yā(\*=yi) toyd.  
she(\*=3<sub>SG</sub>) leave.PST.F  
‘She **left**.’

##### b. Object of transitive ( $O$ ): *žėydw* ‘read’

Wuz=um(\*=i) tu kitob žėyd.  
I=1<sub>SG</sub>(\*=3<sub>SG</sub>) your book read.PST  
‘I **read** your book.’

This pattern is summarized in Table 4.14. The cells of argument types which pattern alike are organized based on shading: those of  $A$  and  $S_A$  are unshaded, while those of  $S_O$  and  $O$  are shaded.

I take the incompatibility with the clitic =(y)i to be the definitive boundary between transitive and unergative verbs, on the one hand, and unaccusative verbs on the other. That is, any verb which does not require this clitic with third-singular past-tense subjects is necessarily unaccusative, and any verb which requires this clitic is either transitive

**Table 4.13:** Verb classes and compatibility with the third-singular past-tense clitic =(y)i.

<u>ARG. TYPE</u>	<u>LABEL</u>	<u>TRIGGERS =(y)i?</u>
<b>Subj. of transitive</b>	A	✓
<b>Subj. of unergative</b>	S <sub>A</sub>	✓
<b>Subj. of unaccusative</b>	S <sub>O</sub>	✗
<b>Obj. of transitive</b>	O	✗

or unergative. In other words, compatibility with the clitic =(y)i is the definitive diagnostic used to distinguish the two types of intransitive verbs, unergative and unaccusative. It will be seen below that the remaining three characteristics of unaccusative verbs to be discussed, although largely consistent, are imperfect diagnostics.

#### 4.4.2.2 Diagnostic 2: Compatibility with resultative (passive-like) participles

Both transitive and unaccusative verbs, but not unergative verbs, have resultative participial forms (examined in Section 11.1). Resultative participles come in two types: (i) *verbal participles* built on a verb's perfect stem (uninflected for gender) with the addition of the suffix *-ak* (e.g. *rinũxč-ak* 'forget.PRF-V.PTPL') and (ii) *adjectival participles* built on a verb's perfect stem (inflected for gender, where possible) together with the suffix *-in* (e.g. *rinũxč-in* 'forget.PRF-A.PTPL'). Verbal participles are used with the auxiliary verb *sittow* 'become' and resemble a passive, although the status of this construction as a true passive is called into question in Section 11.1. Adjectival passive participles, for their part, may be used either attributively within a noun phrase or predicatively in a copular construction.

This diagnostic is grounded in the idea that both transitive and unaccusative verbs have arguments which are semantic patients, whereas unergative verbs have only a single agentive argument. The object of a transitive and the subject of an unaccusative verb, both semantic patients, are capable of being subjects of verbal resultative participial constructions or being modified by adjectival resultative participles. Neither transitive nor unergative subjects – both semantic agents – are compatible in these positions.

The examples in (19)–(21) illustrate this pattern. First consider those in (19). The sentence in (19a) shows a canonical transitive construction with the complex verb *joy čīdow* 'to hide'. The latter two examples show that the object (semantic patient) of the first example is capable of being the subject of a construction with a verbal



resultative participle (19b) and of being modified by an adjectival resultative participle (19c).

(19) **Resultative participles with the transitive verb *joy čīdow* ‘hide’**

a. **Transitive construction**

yulla-yen=en [wev šīrīni-yen]<sub>PAT</sub> **joy čūd.**  
 adult-PL=3PL those.OBL sweet-PL place do.PST  
 ‘The adults hid the candies.’

b. **Verbal resultative participle *joy čūčj-ak***

[Wāð šīrīni-yen]<sub>PAT</sub>=en **joy čūčj-ak** sat.  
 those.DIR sweet-PL=3PL place do.PRF-V.PTPL become.PST.PL  
 ‘The candies were hidden.’

c. **Adjectival resultative participle *joy čūčj-in***

Kūdak-en=en [wev **joy čūčj-in** šīrīni-yen]<sub>PAT</sub> virūd.  
 child-PL=3PL those.OBL place do.PRF-ADJ.PTPL sweet-PL find.PST  
 ‘The children found the hidden candies.’

The examples in (20) show that the unaccusative verb *načtīdow* ‘exit; leave’, a canonical active construction of which is shown in (20a), can also serve as the main verb in a verbal participial construction (20b) and as an adjectival participle (20c).

(20) **Resultative participles with the unaccusative verb *načtīdow* ‘exit; go out’**

a. **Canonical unaccusative construction**

[Yā xāc]<sub>PAT</sub> az wam baklaška **načtoyd.**  
 that.DIR.F water from that.OBL.F bottle exit.PST.F  
 ‘The water came out of that bottle.’

b. **Verbal resultative participle *načtūj-ak***

[Yā xāc]<sub>PAT</sub> **načtūj-ak** sat.  
 that.DIR.F water exit.PRF-V.PTPL become.PST.F  
 ‘The water came out (lit. got/became come out).’

c. **Adjectival resultative participle *načtīc-in***

Tu=t [wam **načtīc-in** xāc]<sub>PAT</sub> wīnt o?  
 you=2SG that.OBL.F exit.PRF.F-ADJ.PTPL see.PST PQ  
 ‘Did you see the water that came out (lit. the come-out water).’

Lastly, the examples in (21) exhibit the incompatibility of the unergative verb *žěxtow* ‘run’ with resultative participles. The first example exhibits a canonical active construction with this verb and shows that it requires the third-singular clitic  $=\langle y \rangle i$  with past-tense verbs. The latter two examples show that this verb cannot be the main verb in resultative participial constructions (21b) and cannot serve as an adjectival participial modifier (21c).

(21) **Resultative participles with the unergative verb *žěxtow* ‘run’ – Illicit**

a. **Canonical unergative construction**

[Yā      ŷinik]<sub>AGT=i</sub> **žěxt.**  
 that.DIR.F woman=3SG run.PST  
 ‘That woman ran.’

b. **Verbal resultative participle *žěxč-ak* – Illicit**

\*[Yā      ŷinik]<sub>AGT</sub> **žěxč-ak**      sat.  
 that.DIR.F woman    run.PRF-V.PTPL become.PST  
 Intended: ‘That woman ran (or *was run*).’

c. **Adjectival resultative participle *žěxč-in* – Illicit**

\*[Yā      **žěxč-in**      ŷinik]<sub>AGT</sub> tar kā      sat?  
 that.DIR.F run.PRF-A.PTPL woman    to    where go.PST.F  
 Intended: ‘Where did that woman who ran go? (lit. that run woman)’

A summary of this pattern is given in Table 4.14. Again, we find that subjects of transitive and unergative verbs pattern alike in this regard, while subjects of unaccusatives and objects (patients) of transitives pattern in a different way. The latter two are capable of being subjects of verbal resultative participial constructions and of being modified by adjectival resultative participles, while the former two are incapable of appearing in such positions.

**Table 4.14:** Verb classes and compatibility with the third-singular past-tense clitic  $=\langle y \rangle i$ .

<u>ARG. TYPE</u>	<u>LABEL</u>	<u>CAN BE SUBJECT OF RESULTATIVE PTPL. WITH -ak OR MODIFIED BY ADJECTIVAL PTPL. WITH -in?</u>
<b>Subj. of transitive</b>	A	✗
<b>Subj. of unergative</b>	S <sub>A</sub>	✗
<b>Subj. of unaccusative</b>	S <sub>O</sub>	✓
<b>Obj. of transitive</b>	O	✓

On the diagnostic of resultative participles, if an intransitive verb has a verbal and/or adjectival participle, then it is unaccusative. However, note that not all unaccusative verbs have verbal participles. For instance, the verb *yattow* ‘come’ is unaccusative by virtue of the fact that it does not trigger the third-singular clitic  $=(y)i$ , but it does not have a verbal participle *\*yaθč-ak*. Ultimately, therefore, compatibility with resultative participles is an imperfect diagnostic. Any verb which has such participles is necessarily unaccusative, but even if an intransitive verb does not have such participles, other diagnostics should be used to confirm its status as unaccusative or unergative.

#### 4.4.2.3 Diagnostic 3: Gender distinction in past and perfect stems

Many unaccusative verbs distinguish gender in their past and perfect stems, whereas virtually no transitive or unergative verbs do so. Examples exhibiting the (lack of) gender distinction for each type of verb are given in (22)–(24).

##### (22) Transitive *ǰêydow* ‘read’: No gender distinction

###### a. Feminine subject: *ǰêyd*

Yā=yi yi kitob ǰêyd.  
she a book read.PST  
‘She read a book.’

###### b. Masculine subject: *ǰêyd*

Yu=yi yi kitob ǰêyd.  
he a book read.PST  
‘He read a book.’

##### (23) Unergative *ǰêxtow* ‘run’: No gender distinction

###### a. Feminine subject: *ǰêxt*

Yā=yi lap ǰêxt.  
she=3SG much run.PST  
‘She ran a lot.’

###### b. Masculine subject: *ǰêxt*

Yu=yi lap ǰêxt.  
he=3SG much run.PST  
‘He ran a lot.’

##### (24) Unaccusative *tīdow* ‘leave’: Gender distinction

###### a. Feminine subject: *toyd*

Yā toyd.  
she leave.PST.F  
‘She left.’

###### b. Masculine subject: *tūyd*

Yu tūyd.  
he leave.PST.M  
‘He left.’

Note, however, that not all unaccusative verbs distinguish gender. For instance, the unaccusative verbs *yattow* ‘come’, *firīptow* ‘arrive’, and *bedow* ‘disappear’ all lack the third-singular clitic in the past tense but do not distinguish gender in the past or perfect stems. This diagnostic, therefore, like the one above regarding resultative participles, is an imperfect one. If an intransitive verb has gender-distinguishing forms in its past and/or perfect stems, it is necessarily unaccusative. However, even if it does not have gender-distinguishing forms, other diagnostics must be performed to determine whether it is unergative or unaccusative. Note, further, that subjects of unaccusative verbs do not align with objects of transitive verbs in this regard, as objects are never capable of gender agreement in this way.

#### 4.4.2.4 Diagnostic 4: Availability of morphological causative forms

The fourth and final diagnostic regards morphological causatives. Most unaccusative verbs have morphological causative counterparts, whereas morphological causatives are generally not found with unergatives or transitives (with a few possible exceptions, discussed below). Morphological causative forms in Shughni seem to be available only for verbs whose subjects are semantic patients – i.e. unaccusative verbs – and not for verbs whose subjects are semantic agents – i.e. transitive and unergative verbs. This interpretation is supported by the notion that morphological causatives encode direct causation, in which the causee is non-agentive (see Section 11.2.3 on the semantics of morphological causatives versus that of periphrastic causatives).

Morphological causatives may be formed either via suffixation, often with the addition of *-en* or *-ûn* to the present stem, or by stem-internal ablaut, usually where an unaccusative present stem contains *-ā-* and the transitive counterpart contains *-ê-*. For instance, the unaccusative verb with present stem *firāp-* ‘arrive’ has the ablaut causative *firēp-* ‘deliver; take’, and the unaccusative verb with present stem *warv-* ‘boil’ has a suffixal causative in *warv-en* ‘cause to boil’. (See Section 11.2.1 for a more detailed discussion of morphological causatives.)

Transitive verbs and unergative verbs have with morphological causative counterparts are exceedingly rare. To my knowledge, the only transitive verb with a morphological causative counterpart is *fām-* ‘know’ > *fām-ûn-* ‘explain’, likely borrowed from Tajik *fahm-on* ‘explain’, also a morphological causative with the Tajik causative suffix *-on*. Regarding unergatives with morphological causatives, I am only aware of two: *nāw-* ‘cry.PRS’ → *rinêw-* ‘make cry’ and *šānd-* ‘laugh.PRS’ → *šand-ûn-* ‘make laugh’, both of which are marginal forms in the language.<sup>4</sup>

<sup>4</sup>The verb *rinêw-* ‘make cry’ contains an extra syllable *ri-*, while, as far as I am aware, no causatives formed from unaccusatives have such

In any case, the vast majority of transitive and unergative verbs lack morphological causative counterparts, while many unaccusative verbs possess morphological causatives formed from their present stems. The schemata in (25)–(26) show the lack of morphological causative for transitive *mīzdow* ‘build’ (prs. stem *moz-*) and unergative *žēxtow* ‘run’ (prs. stem *žoz-*). Note that neither the ablaut causative (top form in each column) nor the suffixes *-en* or *-ūn* (both commonly used with unaccusatives) are possible with these verbs. The schema in (27) shows the existence of a morphological (ablaut) causative for the unaccusative verb *riwixtow* (prs. stem *riwāz-* ‘fly off’).

(25) **Tr. *mīzdow* ‘build’**

(26) **Unerg. *žēxtow* ‘run’**

(27) **Unacc. *riwixtow* ‘fly off’**

<i>moz-</i> (PRS.)	→	<i>*mēz-</i>		<i>žoz-</i> (PRS.)	→	<i>*žēz-</i>		<i>riwāz-</i> (PRS.)	→	✓ <i>riwēz-</i> ‘make fly off’
		<i>*moz-ūn-</i>				<i>*žoz-ūn-</i>				
		<i>*moz-en-</i>				<i>*žoz-en-</i>				
		Int.: ‘make build’				Int.: ‘make run’				

This diagnostic is also imperfect, however. Not only are there at least three apparent exceptions (mentioned above) to the notion that morphological causatives are only available with unaccusative verbs. There are also instances in which unaccusative verbs either do not have a morphological causative counterpart, or else the relationship between an unaccusative verb and its purported causative counterpart is unclear. For instance, the unaccusative verb *tīdow* ‘go; leave’ (pres. stem *tī-*) has an apparent causative counterpart in *tēzdow* ‘filter’. However, not only is a semantic relationship between these two verbs not immediately apparent (albeit not inconceivable), but the form of the causative has an additional consonant in its stem (possibly corresponding to the third-singular form of *tīdow*: *tīzd* ‘goes’). This diagnostic should therefore be employed along with others to determine whether a given intransitive verb is unaccusative or unergative.

#### 4.4.2.5 Summary: Transitive, unergative, and unaccusative verbs

This section has introduced three classes of verbs in Shughni: transitive, unergative, and unaccusative. These classes of verbs are distinguished by the syntactic properties of the verbs themselves (e.g. ability to exhibit agreement in gender), as well as the type of arguments with which they are compatible. Transitive and unergative verbs

---

an extra syllable. Moreover, although *rinēw-* is listed in a number of publications on Shughni (e.g. Sokolova 1967; Dodykhudoeva 1988, a.o.), it apparently is not used in the modern language, at least in the dialect spoken in Khorog. The verb *šandūn-* is likewise not accepted by all speakers and is possibly a calque of Tajik *xand-on-* ‘make laugh’, from present stem *xand-* ‘laugh’, formed with the Tajik causative suffix *-on-*.

have subjects which are semantic agents, while unaccusative verbs have subjects which are semantic patients. In certain regards, unaccusative subjects pattern with transitive objects, which are also semantic patients. For instance, neither unaccusative subjects nor transitive objects trigger the third-person singular past-tense enclitic  $=(\nu)i$ , whereas this clitic is obligatory with third-singular past-tense unergative and transitive subjects.

A summary of the properties attributed to transitive, unergative, and unaccusative verbs is provided in the Table in 4.15. To reiterate, the syntactic property which most clearly and consistently distinguishes unaccusative verbs from unergative verbs is the lack of the clitic  $=(\nu)i$  with third-person singular past-tense subjects. Gender agreement, resultative participles, corresponding morphological causatives are found with many, but not all, unaccusative verbs.

**Table 4.15:** Properties of transitive, unergative, and unaccusative verbs.

<u>CLASS</u>	<u>DIRECT OBJECT</u>	<u>3SG-PAST PARTICIPLES</u>	<u>RESULTATIVE CLITIC</u>	<u>GENDER AGR.</u>	<u>MORPH. CAUS.</u>	<u>EXAMPLE</u>
<b>Trans.</b>	✓	✓	✓	✗	✗	<i>qīwdow</i> ‘call’
<b>Unerg.</b>	✗	✓	✗	✗	✗	<i>žēxtow</i> ‘run’
<b>Unacc.</b>	✗	✗	✓	✓	✓	<i>sittow</i> ‘go’

### 4.4.3 Verbal inflection

This subsection presents verbal inflection. Person-number agreement is presented in Section 4.4.3.1, and gender agreement is presented in Section 4.4.3.2.

#### 4.4.3.1 Person and number agreement.

Shughni uses a system of person-number verbal inflection found in many Iranian languages, including all Pamir languages except Munji (Sokolova 1973), whereby present stems combine with *present-tense agreement affixes*, while past and perfect stems combine with *past-tense agreement clitics*. Both present-tense affixes and past-tense clitics serve essentially the same purpose of indexing the person-number features of the subject, but they differ in

both form and distribution.

Whereas agreement suffixes always attach directly to the right edge of a present verb stem, past-tense clitics pattern as second-position morphophonological clitics. As such, they appear immediately following the first constituent in a clause and are free to attach to constituents of almost any category, excluding complementizers and conjunctions. The examples of present-tense and past-tense sentences in (28) and (29) illustrate these differences. Note that the past-tense clitic *=at* in (29) does not attach directly to the verb stem; indeed, past-tense clitics most often do not attach to the verb stem.

(28) **Present stem: Suffix**

Tu tar k̄a s̄aw-i?  
 you to where go.PRS-2SG  
 ‘Where are you going?’

(29) **Past stem: (Second-position) clitic**

Tu=t tar k̄a sat?  
 you=2SG to where go.PST.F  
 ‘Where did you go?’

Agreement suffixes and clitics display a lot of overlap in form. The paradigms of each, given in Tables 4.16 and 4.17 differ only in two (of six) person-number combinations, namely the second- and third-person singular cells. These cells are grayed out.

**Table 4.16:** Present-tense agreement suffixes.

	SG	PL
1	-um	-ām
2	-i	-et
3	-t/-d	-en

**Table 4.17:** Past-tense clitics.

	SG	PL
1	=um	=ām
2	=(a)t	=et
3	=i/∅	=en

Note that, as discussed in Section 4.4.2, the third-singular past-tense clitic is used with transitive and unergative verbs but not with unaccusative verbs. This phenomenon is a crucial component of the language’s alignment system and is thought to be a vestige of a once more robust ergative system in the past tense (e.g. Stump & Hippisley 2011; see Section 12.4 for a thorough treatment of vestigial ergativity).

#### 4.4.3.2 Gender agreement in verbs

In addition to person-number agreement expressed via the suffixes and clitics presented above, a subset of unaccusative verbs also shows gender (and number) agreement via stem-vowel and stem-final consonant alternations.

In modern Shughni, verbs which distinguish gender generally have two forms for each of the past and the perfect stems: a masculine form and a feminine/plural form.<sup>5</sup> Vowel alternations are found in both past and perfect stems, but stem-final consonant alternations are found only in perfect stems. Here, the masculine form typically contains alveopalatal *j* or *č*, while the feminine form typically contains alveolar *dz* or *c*. Common vowel-alternation patterns found in the past and perfect stems of Shughni unaccusative verbs are given in 4.18.

**Table 4.18:** Common vowel alternations in gender-distinguishing verb stems.

Vowel Alternation (M~F)		VERB	GLOSS	Example	
PAST	PERF			PAST	PERF
<i>u~a</i>	<i>u~i</i>	<i>vidow</i>	‘be’	<i>vud~vad</i>	<i>vuđj~vic</i>
<i>ī~ā</i>	<i>ī~ī</i>	<i>naŷjīdow</i>	‘pass’	<i>naŷjīd~naŷjād</i>	<i>naŷjīđj~naŷjīč</i>
<i>ū~o</i>	<i>ū~ī</i>	<i>nīstow</i>	‘sit’	<i>nūst~nost</i>	<i>nūsč~nīsc</i>
<i>o~o</i>	<i>o~ē</i>	<i>xičidow</i>	‘freeze’	<i>xicod~xicod</i>	<i>xičođj~xicêc</i>

More details on the morphological expression of gender in verbs, as well as the historical processes which have led to the current state of affairs, are given in Section 9.2.

#### 4.4.4 Complex verbs

The previous subsections have dealt primarily with the structure of simplex verbs in Shughni. However, a major role in the grammar is also played by complex verbs, which are built on a non-verbal component which is uninflected (except in the case of gender-distinguishing adjectives) and a light verb. In complex verbs, the non-verbal component typically contributes the core semantic information, while the light verb undergoes any inflection and indicates information related to transitivity and TAM.

The non-verbal component in complex verbs can be from a variety of syntactic categories but is most commonly a noun or adjective. In many cases, non-verbal components constitute borrowings from other languages, most commonly from Arabic or Tajik, but also increasingly from Russian and even English. Light verbs, for their part,

<sup>5</sup>Older publications, including Karamshoev’s (1988a) dictionary, list plural perfect forms which are distinct from both feminine and masculine singular forms. These forms, as a rule, have the vowel of the feminine singular past stem but the consonant cluster of the masculine singular perfect stem. For more on these historical forms, see Section 9.3.2.1.



may be either transitive or intransitive (except for *čīdow* ‘do’, which apparently has both transitive and intransitive variants, as shown above) and typically come from a pool of five or six commonly used verbs in the language. Table 4.19 shows the most common light verbs in Shughni along with their present stems, past stems, and examples of compound verbs on which they are built.

**Table 4.19:** Common light verbs used in complex verbs.

LIGHT VERB	LIT. MEANING	PRS STEM	PST STEM	EXAMPLE	GLOSS
<i>čīdow</i>	‘do’	<i>kin-</i>	<i>čūd</i>	<i>kor čīdow</i> <i>islo čīdow</i>	‘to work’ ‘to correct’
<i>ōēdow</i>	‘hit’	<i>ōāō-</i>	<i>ōod</i>	<i>gāp ōēdow</i> <i>kud ōēdow</i>	‘to speak’ ‘to scold’
<i>sittow</i>	‘become’	<i>sāw-</i>	<i>sut (M)/</i> <i>sat (F)</i>	<i>kirx sittow</i> <i>ošiq sittow</i>	‘to slip’ ‘to fall in love’
<i>ōēdow</i>	‘fall’	<i>ōi-</i>	<i>ōod</i>	<i>xoŋ ōēdow</i> <i>patrā ōēdow</i>	‘to fear’ ‘to pout’

This aspect of Shughni grammar is similar to other languages of the area, which also employ extensive use of complex verbs, including both Indo-Aryan languages such as Hindi and Iranian languages such as Pashto and Persian. In fact, many complex verbs in Shughni have identical or near-identical counterparts in Persian. Compare, for instance, the Shughni complex verb *gāp ōēdow* ‘speak’ with Tajik Persian *gap zadan* – *gap zadan* ‘speak’, where both the Shughni verb *ōēdow* and the Tajik verb *zadan* mean ‘hit’ when appearing as stand-alone verbs. In the past, this seems to have made the borrowing of verbs from other languages into Shughni rather straightforward. More recently, Russian infinitives have come into the language as the non-verbal components of complex verbs (e.g. *motivirovat’ čīdow* ‘to motivate’, which includes the Russian infinitive *мотивировать* ‘to motivate’). The reader is directed to Section 11.3 for more on complex verbs.

#### 4.4.5 Passive-like constructions and causatives

Shughni does not have a dedicated applicative construction, but the language possesses both passive-like constructions built on resultative participles and dedicated causative morphology. This subsection offers a brief overview

of each (passive-like constructions in Section 4.4.5.1 and causatives in Section 4.4.5.2). A thorough presentation of passive-like constructions and causatives in Shughni can be found in Sections 11.1 and 11.2, respectively.

#### 4.4.5.1 Resultative participial (passive-like) constructions

Shughni uses two types of resultative participles: a verbal resultative participle (VRP) and an adjectival resultative participle (ARP), each used in different types of constructions. Both types of participle are formed via a verb's perfect stem together with a suffix: *-ak* in the case of the VRP (e.g. *mīzj* 'built.PRF > *mīzj-ak* 'build.PRF-V.PTPL') and *-in* in the case of the stative passive (e.g. *mīzj-in* 'build.PRF-A.PTPL').

As discussed in Section 4.4.2.2, resultative participles are only compatible with transitive and unaccusative verbs, to the exclusion of unergatives. Note, moreover, that whereas ARPs with *-in* are capable of inflecting for gender (in cases where an unaccusative verb has gender-distinguishing perfect stems), VRPs never inflect for gender and appear invariably in a form equivalent to the masculine. Thus, whereas the unaccusative verb *naǰtīdow* 'go out; exit' has two possible ARPs – *naǰtīc-in* (F/PL) and *naǰtūyǰ-in* (M) – the same verb has only a single VRP *naǰtūyǰ-ak*.

VRPs are used in passive-like constructions formed with the auxiliary verb *sittow* 'become', of which an initial example is given in (30).

#### (30) Verbal resultative participial construction: Initial example

Yam            čīd    **mīzj-ak**            ca    sut,            māš=ām fukaθ ar-ūd    yat.  
 DEM.PROX.DIR house build.PRF-DYN.PASS SUBR become.PST.M we=1PL all    to-here come.PST  
 'When this house was built, we all came here.'

Constructions of the kind exhibited in (30) bear a number of similarities to passives. In particular, they are intransitive constructions with a patientive subject, are often formed via a transitive verb, and bear special morphology which distinguish them from canonical transitive and unaccusative active constructions. However, as discussed in detail in Section 11.1, these constructions differ from canonical passives in important ways. First, an agent is often not implied and almost never expressed in such constructions. And moreover, they do not always promote a semantic patient to grammatical subject, as verbal resultative participial constructions with unaccusative verbs have the same subject as in their "active" counterpart constructions.

Adjectival passive participles built on the suffix *-in* behave syntactically as adjectives and can be used either attributively or predicatively. Initial examples is given in (31).

(31) **Adjectival resultative participle in *-in*: Initial examples**

a. **Attributive ARP**

Wev      mīzj-in      čīd-en=ta      mam      sol-and      parđen.  
 DEM.OBL.PL build.PRF-A.PTPL house-PL=FAC DEM.OBL.F year-LOC sell.PRS-3PL  
 ‘They will sell the houses which have been built (lit. the built houses) this year.’

b. **Predicative ARP**

Wāð      čīd-en=en      ūži      mīzj-in      vad      o?  
 DEM.DIR.PL house-PL=3PL already build.PRF-STAT.PTCPL be.PST.PL PQ  
 ‘Were those houses already built?’

#### 4.4.5.2 Causative

Causatives in Shughni may be either morphological or syntactic. As discussed in Section 4.4.2, morphological causatives occur almost exclusively with unaccusative verbs. They can be subdivided into two subtypes: (i) those formed by *ablaut* and (ii) those formed by *suffixation*. The former type generally involves an intransitive present stem with a stem vowel *ā* and a causative counterpart with a stem vowel in *ê* (e.g. *fīrāp-* ‘arrive’ > *fīrêp-* ‘deliver’). This correspondence has been traced to an umlaut process of the stem vowel caused by the once productive Iranian causative suffix *\*aya* (e.g. Dodykhudoeva 1988).

Suffixal causatives, for their part, generally involve a suffix *-en* or *-in* attached directly to the intransitive stem (e.g. *šand-* ‘laugh’ > *šand-en-* ‘make laugh’). This type of suffix is relatively new and is thought to be a borrowing from Tajik, which has more productive suffixal causatives of this type (e.g. Sokolova 1967). Examples of both ablaut and suffixal causatives are presented in Table 4.20.

Syntactic causatives are either built on the verb *majbūr čīdow* ‘to force’, which takes either a finite or non-finite clausal complement, or on the locative suffix *-ti*. The suffix *-ti* attaches to the causee and no additional morphological indications of causation appear in the clause. Examples of each type are given in (32) and (33), respectively.

**Table 4.20:** Intransitive~causative pairs.

PRES. STEM (INTR.~CAUS.)	GLOSS (INTR.~CAUS.)	INTRANSITIVE INFINITIVE	TRANSITIVE INFINITIVE
ABLAUT CAUSATIVES			
<i>firāp-</i> ~ <i>firēp-</i>	arrive ~ deliver	<i>firīptow</i>	<i>firēptow</i>
<i>θāw</i> ~ <i>θēw-</i>	burn (INTR.) ~ burn (TR.)	<i>θīdow</i>	<i>θēwdow</i>
<i>niθ-</i> ~ <i>nēđ-</i>	sit ~ set; plant	<i>nīstow</i>	<i>nēđdow</i>
SUFFIXAL CAUSATIVES			
<i>wary-</i> ~ <i>warven-</i>	boil ~ make boil	<i>wīrvdow</i>	<i>warventow</i>
<i>fām-</i> ~ <i>fāmūn-</i>	know ~ explain	<i>fāmtow</i>	<i>fāmūntow</i>
<i>šānd-</i> ~ <i>šāndūn-</i>	laugh ~ make laugh	<i>šīntow</i>	<i>šāndūntow</i>

**(32) Syntactic causative with *majbūr čīdow* ‘force’**

Wuz=um wi      **majbūr čūd**      [awqot xīr-t].  
 I=1SG    3SG.OBL.M force    do.PST food    eat.PRS-3SG  
 ‘I forced him to eat.’

**(33) *ti*-causative construction with *žēxtow* ‘run’**

Yā      mollimā=yi    māš-**ti**    dis      žēxt.  
 DEM.DIR.F teacher.F=3SG US-LOC SO.much run.PST  
 ‘That teacher made us run so much.’

There are a number of semantic peculiarities associated with morphological causatives, on the one hand, and syntactic causatives on the other. Moreover, syntactic causatives of the *majbūr čīdow* ‘force’ type display important semantic differences from those built on the locative suffix *-ti*. These semantic differences, in addition to the formation of each type of causative, are also addressed in Section 11.2.

## 4.5 Complex clauses and question formation

This final section of the Grammatical Overview chapter presents complex clauses (Section 4.5.1) and question formation (Section 4.5.2). Both of these issues, as well as information structure and morphosyntactic alignment,

are addressed further in Chapter 12.

### 4.5.1 Complex clauses

In Shughni complex clauses, the subordinate clause often, but not always, contains an overt subordinating element. In complement clauses, this is typically the complementizer *idi*, which appears at the left edge of the subordinate clause. In relative clauses and temporal adverbial clauses, the subordinator *ca* is used, which targets the left edge of the finite verb in the subordinate clause. If an over finite verb is not present, as in the case of present-tense copular constructions, *ca* appears at the right edge of the subordinate clause.

An example of a **complement clause** with the complementizer *idi* is given in (34). This example contains the verb *lūvdow*, which takes a clausal complement. Note that the example here is grammatical with or without the complementizer *idi*. Indeed, this complementizer is often optional. However, in the case of dependent speech, the use of the complementizer *idi* forces an indirect-speech reading. Section 12.1.4 looks at the use of this complementizer with subordinate clauses containing dependent speech, as well as dependent speech in general.

(34) **Complement clause with *idi***

Daler=i lūd [(idi) tu=ta na-yad-i]<sub>cc</sub>.  
 Daler=3SG say.PST COMP you=FAC NEG-come.PRS-2SG  
 ‘Daler said (that) you aren’t coming.’

There are three types of **relative clauses** found in Shughni: (i) **externally headed**; (ii) **internally headed** (specifically, a subtype of internally headed relative clause known as a *correlative clause*), and (iii) **free (headless) relatives** (specifically, *free-choice* free relatives). Initial examples of each type are given in (335). The head noun in each example is bolded, where applicable, and square brackets are placed around the relative clause.

(35) **Shughni relative clauses: Initial examples**

a. **Externally headed relative clause**

Ik=ā **ŷinik<sub>i</sub>** [biyor=at wam<sub>i</sub> ca wīnt]<sub>rc</sub> yā<sub>i</sub> mis ikūdand kor kiṣt.  
 PREC=DEM.DIR.F woman yesterday=2SG her REL see.PST she also here work do.PRS.3SG  
 ‘The woman you saw yesterday also works here.’

b. **Internally headed (correlative) relative clause**

[Ik=wam      **ŷinik**<sub>i</sub>=at    biyor    ca    wīnt]<sub>RC</sub>    yā<sub>i</sub>    mis    ikūdand    kor    kiŷt.  
 PREC=DEM.OBL.F woman=2SG yesterday REL see.PST she also here    work do.PRS.3SG  
 ‘The woman you saw yesterday also works here.’

c. **Free-choice free relative clause**

[Arčīz    wev    tar    ḏust    ca    ḏod]<sub>RC</sub>    zoŷt=en.  
 whatever them into hand REL fall.PST take.PST=3SG  
 ‘They took whatever fell into their hands.’

(Karamshoev 1988a: 135)

Like relative clauses, finite **temporal adverbial clauses** generally contain the subordinator *ca*, although in some instances, specifically in future temporal adverbials, they may contain a distinct subordinator *di*. Non-finite temporal adverbial clauses are also used. In these cases, the subordinate clause contains an infinitive stem together with the converb suffix *-ow*. Examples of finite and non-finite temporal adverbial clauses are given in (36) and (37), respectively.

(36) **Finite temporal adverbial clause**

[Tu=t ar-ūd    ca    yat]<sub>TAC</sub>    wuz=um    tar    magazīn    rawūn    vud.  
 you to-here SUBR come.PST I=1SG to store going be.PST.M  
 ‘When you came by here, I was on my way to the store.’

(37) **Non-finite temporal adverbial clause**

Wuz=ta [piro    az    xêvd-ow]<sub>TAC</sub>    kitob    ŷoy-um.  
 I=FAC before from sleep.INF-PURP book read.PRS-1SG  
 ‘I read a book before sleeping.’

Relative clauses, temporal adverbial clauses, and other types of subordinate clauses (such as purposive, causal, and concessive clauses) are discussed in the first three subsections of Section 12.1.

## 4.5.2 Question formation

Shughni is a *wh-in-situ* language. In neutral word order, the *wh*-element appears in the same position in the clause as it would in a statement. This is exhibited in the examples of *wh*-questions in (38).

(38) **Neutral word order in *wh*-questions**

- a. Čāy biyor tar kor sut?  
 who.DIR yesterday to work go.PST.M  
 ‘Who went to work yesterday?’
- b. Tu=t cawaxt tar kor sut?  
 you=2SG when to work go.PST.M  
 ‘When did you go to work?’
- c. Tu=t biyor tar kā sut?  
 you=2SG yesterday to where go.PST.M  
 ‘Where did you go yesterday?’

Non-canonical word order in *wh*-questions does occasionally occur, however. It is generally connected to the expression of topic and focus. As in statements, certain topicalized elements may appear in clause-initial position in questions. In questions, fronted elements are most often contrastive topics – i.e. those which are picked out of a small subset of alternatives known to both speaker and addressee. In certain instances of focus – particularly when the addressee has not heard the question properly – the *wh*-word itself may be fronted. These issues are expanded upon in Section 12.3.1.

**Polar questions** in Shughni are formed with the polar-question particle (y)o, which appears clause-finally. Examples of polar questions are given in (39)

(39) **Polar-question particle *o***

- a. Awqot=at tayor čūd o?  
 food=2SG prepare do.PST PQ  
 ‘Did you finish making the food?’
- b. Wam qawmiyot mis yamand zindagi-yen o?  
 her relatives also there life-3PL PQ  
 ‘Do her relatives live there too?’

Other types of questions, such as tag questions and echo questions, are also formed with the use of clause-final particles. These types of questions, in addition to basic *wh*-questions and polar questions, are investigated in Section 12.3.





## **Part II**

# **Nominals**



## Chapter 5

# Noun phrases, nominal inflection, and nominal derivation

This chapter is the first in a series of three on Shughni nominals. Here, we look at a number of fundamental aspects of Shughni nouns and noun phrases, including the content of noun phrases, issues in inflection and derivation, and numerals.

The chapter is organized as follows. First, Section 5.1 looks at the **content and order of elements in noun phrases**, as well as the **reference of noun phrases** (generic and (in)definite) and the shape (i.e. content) of noun phrase compatible with each type of reference. Section 5.2 turns to **nominal inflection**, specifically the expression of number in nominals. Here, I examine four types of plural: the additive plural (i.e. the most basic type of plural), associative plural, similitive plural, and collective plural. Each type of plural has peculiarities in its formation and usage. Note that grammatical gender, another major component of nominal inflection in Shughni, is treated in full in Chapter 7, which is the third and final chapter in the series on nominals. Section 5.3 examines issues in **nominal derivation**, looking specifically at the formation of nouns and adjectives through affixation and compounding. And finally, Section 5.4 presents **numerals**, a topic made more complex by the fact that the language has borrowed Tajik numerals for use alongside native numerals. Importantly, however, the choice of whether a Shughni or Tajik numeral is used is not random, but is systematic and dependent upon grammatical context.

## 5.1 Content and reference of noun phrases

This section examines the content and order of elements in Shughni noun phrases, as well as the expression of different kinds of reference – generic, indefinite (specific and non-specific), and definite.

In the discussion here, I assume the interpretation of definiteness developed by Lyons (1999). In particular, I take definiteness to be a property of a noun phrase which comes about due to its *identifiability*. That is, if a noun phrase is definite, the speaker expects that the addressee will be able to readily identify its referent. The notion of *uniqueness* may also factor into definiteness, particularly with respect to entities which are globally unique, including the sun and moon, or unique within a given speech context.

Indefinite noun phrases, on the other hand, are those whose referent the speaker does not expect to be readily identifiable by the addressee. Indefinite noun phrases may be either specific, where the speaker has a unique referent in mind, or non-specific, where there is no unique referent. In many cases, an indefinite noun phrase is used to introduce a new entity into the discourse, after which it is referred to with a definite noun phrase.

And lastly, noun phrases with generic reference, which may be considered a subtype of indefiniteness, make generalizations regarding a class of entities. Noun phrases with generic reference differ from those with a universal quantifier such as *all* (Shughni *fuk*), in that the former may allow for exceptions to the generalization they make, whereas the use of a universal quantifier does not allow for exceptions.

The remainder of this subsection is organized as follows. The discussion begins in Section 5.1.1 with an introduction to the possible sizes, shapes, and content of noun phrases. Then, the following three subsections examine each type of reference in turn. Generic reference, expressed through the use of bare nouns or nouns with a plural suffix, is discussed in Section 5.1.2. Section 5.1.3 then looks at the expression of indefiniteness through the indefinite marker *yi* and other indefinite quantifiers. And finally, Section 5.1.4 examines different types of definiteness and their grammatical expression – in particular, whether each type of definiteness allows the use of a bare noun or requires the use of a demonstrative.

### 5.1.1 Content, size, and order of elements in noun phrases

Noun phrases in Shughni are head-final and consist minimally of a bare noun. They may also contain a demonstrative, possessive adjective, quantifier, numeral, one or more adjectives, and plural or locative suffixes which appear directly on the noun itself. The order of these elements is schematized in (40).

(40) **Order of elements in noun phrases**

[ (DEM) (POSS.ADJ) (NUM/QUANT) (ADJ) **Noun**(-PL)(-CASE) ]<sub>NP</sub>

Maximally, a noun phrase may contain a noun plus six additional elements: a (plural) demonstrative, possessive adjective, numeral, adjective, plural suffix, and case-like suffix, in that order. Minimal and maximal noun phrases in Shughni are schematized, together with examples, in (41).

(41) **Minimal and maximal noun phrases in Shughni**

a. **Minimal NP**

Schema: [ **Noun** ]<sub>NP</sub>

Example: [ *mūn* ]<sub>NP</sub>

‘apple(s)’

b. **Maximal NP**

Schema: [ DEM + POSS.ADJ + NUM + ADJ + **Noun**-PL.SFX-CASE.SFX ]<sub>NP</sub>

Example: [ *wev mu cavor yullā mūn-en-ard* ]<sub>NP</sub>

‘for those four big apples of mine’

In some cases where the referent of a noun is understood from context, Shughni also allows for noun phrases without an overt head noun. Such noun phrases may consist of any combination of a demonstrative, numeral/quantifier, and adjective. They may also include a plural suffix which attaches to the adjective. This is schematized in (42a), and an example is given in (42b).

## (42) Noun phrases with no overt head noun

- a. [ (DEM) (QUANT/NUM) (ADJ) **Noun-PL** ]<sub>NP</sub>
- b. [wev        **cavor yullā-yen**]<sub>NP</sub> mu-rd    dāk  
 DEM.OBL.PL four    big-PL        me-DAT give.2SG.IMP  
 ‘give me those four big ones (e.g. onions)’

A couple permissions on the content of noun phrases are worthy of note. First, oblique personal pronouns used as possessive adjectives are compatible with demonstratives, as in the examples in (43).

## (43) Possessive adjective used with demonstratives

- a. [Yā        **mu rizīn**]<sub>NP</sub> kačād?  
 DEM.DIR.F my daughter where  
 ‘Where is that daughter of mine (lit. that my daughter)?’
- b. [Yid        **tu kurtā**]<sub>NP</sub> dis xušrūy!  
 DEM.DIR.MED your shirt        so beautiful  
 ‘Your shirt (lit. that your shirt) is so beautiful!’

And second, possessive adjectives may also be used with the singular indefinite marker *yī*, as shown in (44), a construction which results in a partitive-like construction meaning ‘one of my N’. The same semantic idea, i.e. reference to one member of a defined group, can be expressed through a construction with the full numeral *yīw* and a prepositional phrase with *az* ‘from; of’. Note that in the latter type of construction, an example of which is given in (45), the use of the locative suffix *-and* on the noun denoting the group is required. This type of construction is more common with numerals greater than one.<sup>1</sup>

(44) Possessive adjective with indefinite (singular) *yī*

- [**Mu yī oḫno**]<sub>NP</sub>=ta arūd    yoḏd.  
 my a friend=FAC to.here come.PRS.3SG  
 ‘A friend of mine is going to come here.’

<sup>1</sup>The (second-position) factual enclitic =ta (glossed FAC) in these examples is commonly found together with present stems. The semantic contribution of =ta is quite varied, but for now it can be said that this enclitic encodes future temporal reference and/or habitual aspect. It is examined in more detail in Section 10.2.1.2 of Chapter 10 on tense, aspect, and mood.

(45) **Genitive alternative to (44)**

[Yīw az mu oḵno-yen-and]<sub>NP</sub>=ta arūd yoḏd.  
 one of my friend-PL-LOC=FAC to.here come.PRS.3SG  
 ‘One of my friends is going to come here.’

5.1.2 **Generic reference**

A noun phrase with generic reference makes a generalization regarding a class of entities (cf. Lyons 1999: 179). In Shughni, generic reference is expressed in two ways: (i) through bare nouns, such as *kud* ‘dog(s)’, and (ii) with nouns containing a plural suffix, as in *kud-en* ‘dogs’. Both types of generic-reference NP may contain adjectives, and hence *safed kud* ‘white dog(s)’ and *safed kud-en* ‘white dogs’ may also be used with generic reference.

Examples of noun phrases with generic reference are given in (46). In these examples, square brackets have been placed around each noun phrase with generic reference. In the first two examples, the noun phrase with generic reference consists of a single noun, while in (46c), it consists of an adjective and a noun. Note that a context has been provided for the sentence in (46a) so that it may be contrasted with the sentence in (47), which contains an indefinite marker and is felicitous only in a different type of context. Note further that the plural suffix on the count noun *kud* ‘dog’ in (46a) is optional; this sentence would be equally grammatical if *kud* appeared as a bare noun.

(46) **Noun phrases with generic reference**

- a. **Context:** A woman with a wound on her arm is walking her dog down the street. A passerby sees the pair and remarks to his friend that he thinks the dog bit the woman. His friend says the following in response:

[Kud-en]=ta xu soyib-en na-piren-en.  
 dog-PL=FAC REFL OWNER-PL NEG-bite.PRS-3PL  
 ‘Dogs don’t bite their owners.’

- b. [Ḫūvd] [sitxūn]-ard dis foydā.  
 milk bone-DAT so useful  
 ‘Milk is very good for bones.’

- c. [Gandā odam]=ta der mīrt.  
 bad person=FAC late die.PRS.3SG  
 ‘Only the good die young (lit. bad people die late).’

Importantly, generic reference, such as that of each bracketed noun in the sentences in (46) cannot be expressed with the indefinite article *yi*. Thus, although the sentence in (47) is grammatical, it is not felicitous in the context

given for the sentence in (46a) above. Rather, it would be felicitous, for instance, in a context in which two participants at an animal obedience training are talking, and one participant says to another that all dogs at this kind of training bite their owners. The other participant, who believes she has a very good dog, takes exception to this remark and responds with the sentence in (47) that one (specific) dog won't bite its owner.

(47) **With article *yi*: Only indefinite specific**

[**Yi** **kud**]=ta xu soyib na-piren-t.

ART.SG dog=FAC REFL OWNER NEG-bite.PRS-3SG

Possible: 'One dog doesn't bite its owner. (I have a very good dog.)'

Impossible: 'A dog doesn't bite its owner. (This is a general property of the dog species.)'

### 5.1.3 Indefinite reference

Shughni lacks a dedicated indefinite article compatible with both singular and plural referents. Two fundamental ways of showing indefiniteness in the language are through the article *yi*, a truncated version of the numeral *yīw* 'one', and through the quantifier *yičand* 'some; several', which is used with plural nouns. Singular *yi* and plural *yičand* are first examined in Section 5.1.3.1, and then Section 5.1.3.2 looks at specific and non-specific indefinite reference.

#### 5.1.3.1 Singular and plural indefinite reference

The article *yi* is perhaps the closest approximation Shughni has to an indefinite article, but like English *a(n)*, Shughni *yi* encodes not only indefiniteness, but also singular number. Because *yi* is only used with singular nouns, I analyze it here as a *cardinal article* (glossed ART.SG) similar to English *a(n)* (cf. Lyons 1999 for a discussion on this term and the distinction between cardinal articles and pure indefinite articles).

Initial examples illustrating that this morpheme is incompatible with plural nouns are given in (48). The sentence in (48a) shows the incompatibility of *yi* with the plural suffix *-en*. Nonetheless, because I have analyzed *-en* and other additive plural suffixes as encoding either definiteness or general reference, it might be expected that an indefinite article be unable to appear in the same noun phrase. The example in (48b) shows that even when a plural suffix is not present, a plural reading of the noun phrase with *yi* is still unavailable.



(48) **Article *yī* encodes singular number**

- a. Māš=ām pis pūnd-aθ [yī yullā kū(\*-yen)] wīnt.  
 we=1PL on road-ADV ART.SG big mountain(\*-PL) see.PST  
 ‘We saw a big mountain(\*-s) on the road.’
- b. Māš=ām pis pūnd-aθ [yī yullā kū] wīnt. (#Wāð=en dis xušrūy vad.)  
 we=1PL on road-ADV ART.SG big mountain see.PST (they=3PL so pretty be.PST.PL)  
 ‘We saw a big mountain on the road. (#They were so pretty.)’

Moreover, the cardinal article *yī* exhibits important syntactic differences from the full numeral *yīw* ‘one’. Specifically, the cardinal article *yī*, rather than the numeral *yīw*, is required when preceding either the noun it modifies or another modifier within the same noun phrase. The numeral *yīw*, by contrast, is used when no overt noun follows, as in contexts in which the noun is elided. This is exhibited in the examples in (49) and (50). In the former, only the cardinal article *yī* is available, as it is preceded by additional material within the noun phrase, namely the adjective *rūšt* ‘red’ and the noun *mālā* ‘house’. In (50b), which is the response to the question in (50a), only the full numeral *yīw* is available, as it is the only element in its phrase.

(49) **Complementary distribution of cardinal article *yī* and numeral *yīw* ‘one’**

- Tu=ta az [yī (\*yīw) rūšt mālā]<sub>NP</sub> načjīs-i.  
 you=FAC from ART.SG (\*one) red house pass.PRS-2SG  
 ‘You’ll pass a red house.’

(50) a. **Question preceding the response in (50b)**

- Tama-rd cūnd kilo yočj darkor?  
 you.PL-DAT how.many kilogram flour necessary  
 ‘How many kilos of flour do you need?’

- b. **Yīw (\*yī)**  
 one (\*ART.SG)  
 ‘One.’

The numeral *yīw* is not the only numeral to have a corresponding cardinal article, however. The numerals *duyūn* ‘two’ and *aray* ‘three’ have shortened forms *du* and *ara*, respectively, which exhibit identical syntactic behavior as the singular cardinal article *yī*. Numerals beyond *aray* ‘three’ do not have shortened forms. This pattern is shown in Table 5.1. Note that, as discussed in Section 5.4 on numerals, Shughni speakers use Tajik for numerals greater than ten.

**Table 5.1:** Cardinal articles: Shortened forms of Shughni numerals.

<u>Full numeral</u>	<u>Cardinal article</u>	<u>Gloss</u>
<i>yīw</i>	<i>yi</i>	one
<i>ḍuyūn</i>	<i>ḍu</i>	two
<i>aray</i>	<i>ara</i>	three
<i>cavor</i>	—	four
<i>pīndz</i>	—	five
...	—	...

Indefinite plural noun phrases are often expressed using the quantifier *yičand* ‘some; several’. This quantifier behaves similarly to numerals with respect to its compatibility with additive plural suffixes. Specifically, when no demonstrative is present, *yičand* cannot appear with a plural suffix, as illustrated in (51a). However, if a demonstrative is present in the same noun phrase as *yičand*, an additive plural suffix is obligatory, as shown in (51b). This quantifier is therefore not inherently indefinite, like English *some*, but can be used in both definite and indefinite noun phrases.

(51) **Quantifier *yičand* ‘some; several’**

- a. Stūdent-en=en [*yičand* kitob(\*-en)]<sub>NP</sub> xarīd čūd.  
 student-PL=3PL some book(\*-PL) purchase do.PST  
 ‘The students bought some books.’
- b. Stūdent-en=en [wev *yičand* kitob\*(-en)]<sub>NP</sub> xarīd čūd.  
 student-PL=3PL DEM.OBL.PL several book\*(-PL) purchase do.PST  
 ‘The students bought those several books.’

**5.1.3.2 Specific and non-specific indefinite reference**

The cardinal article *yi* is compatible with both specific and non-specific indefinite reference. Examples in which *yi* has specific indefinite reference are given in (52), and examples in which it has non-specific indefinite reference are given in (53).

(52) **Specific indefinite reference with the cardinal article *yi***

- a. [*Yi* *divūsk*]=i wi čorik žiroxt.  
 ART.SG snake=3SG DEM.OBL.M man bite.PST  
 ‘A snake bit that man.’

(Parker 2020: 17)

- b. Mu virod=ta [yi Pomīri yin] vīrt. Wuz=um ūži wam qati balad.  
 my brother=FAC ART.SG Pamiri woman bring.PRS.3SG I=1SG already her with acquainted  
 ‘My brother is getting married to a Pamiri woman. I already know her.’

(53) **Non-specific indefinite reference with the cardinal article *yi***

**Context for (53a):** Mawlodod’s parents were out of town last week, so they asked the administrators at his school if someone could take him to school each day. There are multiple teachers at Mawlodod’s school, and one of them (but a different one each day) took him to school each day.

- a. [Yi mollim]=i ar meθ Mawlodod tar maktab yod.  
 ART.SG teacher=3SG every day Mawlodod to school take.PST  
 ‘A teacher took Mawlodod to school everyday.’ (Parker 2020: 17)
- b. [Yi padūška] xu-rd zi!  
 ART.SG pillow REFL-DAT take.2SG.IMP  
 ‘Grab a pillow for yourself.’

Indefinite non-specific reference can be emphasized with the use of free-choice indefinite determiners such as *yičīz*, as in (54), or with the use of free-choice free relatives, which are built on special *wh*-words such as *arčidūm* ‘whichever’, as in (54b).

(54) **Free-choice indefinite determiners *yičīz* and *arčidūm***

- a. Ded-ow-ard tu boyad [yičīz kalīmā] lu.  
 enter-NMZ-DAT you must any word say.2SG.IMP  
 ‘In order to enter, you must say any word.’
- b. [Arčidūm (kitob)] tu-rd ca fort, zi.  
 any book you-DAT REL want.3SG.PRS take.2SG.IMP  
 ‘Take any (book) you want.’

#### 5.1.4 Definite reference

Shughni lacks a dedicated definite article. Definite noun phrases often have a demonstrative, but in many cases bare nouns with definite interpretation are also acceptable. This subsection examines several types of definiteness in Shughni and the types of noun phrases with which each type is compatible. Definiteness may come about due to a variety of factors and can be broken down into multiple types. Here, I describe four types of definiteness outlined by Lyons (1999: 4-12): (i) anaphoric definiteness, (ii) bridging cross-reference (or associative) definiteness, (iii)

modificational definiteness, and (iv) situational definiteness. Special attention will be paid to whether noun phrases with each type of definiteness require the use of a demonstrative, or whether a demonstrative is merely optional.

In **anaphoric definiteness**, the referent is identifiable for the addressee because it has been previously mentioned in the discourse. In Shughni, this is the type of definiteness with the most stringent restrictions on bare nouns. Noun phrases with anaphoric definiteness invariably require a demonstrative, as shown in the example in (55). Note that in (55b), the referents of the definite noun phrases in the second clause are introduced with the (indefinite) singular article *yi* in the first clause. Once they are mentioned the second time, they have definite reference.

(55) **Anaphoric definiteness: Demonstrative required**

- a. Tama=ta taram mošin qati ca set, [\***(yā)** **mošin**] boyad zūr vid.  
 you.PL=FAC to.there car with SUBR go.PRS.2PL DEM.DIR.F car must good be.PRS.3SG  
 ‘If you are going to go there with a car, **the car** must be high quality.’
- b. Māš=ām biyor yi čorik=at yi ŷinik qati balad sat. [\***(Yu)** **čorik**]  
 we=1PL yesterday ART.SG man=and ART.SG woman with familiar become.PST.PL DEM.DIR.M man  
 az Maryov vud=at [\***(yā)** **ŷinik**] az Iškošim vad.  
 from Murghob be.PST.M=and DEM.DIR.F woman from Ishkashim be.PST.F  
 ‘We met a man and a woman yesterday. **The man** was from Murghob, and **the woman** was from Ishkashim.’

**Bridging cross-reference definiteness**, also known as **associative definiteness**, comes about not because the referent of the noun phrase has been previously mentioned, as in anaphoric definiteness. Rather, this type of definiteness occurs because another referent or context has been introduced in the preceding discourse through which the existence and uniqueness of the newly introduced referent can be inferred. In other words, the previously unmentioned referent gains its definiteness *via association with* an aforementioned referent or context. For instance, if the speaker mentions that someone has bought a car, then the existence of a steering wheel in the same car can be inferred, and the steering wheel can thus be expressed in a definite noun phrase.

In Shughni, noun phrases which are definite due to bridging cross-reference also generally require a demonstrative. However, there are some instances of associative definiteness in which the use of a demonstrative may not be required, and a bare noun phrase may be used with definite reference. The pattern is roughly as follows. If the referent of the noun phrase is definite by association with a previously mentioned noun phrase of which it is an integral part, especially a physical part of a whole, then a demonstrative functioning as a possessive adjective is

typically required. This is the case, for instance, in the example given above of a steering wheel gaining definite reference due to its association with a previously mentioned car. Such an example is given in (56a).

On the other hand, if the newly introduced referent is more loosely affiliated with the referent of a previously mentioned noun phrase, particularly if it is not a physical part of the referent, then a bare noun is more acceptable, although a demonstrative may optionally be used. An example of this kind, involving the relation between an airplane and a previously mentioned journey, which is understood by both speaker and addressee to generally involve an airplane, is shown in (56b).

(56) **Bridging cross-reference definiteness**

a. **Bridging cross-reference definiteness: Demonstrative required**

Mu tāṭ=i      biyor      yi      mošīn zoḫt=atā      [\*(wam)      rūl]      viruḫčīn vud.  
 my father=3SG yesterday ART.SG car      take.PST=but DEM.OBL.F steering.wheel broken.M be.PST  
 ‘My father bought a car yesterday, but **the (/its) steering wheel** was broken.’

b. **Bridging cross-reference definiteness: Demonstrative optional**

Biyor=en      az      Dūšanbi tar Istanbūl či      tīd      vad=atā      [(yā)      samalyot] az  
 yesterday=3SG from Dushanbe to Istanbul PROSP go.INF be.PST.PL=but DEM.DIR.M plane      from  
 Almaty der firīpt.  
 Almaty late arrive.PST  
 ‘They were going to go from Dushanbe to Istanbul yesterday, but **the plane** from Almaty arrived late.’

Definiteness may also come about due to the presence of **modifiers which pick out unique referents**, such as *the only*, *the first*, and uses of the superlative such as *the best*. In Shughni, noun phrases which are definite due to the presence of modifiers of this type generally do not require demonstratives. Their definiteness is encoded in the semantics of the modifiers themselves. Such examples are shown in (57).

(57) **Modifiers which pick out unique referents**

a. **dūnd – ‘(the) only’**

Pi      Xaray      mu tāṭ      [(\*yu)      dūnd kanādec].  
 up.in Khorog my father DEM.DIR.M only Canadian  
 ‘My dad is the only Canadian in Khorog.’

b. *yakum* – ‘(the) first’

John [(\*)yu] **yakum kanadec** tar māš ūniversitet dars ca ðīd.  
 John DEM.DIR.M first Canadian in our university class REL give.PRS.3SG  
 ‘John is **the first Canadian** to teach at our university.’

c. *sar bašānd* – ‘(the) best’

Ominā [(\*)yā] **sar bašānd talabā** māš maktab-and.  
 Omina DEM.DIR.F SUP good student our school-LOC  
 ‘Omina is **the best student** in our school.’

Noun phrases meaning *the next* and *the previous* constitute an exception to this generalization in that they require a demonstrative. This is likely due to the fact that they are interpreted deictically. Examples are shown in (58); note that in each example, a lack of the demonstrative does not result in ungrammaticality, but rather in a difference in meaning. In the (58a), a lack of the demonstrative *yā* would result in a reading in which another exam, but not necessarily the (very) next exam, will occur in March. In (58b), a lack of the demonstrative would result in a reading in which the group was in the Pamirs during another week in the past, but not necessarily the week immediately preceding the week mentioned in the first sentence.

(58) *next and previous* – demonstrative required

- a. [Yā **yiga ekzamin**]=ta mart-and sūd.  
 3SG.DIR.F other exam=FAC March-LOC become.PRS.3SG  
 ‘**The next exam** will be in March.’
- b. Māš=ām ðu aftā tar piro pi Pomir vad. [Wam **yiga aftā**]=yām tar Khujand vad.  
 we=1PL two week in past up.to Pamir be.PST.PL DEM.OBL.F other week=1PL in Xujand be.PST.PL  
 ‘Two weeks ago we were in the Pamirs. **The previous week** we were in Khujand.’

The fourth and final type of definiteness to be discussed here is that of **situational definiteness**, in which the referent is identifiable due to the physical surroundings in which the speech participants find themselves. Situational definiteness can be graded based roughly on the notions of the uniqueness of the referent within the speech context and the inherent connection it has to that context. Lyons (1999: 4) conceptualizes this uniqueness and inherent connection with the term *general knowledge*.

In Shughni, the notion of general knowledge, as it pertains to situational definiteness, has concrete grammatical outcomes. In particular, if the referent of a given noun falls within the general knowledge held by speaker and

addressee regarding the types of entities which are found within their physical location, then it is more likely to be grammatical as a bare noun phrase. If it does not fall within this general knowledge, it is less likely to be referred to with a bare noun phrase and is more likely to be accompanied by a demonstrative.

The physical contexts in which general knowledge comes into play range from large to small. On one end of the spectrum are objects which have global uniqueness such as *the sun* and *the moon*. These are the objects which are most likely to be referred to with a bare noun. On a smaller scale, *the president* may be understood to be unique within a particular country, and on an even smaller scale, *the kitchen* may be understood to be unique if the speaker and addressee are located in a house. In Shughni, nouns such as these, when understood to be unique due to the context and form part of the general knowledge shared by speaker and addressee about that particular context, are generally allowed to appear as bare noun phrases, but are less likely to appear as such than more globally unique referents such as the sun and moon.

Examples of situational definiteness are given in (59). The sentences in (59a) and (59b) contain the globally unique referents *xīr* ‘sun’ and *mêst* ‘moon’. The sentences in (59c) and (59d) contain referents which constitute part of the general knowledge regarding smaller physical contexts. In the former, the noun *daryo* ‘river’ might be unique within a community such as Khorugh, Tajikistan. In the latter, *banya* ‘bathroom’ might be unique because the speaker and addressee are located within a house, and they share the general knowledge that houses contain bathrooms.

(59) **Bare nouns with unique referents due to general knowledge**

- a. [(?Yu)    **xīr**]=ta    bādi yi    soat pal kixt.  
          DEM.DIR.M sun=FAC after one hour rise do.PRS.3SG  
          ‘The sun will rise after an hour.’
- b. [(?Yā)    **mêst**] naxtīc.  
          DEM.DIR.F moon go.OUT.PRF.F  
          ‘The moon has come out.’
- c. Tu=t    čīz-ard    ar    [**daryo**] rawûn?  
          you=2SG what-DAT down.to river    going  
          ‘Why are you going to the river?’

- d. [Banya] kā-ndi?  
 bathroom where-LOC  
 ‘Where is the bathroom?’

On the other hand, the referent of a noun phrase may be unique and identifiable within a specific physical context, but not otherwise intimately linked to or unique within the space where the discourse takes place. Here, the noun phrase containing such a referent may be definite due to situational definiteness, but if it falls outside the general knowledge shared by speaker and addressee, it is less likely to be expressed as a bare noun and more likely to be accompanied by a determiner.

An example of this type of noun phrase is given in (60). This example pertains to a situation in which the speaker is instructing the addressee how to clean a room in a house in which they are both standing. The speaker tells the addressee to organize the pillows and clean the table, where the shelves and the table are identifiable due to the physical surroundings. However, because shelves and pillows do not fall into the shared general knowledge of the speaker and addressee regarding what is unique within a house, a demonstrative is more likely to be used.

(60) **Situational definiteness without general knowledge: Demonstrative likely**

Ku [(di) **polka**] zidi=yat [(dam) **bolaŋ**]=ga rost ki.  
 please DEM.OBL.M shelf wipe2SG.IMP=and DEM.OBL.F pillow=also straight do.2SG.IMP  
 ‘Please wipe off **the shelf** and straighten up **the pillow**.’

In sum, we have looked at four different types of definiteness in Shughni and seen that each type is distinct with respect to whether it allows a bare noun or requires a demonstrative. First, if a noun phrase is definite due to its referent being previously mentioned in the discourse (anaphoric definiteness), then a demonstrative is required. Second, if the referent of a noun phrase is definite due to its inherent association with the referent of a previously mentioned noun phrase (i.e. associative definiteness), or with a previously described context, then a demonstrative may or may not be required. A demonstrative is typically used when the newly mentioned referent is an integral, physical part of the previously mentioned referent with which it is associated, while a more loose association between the two may allow for a bare noun to be used. Third, if definiteness occurs due to the semantics of a modifying element within a noun phrase (i.e. modificational definiteness), then the use of a demonstrative is generally not required. An exception to this generalization are modifiers which mean *the next* or *the previous*, which require demonstratives due to their deictic semantics. And lastly, if a noun phrase is definite due to its



uniqueness within the particular physical context in which the utterance takes place (situational definiteness), a demonstrative may or may not be required. In general, if the uniqueness of the referent is part of the general knowledge shared by speaker and addressee about the physical context of the utterance (e.g. a bathroom in a house, a river in particular city, or the sun or moon), then a demonstrative is not required. If no such association exists within their shared general knowledge, then a demonstrative is required. This information is summarized in Table 5.2.

**Table 5.2:** Summary: Definite types and demonstrative use in Shughni.

<b>Anaphoric</b>	Referent previously mentioned	<b>Yes</b>
<b>Associative</b>	Referent associated with previously mentioned NP or previously described context	<b>Yes</b> (if integral, physical part of referent of associated NP), <b>No</b> (else)
<b>Modification</b>	Modifier is inherently picks out unique referent	<b>No</b> (with the exception of <i>the next / previous</i> )
<b>Situational</b>	Referent unique within physical context of utterance	<b>No</b> (less likely if referent associated with general knowledge about particular physical context of utterance, more likely if not)

## 5.2 Number in nominals

This section looks at four types of plural formation in Shughni: (i) the additive plural (Section 5.2.1), (ii) the associative plural (Section 5.2.2), the simulative plural (Section 5.2.3), and (iv) the collective plural (Section 5.2.4). Only the additive plural and one type of collective plural are formed via affixation, while the remaining types of plural are formed through other processes, including the addition of a demonstrative pronoun for the associative plural, reduplication for the simulative plural, and a shift in gender agreement for one type of the collective plural. Each process is examined in turn.

### 5.2.1 Additive plural

The additive plural, the most fundamental type of plural, denotes a homogenous set of entities (cf. Barulin 1980; Daniel and Moravcsik 2013). Outside of this section, I refer to the Shughni additive plural as simply ‘the plural suffix’, but in this section I use the label *additive* to distinguish it from the other types of plural discussed below.

The most commonly used and productive additive plural suffix in Shughni is *-(y)en*, which can attach to the vast majority of singular nouns (e.g. *yāc* ‘girl’ > *yāc-en* ‘girls’; *mālā* ‘dwelling’ > *mālā-yen* ‘dwellings’; etc.). Certain nouns, often those which refer to genealogical and family relations, may alternatively take a plural suffix other than *-en*. For instance, the plural of *xiyūn* ‘sister-in-law’ may be either *xiyūn-en* or *xiyūn-orĵ*, and the plural of *amak* ‘(paternal) uncle’ may be either *amak-en* or *amak-ūn*. The suffix *-jev*, for its part, attaches to certain nouns denoting periods of time to form plurals which also function as temporal adverbs. For example, from *sāraki* ‘morning’ we get *sāraki-jev* ‘in the mornings’, and from *zimistūn* ‘winter’ we get *zimistūn-jev* ‘in the winters’. Table 5.3 shows the most common additive plural suffixes.

**Table 5.3:** Additive plural suffixes in Shughni.

SUFFIX	TYPE	EXAMPLE	GLOSS
<i>-(y)en</i>	all-purpose	<i>daryo-yen</i>	‘rivers’
		<i>ūniversitet-en</i>	‘universities’
<i>-orĵ</i>	in-laws	<i>dūmod-orĵ</i>	‘sons-in-law’
<i>-ūn</i>	uncles, cousins	<i>pitiš-ūn</i>	‘cousins’
<i>-dzinen</i>	sisters	<i>yax-dzinen</i>	‘sisters’
<i>-(a)jev</i>	times of day; times of year	<i>šab-a-jev</i>	‘(in the) evenings’
		<i>tobistūn-jev</i>	‘(in the) summers’

In a few cases, an originally Arabic suffix *-(w)ot*, *-(ĵ)ot* is used in plural formation, as in *sabza-wot* ‘vegetables’ and *mewa-ĵot* ‘fruits’. Note that the interpretation of these suffixes as plural markers is questionable, as the productive suffix *-en* may optionally be added on, hence *sawza-wot-en* ‘vegetables’.

For a handful of nouns, the stem vowel undergoes ablaut when the plural suffix is added – in these cases, always *-en*. The vowel change in question is either from *u* or *ī* to *a*, as in *lunĵ* ‘cheek’ > *lanĵen* ‘cheeks’, and is a result of a historical linguistic process whereby the inflectional ending shifted stress away from the stem vowel and resulted in a distinct vowel reflex in the plural form (see Sokolova 1967: 37). Nouns which undergo this type of stem-vowel

change during pluralization are shown in Table 5.4:

**Table 5.4:** Vowel alternations in plural forms.

SINGULAR	PLURAL	GLOSS
<i>luŋʃ</i>	<i>lanʃen</i>	‘cheeks’
<i>kud</i>	<i>kaden</i>	‘dogs’
<i>puc</i>	<i>pacen</i>	‘sons’
<i>ʃtʃg</i>	<i>ʃagen</i>	‘calves (of a cow)’
<i>diʃtʃd</i>	<i>diʃaden</i>	‘roofs’

Regarding the distribution of the plural suffix, Shughni shows a pattern comparable to that found in Persian, in which the use of a plural morpheme *-hā* is linked to definiteness (see Gebhardt 2009 and references therein). In particular, in Shughni, if a noun phrase contains a plural demonstrative (but no numeral), then the use of the plural suffix is required. This is shown in (61).

(61) **Plural demonstrative and noun (only): Plural suffix required**

Tu=t dev          mūn\*(-en) wīnt o?  
 you DEM.OBL.PL apple-PL see.PST PQ  
 ‘Did you see those apples?’

On the other hand, if a semantically plural noun phrase contains only a numeral greater than one and a noun, then the use of an additive plural suffix is illicit, as demonstrated in (62).

(62) **Numeral and noun (only): Plural suffix illicit**

Mu-rd cavor mūn\*(-en) darkor.  
 me-DAT four apple-PL necessary  
 ‘I need four apples.’

More complex patterns occur when a noun phrase contains both a numeral and a plural demonstrative. In a noun phrase which contains *only* a plural demonstrative, numeral, and noun (and no adjective), a plural suffix is optional, as in *wāð cavor mūn(-en)* ‘those four apples’. But if it contains an adjective in addition to the demonstrative and numeral, then the plural suffix appears to be obligatory (cf. *wāð cavor yullā mūn\*(-en)* ‘those four big apples’). The compatibility of the plural suffix with numerals and demonstratives in noun phrases is summarized in Table

5.5. More research is needed to determine the precise contexts in which the plural suffix is used, particularly with respect to the interaction of numerals, plural demonstratives, and adjectives.

**Table 5.5:** Compatibility of the additive plural with different types of NPs.

<u>Content</u>	<u>Example</u>	<u>Gloss</u>
NUM NOUN(*-PL)	<i>cavor mūn*(-en)</i>	‘four apples’
DEM.PL NOUN*(-PL)	<i>wāḏ mūn*(-en)</i>	‘those apples’
DEM.PL NUM NOUN(-PL)	<i>wāḏ cavor mūn(-en)</i>	‘those four apples’
DEM.PL NUM ADJ NOUN*(-PL)	<i>wāḏ cavor yullā mūn*(-en)</i>	‘those four big apples’

Additive plural suffixes are also permitted, though not required, on nouns with general reference, as in (63). As this example indicates, when a plural noun is used with generic reference, it often denotes different types or kinds of the noun in question. As discussed above, a bare noun is also permitted in the case of general reference, and hence the bare noun *mūn* is possible in (63).

(63) **Generic reference: Allowed**

Wuz=um mūn(-en) dis žīwǵ.  
 I=1SG apple-PL SO LOVE.PRF  
 ‘I love (different kinds of) apples so much.’

Importantly, additive plural suffixes are generally not compatible with indefinite specific reference, unless they receive a type or kind reading. Thus, in (64), the plural noun *mūn-en* ‘(the) apples’ may refer to a variety of types of apples, but it may not refer to an indefinite set of apples, as in ‘I ate (some) apples’. To create the latter reading, the additive plural must be dropped.

(64) **Indefinite specific reference: Prohibited**

Wuz=um yičand mūn(-en) xūd.  
 I=1SG some apple-PL eat.PST  
**With PL suffix:** ‘I ate some different kinds of apples.’  
**Without PL suffix:** ‘I ate some apples (without reference to kind).’

### 5.2.2 Associative plural

The associative plural in Shughni, like associative plurals cross-linguistically, is most commonly used with a proper noun denoting a human (the focal referent) to form a noun denoting a group of people including the focal referent and others associated with this person (cf. Corbett 2000: 101-111; Moravcsik 2003; and Daniel and Moravcsik 2013, and references cited therein; Edelman & Dodykhudoeva (2009b: 793-794) also include a few sentences on the associative plural in Shughni).

The Shughni associative plural is formed with the distal plural demonstrative (DIR *wāḍ* / OBL *wev*) placed after the focal noun, as in *Mawlonazar wāḍ* ‘Mawlonazar and company’. Proximal and medial demonstratives are not compatible with associative plural readings, even in contexts where the speech situation would generally call for their use. Noun phrases formed with the associative plural may be either in direct (subject) position, in which case the direct demonstrative *wāḍ* is used, or in oblique position, in which case the oblique demonstrative *wev* is used. Examples are given in (65); note that associative plurals in Shughni obligatorily exhibit plural agreement.

#### (65) Associative plural

##### a. Direct associative plural: Noun + *wāḍ*

**Mawlonazar wāḍ=ta**      tar sūr      yad-en      (/ \*yoḍd).  
 Mawlonazar ASS.P.DIR=FAC to wedding come.PRS-3PL (/ \*come.PRS.3SG)  
 ‘Mawlonazar and company are going to come to the wedding.’

##### b. Oblique associative plural: Noun + *wev*

Wuz=um nur **Mawlonazar wev**      tar sūr      wīnt.  
 I=1SG today Mawlonazar ASS.PL.OBL at wedding see.PST  
 ‘I saw Mawlonazar and company at the wedding today.’

The referents of the associative plural in Shughni are determined contextually and are not restricted to a single type of association, as in some languages. For instance, depending on context, the reference of *Mawlonazar wāḍ* in (65a) might be Mawlonazar and his relatives, friends, or some other occasional associates (e.g. soccer teammates).

Note that the additive plural suffix *-en* is also compatible with proper names such as *Mawlonazar*, but its meaning differs in the same way as English ‘John and company’ and ‘(the) Johns’. Thus, whereas the associative plural *Mawlonazar wāḍ* refers to a heterogenous set (i.e. not all referents are named Mawlonazar), the set of people de-

noted by the use of the additive plural in *Mawlonazar-en* is homogenous (i.e. all referents are named Mawlonazar). The sentence in (66) shows an example of the additive plural suffix being used with a proper name. This sentence might be uttered, for instance, when the speaker knows a group of multiple people named Mawlonazar, and all of them are coming to a party.

(66) **Additive plural *-en* with a proper name**

**Mawlonazar-en**=ta yad-en.  
 Mawlonazar-PL=FAC come.PRS-3PL  
 ‘The Mawlonazars are coming’

### 5.2.3 Similative plural

The similative plural in Shughni is formed via reduplication and, like the associative plural, denotes a heterogenous group of referents. Unlike the associative plural, however, the similative plural “denotes a class of objects sharing similar features rather than a group of closely related associates” (Daniel and Moravcsik 2013). The similative plural in Shughni is incompatible with proper names of humans.<sup>2</sup> It is formed by a process of reduplication which is virtually identical to that of Persian echo reduplication (cf. Ghaniabadi et al. 2006) and similar to other reduplicative processes involved in the formation of similatives in South Asian languages such as Hindi-Urdu, Tamil, Telugu, etc. (e.g., Abbi 1985; Keane 2001).

The reduplicative process by which the similative plural is formed in Shughni is a type of echo reduplication which targets single words. The entire word is reduplicated with the exception that initial consonant is replaced with a *m* or *p* (in the case of consonant-initial words) or else a *m* or *p* is added to the beginning of the word (in the case of vowel-initial words). Thus, from *toθč* ‘bowl’, we get *toθč-moθč* or *toθč-poθč* ‘bowls and what not’, and from *kitob* ‘book’ we get *kitob-mitob* or *kitob-pitob* ‘books and what not’. If the base word begins with an *m*, a *p* is obligatory in the reduplicated portion, as in *moθ* ‘stick’ > *moθ-poθ* ‘sticks and what not’; likewise, if the base word begins with a *p*, a *m* is obligatory in the reduplicated portion, as in *pīc* ‘face’ > *pīc-mīc* ‘face and what not (or

<sup>2</sup>A consultant indicates that while similative plurals based on human names, such as *Mawlonazar-Pawlonazar*, are unacceptable, those based on names of countries (e.g. *Tojikistūn-Mojikistūn* ‘Tajikistan and all that’), adjectives (e.g. *kasal-masal* ‘sick and all that’), and verbs (e.g. *žēxt-mēxt* ‘running (around) and all that’) are acceptable, but with distinct semantic nuances not found in the reduplication of common nouns. For instance, *Tojikistūn-Mojikistūn* is generally used ironically or pejoratively, whereas this same semantic tinge is not present in, say, *mošīn-pošīn* ‘cars and what not’.

face and the area around the face)’. This process of reduplication is schematized in Table 5.6, and examples are given in (67).

**Table 5.6:** Formation of the similitive plural via reduplication.

WORD SHAPE	BASE	SMLV. PLURAL	EXAMPLE	GLOSS
Word is consonant-initial and does not begin with <i>m</i> or <i>p</i> .	CVC... →	CVC...- <i>m/p</i> VC...	<i>kitob-mitob</i>	‘books and what not’
Word is vowel-initial	VC... →	VC...- <i>m/p</i> VC...	<i>asal-pasal</i>	‘honey and what not’
Word begins with <i>m</i>	<i>m</i> VC... →	<i>m</i> VC...- <i>p</i> VC...	<i>marûb-parûb</i>	‘cream and what not’
Word begins with <i>p</i>	<i>p</i> VC... →	<i>p</i> VC...- <i>m</i> VC...	<i>pût-mût</i>	‘balls and what not’

(67) **Similitive plural**

- a. Māš=ām **pût-mût** vūd o?  
 we=1PL ball-SMLV bring.PST PQ  
 ‘Did we bring balls and what not?’
- b. Tu=t **ūžin-pūžin** čūžj o?  
 you=2SG dinner-SIMV do.PRF PQ  
 ‘Have you eaten dinner and what not?’
- c. Yu **dek-mek** kādand?  
 DEM.DIR.M pot-SMLV where?  
 ‘Where are those pots and what not?’

Two further aspects of the similitive plural, both apparent in these examples, are of note. First, this process is extremely productive and applies easily to recently borrowed words such as Russian *ūžin* ‘dinner’ (ex. 67b). And second, similitive plurals do not agree as plurals, but rather as masculine singular. Even when the base noun is feminine, as with *dek* ‘pot’ in (67c), the similitive plural formed from the base is invariably masculine. This fact is related to the notion that the semantic concepts of abstractness and collectiveness are linked to masculine gender in Shughni. (On the role of abstractness and collectiveness in gender assignment, see Section 7.4.3, especially Subsection 7.4.3.2.)

### 5.2.4 Collective plural

The final type of plural to be discussed here is the collective plural. The collective plural is similar to the additive plural in that it denotes homogenous sets of entities, but it is similar to the associative and similative plurals in that it denotes sets of entities which have some internal cohesion (Daniel and Moravcsik 2013). That is, whereas the associative plural *Shahlo wāḏ* ‘Shahlo and company’ and the similative plural *toθč-moθč* ‘bowls and what not’ are heterogeneous (not all their referents are Shahlos or bowls, respectively), the collective plural *kitob-xel* ‘(mass of) books’ is homogenous in that all its referents are necessarily books. But whereas the referents of the noun with the additive plural suffix *kitob-en* ‘books’ may not have any inherent relation to each other, those denoted by *kitob-xel* are generally located in the same physical area.

There are two ways which a collective plural shows up in Shughni. The first is through the suffix *-xel* ‘group; bunch’, seen in the example *kitob-xel* ‘(mass of) books’ mentioned above. From a grammatical point of view, the suffix *-xel* differs from the additive plural suffixes discussed in Section 5.2.1 above in that while nouns with an additive plural suffix behave as plural with respect to agreeing elements such as demonstratives, verbal conjugation, etc., nouns which combine with *-xel* are treated as singular. Examples of the collective plural formed through *-xel* are given in (68).

(68) **Collective plural suffix *-xel***

- a. Yid     **kitob-xel**   yedand   čīr   kiχt(/\*kin-en)?  
 DEM.DIR book-COLL there.MED what do.PRS.3SG(\*do.PRS-3SG)  
 ‘What are those books doing there?’
- b. Tu=ta   di(/\*dam)                     **kitob-xel**   čīr   kin-i?  
 you=FAC DEM.OBL.M(/\*DEM.OBL.F) book-COLL what do.PRS-2SG  
 ‘What are you going to do with that (pile of) books?’

The example in (68b) illustrates another aspect of the collective plural which aligns it with similative plurals discussed in Section 5.2.3. In particular, collective plural nouns are treated as grammatically masculine, even when the noun in question is typically treated as feminine, as is the case with the *kitob* ‘book’ in (68b). In this sense, in collective plurals, as in similative plurals, the semantic notion of collectiveness apparently forces any collective plurals to behave as masculine.<sup>3</sup>

<sup>3</sup>I follow Edelman & Dodykhudoeva (2009b: 793) in glossing *-xel* as a suffix, but the status of this morpheme as suffix, clitic, or stand-



Grammatical gender is likewise implicated in the second way collective plurals are formed in Shughni. Here, canonically feminine nouns may receive a mass or plural reading if used as masculine. This phenomenon, also discussed in greater detail in Section 7.4.3 in Chapter 7 on grammatical gender, is exhibited in (69). The noun *mūn* ‘apple’ is canonically a feminine noun in Shughni, as shown in (69a), where it agrees with the feminine demonstrative *yā*. However, in (69b), the same noun is grammatically masculine (cf. the masculine demonstrative *yu*) and receives a collective or mass reading. This sentence might have as its referent, for instance, a large pile of apples stored in a pantry.

(69) **Collective plural via grammatical gender shift**

a. **Individual (feminine) reading of *mūn* ‘apple’**

**Yā**            **mūn** čī-nd?  
 DEM.DIR.F apple who.OBL-POSS  
 ‘Whose apple is that?’

b. **Collective (masculine) reading of *mūn* ‘apple’**

**Yu**            **mūn** čī-nd?  
 DEM.DIR.M apple who.OBL-POSS  
 ‘Whose (pile of) apples is that?’

## 5.3 Nominal derivation

Nominal derivation in Shughni is carried out most commonly by suffixation and compounding. These processes are common in the formation of both nouns and adjectives. To my knowledge, virtually all *noun*-forming derivational affixes are suffixes, while a number of prefixes exist which are used in the formation of adjectives. This section looks in turn at the derivation of nouns (5.3.1) and the derivation of adjectives (5.3.2).

---

alone word is not clear. Preliminary indications nonetheless point to its status as a suffix, rather than an independent word. For instance, although *-xel* is identical in form with the Tajik word *хел* – *xel* ‘sort; kind; state’, Shughni *-xel* is not typically pluralized. Thus, forms like *wev čêd-xel* ‘those knife-*xel*’ are uncommon, and the singular form *wi čêd-xel* ‘that knife-*xel*’ may denote more than one group of knife. Ultimately, more research is needed to pinpoint the grammatical status of *xel* in Shughni.

### 5.3.1 Formation of nouns

This subsection looks first at the derivation of (i.e. formation of) nouns via suffixation and compounding. This topic is divided into four subsections, the first three of which deal with affixation, and specifically the formation of nouns **denominally** (Section 5.3.1.1), **deadjectivally** (Section 5.3.1.2), and **deverbally** (Section 5.3.1.3). Section 5.3.1.4 then turns to the **formation of nouns through compounding**.

#### 5.3.1.1 Denominal nouns via affixation

Here we examine a number of suffixes which attach to a noun to form another noun, the most productive of which are the following: (i) the abstract-noun forming suffix  $-(y)i$ , which is perhaps more productive with adjectives than nouns, and (ii) a series of diminutive-like suffixes which are similar in form but exhibit distinctions in their usage and meaning. We also look at a couple of less productive suffixes, including the suffix *-and* which denotes types of meat and the gender-distinguishing suffixes  $-edz$  (F) /  $-ej$  (M) which denote step-relations among family members.

The **abstract-noun forming suffix**  $-(y)i$  attaches to a number of nouns with comparatively concrete meanings to form an abstract noun denoting a quality or concept. In a few cases, mostly when the noun to which this suffix attaches, the connecting palatal glide  $-y$  is instead a  $g$ , as in *amsoyā* ‘neighbor’ > *amsoyagi* ‘good relations among neighbors; neighborliness’. Examples of this suffix are given in Table 5.7.

**Table 5.7:** Abstract-noun forming suffix  $-(y)i$  with nouns.

BASE NOUN	GLOSS	DERIVED NOUN	GLOSS
<i>josūs</i>	‘spy’	<i>josūs-i</i>	‘espionage’
<i>duzd</i>	‘thief’	<i>duzd-i</i>	‘theft’
<i>asgar</i>	‘soldier’	<i>asgar-i</i>	‘military service’
<i>bačā</i>	‘child’ (<Tj.)	<i>bača-gi</i>	‘childhood’
<i>amsoyā</i>	‘neighbor’	<i>amsoya-gi</i>	‘neighborliness’

Secondly, Shughni possesses two distinct noun-forming diminutive suffixes with similar forms:  $-(y)ik$  and  $-(y)ak$ . The common semantic thread among these suffixes is the notion of diminutive, but each has distinct semantic connotations. The **endearment suffix**  $-(y)ik$ , examples of which are given in (70), attaches to nouns to express

the speaker's positive emotion toward the noun's referent, as in (70a) or the speaker's pity for the referent of the noun, as in (70b). Note that the endearment suffix can also attach to pronouns, as in (70c).

(70) **Diminutive suffix -(y)ik: Endearment**

- a. Mu puc-**ik**=ta    katarameθ kor    kiřt.  
my son-DIM=FAC all.day    work do.PRS.3SG  
'My dear son works all day.'
- b. Yā            kampīr-**ik**            dis bečorā.  
DEM.DIR.F old.woman-DIM so    pitiful  
'That old woman is so pitiful.'
- c. Tu-**yik**=ta    mot sāw-i.  
you-DIM=FAC tired become.PRS-2SG  
'(Poor little) you will become tired.'

The suffix -(y)ak, for its part, has two possible semantic nuances. First, as a **(pure) diminutive suffix**, -(y)ak attaches to a noun and encodes a more literal diminutive meaning than the endearment suffix -(y)ik, indicating that a noun is physically small. It lacks the same sense of endearment as -ik. Examples of the diminutive -(y)ak on nouns are given in (71).<sup>4</sup>

(71) **(Pure) diminutive suffix -ak**

- a. Ar        wi            boy    yi dzulik čīd-**ak**    vud.  
down.in DEM.OBL.M garden a    small house-DIM be.PST.M  
'There was a small house in the garden.'
- b. Yax    ku        dam        mu sūmka-**yak** ca    dāk-i.  
sister please DEM.OBL.F my bag-DIM    SUBR give.PRS-2SG  
'Sis, could you give me my (little) bag, please?'

In addition to its diminutive sense, however, -ak can also be used as a **pejorative suffix** to indicate a negative attitude toward the referent of the noun in question. The use of -ak as a pejorative is shown in (72).

---

<sup>4</sup>A consultant points out that the suffix -(y)ak is commonly used among younger populations to indicate a positive attitude toward a noun, though not necessarily endearment, as with -ik. Such a reading is possible for (71b).

(72) **Pejorative suffix -ak**

lk=u                      tu   virod-**ak**   mu-rd   ačaθ   nist     xuš.  
 PREC=DEM.DIR.M   your brother-PEJ   me-DAT   at.all   COP.NEG   pleasant  
 ‘I don’t like that brother of yours at all.’

Shughni has two further noun-forming suffixes which select for nouns, both of which are much less productive than the diminutive-like suffixes discussed above. The first is the suffix *-and*, which attaches to animals and forms types of meat. Hence, we have *čax-and* ‘chicken meat’ from *čax* ‘chicken’, *žow-and* ‘beef’ from *žow* ‘cow’, and *may-and* ‘mutton’ from *may* ‘sheep’.<sup>5</sup> This suffix is at least somewhat productive in that it can attach to nouns denoting more animals whose meat is more exotic, such as *divūsk* ‘snake’ > *divūsk-and* ‘snake meat’ and *nahang* ‘whale’ > *nahang-and* ‘whale meat’.

The second is the gender-distinguishing suffix *-edz* (F) / *-eĵ* (M), which attaches to nouns denoting familial relations to form nouns denoting step-relations. Thus, for instance, from *yax* ‘sister’, we get *yax-edz* ‘step-sister’, from *virod* ‘brother’ we get *virod-eĵ* ‘step-brother’, and from *nān* ‘mother’ and *tāt* ‘father’ we get *nān-edz* ‘step-mother’ and *tāt-eĵ* ‘step-father’.

A summary of noun-forming suffixes which select for nouns is given in Table 5.8.

**Table 5.8:** Noun-forming suffixes which select for nouns.

SUFFIX TYPE	FORM	EXAMPLE	GLOSS
Abstract quality	-(y) <i>i</i>	<i>duzd-i</i>	‘theft’
Endearment	- <i>ik</i>	( <i>mu</i> ) <i>puc-ik</i>	‘(my) dear son’
Diminutive	- <i>ak</i>	<i>daryo-yak</i>	‘little river’
Pejorative	- <i>ak</i>	( <i>mu</i> ) <i>virod-ak</i>	‘stinking brother (of mine)’
Meat-type	- <i>and</i>	<i>žow-and</i>	‘beef’
Step-relation	- <i>edz</i> (F)	<i>yax-edz</i>	‘step-sister’
	- <i>eĵ</i> (M)	<i>virod-eĵ</i>	‘step-brother’

<sup>5</sup>Note that this suffix is string-identical to the locative case-like suffix *-and*. These two suffixes differ only in their stress: the suffix denoting types of meat is stressed, while the case-like suffix is unstressed.

### 5.3.1.2 Deadjectival nouns via affixation

Shughni possesses a single productive suffix used in the formation of nouns from adjectives: the suffix *-(y)i*, which is identical in form to that which attaches to nouns to form abstract qualities, as discussed in Section 5.3.1.1 above. Also similar to its use with nouns, we see a few instances in which it contains an initial *g*, as in *maǰdzūnj-gi* ‘hunger’. This suffix attaches to a much larger number of adjectives than nouns; nonetheless, it forms the same type of abstract nouns when used with adjectives as when used with nouns. Examples are given in Table 5.9.

**Table 5.9:** Abstract-noun forming suffix *-(y)i* with adjectives.

<u>BASE ADJECTIVE</u>	<u>GLOSS</u>	<u>DERIVED NOUN</u>	<u>GLOSS</u>
<i>dzulik</i>	‘small’	<i>dzulik-i</i>	‘childhood’
<i>baland</i>	‘high’	<i>baland-i</i>	‘height’
<i>tez</i>	‘fast’	<i>tez-i</i>	‘speed’
<i>qaror</i>	‘calm’	<i>qaror-i</i>	‘calmness’
<i>maǰdzūnj</i>	‘hungry’ (<Tj.)	<i>maǰdzūnj-gi</i>	‘hunger’
<i>kasal</i>	‘sick’	<i>kasal-i</i>	‘sickness’
<i>šitto</i>	‘cold’	<i>šitto-yi</i>	‘cold(ness)’
<i>pīr</i>	‘old’	<i>pīr-i</i>	‘old age’
<i>qarzdor</i>	‘indebted’	<i>qarzdor-i</i>	‘indebtedness’
<i>xuš</i>	‘happy’	<i>xuš-i</i>	‘happiness’

### 5.3.1.3 Deverbal nouns via affixation

Shughni has a variety of affixal means for deriving nouns from verbs. Here, I will discuss four affixes which target specific types of verb stems and form nouns which are not compounds: the occupational suffix *-ǰ*, the processual suffix *-idz*, the frozen (non-productive) instrument-forming suffix *-īdz*, and the gerund-forming suffix *-ak*. The first three suffixes attach to a verb’s present stem, while the fourth attaches to a verb’s infinitive stem.

Perhaps the most productive noun-forming suffix which targets verbs is the **occupational suffix *-ǰ***, which targets a verb’s present stem. Nouns derived via this suffix may denote a person who does the action in question professionally or merely as a hobby. Examples are given in Table 5.10.

**Table 5.10:** Occupational nouns with the suffix *-tj*.

<u>PRS. STEM</u>	<u>GLOSS</u>	<u>DERIVED NOUN</u>	<u>GLOSS</u>
<i>ǰoy-</i>	‘read’	<i>ǰoy-tj</i>	‘reader’
<i>niviš-</i>	‘write’	<i>niviš-tj</i>	‘writer’
<i>ancāv-</i>	‘sew’	<i>ancāv-tj</i>	‘seamstress’
<i>parōāō-</i>	‘sell’	<i>parōāō-tj</i>	‘seller’
<i>pīdz-</i>	‘cook’	<i>pīdz-tj</i>	‘cook’

The **processual suffix** *-idz* also attaches to a verb’s present stem to form a noun denoting a process. This suffix is much less productive than the occupational suffix *-tj*. Examples are given in Table 5.11.

**Table 5.11:** Processual nouns with the suffix *-idz*.

<u>PRS. STEM</u>	<u>GLOSS</u>	<u>DERIVED NOUN</u>	<u>GLOSS</u>
<i>zidār-</i>	‘sweep’	<i>zidār-idz</i>	‘(process of) sweeping’
<i>čēr-</i>	‘write’	<i>čēr-idz</i>	‘(process of) ploughing’

A small number of nouns in Shughni exhibit the frozen form of a **once productive agentive suffix** *-īdz* (< *\*-ačī*), which has cognate suffixes with varying degrees of productivity in the other languages of the Shughni-Rushani group. This suffix apparently once productively attached to present verb stems to form nouns denoting an instrument or agent. In a few cases, we can still recognize the present verb stem to which it attached (e.g. *ǰābīdz* ‘stick; switch’ < *ǰāb-* ‘beat’), while in others we have only oblique evidence that the present stem once existed in an ancestral language of Shughni. For instance, the noun *zarīdz* ‘partridge’ apparently contains this suffix, but no verb with present stem *zar-* exists in modern Shughni. Nonetheless, this exact stem with the meaning ‘to sing’ is found in the Eastern Iranian language Ossetian, and a reconstructed form for *zarīdz*, *\*zaračī*, is thus considered to be reliable (see Karamshoev 1978: 212). Examples of nouns purported to contain this suffix are given in Table 5.12; those which are derived from a verb which is no longer used in modern Shughni have an asterisk.

**Table 5.12:** Nouns containing the once productive agentive suffix *-idz*.

<u>PRS. STEM</u>	<u>GLOSS</u>	<u>DERIVED NOUN</u>	<u>GLOSS</u>
<i>ǰāb-</i>	‘beat; hit’	<i>ǰāb-īdz</i>	‘stick; switch’
<i>šarð-</i>	‘defecate’	<i>šarð-īdz</i>	‘butt; behind’
<i>*wiǰi-</i>	‘unlock’	<i>wiǰ-īdz</i>	‘key’
<i>*zar-</i>	‘sing?’	<i>zar-īdz</i>	‘partridge’

Lastly, the **gerundial suffix -ak** attaches productively to infinitive stems to form deverbal nouns – i.e. *gerunds*. Example sentences containing this suffix are given in (73a) and (73b). Further examples of gerunds derived from verbs using this suffix are given in Table 5.13.

(73) **Gerundial suffix -ak**

- a. *Xīd-ak*      *ūži*      *sar*      *o?*  
eat.INF-GER already start PQ  
‘Has the eating already started?’
- b. *Ded-ak*      *di*      *sut*      *xabar mu kin-et.*  
enter.INF-GER SUBR become.PST.M news me do.IMP-2PL  
‘When you enter (lit. when the entering has happened), let me know.’

**Table 5.13:** Gerunds formed with the suffix *-ak*.

<u>INFINITIVE</u>	<u>GERUND</u>	<u>GLOSS</u>
<i>ded-</i>	<i>dedak</i>	‘entering’
<i>xīd-</i>	<i>xīdak</i>	‘eating’
<i>naǰtīd-</i>	<i>naǰtīdak</i>	‘leaving’
<i>xêvd-</i>	<i>xêvdak</i>	‘sleeping’
<i>mīzd-</i>	<i>mīz dak</i>	‘building’

A summary of Shughni noun-forming suffixes which attach to verb stems is given in Table 5.14.

**Table 5.14:** Suffixes used in de-verbal derivation of nouns.

SUFFIX TYPE	FORM	SELECTS	EXAMPLE	GLOSS
<b>Occupational; Agentive</b>	<b>-ĩĵ</b>	PRS STEM	<i>nivišĩĵ</i>	‘writer’
<b>Processual</b>	<b>-idz</b>	PRS STEM	<i>zidāridz</i>	‘cleaning; tidying’
<b>Instrument; Agentive (unproductive)</b>	<b>-ĩdz</b>	PRS STEM	<i>ṡābĩdz</i>	‘stick; switch’
<b>Gerundial</b>	<b>-ak</b>	INF STEM	<i>xĩdak</i>	‘eating’

#### 5.3.1.4 Formation of nouns via compounding

We now turn to the formation of nouns through compounding. Compound nouns in Shughni are fundamentally of two types: those which combine two nominal elements (most commonly two nouns, less commonly an adjective and a noun), and those which combine a nominal element with a verb. In the latter case, the nominal element always precedes the verbal element. Each type of compound is schematized in (74). This subsection looks at each in turn.

(74) a. **Noun-based compound noun**

NOUN / ADJ + HEAD N

*civĩnc* ‘bee’ + *rēdz* ‘hive’ > *civĩnc-rēdz* ‘beehive’

b. **Verb-based compound noun**

NOUN / NUM + VERB

*yoc* ‘fire’ + *kun* ‘make’ > *yoc-kun* ‘firekeeper’

We begin the discussion in Section 5.3.1.4.1 with noun-based compounds – i.e. those which contain a noun or adjective and a head noun. Section 5.3.1.4.2 then turns to a presentation of verb-based compound nouns.

**5.3.1.4.1 Noun-based compound nouns.** Compound nouns which consist of two nouns – or of an adjective and a noun – are relatively few in comparison with those consisting of a noun and a verb. Compounds of the



NOUN + NOUN type are always right-headed and retain the gender of the head noun. Thus, the noun *đust-ǰac* ‘water for washing hands’ (lit. hand-water’) is feminine by virtue of the fact that the noun *ǰac* ‘water’ is feminine. Similarly, the compound *kāl-dārđ* ‘headache’ (lit. ‘head-pain’) is masculine due to the fact that the noun *dārđ* ‘pain’ is masculine. Examples of compounds of this kind are given in Table 5.19. Note that the last compound *têr-čoy* is an example of a compound of the ADJ – NOUN type.

**Table 5.15:** NOUN + NOUN compounds in Shughni.

COMPOUND	ROOT-BY-ROOT	GLOSS
<i>đust-ǰac</i>	hand-water	water for washing hands
<i>kāl-dārđ</i>	head-pain	headache
<i>civīnc-rêdz</i>	bee-nest	beehive
<i>takyā-parwīn</i>	pillow-cover	pillowcase
<i>tīr-čīd</i>	top-house	house at the top of a town
<i>šīr-čoy</i>	milk(Tj.)-tea	milktea
<i>*têr-čoy</i>	black-tea	black tea

A curious subtype of NOUN–NOUN compounds are those which have a second component which is clearly etymologically related to an independent Shughni noun, but which has undergone phonological, and in some cases grammatical, changes. This type of compound is seen more in the formation of compound adjectives, but one compound-forming element of this kind is commonly used in the formation of nominal compounds. This is the gender-distinguishing element *-bic* (F), *-buc* (M), which is used as a diminutive on certain nouns expressing familial relations, as well as on nouns denoting animals to indicate that the animal is a youth. In its former usage, this diminutive is used strictly to signal the age relation of one family member with respect to another, rather than to indicate endearment. Examples of each usage are given in Table 5.16.<sup>6</sup>

Etymologically, this compound-forming element is linked to the Shughni word *puc* ‘son’ and derives ultimately from older Iranian *\*puθra-* ‘son’ Karamshoev (1978: 265–275). However, unlike the word *puc*, the morpheme *-buc/-bic* (with an initial voiced plosive) cannot be used as an independent noun. According to Karamshoev,

<sup>6</sup>When used with animals, this suffix typically selects a specific gender of the animal (e.g. *kud-buc* ‘puppy’, rather than *kid-bic*, and *piš-bic* ‘kitten’, rather than *puš-buc*). In these cases *kud* and *piš* are typically used to refer to ‘dog’ and ‘cat’, respectively, regardless of an animal’s gender. On a few nouns denoting animals, this suffix may have a pejorative meaning, such as in *žow-bic* ‘little (female) cow’ and *markāb-buc* ‘little (male) donkey’, which are used to mean ‘stupid person’.

**Table 5.16:** Genealogical diminutive *-bic/-buc*.

<u>Familial relations</u>		<u>Animal young</u>	
FORM	GLOSS	FORM	GLOSS
<i>yax-bic</i>	‘little sister’	<i>kud-buc</i>	‘puppy’
<i>virod-buc</i>	‘little brother’	<i>piš-bic</i>	‘kitten’
<i>pitiš-bic</i>	‘little cousin (F)’	<i>šīg-buc</i>	‘calf’
<i>pitiš-buc</i>	‘little cousin (M)’	<i>xirs-buc</i>	‘bear cub’
<i>xêr-bic</i>	‘little niece’	<i>mohi-buc</i>	‘fry (young fish)’
<i>xêr-buc</i>	‘little nephew’		

the development of this suffix into a gender-distinguishing derivational formant occurred in later stages of the language’s development via analogy with already established gender-distinguishing vowel alternations.

This compound-forming element is clearly distinct from its independent-noun counterpart, however, both phonologically and grammatically. Phonologically, its first consonant has become voiced, a phenomenon which is seen in other compound-forming elements, as discussed in the section on adjectival compounds below. And grammatically, it has developed the feminine counterpart *-bic*, where no such feminine counterpart exists for the noun *puc* ‘son’ (cf. *rizīn* ‘daughter’). For these reasons, the compound forming component *-buc/-bic* might be considered to be in the process of transition from independent noun to suffix, and its development is of particular interest in future research on the language.

**5.3.1.4.2 Verb-based compound nouns.** There are fundamentally three types of verb-based nominal compounds in Shughni, each of which involve a nominal element – generally a noun or numeral – preposed to a verbal element – either a present stem or an infinitival stem. This subsection examines each of the three types.

The first type of verb-based compound noun is that in which a **nominal element is preposed to borrowed verbal component**, such as the Tajik present stem *kun-* (from *kardan* ‘do’). Although the verbal component is a borrowed verb stem, the nominal element is nonetheless in most cases a native Shughni morpheme. This is the situation, for instance, in *yoc-kun* ‘firekeeper’ (lit. fire-make.PRS), where the noun *yoc* is a native Shughni word.

This type of compound noun may be further subdivided into compounds which consist only of a nominal component and a verbal component, as with *yoc-kun*, and compounds which take a further suffixal element, namely

the suffix *-ak*. The former type generally denotes a person with a particular occupation, while the latter typically denotes a type of tool or other inanimate man-made creation. This first type of verb-based compound noun is schematized with examples in (75).

(75) **Type 1: Borrowed present stem**

Noun / Num	+	(Borrowed) Verb.Prs	(+ <i>-ak</i> )
<i>yoc</i> ‘fire’	+	<i>kun</i> ‘do.PRS (Taj.)’	> <i>yoc-kun</i> ‘firekeeper’
<i>dastā</i> ‘handle’	+	<i>dor</i> ‘have.PRS (Taj.)’	+ <i>-ak</i> > <i>dasta-dor-ak</i> ‘(cup with a long handle)’

Further examples of this type of compound noun, including examples with and without the suffix *-ak*, are given in Table 5.19. These examples involve the borrowed Tajik present stems *kun-* ‘do’ and *xon-* ‘read; study’ (with the further change of the vowel /o/ before a nasal to /ũ/, a regular process in Shughni). Note that the nouns *du-xũn* ‘poor student’ (lit. two-study) and *panj-xũn* ‘good student’ (lit. five-study) make reference to the grading system in Tajik schools, in which a score of *five* is the highest possible mark. Hence, a poor student is one who receives a two, and a good student is one who receives a five.

**Table 5.17:** Verb-based compound nouns with a borrowed (Tajik) present stem.

<u>Simple (no suffix)</u>		
<u>COMPOUND</u>	<u>ROOT-BY-ROOT</u>	<u>GLOSS</u>
<i>yoc-kun</i>	fire-do	‘firekeeper’
<i>du-xũn</i>	two-study	‘poor student’
<i>panj-xũn</i>	five-study	‘good student’
<i>namoz-xũn</i>	prayer-read	‘one who prays regularly’
<u>Complex (with suffix)</u>		
<u>COMPOUND</u>	<u>ROOT-BY-ROOT</u>	<u>GLOSS</u>
<i>dasta-dor-ak</i>	handle-have.PRS	‘cup with long handle’
<i>ǰūdž-kun-ak</i>	wind-make.PRS	‘fan’
<i>yet-kun-ak</i>	open-make.PRS	‘opener’

The second type of verb-based compound noun is also right-headed and which consists of **a noun and native Shughni present stem**. This type differs crucially from the first type in that it *requires* additional suffixal material

beyond the two stems. Compound nouns of this type may take only the suffix *-īj*, an agentive suffix denoting humans with a specific occupation or predilection (which, as discussed in Section 5.3.1.3, is also compatible with plain nouns without objects). An example of this type is *sūrat-zêz-īj* ‘photographer’ (lit. ‘photo-take.PRS-AGT.SFX’). They may also take both the suffix *-īj* and the same instrument-forming suffix *-ak* as discussed above, in which case the gender of the noun may be either masculine or feminine. An example of this type is *sūrat-zêz-īj-ak* ‘camera’ (lit. ‘photo-take.PRS-AGT.SFX-INS.SFX’). The structure of compound nouns built on *-īj* and *-ak* are schematized in (76).

(76) **Type 2: Native present stem**

<b>Nominal</b>	+	<b>(Native) Verb.Prs</b>	+	<b>-īj</b>	+	<b>(-ak)</b>	
<i>sūrat</i> ‘photo’	+	<i>zêz</i> ‘take.PRS	+	<i>-īj</i> (AGT.SFX)	>	<i>sūrat-zêz-īj</i>	‘photographer’
<i>sūrat</i> ‘photo’	+	<i>zêz</i> ‘take.PRS	+	<i>-īj</i> (AGT.SFX)	+	<i>-ak</i>	> <i>sūrat-zêz-īj-ak</i>
							‘camera’

Further examples of the second type of verb-based compound noun are given in Table 5.18. These include both human agentive compounds with the suffix *-īj*, as well as inanimate compounds denoting tools built with the suffix *-īj* and the suffix *-ak*.

**Table 5.18:** Occupational compounds with the suffix *-īj*.

<b>Human agentive compounds with <i>-īj</i></b>					
<u>NOUN</u>	<u>GLOSS</u>	<u>VRB.PRS</u>	<u>GLOSS</u>	<u>COMPOUND</u>	<u>GLOSS</u>
<i>ǰūvd</i>	‘milk’	<i>parōād-</i>	‘sell’	<i>ǰūvd-parōād-īj</i>	‘milk seller’
<i>garōā</i>	‘bread’	<i>pīdz-</i>	‘cook’	<i>garōā-pīdz-īj</i>	‘baker’
<i>šīrčoy</i>	‘milk tea’	<i>weđ-</i>	‘put’	<i>šīrčoy-weđ-īj</i>	‘milk tea maker’
<b>Inanimate instrument compounds with <i>-īj</i> and <i>-ak</i></b>					
<u>NOM.</u>	<u>GLOSS</u>	<u>VRB.PRS.</u>	<u>GLOSS</u>	<u>COMPOUND</u>	<u>GLOSS</u>
<i>sūrat</i>	‘photo’	<i>zêz-</i>	‘take’	<i>sūrat-zêz-īj-ak</i>	‘camera’
<i>kaš</i>	‘hot’	<i>anjāv-</i>	‘grab’	<i>kaš-anjāv-īj-ak</i>	‘cloth for handling hot items’
<i>yubor</i>	‘dust’	<i>zidār-</i>	‘sweep’	<i>yubor-zidār-īj-ak</i>	‘duster’

And third, there are compound nouns which consist of a **nominal element and a native Shughni (short) infinitive stem**. This type forms gerunds, as in *luq-zinêd* ‘clothes washing’ (lit. clothes-wash.INF) and consists of a semantic

object, such as *luq* ‘clothes’ in the example *luq-zinêd*, and a transitive verb. This third type of verb-based nominal compound is schematized in (77).

(77) **Type 3: Native infinitive stem**

**Nominal** + **(Native) Verb.Inf**

*luq* ‘clothes’ + *zinêd* ‘wash.INF’ > *luq-zinêd* ‘clothes washing’

Further examples of the third type of compound are given in Table 5.19.

**Table 5.19:** NOUN + INF. STEM compounds in Shughni.

COMPOUND	ROOT-BY-ROOT	GLOSS
<i>awqot-pêxt</i>	food-cook	(food-)cooking
<i>luq-zinêd</i>	clothes-wash	clothes-washing
<i>kitob-ǰêyd</i>	book-read	book-reading
<i>palow-xīd</i>	pilaf-eat	pilaf-eating
<i>ǰac-birêǰt</i>	water-drink	water-drinking

This concludes the discussion on verb-based compound nouns in Shughni. A summary of this type of compound noun is given in Table 5.20.

**Table 5.20:** Verb-based compound nouns: Summary.

TYPE	FORMATION	EXAMPLE	GLOSS
1	<b>Nominal + Borrowed (Tajik) Present Stem</b>	<i>namoz-xūn</i> <i>dasta-dor-ak</i>	‘one who prays diligently’ ‘cup with a long handle’
2	<b>Nominal + Native Present Stem</b>	<i>kitob-ǰoy-īǰ</i> <i>noxūn-zêz-īǰ-ak</i>	‘book reader’ ‘nailclipper’
3	<b>Nominal + Native Infinitive Stem</b>	<i>kitob-ǰêyd</i> <i>noš-xīd</i>	‘book reading’ ‘apricot eating’

### 5.3.2 Formation of adjectives

The derivation of adjectives, like the derivation of nouns, often occurs through suffixation. Unlike in the derivation of nouns however, there are a handful of productive prefixes which are used in the formation of adjectives. Adjectives may also be formed via compounding, most notably with an adjective preposed to a noun. This process is not as widespread as the formation of compound nouns, but it shows a number of peculiarities and is worthy of examination here. This section looks at the formation of adjectives through each of these processes and is structured in a similar way to Section 5.3 above on nominal derivation.

The first three subsections here examine the **derivation of adjectives from nouns** (Section 5.3.2.1), **from other adjectives** (Section 5.3.2.2), and **from verbs** (Section 5.3.2.3). Each of these sections looks at both suffixation and prefixation processes, as applicable. Lastly, Section 5.3.2.4 turns to the formation of adjectives via compounding.

#### 5.3.2.1 Denominal formation of adjectives

The formation of adjectives from nouns occurs through both suffixation and prefixation. Common adjective-forming suffixes include: (i) the **relational adjective-forming** *-(y)in*, generally attaching a noun denoting a material, as in *ǝorg* ‘wood’ > *ǝorg-in* ‘wooden’; (ii) the suffix *-gār*, which forms adjectives denoting humans with **a certain nature or tendency**, as in *noz* ‘whim’ > *noz-gār* ‘whimsical’; (iii) the suffix *-rāng* which signifies the **resemblance of a human to another**, as in *dod* ‘father’ > *dod-rāng* ‘resembling (one’s) father’; and (iv) the **temporal relational adjective-forming** *-īnj*, which attaches to a noun denoting a period of time, as in *biyor-īnj* ‘yesterday’s’. Examples of each within sentences are given in (78).

#### (78) Derived (suffixal) adjectives in context

##### a. Suffix *-in*: Relational Adjective

Wuz=um **wūrǝ-in** kurtā zoǝt.  
I=1SG wool-REL.ADJ dress take.PST  
‘I bought a woolen dress.’

##### b. Suffix *-gār*: Nature or proclivity

Mu nān=i **naqli-gār** odam žīwǝ.  
my mother=3SG speech-ADJ person love.PRF  
‘My mom loves talkative people.’

c. **Suffix *-rāng*: Resemblance to another human**

Yā dis **nān-rang**, wam xarākteṛ=xu gāp-ðêd.  
 she so mother-like her character=and word-hit.INF  
 ‘She is so much like her mother in her personality and way of speaking.’

d. **Suffix *-īnĵ*: Temporal relational adjectives**

Wāð=en wi **biyor-īnĵ** awqot kaš čūd=xu xūd=en.  
 they=3PL DEM.OBL.M yesterday-REL.ADJ food hot make.PST=and eat.PST=3PL  
 ‘They heated up the food from yesterday and ate it.’

A summary of denominal-adjective-forming suffixes with examples is given in Table (5.21).

**Table 5.21:** Denominal adjective-forming suffixes.

FORM	MEANING	EXAMPLE	GLOSS
-(y) <i>in</i>	‘made of N’	<i>ðorg-in</i> <i>xeṣa-in</i> <i>tillo-wūrĵ</i>	‘wooden’ ‘glass’ ‘woolen’
- <i>gār</i>	‘prone to N’	<i>naqli-gār</i> <i>ðêd-gār</i> <i>noz-gār</i>	‘talkative’ ‘combative’ ‘whimsical’
- <i>rāng</i>	‘looking like N’	<i>nān-rāng</i> <i>dod-rāng</i> <i>bob-rāng</i>	‘looking like (one’s) mother’ ‘looking like (one’s) father’ ‘looking like (one’s) grandfather’
- <i>īnĵ</i>	‘from ADV of time’	<i>biyor-īnĵ</i> <i>asūd-īnĵ</i>	‘yesterday’s’ ‘last year’s’

Common denominal-adjective-forming prefixes include: (i) **ba-**, which forms an adjective **possessing the quality** denoted by the noun; (ii) **be-**, which forms an adjective **lacking the quality** denoted by the noun; (iii) **ser-**, which creates an adjective with an abundant amount of the quality or substance denoted by the noun; and (iv) **kam-**, which creates an adjective with a small amount of the quality or substance denoted by the noun. Examples of each in context are given in (79).

(79) **Derived (prefixal) adjectives in context**a. **Prefix *ba-*: Possessing a quality**

Ominā dis **ba-barakat**, xoli=ta mardum-ard yordam kiṣt.  
 Omina so with-blessing, always=FAC people-DAT help do.PRS.3SG  
 ‘Omina is so blessed; she’s always helping people.’

b. **Prefix *be-*: Lacking a quality**

**Be-ṣarm** (odam)=um wīnč=atā tu ḵino nāy!  
 without-shame (person)=1SG see.PRF=but you like no  
 ‘I’ve seen shameful people, but none like you!’

c. **Prefix *ser-*: Full of a quality/substance**

Mam sol māš zamīn **ser-osil** vud.  
 DEM.OBL.F year our land full-harvest be.PST.F  
 ‘Our land was very fertile this year.’

d. **Prefix *kam-*: Having only a small amount of a quality/substance**

Yūd-and odam **kam-gāp** ca vid, bīdi.  
 here person little-word SUBR be.PRS.3SG better  
 ‘Here, if a person is reticent, it’s a good thing.’

A summary of denominal-adjective-forming prefixes, together with examples, is given in Table 5.22.

**Table 5.22:** Denominal adjective-forming prefixes.

FORM	MEANING	EXAMPLE	GLOSS
<i>ba-</i>	‘with, having N’	<i>ba-mazā</i> <i>ba-barakat</i>	‘delicious’ (lit. ‘with flavor’) ‘blessed’ (lit. ‘with blessing’)
<i>be-</i>	‘without N’	<i>be-ṣarm</i> <i>be-rām</i>	‘shameless’ ‘merciless’
<i>ser-</i>	‘full of N’	<i>ser-osil</i> <i>ser-namak</i>	‘fertile’ (lit. harvest-ful) ‘salty’
<i>kam-</i>	‘with little N’	<i>kam-gāp</i> <i>kam-wūn</i>	‘reticent’ (lit. ‘with little word’) ‘with little wool’ (of a sheep)



### 5.3.2.2 Deadjectival formation of adjectives

Three of the most commonly used suffixes in the deadjectival formation of adjectives are the following: (i) the suffix **-rāng**, which describes a noun which possesses **a small amount of the quality** denoted by the base adjective, as in *šitto* ‘cold’ > *šitto-rāng* ‘a bit cold’; (ii) the **augmentative suffix -(y)izor**, which describes a noun possessing a large amount of the quality in question, as in *mot-izor* ‘very tired’; and (iii) the diminutive-like suffix **-ak**, which requires an additional suffix **-aθ**, as in *maḃdzūnj* > *maḃdzūnj-ak-aθ* ‘a little hungry’. A single prefix is used in the derivation of adjectives: the prefix **no-** derives an adjective with the meaning ‘not ADJ’. Examples of words formed with these affixes are given in context in (80).

#### (80) De-adjectival (suffixal) derived adjectives in context

##### a. Suffix **-rāng**: Somewhat ADJ

Mu kamānda nur **sust-rāng** vud.  
my team today weak-DIM be.PST.M  
‘My team was somewhat weak today.’

##### b. Suffix **-(y)izor**: Very ADJ

Māš=ām bādi kor **mot-izor** sat.  
we=1PL after work tired-AUG become.PL  
‘We got really tired after work.’

##### c. Suffix **-(y)akaθ**: A little ADJ

Wāð sawol-en=en **qīn-akaθ** vad.  
DEM.DIR.PL question-PL=3PL difficult-DIM be.PST.PL  
‘Those questions were a bit difficult.’

##### d. Prefix **no-**: Not ADJ

Wuz=ta **no-balad** odam-en qati gāp na-ðām.  
I=FAC un-familiar person-PL with word NEG-hit.PRS.1SG  
‘I don’t talk to unfamiliar people.’

A summary of deadjectival-adjective-forming prefixes, together with examples, is given in Table 5.23.

**Table 5.23:** Deadjectival adjective-forming suffixes.

<u>Suffixes</u>			
<u>FORM</u>	<u>MEANING</u>	<u>EXAMPLE</u>	<u>GLOSS</u>
<i>-rāng</i>	‘somewhat ADJ’	<i>šitto-rāng</i> <i>sust-rāng</i>	‘a bit cold’ ‘a bit weak’
<i>-(y)izor</i>	‘very ADJ’	<i>maǰdzūnj-izor</i> <i>mot-izor</i>	‘very hungry’ ‘very tired’
<i>-ak(-aθ)</i>	‘somewhat ADJ’	<i>maǰdzūnj-ak-aθ</i> <i>qīn-ak-aθ</i>	‘a bit hungry’ ‘a bit difficult’
<u>Prefixes</u>			
<i>no-</i>	‘not ADJ’	<i>no-balad</i> <i>no-darkor</i>	‘unfamiliar’ ‘unnecessary’

### 5.3.2.3 Deverbal formation of adjectives

There is a single productive means of producing deverbal adjectives in Shughni: the suffix *-in*, which attaches to perfect stems to form adjectival resultative participles. (See Section 11.1 on passive-like constructions, and specifically Section 11.1.2, which goes into more detail on the adjectival resultative participle.)

This suffix may attach to either unaccusative verbs or transitive verbs, but not unergative verbs (see Section 4.4.2 for more on this distinction), and it is compatible with both simplex and complex verbs. The examples in (81) illustrate.

#### (81) Adjectival resultative participles with *-in*

##### a. Unaccusative Verb

Tar **virīč-in**                      yēnak=ta    xu    na-čis-en  
 in    break.PRF.F-PASS.ADJ    MIRROR=FAC    REFL    NEG-look.at.PRS-3PL  
 ‘People shouldn’t look at themselves in a broken mirror.’

b. **Simplex Transitive Verb**

Ku fuk **ǰēyǰ-in** kitob-en dam sitol-ti ribi.  
 please all read.PRF-PASS.ADJ book-PL DEM.OBL.F table-LOC put.2SG.IMP  
 ‘Please put all the books which have been read on that table.’

c. **Complex Transitive Verb**

Tar mi mu **zār čūǰǰ-in** ǰust mā-ǰi.  
 in DEM.OBL.M my injury do.PRF-PASS.ADJ hand PROH-hit.2SG.IMP  
 ‘Don’t hit me on my injured hand.’

5.3.2.4 **Formation of adjectives via compounding**

Compound adjectives in Shughni are generally formed with an adjective preposed to a noun. No additional morphology, such as an adjective-forming derivational suffix, is used in the formation of compound adjectives. The structure of compound adjectives in Shughni is schematized in (82). Initial examples of compound adjectives are given in Table 5.24.

(82) **Structure of compound adjectives**

ADJECTIVE + NOUN > ADJECTIVE-NOUN  
*daroz* ‘long’ *zūǰ* ‘sleeve’ > *daroz-zūǰ* ‘long-sleeved’

**Table 5.24:** Compound adjectives: Examples.

<u>ADJECTIVE</u>	<u>MORPHEME-BY-MORPHEME</u>	<u>GLOSS</u>
<i>daroz-zūǰ</i>	long-sleeve	‘long-sleeved’ (e.g. of a shirt)
<i>rūšt-kurtā</i>	red-shirt	‘wearing a red shirt’ (e.g. of a man)
<i>kut-yūnǰ</i>	short-hair	‘short-haired’ (e.g. of a boy)
<i>kut-ǰum</i>	short-tail	‘short-tailed’ (of an animal)

Two aspects of Shughni compound adjectives are worthy of note, both of which relate to the expression of gender. The first regards gender-distinguishing adjectives when used as the adjectival component of a compound adjective, and in particular what determines whether the feminine or masculine form of the adjective is used. In my fieldwork

on the language, I have noticed a different pattern from that described by Karamshoev (1978: 284-289); this may be due to dialectal differences, language change, or both.

In any case, whereas Karamshoev describes a pattern in which the gender of the adjectival component of the compound is determined by the gender of the noun modified by the compound, I have found that the gender of the adjectival component is determined by the gender of the nominal component of the compound. As an example, consider the compound adjective formed with the gender-distinguishing adjective *baq* (F) ~ *buq* (M) ‘convex’ and the feminine noun *nêdz* ‘nose’. According to Karamshoev, both *baq-nêdz* ‘big-nosed (F)’ and *buq-nêdz* ‘big-nosed (M)’ are possible forms and agree in gender with the noun they modify. My consultants indicate, however, that only the form with the feminine adjective *baq-nêdz* is possible, and that this is because the nominal component of the compound *nêdz* ‘nose’ is feminine. To be sure, it is not the case that only a single form of gender-distinguishing adjectives has been frozen for use in compounds. The masculine form of this same adjective is used, for instance, in the compound *buq-mîð* ‘hunch-backed’, where the nominal component *mîð* ‘back’ is masculine.

Compound adjectives with gender-distinguishing adjectival components are relatively few. Nonetheless, a couple further examples of this phenomenon are provided in Table 5.25.

**Table 5.25:** Compound adjectives with gender-distinguishing adjectival components.

GENDER-DISTINGUISHING ADJ.	NOUN	GLOSS	GDR	COMPOUND	GLOSS
<i>baq</i> (F) ~ <i>buq</i> (M) ‘convex’	<i>nêdz</i> <i>mîð</i>	‘nose’ ‘back’	F M	<i>baq-nêdz</i> <i>buq-mîð</i>	‘big-nosed’ ‘hunch-backed’
<i>žarn</i> (F) ~ <i>žurn</i> (M) ‘round’	<i>pîc</i>	‘face’	M	<i>žurn-bîc</i>	‘round-faced’
<i>žilaq</i> (F) ~ <i>žiluq</i> (M) ‘vertical’	<i>cem</i>	‘eye’	M	<i>žiluq-dzem</i>	‘big-eyed’

The second noteworthy aspect of compound adjectives in Shughni has to do with the second component of compound adjectives, and in particular those in which the nominal component derives from an independent noun whose initial (voiceless) consonant has undergone voicing. This phenomenon was already noted in Section 5.3.1.4.1 on

compound nouns as it relates to the use of the morpheme *bic* (F)  $\sim$  *buc* (M), which derives from the noun *puc* ‘son’ and is used in the formation of compound nouns denoting animal young and younger relatives.

The process whereby the initial consonant of a noun becomes voiced when used as the nominal component of a compound affects many, but not all nouns (cf. *kurtā* ‘shirt’  $>$  *rūšt-kūrtā* ‘red-shirted’,  $*rūšt-gurtā$ ). It appears to affect those nouns which are most commonly used in compounds. Three such nouns which are used in the formation of compound adjectives are given, together with examples, in Table 5.26.

**Table 5.26:** Denominal components of compound adjectives with voicing of initial consonant.

NOUN	IN COMPOUND	GLOSS	EXAMPLE	GLOSS
<i>pīc</i>	<i>-bīc</i>	‘face’	<i>žurn-bīc</i>	‘round-faced’
<i>cem</i>	<i>-dzem</i>	‘eye’	<i>žiluq-dzem</i>	‘big-eyed’
<i>kīl</i>	<i>-gīl</i>	‘head’	<i>safed-gīl</i>	‘with gray hair’ (lit. ‘white-headed’)

Whereas from the noun *puc* a feminine formant *-bic* developed alongside the masculine *-buc*, the compound components shown in Table 5.26, with the possible exception of *-gīl* ( $<$  *kīl* ‘head’), do not distinguish gender. Here again, the pattern suggested by my fieldwork data differs from that described by Karamshoev (1978), likely due to the same factors of dialectal differences and/or language change.

Specifically, whereas Karamshoev indicates that a number of gender-distinguishing compound adjectives exist based on the formant *-gīl* (M) and its feminine counterpart *-gāl*, which purportedly developed via analogy with gender-distinguishing vowel patterns much in the same way as the feminine formant *-bic*. He lists, in addition to the aforementioned *safed-gīl* ‘gray-haired’ (lit. white-head), *kata-gīl* (M)  $\sim$  *kata-gāl* (M) ‘big-headed’, *tēr-gīl* (M)  $\sim$  *tēr-gāl* (M) ‘black-haired’ (lit. ‘black-head’), and *zīrd-gīl* (M)  $\sim$  *zīrd-gāl* (M) ‘blond’ (lit. yellow-head). Of these, my consultants have not heard the feminine counterparts of the latter two. For the former two, they indicate that the feminine counterpart is acceptable but rare, and the masculine form with the stem vowel *ī* is by far the most common.

The discrepancies between my data and that of Karamshoev (1978) regarding gender distinction in the first and second components of compound adjectives remain to be explained. They may indicate, on the one hand, that the expression of grammatical gender in certain areas of Shughni grammar may be undergoing rapid changes. This

could be the case despite the fact that, as will be seen in Chapter 7, the system whereby gender is assigned to nominals appears to be relatively stable. On the other hand, these discrepancies may be due to dialectal differences. In any case, both language change and dialectal differences in the sphere of grammatical gender are rich areas for future research in Shughni.

## 5.4 Numerals

The final topic to be explored in this initial chapter on Shughni nominals is that of numerals. Shughni employs a system of numerals which makes significant use of both native and borrowed (Tajik) elements. Here, **the form of numerals** is examined first in Section 5.4.1, which includes a look at both Shughni numerals and borrowed Tajik numerals, which are generally adapted to the phonotactics of Shughni. Both native numerals and Tajik numerals have widespread usage in Shughni, though they are used in generally mutually exclusive environments. Section 5.4.2 provides a discussion of the precise **contexts in which each group of numerals is used**.

### 5.4.1 Numerals: Form

This subsection first looks at the form of native Shughni numerals (Section 5.4.1.1) and then turns to the form of borrowed Tajik numerals used in Shughni (Section 5.4.1.2).

#### 5.4.1.1 Native Shughni Numerals

Shughni has an indigenous system of numerals which have undergone the sound changes typical of Shughni and other Eastern Iranian languages (e.g. initial *\*d* > *ǰ* – cf. Shughni *ǰīs* ‘ten’, Persian *dah*). As we will see below, however, only Shughni numerals between one and ten – shown in Table 5.27 – are used in everyday speech.

Indigenous Shughni numerals beyond ten, though not typically used in modern speech, are attested in a number of previous works on the language (see, e.g., Bakhtibekov 1979: 20-23; Edelman & Dodykhudoeva 2009b: 797). According to these authors, Shughni uses a base-ten system, where for each multiple of ten (beginning at 20), *ǰīs* ‘ten’ is preceded by the number it is multiplied by (e.g. *ǰu ǰīs* ‘twenty’ – lit. ‘two ten(s)’). The same pattern follows for the numbers ‘one hundred’ – *ǰīs-ǰīs-(ak)* and ‘one thousand’ – *ǰīs-ǰīs-ǰīs-(ak)*, which optionally take

**Table 5.27:** Shughni numerals 1-10.

1	<i>yīw</i>	6	<i>xoǵ</i>
2	<i>ǫuyūn</i>	7	<i>wūvd</i>
3	<i>aray</i>	8	<i>waǵt</i>
4	<i>cavor</i>	9	<i>now</i>
5	<i>pīndz</i>	10	<i>ǫīs</i>

the suffix *-ak*. For numbers between multiples of ten, the enclitic *=at* ‘and’ follows the multiple of ten and links it to the remaining numeral, a strategy which is used both between ten and twenty (e.g. *ǫīs=at cavor* ‘fourteen’ – lit. ‘ten=and four’) and above twenty (e.g. *ǫu ǫīs=at yīw* ‘twenty-one’ – lit. ‘two ten(s)=and one’). Shughni numerals above ten are shown in Table 5.28.

**Table 5.28:** Shughni numerals above 10.

11-20		20-100		21-29	
11	<i>ǫīs=at yīw</i>	20	<i>ǫu ǫīs</i>	21	<i>ǫu ǫīs=at yīw</i>
12	<i>ǫīs=at ǫu</i>	30	<i>ara ǫīs</i>	22	<i>ǫu ǫīs=at ǫu</i>
13	<i>ǫīs=at aray</i>	40	<i>cavor ǫīs</i>	23	<i>ǫu ǫīs=at aray</i>
14	<i>ǫīs=at cavor</i>	50	<i>pīndz ǫīs</i>	24	<i>ǫu ǫīs=at cavor</i>
15	<i>ǫīs=at pīndz</i>	60	<i>xoǵ ǫīs</i>	25	<i>ǫu ǫīs=at pīndz</i>
16	<i>ǫīs=at xoǵ</i>	70	<i>wūvd ǫīs</i>	26	<i>ǫu ǫīs=at xoǵ</i>
17	<i>ǫīs=at wūvd</i>	80	<i>waǵt ǫīs</i>	27	<i>ǫu ǫīs=at wūvd</i>
18	<i>ǫīs=at waǵt</i>	90	<i>now ǫīs</i>	28	<i>ǫu ǫīs=at waǵt</i>
19	<i>ǫīs=at now</i>	100	<i>ǫīs-ǫīs(-ak)</i>	29	<i>ǫu ǫīs=at now</i>

In addition to the additive formula (of the type ‘ten and one’) described above and exhibited by the numerals in Table 5.28, both Bakhtibekov (1979) and Edelman & Dodykhudoeva (2009b) note that in the upper reaches of each multiple of ten (e.g. 8-9, 18-19, 28-29, and so on), it is possible to use subtraction. When this occurs, the operator *kam* ‘little; less’ is used in the place of *=at* ‘and’, and the order of numerals is reversed. That is, the remainder numeral – to be subtracted – is placed before *kam*, which is followed by the multiple of ten. For instance, ‘nine’ may be expressed as either *now* or *yīw kam ǫīs* – lit. ‘one less ten’, and ‘nineteen’ may be expressed as either *ǫīs=at now* or *yīw kam ǫu ǫīs* – lit. ‘one less twenty’.

Finally, as discussed in Section 5.1 above, each of the native numerals from one to three have short forms; namely, *yīw* ‘one’ has short form *yi*; *ǫuyūn* ‘two’ has short form *ǫu*; and *aray* ‘three’ has short form *ara*. The short forms

of these numerals are used when they precede either the noun they are quantifying or another modifier within the same noun phrase. The long form, on the other hand, is used when no overt noun follows them, as in contexts in which the noun is elided. Compare, for instance, the examples in (83); in (83a), the short numeral *ǫu* is directly followed by the noun it quantifies, whereas in (83b), the long form *ǫuyǫn* is not followed by any material in its same NP, as the noun is understood from context and has been elided.

(83) **Short and long forms of the numeral *ǫuyǫn* ‘two’**

a. **Noun follows numeral > Short form**

Dev        [ǫu    mǫn]<sub>NP</sub> mu-rd   dāk.  
DEM.OBL.PL two.SF apple   me-DAT give.2SG.IMP  
‘Give me those two apples.’

b. **No noun follows numeral > Long form**

Dev        ǫuyǫn mu-rd   dāk.  
DEM.OBL.PL two.LF me-DAT give.2SG.IMP  
‘Give me those two.’

### 5.4.1.2 Borrowed Tajik Numerals

All Tajik numerals – including those from 1-10 – have been borrowed into Shughni and are used regularly in everyday speech. As will be discussed in the following subsection, there are areas in which only Tajik numerals are used, where Shughni numerals, even those between one and ten, are infelicitous. Hence, Shughni employs both the native numerals and borrowed Tajik numerals from one to ten.

However, it is important to note that borrowed Tajik numerals are often modified to conform more closely to the Shughni sound system. Such “Shughnification” of Tajik numerals includes a number of processes, many of which are regular sound processes in the language. These are enumerated below:

- (i) *h* deleted in all positions;
- (ii) *a* lengthened to *ā* when followed by deleted *h* in word-final position;
- (iii) *b* > *v* in the cluster *bd*;
- (iv) *o* > *ū* before *n*;
- (v) *v* > *w* intervocalically.



Borrowed Tajik numerals in Shughni are exhibited in Table 5.29, and any sound changes are noted in the ‘Change’ column, with the numeral matching that in the list above.

**Table 5.29:** Borrowed Tajik numerals in Shughni.

Num.	Tajik		Shughni	Change
	Cyr.	Rom.	BORROWED FORM	
1	як	<i>yak</i>	<i>yak</i>	—
2	ду	<i>du</i>	<i>du</i>	—
3	се	<i>se</i>	<i>se</i>	—
4	чор	<i>čor</i>	<i>čor</i>	—
5	панҷ	<i>pañĵ</i>	<i>pañĵ</i>	—
6	шаш	<i>šaš</i>	<i>šaš</i>	—
7	ҳафт	<i>haft</i>	<i>aft</i>	(i)
8	ҳашт	<i>hašt</i>	<i>ašt</i>	(i)
9	нӯҳ	<i>nūh</i>	<i>nū</i>	(i)
10	даҳ	<i>dah</i>	<i>dā</i>	(i, ii)
15	понздаҳ	<i>ponzdah</i>	<i>pūnzdā</i>	(i, ii, iv)
17	ҳабдаҳ	<i>habdah</i>	<i>avdā</i>	(i, ii, iii)
21	бисту як	<i>bistu yak</i>	<i>bistu yak</i>	—
90	навад	<i>navad</i>	<i>nawad</i>	(v)
100	сад	<i>sad</i>	<i>sad</i>	—
1000	ҳазор	<i>hazor</i>	<i>azor</i>	(i)
1m	миллион	<i>million</i>	<i>milliūn</i>	(iv)
1b	миллиард	<i>milliyard</i>	<i>milliyard</i>	—

### 5.4.2 Usage: Shughni vs. Tajik numerals

As mentioned above, speakers of modern Shughni use both native Shughni numerals and Tajik numerals on a regular basis, but in mutually exclusive contexts. Importantly, however, even in those contexts which call for the use of Shughni numerals, only those from one to ten are used (see Table 5.27); in all cases involving numbers greater than ten, regardless of context, the Tajik numeral is used. Russian numerals, for their part, are rarely, if ever used. We only see a Russian numeral in a single instance: Shughni *nol* ‘zero’, a borrowing of Russian нуль (*nul’*). The following sections examine the contexts which call for Shughni numerals (Section 5.4.2.1) and those which call for Tajik numerals (Section 5.4.2.2).

### 5.4.2.1 Contexts for Shughni Numerals

Shughni numerals are generally used to express specific quantities, whether dealing with concrete objects or abstract, quantifiable concepts, as in measurements of time (e.g. seconds, minutes, weeks, etc.) or actions (e.g. ‘an instance of reading’). Shughni numerals also are used to express years of age. These contexts are summarized and elaborated upon with examples in Table 5.30.

**Table 5.30:** Contexts/quantities for Shughni numerals (1-10).

CONTEXT	QUANTITY	EXAMPLES			
Objects/people	1-10	<i>ara mūn</i>	3 apples	<i>waxt nafar</i>	8 people
Time periods	1-10	<i>du soat</i>	‘2 hours’	<i>wūvd mēst</i>	7 months
Years of age	1-10	<i>cavor sola</i>	4 years old	<i>xoŷ sola</i>	6 years old
Actions	1-10	<i>yi xēydow</i>	one (instance of) reading		

As indicated in Table 5.30, even in these contexts Shughni numerals are only used when the quantity in question is between one and ten. When the quantity is greater than ten, the Tajik numeral is invariably used. This is shown in Table 5.31

**Table 5.31:** Quantities expressed by Tajik numerals.

CONTEXT	QUANTITY	EXAMPLES			
Objects/people	> 10	<i>sezda mūn</i>	13 apples	<i>bīst nafar</i>	20 people
Time periods	> 10	<i>yoza soat</i>	‘11 hours’	<i>duwozda mēst</i>	12 months
Years of age	> 10	<i>bīstučor sola</i>	24 years old	<i>šast sola</i>	60 years old

In all instances in which a count noun is quantified, the classifier *dūnā* ‘piece; unit’ may follow the numeral and precede the noun, as in *du dūnā marūd* ‘two pears’, lit. ‘two units of pear’. It may also be used when the noun has been elided – e.g. *du dūnā* ‘two (units)’ – which might be a response to the question *turd cūnd marūd darkor?* ‘how many pears do you need?’. Its usage is always optional, and it can be used with virtually any count noun. Noun phrases which contain the classifier *dūnā* are typically perceived as more formal than those without it.

### 5.4.2.2 Contexts for Tajik Numerals

Whereas Shughni numerals 1-10 are used in virtually all instances of quantifying nouns, the use of Tajik numerals 1-10 is confined to specific contexts and expressions, including the following: (i) telling time; (ii) days of the week; (iii) calendar days; (iv) calendar years; and (v) ordinal numbers.<sup>7</sup>

The time of day in Shughni, as in Tajik, is expressed using the word *soat(-i)* ‘hour(-EZ)’, followed by the numeral. Days of the week are generally expressed with *rūz-i* ‘day-EZ’ followed by the ordinal number of the day, beginning with the first day as Monday. The word for Sunday – *o(t)dix* is borrowed from Russian *отдых – otдых* ‘rest’.<sup>8</sup> Calendar days are also expressed with an ordinal number, connected to the month with an *izofat*. Calendar years are as in Tajik, with the number spelled out beginning with the thousands, then hundreds, and so on, each connected by *=u* ‘and’. Ordinal numbers are formed with the suffix *-um* when following the noun they modify, and with the suffix *-umīn* when preceding the noun they modify. In all these contexts, the use of Shughni numerals is infelicitous. Examples are given in Table 5.32.

**Table 5.32:** Contexts requiring Tajik numerals in Shughni.

CONTEXT	TAJIK NUMERAL (felicitous)	SHUGHNI NUMERAL (infelicitous)	GLOSS
telling time	<i>soat(-i) se</i>	* <i>soat(-i) aray</i>	three o’clock
days of the week	<i>rūz-i se</i>	* <i>rūz-i aray</i>	Wednesday
calendar days	<i>se-yum-i iyūl</i>	* <i>aray-um-i iyūl</i>	July 3rd
calendar years	<i>azor=u nūsad=u panjo</i>	* <i>ōīs-ōīs-ōīs=at(ā) ...</i>	1950
ordinal numbers	<i>se-yum(-īn)</i>	* <i>aray-um(-īn)</i>	third

<sup>7</sup>Bakhtibekov (1979: 23) suggests that Shughni numerals can be used together with the ordinal-forming suffix *-um*, as found in Tajik: e.g. *yīw-um* ‘first’, *ōuyūn-um* ‘second’, *aray-um* ‘third’, etc. In modern speech, however, ordinal numbers are much more commonly formed with the Tajik numeral: *yak-um* ‘first’, *du-yum* ‘second’, *se-yum* ‘third’.

<sup>8</sup>Despite the fact that all elements in constructions such as *rūz-i yak* ‘Monday’ are borrowed from Tajik (Tj. *rūz* ‘day’), this way of expressing the days of the week is not used in Tajik, where the standard Persian days based on *šanbe* are used: *šanbe* ‘Saturday’, *yakšanbe* ‘Sunday’, . . . *jum’a* ‘Friday’, etc. Some Shughni speakers opt to use the Tajik days of the week based on *šanbe*.



## Chapter 6

# Pronouns and Demonstratives

This chapter, the second of three in the series on nominals, looks at pronouns and demonstratives in Shughni. Each topic is treated in detail, with the **various types of pronouns, including personal pronouns, interrogative pronouns, indefinite pronouns, and reflexive and emphatic pronouns**, each examined in turn in Section 6.1. This section gives special attention to the treatment of the subject-oriented reflexive pronoun *xu* and the emphatic pronoun *xubaθ*. It has generally been assumed in the literature on Shughni that the latter constitutes the direct-case counterpart to the former. I demonstrate here that this is not the case, however, and that the difference between the two is purely semantic and pragmatic rather than grammatical.

Section 6.2 turns to **demonstrative pronouns** in Shughni, another topic which has yet to be studied in any detail. In this chapter, I look at both *exophoric* uses of demonstratives – i.e. where they are used to pick out something in the physical space surrounding the speech situation – and *endophoric* uses, where they are used to pick out some aspect of the discourse itself. To my knowledge, this aspect of Shughni demonstratives has not been studied before. Moreover, I show that the language’s triple-deictic system of demonstratives has a fundamentally addressee-oriented medial form, but with a twist: objects which are perceived to be within the *shared interactional space* of both speaker and addressee may be referred to with a medial demonstrative, even though they are outside the immediate space of the addressee.

## 6.1 Pronouns

This section on pronouns is divided into four subsections as follows. Section 6.1.1 looks at personal pronouns, which in Shughni includes only first- and second-person pronouns. Section 6.1.2 then looks at interrogative pronouns and *wh*-phrases, including both simplex and complex interrogative pronouns, and Section 6.1.3 turns to indefinite pronouns, which are built on interrogative pronouns. Finally, Section 6.1.4 takes an in-depth look at ‘self-’ pronouns, including the subject-oriented reflexive pronoun *xu* and the emphatic pronoun *xubaθ*.

### 6.1.1 Personal pronouns

As seen in Section 4.2.1.1, Shughni makes use of personal pronouns for the first- and second-person singular and plural. Object pronouns are always overt, but subject pronouns may be unrealized in cases where they are not focused, contrastively topicalized, or otherwise emphasized.

As elsewhere in the language, the direct case is reserved for subjects, while the oblique case is used for non-subjects, including direct and indirect objects, objects of adpositions, and possessors. Note that only in the first-singular is the direct form of a personal pronoun distinct from its oblique form. Note also that the first-person singular direct pronoun *(w)uz* is variably pronounced with or without an onset glide. In the third-person, demonstrative pronouns, discussed in Section 6.2, are used in the function of personal pronouns.

Shughni personal pronouns are given in Table 6.1. In this table, the cells of pronouns which have distinct forms in the direct and oblique cases – i.e. first-person singular – are shaded:

**Table 6.1:** Personal pronouns.

	<b>Direct</b>		<b>Oblique</b>	
	SING	PLUR	SING	PLUR
1	<i>(w)uz</i>	<i>māš</i>	<i>mu</i>	<i>māš</i>
2	<i>tu</i>	<i>tama</i>	<i>tu</i>	<i>tama</i>

A couple notes are in order regarding the second-person plural pronoun *tama*. First, this pronoun can be used both as a second-person plural pronoun and as a second-person singular honorific pronoun. In this sense, it is

similar in function to Tajik *уҷмо* (*šumo*) and Russian *вы* (*vy*).<sup>1</sup> An example of its use as a second-person-singular honorific is given in (84). Note that *tama* requires the same agreement morphemes, present suffix *-et* or past-tense second-position clitic *=et*, in both its singular and plural usages.

(84) **Singular honorific usage of *tama***

Clinton, **tama**=yet                      mu-rd lūd    idi,    guyo,    wuz **tama**-rd    yi sūg lūm.  
 Clinton you.SG.HON=2SG.HON me-DAT say.PST COMP it.seems I    2SG.HON-DAT a story tell.PRS.1SG  
 ‘Clinton, you asked me to tell you a story.’

And second, the pronoun *tama* does not descend from an any known ancient Iranian second-plural stem or enclitic pronoun. Instead, it is likely the outcome of a word-formational model found in a substrate language, whereby the second-person singular form was attached to a plural marker. A similar formation is found in a number of languages which make up the Central Asian Sprachbund, including both Indo-European languages such as the Iranian languages Pashto, the other Shughni-Rushani languages, Yazghulami, and Ishkashimi, and the (possibly) Indo-Aryan Dardic languages, as well as non-Indo-European languages such as Dravidian, Himalayan, and possibly Burushaski (Edelman 1980b: 22–23; see also Sims-Williams 1996: 651 and Wendtland 2009: 179–180).

## 6.1.2 Interrogative pronouns and *wh*-phrases

This section examines interrogative pronouns and *wh*-words in Shughni. Section 6.1.2.1 discusses simplex interrogative pronouns, and Section 6.1.2.2 looks at the complex interrogative pronouns formed on the basis of simplex interrogative pronouns. Subsection 6.1.2.3 then provides a brief look at *wh*-phrases, where an interrogative pronoun combines with other material to form a complex syntactic phrase. For a discussion of the distribution of *wh*-words in Shughni, as well as non-canonical types of *wh*-questions, such as echo questions, see Section 12.3.

### 6.1.2.1 Simplex interrogative pronouns

Shughni has eight simplex interrogative pronouns. These, along with examples, are shown in Table 6.2. I assume this list to be exhaustive, but the status of two interrogative pronouns, *ca* and *car*, as discussed below, is noteworthy.

<sup>1</sup>A consultant believes that the honorific use of *tama*, as shown in (84) is a relatively new development which has resulted from the influence of Persian and/or Russian. This is an intriguing prospect, but one which needs further research to be confirmed.

Note also that only one interrogative pronoun has distinct direct and oblique forms: *čay* (DIR) ‘who’ and *či* (OBL) ‘whom’.

**Table 6.2:** Fundamental interrogative pronouns.

FORM	GLOSS	EXAMPLE	TRANSLATION
<i>čřz</i>	what (thing; reason)	<i>tu čřz xāri?</i>	<b>what</b> are you eating?
<i>čřr</i>	what (action; goal)	<i>čřr kini?</i>	<b>what</b> are you doing?
<i>čřdũm</i>	which	<i>čřdũm kitob fort turd?</i>	<b>which</b> book do you want?
<i>kā</i>	where	<i>tu tar kā (sāwi)</i>	<b>where</b> are you going?
<i>čay</i> (DIR) <i>či</i> (OBL)	who(m)	<i>čay tu qati</i> <i>tu=t či qati</i>	<b>who</b> is with you <b>whom</b> are you with?
<i>cũnd</i>	how much	<i>tu ja cũnd sũm?</i>	<b>how many</b> Somoni do you have?
<i>car</i> * *See note in prose on <i>car</i> .	why (not)	<i>tu car tar kor na-deđi?</i>	<b>why don’t</b> you get work?
<i>ca</i> ** (+N) **See note in prose on <i>ca</i> .	which/what	<i>cawaxt ta yādi?</i>	<b>when</b> are you coming?

The morpheme *ca*, with respect to its use in *wh*-words, is found most often as the first member of a complex *wh*-word, as in *ca-waxt* ‘when’ (lit. ‘what time’) and *ca-rāng* ‘how’ (lit. ‘what manner’).<sup>2</sup> I consider *ca* in these cases to be a bound interrogative pronoun, while the remaining pronouns in Table 6.2 are free. Nonetheless, *ca* is also found apparently as a free morpheme in one isolated instance, namely the question *ca sut?* ‘what happened?’.

The interrogative pronoun *car* is noteworthy for two reasons. The first regards its limited usage: *car* is only used with negated verbs in constructions that serve as suggestions, such as ‘*why don’t* you get work’, as in (85a). *Car* is also noteworthy for its form, in particular its similarity to *carāng*. Given both its form and restricted usage, one may wonder whether *car* has come about via the shortening of the complex *wh*-word *carāng* ‘how’. The shortening of word forms is found elsewhere in the language, such as with past-tense verb stems (e.g. *čũd* > *ču* ‘did’) and imperative forms (*sāw* > *sa* ‘go!’). From a synchronic standpoint, however, *car* and *carāng* have clearly distinct

<sup>2</sup>A morpheme with the form *ca* is also used as a subordinator in relative, adverbial, and other subordinate clauses. Though an etymological link between the two is likely, I am unable to comment on such a connection here.



and generally non-overlapping uses. If we place these words in minimally different (negated) environments, as is done in the examples in (85), we can see that *car* is used to express a suggestion, while *carāṅg* is more likely to express surprise. Therefore, although *car* may be both etymologically and semantically linked to *carāṅg*, I conclude that in modern Shughni it is a simplex interrogative pronoun.

(85) **Juxtaposition of *car* and *carāṅg***

- a. Tu *car* tar kor na-deð-i?  
 you why to work NEG-enter.PRS-2SG  
 ‘Why don’t you get some work?’

Implication: ‘I suggest you do so.’

- b. Tu *carāṅg* tar kor na-deð-i?  
 you how to work NEG-enter.PRS-2SG  
 ‘How do you not get work?’

Implication: ‘I’m surprised (e.g. because you are smart and capable).’

Finally, note that both *čīz* and *čīr* can be translated as ‘what’, but the former tends to inquire about a (concrete or abstract) object, while the latter tends to inquire about an action. Compare, for instance, the examples in (86).

(86) **Wh-words *čīz* and *čīr* as ‘what’**

- a. Tu=ta *čīz* (#*čīr*) lûv-i?  
 you=FAC what.thing (#what.action) say.PRS-2SG  
 ‘What will you say?’

- b. Tu=ta *čīr* (#*čīz*) kin-i?  
 you=FAC what.action (#what.thing) do.PRS-1SG  
 ‘What will you do?’

Further, in some instances, both *čīz* (also *čīz-ard*, *čīz jāt*) and *čīr* can be translated as ‘why’, and in these cases the same distinction between object (or reason) and action holds. In particular, *čīr* is typically used to inquire about the reason for a verb involving directed movement, while *čīz* is used to inquire about the reason for other verbs. This distinction is exhibited in (87).

(87) *Wh-words čīz and čīr as ‘what’*

- a. Yu **čīz**                    (#čīr)                    awqot na-xīr-t?  
     he what.reason (#what.action) food NEG-eat.PRS-3SG  
     ‘Why isn’t he eating?’
- b. Yu **čīr**                    (#čīz)                    pi Pomir yat?  
     he what.action (#what.reason) to Pamirs come.PST  
     ‘Why did he come to the Pamirs?’

## 6.1.2.2 Complex interrogative pronouns

Simplex interrogative pronouns combine with other elements to form complex interrogative pronouns. Here, I present three types of complex interrogative pronouns, including those based on (i) the bound interrogative pronoun *ca-*, (ii) dative and locative suffixes, and (iii) derivational suffixes. Section 6.1.2.3 turns to a brief presentation of *wh*-phrases based on adpositions.

As mentioned in the discussion on simplex interrogative pronouns, the bound interrogative pronoun *ca-* combines with other elements to form complex interrogative pronouns. The use of *ca-* is limited, however, as it only combines with two words, namely *waxt* ‘time’ to form *cawaxt* ‘when’ and *rāng* ‘manner’ to form *carāng* ‘how’.

Another type of complex interrogative pronoun is based on the addition of the dative and locative suffixes to a simplex interrogative pronoun. These are the suffixes *-(a)rd* (alternative form *-ra*) and *-(a)nd* (alternative form *-ndi*), respectively.

Regarding productivity, the use of these suffixes in the formation of complex interrogative pronouns is much less restricted than that of *ca-*, although it is not entirely productive. For instance, the form *cawaxt-ard* (DAT) ‘around what time’ is acceptable, but *cawaxt-and* (LOC) is not acceptable (cf. *cūnd-and* ‘at what time’).

Regarding form, when the vowel-initial suffixes *-ard* and *-and* combine with *kā* ‘where’, an epenthetic consonant *-d-* appears between the interrogative pronoun and the suffix, hence *kā<sub>d</sub>ard* and *kā<sub>d</sub>and*, although both *kā<sub>d</sub>ndi* ‘where (at)’ and *kā<sub>d</sub>ra* ‘around where’ are also possible. The use of a consonant with these suffixes to resolve vowel hiatus is not uncommon and typically occurs after a long vowel (e.g. *malā-y-ard* ‘for the house’), but the fact that we get /d/ rather than the palatal glide /y/ is surprising. To my knowledge, in all cases of hiatus-resolving epenthesis other than with *kā*, the palatal glide is used. I take this anomaly, particularly in combination with these

suffixes' lack of complete productivity noted above, as evidence for the notion that they share a greater degree of fusion with *wh*-words than do the adpositions in *wh*-phrases discussed below.

Both the locative and dative suffixes, when combining with interrogative pronouns, have virtually the full range of meanings ascribed to these suffixes on non-*wh*-words (see Section 4.3.1 for an overview). In particular, the dative suffix is used both in a canonical case-like sense, as in *či-rd* 'to whom' and *čīz-ard* 'why' (lit. what-DAT), as well as in more idiosyncratic ways. For instance, it is used to request an approximate answer, such as in *kād-ard* 'around where', as well as to request a rate (e.g. *cūnd-ard* 'at how much (e.g. dollars per kilogram)'). The locative suffix, for its part, also has canonical possessive (e.g. *či-nd* 'whose') and locative functions (e.g. *kād-and* 'where'). Unlike the dative suffix on interrogative pronouns, the use of the locative suffix typically indicates a request for a precise response. Compare, for instance, *cūnd-ard* (DAT, 'around what time') and *cūnd-and* (LOC, 'at (precisely) what time').

The third and final type of complex interrogative pronoun to be discussed is that formed by adding adjective-forming derivational suffixes to an interrogative pronoun. Three derivational suffixes may participate in the formation of complex interrogative pronouns: the kind-forming suffix *-in*, temporal-adjective-forming suffix *-īnj*, and the diminutive suffix *-ik*.

The suffix *-in* combines with both *čīz* and *carāng* to form *čīzin* and *carāngin*. Both of these derived *wh*-words have the meaning 'what kind (of)', and *carāngin* has the additional meaning 'what color'. Examples of these complex *wh*-words in use are given in (88).

(88) **Complex *wh*-words *carāng-in* and *čīz-in***

- a. Tu-nd     **carāng-in** mošin?  
you-POSS how-ADJ car  
'What kind of car do you have?' OR 'What color car do you have?'
- b. Tu-nd     **čīz-in** žemper?  
you-POSS what-ADJ sweater  
'What kind of sweater do you have?' (i.e. what material is it made of?)

The suffix *-īnj*, which attaches to temporal adverbs and is found in expressions such as *biyorīnj axbor* 'yesterday's news', combines with *cawaxt* to form *cawaxtīnj* 'of when'. An example of this pronoun is shown in (89).

(89) **Complex *wh*-word *cawaxt-ĩñj* ‘from what time’?**

Yam **cawaxt-ĩñj** awqot?  
 this when-ADJ food  
 ‘When is this food from?’

Lastly, the diminutive *-ik* may attach to an interrogative pronoun to show endearment or compassion on the part of the speaker. This use is illustrated in (90). The former may be uttered, for instance, by a father asking his small child what she has done that day. The latter may be uttered by someone responding to a friend who has just indicated she has a headache.

(90) ***Wh*-words formed with the diminutive *-ik***

- a. Čĩz-**ik**=at      nur      čũd?  
       what-DIM=2SG today do.PST  
       ‘What (little things) did you do today’
- b. Čĩz-**ik**-ard    tu      kāl      dārð kiřt?  
       why-DIM-DAT your head pain do.PRS.3SG  
       ‘(Oh no!) Why do you have a headache?’

Note that the diminutive *-ik* exhibited in (90) should not be confused with the suffix *-ik* in *cũndik*, which inquires about degree, as in ‘how (big, tall, nice, etc.)’. Nor should it be confused with the question-final particle *-ik*, which is used to signal an echo question, as discussed in Section 12.3.3.

A summary of complex interrogative pronouns is given in Table 6.3.

**6.1.2.3 *Wh*-phrases with adpositions**

Both simplex and complex interrogative pronouns may combine with adpositions to form complex *wh*-phrases. Unlike the dative and locative suffixes discussed above, the combination of adpositions with interrogative pronouns is virtually unrestricted. Examples of complex *wh*-phrases with prepositions include *az kã* ‘from where’ and *tar čidũm čĩd* ‘to which house’. Examples of *wh*-phrases based on postpositions include *čĩz jãt* ‘why; what for’ and *čĩ qati* ‘with whom’. These phrases behave as a cohesive syntactic unit and follow the word order norms for *wh*-questions, as discussed in Section 12.3.

Table 6.3: Complex interrogative pronouns.

With bound morpheme <i>ca-</i>			
WH-WORD	GLOSS	EXAMPLE	GLOSS
<i>cawaxt</i>	‘when’	<i>Cawaxt=ta yādi?</i>	‘When are you coming?’
<i>carāng</i>	‘how’	<i>Tu tāt=at nān=en carāng?</i>	‘How are your parents?’
With dative <i>-a(rd)</i> and locative <i>-a(nd)</i>			
WH-WORD	GLOSS	EXAMPLE	GLOSS
<i>či-rd</i>	‘(to) whom’	<i>Čird=en wi pūl dākt?</i>	‘Who did they give the money to?’
<i>kād-ard</i>	‘(approx.) where’	<i>Tama=yet kādard vad?</i>	‘(Around) where were you?’
<i>cūnd-ard</i>	‘(approx.) what time’	<i>Cūndard=ta firāpen?</i>	‘(Around) what time will they arrive?’
	‘what price (per unit)’	<i>Cūndard=at wev zoxt?</i>	‘How much did you pay for them?’
<i>cūnd-and</i>	‘(exactly) what time’	<i>Cūndand ta mu-rd qīwi?</i>	‘What time will you call me?’
<i>či-nd</i>	‘whose’	<i>Yam kitob Čind?</i>	‘Whose book is this?’
<i>kād-and</i>	‘(exactly) where’	<i>Tu=t šič kādand?</i>	‘Where are you now?’
With derivational suffixes			
WH-WORD	GLOSS	EXAMPLE	GLOSS
<i>čīz-in</i>	‘what kind (of)’	<i>Tu-nd čīzin žemper?</i>	‘What kind of sweater do you have?’
<i>carāng-in</i>	‘what kind (of)’	<i>Tu-nd carāngin mošin?</i>	‘What kind of car do you have?’
	‘what color’	<i>Tund carāngin mošin?</i>	‘What color car do you have?’
<i>cawaxt-īnj</i>	‘from when’	<i>Yam cawaxtīnj garōā?</i>	‘When is this bread from?’
<i>čīz-ik</i>	‘what (DIM)’	<i>Čīzik=at nur čūd?</i>	‘What did you do today (endearingly)?’

Examples of *wh*-phrases with adpositions are given in (91). Note that in (91c), the *wh*-phrase involves a circumfix-like combination of a preposition and a suffix – *to . . . -ec*, which combines to mean ‘until’.

(91) *Wh*-phrases formed with *wh*-words and adpositions

- a. Tu=t biyor [tar kā] sat.  
you=2SG yesterday to where go.PST.F  
‘Where did you go yesterday.’
- b. Wam puc=i [čīz jāt] ispānski ziv xeyd?  
her son=3SG what for Spanish language study.PST  
‘What did her son study Spanish for?’
- c. Māš=ta [to cawaxt-ec] futbol bozi kin-ām.  
we=FAC until when-until football play do.PRS-1PL  
‘Until when will we play football?’

### 6.1.3 Indefinite pronouns

Indefinite pronouns in Shughni are most often predictably derived from interrogative pronouns.<sup>3</sup> The usage of indefinite pronouns is restricted based on both polarity (affirmative vs. negative) and mood (imperative and interrogative vs. otherwise). They can be divided into the following four types (with a corresponding example based on ‘something’, ‘anything’, etc.): (i) *assertive existential* – roughly ‘something’ – used in affirmative declarative contexts; (ii) *elective existential* – roughly ‘anything’ – used in interrogative and imperative contexts; (iii) *negative* – roughly ‘nothing’ – used in negative polarity contexts; and (iv) *universal* – roughly ‘everything’ – used in all three of the preceding contexts.

Assertive existential pronouns are formed with an interrogative pronoun with a prefixed *ar-* (lit. ‘every, each’) and a suffixed *-ca* (a complementizer and basic *wh*-word component), as in *ar-čīz-ca* ‘something’. Elective existential pronouns are formed with the indefinite determiner *yi-* prefixed to an interrogative pronoun, as in *yičīz* ‘anything’. Negative indefinite pronouns are identical to elective existential pronouns but may optionally take the augmentative suffix *-aθ*, as in *yičīz(aθ)* ‘nothing’. Finally, universal pronouns are formed with the quantifier *fuk* ‘all; every’ prefixed to the interrogative pronoun, with the augmentative suffix *-aθ* optionally suffixed to it, as in *fukčīz(aθ)* ‘everything’. The formation of the different types of indefinite pronouns is schematized in (92), and examples of each type of pronoun are given in (93):

(92) **Formation of indefinite pronouns**

a. **Assertive existential**

*ar* + *wh*-WORD + *ca*

c. **Negative existential**

*yi* + *wh*-WORD (+ *-aθ*)

b. **Elective existential**

*yi* + *wh*-WORD

d. **Universal**

*fuk* + *wh*-WORD (+ *-aθ*)

(93) a. **Assertive existential – *arčīzca* ‘something’**

Maŷdzūnj=um nist,     **arčīzca**=yum   xūd   ūži.  
 hungry=1SG   NEG.COP something=1SG eat.PST already  
 ‘I’m not hungry. I ate something already.’

<sup>3</sup>A notable exception to this tendency is that the interrogative pronoun *kā* ‘where’ is not used in the formation of indefinite pronouns, with *joy* ‘place’ used in its stead. Hence, we have *arjoyca* ‘somewhere’, rather than *\*arkaca*.

b. **Elective existential – *yičt̥z* ‘anything’**

Tu=t    **yičt̥z**    xūd    o?  
 you=2SG anything eat.PST PQ  
 ‘Did you eat anything?’

c. **Negative existential – *yičt̥z-aθ* ‘anything’**

Wuz=um **yičt̥z-aθ**    na-xūd.  
 I=1SG nothing-AUG NEG-eat.PST  
 ‘I didn’t eat anything (at all).’

d. **Universal– *fukičt̥z-aθ* ‘everything’**

Māš=ām **fukičt̥z-aθ**    xūd.  
 we=1PL everything-AUG eat.PST  
 ‘We ate everything.’

Table 6.4 shows some of the most commonly used indefinite pronouns in Shughni. Note that in the case of complex interrogative pronouns formed with *ca* + NOUN (e.g., *cawaxt* ‘when’), the *ca* is dropped in the formation of the universal indefinite pronoun – thus *fukwaxt(aθ)* ‘all the time’ and *fukrāng(aθ)* ‘in every way’.

**Table 6.4:** Common indefinite pronouns in Shughni.

BASE		ASSERTIVE EX.	ELECTIVE EX.	NEGATIVE	UNIVERSAL
<i>čt̥z</i>	‘what’	<b><i>arčt̥zca</i></b> ‘something’	<b><i>yičt̥z</i></b> ‘anything’	<b><i>yičt̥z(aθ)</i></b> ‘nothing’	<b><i>fukčt̥z(aθ)</i></b> ‘everything’
<i>čay</i>	‘who’	<b><i>arčayca</i></b> ‘someone’	<b><i>yičay</i></b> ‘anyone’	<b><i>yičay(aθ)</i></b> ‘no one’	<b><i>fukčay(aθ)</i></b> ‘everyone’
<i>carāng</i>	‘how’	<b><i>arcarāngca</i></b> ‘somehow’	<b><i>yicarāng</i></b> ‘in any way’	<b><i>yicarāng(aθ)</i></b> ‘in no way’	<b><i>fukrāng(aθ)</i></b> ‘in every way’
<i>cawaxt</i>	‘when’	<b><i>arcawaxtca</i></b> ‘sometime’	<b><i>yicawaxt</i></b> ‘at any time’	<b><i>yicawaxt(aθ)</i></b> ‘never’	<b><i>fukwaxt(aθ)</i></b> ‘all the time’
<i>joy</i>	‘place’	<b><i>arjoyca</i></b> ‘somewhere’	<b><i>yijoy</i></b> ‘anywhere’	<b><i>yijoy(aθ)</i></b> ‘nowhere’	<b><i>fukjoy(aθ)</i></b> ‘everywhere’

### 6.1.4 Reflexive and emphatic pronouns

The final type of pronoun to be discussed in this section is that of reflexive and emphatic pronouns, which roughly correspond to English ‘-self’ pronouns (e.g. *myself*, *herself*, etc). Shughni uses both (i) a subject-oriented oblique reflexive pronoun *xu* (glossed REFL), and (ii) an emphatic pronoun *xubaθ* (glossed PRON.EMPH). The former is comparable to reflexive uses of English ‘-self’ pronouns, such as in ‘you saw *yourself*’, and is used with possessors who are also subjects in the same clause), as in ‘John saw *his (own)* father’. The latter, *xubaθ*, is comparable to emphatic uses of English ‘-self’ pronouns, as in ‘I saw you *myself*’ or ‘I saw the president *himself*’. This subsection discusses each of these pronouns in turn, beginning in Section 6.1.4.1 with the reflexive *xu*, and then turning in Section 6.1.4.2 to the emphatic pronoun *xubaθ*.

#### 6.1.4.1 Reflexive oblique pronoun *xu*

The reflexive pronoun *xu* takes on virtually the same grammatical functions as personal and demonstrative oblique pronouns in Shughni. Each use is demonstrated in (94). It can function as a direct object (i.e. as the pronominal component of reflexive verbs), as in (94a); an indirect object with the dative suffix *-(a)rd*, as in (94b); the object of an adposition, as in (94c); or as a possessive adjective, as in (94d).

#### (94) Range of uses of the oblique anaphor *xu*

##### a. Direct object (of a reflexive verb)

Wuz=um tar yenak **xu** wīnt.  
I=1SG in mirror REFL see.PST  
‘I saw myself in the mirror.’

##### b. Indirect object

Wuz=um **xu**-rd awqot pêxt.  
I=1SG REFL-DAT food cook.PST  
‘I cooked myself some food.’

##### c. Object of an adposition

Wuz=um **xu** qati gāp ðod.  
I=1SG REFL with word hit.PST  
‘I spoke with myself.’

##### d. Possessive adjective

Wuz=um **xu** tāt stiklo-yand wīnt.  
I=1SG REFL father glass-LOC see.PST.  
‘I saw my father in the glass.’

*Xu* has the following five defining properties: (i) subject orientation, (ii) restriction to oblique position, (iii) lack of inflection for person, number, or gender, (iv) (partial) complementary distribution with pronominal and demonstrative pronouns, and (v) domain restriction (to the minimal clause of its co-indexed subject). Each are discussed



in turn below.

**1. Subject orientation.** *Xu* must have an antecedent which, with the possible exception of certain infinitival clauses (see fn. 4), is the grammatical subject of its local clause. This means that in all cases where *xu* appears in a finite clause, it must have an antecedent in the direct case. The subject-orientation of *xu* is exhibited in the examples in (97), in which it functions as a possessive adjective and is necessarily co-indexed with the subject. In examples like (97b), in which both the subject and oblique pronoun are third-person, the use of *xu* indicates that the identity of the oblique pronoun is the same as that of the subject pronoun. The use of a personal pronoun would indicate that the identity of the two are different.<sup>4</sup>

(97) **Subject orientation of the anaphor *xu***

- a. Wuz<sub>i</sub>=um tu<sub>j</sub>-rd **xu<sub>i</sub>**/<sub>\*j</sub> čĩd diviřt.  
 I=1SG YOU-DAT REFL house show.PST  
 ‘I showed you my (\*your) house.’

---

<sup>4</sup> A compelling area for future research regards the use of *xu* in the infinitival complement clauses of dative-first constructions, such as those with the verb *fort* ‘want; be desirous to’ (see Section 11.5 on oblique-first constructions). In these constructions, the dative argument – usually a semantic experiencer – precedes the nominative theme, and verbal agreement occurs with the latter. The status of either the dative or nominative arguments as the grammatical subject in such constructions remains an open question, however, and the reflexive pronoun *xu* may shed light on the matter. In particular, initial data suggest that if there is an oblique pronoun inside the infinitival complement clause of a dative-first construction, and this pronoun is co-referential with the dative experiencer, then it may be either the reflexive pronoun *xu* or a personal pronoun, as shown in (95).

(95) **Infinitival complement of dative-first construction: *xu* OR personal pronoun**

- Māř<sub>i</sub>-ard [māř<sub>i</sub> (**xu<sub>i</sub>**) tāt-en wĩnt-ow] fort.  
 1PL.OBL-DAT 1PL.OBL (/REFL) father-PL see-NMZ want.PRS.3SG  
 ‘We want to see our fathers.’

This contrasts with infinitival complement clauses in non-dative-first constructions, in which the use of the reflexive pronoun is obligatory if it is co-referential with the subject of the matrix clause. This is shown in (96) with the verb *kuřĩř čĩdow* ‘try’. Compare the obligatory use of *xu* in this example with its interchangeability with a personal pronoun in (95).

(96) **Infinitival complement of direct-first construction: *xu* only**

- Māř<sub>i</sub>=ta kuřĩř kin-ām [**xu<sub>i</sub>** (/māř<sub>i</sub>) tāt-en wĩnt-ow].  
 we=FAC try do.PRS-1PL REFL (/1PL.OBL) father-PL see-PURP  
 ‘We will try to see our fathers.’

It should be noted that this data is taken from elicited fieldwork, however, and the optionality of the personal pronoun and reflexive pronoun in examples such as (95) is not yet confirmed through natural speech.

- b. Davlat<sub>i</sub>=i Nekrūz<sub>j</sub>-ard **xu**<sub>i/\*j</sub> čīd divīxt.  
 Davlat=3SG Nekruz-DAT REFL house showed  
 ‘Davlat<sub>i</sub> showed Nekruz<sub>j</sub> his<sub>i/\*j</sub> house.’

In cases where the subject NP is complex – i.e. consists of more than one NP – *xu* must be co-indexed with the subject, rather than an NP inside the subject:

- (98) [Alī<sub>i</sub> tāt]<sub>j</sub>=i xu<sub>\*i/j</sub> čīd parđod.  
 Ali father=3SG REFL house sell.PST  
 ‘Ali’s<sub>i</sub> father<sub>j</sub> sold his<sub>\*i/j</sub> house.’

**2. Oblique position.** *Xu* must be in an oblique-case position. This restriction rules out its usage as a coordinated component of a subject NP, even when it is co-referential with the other coordinated component of the subject NP, as shown in (99):

- (99) [Wuz=at mu/(**\*xu**) tāt]<sub>SUBJ</sub> qatiyaθ zindagi kin-ām.  
 I=and my/(**\*REFL**) father together life do.PRS-1PL  
 ‘Me and my dad live together.’

Importantly, it is not the presence of *xu* in a coordinated NP which is at issue, but rather the fact that the coordinated NP is in a direct-case (subject) position. Consider example (100), in which the use of *xu* in an oblique-case coordinated NP is grammatical:

- (100) Tu=ta [mu=at **xu**/(**\*tu**) tāt]<sub>OBL</sub> qati zindagi kin-i.  
 you=FAC [me=and REFL/(**\*your**) dad] with life do.PRS-2SG  
 ‘You will live with me and your dad.’

**3. Lack of inflection for person, number, or gender.** *Xu* does not overtly agree with its antecedent in person, number, or gender. This is exhibited in the examples in (101). In (101a), it is co-indexed with a second-singular subject, while in (101b) it is co-indexed with a third-plural subject. In both instances, it has the same form:

(101) **Lack of inflection in the anaphor *xu***

- a. Tu=t     **xu**   nān   qati ar bozor vad.  
 you=2SG REFL mother with at market be.PST.F  
 ‘You were with your mom at the market.’
- b. Wāð=en **xu**   čīd-and   vad.  
 they=3PL REFL house-LOC be.PST.PL  
 ‘They were in their house.’

**4. (Partial) complementary distribution with pronominal and demonstrative pronouns.** In finite clauses, only the first-person-singular personal pronoun *mu* is fully interchangeable with the anaphor *xu* in contexts where the use of *xu* is grammatical (i.e. where it is in an oblique position and co-referential with the subject). Thus, either *xu* or *mu* are equally felicitous in the sentences in (102):

(102) **Interchangeability of *xu* and first-person singular oblique *mu***

- a. Wuz=um **xu/(mu)** tāt   ar bozor wīnt.  
 I=1SG     REFL/my father at market see.PST  
 ‘I saw my dad at the market.’
- b. Wuz=um **xu/(mu)** tāt   qati ar bozor sat.  
 I=1SG     REFL/my father with to market go.PST.F  
 ‘I went with my father to the market.’

However, when the subject is first-person-plural or second-person (singular or plural), only the reflexive pronoun *xu*, not a personal pronoun, may be used as a possessive pronoun or object co-indexed with the subject:

(103) **Lack of interchangeability of *xu* with non-1sg**

- a. Māš=ām **xu/(māš)** tāt-en   ar bozor wīnt.  
 we=1PL   REFL/(our) father-PL at market see.PST  
 ‘We saw our fathers at the market.’
- b. Tu=t     **xu/(tu)**   tāt-en   ar bozor wīnt.  
 you.SG=2SG REFL/(your) father-PL at market see.PST  
 ‘You (sg.) saw your father at the market.’
- c. Tama=yet **xu/(tama)**   tāt-en   ar bozor wīnt.  
 you.PL=2PL REFL/(your.PL) father-PL at market see.PST  
 ‘You (pl.) saw your fathers at the market.’

The choice of *xu* versus a personal pronoun is likewise inflexible in clauses where the subject is in the third person. If *xu* occurs in a clause with a third-person subject, it is obligatorily co-referential with the subject, whereas if a third-person oblique pronoun is used in a clause with a third-person subject, it is obligatorily *not* co-indexed to the subject:

(104) ***Xu* with third-person subjects**

- a.  $Y\bar{a}_i=yi$   **$xu_{i/*j}$**   $t\bar{a}t$  ar bozor  $w\bar{i}nt$ .  
 she=3SG REFL father at market see.PST  
 ‘She<sub>i</sub> saw her<sub>i/\*j</sub> father at the market.’
- b.  $Y\bar{a}_i=yi$   **$wam_{*i/j}$**   $t\bar{a}t$  ar bozor  $w\bar{i}nt$ .  
 she=3SG her father at market see.PST  
 ‘She<sub>i</sub> saw her<sub>\*i/j</sub> father at the market.’

**5. Domain restriction.** In finite clauses, the antecedent of *xu* must be the subject of its minimal clause. We have seen above that in simplex finite clauses, *xu* must be co-indexed with a direct-case argument. In complex clauses, *xu* must be co-indexed to the subject of whichever clause it appears in, whether that is the matrix clause or the subordinate clause. This is the case whether we are dealing with basic complement clauses, as in (105a), with adverbial clauses, as in (105b) or with relative clauses, as in (105c). Subordinate clauses are in square brackets.

(105) **Domain restriction of *xu* in complex clauses**

a. **Complement clause**

$Yu_i=yi$   $l\bar{u}d$  [ $i d_i$   $t u_j=t$   $xu_{*i/j}$   $k i t o b$   $\check{x} e y \check{y}$ ].  
 he=3SG say.PST COMP you=2SG REFL book read.PST  
 ‘He said that you read his(\*your) book.’

b. **Adverbial clause**

$F\bar{i}r\bar{u}z\bar{a}_i$   $d i s$   $x u \check{s}$   $v a d$ , [ $M a w l o d_j=i$   $x u_{*i/j}$   $k i t o b$   $c a$   $\check{x} e y d$ ].  
 Firuza so happy be.PST.F Mawlod=3SG REFL book SUBR read.PST  
 ‘Firuza was so happy when Mawlod read his(\*her) book.’

c. **Relative clause**

$W u z_i=u m$   $w a m$   $\check{y} i n i k_j$   $w \bar{i} n t$  [ $x u_{*i/j}$   $\check{c} i d=i$   $m u$   $t \bar{a} t-a r d$   $c a$   $p a r \check{d} o d$ ].  
 I=1SG 3SG.OBL.F woman see.PST REFL house=3SG 1SG.OBL dad-DAT REL sell.PST  
 ‘I saw the woman who sold her(\*my) house to my dad.’

### 6.1.4.2 Emphatic anaphor *xubaθ*

The emphatic anaphor *xubaθ* corresponds closely in meaning to the emphatic usage of ‘-self’ pronouns in English, which can be used in either subject (direct) position, as in ‘I *myself* went to the store’ (as opposed to someone else), or in object (oblique) position, as in ‘You know a lot about Montreal’s universities, but you don’t know much about the city *itself*’.

Several previous authors have described *xu* and *xubaθ* as two reflexive pronouns in complementary distribution with one another, where *xubaθ* is the direct (i.e. subject) pronoun and *xu* is the oblique pronoun (e.g. Bakhtibekov 1979: 30-31; Karamshoev 1988b: 224-226; Alamshoev 1994: 71-76). The analysis presented here therefore departs somewhat from those of these authors, in that *xubaθ* is taken not to be a direct-case version of *xu*, but rather an emphatic pronoun which serves an entirely different grammatical and pragmatic function than *xu*.

In Shughni, the most common use of the emphatic pronoun *xubaθ* is perhaps that in which it is co-indexed to the subject of a sentence and emphasizes that an action or state is true of the subject without intervention or interference from anyone or anything else. This is precisely how *xubaθ* is described by Alamshoev (1994: 71). Such usages of *xubaθ* can be appreciated in (106):

#### (106) Emphatic anaphor *xubaθ* in direct position

- a.  $Tu_i$  ***xubaθ\_i***    *bīdi*    *wizūn-i*.  
you PRON.EMPH better know.PRS-2SG  
‘You yourself know better.’ (Karamshoev 1988b: 228)
- b.  $(Tu_i)$  ***xubaθ\_i***    *sa*            *pis xūvd xu ya*.  
(you) PRON.EMPH go.2SG.IMP for milk and come.2SG.IMP  
‘Go get some milk yourself and come back.’ (Bakhtibekov 1979: 30)

Examples such as these, in which *xubaθ* is used to emphasize the subject of the clause, are indeed rather abundant in the language, but it is nonetheless possible for *xubaθ* to be co-indexed with a noun phrase which is in an oblique position, as shown in the sentences in (107). These examples show this pronoun as a direct object, as in (107a), and as the object of an adposition, as in (107b) and (107c):

(107) **Emphatic anaphor *xubaθ* in oblique position**

- a.  $Tu_i$   $tāt$   $at$   $nān=um$   $wīnt$ ,  $ammo$   $tu_i=yum$  ***xubaθ\_i***  $na-wīnt$ .  
 your father and mother=1SG see.PST but you=1SG PRON.EMPH NEG-see.PST  
 ‘I saw your mother and father, but I didn’t see you (yourself).’
- b.  $Wuz$   $dar.borai$   $di$   $ūniversitet$   $fām-um$ ,  $ammo$   $dar.borai$   $Makedoniya_i$  ***xubaθ\_i***  $yičīzaθ$   
 I about the university know.PRS-1SG, but about Macedonia PRON.EMPH nothing  
 $na-fām-um$ .  
 NEG-know.PRS-1SG  
 ‘I know about the university, but I don’t know anything about Macedonia itself.’
- c.  $Wi$   $oχno$   $qati=yum$   $ar$   $Dušanbe$   $vud$ ,  $ammo$   $wi_i$   $qati=yum$  ***xubaθ\_i***  $yamand$   $na-vud$ .  
 his friend with=1SG in Dushanbe was.PST.M, but him with=1SG PRON.EMPH there NEG-be.PST.M  
 ‘I was in Dushanbe with his friend, but I wasn’t with him (himself).’

I therefore conclude that a more accurate descriptive term for *xubaθ* is that of an emphatic anaphor, rather than a direct-case reflexive pronoun in complementary distribution with an oblique reflexive counterpart.

## 6.2 Demonstratives

This section presents demonstratives in Shughni. I adopt Diessel’s (1999: 2) definition of demonstratives as deictic expressions which have specific syntactic and semantic functions, and I make use of strategies outlined in Wilkins’ (2018) demonstratives questionnaire, particularly to understand exophoric uses of demonstratives in Shughni.

As deictics, demonstratives encode features of the context in which an utterance takes place “and thus also [concern] the ways in which the interpretation of utterances depends on the analysis of that context (Levinson 1983: 54)”. Demonstratives have multiple syntactic, semantic, and pragmatic uses, including to draw the hearer’s attention to objects or locations in the speech situation (i.e. *exophoric* uses), as well as to maintain a working record of prior discourse participants and refer to specific chunks of speech within the discourse (i.e. *endophoric* uses). (See Halliday & Hassan 1976 for more on exophoric and endophoric uses of demonstratives.) Both exophoric and endophoric uses of Shughni demonstratives are addressed in this section.

I also follow Diessel (1999: 57) in assuming that demonstratives have the following four fundamental syntactic uses: (i) *adnominal*; (ii) *pronominal*; (iii) *identificational*; and (iv) *adverbial*. This section is concerned only with

the first three uses, each of which is described and discussed below. Adverbial demonstratives, which are formed from distinct deictic stems, are presented in Section 8.4 in Chapter 8 on verbs.

The remainder of this section is organized as follows. First, Section 6.2.1 gives the forms and basic syntax of demonstratives in each of their adnominal, pronominal, and identificational uses. Then, Section 6.2.2 examines exophoric uses of demonstratives, including a description of the triple deictic system of Shughni demonstratives and the factors which contribute to whether a speaker uses a proximal, medial, or distal form in a given context. Section 6.2.3 looks at endophoric uses of Shughni demonstratives, including both anaphoric and discourse deictic uses, and finally, Section 6.2.4 looks at demonstrative presentatives in Shughni, which are similar in function to French *voici/voilà* and Russian *vot/von*.

### 6.2.1 Form and syntax of demonstratives

Shughni demonstratives exhibit a system of triple deixis with proximal, medial, and distal degrees. Three grammatical categories are morphologically distinguished: case, number, and in a subset of demonstratives, gender. Specifically, gender is distinguished in all singular oblique demonstratives, as well as in the singular direct distal forms.<sup>5</sup> Gender is not distinguished in any plural demonstratives or in the singular direct of proximal and medial grades. The paradigm of Shughni demonstratives is given in Table 6.5; cells where gender is distinguished are shaded.

**Table 6.5:** Demonstrative pronouns and grammatical gender.

	Proximal		Medial		Distal	
	DIR	OBL	DIR	OBL	DIR	OBL
SG.F						
SG.M	<i>yam</i>	<i>mam</i> <i>mi</i>	<i>yid</i>	<i>dam</i> <i>di</i>	<i>yā</i> <i>yu</i>	<i>wam</i> <i>wi</i>
PL	<i>māō</i>	<i>mev</i>	<i>dāō</i>	<i>dev</i>	<i>wāō</i>	<i>wev</i>

The demonstratives given in Table 6.5 are used in three syntactic contexts: (i) *pronominally*, as the sole element in

<sup>5</sup>Shughni, including its Bajuwī dialect, is unique among languages of the Shughni-Rushani group in having a pair of gender-distinguishing direct pronouns *yā* (F) and *yu* (M). All other varieties of the group, with the exception of Sarikoli, which does not have grammatical gender, distinguish gender only in the oblique forms of demonstrative pronouns.

a noun phrase; (ii) *adnominally*, as a modifying element of a complex noun phrase; and (iii) *identificational*, in copular and nonverbal clauses (cf. English ‘*this* is my father’). Note that whereas some languages employ separate forms for each of these three uses (Diessel 1999: 57-92), the same Shughni demonstrative forms are used in each of these contexts. Examples of pronominal, adnominal, and identificational uses of Shughni demonstratives are given in examples (108). The demonstrative in each example is bolded, and the noun phrase in which it is contained is in square brackets.

(108) **Demonstratives in different syntactic uses**

a. **Pronominal**

[**Yid**            mešāt kiřt!  
DEM.DIR.MED bother do.3SG.PRS  
‘That (e.g. light) is bothersome!’

b. **Adnominal**

[**Mev**            toθč-en]=at tozā čūd o?  
DEM.OBL.PROX.PL bowl-PL=2SG clean do.PST PQ  
‘Did you clean these bowls?’

c. **Identificational**

[**Yam**            mu ořno.  
DEM.DIR.PROX my friend  
‘This is my friend.’

Note that Shughni demonstratives in their adnominal usage are compatible with possessive adjectives, as shown in (109).

(109) **Demonstrative with possessive adjective**

[**Yid**            tu tilifũn] cũnd sũm vad?  
DEM.DIR.MED your telephone how.many Somoni be.PST.F  
‘How many Somoni (currency) was your phone?’ (lit. how many Somoni was that your phone)

The following subsection turns to exophoric uses of Shughni demonstratives, including a detailed look at the triple deictic system and the contextual factors determining the use of proximal, medial, and distal forms.



### 6.2.2 Exophoric uses of demonstratives

In their exophoric uses, demonstratives refer to non-linguistic entities – in the case of Shughni, both objects and people – in the context in which an utterance occurs. As noted above, Shughni uses a system of triple deixis with proximal, medial, and distal forms. The medial form in Shughni is at its core *person-oriented*, rather than *distance-oriented*. That is, if an object is perceived to be in the personal space of the addressee, especially if it is also not within the personal space of the speaker, then the medial form is virtually always used.<sup>6</sup> In this way, the Shughni demonstrative system resembles those of Japanese (e.g. Kuno 1973) and Korean (e.g. Sohn 1994), which are also person-oriented, rather than that of Spanish, which is distance-oriented (e.g. Anderson & Keenan 1985: 283-285).

However, as will be seen below, the medial form in Shughni is not confined to use in situations where the referent is located within the personal space of the addressee and outside that of the speaker. Instead, Shughni also appears to make use of the notion of a *shared interactional space* between speaker and addressee. In a speech situation in which speaker and addressee are physically present together, the shared interactional space is roughly the area between them and their shared field of vision. If the referent is perceived to be within this space (e.g., if it is between and/or visible to both speaker and addressee), it may also be referred to with a medial form even if it is not within the personal space of the addressee. In such cases, additional factors such as distance and visibility of the referent, as well as the attention of the addressee, are also at play. The idea of the shared interactional space, as well as the way it factors into the choice of demonstrative degree, will become clearer in the following discussion.

In the remainder of this subsection, I describe the kinds of contexts in which each of the proximal, medial, and distal demonstrative forms in Shughni are used exophorically. The discussion is divided into non-contrastive exophoric uses of demonstratives (Section 6.2.2.1) and contrastive exophoric uses of demonstratives (Section 6.2.2.2). This description of Shughni demonstratives is informed both by fieldwork with native speakers, as well as by my observations of naturally occurring speech over the course of several months in the field.

---

<sup>6</sup>It can be said that if speaker and addressee are relatively close to one another – say, within two or three meters – then any object within reach of the speaker is in the personal space of the speaker (and is thus compatible with a proximal demonstrative), while any object within the reach of the addressee is within the personal space of the addressee (and is thus compatible with a medial demonstrative). As speaker and addressee get farther from one another, the personal space associated with each also extends outward, such that if they are at some distance from one another, an object may be conceived of as within one's personal space even if it is not within reach. The linguistic upshot of such a relative notion of personal space is that as speaker and addressee become farther from one another, proximal and medial demonstratives forms are more readily used to refer to entities which are farther from the speaker and addressee, respectively.

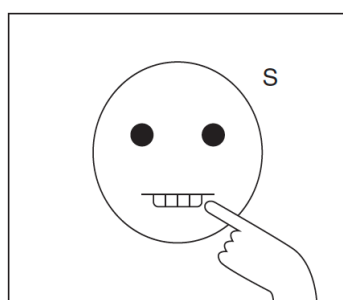
### 6.2.2.1 Non-contrastive exophoric uses of demonstratives.

I take non-contrastive uses of demonstratives to be scenarios in which a demonstrative is used to pick out a referent in the speech situation *without* distinguishing it from or comparing it to other referents within the same speech situation. This opposes contrastive uses of demonstratives, examined in Section 6.2.2.2 below, in which multiple referents are compared to one another.

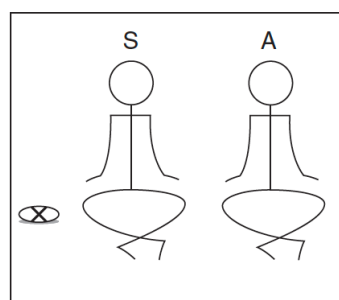
Within non-contrastive exophoric uses, each demonstrative degree generally corresponds to both “canonical” contexts, in which its use is obligatory, and “peripheral” contexts, in which the use of a demonstrative of a different degree may also be possible. Each type of situation is addressed here. The discussion and presentation of data here is inspired in large part by the demonstrative questionnaire developed by Wilkins (2018). Throughout the remainder of the subsection, I will use certain images from this questionnaire to distinguish between different demonstrative uses.

**Proximal forms of demonstratives**, in their non-contrastive uses, canonically refer to entities (including people) which are perceived as being within the personal space of the speaker. The use of a proximal form is especially clear-cut when the referent is both within the personal space of the speaker and outside that of the addressee. Such contexts are illustrated by the images in Figure 6.1. Examples corresponding to each are given in (110).

**Figure 6.1:** Canonical contexts for the use of proximal demonstratives (Image source: Wilkins 2018)



**Context 1**



**Context 2**

(110) a. **Sentence for Context 1**

**Yam**            mu ðindũn darð kiřt.  
 DEM.DIR.PROX my tooth    pain do.3SG  
 ‘This tooth hurts.’ (lit. this my tooth hurts)

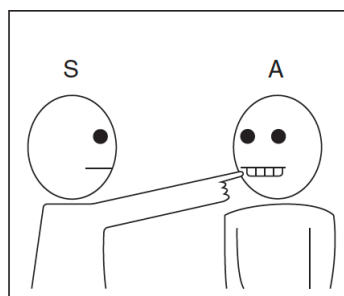
b. **Sentence for Context 2**

**Mam**            kitob=at    řeyř    o?  
 DEM.OBL.PROX.F book=2SG read.PRF PQ  
 ‘Have you read this book?’

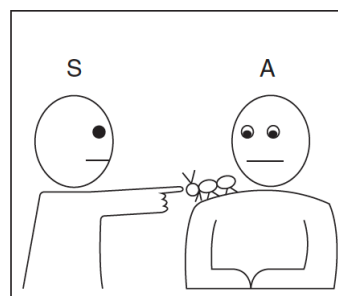
Speakers are virtually required to use a proximal form when referring to their own body parts, and hence, in Context 1 above, the use of a medial or distal form in a non-contrastive situation is illicit, regardless of additional factors such as the attention of the speaker. The use of any form other than the proximal is likewise not permitted in a (non-contrastive) situation like that exhibited in Context 2, where the object is clearly within the personal space of the speaker but not that of the addressee.

Even in cases where the referent is within the personal space of the addressee, a proximal form may be used if the speaker makes gestures, such as pointing with the finger or face, which bring her closer to contact with the referent. Thus, in Contexts 3 and 4, shown in Figure 6.2, both proximal and medial forms are possible. In both, a proximal form becomes more likely the closer the speaker's pointing gesture brings her to the referent. If the speaker's finger is contact with the referent, only the proximal is possible. Example sentences corresponding to Contexts 3 and 4 are given in (111).

**Figure 6.2:** Peripheral contexts for the use of proximal demonstratives.



**Context 3**



**Context 4**

(111) a. **Sentence for Context 3**

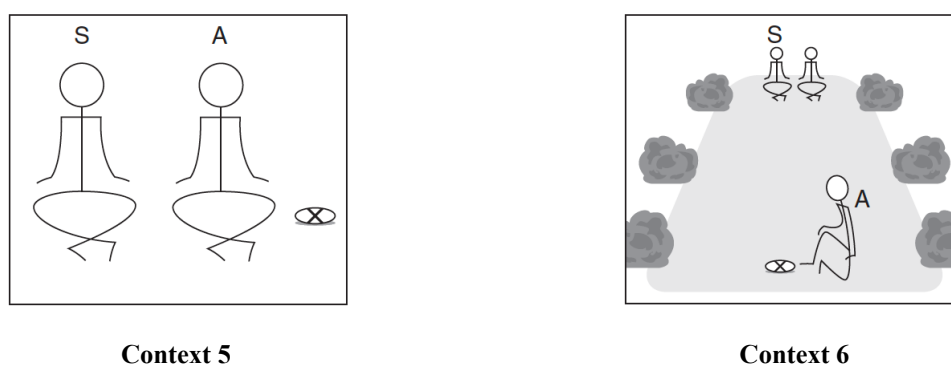
**Yam**            **(yid)**            tu    ðindûn viriřcin.  
 DEM.DIR.PROX (DEM.DIR.MED) your tooth    broken.F  
 ‘This (that) tooth is broken.’ (lit. this/that your tooth is broken)

b. **Sentence for Context 4**

**Yam**            **(yid)**            mürdzak=i tu    garang čūd    o?  
 DEM.DIR.PROX (DEM.DIR.MED) ant=3SG    you upset    do.PST PQ  
 ‘Is this (that) ant bothering you?’

**Medial forms of demonstratives**, for their part, are canonically used when the referent is perceived as being within the personal space of the addressee and outside that of the speaker. Contexts 3 and 4 above are therefore possible situations in which a medial form is felicitous, particularly if the speaker does not bring her hand into contact with the referent. Speech situations in which the use of a medial form is even more clear-cut are given in Figure 6.3. In these scenarios, the distance between the addressee and the referent is roughly the same, but the distance between the speaker and the referent is much greater in Context 6 than in Context 5. In both scenarios, however, only the use of the medial form is possible. These situations thus illustrate the notion that the medial form in the Shughni demonstrative system is fundamentally person-oriented, rather than distance oriented.

**Figure 6.3:** Canonical contexts for the use of medial demonstratives.



(112) a. **Sentence for Context 5**

**Yid** (\*yam, \*yā) lap xušrūy tilifūn. Yid tu-nd o?  
 DEM.DIR.MED (\*PROX, \*DIST) very beautiful telephone DEM.DIR.MED you-POSS PQ  
 ‘That’s a beautiful phone. Is it yours?’

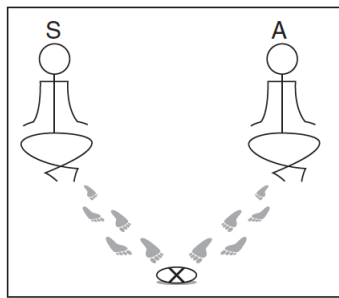
b. **Sentence for Context 6**

Jamšed, ku **dam** (\*mam, \*wam) xu pūt va.  
 Jamshed please DEM.OBL.MED.F (\*PROX, \*DIST) REFL ball bring.2SG.IMP  
 ‘Jamshed, please bring your ball (here).’

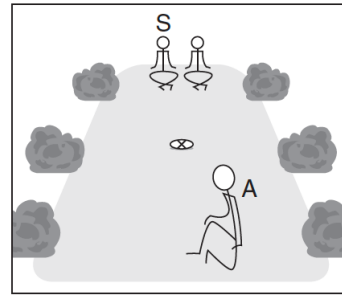
In addition to situations like those in Contexts 5 and 6, however, medial demonstrative forms may also be used if the referent is perceived to be within the shared interactional space of the speaker and addressee, but not necessarily within the personal space of either. This can be seen in situations such as those in Figure 6.4. In the speech

situation illustrated in Context 7, the referent is located at a roughly equal distance from speaker and addressee, but not within immediate reach. Here, either a medial or distal form is much more likely than a proximal form, and the choice between a medial and distal form depends on at least two factors: (i) the distance between the referent, on the one hand, and the speaker and addressee, on the other, and (ii) the attention of the addressee. A distal demonstrative becomes more likely as the referent becomes farther away, and a distal form is more likely if the speaker believes the addressee's attention is not yet on the referent. In Context 8, where the referent is clearly in between speaker and addressee, the medial form is the only choice. Examples for each context are given in (113).

**Figure 6.4:** Peripheral contexts for the use of medial demonstratives.



**Context 7**



**Context 8**

(113) a. **Sentence for Context 7**

**Dam** (wam) pūt qatī=ta bozi-yām, ani?  
 DEM.DIR.MED (DIST) ball with=FAC play-1PL right?  
 ‘We’re going to play with that ball, right?’

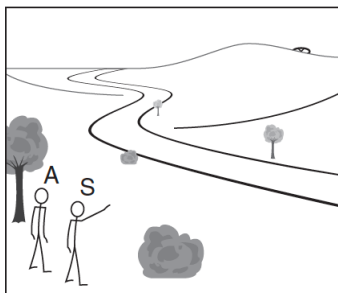
b. **Distal demonstrative in relative clause Sentence for Context 8**

Daler, **dī** (\*mī, \*wī) xu kud čis!  
 Daler DEM.OBL.MED.M (\*PROX, \*DIST) REFL dog look.at.2SG.IMP  
 ‘Daler, look at your dog!’

Finally, **distal forms of demonstratives** are canonically used when the referent is not physically present in the speech situation and not visible to either speaker or addressee. Such a situation is exhibited by Context 9 in Figure 6.5, with the corresponding example sentence in (114a). In the same vein, distal demonstratives are commonly

used with the head of a restrictive relative clause whose referent is not present in the speech situation. Such an example is given in (114b).

**Figure 6.5:** Canonical context for the use of distal demonstratives.



**Context 9**

(114) a. **Sentence for Context 9**

Tu=t pi **wi** (\*mi, \*di) puxtā sifīc o?  
 you=2SG up.to DEM.OBL.DIST.M (\*PROX, \*DIST) mountain.meadow go.up.PRF.F PQ  
 ‘Have you gone up to that mountain meadow before?’ (meadow is not visible to speaker or addressee)

- b. **Wam** ŷinik=um wuz biyor ca wīnt yā az Xaray vad.  
 DEM.OBL.DIST.F woman=1SG I yesterday REL see.PST DEM.DIR.DIST.F from Khorog be.PST.F  
 ‘That woman I saw yesterday was from Khorog.’

Summarizing this subsection, the primary factors which contribute to the choice of demonstrative degree are the following: (i) the personal space of speaker and addressee, (ii) the shared interactional space between speaker and addressee, (iii) the distance of the referent from speaker and addressee, and (iv) attention of the addressee. Most fundamentally, proximal and medial demonstrative forms are reserved for referents which are perceived as being in the personal space of the speaker and addressee, respectively. In cases where the referent is outside the speaker’s personal space, gestures such as pointing with the finger or face contribute to the availability of a proximal form. In addition to the personal spaces of speaker and addressee, Shughni also makes use of a shared interactional space between the two. Objects which are readily visible to both speaker and addressee, especially if they are located between the two, are generally perceived as being within this space. In cases where the referent is within this shared interactional space, but not within the personal space of either speaker or addressee, the use of either a

medial or distal form is likely. Medial forms are preferred when a speaker believes the addressee's attention is already on the referent, and a distal form is preferred if not. Distal forms are also preferred when a referent is not readily visible to speaker or addressee. This information is presented in the form of a decision table in Figure 6.6.

### 6.2.2.2 Contrast and precision in exophoric uses of demonstratives

This subsection explores contrastive exophoric uses of demonstratives in Shughni – that is, instances in which the referent is distinguished from or compared to other entities in the speech situation which either have been or will be referred to in the surrounding discourse. I also present here the precision particle *ik*, which, among other uses, precedes demonstratives and aids in picking out a referent from a set of possible referents.

There are two common phenomena found in the contrastive use of demonstratives in Shughni. First, demonstratives may be used contrastively without any additional morphological material. In such cases, the typical criteria used for selecting a demonstrative degree in non-contrastive environments, as discussed above in Section 6.2.2.1, are relaxed, and the most important criteria become the relative distance of each contrasted referent from the deictic center (i.e. the speaker). Thus, if something is predicated about three referents in succession, the proximal must be used for the referent nearest the deictic center (i.e. the speaker), the medial for the second closest, and the distal for the farthest. This is the case even in cases where the farthest referent is only a few meters away from and clearly visible to the speaker.

This is exhibited in (115). In this example, the speaker and addressee are organizing a group of books. They are sitting next to each other, and three books lie in front of them at different distances. The speaker utters the sentence in (115) in response to a question by the addressee inquiring whether the speaker has read these books.

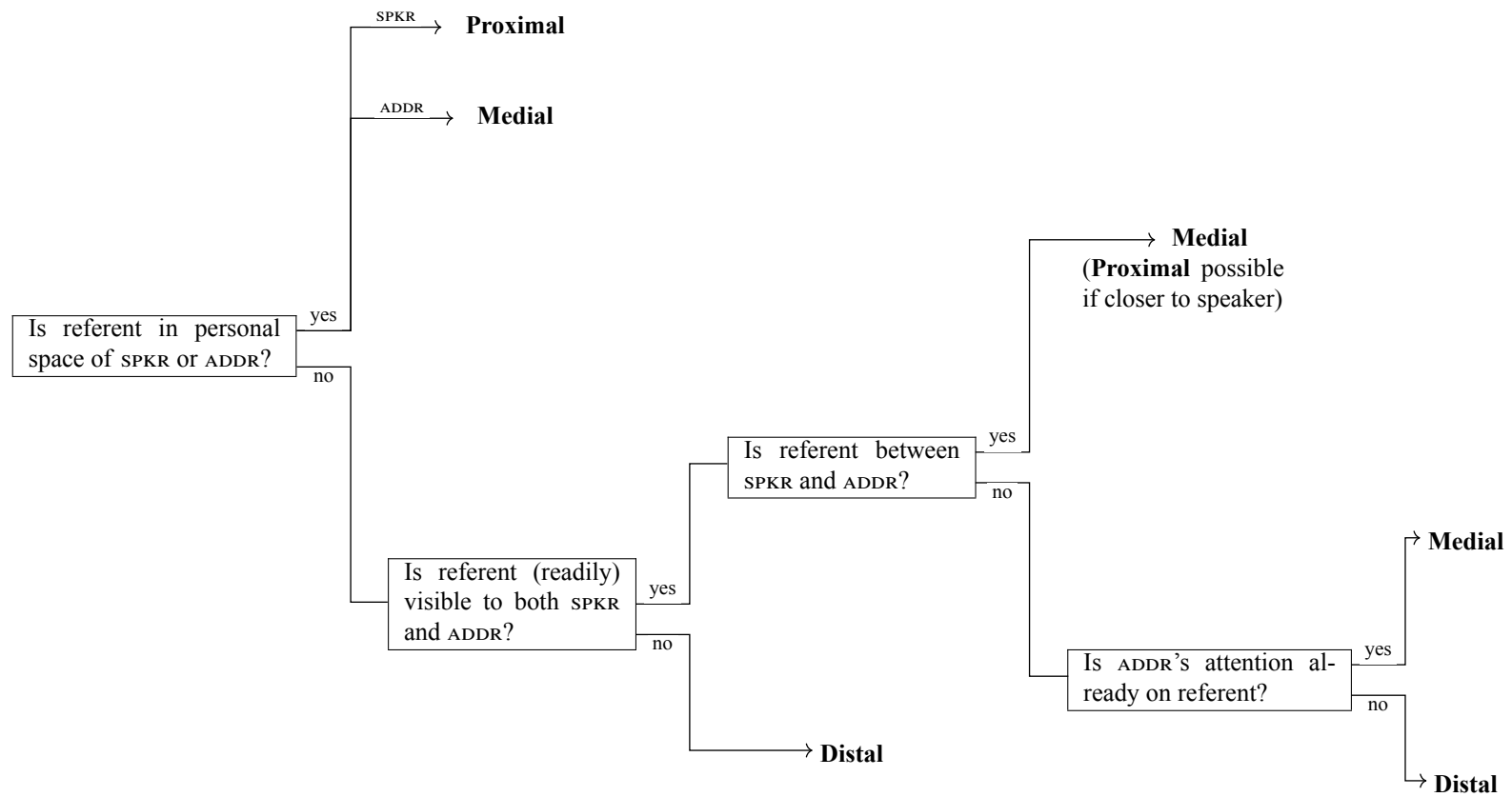
#### (115) Contrastive use of demonstratives without additional morphological material

**Mam**=um      **ǰeyj**=atā      **dam**=um      na-**ǰeyj**=at      **wam**=um      mis  
 DEM.OBL.DIST.F=1SG read.PRF=but DEM.OBL.MED.F=1SG NEG-read.PRF=and DEM.OBL.DIST.F=3SG also  
 na-**ǰeyj**.  
 NEG-read.PRF

‘I’ve read this (closest) one, but I haven’t read that one (in the middle), and I also haven’t read that (farthest) one.’

Note that in the context in which the sentence in (115) is produced, if the speaker were to refer to the farthest

**Figure 6.6:** Decision tree: Non-contrastive uses of exophoric demonstratives





book non-contrastively, the medial form, rather than the distal form, would most likely be used. This is because the referent is within the shared interactional space of the speaker and addressee, and the addressee's attention is already on the referent.

In fact, even in a contrastive environment, the use of a medial form is possible if additional material is used, namely the particle *iga* 'other'. This is the second common phenomenon found in contrastive uses of demonstratives. In a contrastive environment, a demonstrative of one degree, such as *dam* in the example in (115), may be followed by a demonstrative of the same degree, as long as the second instance is followed by the particle *iga*. Such an example is shown in (116).

(116) **Contrastive use of demonstratives without additional morphological material**

**Mam**=um      **ǰeyj**=atā      **dam**=um      na-**ǰeyj**=at      **dam**      **iga**=yum  
 DEM.OBL.DIST.F=1 SG read.PRF=but DEM.OBL.MED.F=1 SG NEG-read.PRF=and DEM.OBL.MED.F other=1 SG  
 mis na-ǰeyj.  
 also NEG-read.PRF  
 'I've read this (closest) one, but I haven't read that (medial), and I also haven't read that other (medial) one.'

In addition to contrast, the notions of *precision* or *emphasis*, roughly glossed 'this/that very', may also be marked grammatically with demonstratives via the particle *ik* (glossed *PREC*), which immediately precedes the demonstrative it modifies (cf. Edelman & Dodykhudoeva 2009b: 794). This same particle also occurs in certain instances where demonstratives are used anaphorically, in which cases it receives a gloss of 'this/that same'. The compatibility of *ik* with anaphoric uses of demonstratives is discussed in more detail in Section 6.2.3 below on endophoric uses of demonstratives.

The particle *ik* is syntactically noteworthy in that it appears to cliticize to all glide-initial demonstratives, namely the direct forms *yam* (PROX), *yid* (MED), *yā* (DIST.F) and *yu* (DIST.M). Here, the initial glide is deleted (e.g. *ik* + *yam* > *ik=am*), and the *k* is resyllabified as the onset of the second syllable. The forms of each demonstrative together with the precision marker *ik* are given in Table 6.6. Forms in which *ik* cliticizes to the demonstrative are shaded.

When appearing with a demonstrative used exophorically, *ik* is used to aid in picking a referent out of a set of other possible referents in the speech situation. Thus, a sentence like the one in (117) is likely to be uttered in a context where there are several other books lying around, and it is likely to be accompanied by a pointing gesture.

**Table 6.6:** Demonstrative pronouns and the precision particle *ik*

	Proximal		Medial		Distal	
	DIR	OBL	DIR	OBL	DIR	OBL
SG.F	<i>ik=am</i>	<i>ik mam</i>	<i>ik=id</i>	<i>ik dam</i>	<i>ik=ā</i>	<i>wam</i>
SG.M		<i>ik mi</i>		<i>ik di</i>		<i>ik wi</i>
PL	<i>ik māḍ</i>	<i>ik mev</i>	<i>ik dāḍ</i>	<i>ik dev</i>	<i>ik wāḍ</i>	<i>ik wev</i>

**(117) Precision particle *ik* with exophoric demonstrative**

**Ik=id**                      kitob tu-nd      o?  
 PREC=DEM.DIR.MED book you-POSS PQ  
 ‘Is that book (right there) yours?’

This particle is remarkably productive in Shughni. In addition to its compatibility with both exophoric and endophoric demonstratives, it is also used to mark precision with demonstrative adverbs, as in *tarūd* ‘here’ > *ik-tarūd* ‘(right) here’, and with other deictic elements such as *ik-šič* ‘right now’ and *ik-nur* ‘(exactly) today’.

**6.2.3 Endophoric uses of demonstratives**

Endophoric uses of demonstratives in Shughni are divided into two types: anaphoric and discourse deictic. *Anaphoric* demonstratives, examined in Section 6.2.3.1, are used to refer to an NP which was previously mentioned in the discourse and to “keep track of discourse participants” (Diessel 1999: 93). *Discourse deictic* demonstratives, examined in Section 6.2.3.2, are used to refer to another chunk of speech within the discourse, which either precedes or follows the utterance in which the demonstrative is contained (see Diessel 1999; Himmelmann 1996, 1997; and Levinson 1983, a.o., on endophoric uses of demonstratives). This subsection also looks in Section 6.2.3.3 at *pure text deictic* uses of demonstratives. In this type of use, which Diessel (1999: 101) categorizes as a “special instance of the exophoric usage”, a demonstrative is used to refer to “the material side of” the language, rather than to some aspect of meaning, as in discourse deictic uses (see also Lyons 1977: 668 on pure text deixics).

### 6.2.3.1 Anaphoric uses of demonstratives

Shughni, like all other Pamir languages except Yazghulami, lacks specialized third-person pronouns, and distal forms are generally used in cases where a third-person pronoun would be used in English (Edelman & Dodykhudoeva 2009b: 794; Edelman & Dodykhudoeva 2009a: 781). The anaphoric use of a demonstrative is exhibited in the example sentences in (118). In (118a), there is only a single discourse participant in the first clause, and there is no question about the identity of the referent of the third-singular masculine distal form *yu* in the second clause. In (118b), however, there are two male discourse participants in the first clause and thus two possible referents for the same demonstrative *yu* in the second clause. As the example shows, however, the distal form is required regardless of which discourse participant is the referent. In general, Shughni does not use a distinct demonstrative to mark the introduction of a new topic versus the continuation of an already established topic.

#### (118) Anaphoric uses of third-singular demonstratives

- a. Wuz=um biyor pi Nuriddin]<sub>i</sub> xez sut. **Yu<sub>i</sub>** (\*yam, \*yid) yi mošin soz  
 I=1SG yesterday up.at Nuriddin location go.PST.M DEM.DIR.DIST.M (\*PROX, \*MED) a car fixed  
 čīd-ow-and vud.  
 do.INF-NMZ-LOC be.PST  
 ‘I went to Nuriddin’s yesterday. He was repairing a car.’
- b. Nuriddin<sub>i</sub> biyor pi Mawlonazar<sub>j</sub> xez sut. **Yu<sub>i/j</sub>** (\*yam, \*yid) kor  
 Nuriddin yesterday up.at Mawlonazar location go.PST.M DEM.DIR.DIST.M (\*PROX, \*MED) work  
 čīd-and vud.  
 do.INF-LOC be.PST.M  
 ‘Nuriddin went to Mawlonazar’s yesterday. He was working.’

Nonetheless, two aspects of the *demonstrative-as-3sg-pronoun* phenomenon in Shughni are noteworthy. First, in some cases where a third-person pronoun would be used in a language like English, it is not the distal form which is used in Shughni, but rather the medial or proximal form. This occurs when the referent, already mentioned in the discourse, is physically present in the speech situation and the spatial arrangement of speech participants with respect to the referent is such that a medial or proximal demonstrative would be used exophorically. Here, the use of the medial or proximal form is required. In other words, in contexts where a demonstrative could be used either exophorically, making use of the physical position of speech participants and referent(s), or anaphorically, making use of the previous mention of an NP, it is the form required by the exophoric use which is selected. This is shown in example (119), which occurs in a context with four speech participants, including the speaker, the addressee,

and the referents, *mu tāt=at nān* ‘my mother and father’, who are physically present and part of the conversation. Note the gloss of the medial demonstrative at the start of the second clause as the English third-person pronoun *they*.

(119) **Ambiguous exophoric/anaphoric use of a demonstrative**

Mu [tāt=at nān]<sub>i</sub>=ta yičīzaθ na-xen. **Dāð<sub>i</sub>**=en (?wāð<sub>i</sub>=en) maŷdzūnǝ nist.  
 my father=and mother=FAC nothing NEG-eat.PRS.3PL DEM.DIR.MED.PL (?DIST=3PL) hungry NEG.COP  
 ‘My parents won’t eat anything. **They** aren’t hungry.’

And second, the use of a medial demonstrative anaphorically, and in fact the use of the medial more generally, appears to be more widespread in the narrative style of speech, in comparison with other styles of speech, where the distal form is more common. A sample of a narrative is provided in (120), in which a series of hypothetical elements is listed, also using the medial, and then referred to anaphorically with the medial demonstrative plural medial form *ik-dāð*.

(120) **Narrative excerpt with medial demonstrative used anaphorically**

Māš-and kadūm sanad-i tārīxi, **yid** čidūm podxo vuðǝ, **di** nūm čāy  
 we-POSS any document-EZ historical DEM.DIR.MED which king be.PRF.M DEM.OBL.MED.M name who  
 vuðǝ, cawaxt vuðǝ, čidūm odam-ti **ik=id** kor suðǝ yo na-suðǝ,  
 be.PRF.M when be.PST.M which person-LOC PREC=DEM.DIR.MED work become.PRF.M OR NEG-PRF.M  
**ik=dāð** asnod-en=en māš-and nist.  
 PREC=DEM.DIR.MED.PL document.PL-PL=3PL.COP we-POSS NEG.COP  
 ‘We don’t have any document (saying) which king this was, what was his name, when it was, which person  
 this thing happened or didn’t happen to. We don’t have those documents.’

The common use of medial demonstratives anaphorically is not the only distinguishing feature of narrative-style speech in Shughni. Note also the use of the repeated use of the perfect verb forms *vuðǝ* ‘be.PRF.M’ and *suðǝ* ‘go.PRF.M’. The use of perfect verb stems in the place of past verb stems is another feature of the narrative speech style in Shughni and may be connected to the use of the perfect as an evidential marker expressing indirect evidence (see Section 10.2.3 in Chapter 10 on TAM and Evidentiality). The widespread use of medial demonstratives in narrative speech will not be explored further in this thesis, but I wish to note it as a potentially fruitful topic for future investigation.

### 6.2.3.2 Discourse deictic uses of demonstratives

When used as discourse deictics, demonstratives in Shughni virtually always appear in a singular masculine form. The singular masculine medial form *di* is most commonly used, but the distal form *wi* is also possible in certain scenarios, as discussed below. The proximal form *mi* is virtually never used in this capacity.

First, Shughni uses what might be analyzed as a fixed (if not monomorphemic) discourse deictic in *di-jāt* ‘because; for the following reason’ (lit. DEM.OBL.MED.M-for), which commonly also takes the precision marker *ik-*, as in *ik-di-jāt* ‘for that (very) reason’. This form can be used either between the reason and consequent clauses, as in (121a) or preceding the reason clause, in which case a consequent clause beginning with the complementizer *idi* is anticipated, as in (121b).

(121) **(Ik=)di-jāt ‘because; for that reason’**

- a. Wuz=ta der az kor yad-um, (ik=)di-jāt=ta tar večer na-sām.  
 I=FAC late from work come.PRS-1SG (PREC=)DEM.OBL.MED.M-for=FAC to party NEG-go.PRS.1SG  
 ‘I’m going to come home from work late; that’s why I won’t go to the party.’
- b. (Ik=)di-jāt=ta tar večer na-sām idi wuz=ta az kor der  
 (PREC=)DEM.OBL.MED.M-for=FAC to party NEG-go.PRS.1SG COMP I=FAC from work late  
 yad-um.  
 come.PRS-1SG  
 ‘I’m not going to the party because I will come home from work late.’

Discourse deictics also appear in constructions which I call *anticipatory discourse deictic constructions*. Here, the demonstrative appears in the matrix clause and precedes an embedded clause, the information in which is the referent of the demonstrative. Demonstratives in this type of construction are therefore used cataphorically. In general, a medial masculine form *yid* (DIR) / *di* (OBL) is used when the action or state denoted by the verb in the matrix clause holds in the present or future, while the distal masculine form *yu* (DIR) / *wi* (OBL) is used when it holds in the past. This construction is schematized in (122)

(122) **Anticipatory discourse deictic construction**

DEM<sub>i</sub> . . . V [ EMBEDDED CLAUSE ]<sub>i</sub>

Examples are given in (123). Note that in (123a), it is the fact that the speaker’s state of knowing takes place in the

present which requires the use of the medial form. In (123b), the knowing instead takes place in the past (cf. *fāmt*, the past stem of *fāmtow* ‘know’), and the distal form *wi* is therefore required. In each example, the demonstrative used as a discourse deictic is bolded.

(123) **Anticipatory discourse deictic uses of oblique demonstratives**

a. **Anticipatory *dī* (MED)**

**Di** (\*mi, \*wi) mis fām-um, gīr sit wi dow-aθ  
 DEM.OBL.MED.M (\*PROX, \*DIST) also know.PRES-1SG stop become.INF DEM.OBL.DIST.M intention-AUG  
 nist.  
 COP.NEG

‘I also know that it is not his intention to stop.’ (lit. I already know it, stopping is not his intention)

b. **Anticipatory *wi* (DIST)**

**Wi**=yum (\*mi=yum, \*di=yum) mis fāmt, gīr sit wi  
 DEM.OBL.DIST.M=1SG (\*PROX=1SG, \*MED=1SG) also know.PST stop become.INF DEM.OBL.DIST.M  
 dow-aθ nist.  
 intention-AUG COP.NEG

‘I also knew that it isn’t (wasn’t) his intention to stop.’ (lit. I already knew it, stopping was not his intention)

Discourse deictic forms in Shughni are not only used in anticipatory constructions of the kind shown in (123). They may also be used anaphorically to refer to the information in a preceding clause, as shown in (124). Regarding the use of medial and distal forms, these examples follow the same pattern as those above. That is, in (124a), the medial form is required because the event predicated in the clause containing the discourse deictic (i.e. the speaker’s becoming angry) will hold in the future. In (124b), on the other hand, the event predicated in this clause is in the past. Hence the use of a medial demonstrative in the former, but a distal demonstrative in the latter.

(124) **Discourse deictic uses in non-anticipatory constructions**

a. Jahongīr=ta der yoðd. **Yid**=ta mis mu qār sifen-t.  
 Jahongir=FAC late come.PRS.3SG DEM.DIR.MED=FAC also 1SG.OBL anger raise.PRS-3SG

‘Jahongir is going to arrive late. That’s also going to make me angry.’

- b. Jahongīr der yat=xu      dāži baxiǰ=i      na-tilāpt.      **Ik=u=yi**      mu qār  
 Jahongir late come.PST=and even apology=3SG NEG-ask.for.PST PREC=3SG.OBL.DIR.M=3SG my anger  
 sifent.  
 raise.PST  
 ‘Jahongir arrived late and didn’t even apologize. That made me angry.’

### 6.2.3.3 Pure text deictic uses of demonstratives

Pure text deixis involves the use of demonstratives to refer to a chunk of language itself (e.g. a word or sentence) – i.e. the “material side of language” (Diessel 1999: 101) – rather than the meaning of that language. Thus, for instance, the demonstrative *that* in the question ‘Could you repeat *that*?’ is an instance of pure text deixis.

Diessel (1999) considers pure text deixis to be a special case of exophoric use. However, I have chosen to consider pure text deixis in Shughni in this subsection on endophoric uses of demonstratives because, unlike exophoric uses of demonstratives, as examined in Section 6.2.2, pure text deictics do not pick out a referent in the physical speech situation. In pure text deixis in Shughni, the choice of demonstrative degree depends on the point in the discourse at which its referent was uttered (or will be uttered). In this way, pure text deictics in Shughni resemble discourse deictics, in which the choice of medial versus distal degree similarly depends on the *temporal* location at which an event occurred.

Nonetheless, pure text deixis in Shughni differs from discourse deictics in that the use of proximal demonstratives is permitted. Specifically, if the referent is a chunk of speech which will immediately follow the utterance containing a pure text deictic, then the proximal form is used. This is shown in the sentence in (125a). If the referent is a chunk of speech which immediately precedes the utterance containing a discourse deictic, then the distal form is used. This can be seen in (125b), which is a response to (125a).

#### (125) Pure text deictic demonstrative use: Proximal and medial

- a. Tu dar-borai **ik=mi**      (\*di, \*wi) jumlā čīz fīkri kin-i?      . . .  
 you about      PREC=3SG.PROX.OBL.M (\*MED, \*DIST) sentence what thought do.PRS-2SG  
 ‘What do you think about this sentence.’ . . . (speaker then says sentence)
- b. **Response to (125b)**
- Wuz fīkri-yum **yid**      (\*yam, \*yu) zūr jumlā!  
 I thought-1SG DEM.MED.DIR (\*PROX, \*DIST) great sentence  
 ‘I think that’s a great sentence!’

However, if the referent is a piece of speech which is further removed in time from the utterance containing the discourse deictic, whether in the past or the future, then a distal form is required. This is exhibited by the sentences in (126), where the referent of the distal demonstrative *wi*, used in both (126a) and (126b), is a sentence which was uttered the day before.

(126) **Pure text deictic demonstrative use: Distal**

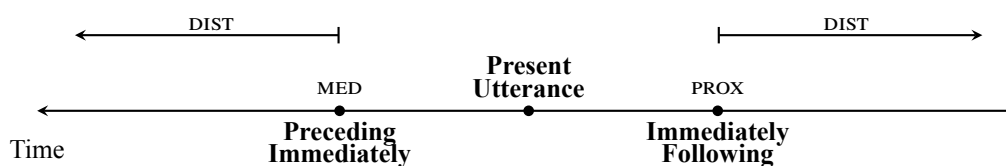
- a. Tu ba yoð o, biyor=um wi (\*mi, \*di) jumla tu-rd ca lud?  
 your to memory PQ yesterday=1SG DEM.DIST.OBL.M (\*PROX, \*MED) sentence 2SG.OB-DAT C say.PST?  
 Tu dar-borai wi čiz fikri kin-i?  
 what about DEM.DIST.OBL.M thought do.PRS-2SG  
 ‘Remember the sentence I told you yesterday? What do you think about that sentence?’

b. **Response to (126a)**

- Wuz fikri-yum yu (\*yam, \*yid) zur jumla!  
 I thought-1SG DEM.DIST.DIR.M (\*PROX, \*MED) great sentence  
 ‘I think that’s a great sentence!’

The choice of demonstrative degree in pure text deixis is schematized in Figure 6.7. The points (and intervals) in this figure represent the temporal location of the piece of speech to which a pure text deictic refers.

**Figure 6.7:** Choice of demonstrative degree in pure text deixis.



## 6.2.4 Deictic presentatives

In addition to demonstratives used in adnominal, pronominal, and identificational functions, which share the forms given in Table 6.5 above, Shughni also possesses a set of deictic presentatives (glossed *PRSV*), which are similar in use to French and Russian presentatives *voici/voilà* and *vot/von*, respectively (see Grenoble & Riley 1996 and references therein on the Russian and French presentatives). To my knowledge, there is no comprehensive descrip-



tion of Shughni deictic presentatives (but see Karamkhudoev 1973: 237; Kurbanov 1976: 220; and Karamshoev 1963a: 236, 1978: 200-202 for brief discussions).

Shughni deictic presentatives have distinct forms for each of the proximal, medial, and distal degrees. These are given in Table 6.7. As noted by Karamshoev (1978: 200-202), the forms of these deictics suggest an etymological connection to the direct demonstratives *yam* (PROX), *yid* (MED), and *yu* (DIST.M). Note further that a separate distal form *yuwa* has been attested and described, for instance, by Karamshoev (1978), according to whom it is traditionally used with feminine referents. Although my consultants have heard this form, they generally agree that it is not used in the modern language. Indeed, Karamshoev himself remarks that *yuwa* is unstable and is not consistently used with feminine referents.

**Table 6.7:** Shughni deictic presentatives.

<u>Degree</u>	<u>Form</u>
PROX	<i>yima</i>
MED	<i>yida</i>
DIST	<i>yuwu</i> ( <i>yuwa</i> )

Shughni deictic presentatives have both exophoric and endophoric uses, similar to the exophoric and discourse deictic uses of French and Russian presentatives, as described by Grenoble and Riley (1996). When used exophorically, Shughni deictic presentatives typically accompany the introduction of a new entity into the discourse. Presentatives in Shughni may or may not be accompanied by additional material, as the newly introduced entity may already be visible and apparent to both speaker and addressee. Hence, *yima*, *yida*, and *yuwu* are all perfectly acceptable minimal utterances in their own right. In many cases, however, as in the examples in (127) below, the presentative is followed immediately by a noun (phrase) which constitutes the topic.

The choice of degree corresponds to the same set of criteria used in choosing the degree of other demonstratives, as described in Section 6.2.2. In general, if the new entity being introduced into the discourse is in the personal space of the speaker, the proximal form *yima* is used. If it is in the personal space of the addressee or elsewhere in the shared interactional space, the medial *yida* is used. In cases where the referent is far from both and outside their shared interactional space, the distal *yuwu* is used. Examples of each degree of deictic presentative are given in (127), along with a relevant context.

- (127) a. **Proximal, Context:** The speaker, who is hosting the addressee for the night, hands him a pillow.

**Yima** tu takyā.  
 PRSV.PROX your pillow  
 ‘Here’s your pillow.’

- b. **Medial, Context:** The waiter at a restaurant brings the addressee some water, and the speaker notifies the addressee that her water has arrived.

**Yida** ǰac tu-rd.  
 PRSV.MED water you-DAT  
 ‘There’s some water for you.’

- c. **Distal, Context:** Speaker and addressee are sitting on a terrace looking out onto the city streets below. The speaker sees his best friend walking a couple blocks away and points this out to the addressee.

**Yuwu** mu sar bašānd oǰno!  
 PRSV.DIST my SUP good friend  
 ‘There’s my best friend!’

Only the medial deictic presentative in Shughni is commonly used endophorically. Here, it has a usage similar to the presentatives of French and Russian, in that it serves “to mark the openings and closings of topics” (Grenoble and Riley 1996: 820) and “bracket units of talk” (Schriffin 1987: 71). (See also Schegloff & Sacks 1973.) In Shughni, for instance, a speaker might utter the sentence in (128) after giving several sentences describing the chaos after a snowstorm which caused her travel to be delayed. Note that the presentative can be used at either end of the sentence, or at both ends. Such a sentence would signal that the speaker has finished her discussion on the preceding topic, and is prepared either to start a new topic or to hear the reaction of the addressee. The free gloss in (128) is simply an approximation. Other glosses may be more accurate depending on the circumstances in which the sentence is uttered.

- (128) **Endophoric use of a the medial presentative *yida***

**Yida** ik=disga (yida).  
 PRSV.MED PREC=like.this PRSV.MED  
 ‘So, there you have it.’

## Chapter 7

# Grammatical Gender in Nominals

This chapter, the final chapter in the series on nominals, examines the system of binary grammatical gender (masculine and feminine) in Shughni nominals. It is specifically concerned with three fundamental aspects of this topic: (i) the historical linguistic developments which have led to the way grammatical gender is expressed in modern Shughni; (ii) the morphosyntactic expression of gender on nouns and adjectives (i.e. gender concord; see below for more on this term); and (iii) the assignment of gender to Shughni nouns, including the complex set of semantic, formal, and historical factors which contribute to it. Note that the expression of grammatical gender in Shughni verbs is not treated here, but rather in Section 9.2 in Chapter 9 on regular and irregular verbs. Nonetheless, much of the information here, especially the historical context provided in Section 7.2, may serve as helpful background for understanding gender in Shughni verbs.

The chapter is organized in line with these goals. First, Section 7.1 offers **introductory remarks and an overview of the topic** which may serve as a summary for the reader looking only for an outline. Next, Section 7.2 examines the **diachronic sound changes**, starting in the Old Iranian period, which have led to the vowel and consonant alternations used to distinguish gender formally in modern Shughni. These **gender-distinguishing vowel and consonant alternations**, as well as patterns of **gender concord**, are presented fully in Section 7.3. Lastly, Section 7.4 looks at the complex set of **factors mediating gender assignment**.

## 7.1 Overview of grammatical gender in Shughni

Shughni possesses a system of binary grammatical gender in which all nouns are specified as either feminine or masculine. On nouns themselves, gender is distinguished morphologically only in a handful of pairs denoting higher level animals, such as *čaǰ* ‘hen’ and *čuǰ* ‘rooster’. The gender of the head noun in noun phrases is expressed morphosyntactically on a subset of modifying demonstratives and adjectives, and the gender of subject nouns is expressed on a subset of past and perfect verb stems. Note that I refer to the phenomenon whereby the grammatical gender of a head noun co-varies with the form of its modifiers (demonstratives and adjectives) as *nominal concord*, and I refer to the expression of gender on verbs as *agreement* (see Norris 2014, 2019 for an overview these terms and on the differences between these two phenomena).

Initial examples illustrating the grammatical expression of gender are given in (129). In (129a), the feminine noun *yāc* ‘girl’ exhibits concord with the demonstrative determiner *yā* and the adjective *dzalik* ‘small’, and it agrees with the verb *sat* ‘went’. The example in (129b) follows the frame of (129a), but with the masculine noun *yiǰā* ‘boy’ and the corresponding masculine forms of the concord elements, *yu* and *dzulik*, and the masculine form of the verb *sut* ‘went’. In these examples and others throughout the chapter, the controlling noun is bolded and agreeing (or concord) elements are underlined.

### (129) Grammatical gender: Initial examples

#### a. Feminine noun – *yāc* ‘girl’

Yā            dzalik **yāc** xubaθ        tar maktab sat.  
 DEM.DIR.F little.F girl PRON.EMPH to school go.PST.F  
 ‘That little girl went to school by herself.’

#### b. Masculine noun – *yiǰā* ‘boy’

Yu            dzulik **yiǰā** xubaθ        tar maktab sut.  
 DEM.DIR.M little.M boy PRON.EMPH to school go.PST.M  
 ‘That little boy went to school by himself.’

The binary gender system of Shughni is ultimately inherited from the ternary gender system of the Old Iranian period, but the erosion of the Old Iranian case system and the leveling of Old Iranian inflectional paradigms have made grammatical gender a comparatively less salient, if no less important, feature of Shughni grammar. Unlike

in Old Iranian, where a stem's phonological shape was correlated with its gender (e.g., Sokolova 1967; 1973; Karamshoev 1978), the form of a Shughni noun is not a reliable predictor of its gender classification. Moreover, unlike gender agreement in Old Iranian and in many modern Indo-European languages, the morphological expression of gender in Shughni nouns, adjectives and verbs is not carried out by suffixation, but rather via stem-internal vowel alternations, as in the adjectives *dzalik* 'small (F)' and *dzulik* 'small (M)', and in the verb stems *sāt* 'went (F)' and *sūt* 'was (M)' (cf. the examples in (129) above).

The Shughni vowel alternations which distinguish gender are complex. Several pairs of vowels are implicated in the expression of gender, and some vowels may express feminine gender in one word, but masculine in another. For example, the vowel *o* encodes feminine gender in *nōst* 'sat (F)' (cf. masc. *nūst*) but masculine gender in *vorj* 'horse (M)' (cf. fem. *vērdz*). Nonetheless, a number of patterns may be distinguished, and certain vowels are correlated with one or the other gender.

Regarding the assignment of gender, the semantic core of the Shughni system is the natural sex of humans and higher animals: nouns denoting female humans (e.g. *šinik* 'woman') and higher female animals (e.g. *vērdz* 'mare') are assigned feminine gender, while nouns denoting male humans (e.g. *čorik* 'man') and higher male animals (e.g. *vorj* 'stallion') are assigned masculine gender. In this respect, Shughni is similar to most Indo-European languages and a large number of non-Indo-European languages which also possess systems of grammatical gender or nominal classification with a natural-sex-based semantic core (Corbett 1991; 2014; Dahl 2000; Kramer 2020, *inter alia*).

However, the semantic residue in Shughni (i.e. nouns denoting lower animals such as *tīvdak* 'mosquito', inanimate entities such as *šac* 'water', and abstract concepts such as *žīwgax* 'love') is dealt with according to a *semantic* system of classification, rather than a *formal* system of the type common to Indo-European languages, such as the largely phonologically based system of French (e.g., Tucker et al. 1977), or the largely morphologically based system of Russian (e.g., Corbett 1982; 1991).

Specifically, the semantic residue in Shughni is grouped by and large according to the following four semantic factors: (i) **thematic (or conceptual) groupings** (Section 7.4.2), such that, for instance, sicknesses and milk products are masculine, while bodies of water and kitchen utensils are feminine; (ii) **abstractness vs. concreteness** (Section 7.4.3.1), such that virtually all Shughni nouns denoting abstract concepts are masculine, and in pairs of semantically related homophones which differ only in their gender, the feminine noun denotes the more concrete concept, while the masculine noun denotes the more abstract concept; (iii) **individual vs. mass** (Section 7.4.3.2),

such that traditionally feminine nouns denoting discrete entities may be used as masculine when referring to a general concept or when creating a mass reading; and (iv) **part vs. whole** (Section 7.4.3.3), where a word which expresses an item perceived as a whole is feminine, while its parts are masculine. Other factors, such as a word's historical gender (i.e. the gender of its ancient Iranian etymon) or its phonological shape, play only backgrounded roles in the determination of a noun's gender.

In gender assignment to categories besides humans and higher animates, Shughni resembles the semantically based classification systems of a wide range of non-Indo-European languages such as the Pama-Nyungan language Dyirbal (Dixon 1972), the Caucasian language Archi (Kibrik et al. 1977), and especially the language isolate Burushaski (Lorimer 1935–1938; Munshi 2006; 2019), which is spoken in northern Pakistan not far from the Shughni-speaking regions of Afghanistan and Tajikistan. The Burushaski gender system shares with Shughni gender assignment which occurs at least partially along the lines of semantic notions of abstract vs. concrete and mass vs. individual (e.g. Munshi 2019: 133–141). Indeed, the Shughni system has been posited to be the result of the substratal influence of neighboring non-Indo-European languages, perhaps Burushaski or bygone languages which were genetically related to it (see especially Edelman 1980b and references therein).

Before moving forward, it is worth pointing out that the most comprehensive source on grammatical gender in the Shughni-Rushani languages is Dodkhudo Karamshoev's (1978; 1986) two-volume work. However, many of the Shughni words and phrasal examples he cites appear to be archaic or no longer used in the language. This is at once an advantage of his work, as it provides documentation for a number of historical forms, and a disadvantage, in that many younger Shughni speakers may not recognize some of the language represented in the book. In any case, his work represents a useful starting point for grammatical gender in Shughni, a topic which is not fully understood and which presents many questions requiring extensive investigation.

## 7.2 Historical considerations in Shughni grammatical gender

The morphological expression of gender in modern Shughni nouns, adjectives, and past and perfect verb stems is perhaps best understood through a historical lens. Historical developments which helped shape the modern Shughni vowel system shed light on many of the complexities in the way gender is expressed in modern Shughni, and moreover, they help make sense of key differences between Shughni and the other languages of the Shughni-

Rushani group.

For instance, the history of the vowel system helps explain why the vowel *o* represents feminine gender in the past stem *nost* ‘sat’ and the adjective *rošt* ‘red’ (masculine forms *nūst* and *rūšt*), while it represents masculine gender in the noun *vorj* ‘(male) horse’ and the verb stem *xicođj* ‘frozen’ (feminine forms *vêrdz* and *xicêc*). It helps explain why the vowel *ī* is found in both the feminine and masculine perfect stems of the verb *naŷjīdow*, namely *naŷjīc* (F) and *naŷjīđj* (M), while its past stems have separate vowels, namely *naŷjīd* (F) and *naŷjīd* (M). And from a cross-linguistic perspective, it clarifies why certain non-gender-distinguishing animate nouns in Shughni like *šīg* ‘calf (M or F)’ have gender-distinguishing counterparts in other Shughni-Rushani varieties – cf. Rushani *šog* (M) and *šēg* (F), Bartangi *šōg* and *šēg*. We will return to these questions in due course.

The remainder of this section provides an overview of the historical developments which led to the current system by which gender is formally expressed in Shughni. Section 7.2.1 provides a brief introduction to the Old Iranian sound system; Section 7.2.2 then looks at the factors which governed sound changes in modern Shughni; and Section 7.2.3 provides a summary of those changes. The subsequent sections explain how these sound changes have shaped the expression of gender in the language.

### 7.2.1 Old Iranian basics

To understand the relevant changes which took place between the ancient Iranian period and the modern period, a basic understanding of the vowel systems of ancient Iranian languages is a useful starting point. Proto-Iranian had a system of at least four vowel phonemes: *\*a*, *\*ā*, *\*i*, and *\*u*, with the possible phonemic status of two further vowels *\*ī* and *\*ū* (e.g. Cantera 2017).<sup>1</sup> These vowels are exhibited in the Table 7.1.

In the attested Old Iranian languages, Old Persian and Avestan, the phonological shape of a noun’s stem was correlated with its grammatical gender specification (e.g. Karamshoev 1978: 27). Thus, nouns ending in *-a*, *-u*, and *-ū* were typically masculine (e.g. Av. *aspa-* ‘horse’; *dāru-* ‘wood’), while nouns ending in *-ā*, *i*, and *ī* were typically feminine (e.g. Av. *būmī-* ‘land’, *axšti-* ‘world’).

<sup>1</sup>The Iranian sounds *\*ī* and *\*ū* are the result of compensatory lengthening following the loss of laryngeals, as in *\*b<sup>h</sup>uh<sub>2</sub>mi-* > Avestan *būmī-* ‘land’. However, the precise time at which this change occurred, and the precise time at which the long vowels *\*ī* and *\*ū* became phonemes, is uncertain (Cantera 2017: 484). This process, whenever it occurred, has shaped the development of all Iranian languages, and in what follows I make reference to the Iranian sounds *\*ī* and *\*ū* while remaining agnostic regarding their phonemic status in Proto-Iranian.

**Table 7.1:** Proto-Iranian vowels (cf. Cantera 2017: 482).

	Front	Central	Back
<b>High</b>	<i>i</i> ( <i>ī</i> )		<i>u</i> ( <i>ū</i> )
<b>Mid</b>			
<b>Low</b>		<i>a</i> ( <i>ā</i> )	

A number of suffixes had gender-distinguishing forms. Those which are most relevant to the development of grammatical gender in Shughni are the past participial suffix *-ta* (M) / *-tā* (F) and the perfect suffix *-aka* (M) and *-ači* (F). As discussed in Section 9.2, these suffixes play an important role in shaping the expression of gender in Shughni verbs. However, the perfect stems also show up in at least one gender-distinguishing Shughni noun: *vorj* ‘(male) horse’ (< \**bār-aka*) and *vêrdz* ‘(female) horse’ (< \**bār-ači*-). This basic information on the Proto-Iranian vowel system and the Old Iranian nominal system sets the stage for the changes which lead to modern Shughni.

### 7.2.2 Factors moderating sound change.

A detailed examination of the correspondences between ancient Iranian vowels and the vowels of the Shughni-Rushani languages is given in Sokolova (1967: 24-63), according to whom there were five primary factors which affected an ancient Iranian vowel’s development in a given word:

- (i) **Stress** (stressed or unstressed position?)
- (ii) **Umlaut** (*i*- or *a*-umlaut, or no umlaut?)
- (iii) **Syllable shape** (is the vowel followed by one or two consonants?)
- (iv) **Surrounding consonants**
- (v) **Word-final or word-initial position**

For the purposes of the present discussion on gender in Shughni, factors (i)–(iii) will be the most important. We will focus only on vowels in stressed position while examining the effects of umlaut and the presence of one or two following consonants. The fourth factor, the effects of a specific surrounding consonant, comes into play only once for the purposes of gender in Shughni, namely where Proto-Iranian \**a* sometimes becomes the Shughni front



vowel *ē* before uvulars, where we would otherwise expect another vowel (e.g. *pēxt* ‘cooked’ < *\*paxta-*, where we would otherwise expect *pūxt*). The fifth factor is not relevant for this discussion because gender-distinguishing vowel alternations occur for stem-internal vowels and due to historical changes in stem-internal ancient Iranian vowels. Although the location of a vowel in word-initial or word-final position is indeed important for the overall development of the Shughni sound system, we may ignore this factor here.

In the history of Shughni, we can define *umlaut* as a process in which the quality of a given vowel is affected by the presence of another vowel or glide later in the word. There are three positions with respect to umlaut:

- (i) ***i*-Umlaut Position** – the vowel is followed by a reconstructed high front vowel *\*i* (*\*ī*) or palatal glide *\*y* in the following syllable, e.g. *\*a* in *\*gari-* > Sh. *žīr* ‘rock’
- (ii) ***a*-Umlaut Position** – the vowel is followed by a reconstructed long low vowel *\*ā* in the following syllable, e.g. *\*ū* in *\*būzā-* > Sh. *vaz* ‘goat (f)’
- (iii) **Neutral position** – all other positions, including vowels in stems ending in a consonant or in the vowels *\*a* or *\*u*, e.g. *\*a* in *\*kaska-* > Sh. *čūšč* ‘barley’.

Recall from Section 7.2.1 above that most Old Iranian masculine and neuter nouns ended in a consonant or in *-a*, and hence the stem vowel was in neutral position. Feminine nouns typically ended in either *-ā*, in which case the stem vowel was in *a*-umlaut position, or *-i*, in which case the stem vowel was in *i*-umlaut position. In modern Shughni gender-distinguishing nouns, which are most often animate nouns whose Old Iranian counterparts differed only in their final vowel, the masculine form always contains a vowel which was historically in neutral position, while the feminine form contains a vowel which was historically in *a*- or *i*-umlaut position (e.g. *kuta-* ‘male dog’ > Sh. *kud*; *\*kuti-* ‘female dog’ > Sh. *kid*). The following section presents the sound changes undergone by each Proto-Iranian vowel as it developed into its modern Shughni counterpart.

### 7.2.3 Sound changes affecting the expression of gender

The vowel changes which took place between the Ancient Iranian period (Proto-Iranian and Old Iranian) and modern Shughni are presented in Table 7.2. Note that in the table, where the fate of an ancient vowel is conditioned

by the presence of one or two following consonants, the result following one consonant is above (and labeled with ‘\_1C’), and the result following two consonants is below, labeled with ‘\_2C’). For instance, Proto-Iranian *\*a* becomes Shughni *ī* before a single consonant (e.g. Av. *\*dasa-*, Sh. *ōīs* ‘ten’), but *\*a* becomes *ū* before two consonants (e.g. *\*kaska-* > Sh. *čūšč* ‘barley’).

In general, *\*a*-umlaut and *\*i*-umlaut positions have the effect of turning the affected vowel into a more *a*-like or *i*-like vowel, respectively. Thus, for instance, the vowel *\*u* in neutral position becomes Shughni *u*, but in *a*-umlaut position it becomes *a* and in *i*-umlaut position *i* (before a single consonant).

There are certain exceptions to the developments presented below, but these general rules are easily sufficient to serve as a basis for the subsequent discussion. One notable exception, already alluded to above, is relevant: Iranian *\*a* in neutral position becomes Shughni *\*ē* before uvulars. That is, we find *ē* under conditions where we would otherwise expect *ī* (before a single consonant, as in *nêx* < *\*naxu* ‘plank bed’) or *ū* (before two consonants, as in *pêxt* < *\*paxta* ‘cooked’).

### 7.3 Morphological expression of gender in Shughni nominals

We have now seen the sound changes which have shaped, among other things, the way gender is expressed in modern Shughni. This section turns to the resulting vowel alternations implicated in the expression of gender in Shughni nominals. The morphological expression of gender in nouns is examined in Section 7.3.1 and in adjectives in Section 7.3.2. Then, Section 7.3.3 provides a closer look at patterns of gender concord in noun phrases. The latter section points out a peculiar difference, to my knowledge not yet noted in the literature, in the way plural nouns exhibit concord, on the one hand, versus agreement with verbs, on the other. Note that the expression of gender in demonstrative pronouns is not examined here, but rather in Section 6.2.1, and, as mentioned above, the expression of gender in verbs is examined in more detail in Section 9.2 of Chapter 9 on regular and irregular verbs.

**Table 7.2:** Historic developments of the Shughni vowel system.

	PROTO-IRANIAN VOWEL	SHUGHNI VOWEL	ANCIENT WORD	EXAMPLE MODERN SHUGHNI	GLOSS
Neutral Pos.	* <i>a</i>	$\xrightarrow{-1C} \bar{i}$	* <i>kata-</i>	<i>č̣īd</i>	‘house’
		$\xrightarrow{-2C} \bar{u}$	* <i>hapta-</i>	<i>wūvd</i>	‘seven’
	* <i>ā</i>	$\longrightarrow o$	<i>dāru-</i> (Av.)	<i>ōrg</i>	‘wood’
	* <i>u</i> , * <i>ū</i>	$\longrightarrow u$	* <i>kuta-</i>	<i>kud</i>	‘dog (M)’
<i>i</i> -Umlaut Pos.	* <i>a</i>	$\longrightarrow \bar{i}$	* <i>gari-</i>	<i>žīr</i>	‘rock’
	* <i>ā</i>	$\longrightarrow \hat{e}$	* <i>māsti-</i>	<i>mēst</i>	‘moon’
	* <i>u</i> , * <i>ū</i>	$\xrightarrow{-1C} i$	* <i>kuti-</i>	<i>kīd</i>	‘dog (F)’
		$\xrightarrow{-2C} i/\bar{i}$	<i>supti-</i> (Av.)	<i>sīvd</i>	‘shoulder’
<i>a</i> -Umlaut Pos.	* <i>a</i>	$\xrightarrow{-1C} \bar{a}$	* <i>kafā-</i>	<i>šāf</i>	‘saliva’
		$\xrightarrow{-2C} o$	* <i>taxtā-</i>	<i>toyd</i>	went (F)
	* <i>ā</i>	$\longrightarrow o$	* <i>dāiā-</i>	<i>ōod</i>	‘fell; hit (PST STEM)’
	* <i>u</i> , * <i>ū</i>	$\longrightarrow a$	* <i>gantumā-</i>	<i>žindam</i>	‘wheat’

### 7.3.1 Expression of gender in Shughni nouns

Gender is morphologically distinguished only in a handful of nouns in modern Shughni, primarily animate nouns denoting higher (domesticated) animals whose natural sex is relevant in the work of those who keep them.<sup>2</sup> In all, there are three primary vowel alternations used to distinguish gender in nouns, given as  $M \sim F$ :

ALTERNATION	EXAMPLE	GLOSS
(i) $u \sim a$	$\check{c}u\check{x} \sim \check{c}a\check{x}$	‘rooster~hen’;
(ii) $u \sim i$	$kud \sim kid$	‘dog (M~F)’;
(iii) $o \sim \hat{e}$	$vorj \sim v\hat{e}rdz$	‘horse (M~F)’

The first two alternations can be traced to ancient Iranian nouns with a stem vowel  $*u$  or  $*\bar{u}$ . In both, the masculine vowel is in neutral position, as it is followed by a stem-final consonant or a vowel  $*a$  or  $*u$  in the following syllable (e.g. *kuta-* ‘male dog’ > *kud*). Following regular sound changes, therefore, the vowel  $*u$  in masculine forms becomes modern Shughni *u*. In the feminine forms, however, the vowel is in different phonetic positions, namely  $*a$ -umlaut in the first type, where the following syllable contained  $*\bar{a}$ , and  $*i$ -umlaut in the second type, where the following syllable contained  $*i$  or  $*\bar{i}$ . Hence, in the first and second types, also via regular sound changes, the feminine vowels become modern Shughni *a* and *i*, respectively. The third type can be traced to ancient Iranian nouns with a stem vowel  $*\bar{a}$ . Here, the vowel in the masculine form is neutral position and becomes modern Shughni *o*, while the vowel in the feminine position is in *i*-umlaut position and becomes modern Shughni  $\hat{e}$ .<sup>3</sup>

These three types, with examples of etymologies where available, are summarized in Table 7.3. Note that reliable etymologies for individual nouns (and adjectives, as discussed below) are scarce. I have provided etymologies where possible, but the reconstruction of protoforms for many gender-distinguishing nouns remains to be done.

<sup>2</sup>Besides gender distinction in animate nouns, there are a few pairs of inanimate nouns which are formally identical except for their stem vowels, which follow one of the known patterns of vowel alternations implicated in gender distinction. For instance, the feminine noun *šac* ‘water’ has the masculine counterpart *šuc* ‘bullion; broth’, where the vowel pair  $u \sim a$  is a common gender-distinguishing alternation.

<sup>3</sup>Historically, two further vowel pairs have been implicated in gender distinction with animate nouns:  $\bar{u} \sim i$  (*wūrj*//*wirdzin* ‘(she-)wolf’, and  $\bar{u} \sim o$  (*vūyd*//*voyd*) ‘evil spirit’. To my knowledge, these are the only attested pairs of nouns which contain these vowels, and consultants indicate that the word *wirdzin* is not used with the meaning ‘she-wolf’, but rather ‘disheveled’, while the pair *vūyd* and *voyd* is no longer used.

**Table 7.3:** Gender-distinguishing noun pairs.

ALTERNATION (M~F)	HISTORICAL VOWEL	POSITION	EXAMPLE	GLOSS	ETYMOLOGY
<b>u~a</b>	<b>*u, *ū</b>	M: Neutral Position F: <i>a</i> -Umlaut Position	M: <i>č<u>u</u>ǰ</i> F: <i>č<u>a</u>ǰ</i>	rooster hen	— —
			M: <i>ǰ<u>u</u>c</i> F: <i>ǰ<u>a</u>c</i>	broth water	cf. Av. <i>xš<u>u</u>ǰra-</i> <i>*xš<u>u</u>ǰrā-</i>
<b>u~i</b>	<b>*u, *ū</b>	M: Neutral Position F: <i>i</i> -Umlaut Position	M: <i>k<u>u</u>d</i> F: <i>k<u>i</u>d</i>	dog (M) dog (F)	<i>*kuta-</i> <i>*kuti-</i>
			M: <i>p<u>u</u>š</i> F: <i>p<u>i</u>š</i>	cat (M) cat (F)	— —
			M: <i>g<u>u</u>ǰ</i> F: <i>g<u>i</u>ǰ</i>	baby goat (M) baby goat (F)	
			M: <i>b<u>u</u>c</i> F: <i>b<u>i</u>c</i>	animal young (M) animal young (F)	cf. Av. <i>puθra-</i> (via analogy)
<b>o~ê</b>	<b>*ā</b>	M: Neutral Position F: <i>i</i> -Umlaut Position	M: <i>v<u>o</u>ǰ</i> F: <i>v<u>ê</u>rdz</i>	horse (M) horse (F)	<i>*bāra<u>ka</u>-</i> <i>*bāra<u>či</u>-</i>
			M: <i>ǰ<u>o</u>ǰ</i> F: <i>ǰ<u>ê</u>ǰ</i>	homestead summer grazing pasture	
			M: <i>p<u>o</u>ǰ</i> F: <i>p<u>ê</u>ǰ</i>	leg animal trap	

### 7.3.2 Expression of gender in Shughni adjectives

Like gender-distinguishing nouns, adjectives which show morphological alternations based on concord with their head noun are relatively few in number. Karamshoev (1978: 30-58) identifies only a few dozen such adjectives, and many of these have apparently fallen out of use in the modern language. The following three vowel alternations are used in gender-distinguishing adjectives:

ALTERNATION	EXAMPLE	GLOSS
(i) $u \sim a$	<i>kut</i> ~ <i>kat</i>	‘short’;
(ii) $\bar{i} \sim \bar{a}$	<i>xīy</i> ~ <i>xāy</i>	‘handsome; pretty; sweet’;
(iii) $\hat{e} \sim \bar{a}$	<i>xinēx</i> ~ <i>xanāx</i>	‘annoying’

Examples showing the use of each adjectives with each of these vowel alternations exhibiting concord with head nouns are given in (130)–(132).

(130) $u \sim a$	(131) $\bar{i} \sim \bar{a}$	(132) $\hat{e} \sim \bar{a}$
a. <i>kut</i> <i>yūnj</i> ‘short hair’	a. <i>xīy</i> <i>yidā</i> ‘handsome boy’	a. <i>xinēx</i> <i>yidā</i> ‘annoying boy’
b. <i>kat</i> <i>vāx</i> ‘short rope’	b. <i>xāy</i> <i>yīnik</i> ‘pretty woman’	b. <i>xanāx</i> <i>yāc</i> ‘annoying girl’

The first alternation  $u \sim a$  is also found among nouns and also represents the reflex of Proto-Iranian  $*u$  and  $*\bar{u}$  in neutral position (masculine) and  $a$ -umlaut position (feminine), respectively. The two subsequent alternations involve the result of Proto-Iranian  $*a$ . The second alternation  $\bar{i} \sim \bar{a}$  involves Proto-Iranian  $*a$  in neutral and  $a$ -umlaut position before a single consonant, and the third type  $\hat{e} \sim \bar{a}$  involves this vowel in neutral position before a uvular, where it becomes  $\hat{e}$ , and in  $a$ -umlaut position where it becomes  $\bar{a}$ .<sup>4</sup>

The vowel alternations used in gender-distinguishing adjective pairs are summarized in Table 7.4. Here, there are no reliable etymologies which clearly demonstrate the development of the historical form into the modern form. However, these words exhibit vowel alternations which are well attested in other words, primarily verbs, for which there is reliable etymological evidence.

To return briefly to an issue mentioned in Section 7.1, note that there are a few nouns and adjectives which do not formally distinguish gender in Shughni but do so in other Shughni-Rushani languages. Many of these cases

<sup>4</sup>A fourth alternation which is apparently no longer used in adjectives is  $\bar{u} \sim o$ , which is found for instance in the adjective *vūydz* ~ *voýdz* ‘oval-shaped’. This alternation comes from  $*a$  in neutral position and  $a$ -umlaut position before two consonants, where it becomes  $\bar{u}$  and  $o$ , respectively. In this case, only the masculine form *vūydz* is used in modern speech.

**Table 7.4:** Gender-distinguishing adjective pairs.

ALTERNATION (M~F)	HISTORICAL VOWEL	POSITION	EXAMPLE (M~F)	GLOSS
<b>u~a</b>	<b>*u, *ū</b>	M: Neutral Position F: <i>a</i> -Umlaut Position	<i>k<u>u</u>t</i> ~ <i>k<u>a</u>t</i> <i>dz<u>u</u>l(ik)</i> ~ <i>dz<u>a</u>l(ik)</i> <i>b<u>u</u>q</i> ~ <i>b<u>a</u>q</i> <i>cil<u>u</u>q</i> ~ <i>cil<u>a</u>q</i> <i>c<u>u</u>q</i> ~ <i>c<u>a</u>q</i> <i>ǰip<u>u</u>ǰ</i> ~ <i>ǰip<u>a</u>ǰ</i> <i>b<u>u</u>f</i> ~ <i>b<u>a</u>f</i> <i>ǰib<u>u</u>q</i> ~ <i>ǰib<u>a</u>q</i>	short small convex sticking out sticking out bright white heavyset heavyset
<b>ī~ā</b>	<b>*a</b>	M: Neutral Position F: <i>a</i> -Umlaut Position	<i>xīǰ̃</i> ~ <i>xāǰ̃</i> <i>cīǰ̃</i> ~ <i>cāǰ̃</i> <i>šīl̃</i> ~ <i>šāl̃</i>	handsome; sweet bitter limping
<b>ê~ā</b>	<b>*a</b>	M: Neutral Position (_uvular) F: <i>a</i> -Umlaut Position	<i>čēxt</i> ~ <i>čāxt</i> <i>xinēǰ̃</i> ~ <i>xanāǰ̃</i>	sweet annoying

are also the result of regular sound change and involve the historical vowel *\*a* and the modern Shughni vowel *ī*. In particular, Proto-Iranian *\*a* becomes Shughni *ī* both in neutral position and *i*-umlaut position, but in other languages of the group there are separate results for this vowel in each position. In Rushani, for instance, *\*a* in neutral position becomes *o*, but in *i*-umlaut position it becomes *ē*. Thus, the Shughni noun *šīg* ‘calf (of a cow – M or F)’ is the result of *\*a* in both neutral and *i*-umlaut position, but in Rushani we have *šog* (M) from neutral position and *šēg* (F) from *i*-umlaut position.

Finally, a number of gender-distinguishing forms, both nouns and adjectives, given by Karamshoev in both volumes of his work on gender (1978; 1986) appear to be either no longer used in modern Shughni, or else one of the gender-distinguishing counterparts is no longer used. In some cases, the feminine form falls out of usage, while in others the masculine form is no longer used. For example, the feminine form for the adjective meaning ‘red’, namely *rošt*, is apparently no longer used, with the masculine form *rūšt* used for all nouns. For the adjective *čūxt* ~ *čāxt* ‘crooked’, it is the masculine form which is no longer used, with *čāxt* used for all nouns. A more detailed study investigating which gender-distinguishing adjectives are still used in the language, ideally corpus-based and ideally with a cross-dialectal component, is needed to better understand this phenomenon.

### 7.3.3 Gender concord in noun phrases (vs. verb phrases)

Gender concord between nouns and their dependents is straightforward. It takes place both between head nouns and their dependents within the same noun phrase, as shown in the examples in (133), and between subject nouns and predicative adjectives, as in (134).

#### (133) NP-internal gender concord

##### a. *Sūg* ‘story’ – Feminine

Wuz=ta tu-rd    yi kat    **sūg** lūm.  
 I=FAC    you-DAT a    short.F story tell.1SG.PRS  
 ‘I’m going to tell you a short story.’

##### b. *Čīd* ‘house’ – Masculine

Tama=yet wi            dzulik **čīd**    xarīd čūd    o?  
 you.PL=2PL DEM.OBL.M small.M house buy    do.PST PQ  
 ‘Did you buy that small house?’



(134) **Gender concord in predicative adjectives**a. *Sûg* ‘story’ – Feminine

Tu **sûg** čīzjāt dis kat?  
 your story why so short.F  
 ‘Why is your story so short?’

b. *Puc* ‘son’ – Masculine

Wev **puc** yal lap dzulik.  
 their son still very small.M  
 ‘Their son is still very small.’

In the sentence in (133a), concord is shown between the head noun *sûg* ‘story’, a feminine noun, and the feminine oblique demonstrative pronoun *wam* and the feminine adjective *kat* ‘short’, all within the same noun phrase. In (133b), concord occurs between the masculine noun *čīd* and its dependents, namely the demonstrative pronoun *wi* and the adjective *dzulik*. In the examples (134), gender concord occurs between the subject nouns *sûg* ‘story’ (F) and *puc* ‘son’ (M) and the predicative adjectives *kat* ‘short’ and *dzulik* ‘small’, respectively.

There is no special form exhibiting concord with plural nouns. The same feminine and masculine forms of adjectives are used for concord with singular and plural nouns. Examples showing concord with a plural feminine and masculine noun are shown in (135a) and (135b), respectively.

(135) **Gender agreement between plural nouns and adjectives**a. **Feminine plural noun → Feminine adjective**

Wāð **dzalik** yāc-en=en fukaθ tar maktab.  
 DEM.DIR.PL small.PL girl-PL=3PL all at school  
 ‘All those small girls are at school.’

b. **Masculine plural noun → Masculine adjective**

Wāð **dzulik** yiðā-yen=en fukaθ tar maktab.  
 DEM.DIR.PL small.PL boy-PL=3PL all at school  
 ‘All those small boys are at school.’

Importantly, however, we find a split with respect to gender agreement with verbs and gender concord with adjectives. With nominal concord, as shown above, the form of the adjective invariably corresponds to the gender of the

plural noun – i.e. feminine adjectives with feminine plural nouns and masculine adjectives with masculine plural nouns. On the other, in the case of verbs, both feminine and masculine plural nouns call for a verb form which is identical to the feminine. This is shown in the sentences in (136).

(136) a. **Feminine plural noun → Feminine/(Plural) past stem**

Wāð        **yāc**-en=en tar maktab sat.  
 DEM.DIR.PL girl-PL=3PL to school go.PST.PL  
 ‘Those girls went to school.’

b. **Masculine plural noun → Feminine/(Plural) past stem**

Wāð        **yīðā**-yen=en tar maktab sat.  
 DEM.DIR.PL boy-PL=3PL to school go.PST.PL  
 ‘Those boys went to school.’

The same pattern occurs not only with past stems, but also with perfect stems, as shown in (137).

(137) a. **Feminine plural noun → Feminine/(Plural) perfect stem**

Wāð        **yāc**-en=en tar maktab sic.  
 DEM.DIR.PL girl-PL=3PL to school go.PRF.PL  
 ‘Those girls have gone to school.’

b. **Masculine plural noun → Feminine/(Plural) perfect stem**

Wāð        **yīðā**-yen=en tar maktab sic.  
 DEM.DIR.PL boy-PL=3PL to school go.PST.PL  
 ‘Those boys have gone to school.’

The situation regarding concord and agreement with singular and plural nouns are summarized in Table 7.5. The examples come from the gender-distinguishing adjective *kat~kut* ‘short’ and the past stems of the gender-distinguishing verb *vidow* ‘be’. Cells which illustrate the irregular agreement pattern – i.e. where a single verb form, identical to the feminine form, is used for both feminine and masculine plural nouns – are shaded.

This pattern can be partially explained through regular sound change. In particular, past verb stems historically have a plural form which, through regular sound change, became identical to the feminine stem. Hence, for verbs

**Table 7.5:** Gender agreement with adjectives and verbs.

Noun (GEN.NUM)	Conc./Agr. Form		Example	
	ADJ	VERB	ADJ	VERB
SG.F	fem.	fem.	<i>kat</i>	<i>vad</i>
SG.M	masc.	masc.	<i>kut</i>	<i>vud</i>
PL.F	fem.	fem.(/pl.)	<i>kat</i>	<i>vad</i>
PL.M	masc.	fem.(/pl.)	<i>kut</i>	<i>vad</i>

like *sat* in (136a), which are identical in form with the feminine form but which agree in gender with a masculine plural noun, we are actually dealing with a plural form which has become syncretic with the feminine singular.

However, the same explanation does not hold for perfect stems. Perfect stems historically have a distinct plural form, which, also due to regular sound changes, has a vowel identical to the past feminine vowel, but stem-final consonant(s) identical to the masculine stem-final consonants. Thus, historically, the verb *sittow* ‘go’ had a plural form *sađj* in addition to feminine *sic* and masculine *suđj*. However, in recent decades this form has virtually fallen out of use, and its function has been picked up by the feminine form of perfect stems. This is presumably a result of analogy with past stems, where speakers transferred the syncretism of the feminine and plural past stems to perfect stems (Dodykhudoeva 1988). (See also Section 9.3.2.1 for a discussion of plural perfect stems and Section 9.2.3.2 on gender in perfect stems.)

## 7.4 Gender assignment in nouns

The assignment of gender to nouns in Shughni, like the morphological expression of gender, differs markedly from the systems of gender assignment in other Indo-European languages. In particular, semantic factors, as opposed to formal factors such as a noun’s phonological shape or its inflectional class, are most influential in determining the gender of a Shughni noun. Importantly, this section demonstrates that the factors which determine a noun’s gender are not only relevant for native Shughni nouns, but also for borrowed lexemes, including both old borrowings from

Tajik and Arabic, as well as new borrowings from Russian. This suggests that the system of gender assignment in Shughni has remained steadfast in the face of an influx of Russian borrowings in the past century. Nonetheless, much remains to be understood about the Shughni system, and future research may not only shed light on how each of the factors discussed here interact with one another; it may also clarify a number of apparent exceptions to the general tendencies of gender assignment described in this section.

The section proceeds as follows. Section 7.4.1 provides an overview of the different factors which may influence a noun's grammatical gender. The following two subsections then explore specific semantic concepts which may affect a noun's gender specification, including thematic groupings (Section 7.4.2) and notions of abstract vs. concrete, mass vs. individual, and part vs. whole (Section 7.4.3).

### 7.4.1 Overview of gender assignment: Three primary factors

The three factors which most influence a noun's gender specification as masculine or feminine are the following.

- (i) **Historical gender** – for a noun inherited from Old Iranian, what was its gender at that time?;
- (ii) **Phonological form** – does the noun's stem vowel align with vowels typical for fem. or masc. nouns?;
- (iii) **Meaning** – e.g., does the noun belong to a thematic class linked to a specific gender?

It was mentioned above that the phonological form of a noun cannot be reliably used in predicting its gender. In fact, none of the three primary factors which may influence a noun's gender specification are foolproof ways of determining its gender. However, if we take each into consideration, it is generally possible, albeit not always, to understand why a given noun belongs to a given gender. Below, I explain why each factor is an important, yet imperfect, criterion for understanding a noun's gender specification.

#### 7.4.1.1 First factor: Historical gender

Many Shughni nouns which have been inherited from ancient Iranian languages have retained their grammatical gender from ancient times. For example, the masculine Proto-Iranian noun *\*kata* 'house' becomes the masculine Shughni noun *čīd*. Other examples of masculine Shughni nouns which have proto-forms which were also masculine include *δust* 'hand' (cf. Av. *zasta-*, masc.); *yūčk* 'tear' (cf. Av. *asru-*, masc.); and *divi* 'door' (cf. Av. *dvar-*, masc.). Shughni feminine nouns which have proto-forms which were feminine in ancient Iranian languages include *ziv*

‘tongue’ (cf. Av. *hizvā-*, fem.); *ǰāb* ‘night’ (cf. Old Prs. *xšap-*, fem.); and *sitan* ‘column’ (cf. Old Prs. *xšap-*, fem.). Most ancient Iranian neuter nouns, for their part, have become masculine nouns in Shughni, including, for instance, *ǰūñǰ* ‘hair’ (cf. Av. *gaona-*, neut.) and *xīr* ‘sun’ (cf. Av. *hvar-*). (See Karamshoev 1986: 133-138 for a more complete list of Shughni nouns and their corresponding etymons.)

Although many Shughni nouns inherited from ancient Iranian follow this pattern, there are multiple exceptions. For instance, the Shughni feminine nouns *ǰitērdz* ‘star’, *yed* ‘bridge’, and *mēst* ‘moon’ are traced to masculine nouns in Old Iranian – cf. Av. *stār-*, *haētū-*, and *mah-*, respectively. Importantly, however, many Shughni nouns are not inherited from ancient Iranian, but have rather been borrowed into the language at various stages of its history. These include Tajik, Arabic (via Tajik), and Russian nouns. For these nouns, particularly Arabic and Russian nouns, there must be a system other than their gender classification in ancient Iranian. This system shows up in the next two factors.

#### 7.4.1.2 Second factor: Phonological form

The second factor which may play a role in determining a noun’s gender is its phonological form. As indicated above, a noun’s phonological form is by no means a reliable means of determining its gender, but it may help explain certain deviations from the other two factors.

We saw in Section 7.3 that in gender-distinguishing nouns and adjectives, certain vowels are correlated with masculine or feminine gender. In particular, the Shughni vowels *u* and *ū* are often found in the masculine correlate of a gender-distinguishing pair of nouns, adjectives, and as will be described in Section 9.2, past and perfect verb stems. Indeed, there are no instances in which we find *u* or *ū* as the stem vowel of a feminine correlate. The vowel *ī*, which is historically associated with masculine nouns with a stem vowel *\*a*, is also correlated with masculine gender. On the other hand, the vowels *a*, *ā*, and *i* are typical of feminine correlates and virtually not present in masculine correlates. Thus, in cases where a noun’s gender does not match the gender of its protoform, or where it does not match gender of the semantic class it is most closely associated with, we often find that it contains a stem vowel correlated with its gender.

As an example, the vast majority of nouns denoting lower animals – i.e. those outside the small group of domesticated animals for which the language has gender-distinguishing forms based on natural sex – belong to the feminine gender. In other words, lower animals constitute a feminine thematic grouping. In three clear exceptions we find

to this pattern, however, namely where a non-gender-distinguishing noun denoting an animal is masculine, the stem vowel is either *ū* or *ī*. This is the case for *ǰirbǰǰ* ‘frog’, *ǰičǰǰ* ‘marmot’, and *pūrg* ‘mouse’ (Karamshoev 1986: 39). Similarly, the thematic class of bedding items is almost exclusively feminine. An exception is the borrowed Russian noun *padūška*, for which the stressed *ū* may explain the fact that it belongs to masculine gender.

#### 7.4.1.3 Third factor: Semantics.

The third factor underlying the gender classification of Shughni nouns is semantics, a broad factor whose many facets will be discussed in turn in the following subsections. As a whole, the semantic factor is by far the most influential factor in a noun’s gender assignment in that it largely overshadows the other two factors mentioned above. First and foremost, of course, is the notion that grammatical gender corresponds to natural sex for many nouns denoting humans and higher animals. Beyond natural sex, many nouns are organized by thematic classes which correspond to either masculine or feminine, as well as by semantic notions of abstractness vs. concreteness; mass vs. individual, and part vs. whole. Each of these categories are discussed in turn below.

### 7.4.2 Thematic classes and grammatical gender

The grammatical gender assignment of many Shughni nouns appears to be rooted in their belonging to a particular thematic (or ‘conceptual’) class. Here, I give an overview of some of the major thematic classes in Shughni nouns. In each case, where applicable, I point out any apparent exceptions. It is worth noting, however, that the labels I give to each class are in many cases quite broad (e.g. technology or transportation vessels), and a key goal for future research is to narrow down the definitions of each class. In doing so, one may be able to determine that certain nouns which seemingly constitute exceptions in fact belong to a separate class.

#### 7.4.2.1 Predominantly masculine categories.

Thematic categories of nouns which are predominantly masculine include the names of populated places (e.g. cities, villages, etc.), sicknesses, footwear, milk products, and vegetables. By far the largest and most cohesive class of masculine nouns in Shughni, however, is that of abstract nouns. Both native and borrowed words denoting abstract concepts are masculine, nearly without exception. This includes abstract concepts such as *ǰīwǰgaǰ* ‘love’,

as well as verbal nouns such as *zidāridz* ‘sweeping’ and infinitives used as gerunds in argument position, such as *kor čīd(ow)* ‘(the act of) working’.

A non-exhaustive list of masculine categories, together with several examples, is given in Table 7.6 below. (A more extensive list can be found in Table A.1 in Appendix A.) Note that because the semantic concepts of abstract vs. concrete constitute a larger phenomenon within the system of gender assignment in Shughni nouns, only a handful of verbal nouns are included in this list. The topic of abstractness and concreteness is taken up again in Section 7.4.3.1.

Importantly, borrowed words – including recent Russian borrowings – are also subject to gender assignment via thematic categorization. For each category, two or three Russian borrowings are listed which ostensibly fall into these same thematic groupings and thus receive the same grammatical gender assignment as their native Shughni groupmates. These Russian borrowings are marked with an asterisk – e.g. *\*girīp* ‘flu’ < Ru. грипп, *gripp*, which falls into the category of sicknesses. Where the Russian source of a borrowed lexeme differs markedly from its realization in Shughni, the Russian source word is provided in parentheses (e.g. Sh. *kapūst* ‘cabbage’ < Ru. капуста, *kapusta*).

#### 7.4.2.2 Predominantly feminine categories.

Examples of predominantly feminine categories include bodies of water, literary and artistic productions, trees and fruits, containers and tableware, bedding items, and technology and transport. Examples of feminine categories are shown in Table 7.7 (with a more extensive list given in Table A.2 in Appendix A).

As was the case with masculine categories discussed above, many Russian borrowings are assigned feminine gender apparently due to their association with feminine thematic categories. Two examples of borrowed Russian nouns are provided for each category in Table 7.7. Thus, for instance, the word *vadapad* ‘waterfall’ (< Ru. водопад, *vodopad*) is feminine due to its association with the feminine category of bodies of water, despite the fact that it is masculine in Russian. Similarly, the word *ramān* ‘novel’ (< Ru. роман, *roman*) is feminine by virtue of its inclusion in the thematic category of literary and artistic items. The source word for *ramān*, like that of *vadapad*, is also masculine in Russian.

**Table 7.6:** Predominantly masculine thematic categories.

<u>PLACE NAMES</u>		<u>SICKNESSES</u>	
<i>Xaray</i>	Khorog	<i>dārō</i>	pain; illness
<i>Riǰūn</i>	Rushan	<i>bođ</i>	eczema
<i>Dūšanbi</i>	Dushanbe	<i>kêxak</i>	cough
<i>*Maskow</i>	Moscow	<i>*girīp</i>	flu (< грипп)
<i>*Sankt Peterburg</i>	St. Petersburg	<i>*tūbirkuloz</i>	tuberculosis

<u>FOOTWEAR</u>		<u>MILK PRODUCTS</u>	
<i>šilopak</i>	sandal	<i>ǰūvd</i>	milk
<i>ǰirīb</i>	(large) sock	<i>qurūt</i>	yogurt ball
<i>kafǰ</i>	(wooden) shoe	<i>ōūy</i>	doogh (drink)
<i>*tāpka</i>	slippers	<i>*s(i)metana</i>	cream
<i>*k(i)rasovka</i>	gym shoe	<i>*yogūrt</i>	yogurt

<u>VEGETABLES</u>		<u>INFINITIVES AND GERUNDS</u>	
<i>piyoz</i>	onion	<i>zidāridz</i>	sweeping
<i>bodrīng</i>	cucumber	<i>čêridz</i>	harvesting
<i>zardak</i>	carrot	<i>kitob ǰêyd</i>	book reading
<i>*kapūst</i>	cabbage (< капуста)	<i>*ǰ(i)dat čīd</i>	waiting
<i>*kartuškā</i>	potato flu	<i>*ispolzuvat čīd</i>	using

A couple of notes are in order about specific words in the feminine categories delineated in Table 7.7. The first regards the gender of compound nouns, which always corresponds to the gender of the head noun. A useful example is the word *šīrčoy* ‘milk tea’, which one may initially link to the masculine category of milk products (cf. Table 7.6). However, because the head noun of this compound, *čoy* ‘tea’, falls into the feminine category of non-milk-based drinks, the noun *šīrčoy* likewise falls into this category and receives feminine gender assignment.

The second note regards apparent exceptions to certain thematic categories which have no immediate explanation. In the case of feminine categories, notable exceptions include the masculine word *padūška* ‘pillow’, which one might expect to fall under the category of bedding items, and the masculine word *kancert* ‘concert’, which one might expect to fall under the feminine category of ‘literary and artistic productions’ along with *kino* ‘movie’ and *kitob* ‘book’. As mentioned above, the gender assignment of *padūška* might be explained by its vowel *ū*, which



**Table 7.7:** Predominantly feminine thematic categories.

<u>BODIES OF WATER</u>		<u>LITERARY AND ARTISTIC ITEMS</u>	
<i>daryo</i>	river	<i>kitob</i>	book
<i>qûl</i>	lake	<i>sûg</i>	tale
<i>šarvidoĵ</i>	mountain stream	<i>xât</i>	letter
<i>*vadapad</i>	(large) waterfall	<i>*pesā</i>	play (< пьеса)
<i>*akeān</i>	ocean	<i>*ramān</i>	novel
<u>TREES</u>		<u>TECHNOLOGY AND TRANSPORT</u>	
<i>daraxt</i>	tree	<i>mošîn</i>	car
<i>yûz</i>	walnut tree	<i>tilifûn</i>	phone
<i>wed</i>	willow	<i>*tilivîzor</i>	television
<i>*yolka</i>	fir	<i>*poyizd</i>	train
<i>*s(i)liva</i>	plum tree	<i>*kampyûtir</i>	computer
<u>CONTAINERS AND TABLEWARE</u>		<u>BEDDING ITEMS</u>	
<i>toθč</i>	bowl	<i>lef</i>	blanket
<i>dek</i>	pot	<i>namad</i>	wool carpet
<i>čêd</i>	knife	<i>bolax</i>	small carpet
<i>*vîlka</i>	fork	<i>*adyāl</i>	blanket (< одеяло)
<i>*tarelka</i>	plate	<i>*matrās</i>	mattress

in Shughni, as in Russian, receives stress. This vowel, as discussed above, is often linked to masculine gender in many gender-distinguishing nominals and verbs in Shughni. No such explanation exists for *kancert*, however, and exceptions like these remain a question for future investigation.

#### 7.4.2.3 Split categories

A final topic in thematic categories is the notion of split thematic categories. These are semantic categories which clearly form a thematic class, but which contain nouns which are both masculine and feminine. Three examples of split categories, namely body parts, buildings, and plants, are exhibited in Table 7.8. (A more full list can be found in Table A.3 in Appendix A.)

Table 7.8: Split thematic categories.

Body parts	Buildings and building parts	Plants
<u>FEMININE</u>	<u>FEMININE</u>	<u>FEMININE</u>
<i>angišt</i> finger	<i>magazīn</i> store	<i>ǰar</i> dogrose
<i>ziv</i> tongue	<i>lafkā</i> store	<i>žāš</i> burdock
<i>yêv</i> mouth	<i>sitan</i> column	
<u>MASCULINE</u>	<u>MASCULINE</u>	<u>MASCULINE</u>
<i>cem</i> eye	<i>čīd</i> house	<i>woǰ</i> grass
<i>dām</i> back	<i>yījīd</i> stable	<i>cūōm</i> wormwood
<i>kīl</i> head	<i>dišīd</i> roof	<i>šūō</i> thorn

The first two categories, body parts and buildings, appear to be more or less evenly split. A possible explanation for this situation is the fact that many of the words in both categories are inherited from ancient Iranian and may have retained their gender classification from that time.

The third thematic group – nouns denoting plants – is an intriguing one, as it may represent a group in the process of transition. Nouns indicating plants which are not typically used in cooking are often inherited from ancient Iranian, and we might therefore expect these words to be evenly split with respect to gender classification. Indeed, Karamshoev (1986: 76-77) provides lists of both feminine and masculine nouns denoting plants. However, my consultants have indicated that many of the nouns he lists as feminine, such as *šitorθk* ‘rhubarb’, are in fact masculine in their dialects. One possible explanation for this, and one which should certainly be explored further, is that this category is undergoing a transition from a split category whose gender assignment is based on historical gender toward a predominantly masculine category where gender assignment is based on conceptual association.

#### 7.4.2.4 Explaining thematic categories: Umbrella nouns

At first glance, the correspondence of many thematic classes to one gender or the other may seem rather arbitrary. Why might nouns denoting sicknesses, settlements, footwear, solid food and milk-based products be masculine, while nouns denoting transportation vessels, technology, bodies of water, and clothing items (except footwear) are feminine?

Indeed, there is no immediately apparent explanation for this state of affairs, but one possible explanation brought

forth by Karamshoev (1986) is that at least some semantic classes are rooted in the grammatical gender of an umbrella noun with which its members are associated. Thus, for example, nouns denoting sicknesses, settlements, and milk-based products are associated with the masculine nouns *darǰ* ‘pain; sickness’, *ǰār* ‘city’, and *ǰūvd* ‘milk’, respectively. Similarly, feminine nouns denoting bodies of water, such as *qūl* ‘lake’, *daryo* ‘river’, and *bār* ‘sea’ are associated with the feminine noun *ǰac* ‘water’. This theoretical prospect is schematized in Figure 7.1.

This is an intriguing prospect, but undoubtedly more work is needed to lend support to it. In particular, there are several thematic categories for which no umbrella noun has been identified, perhaps most notably for the category of technology and transport. As will be seen in Section 7.4.3.3, moreover, many nouns which could easily be understood to be technological items receive masculine gender, apparently due to another semantic notion active in gender assignment, namely that of part vs. whole. For instance, we might expect *karburator* ‘carburetor’ to fall squarely within the conceptual category of technology and transport. However, *karburator* is a masculine noun, ostensibly because it is construed as a part of a car (*mošīn*), which receives feminine gender. Hence, further work is needed not only to confirm or deny the role of umbrella nouns in thematic categorization of this type, but also to determine the nuanced ways in which thematic categorization interacts with other semantic factors in gender assignment.

#### 7.4.2.5 Summary: Thematic categories

In sum, the gender classification of Shughni nouns is largely based on thematic category, where certain thematic groups correspond to masculine gender and others to feminine gender. Bodies of water, trees, and fruits constitute feminine groups, while sicknesses, vegetables, and names of settlements constitute masculine groups. Apparent exceptions in each group are possibly explained by links between certain vowels and a particular gender, or alternatively by interactions with other semantic threads which factor into gender assignment. Other thematic groups appear to be split with respect to gender. For instance, body parts, buildings, and plants all have both masculine and feminine groups. This may be due to the fact that a significant portion of each group is made up of nouns which have been inherited directly from ancient Iranian and have retained their gender from that time. Table 7.9 gives a summary of thematic classes implicated in gender classification.

Figure 7.1: Umbrella nouns and thematic categories.

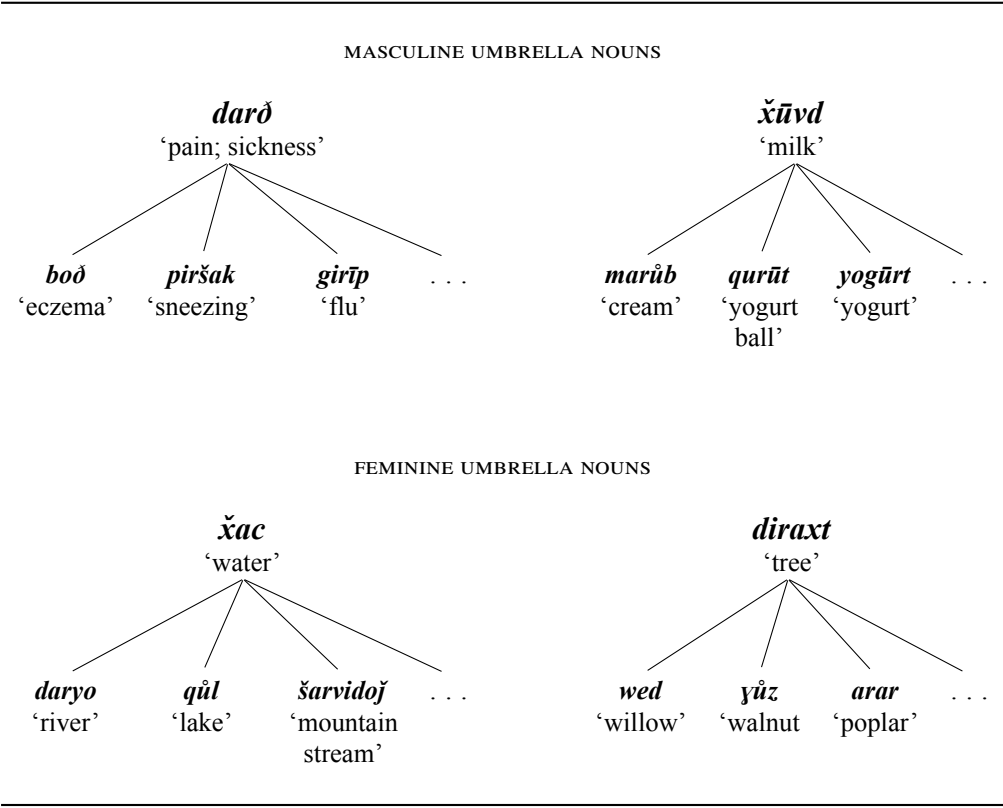


Table 7.9: Summary: Thematic groups of nouns by gender.

FEMININE GROUPS	MASCULINE GROUPS	SPLIT GROUPS
bodies of water	names of settlements	body parts
trees	milk products	buildings and building parts
fruits	vegetables	plants
technology and transport	footwear	
literature and cultural productions	sicknesses	

### 7.4.3 Other semantic factors in gender assignment

Beyond thematic categorization, three further semantic threads are found in the fabric of Shughni gender assignments: (i) abstractness and concreteness, discussed in Section 7.4.3.1, (ii) individual vs. mass, discussed in Section 7.4.3.2, and (iii) part vs. whole, discussed in Section 7.4.3.3.

#### 7.4.3.1 Abstractness and concreteness

The vast majority of nouns denoting abstract concepts in Shughni are masculine. This encompasses both native Shughni nouns as well as borrowed nouns from any language of origin. Table 7.10 provides a few examples of abstract nouns, all of which are masculine, divided into groups based on their language of origin. (A more complete list is found in Table A.4 in Appendix A.)

**Table 7.10:** Examples of (masculine) abstract nouns from various etymological sources.

Shughni			Tajik		
WORD	GLOSS		WORD	GLOSS	ТАЖИК
<i>žwǝgaʃ</i>	love		<i>mehr</i>	kindness	мехр
<i>ʃoj</i>	fear		<i>barobari</i>	equality	баробарӣ
<i>xūðm</i>	sleep		<i>unar</i>	talent	хунар
Arabic			Russian		
WORD	GLOSS	ARABIC	WORD	GLOSS	RUSSIAN
<i>išq</i>	love	عشق	<i>*sipicialnust</i>	specialty	сепциальность (F)
<i>fikri</i>	thought	فكر	<i>*uspex</i>	success	успех (M)
<i>aql</i>	mind	عقل	<i>*zavisimust</i>	addiction	зависимость (F)

Another important way the dichotomy of abstract and concrete figures into grammatical gender is in pairs of homonyms with one member belonging to each gender. Here, it is the feminine member which denotes a more concrete object, while the masculine form denotes a more abstract concept. Thus, for instance, the noun *ziv* means ‘tongue’ (the anatomical organ) when used as a feminine noun, but ‘language’ when used as a masculine noun. Further examples are given in Table 7.11. The list here is likely not exhaustive, and the identification of more such pairs is a task for future research.

Table 7.11: Pairs of FEM~MASC homonyms.

<u>WORD</u>	<u>FEM. GLOSS</u>	<u>MASC. GLOSS</u>
<i>soat</i>	wristwatch	time
<i>ziv</i>	tongue (ANAT)	language
<i>yêv</i>	mouth (ANAT)	opening; origin
<i>žīr</i>	stone (individual)	rock (material)
* <i>zapīs</i>	(e.g. musical) recording	(e.g. doctor's) appointment

Note that the list of pairs of homonyms in Table 7.11 includes the Russian borrowing *zapīs* (< запис ‘recording; appointment’). It is noteworthy, furthermore, that even borrowed Russian feminine nouns denoting abstract concepts, such as *specialnust* ‘specialty’ (from Russ. специальность, a feminine noun), are all masculine. This is despite the fact that the majority of Shughni speakers are at least conversational in Russian and interact with the language on a daily basis at least in the form of social media and entertainment. The fact that borrowed Russian lexemes denoting abstract concepts adhere strictly to the Shughni system of gender assignment, seemingly without exception, lends support to Karamshoev’s (1978: 9) contention that grammatical gender is still a strong and steadfast feature of the grammar of Pamir languages, in spite of the assertions of many scholars the contrary.

#### 7.4.3.2 Mass vs. individual

Similar to abstractness and concreteness discussed above, the semantic opposition of mass vs. individual shows up in Shughni grammatical gender in more than one way. First, a number of nouns which would typically be categorized as feminine may exhibit masculine agreement when referring to a large, indiscriminate group of the object in question. In other words, a feminine noun used as masculine typically creates a mass reading. Not all feminine nouns may be used as masculine in this way, however. There are two traditionally feminine thematic classes – fruits and tableware – whose members undergo this type of gender shift most frequently. Examples with a noun from each class, namely *mūn* ‘apple’ and *toθč* ‘bowl’, are given in (138) and (139).

(138) **Individual~Mass (F~M) readings of *mūn* ‘apple’**a. **Feminine – individual**

Čāy=i    **wam**    mūn    xūd?  
 who=3SG DEM.OBL.F apple eat.PST  
 ‘Who ate that apple?’

b. **Masculine – mass**

**Yu**    mūn    čī-nd?  
 DEM.DIR.M apple who.OBL-POSS  
 ‘Whose (pile of) apple(s) is that?’

(139) **Individual~Mass (F~M) readings of *toθč* ‘bowl’**a. **Feminine – individual**

Tu=t    **wam**    toθč    tar kā    yod?  
 you=2SG DEM.OBL.F bowl to where take.PST  
 ‘Where did you take that bowl?’

b. **Masculine – mass**

Tu=t    **wi**    toθč    tar kā    yod?  
 you=2SG DEM.OBL.M knife to where take.PST  
 ‘Where did you take that (group of) bowl(s)?’

Moreover, traditionally feminine nouns such as *mūn* ‘apple’ and *toθč* ‘bowl’ are only pluralizable in their feminine forms, and only the feminine forms are compatible with numerals. Thus, *wāð* (*cavor*) *dzalik mūn-en* ‘those (four) small.F apples.F’ is possible but \**wāð* (*cavor*) *dzulik mūn-en* ‘those (four) small.M (piles of) apples.M’ is not. This further supports the notion of a mass reading with the masculine forms of such nouns.

Secondly, a mass reading for a noun may be created not only via a shift from feminine to masculine gender, but also by either adding a denominal suffix *-xel* ‘group’ (the collective plural, see Section 5.2.4) or via a process of reduplication creating a similitive plural – e.g. *toθč* ‘bowl’ > *toθč-moθč* ‘bowls and such’ (see Section 5.2.3; this phenomenon has also been called *indefinite collectiveness* by Edelman & Dodykhudoeva 2009b: 793). In the latter case, the resulting noun is understood not only to include the noun from which it is derived, but also other related objects. Thus, for instance, the feminine noun *čêd* ‘knife’ is used in the formation of the nouns *čêd-xel* ‘(group of) knives’ and *čêd-mêd* ‘knives and such’, both of which are masculine. Examples are given in (140).

(140) **Masculine mass readings of *čêd* ‘knife’ via suffixation and reduplication**a. *čêd-xel* ‘(group of) knives’

Yu            *čêd-xel*            *kāču?*  
 DEM.DIR.M knife-PL.MASS where  
 ‘Where did those knives go?’

b. *čêd-mêd* ‘knives and such’ – i.e. indefinite collectiveness

Yu            *čêd-mêd*            *kāču?*  
 DEM.DIR.M knife-PLUR.SIM where  
 ‘Where did those knives and things go?’

Note that the semantic notion of mass vs. individual also applies to Russian borrowings. For instance, the Shughni word *vīlka* (< Ru. вилка, *vīlka* ‘fork’) typically receives feminine gender assignment in Shughni, ostensibly due to its association with the predominantly feminine thematic category of tableware. This is shown in (141a) However, like other tableware items, this word can be used as masculine and receive a mass reading, as shown in (141b).

(141) **Individual~Mass (F~M) readings of *vīlka* ‘fork’**a. **Feminine – individual**

**Wam**            *vīlka* toza kin-um o?  
 DEM.OBL.F fork clean do.PRS-1SG PQ  
 ‘Should I clean this fork?’

b. **Masculine – mass**

**Wi**            *vīlka* toza kin-um o?  
 DEM.OBL.M fork clean do.PRS-1SG PQ  
 ‘Should I clean this (group of) fork(s)?’

Note, further, that the pluralization of Russian borrowings through the group suffix *-xel* and indefinite collective reduplication is also possible. Hence, *vīlka-xel* ‘group of forks’ and *vīlka-mīlka* ‘forks and such’ are both possible formations, and both receive masculine gender assignment. The semantic notion of individual vs. mass has therefore also remained steadfast with the influx of Russian borrowings.



### 7.4.3.3 Part vs. whole

The third and final semantic thread to be discussed here is the opposition of part vs. whole, where we sometimes find a pattern in which nouns denoting whole objects are feminine, while nouns denoting their parts are masculine. This is the case, for instance, for the feminine word *diraxt* ‘tree’, whose parts, such as *xêx̌* ‘branch’, *pāřk* ‘leaf’, *bun* ‘root’, and *wiyês* ‘root’, are all masculine. This example is schematized in Figure 7.2.

**Figure 7.2:** Part/Whole relations in grammatical gender.

<u>Whole (FEM)</u>	<u>Parts (MASC)</u>	<u>GLOSS</u>
<i>diraxt</i> ‘tree’	→ <i>xêx̌</i>	branch
	→ <i>pāřk</i>	leaf
	→ <i>bun</i>	root
	→ <i>wiyês</i>	root

This pattern is found for both native Shughni words such as *diraxt* and its parts, as well as borrowed lexemes. In fact, gender assignment through the relationship of part vs. whole appears to be even more widespread in borrowed words denoting modern technology items than in native Shughni nouns, although this remains to be confirmed through further investigation. Examples of this type abound in Shughni. For instance, the borrowed Tajik word *mošīn* ‘car’ is feminine, as is the borrowed Russian word *kamaz* ‘(semi-) truck’, while their parts, such as *rūl* ‘steering wheel’, are masculine. Likewise, the borrowed Russian nouns *kampyūtir* ‘computer’ and *kofe mašīna* ‘coffee machine’ are feminine, while their parts are masculine. These examples are schematized in Figure 7.3.

The masculine words in Figure 7.3 are arguably exceptions to the feminine “technology and transport” category, discussed in Section 7.4.2.2 and summarized in Table 7.7. It is difficult to argue, for instance, that the word *pracesur* is not a modern technological item. It appears, therefore, that in this case, the part vs. whole aspect of gender assignment wins out against gender assignment by thematic category. However, it is unclear whether part vs. whole always takes precedence over assignment by thematic category, and future studies are therefore needed to explore this and other phenomena in Shughni grammatical gender.

**Figure 7.3:** Part/Whole relations in borrowed Russian lexemes.

<u>Whole (FEM)</u>	<u>Parts (MASC)</u>	<u>GLOSS</u>
<i>moštīn</i> // <i>kamaz</i> 'car // truck'	→ <i>karbūrator</i>	carburetor
	→ <i>rūl</i>	steering wheel
	→ <i>mator</i>	motor
	→ <i>kabīnka</i>	cab (of a truck)
<i>kampyūtir</i> 'computer'	→ <i>pracesur</i>	processor
	→ <i>mīška</i>	mouse
	→ <i>klaviatūra</i>	keyboard
<i>koфе mašina</i> 'coffee machine'	→ <i>kanteynir</i>	container
	→ <i>būnkir</i>	bunker
	→ <i>kapūčīnatur</i>	cappucinator

Indeed, the role of the part vs. whole relationship in the gender classification of Shughni nouns is not thoroughly documented. It is not fully understood, for instance, to what extent gender assignment based on the part vs. whole relation occurs in native Shughni words (or words which were borrowed many centuries ago), and to what extent it occurs in recently borrowed lexemes such as the Russian words given in Figure 7.3. This aspect of gender assignment in Shughni is especially ripe for investigation. Both corpus-based and experimental studies could be used to contribute to a better understanding of this phenomenon.

## **Part III**

# **Verbs**



## Chapter 8

# Verb Stems, Inflection, and Modification

This chapter provides an introduction to the fundamental building blocks of the Shughni verbal system. It is meant to serve as a foundation for the following three chapters, which look in more detail at the various aspects of this system. For instance, this chapter provides a first look at the concept of the *verb stem* in Shughni, thereby setting the stage for the detailed discussion in Chapter 9 on patterns of regular and irregular verbs, as well as the ways the verbal system has changed (and continues to change) in recent times. This section also examines the basic uses of each stem and their correspondences to temporal and aspectual reference, topics which are dealt with fully in Chapter 10 on TAM and evidentiality. The information in this chapter will likewise provide the reader with a base for understanding argument structure and issues of transitivity, which are examined in Chapter 11. Once the reader has read the present chapter, they should be able to skip to any of the following three chapters without any issues.

The remainder of this chapter proceeds as follows. Section 8.1 provides an overview of the most fundamental elements of the verbal system, namely **verb stems** and **verb-stem paradigms**, as well as **the basic uses of different verb stems**. Section 8.2 then discusses **verbal inflection**, including person/number agreement and negation. Section 8.3 examines the **copula**, and finally, Section 8.4 describes **verbal modification**.

## 8.1 Verb stems and usage

Arguably the most basic element of the Shughni verbal system is the verb stem, upon which all possible combinations of tense, aspect, and mood are built, and which serves as the locus for agreement with subjects. This section provides a first look at both the forms and usage of Shughni verb stems.

The succession of topics in this section is as follows. Section 8.1.1 first gives an introduction to verb stems and verb-stem paradigms. The notion of regular and irregular verbs in Shughni is also introduced, and an example of each type of verb is provided. Section 8.1.2 then presents basic information on each of the four types of verb stems found in Shughni. Both historical information and a description of basic uses of each stem is given.

### 8.1.1 Verb-stem paradigms

Modern Shughni verbs have four stems traditionally labeled *present*, *infinitive*, *past*, and *perfect* (the uses of each stem are summarized in Section 8.1.2). Here, I refer to the group of a verb's stems collectively as its *verb-stem paradigm*; thus, each Shughni verb has a paradigm of four stems. To refer to the more abstract concept of the verb itself, I follow native Shughni speakers in using a verb's citation form, which is equivalent to its infinitive stem together with the suffix *-ow*. Outside of example sentences, I do not separate this morpheme from the stem with a hyphen. Thus, the Shughni verb *qīwdow* 'call' has the verb-stem paradigm in (142), and the Shughni verb *vīdow* 'bring' has the verb-stem paradigm in (143).

(142) **Regular verb *qīwdow* 'call'**

PRS	<b><i>qīw-</i></b>
INF	<b><i>qīwd</i></b>
PST	<b><i>qīwd</i></b>
PRF	<b><i>qīwǰ</i></b>

(143) **Irregular verb *vīdow* 'bring'**

PRS	<b><i>vār-</i></b>
INF	<b><i>vīd</i></b>
PST	<b><i>vūd</i></b>
PRF	<b><i>vūǰǰ</i></b>

Shughni verbs may be either regular or irregular. Regular verbs are all alike in following a single recognizable pattern, whereas irregular verbs may deviate from this pattern in a variety of ways. In regular verbs, such as *qīwdow* in (142), the form of each of the four verb stems is predictable given the form of any other verb stem. Past stems and infinitive stems are equivalent to the present stem (*qīw-*) with an additional stop *-t/-d* (*qīwd*), where a *-d* is

found after a voiced obstruent, semivowel (*w* or *y*), or a vowel, and a *-t* is found elsewhere (i.e. after a voiceless consonant, nasal *n* or *m*, or liquid *l* or *r*). Perfect stems are equivalent to the past stem with an affricate *-j/-č* in place of the stop (*q̄t̄wj̄*). The choice of affricate in perfect stems is governed by the same rule as with *d/t* in past and infinitive stems. This pattern, together with examples, is schematized in (144) and (145).

- (144) **Regular verb: Present stem ends in voiced obstruent, semivowel, or vowel**      (145) **Regular verb: Present stem ends in voiceless consonant, nasal, or liquid**

STEM	ENDG.	EX.
PRS	-	<i>q̄t̄w-</i>
INF	<b>-d</b>	<i>q̄t̄w-d̄</i>
PST	<b>-d</b>	<i>q̄t̄w-d̄</i>
PRF	<b>-j̄</b>	<i>q̄t̄w-j̄</i>
GLOSS		‘call’

STEM	ENDG.	EX.
PRS	-	<i>mol-</i>
INF	<b>-t</b>	<i>mol-t̄</i>
PST	<b>-t</b>	<i>mol-t̄</i>
PRF	<b>-č</b>	<i>mol-č̄</i>
GLOSS		‘rub’

An irregular verb, such as *v̄idow* in (143), is one which does not follow one of the two patterns in (144) and (145). Most commonly, the stem vowel or final consonant(s) differ in one or more of an irregular verb’s stems. Compare, for instance, the difference in stem vowel between the infinitive *v̄id̄*, present *v̄āṛ-*, and past *v̄ūd̄*, as well as the presence of a stem-final *-r-* in the present stem, *v̄āṛ-* but its absence elsewhere. Regular and irregular verbs are treated in detail in Chapter 9, specifically in Section 9.1.

Note that in some publications on Shughni – e.g. Bakhtibekov 1979: 43-44; Dodykhudoeva 1988: 53; Edelman & Dodykhudoeva 2009b: 797 – a separate pluperfect stem is listed, which is equivalent to the perfect stem with the addition of the stressed suffix *-at*. For instance, the verb *q̄t̄wdow* would have pluperfect stem *q̄t̄wj̄-at* and the verb *v̄idow* would have pluperfect stem *v̄ūj̄j̄-at*. Bakhtibekov indicates that this “stem” is used in both traditional pluperfect environments – i.e. to situate an event in the past before another event in the past – as well in certain irrealis environments. In the other publications, namely Dodykhudoeva 1988 and Edelman & Dodykhudoeva 2009b, the use of the stem is not treated in any detail.

For the following three reasons, I have decided to leave this form out of the list of verb stems and out of the forms that factor into verb-stem paradigms. First, in my fieldwork on the variety of the language spoken in Khorugh, I have only ever encountered the irrealis usage of these forms, namely with counterfactual desideratives and conditionals (see Section 10.4), and consultants have also indicated that this form is not used in traditional pluperfect

contexts. Second, the suffix *-at* is generally optional. That is, even in irrealis contexts where the suffix *-at* is permitted, a bare perfect stem is equally felicitous as a perfect form suffixed with *-at*. And third, this form is always predictably derived from the perfect with the addition of a suffix, whereas other stems are not always predictably derived from one another. For these reasons, I do not count this form as a separate stem in this dissertation. Henceforth, I refer to it as the *counterfactual perfect*, and I gloss the suffix *-at* as a *counterfactual suffix* (CF).

### 8.1.2 Basic uses of verb stems

The name of each stem generally corresponds to its most fundamental usage. However, all stems have additional, sometimes unexpected uses beyond those exhibited here. In what follows, a brief history of each stem is given along with examples of its basic uses. Stems' non-canonical uses are also mentioned, and the reader is directed to the section(s) of the thesis where more information on these topics can be found.

**Present stems** are the oldest of the four stem types, descending ultimately from Proto-Indo-European present (or imperfective) stems (e.g. Dodykhudoeva 1988: 6). In modern Shughni, present stems, unlike past and perfect stems, agree in person and number via agreement suffixes, rather than clitics. (Section 8.2.1, which follows this section, deals with person and number inflection via suffixes and clitics).

Present stems are most commonly found in non-past contexts and are compatible with both present and future temporal reference. They are also used to form imperatives. The second-singular imperative is generally equivalent to a bare (although sometimes shortened) present stem. Below, example (146) illustrates a present-tense use of the present stem *vār-* 'bring', and example (147) provides an example of its use as an imperative. Note that in the latter example, the present stem is shortened from *vār-* to *va*. Verb-stem shortening of this kind is a rather extensive phenomenon in the language and is the topic of Section 9.1.5.

#### (146) Basic present-stem construction

Māš ǰac az daryo **vār**-ām.  
 we water from river bring.PRS-1PL  
 'We are bringing water from the river.'

#### (147) Second-singular imperative

Ku yi misol **va**, di=ta carāng kor kin-en.  
 please one example bring.2SG.IMP DEM.OBL.M=FAC how work do.PRS-3PL  
 'Please give an example of how it's done.'



**Past stems** in modern Shughni generally descend from Old Iranian participles built on the suffix *\*-ta*, which supplanted the older inflected perfective stems and were initially used in periphrastic constructions, often together with a copula (e.g. Dodykhudoeva 1988; Haig 2008). In some cases, past stems may also descend from verbal nouns with the suffix *\*-ti-*, although these are more commonly the predecessors of infinitive stems. Past stems are most commonly used in constructions with past temporal reference. An example of such a construction is given in (148).

(148) **Basic past-stem construction**

Dev            ṣac-en=en    biyor    **vūd**.  
 DEM.OBL.PL water-PL=3PL yesterday bring.PST  
 ‘They brought those waters yesterday.’

Importantly, both present and past stems are sometimes used in contexts where they seem to deviate from their typical temporal reference. In particular, in some instances, present stems are ostensibly used with past temporal reference, while past stems in some cases seemingly have future temporal reference. It is argued in Section 10.2, however, that present and past stems nonetheless still properly entail non-past and past tenses, respectively, and that apparent mismatches between stem and tense suggest that these stems in fact encode relative tense, rather than absolute tense. Such mismatches are nonetheless not overly common and are restricted to specific types of constructions.

**Perfect stems**, like past stems, descend from Old Iranian participles in *\*-ta-*, but with the addition of an additional suffix *\*-ka-*. The reflex of this suffix in modern Shughni is the affricate *č/j*, which is found in all non-gender-distinguishing perfect stems and the masculine form of gender-distinguishing perfect stems. In many perfect stems, a fricative precedes the final affricate. This is most often *θ/ð* (e.g. *vuðj* ‘been’, cf. past stem *vud* ‘was’), which is the reflex of the *\*t* in *\*-ta-*, but in others it is the velar fricative *ɣ* (as in *vūɣj* ‘brought’). To my knowledge, non-gender-distinguishing modern Shughni perfect stems, as well as the masculine form of gender-distinguishing perfect stems, always contain the same stem vowel as their corresponding past stem. The only factor which renders perfect stems not predictably derived from past stems is the presence or absence of a preceding fricative (and which fricative is found – *ð* or *ɣ* – in cases where one is present).

Modern Shughni perfect stems display many of the uses of the perfect across the world’s languages, such as to convey experientiality, as in (149), or to express a completed eventuality which has relevant consequences for the

present, as shown in (150). Note that I label these two types of readings *experiential* and *resultative*. (See Bertrand et al. 2022 on these two types of uses of the perfect cross-linguistically; see also the discussion in Section 10.2.3 and references cited there.)

(149) **Perfect stem: Experiential usage**

Wuz=um tar Kanada **vuðj**.  
 I=1SG in Canada be.PRF.M  
 ‘I have been to Canada.’

(150) **Perfect stem: Resultative usage**

Ominā=yi ūži pomidor **vūŷj**, ačga nist darkor.  
 Omina=3SG already tomato bring.PRF anymore COP.NEG necessary  
 ‘Omina already brought tomatoes. We don’t need any more.’

Beyond these cross-linguistically common usages, perfect stems in Shughni have two further uses: (i) a *modal* usage to show irrealis mood in counterfactual irrealis clauses, commonly with the counterfactual suffix *-at* as mentioned above, and (ii) an *evidential* usage to indicate that a speaker has indirect evidence for the information she is conveying. These, like apparent temporal-reference mismatches of present and past stems, are addressed in Section 10.2 of Chapter 10 on tense, aspect, and mood.

Finally, **infinitive stems** are commonly used as verbal nouns. They can appear in subject or object position and may take a variety of affixes used with nominals, such as locative and dative postpositions. An example of an infinitive stem as a verbal noun in subject position is shown in (151) (together with an object *ǰac* ‘water’, also bolded).

(151) Mu-rd az daryo **ǰac vīd** nist xuš.  
 me-DAT from river water bring.INF COP.NEG pleasant  
 ‘I don’t like bringing water from the river.’

Infinitive stems are also commonly used together with locative elements and auxiliary verbs in constructions which distinguish certain aspectual nuances, such as inchoative, progressive, and prospective aspects. Aspectual distinctions of this kind are discussed in Section 10.3.

A summary of Shughni verb stems and inflection is provided in Table 8.1.

**Table 8.1:** Shughni verb stems: Form, inflection, and usage.

<u>STEM</u>	<u>FORMAL MARKERS</u>	<u>AGR. TYPE</u>	<u>USES</u>
<b>Present</b>	–	present suffixes	non-past indicative non-past subjunctive imperatives
<b>Past</b>	stem-final <i>-t/-d</i>	past clitics	past indicative future adverbial clauses with complementizer <i>di</i>
<b>Perfect</b>	stem-final <i>č/-j</i>	past clitics	canonical perfect evidential (indirect information) past-tense irrealis resultative participles
<b>Infinitive</b>	stem-final <i>t/d</i> (masc.) suffix <i>-ow</i> (in some contexts)	–	citation form of verbs aspectual constructions with loc. adp. verbal nouns

## 8.2 Verbal inflection

Shughni verbs inflect for four categories: person, number, negation, and in a subset of unaccusative verbs, gender (on other ways unaccusative verbs are distinguished from transitive and unergative verbs, see Section 4.4.2). The expression of gender, unlike the first three categories, is not done by suffixation or cliticization, but rather by stem-internal vowel and consonant alternations. This topic is left aside for now and taken up in Section 9.2. This subsection focuses only on person and number inflection, discussed in Section 8.2.1, and verbal negation, examined in Section 8.2.2.

### 8.2.1 Person and number agreement

Person and number agreement in Shughni verbs occurs through two sets of morphemes: suffixes are used with present stems, and second-position clitics are used with past and perfect stems. The former, which I refer to as

*present suffixes*, attach directly to the verb stem and are never used with any stem other than present stems. The latter, which I refer to as *past clitics*, target the right edge of the first syntactic constituent in a clause and are more free with respect to the type of constituent to which they attach. Hence, in the sense of Zwicky (1977) and Zwicky & Pullum (1983), present suffixes behave as suffixes, while past-tense clitics behave as morphophonological clitics. The examples below provide an initial look at the use of a present suffix (in 152) versus the use of past clitic (in 153). The verb stem in each is bracketed. Note that the past-tense clitic in (153) attaches to the subject *wuz* ‘I’, which is the first syntactic constituent in the clause.<sup>1</sup>

(152) **Present 1sg. suffix -um**

Wuz=ta x̌umne der [andidz]-**um**.  
 I=FAC tomorrow late get.up.PRS-1SG  
 ‘I will get up late tomorrow.’

(153) **Past 1sg. clitic =um**

Wuz=**um** biyor der [andüyɖ].  
 I=1SG yesterday late get.up.PST.M  
 ‘I got up late yesterday.’

Despite their striking syntactic differences, Shughni present suffixes and past clitics are nearly identical in form. They differ only in two cells: second- and third-person singular. The second-person singular present suffix is *-(y)i*, while the past-tense clitic is *=(a)t*. The third-singular present suffix is either *-d* (following voiced obstruents and vowels) or *-t* (elsewhere), as in past and infinitive stems, while the third-singular past clitic is either  $=\emptyset$  (used with unaccusative verbs) or *=(y)i* (used with unergative and transitive verbs). The latter alternation implicating unaccusative verbs, on the one hand, and unergative and transitive verbs, on the other, has been called *vestigial ergativity* (Stump & Hippisley 2011; Hippisley & Stump 2011; Parker 2020) and is examined in detail in Section 12.4 on morphosyntactic alignment. The paradigm of present suffixes and past clitics are given in Table 8.2 and Table 8.3, respectively.

**Table 8.2:** Present suffixes.

	SG	PL
<u>1</u>	<b>-um</b>	<b>-ām</b>
<u>2</u>	<b>-i</b>	<b>-et</b>
<u>3</u>	<b>-d/-t</b>	<b>-en</b>

**Table 8.3:** Past clitics.

	SG	PL
<u>1</u>	<b>=um</b>	<b>=ām</b>
<u>2</u>	<b>=(a)t</b>	<b>=et</b>
<u>3</u>	<b>=i/∅</b>	<b>=en</b>

<sup>1</sup> The phenomenon whereby agreement suffixes are reserved for present stems and second-position clitics for past and perfect stems, is not unique to Shughni. It is found in all other Pamir languages except Munji (Sokolova 1973), as well as a number of Western Iranian languages, including Davani (Moghaddam 2016) and Tati (Taherkhani 2019).

The peculiarities in form and syntax of present suffixes and past clitics can be at least partially explained from a historical linguistic perspective. Whereas present suffixes descend from the verbal agreement suffixes used with ancient Iranian present (or *imperfective*) stems, past clitics descend from second-position pronominal enclitics used in a particular kind of Old Iranian participial construction with a passive or impersonal interpretation, namely the *manā kartam* construction (Dodykhudoeva 1988: 6-7; Haig 2008; Pireiko 1975: 186; Sokolova 1973: 132-134; a.o.). Eventually, the participles of the *manā kartam* construction became modern Shughni past stems, and the second-position pronominal enclitics became Shughni past clitics. Hence, despite their distinct etymological origins, modern Shughni present suffixes and past clitics have converged in their function in the language. (See especially Haig 2008 and Jügel 2012 on the development of the modern Iranian past tense.)

The current formal similarities between present suffixes and past clitics are due to more recent contamination, not only with each other, but also with inflected forms of the Old Iranian copula, which were also used in Old Iranian participial constructions (e.g., Dodykhudoeva 1988: 6–8). Present suffixes and past clitics differ in form where contamination has not occurred completely. Specifically, the second-singular present suffix *-i* is the reflex of the Proto-Iranian imperfective agreement suffix *\*-ahi*, and the second-singular past clitic *=(a)t* is the reflex of the second-position pronominal clitic *=\*tai*. The third-singular present suffix *-t/-d*, for its part, is the reflex of the Proto-Iranian imperfective agreement suffix *\*-ati*, while the third-singular past clitic *=(y)i* is the reflex of the pronominal clitic *=\*hai*. (The full etymology of present suffixes and past clitics is given in Table B.1 in Appendix B.)

Finally, note that the set of past-tense clitics used with unaccusative verbs – i.e. that for which the third-singular form is realized as a zero-morpheme – is identical to the second-position clitics used in copular constructions, as discussed below. Shughni past clitics and copular clitics share the same etymological origin, but in the modern language, they display important semantic and morphosyntactic differences. Section 8.3 below gives a detailed overview of these differences.

### 8.2.2 Verbal negation

Verbal negation in Shughni is carried out by two prefixes, both of which receive primary stress. The usage of each prefix is divided roughly along the lines of realis and irrealis moods. The prefix *na-*, which I refer to as the *realis negation marker*, is used in all indicative environments, while the prefix *mā-*, which I refer to as the

*prohibitive marker*, is used in most irrealis environments, including with all imperatives. This fundamental division is exhibited in examples (154) and (155); the negation prefix in each example is bolded.

(154) **Realis verbal negation marker: *na-***

- a. Pi kū-yen=ta ačaθ malā **na**-moz-en.  
up.in mountain-PL at.all house NEG-build.PRS-3PL  
'They don't build any houses up in the mountains.'
- b. Wi=yen tar ūniversitet qabūl **na**-čūd.  
him=3PL into university accept NEG-do.PST  
'They didn't accept him into university.'

(155) **Prohibitive (irrealis) verbal negation marker: *mā-***

- a. Lap řac birêz, tu lāk kasal **mā**-sāw-i.  
much water drink.2SG.IMP you SUBR sick PROH-become.PRS-2SG  
'Drink a lot of water so you don't get sick.'
- b. Disga mu qatī gāp **mā**-đa!  
like.that me with word PROH-hit.2SG.IMP  
'Don't speak with me like that!'

In at least one instance, we find the prefix *na-*, typically used for indicative clauses, where we might expect to find the prohibitive marker *mā-* instead. This is the case of conditional constructions built on the subordinator *ca*, which targets the left edge of the verb (i.e. second-to-final position in the clause). These constructions pattern as irrealis in that they require the copular verb *vidow* 'be', rather than second-position copular clitics. (Section 8.3 addresses copular forms in more detail.) However, with respect to negation, they resemble indicative clauses in their compatibility with the prefix *na-*, rather than the prohibitive *mā-*. In conditional sentences of this kind, the prefix *na-* is used for negation in both the subordinate (conditional antecedent) and the matrix (conditional consequent) clauses, as shown in (156).

(156) **Negation in *ca* conditionals: Realis prefix *na-***

a. **Present-future conditional**

- Mu nān aga pi čīd ca **na**-vi-d, māš=ta tar mu mūm xez  
my mother if up.in house SUBR NEG-be.PRS-3SG we=FAC to my grandmother location  
**na**-sāw-ām.  
NEG-go.PRS-1PL  
'If my mom isn't at home, we won't go to my grandmother's'

b. **Counterfactual conditional**

Mu nān aga pi čīd ca **na**-vic-at, māš=ām tar mu mūm xez  
 my mother if up.in house SUBR NEG-be.PRF.F-CF we=1PL to my grandmother location  
**na**-sic-at.  
 NEG-go.PRF.PL-CF  
 ‘If my mom hadn’t been at home, we wouldn’t have gone to my grandmother’s’

This distribution is schematized in Figure 8.1.

**Figure 8.1:** Division of labor in verbal negation prefixes.

REALIS		IRREALIS			
		COND	IMP	DESID	PURP
Realis negation prefix <i>na</i> -		Prohibitive prefix <i>mā</i> -			

### 8.3 Copular constructions

The *copula* here is taken to be ‘a linguistic element which co-occurs with certain lexemes . . . when they function as a predicate nucleus’, but which ‘does not add any semantic content to the predicate phrase it is contained in’ (Pustet 2003: 5, drawing from Hengeveld 1992: 32 and Stassen 1997: 65). Within the Indo-European context, the Shughni copula fulfills the familiar role of linking subjects with predicate nominals and adjectives.

Shughni possesses two types of copular constructions which are used in distinct grammatical environments. In present-tense indicative environments, there is no overt copular verb, but rather second-position clitics are used to indicate the person and number values of the subject. An example of this type of construction is given in (157).

(157) **Present indicative: Second-position clitic only**

Wuz=**um** šič tar sūr.  
 I=1SG now at wedding  
 ‘I am at the wedding now.’

In environments which are not present indicative – including past, perfect, infinitival, and present irrealis environments – the copular verb *vidow* is used. An initial example containing the past stem of *vidow* is given in (158).

(158) **Past: Copular verb *vidow* ‘be’**

Yā      ŷinik   biyor   tar sūr      **vad.**  
 DEM.DIR.F woman yesterday at wedding be.PST.F  
 ‘That woman was at the wedding yesterday.’

The remainder of this section examines various aspects of Shughni copular constructions in more detail. First, Sections 8.3.1 and 8.3.2 describe each type of the two types of constructions and the specific grammatical contexts to which it is confined. Section 8.3.3 then looks at the usage of the copula from the standpoint of its compatibility with different lexical classes. Third, Section 8.3.4 presents existential and possessive constructions, which are also formed via copular-like constructions, albeit with important differences. And finally, Section 8.3.5 summarizes.

### 8.3.1 Second-position copular clitics

The second position clitics used in present-tense copular constructions are identical in form to those used with subjects of unaccusative verbs in past and perfect constructions. Their paradigm is given in Table 8.4.

**Table 8.4:** Second-position clitics used in present realis copular constructions.

	SG	PL
1	= <i>um</i>	= <i>ām</i>
2	=( <i>a</i> ) <i>t</i>	= <i>et</i>
3	∅	= <i>en</i>

Regarding their distribution within the clause, the clitics used in copular constructions behave in the same way as past clitics in that they obligatorily attach to the right edge of the first syntactic phrase of the clause. The examples in (159) illustrate this distribution; (159a) shows a clitic attaching directly to the subject, and (159b) shows a clitic attaching instead to a temporal adverb.

(159) **Present-tense indicative copular clitics**

- a. Māš=**ām** šič   tar čīd.  
 we=1PL now at house  
 ‘We are at home now.’



- b. Šič=**ām** tar čīd (māš).  
 now=1PL at house (we)  
 ‘Now we are home.’

Note further that, as with past and perfect stems of unaccusative verbs, there is no overt realization of a clitic in the third-person singular, as shown in (160):

(160) **3sg-copular construction: No clitic**

- a. Yu mollim.  
 he teacher  
 ‘He is a teacher.’
- b. Tu mošīn tar kā?  
 your car at where  
 ‘Where is your car?’

Present-tense copular constructions are negated with the particle *nist*, which in neutral word order occurs clause-finally. This pattern, exhibited in (161), constitutes an important difference between the use of second-position clitics in present-tense copular constructions and their use with past and perfect stems. In the latter usage, the particle *nist* is not used in negation; instead, the verb stem takes the verbal negation prefix *na-*.

(161) **Negation in present-tense realis copular constructions: Particle *nist***

- a. Māš=**ām** šič tar čīd **nist**.  
 we=1PL now at house NEG.COP  
 ‘We are not at home now.’
- b. Yu mollim **nist**.  
 he teacher NEG.COP  
 ‘He is not a teacher.’

On this note, some words are in order regarding the relation of the second-position clitics used in present-tense copular constructions to those used with past and perfect verbs. As mentioned above, clitics used in copular constructions are formally identical to the second-position clitics used with past and perfect stems of unaccusative verbs (see Table 8.3 in Section 8.2.1). Indeed, from a historical linguistic standpoint, both sets of clitics have arisen from the contamination of bound pronominal elements with inflected forms of the ancient Iranian copula

(e.g. Sokolova 1967; 1973; Dodykhudoeva 1988; cf. the discussion on verbal inflection in Section 8.2.1). In Old Iranian, the copula was used as part of the participial constructions which developed into the modern Shughni past tense (e.g. Dodykhudoeva 1988; Haig 2008). At one point in the history of Shughni, therefore, basic copular constructions and past-tense clauses built on the copula and a participle may have been of the same fundamental type.

Despite this shared history, however, two pieces of evidence motivate the separation of these two formally identical sets of clitics in modern Shughni. First, past-tense clitics take a special form *=i* to mark 3rd-person singular subjects of unergative and transitive verbs, whereas this morpheme never appears in copular constructions. And second, as just described, in copular constructions, second-position clitics are negated with the particle *nist*, whereas in their usage with past and perfect stems, negation instead takes the form of a canonical verbal negation suffix *na-* or *mā-* and appears directly on the verb (cf. *wuz=um na-nūst* ‘I didn’t sit’ vs. *\*wuz=um nūst nist*). With this in mind, I take the position that from a synchronic standpoint, second-position clitics used with past and perfect stems are a different kind of morpheme than those used in copular clauses (see Parker 2020, where it is argued that past-tense clitics are the result of clitic doubling).

### 8.3.2 Copular verb *vidow*

The copular verb *vidow* ‘to be’ is of the same root as other familiar Indo-European copulas, including Persian *budan* and English *be* (see, e.g. Dodykhudoeva 1988). This verb is used all copular constructions except present realis, namely present irrealis, past, perfect, past irrealis, and infinitival environments. *Vidow* is a proper verb: it has a full paradigm of inflection and, as any other verb in the language, it is negated with the prefix *na-* in realis environments and with the prohibitive marker *mā-* in irrealis environments. Its verb-stem paradigm is given in Table 8.5, and its present-tense inflectional paradigm is given in Table 8.6.

**Table 8.5:** Verb-stem paradigm of *vidow* ‘be’.

INF	<b><i>vid(ow)</i></b>	
PRS	<b><i>vi-</i></b>	
PST	<b><i>vud</i></b> (M)	<b><i>vad</i></b> (F)
PRF	<b><i>vuđj</i></b> (M)	<b><i>vic</i></b> (F)

**Table 8.6:** Present-tense paradigm of *vidow* ‘be’.

	<u>SG</u>	<u>PL</u>
1	<b><i>vīm</i></b>	<b><i>viyām</i></b>
2	<b><i>viyi</i></b>	<b><i>viyet</i></b>
3	<b><i>vid</i></b>	<b><i>viyen</i></b>

Present-tense irrealis constructions which call for the present stem of the verb *vidow* include subjunctive constructions and imperative constructions. The former may be built on the subordinator conjunction *lāk* ‘so that’, on epistemic particles such as *nog* ‘(I) wish’ or *boyad* ‘must’, or on conditionals with the particle *aga* ‘if’ and subordinator *ca*. Examples of each are given in (162); the copula is bolded. Note that in all of the three examples in (162), the use of a second-position copular clitic in the place of (or alongside) the verb *vidow* is illicit.

(162) **Present-tense irrealis copula: Verb *vidow***

a. **Copula with subordinating conjunction *lāk* ‘so that’**

Māš-ard dūs-ik-aθ xêvd-ow darkor, lāk vegā tar večer mot **mā-vi-yām**.  
 US-DAT little-DIM-AUG sleep-INF-NMZ necessary SUBR evening at party tired PROH-be.PRS.IRR-1PL  
 ‘We need to sleep a bit so we aren’t tired at the party tonight.’

b. **Copula with epistemic *nog* ‘(I) wish’**

Nog Davlat=at Bakhmal lap-aθ boy **vi-yen**.  
 I.hope Davlat=and Bakhmal very-AUG wealthy be.PRS.IRR-3SG  
 ‘I hope Davlat and Bakhmal are very wealthy.’

c. **Copula in present-tense conditional**

Mu nān (aga) tar čīd ca **vi-d**, salūm wam-ard lu.  
 my mother (if) at home SUBR be.PRS.IRR-3SG hello her-DAT say.2SG.IMP  
 ‘If my mom is home, say hello to her.’

The imperative of the copular verb *vidow* is formed as any other verb used in the imperative. That is, its bare present stem, in this case *vi*, is used in second-person singular imperatives, and its full second-person plural form is used in second-person plural imperatives. When negated, it also calls for the prohibitive marker *mā-*. Examples are given in (163). (See Section 10.4 for a more detailed discussion on grammatical mood.)

(163) a. Disga **mā-vi!**

like.that PROH-be.2SG.PRS.IMP  
 ‘Don’t be like that!’

b. Salūmat **vi-yet!**

healthy be.PRS.IMP-2PL  
 ‘Be well!’

In the case of past and perfect copular constructions, the corresponding stem of the verb *vidow* ‘be’ is used. Note that in its past and perfect stems, the verb *vidow* behaves as an unaccusative verb in inflecting for gender and number. It also requires the use of a second-position clitic from the paradigm of unaccusative past-tense clitics to mark the person and number features of the subject. Examples are given in (164).

(164) **Past and perfect copula: Verb *vidow***

- a. Māš=ām biyor tar čīd na-**vad**.  
 we=1PL yesterday at house NEG-be.PST.PL  
 ‘We weren’t at home yesterday.’
- b. Māš=ām uži tar Dūšanbi **vic**.  
 we=1PL already in Dushanbe be.PRF.PL  
 ‘We have already been in Dushanbe.’

The perfect stem of *vidow* may also be used in counterfactual irrealis copular constructions. Here, the irrealis stem, marked by the stressed counterfactual suffix *-at*, is used, as shown in (165):

(165) **Past irrealis copular construction**

- Aga yu ar Dūšanbi ca **vuđj-at**, wuz=um wi **wīnč-at**.  
 if he in Dushanbe SUBR be.PRF.M-CF I=1SG him see.PRF-CF  
 ‘If he were in Dushanbe, I would have seen him.’

Summarizing briefly, the distribution of Shughni copular forms is split along the lines of tense and mood. Second-position copular clitics, which are identical to past clitics used with unaccusative verbs, are used without an overt copular verb in present indicative environments. The verb *vidow* ‘be’, which has all the trappings of any other Shughni verb, including a full paradigm of stems and typical inflectional patterns, is used in other environments. The distribution of Shughni copular forms across grammatical environments is schematized in Table 8.7.

**Table 8.7:** Distribution of the Shughni copula across grammatical environments.

	PRESENT		PAST/PERFECT	
	INDICATIVE	IRREALIS	INDICATIVE	IRREALIS
2nd-position clitics only			<i>vidow</i> ‘be’	

### 8.3.3 Compatibility of the copula with different lexical classes

With respect to its usage with particular lexical classes as predicate nuclei, the Shughni copula is unremarkable within the Indo-European context. It is obligatorily used with both nominal and adjectival predicates, including certain deverbal participles, most notably the adjectival resultative participle and aspectual constructions built on the combination of an infinitive verb and an adposition. Each type of usage is displayed in the examples below.

The examples in (166) show the copula used in nominal predication, with an example of predicative usage – i.e. where the predicate nominal is non-referential – and specificational usage, where the predicate nominal is a definite noun phrase. (For a taxonomy of copular constructions, see den Dikken and O’Neill 2017: 3-22.) Note that the use of the copula is obligatory here and in all environments where it is used. Its obligatory usage is shown in these examples but will not be indicated hereafter.

#### (166) Copula with predicate nominals

##### a. Predicative usage with nominal predicate

Wāð\*(=en) bašānd mollim-en.  
 they=3<sub>PL</sub> good teacher  
 ‘They are good teachers.’

##### b. Specificational usage with nominal predicate

Samīra=yat Sino\*(=yen) māš sār bašānd mollim-en.  
 Samira=and Sino==3<sub>PL</sub> our <sub>SUP</sub> good teacher-PL  
 ‘Our best teachers are Samira and Sino.’

Examples (167) show the use of the copula with adjectival predicates. Note that in (167b), the third-singular copula is realized as a zero-morpheme.

#### (167) Copula with predicate adjectives

a. Tu=t xušrūy.  
 you=2<sub>SG</sub> beautiful  
 ‘You are beautiful.’

b. Yam restoran awqot=∅ bamazā.  
 DEM.DIR restaurant food=COP.3<sub>SG</sub>.PRS tasty  
 ‘This restaurant’s food is tasty.’

Finally, the examples in (168) illustrate the use of the copula in different constructions. In example (168a), it is used with an adjectival resultative participle, syntactically an adjective, in predicate function, and in example (168b), it is used with a locative-infinitival construction expressing prospective aspect.

(168) **Copula in deverbal constructions**

- a. Fuk wāð      čīd-en=**en**      mīzj-in.  
 all DEM.DIR.PL house-PL=3PL build.PRF-PASS  
 ‘All those houses are built.’
- b. Māš=**ām** ūži      či      naχtīd.  
 we=1PL already LOC leave.INF  
 ‘We’re already about to leave.’

### 8.3.4 Existential and possessive constructions

Existential constructions in Shughni, which express the presence or absence of something at a particular location (corresponding to English *there is/are*), are also built on the copula. While existential constructions form a clear semantic subtype of copular construction (Pustet 2003; den Dikken & O’Neill 2017), they exhibit important differences with canonical copular constructions like those discussed in the preceding subsections. First, unlike canonical copular constructions, existential constructions are compatible with the morpheme *yast* – a frozen form of the 3SG copula *\*asti* (Edelman 1974) – which is used as an existential emphatic particle (glossed *EEP*) in indicative, affirmative present-tense contexts. And second, existential copular constructions show definiteness effects (cf. Milsark 1974; 1977), such that definite NPs may not serve as subjects of existential constructions.

Initial examples of existential constructions are given in (169). Note that in the present-tense example in (169a), the emphatic particle *yast* might be used, for instance, by a speaker who disagrees with someone who has claimed that there are currently no good universities in Khorugh. In a neutral context, such as introducing a fact about the city of Khorugh, *yast* would not be used. In the past-tense existential construction in (169b), the use of *yast* is ungrammatical.

(169) **Existential constructions: Initial examples**a. **Present existential construction**

Pi Xaray=**en** bašānd ūniversitet-en (**yast**).  
 up.in Khorog=3PL good university-PL EEP  
 ‘There (*are*) good universities in Khorog.’

b. **Past existential construction**

Pi wam ūniversitet=en bašānd mollim-en **vad** (\***yast**).  
 up.in DEM.OBL.F university=3PL good teacher-PL be.PST.F (\*EEP)  
 ‘There were good teachers at that university.’

The examples in (170) show that constructions with the particle *yast* are subject to definiteness effects. Here, the use of the demonstrative *wāð* ‘those’ would be ungrammatical with the existential particle *yast*, which is permitted only in existential constructions.

(170) **Definiteness effect with the existential particle *yast***

(\*Wāð) bašānd ūniversitet-en=en pi Xaray yast.  
 those.DIR good university-PL=3PL up.in Khorugh EEP  
 ‘There are (\*those) good universities in Khorugh.’

Existential constructions in Shughni are also used to express possession. Similarities between existential, possessive, and locative constructions are cross-linguistically common (Freeze 1992). Indeed, existential constructions are a common means of expressing possession throughout Iranian languages. Possessive constructions built on verbs meaning ‘have’ (e.g. Persian *dāštan*) were apparently virtually non-existent in older stages of Iranian and have only come about in a subset of modern Iranian languages (see Edelman 1974 for an overview).

In Shughni canonical possessive constructions, the possessor is marked by the locative suffix *-(y)and*, or *-nd* with certain pronominals (see Section 4.3.1). Agreement in person and number (and gender in the past tense) is invariably with the possessum (and never the possessor). Examples of possession expressed via existential constructions are given in (171).

(171) **Basic possessive constructions**a. **Present-tense possession**

Mu-nd=**en** lap-aθ kitob-en.  
 me-POSS=3PL many-AUG book-PL  
 ‘I have a lot of books.’

b. **Past-tense possessive construction**

Mu-nd kitob **na-vad**.  
 me-POSS book NEG-be.PST.PL  
 ‘I had no book.’

Like basic existential constructions, present-tense indicative existential clauses indicating possession can also be emphasized with the particle *yast*. This is shown in (172), which corresponds to the neutral example in (171a). The particle *yast* may be used, for instance, in response to an assertion that the speaker does not have any books.

(172) **Emphatic present-tense possession**

Mu-nd=**en** lap kitob-en **yast**.  
 me-POSS=3PL many book-PL EEP  
 ‘I *do* have a lot of books.’

**8.3.5 Summary: Copular constructions**

As seen in the preceding subsections, copular constructions in Shughni are realized as two etymologically and formally distinct sets of morphemes: second-position clitics and a full verb *vidow* ‘be’. However, the division of labor between each is clearly defined, and the use of one or the other copular form is never optional. Second-position copular clitics are used in present-tense indicative environments and negated with the clause-final particle *nist*. On the other hand, the verb *vidow* is used in present-tense irrealis environments, where its present stem is used, and in past and perfect environments, where its past and perfect stems are used.

Regarding its inflection, present-tense copular clitics show fewer verb-like qualities than *vidow*. The former have only a single paradigm and are negated via the clause-final particle *nist*, while the latter has a full inflectional paradigm is negated via the canonical verbal negation prefixes, namely the indicative *na-* and prohibitive *mā-*. The form and distribution of the copula more generally is summarized in Table 8.8.



**Table 8.8:** Summary of copula form and distribution.

	<u>ENVIRONMENT</u>	<u>POSITION</u>	<u>NEGATION</u>
<b>2nd-position clitics</b>	present indicative	right edge of first syntactic phrase	<i>nist</i> (clause-finally in neutral word order)
<b><i>vidow</i> (present stem)</b>	present irrealis	clause-finally (neutral word order)	PFX <i>mā-</i>
<b><i>vidow</i> (past/perfect stem)</b>	past tense indicative/irrealis	clause-finally (neutral word order)	Indicative: PFX <i>na-</i> Subjunctive: PFX <i>mā-</i>

The Shughni copula is also used in existential constructions, which are often used to encode possession. Here, the copula behaves as it does in basic copular constructions, with the exception that in present-tense indicative environments – i.e. with copular clitics – an affirmative existential construction may be emphasized with the particle *yast*.

## 8.4 Verbal and predicate modification

This section examines the modification of verbs and larger predicates with adverbials. The discussion here is organized as follows. Section 8.4.1 provides an introduction to types of adverbs and looks briefly at the order of adverbs in the clause with respect to the object, verb, and one another. Section 8.4.2 examines adverbial derivation and the various morphemes which may attach to adverbs. The final three subsections look in detail at specific types of adverbs, including adverbs of degree (Section 8.4.3), epistemic adverbs (Section 8.4.4), and deictic adverbs (Section 8.4.4).

### 8.4.1 Types and order of adverbs

This section examines six types of adverbs in Shughni. For the following four types of adverbs, a rough classification can be given by the question they answer:

<u>ADV TYPE</u>	<u>QUESTION ANSWERED</u>	<u>EXAMPLE</u>
<b>Frequency adverbs</b> (a.k.a. adverbs of frequency)	‘how often (does it occur)?’	<i>yabori</i> ‘sometimes’
<b>Manner adverbs</b> (a.k.a. adverbs of manner)	‘how; in what way (does it occur)?’	<i>ostā</i> ‘slowly’
<b>Temporal adverbs</b> (a.k.a. adverbs of time)	‘when (does it occur)?’	<i>biyor</i> ‘yesterday’
<b>Locational adverbs</b> (a.k.a. locational deictics)	‘where (does it occur)?’	<i>azūd</i> ‘from here’

Each of these four types of adverbs behave differently with respect to their compatibility with certain types of adverbial derivation. The compatibility of each with different adverbial suffixes and other derivational processes is the topic of the present subsection.

Two further types of adverb are distinguished: *degree adverbs* modify another adverb, such as *dis ostā* ‘very slowly’, and *epistemic adverbs* such as *mumkin* ‘perhaps’ communicate the degree of certainty a speaker has about the information in question. These are examined in Sections 8.4.3 and 8.4.4, respectively.

Another class of adverbs is that of deictic adverbs, which in Shughni includes certain adverbs of location, manner, and time. These are adverbs whose root is a deictic element, that is, a linguistic expression whose meaning is interpreted in relation to the discourse environment (see Diessel 1999 and references therein on deixis and deictic adverbs; see also section 6.2 in this thesis for an overview). For instance, temporal adverbs such as *yesterday*, *today*, and *tomorrow*, are deictic adverbs, as their meaning is interpreted based on the time when the utterance takes place. Similarly, adverbs of location such as *yūdand* ‘here’ are deictic adverbs because their meaning is interpreted based on the physical location of the speech participants. Deictic adverbs are the topic of Section 8.4.5.

It will be shown in Section 8.4.2 that the various types of adverbs presented here behave in specific ways with respect to their morphosyntax. Moreover, not all adverbs are a single word. Prepositional phrases such as *bād-i yi mēst*, built on the preposition *bād* ‘after’, are considered here to function as adverbs of time. Examples of adverbs of various kinds are given in Table 8.9.

In neutral word order, adverbials adhere to a certain position in the clause, both with respect to the verb and object (if present), as well with respect to one another. This order is schematized in (173).

**Table 8.9:** Types of adverbs and examples.

FREQUENCY		MANNER		TIME	
<i>bašānd(-aθ)</i>	well	<i>yabori</i>	sometimes	<i>šič</i>	now
<i>tez(-aθ)</i>	fast	<i>fukwaxt(-aθ)</i>	all the time	<i>ǰumne</i>	tomorrow
<i>ostā(-yaθ)</i>	slowly	<i>redki(-yaθ)</i>	rarely	<i>bādi yi mēst</i>	in a month (from now)

LOCATION		DEGREE		EPISTEMIC	
<i>yūdand</i>	(in) here	<i>dis</i>	very; so	<i>boyad</i>	must; certainly
<i>yed(-ard)</i>	(around) there	<i>sof</i>	too	<i>mumkin</i>	perhaps
<i>padam</i>	(up) there	<i>lap</i>	very	<i>ukmand-aθ</i>	definitely

(173) **Neutral order adverbs**

SUBJ    ADV.FREQ    ADV.TEMP    (OBJ)    ADV.DEG    ADV.MANNER    VERB

An example of a sentence containing such an order is given in (174). Adverbs are bolded.

- (174) Yu **abīčni sāraki**    xu    awqot **dis tez** xīrt.  
 he usually in.morning REFL food    so    fast eat.3SG.PRS  
 ‘He usually eats his food so fast in the morning.’

Nonetheless, with the exception of degree words, adverbs may generally partake in scrambling to carry out notions of emphasis and focus. (For a detailed discussion of topic, focus, and their relation to word order, see Section 12.2.)

**8.4.2 Adverbial derivation and morphology**

This subsection looks at the derivation of adverbs and adverbial morphology. It will be seen that certain morphemes and morphosyntactic processes target only some (or one) of the specific types of adverbs presented above. The first three subsections look at affixes and particles which may attach to adverbs, namely the **augmentative -aθ** (Section 8.4.2.1); the **comparative suffix -di** (Section 8.4.2.2); and the **superlative particle sar** (Section 8.4.2.3), which precedes the adverb it modifies. Section 8.4.2.4 then looks at the derivation of adverbs via **reduplication**, a process which is common for a small subset of manner adverbs.

### 8.4.2.1 Augmentative suffix *-aθ*

Perhaps the most versatile and commonly used morpheme with adverbs is the augmentative suffix *-aθ*, which carries certain nuances in meaning depending on the type of adverb with which it is used. For instance, when used with adverbs of time, *-aθ* is used to indicate preciseness or punctuality. Examples (175) and (176), with the help of the contexts provided for each, illustrate this point. In (175), the adverb *sāraki* is used without *-aθ* in a neutral context in which the speaker is merely listing a sequence of events. In (176), the use of the augmentative emphasizes that the action was completed at this time and no later.

#### (175) Without *-aθ*: Neutral reading

**Context:** As part of a response to the question *čīr=at nur čūd?* ‘what did you do today?’

Wuz=um sāraki čīd zidūd.  
 I=1SG in.morning house sweep.PST  
 ‘In the morning I swept the house.’

#### (176) With *-aθ*: Punctuality

**Context:** Usmon knows Miskin played basketball this afternoon, but he thought Miskin had to sweep the house, a chore which he normally does in the afternoons. Usmon is surprised and asks Miskin how he managed to play basketball in the afternoon, given that he was supposed to sweep the house.

Wuz=um sāraki-**yaθ** čīd zidūd.  
 I=1SG in.morning-AUG house sweep.PST  
 ‘I swept the house (already) in the morning.’

The augmentative is also commonly used with adverbs of manner. Here, its usage is obligatory in some grammatical contexts. In particular, it is required on manner adverbs in imperative sentences, except when the adverb is reduplicated (see Section 8.4.2.4 below on reduplicated adverbs). This requirement is illustrated in examples (177) and (178). The former shows the optional use of *-aθ* in declarative sentences, while the latter shows its obligatory use in imperatives.

#### (177) Declaratives: augmentative *-aθ* optional

- a. Somon=i nur bašānd(**-aθ**) fūtbol bēxt.  
 Somon today good(-AUG) football play.PRS.3SG  
 ‘Somon played football well today.’ cf. (178a)
- b. Māš=ām tez(**-aθ**) andoyd=xu bād=ām zavtrak čūd.  
 we early(-AUG) get.up.PST.PL=and later=1PL breakfast do.PST  
 ‘We got up early and then had our breakfast.’ cf. (178a)

(178) Imperatives: augmentative *-aθ* obligatory

a. Bašānd\*(-aθ) fūtbol boz nur!  
 good\*(-AUG) football play.2SG.IMP today  
 ‘Play football well today!’ cf. (177a)

b. Tez\*(-aθ) andidz-et=xu zavtrak kin-et!  
 early\*(-AUG) get.up-2PL.IMP=and breakfast do.IMP-2PL  
 ‘Get up (PL) early and have breakfast!’ cf. (177b)

The augmentative is also found on adverbs of frequency, such as *xoli-yaθ* ‘always; constantly’ and *abīčni-yaθ* ‘usually’, and may even be used on the degree adverb *dis* ‘so; very’, as in *dis-aθ* ‘just like that (and no more)’, as shown in example (179).

(179) Yu=ta **dis-aθ** na-qīw-d, wi-rd fukwaxt-aθ arčīzca az māš darkor.  
 he=FAC SO-AUG NEG-call.PRS-3SG. him-DAT always-AUG something from us necessary  
 ‘He never *just* calls (to check in); he always needs something from us.’

8.4.2.2 Comparative suffix *-di* with adverbs

The comparative suffix *-di* is used with adverbs of manner only. In declarative statements, it can only be used with bare adverbs – i.e. not on adverbs with the augmentative *-aθ* and not on reduplicated adverbs. This is shown in (180) and (181).

(180) Mu-nd lap kor, xūmne=ta **tez-di(\*-yaθ)** andidz-um.  
 I=1SG much work, tomorrow=FAC early-CMPR(\*-AUG) get.up.PRS-1SG  
 ‘I have a lot of work; tomorrow I’ll get up later.’

(181) Mu bob az mu tāt **xaǰ-di** (\*xaǰ-xaǰ-di) gāp ōed.  
 my grandfather from my father loud-COMP (\*loud-loud-CMPR) word hit.PRS.3SG  
 ‘My grandfather speaks louder than my father.’

However, the use of the comparative suffix and the augmentative together is permissible in imperative sentences, as shown in (182) and (183).

(182) **Tez-di-yaθ** andidz!  
 early-CMPR-AUG get.up.2SG.IMP  
 ‘Get up earlier!’

- (183) **Xaǰ-di-yaθ** gāp ða!  
 loud-CMPR-AUG speak hit.2SG.IMP  
 ‘Speak louder.’

#### 8.4.2.3 Superlative particle *sar* with adverbs

The superlative particle *sar*, which appears immediately before the adverb it modifies, is also compatible with manner adverbs. The superlative particle does not occur together with either suffixed adverbs or reduplicated adverbs. Examples are given in (184).

(184) **Adverbs with the superlative particle *sar***

- a. Čāy az fuk-aθ **sar bašānd(\*-aθ)** anglīsi-ti gāp ðed?  
 who.DIR from all-AUG SUP good(\*-AUG) English-LOC word hit.PRS.3SG  
 ‘Who speaks the best English (out of everyone)?’
- b. Yu=ta az fuk-aθ **sar lap(\*-aθ)** awqot xīrt.  
 he=FAC from all-AUG SUP much(\*-AUG) food eat.PRS.3SG  
 ‘He will eat the most out of anyone.’

#### 8.4.2.4 Reduplicated adverbs

A subset of commonly used manner adverbs may be reduplicated, including *tez* ‘fast’, *qaror* ‘quietly’, *xaǰ* ‘tightly; loudly’, and *ostā* ‘quietly’. Importantly, the reduplicated form of an adjective in some instances has a distinct meaning than its suffixed and/or bare forms. Thus, for instance, the adverb *qaror-aθ* (suffixed form) has the meaning ‘quietly’, while the adverb *qaror qaror* has the meaning ‘secretively; furtively’. Reduplicated adverbs in Shughni, including nuances in their meanings, are summarized in Table 8.10.

**Table 8.10:** Suffixed and reduplicated adverbs.

<u>BARE FORM</u>	<u>GLOSS</u>	<u>SUFFIXED FORM</u>	<u>GLOSS</u>	<u>REDUPLICATED FORM</u>	<u>GLOSS</u>
<i>qaror</i>	quietly	<i>qaror-aθ</i>	quietly	<i>qaror-qaror</i>	secretively
<i>ostā</i>	–	<i>osta-yaθ</i>	quietly	<i>ostā-ostā</i>	slowly
<i>tez</i>	early	<i>tez-aθ</i>	fast; early	<i>tez-tez</i>	fast; often
<i>xaǰ</i>	loudly; rudely	<i>xaǰ-aθ</i>	tightly; loudly	<i>xaǰ-xaǰ</i>	loudly

Examples illustrating the difference in possible meanings between a suffixed adverb – *xaǰ-aθ*, which can mean either ‘tightly; loudly’ – and its reduplicated counterpart – *xaǰ-xaǰ* – which can only mean ‘loudly’ – are given in (185) and (186).

- (185) Tu boyad wam butilka **xaǰ-aθ** (# **xaǰ-xaǰ**) anja!  
 you must DEM.OBL.F bottle tight-AUG (# loud-loud) grab.2SG.IMP  
 ‘You need to grab that bottle tightly!’
- (186) Wuz tu gāp-en na-ǰin-um! Ku **xaǰ-xaǰ** (**xaǰ-aθ**) gāp ða!  
 I your word-PL NEG-hear.PRS-1SG please loud-loud (loud-AUG) word hit.2SG.IMP  
 ‘I can’t hear you! Please speak loudly!’

Reduplicated adverbs are not compatible with the augmentative or comparative suffixes, or with the superlative particle *sar*.

#### 8.4.2.5 Adverbial morphosyntax: Summary

In this subsection we have seen three morphemes which are used with Shughni adverbs. The augmentative *-aθ* combines with adverbs of frequency to emphasize the temporal precision or punctuality with which an eventuality occurred. This suffix is also compatible with adverbs of manner, where it is optional in declaratives but obligatory in imperatives. The comparative suffix *-di* and superlative particle *sar*, which appears before adverbs, are used only with adverbs of manner. The suffix *-di* may be combined with the augmentative *-aθ* in imperatives, but the superlative *sar* is not compatible with the augmentative.

The process of reduplication in adverbs was also examined. Only a select few adverbs of manner, such as *ostā* ‘slowly’ and *tez* ‘fast’ may undergo reduplication, and in some cases the reduplicated form has a different meaning than its non-reduplicated counterpart. Reduplicated adverbs are not compatible with any of the other adverbial morphemes discussed here.

### 8.4.3 Adverbs of degree

Adverbs of degree specify the degree to which the action or state denoted by a verb is carried out. Certain adverbs of degree in Shughni can directly modify verbs, while others are used only before adverbs of manner. Shughni

degree adverbs are given in Table 8.11.

**Table 8.11:** Degree adverbs in Shughni.

<u>USED DIRECTLY BEFORE VERBS</u>	
<i>kām</i>	‘a little’
<i>dūs</i>	‘a little’
<i>lap</i>	‘a lot’
<i>qarīb</i>	‘almost’
<i>bailā</i>	‘barely; with a lot of difficulty’
<i>ačaθ</i>	‘(not) at all’
<u>USED DIRECTLY BEFORE MANNER ADVERBS</u>	
<i>dīs</i>	‘very’
<i>sof</i>	‘very; too’
<i>dūnd</i>	‘(not) very’
<i>lap</i>	‘very’
<i>dūs</i>	‘slightly’

Examples with the degree adverbs *qarīb* ‘almost’ and *kām* ‘a little’ are given in (187) and (188).

(187) *qarīb* ‘almost’

Yā      čaynak *qarīb* az      wam      sitol-ti      wêxt!  
 DEM.DIR.F teapot almost from DEM.OBL.F table-LOC fall.PST  
 ‘The teapot almost fell off the table!’

(188) *kām* ‘a little’

Tar māš ūniversitet rūsi      dars-en *kām* naŷdzimb-en.  
 in our university Russian course-PL little pass.CAUS-3PL  
 ‘In our university they offer few Russian courses.’

Note that the degree adverbs *lap* ‘a lot; very’ and *dūs* ‘a little; slightly’ can modify either verbs or manner adverbs.

This is illustrated in examples (189) and (190).



(189) a. *dūs* ‘a little’: Verbal modification

Yu=ta **dūs** naqli kišt.  
 he=FAC little speak do.3SG.PRS  
 ‘He (only) speaks a little.’

b. *dūs* ‘a little’: Adverbial modification

Tu=t **dūs** der-rang yat.  
 you=2SG a.little late-like arrive  
 ‘You arrived a bit late.’

(190) a. *lap* ‘a lot’: Verbal modification

Tar fūtbol tu boyad **lap** žoz-i.  
 in football you must much run.PRS-2SG  
 ‘In football you have to run a lot.’

b. *lap* ‘very’: Adverbial modification

Vorj qati=ta **lap** tez pi wi kū sifān-i.  
 horse with=FAC very quickly up.to DEM.OBL.M mountain go.up.PRS-2SG  
 ‘With a horse you’ll get up that mountain very quickly.’

Finally, the adverbs *dis* ‘very’; *sof* ‘a lot; too much’; and *dūnd* ‘(not) very’ are used only to modify adverbs of manner, which they obligatorily precede directly. The adverb *dūnd* ‘(not) very’ is used only in negated contexts and can be considered a negative polarity item. *Dis* and *sof*, but not *dūnd*, are compatible with the augmentative *-aθ*, but none are compatible with reduplicated adverbs. Examples of each are given in (191)–(193).

(191) *dis* ‘so; very’

Tu **dis** sitirob-aθ xu čimadān-en jām kin-i, mu-rd nist xuš disga.  
 you so hurriedly-AUG REFL suitcase-PL together do.PRS-2SG me-DAT COP.NEG pleasant like.this  
 ‘You pack your bags so hurriedly; I don’t like it like this.’

(192) *sof* ‘very; too’

Di morožini **sof** tez mā-xa, tu noy=ta pi dārō đed.  
 DEM.OBL.M ice.cream too fast PROH-eat.2SG.IMP 2SG.OBL throat=FAC up.to pain hit.PRS.3SG  
 ‘Don’t eat that ice cream too fast; your throat will hurt!’

(193) *dünd* ‘(not) very’

Wuz=ta **dünd** lap šīrčoy na-birêz-um.  
 I=FAC this much milk.tea. NEG-drink.PRS-1SG  
 ‘I won’t drink very much milk tea.’

## 8.4.4 Epistemic adverbs

Epistemic adverbs communicate the degree of certainty a speaker has about the information in question. Three common Shughni epistemic adverbs are given in Table 8.12.

**Table 8.12:** Common epistemic adverbs in Shughni.

<i>mumkin</i>	possibly
<i>boyad</i>	probably (// must)
<i>ukmand-aθ</i>	definitely

These epistemic adverbs obligatorily modify an entire clause. Examples are given in (194)–(196).

(194) *mumkin* ‘possibly; maybe’

Yā=ta **mumkin** der kiġt.  
 she=FAC possibly late do.PRS.3SG  
 ‘She might be (running) late.’

(195) *boyad* ‘probably; must’

Boyad=en **wāθ** tar dars sat.  
 must=3PL they to class go.PST.PL  
 ‘They must have gone to class’ or ‘They probably went to class.’

(196) *ukmand-aθ* ‘definitely; certainly’

Yā biyor nala wuz=ta **ukmand-aθ** yad-um.  
 she yesterday say I=FAC certainly-AUG come.PRS-3SG  
 ‘She said yesterday she will definitely come.’

### 8.4.5 Deictic adverbs

Deictic adverbs, like demonstrative pronouns, are interpreted based on the discourse context. In Shughni, deictic adverbs include three of types of adverbs, including locative and manner adverbs, where the deictic center is understood to be the speaker, and temporal adverbs, where the deictic center is understood to be the utterance time. This section examines each type of deictic adverb in turn. Section 8.4.5.1 first looks at **deictic locative adverbs**, followed by **deictic manner adverbs** in Section 8.4.5.2 and **deictic temporal adverbs** in Section 8.4.5.3. A summary is given in Section 8.4.5.4.

#### 8.4.5.1 Deictic locative adverbs

Deictic locative adverbs specify the location where the eventuality denoted by a verb takes place. In Shughni, deictic locative adverbs exhibit both important similarities and important differences with demonstrative pronouns. Like demonstrative pronouns, deictic locative adverbs follow a system of triple deixis with proximal, medial, and distal degrees. The deictic center for demonstrative adverbs, as for demonstrative pronouns, is the speaker, and thus the proximal degree denotes a location near the speaker. The medial degree in demonstrative adverbs, as in demonstrative pronouns, is *person-oriented*, rather than *distance-oriented* (cf. Diessel 1999: 39; Fillmore 1982: 49-50). That is, it is used to refer to a location which is near the addressee (the so-called *du*-deixis; e.g. Edelman & Dodykhudoeva 2009b: 794-795), rather than one which is at a middle distance from the speaker. The distal degree is used to indicate a location away from both the speaker and the addressee.

Deictic locative adverbs are also built on the same *deictic roots* as demonstrative pronouns, with one striking difference: the roots which denote proximal and distal degrees in demonstrative pronouns are reversed in deictic locative adverbs (cf. Wendtland 2009: 183, fn. 99).<sup>2</sup> That is, the root of the proximal demonstrative pronoun *yam* is the same as that used for the distal locative adverb (cf. *yam-and* '(in) there.DIST'), while the root of the distal demonstrative pronoun *yu* is the same as that used for the proximal locative adverb (cf. *yû-va* '(along) here.PROX').

---

<sup>2</sup>Although the deictic *roots* used in deictic locative adverbs are assumed to be identical to those in demonstrative pronouns, the deictic *stems* used in each differ slightly in form in some cases. This issue is addressed below; see especially fn. 3. Nonetheless, the proximal demonstrative pronoun *yam* is found in this precise form in all distal deictic locative adverbs which take suffixes (e.g. *yam-and* '(in) there.DIST'). In those which take prefixes, the initial glide is not present. Nonetheless, it is apparently uncontroversial (and unexplained) that the roots corresponding to proximal and distal degrees is reversed for demonstrative pronouns and deictic locative adverbs.

The root for the medial degree is the same in both – cf. demonstrative pronoun *yid* and distal locative adverb *yedand* ‘(in) there.MED’. This reversal of deictic degrees between demonstrative pronouns and deictic locative adverbs is schematized in Figure 8.2.

**Figure 8.2:** Reversal of degrees deictic in pronouns and adverbs.

<u>DEGREE</u>	<u>DEMONSTRATIVE PRONOUNS</u>	<u>DEICTIC ADVERBS</u>
PROX.	<i>yam</i>	<i>yũ(d)-</i>
MED.	<i>yid</i>	<i>ye(d)-</i>
DIST.	<i>yu</i>	<i>yam-</i>

Another key difference between demonstrative pronouns and deictic locative adverbs is that whereas demonstrative pronouns are syntactically independent and can be the sole element of a noun phrase, deictic locative adverbs consist of a bound deictic stem and a separate locative element. In other words, demonstrative pronouns are morphologically simplex, while deictic adverbs are morphologically complex.

The locative element which attaches to the bound deictic stem of a locative adverb may be either a preposition or a suffix. The form of the deictic stem differs slightly in each case. When combining with a preposition – i.e. where the deictic stem is to the right of the locative element – it lacks the initial glide *y* which is present when it combines with a suffixal element (cf. preposition *tar* ‘to’ + stem *-(y)am* > *taram* ‘to there’, but stem *yam* + suffix *-and* > *yamand* ‘(in) there’). Moreover, when the combining of a deictic stem and preposition or suffix results in two adjacent vowels, these are connected by the alveolar stop *d*, as in stem *ye-* + suffix *-and* > *yedand*.<sup>3</sup> The

<sup>3</sup>Determining the precise shape of the deictic stems used in locative adverbs is not trivial. In particular, it is not immediately clear whether the *-d-* in medial forms such as *yedand* ‘in there.MED’ is part of the stem or merely a connective element. On the one hand, there is a final *-d* in the corresponding medial demonstrative pronoun *yid*. On the other, there is no connecting *-d-* when the following locative element begins with a consonant, as in *ye-va* ‘along there.MED.’ Moreover, there is also a *-d-* in the proximal adverb *yũdand* ‘in here’, despite the fact that the corresponding demonstrative pronoun *yu* does not contain a final *-d*.

Hence, there is reason to believe the *-d-* in medial locative adverbs is part of the deictic stem, and there is also reason to believe it is simply a connective element. I make no final judgment on this matter here. I simply note that the structure of deictic stems in locative adverbs is a

structure of deictic locative adverbs is schematized in Figure 8.3.

**Figure 8.3:** Structure of deictic locative adverbs.

DEGREE	WITH PREPOSITION	WITH SUFFIX
PROXIMAL	PREP-(d) <i>ûd</i>	<i>yû(d)</i> -SFX
MEDIAL	PREP-(d) <i>ed</i>	<i>ye(d)</i> -SFX
DISTAL	PREP-(d) <i>am</i>	<i>yam</i> -SFX

Four prepositions are used in the formation of deictic locative adverbs: *az* ‘from’, *ar* ‘(down) to/at’, *pi* ‘(up) to/at’, and *tar* ‘(horizontally) to/at’. The preposition *pi* ‘up (to)’ has the form *pa-* when combined with deictic stems. Locative adjectives formed from these prepositions are shown in Table 8.13.

**Table 8.13:** Deictic adverbs with prepositional locative elements.

DEGREE	DEICTIC STEM	COMBINING PREPOSITION			
		<i>ar</i> ‘(down) to’	<i>tar</i> ‘(horiz.) to’	<i>pi</i> ‘(up) to’	<i>ar</i> ‘from’
PROX.	-(d) <i>ûd</i>	<i>arûd</i>	<i>tarûd</i>	<i>padûd</i>	<i>azûd</i>
MED.	-(d) <i>ed</i>	<i>ared</i>	<i>tared</i>	<i>paded</i>	<i>azed</i>
DIST.	-(d) <i>am</i>	<i>aram</i>	<i>taram</i>	<i>padam</i>	<i>azam</i>

Three locative suffixes participate in the formation of deictic adverbs: *-and* ‘inside; at’, *-ard*(//*-ra*) ‘toward; around’, and *-va* ‘along’. The suffix meaning ‘toward; around’ indicates either movement toward a location or approximate location. This suffix has two possible realizations. The form *-ra* is sometimes used when following a vowel (e.g. *yû-ra* ‘around here’), but its use is illicit with the distal stem *yam* (\**yam-ra*). In all cases where the use of *-ra* is possible, an alternative form with *-ard*, preceded by *-d-* is possible, and hence *yûd-ard* ‘around here’ is also used. Locative deictic adverbs built on suffixes are shown in Table 8.14.<sup>4</sup>

matter for further research. When referring to the proximal and medial stems of locative adverbs, I indicate this *-d-* in parentheses, as in *yû(d)-* and *ye(d)-*

<sup>4</sup>The preposition *az* and the locative suffix *-and* are permitted to occur together on the same deictic stem, and hence we get the forms

**Table 8.14:** Deictic adverbs with suffixal locative elements.

DEGREE	DEICTIC STEM	COMBINING LOCATIVE SUFFIX		
		<i>-and</i> ‘in(side)’	<i>-ard / -ra</i> ‘toward; around’	<i>-va</i> ‘along’
PROX.	<b>yû(d)-</b>	<i>yûdand</i>	<i>yûdard // yûra</i>	<i>yûva</i>
MED.	<b>ye(d)-</b>	<i>yedand</i>	<i>yedard // yera</i>	<i>yeva</i>
DIST.	<b>yam-</b>	<i>yamand</i>	<i>yamard</i>	<i>yamva</i>

The following examples illustrate the use of the locative deictic adverbs *tarûd* ‘here’, *tared* ‘there.MED’, and *taram* ‘there.DIST’. A context is provided for each pair of examples which help show that the proximal degree is used for a location near the speaker, the medial degree for a location near the addressee, and the distal degree for a location away from both speaker and addressee. The first two examples in (197) show the difference between proximal *tarûd* and medial *tared*.

(197) **Deictic locative adverbs with *tar* ‘at/to’: Medial usage**

**Context:** Hakim is at school when his dad calls from their home to ask about his plans. Hakim tells his dad he plans on studying the rest of the afternoon, and his dad inquires whether he will study at school (197a). (The # indicates that the use of distal *tar-am* would be infelicitous if the location Hakim’s father inquires about is Hakim’s current location.) Hakim’s response is the sentence in (197b).

a. ***Tared* – MEDIAL: Addressee’s location**

Tu=ta **tar-ed** (# tar-am) xu dars xoy-i yo?  
 you=FAC in-MED (# in-DIST) REFL lesson study.PRS-2SG PQ  
 ‘Are you going to study there (at your current location)?’

b. ***Tarûd* – PROXIMAL: Speaker’s location**

Ûn, wuz=ta **tar-ûd** xu dars xoy-um. Taredira=ta pi čîd yad-um.  
 yes I=FAC in-PROX REFL lesson study.PRS-1SG later=FAC up.to house come.PRS-1SG  
 ‘Yes, I’m going to study here (at my current location). I’ll come home later.’

The examples in (198) show the difference in usage between medial *tared* and distal *taram*.

(198) **Deictic locative adverbs with *tar* ‘at/to’: Distal usage**

**Context:** Hakim and his father are at home together when his father asks Hakim about his plans for the day. Hakim tells his dad he plans on going to the library to get a book and then studying the rest of the afternoon. His dad inquires whether he will study at the library or another place (197a). (The # indicates that the use

---

*az-ûd-and* ‘from (in) here’, *az-ed-and* ‘from (in) there.MED’, and *az-am-and* ‘from (in) there.DIST’.

of medial *tar-ed* would be infelicitous if the location Hakim’s father inquires about is the library, which is away from both Hakim’s and his father’s location.) Hakim’s response is the sentence in (197b).

- a. Tu=ta **tar-am** (#*tar-ed*) xu dars xoy-i yo?  
 you=FAC in-DIST (#in.MED) REFL lesson study.PRS-2SG PQ  
 ‘Are you going to study there (at the library)?’
- b. Ūn, wuz=ta **tar-am** xu dars xoy-um. Taredira=ta pi čīd yad-um.  
 yes I=FAC in-DIST REFL lesson study.PRS-1SG later=FAC up.to house come.PRS-1SG  
 ‘Yes, I’m going to study there (at the library). I’ll come home later.’

#### 8.4.5.2 Deictic manner adverbs

In addition to the deictic locative adverbs discussed above, Shughni also has a set of deictic manner adverbs roughly meaning ‘in this/that way’. Unlike deictic locative adverbs, whose deictic stems correspond to the direct forms of demonstrative pronouns, deictic manner adverbs have stems which are identical to the oblique forms of demonstrative pronouns. Also unlike locative adverbs, the deictic degrees in these manner adverbs are not reversed with respect to demonstrative pronouns. The set of deictic manner adverbs is shown in Table 8.15:

**Table 8.15:** Deictic adverbs of manner.

DEGREE	OBLIQUE.MASC DEM. PRON.	DEICTIC MANNER ADV.	GLOSS
PROX.	<i>mi</i>	<i>mi-ti</i>	‘like this (as the speaker does it)’
MED.	<i>di</i>	<i>di-ti</i>	‘like that (as the addressee does it)’
DIST.	<i>wi</i>	<i>wi-ti</i>	‘like that (as a third party does it)’

Deictic manner adverbs also have person-oriented medial degree; that is, the medial form is used to refer to the manner in which the addressee carries out an action. The examples in (199) illustrate the use of these adverbs. The sentences here are uttered one after another by the same speaker, and thus a single context is provided for all three.

#### (199) Deictic manner adverbs

**Context:** Samir’s father is teaching him how to iron his shirts. They’re both standing by the iron, and Samir has just incorrectly tried to iron his shirt without water. When his father sees this, he utters the sentence in (199a). His father then shows him how it should be done and utters the sentence in (199b). After doing so, he notes that Samir’s mother employs a different, but equally viable, ironing technique, which Samir has seen before. This is conveyed in (199c).

a. **Medial deictic adverb of manner: *di-ti***

**Di-ti**                      rost      na-sūd  
 DEM.OBL.MED-LOC correct NEG-become.PRS.3SG  
 ‘It won’t be right like that (i.e. the way you are doing it).’

b. **Proximal deictic adverb of manner: *mi-ti***

Tu boyad **mi-ti**                      řac      qati di                      kurtā wutūk đāđ-i.  
 you must DEM.OBL.PROX-LOC water with DEM.OBL.M shirt iron give.PRS-2SG  
 ‘You have to iron your shirt like this (i.e. the way I am doing it) with water.’

c. **Distal deictic adverb of manner: *wi-ti***

Tu nān=i              wi-ti                      ca đod,      **wi-ti**                      mis boft.  
 your mother=3SG 3SG.OBL.DIST-LOC C give.PST DEM.OBL.DIST-LOC also function.PRS.3SG  
 ‘The way your mom did it – that way also works.’

## 8.4.5.3 Deictic temporal adverbs

Deictic temporal adverbs have as their deictic center the time of the utterance. Thus, for instance, a word such as *biyor* ‘yesterday’ denotes the day before the utterance took place. Two basic types of deictic temporal adverbs are distinguished here. The first type of deictic temporal adverb is that which is not built on a demonstrative pronoun. These are single lexical items which contain are inherently deictic. Examples of this type of adverb are shown in Table 8.16.

**Table 8.16:** Deictic temporal adverbs in Shughni: Single lexemes.

<i>nur</i>	today	<i>biyor</i>	yesterday	<i>řumne</i>	tomorrow
<i>ařtb</i>	the day before yesterday	<i>parařtb</i>	three days ago	<i>řič</i>	now
<i>afař</i>	the day after tomorrow	<i>parwoz</i>	last year	<i>widir</i>	three days from now

The second type of deictic temporal adverb is that which is built on an oblique demonstrative pronoun and a separate word denoting the period of time in question (e.g. *mam sol* ‘this year’). When used in temporal adverbs, demonstrative pronouns still ostensibly exhibit a system of triple deixis. The proximal demonstrative pronoun is used to indicate a time which overlaps with the utterance time (e.g. *mam sol* ‘this year, *mam ařtā* ‘this week’). The medial demonstrative pronoun is used to indicate a time period which follows the utterance time (e.g. *dam sol* ‘next year’, *dam ařtā* ‘next week’). The distal demonstrative pronoun is used to indicate a time period which precedes the utterance time (e.g. *wam sol* ‘last year’, *wam ařtā* ‘last week’). Examples of of deictic temporal adverbs in



context are shown in (200).

(200) **Deictic temporal adverbs**

**Context:** Shahlo's mother is interested in the weather in Dushanbe, where Shahlo lives. Shahlo gives her the following information. The sentences in (200 a–c) are uttered in sequence.

a. **Distal temporal adverb: *wam aftā* 'last week'**

**Wam**            **aftā** xub gārm vud=xu      yabor-aθ      di čūd šitto.  
DEM.OBL.DIST.F week quite warm be.PST.M=then suddenly-AUG MIR do.PST cold  
'Last week it was quite warm, and then all of a sudden it got cold.'

b. **Proximal temporal adverb: *mam aftā* 'this week'**

**Mam**            **aftā** xoli borūn ōd.  
DEM.OBL.PROX.F week only rain hit.PST  
'This week it just rained.'

c. **Medial temporal adverb: *dam aftā* 'next week'**

**Dam**            **aftā**=ta šitto-di sūd.  
DEM.OBL.MED.F week=FAC cold-COMP become.PRS.3SG  
'Next week it will be even colder.'

Note, however, that although the distal demonstrative pronoun used with a time period refers to the period immediately preceding that which contains the utterance time, this pronoun may also be used to indicate a point in time not relative to the utterance time, but rather relative to some other point in time which has been specified previously in the discourse. In the terminology of Jakobson (1957), in the former case – i.e. when the point of reference is the utterance time and the pronoun shifts the absolute time reference – the distal pronoun is used as a *shifter*, while in the latter use, where the change in temporal reference is relative, it is used as a *connector*.

When used as a connector, the distal pronoun is generally accompanied by additional lexical items, such as *iga* 'another; next', *piro* 'before', and *ik* (anaphoric marker). For instance, if a speaker is telling a friend her plans and sets the time being talked about five weeks from the utterance time, *wam iga aftā* 'the next week' would refer the week four weeks from the utterance time, *wam piro aftā* would refer to the week six weeks from the utterance time, and *ik wam aftā* would refer to the same week indicated in the first sentence (i.e. five weeks from the utterance time).

#### 8.4.5.4 Summary: Deictic Adverbs

This subsection has examined deictic adverbs in Shughni, including deictic adverbs of location, manner, and time. Deictic locative adverbs consist of a bound root, which corresponds to the direct form of a demonstrative pronoun, and a bound form of a preposition or locative affix. The deictic system of all deictic adverbs, like that of demonstrative pronouns, is tripartite with a person-oriented medial form. Deictic locative adverbs may therefore indicate a location near the speaker (proximal), near the addressee (medial), or away from both speaker and addressee (distal). However, Shughni exhibits a curious idiosyncrasy with respect to its deictic stems: the proximal and distal stems used in deictic location adverbs are reversed with respect to those used in demonstratives.

Shughni has three deictic adverbs of manner which are built on oblique (masculine) demonstrative pronouns. These also employ a person-oriented system of triple deixis, where the proximal form *mi-ti* ‘like this’ refers to the way a speaker carries out an action, the medial form *di-ti* refers to the way the addressee carries out an action, and the distal form *wi-ti* refers to the way something is done by a third person, who is neither speaker nor addressee.

Deictic temporal adverbs may be either a single lexeme (e.g. *nur* ‘today’) or else consist of a noun indicating a time period together with an oblique demonstrative pronoun which agrees with it in gender and number. In general, the distal demonstrative pronoun refers to the period immediately preceding the utterance time (e.g. *wam mēst* ‘last month’), the proximal pronoun to the period including the utterance time (e.g. *mam mēst* ‘this month’), and the distal pronoun to the immediately following the utterance time (e.g. *dam mēst* ‘next month’). Additional lexemes such as *iga* ‘another; next’, *piro* ‘before’ and the anaphor marker *ik* are used to convey time periods relative to a time other than the utterance time (e.g. *wam iga mēst* ‘the next month’).

This subsection concludes the chapter on foundations of the Shughni verbal system. This chapter looked at the basics of verb stems, including verb-stem paradigms and verbal inflection, as well as the copula and verbal modification. The next chapter delves deeper into the nuts and bolts of verb-stem paradigms, verb forms, and the notion of regular and irregular verbs.

## Chapter 9

# Regular and Irregular Verbs

This chapter presents a detailed investigation of Shughni verb forms. The discussion begins with a more precise definition of regular and irregular verbs than that given in the previous chapter, and then turns to an examination of the myriad ways a verb may be irregular. Historical explanations for these irregularities, as well as recent and ongoing changes in the language's verbal system, are also discussed.

The verbal system of modern Shughni and other Pamir languages has been studied almost exclusively from a historical linguistic point of view (e.g., Dodykhudoeva 1988; Edelman 1975; Morgenstierne 1974; Shaw 1877; Tomaschek 1880; Sokolova 1967; 1973), and the discussion in this chapter draws from and builds upon this research. The historical linguistic approach provides context for many of the irregularities in the forms of verb stems. However, this chapter not only synthesizes and updates much of the previous research on Shughni verbs, it also points out a number of phenomena which have yet to be described, most notably verb-stem shortening.

The chapter is organized as follows. Section 9.1 begins with a definition of **regular and irregular verb stems** before examining various ways in which Shughni verbs may be irregular. Section 9.2 then provides a description of **the expression of gender in Shughni verbs**. The chapter concludes in Section 9.3 with a look at **processes of change in Shughni verbs** which have either recently occurred in the language or are still ongoing. (Note that in this chapter, knowledge of the basic changes affecting the Shughni sound system from the time of Old Iranian is assumed. This information can be found in Section 7.2 in Chapter 7 on grammatical gender in nominals.)

## 9.1 Regular and irregular verbs

Having seen the fundamentals of the form and uses of verb stems in Section 8.1 of the previous chapter, we now turn to a more detailed treatment of the form of Shughni verb stems. The discussion here begins in Section 9.1.1 with a look at one key way in which verb stems differ from one another, namely monosyllabic and polysyllabic verb stems, where the latter most often contain the reflex of a formerly productive pre-verbal element. The remaining subsections give a detailed look at regular and irregular verbs, including the various ways in which a verb can be irregular. Section 9.1.2 sets the stage for this discussion by providing a formal definition of regular and irregular verbs. Section 9.1.3 looks at irregularities in the consonants of verb stems and the historical motivations behind such irregularities. Section 9.1.4 does the same for irregularities in stem vowels. Finally, Section 9.1.5 looks at a different type of irregularity, namely the shortening of stems, including shortening via the fusing of a stem with its inflectional ending.

### 9.1.1 Monosyllabic and polysyllabic stems

Shughni verb stems may be either monosyllabic or polysyllabic. The majority of polysyllabic stems contain a distinctive first syllable which unites a particular group of polysyllabic verbs. For instance, the initial syllable *an-* in *anjīvd* ‘grab’ is also found in *ancīvd* ‘sew’, *angaxct* ‘become stuck’, and *andīd* ‘get up’, and the initial syllable *si-* in *sifīd* ‘rise’ is also found in the verbs *sitīvdow* ‘fry’; *siportow* ‘offer’, and *siridow* ‘separate’. This is not accidental; rather, these initial syllables of polysyllabic verb stems are in most cases the reflexes of ancient Iranian pre-verbs (e.g., Sokolova 1967) – i.e. prepositional elements used as verbal prefixes – which are no longer productive in modern Shughni and generally not recognized by native speakers as separate morphemes (Shahlo Shomansurova, p.c.).

A number of pre-verbs have multiple reflexes in modern Shughni, and in many cases these are clearly phonologically conditioned. For example, the pre-verb *\*fra-* ‘forth; forward’ has reflexes *fir-* (e.g. *firāp* < *\*fra-ap-*) and *ri-* (e.g. *riwāz-* ‘fly away’ < *\*fra-waz-*). Here, the former reflex occurs before a stem-initial vowel, while the latter occurs before a stem-initial consonant. In many cases, however, there are exceptions to phonologically conditioned developments of this kind. Thus, the same pre-verb *\*fra-* in some cases may have *far-* as its reflex before a stem-initial consonant – i.e. where would expect *ri-* – such as in the verb *farsêp-* < *\*fra-sāp-aya-* ‘move’

(Dodykhudoeva 1988: 125).

Nonetheless, the reflexes of ancient Iranian pre-verbs in modern Shughni verb stems are by and large regular. Several examples are given in Table 9.1. In this table, the reconstructed pre-verb and the Shughni etymology come from Dodykhudoeva (1988: 120-145). A rough gloss of each historical morpheme is provided along with its realization as attested in Avestan; these come from Beekes (1988) and Martínez & de Vaan (2014).

**Table 9.1:** Polysyllabic verb stems and their ancient Iranian pre-verb counterparts.

PREVERB	MODERN REFLEX(ES)	EXAMPLE	GLOSS	RECONSTRUCTED STEM
<b>*ni-</b> 'down' Av. <i>nī-</i>	→ <i>ni-</i>	<i>niyūy-</i> <i>niḍaḥ-</i> <i>niḥpār-</i>	listen stick step on	<i>*ni-gauš-a-</i> <i>*ni-daf-sa-</i> <i>*ni-spar-a-</i>
<b>*pati-</b> 'against; through' Av. <i>paiti-</i>	→ <i>pi-</i>	<i>piḍis-</i> <i>pinidz-</i> <i>pittew-</i>	ignite dress throw	<i>*pati-di-sa-</i> <i>*pati-muč-sa-</i> <i>*pati-tāw-aya-</i>
<b>*us-</b> 'out; up' Av. <i>nī-</i>	→ <i>si-</i>	<i>sifān-</i> <i>sirāw-</i> <i>sitaḥ-</i>	rise separate fry	<i>*us-fan-a-</i> <i>*us-rāw-a-</i> <i>*us-taf-sa-</i>
<b>*fra-</b> 'forth; forward' Av. <i>frā-</i>	→ <i>ri-</i> → <i>fir-</i>	<i>riwāz-</i> <i>firāp-</i>	fly away reach	<i>*fra-waz-a-</i> <i>*fra-ap-a-</i>
<b>*ham-</b> 'together' Av. <i>hām-</i> , <i>ham</i> , ...	→ <i>am-</i> → <i>an-</i>	<i>ambīθ-</i> <i>anjāv-</i>	collapse grab	<i>*ham-pad-s-ya-</i> <i>*ham-kap-a-</i>
<b>*nir-</b> 'out' Av. <i>nīš-</i> , <i>nīž-</i>	→ <i>naḥ-</i> → <i>naḥ-</i>	<i>naḥfīθ-</i> <i>naḥjīs-</i>	fall out pass	<i>*nir-fat-s-ya-</i> <i>*nir-ga-sa-</i>

Two caveats related to ancient pre-verbs and polysyllabic verb stems in modern Shughni are in order. First, not all Shughni verb stems which contain the reflex of a pre-verb are polysyllabic. For instance, the verbs with infinitive stems *bed* 'disappear', *bīft* 'be able to', and *bēxt* 'send', are all the reflexes of a vowel-initial verb stem preceded by *\*upa-* 'towards; up to', where the initial short vowel of the pre-verb has since dropped off. And secondly, not all polysyllabic verb stems in modern Shughni descend from a verb with a prefixal element. In some cases, albeit

relatively few, the first syllable of a polysyllabic stem is the result of epenthesis, which was used to bring ancient stems with initial consonant clusters into conformity with the modern Shughni ban on syllable-initial consonant clusters. This is the case, for example, with the verbs whose present stems are *virayǎ*- ‘break’ (< \**brauš*), *žirayǎ*- ‘bite; sting’ (< \**grauš*) and *ǎici*- ‘freeze’ (< \**strā*).

### 9.1.2 Regular verbs

Having looked at the issue of monosyllabic and polysyllabic verb stems, we now turn to a definition of precisely what constitutes a regular and an irregular verb. In regular verbs, the form of any stem is predictable given any other stem. That is, if we know the present stem, we can determine the form of all other stems; if we know the past stem, we can determine the form of all other stems, and so on.

In a regular verb with present stem X – which may take any phonological shape – the past and infinitive stems are equivalent to X-*t* or X-*d*. As described in Section 8.1, verbs with a final vowel, semivowel, or voiced obstruent take a final -*d*, while all other verbs take a final -*t*. The perfect stem is equivalent to X-*č* or X-*ǰ*, depending in the same way on the final consonant of X. This is schematized in (201).

#### (201) Stems of regular verbs

##### a. Stem-final voiceless C, liquid, or nasal

PRS:	<b>X-</b>
PST:	<b>X-<i>t</i></b>
INF:	<b>X-<i>t</i></b>
PRF:	<b>X-<i>č</i></b>

##### b. Stem-final vowel, semi-vowel, or voiced obstruent

PRS:	<b>X-</b>
PST:	<b>X-<i>d</i></b>
INF:	<b>X-<i>d</i></b>
PRF:	<b>X-<i>ǰ</i></b>

A regular verb is therefore any verb whose stems follow one of the two patterns in (201). An irregular verb is any verb whose stems do not follow one of these two patterns. This is laid out in (202).<sup>1</sup>

<sup>1</sup> The interpretation of regular and irregular verbs in (202) is somewhat more stringent than that of Dodykhudoeva (1988). In particular, she admits verbs such as *cidow* ‘harvest’ and *naǎdzim(p)ow* ‘cause to pass; impart’ into the category of regular verbs, although they do not conform to the definition I have given here. The former verb contains an interdental fricative *ð* before the final affricate of its perfect stem, i.e. *ciðǎ* ‘(has) harvested’, where this consonant is not found in any other of its stems. The second verb has a somewhat more subtle deviation from a regular pattern: its present stem contains a final voiced *b* *naǎdzimb-* not found in its other stems. Instead, its other stems pattern as if

(202) **Definitions of Shughni regular and irregular verbs**

- a. **Regular verb:** Any verb whose four verb stems follow one of the two patterns in (201).
- b. **Irregular verb:** Any verb whose four verb stems do not follow one of the two patterns in (201).

A regular verb which illustrates the pattern in (201a) is *binêstow* ‘to lose’. This verb has present stem *binês-*, where the final voiceless consonant *-s* dictates that the final consonants of all other stems will be voiceless: PST/INF *binêst*, PRF *binêsč*. A regular verb which illustrates the pattern in (201b) is *qîwdow* ‘call’, whose present stem is *qîw-*. The final voiced consonant of this stem *w*, a voiced semi-vowel, dictates that the final sound of the remaining stems will be voiced: PST/INF *qîwd*, PRF *qîwǰ*. The sets of stems of these two regular Shughni verbs, *binêstow* ‘to lose’ and *qîwdow* ‘to call’, are given in (203).

(203) **Regular verb stems: *qîwdow* ‘call’ and *binêstow* ‘lose’**

INF	<i>qîwd(ow)</i>	INF	<i>binêst(ow)</i>
PRS	<i>qîw-</i>	PRS	<i>binês-</i>
PST	<i>qîwd</i>	PST	<i>binêst</i>
PRF	<i>qîwǰ</i>	PRF	<i>binêsč</i>

Regular verbs in Shughni are varied with respect to both their phonological shape and their etymological origin. While most recently borrowed simplex verbs are regular, it is not the case that all regular verbs are borrowed. A number of regular verbs are the result of the leveling of the paradigms of native Shughni verbs, such as with the verb *ciftow* ‘steal’, whose reconstructed present stem is *\*tʃf-ya-* (Dodykhudoeva 1988: 38, 123). This verb displays well-known sound changes such as Ir. *\*tr* > Sh. *c* and the *i*-umlaut of a present-stem vowel when followed by the suffix *-ya-*. Others are onomatopoeic verbs whose stems have been formed on analogy with other regular verbs in the language (e.g. *čiktow* ‘drip’).

Another large cohort of regular verbs are causatives formed from their non-causative counterparts either via stem-vowel alternation (*riwāz-* ‘fly off’ > *riwêz-* ‘make fly off’) or by the addition of a causative suffix *-en/-ûn* (*fām-* ‘know’ > *famûn-* ‘explain’). (See Section 11.2 on causatives, especially Subsection 11.2.1 on morphological causatives.)

---

the consonant preceding the *t/d* were the nasal *m* or voiceless stop *p* (cf. PST/INF *naǰdzim(p)t*). If this verb were regular on the definition given here, we would expect either past/infinite stem *naǰdzim(b)d*, or else a present stem *naǰdzimp-*.

Borrowed regular verbs include a number of borrowings from Tajik, including *moltow* ‘rub’ (cf. Tajik молидан, *molidan*); *nivištow* ‘write’ (cf. Tajik навиштан, *naviştan*), and *fortow* ‘want; be desirable to’ (cf. Tajik форида, *foridan*). Note, however, that in modern times, most new verbal borrowings are entering the language from Russian, and to a much lesser extent, English. In these cases, verbs are borrowed not as simplex verbs, such as those Tajik borrowings listed above, but rather as complex verbs built on light verbs, such as *mešat čīdow* ‘bother; impede’ – cf. Russ. мешать *mešat*, ‘hinder’. (See section 11.3 on complex verbs.) Examples of regular Shughni verbs are given in 9.2.

**Table 9.2:** Regular Shughni verbs.

<u>PRS</u>	<u>PST/INF</u>	<u>PRF</u>	<u>GLOSS</u>
<i>cif-</i>	<i>cift</i>	<i>cifč</i>	steal
<i>fām-</i>	<i>fām</i>	<i>fāmč</i>	know; understand
<i>famūn-</i>	<i>famūnt</i>	<i>famūnč</i>	explain
<i>jaq-</i>	<i>jaqt</i>	<i>jaqč</i>	bark
<i>mol-</i>	<i>molt</i>	<i>molč</i>	rub
<i>niviš-</i>	<i>nivišt</i>	<i>nivišč</i>	write
<i>peḫc-</i>	<i>peḫct</i>	<i>peḫcč</i>	ask
<i>qilāp-</i>	<i>qilāpt</i>	<i>qilāpč</i>	tinker; search aimlessly
<i>qīw-</i>	<i>qīwd</i>	<i>qīwč</i>	call
<i>riwēz-</i>	<i>riwēzd</i>	<i>riwēzč</i>	make fly away
<i>sipor-</i>	<i>siport</i>	<i>siporč</i>	entrust; pay; apply
<i>tār-</i>	<i>tārt</i>	<i>tārč</i>	clean up
<i>warven-</i>	<i>warvent</i>	<i>warvenč</i>	boil (tr.)
<i>žaq-</i>	<i>žaq</i>	<i>žaqč</i>	push; press

The remainder of this section examines three types of verb-stem irregularities. First, Section 9.1.3 looks at irregularities in the consonants of verb stems, where the stem-final consonant(s) of a verb’s present stem differ from those found in its other stems. Section 9.1.4 then turns to irregularities in the vowels of verb stems. Each type of irregularity is presented primarily within a historical linguistic framework, drawing on the works of previous authors who have examined the Shughni verb system in a historical light, especially Dodykhudoeva (1988), Karamshoev (1978), and Sokolova (1967; 1973). Section 9.1.5 departs from the historical framework to look at the phenomenon of stem shortening, where a verb stem is truncated and/or fused with the inflectional ending for some cells in a verb’s inflectional paradigm.



### 9.1.3 Irregular verbs: Consonant alternations

This subsection details verbs which exhibit irregularities in their stem consonants. Section 9.1.3.1 first lays out precisely what constitutes irregularities in the consonants of a verb's stems and puts forth two basic types. Section 9.1.3.2 then looks at the types of consonant irregularities found in modern Shughni verb stems and the historical developments which have led to them.

#### 9.1.3.1 Fundamentals of consonant-related irregularities in verb stems.

Irregularities in the consonants of Shughni verb stems are almost always related to the consonant(s) found at the end of the stem, i.e. those which follow the stem vowel.<sup>2</sup> Furthermore, they almost always concern the consonants of a present stem, which exhibit one pattern, versus those of the past and infinitive stems, which exhibit another.

If we recall our formulation of regular verbs given in (201) and (202), we would expect any consonants in the present stem of a regular verb to also be found in its past and infinitive stems. If a regular verb's present stem has pattern  $C_1V_1C_2$ , we would expect the past stem to be of shape  $C_1V_1C_2-t$  or  $C_1V_1C_2-d$ . Note that in the discussion here, the final stop  $t/d$  of past and infinitive stems is excluded from calculations of the number of consonants present in the past and infinitive stems. Thus, we would say here that the infinitive stem  $\check{z}\acute{e}\check{x}t$  'run' contains one final consonant.

In Shughni verbs with irregular consonant patterns, we typically find one of two situations. In the first situation, a verb's present stem contains more consonants than its past stem. For instance, in the verb meaning 'eat', the present stem  $x\bar{a}r-$  contains a stem-final  $r$  not found in the verb's past or infinitive stems,  $x\bar{u}d$  and  $x\bar{i}d$ , respectively. In the second situation, the number of consonants in each stem is the same, but they differ in one way or another. This is the case, for instance, in verb meaning 'fly away', whose present stem  $riw\bar{a}z-$  contains a final  $z$ , whereas its past and infinitive stems instead contain a  $\check{x}$  preceding the stop  $t$ :  $riwu\check{x}t$  and  $riwi\check{x}t$ , respectively. Each type of irregularity is schematized in (204).

---

<sup>2</sup>To my knowledge, only a single verb exhibits distinct stem-initial consonants across its stems:  $\check{c}\bar{i}dow$  'do', PST  $\check{c}\bar{i}d$ , INF  $\check{c}\bar{i}d$ , but PRS  $\check{k}in-$ .

(204) **Types of consonant irregularities in verb stems**

- a. **Type 1** : Present stem contains more consonants than past/infinite stems.

<u>PRS</u>	<u>PST/INF</u>
<b>C<sub>1</sub>VC<sub>2</sub></b>	<b>C<sub>1</sub>V-t/d</b>
<i>xār-</i>	<i>xūd</i>

- b. **Type 2**: All stems contain the same number of consonants, but the present-stem consonants differ from those in the other stems.

<u>PRS</u>	<u>PST/INF</u>
<b>C<sub>1</sub>VC<sub>2</sub></b>	<b>C<sub>1</sub>VC<sub>3</sub>-t/d</b>
<i>riwāz-</i>	<i>riwixt</i>

The majority of verbs which exhibit irregularities in consonants also exhibit irregularities in their vowels, as is the case with the two examples provided here. Although irregularities in verbs' consonants and vowels are related phenomena, they are generally the result of fundamentally different historical processes and are therefore treated separately here. (Section 9.1.4 deals with vowel irregularities in verb stems.)

### 9.1.3.2 Historical developments leading to consonant irregularities in verb stems.

Irregularities in verb-stem consonants have two fundamental historical sources. The first source relates to erstwhile productive affixes which attached to present stems but not other stems. As a result, the consonantal reflex of such an affix appears in a modern Shughni verb's present stem but not in its other stems.

The two historical affixes whose reflexes are most commonly found in modern Shughni present stems are Proto-Indo-European *\*-skē-*, thought to once mark iterative or inchoative verbs, and a nasal infix *\*-n-* or suffix *\*-n(a)u-* (Dodykhudoeva 1988: 29-38). The former came to be a marker of intransitivity in Shughni verbs as the suffix *\*-sa-*, which shows up in the present stems of many modern intransitive verbs as a stem final *-s-*, *-c-*, or *-θ-*, as modulated by phonetic position. In particular, this suffix appears as *-s-* when following a vowel, as *-c-* when following a consonant, and as *-θ-* as a result of the historical combination of a preceding stem-final *-d-*. The reflex of the nasal affix, for its part, is found in a more diverse range of verbs, but is thought to have been a marker of transitivity at one point in the language's history. Examples of the reflex of each historical affixes are given in

Table 9.3. The past stem of each verb, which does not contain the reflex of the corresponding affix, is provided for comparison.

**Table 9.3:** Historical affixes and consonant irregularities in verb stems.

AFFIX	REFLEX	Present		Past		
		STEM	HISTORICALLY	STEM	HISTORICALLY	GLOSS
* <i>-ské-</i> > * <i>-s(a)-</i>	-s-	<i>naǰǰts-</i> <i>piðis-</i>	* <i>nir-ga-sa-</i> * <i>pati-di-sa-</i>	<i>naǰǰtd</i> <i>piðid</i>	* <i>nir-ga-ta-</i> * <i>pati-di-ta-</i>	‘pass’ ‘ignite’
	-c-	<i>wirāfc-</i> <i>biðafc-</i>	* <i>awi-rab-s-ya</i> * <i>upa-dab-sa-</i>	<i>wirūvd</i> <i>biðūvd</i>	* <i>awi-rab-ta-</i> * <i>upa-dab-ta-</i>	‘stand’ close
	-θ-	<i>ricθ-</i> <i>niθ-</i>	* <i>frat-rad-s-ya</i> * <i>ni-had-s-ya</i>	<i>ricūst</i> <i>nūst</i>	* <i>frat-rad-ta-</i> * <i>ni-had-ta-</i>	‘flee’ ‘sit’
* <i>-n(a)u-</i> / * <i>-na-</i>	-n-	<i>ǰin-</i> <i>piðin-</i> <i>yān-</i>	* <i>sǰ-nau-</i> * <i>pati-di-na-</i> * <i>ar-na-</i>	<i>ǰud</i> <i>piðid</i> <i>yūd</i>	* <i>sru-ta-</i> * <i>pati-di-ta-</i> * <i>ar-ta-</i>	‘hear’ ‘ignite’ ‘grind’
* <i>-n-</i>	-n-	<i>ǰičand-</i>	* <i>skǰ-n-t-</i>	<i>ǰičuxť</i>	* <i>skart-ta-</i>	‘cut’

The second source of discrepancies in the consonants of verb stems relates to regular sound changes in stems which did not contain extra morphological material. As a rule, the stem-final consonant of a present stem was generally found in a position before a vowel, as most inflectional affixes are vowel-initial, while the same consonant appeared before the consonant \**t/d* in past and infinitive stems. These differing phonetic environments provided the preconditions for the distinct development of certain stem-final sounds.

Three regular sound-change rules which have resulted in distinct consonants in present and past stems are the following. First, the consonants \**s*, *š*, and \**z* become modern Shughni *ǰ* when before *t*, which has resulted in the alternations (PRS~PST/INF) *s*~*ǰ* and *z*~*ǰ*. This is the case, for instance, in the verbs *diveš*~*diviǰť* ‘show’ and *raz*~*riǰť* ‘fall’. Second, the clusters \**rt* and \**rd* become \**ǰ* when before *t*. This has resulted in the alternation *rð*~*ǰ*, such as in the verbs *tidarð*~*tidiǰť* ‘tear’ and *šarð*~*šiǰť* ‘defecate’. And third, the glide \**w* and liquid \**r* were often deleted when preceding a *t/d* (Sokolova 1967: 46, 58). This has resulted in alternations of the type

$w \sim \emptyset$  and  $r \sim \emptyset$ . These alternations are found, for instance, in the verbs  $s\bar{a}w\sim sut$  ‘go; become’ and  $x\bar{a}r\sim x\bar{u}d$  ‘eat’.

Finally, in some cases the final consonant of the present and past stems share place and manner of alternations, but differ in voicing. When this occurs, it is always the present stem whose final consonant is voiced – presumably due to assimilation with surrounding vowels – while the final consonant of the past and infinitive stems is voiceless.

Examples of each consonant alternation due to regular sound changes, as well as voicing alternations, are given in Table 9.4; a more complete list can be found in Table B.2 in Appendix B.

**Table 9.4:** Consonant alternations in verb stems due to regular sound changes.

<u>SOUND CHANGE</u>	<u>PRS. STEM</u>	<u>HISTORICALLY</u>	<u>PST STEM</u>	<u>HISTORICALLY</u>	<u>GLOSS</u>
$*s, \check{s}, z > \check{x}, \_t$	<i>aboz-</i>	<i>*apa-āz-a-</i>	<i>abē<math>\check{x}</math>t</i>	<i>apa-āš-ti</i>	swallow
	<i>dives-</i>	<i>*ati-dais-a-</i>	<i>divi<math>\check{x}</math>t</i>	<i>de-diš-ta-</i>	show
	<i>žoz-</i>	<i>*gāz-a-</i>	<i>žē<math>\check{x}</math>t</i>	<i>gāš-ti-</i>	run
$*rt, rd > \check{x}, \_t$	<i>rarō-</i>	<i>*fra-rd-</i>	<i>ru<math>\check{x}</math>t</i>	<i>fra-ṛd-ta</i>	take apart
	<i>tidarō-</i>	<i>*ati-tard-a-</i>	<i>tidu<math>\check{x}</math>t</i>	<i>ti-tard-ta</i>	rip
	<i>zidarō-</i>	<i>*us-tard-a-</i>	<i>zidu<math>\check{x}</math>t</i>	<i>us-tard-ta</i>	unravel
$*w, r > \emptyset, \_t$	<i>sāw-</i>	<i>*čyaw-a-</i>	<i>sut</i>	<i>č(y)u-ta</i>	go; become
	<i>vār-</i>	<i>*bar-a-</i>	<i>vūd</i>	<i>bar-ta</i>	bring
	<i>xār-</i>	<i>*xwar-a-</i>	<i>xūd</i>	<i>xwar-ta</i>	eat
Voicing alternations	<i>tilāb-</i>		<i>tilāpt</i>		ask for
	<i>šānd-</i>		<i>šīnt</i>		laugh
	<i>niyũ<math>\check{y}</math>-</i>		<i>niyu<math>\check{x}</math>t</i>		listen

### 9.1.4 Irregular verbs: Vowel alternations

In regular verbs such as *fāmtow* ‘know’, all stems have the same stem vowel.<sup>3</sup> Thus, the stem inventory of a regular verb with present-stem of shape CVC is as in (205):

(205) **Vowel pattern in the stems of a regular verb**

<u>PRS</u>	<u>PRS.3SG</u>	<u>PST</u>	<u>INF</u>
CV <sub>1</sub> C	CV <sub>1</sub> C-t/d	CV <sub>1</sub> C-t/d	CV <sub>1</sub> C-t/d
<i>fām-</i>	<i>fāmt</i>	<i>fāmt</i>	<i>fāmt</i>

Irregular verbs may stray from this pattern in that the following vowels may differ from one another:

- (i) **the present-stem vowel;**
- (ii) **the present-stem vowel in the third-singular form;**
- (iii) **the past-stem vowel**, which is always identical to the perfect-stem vowel;<sup>4</sup>
- (iv) **the infinitive-stem vowel.**

Verbs which are irregular with respect to their stem vowels show a variety of patterns. A verb may have as many as four different vowels in its verb stems – that is, a different vowel in each of its present (non-3sg), present 3sg, past/perfect, and infinitive stems. Alternatively, an irregular verb may have only one stem in which the vowel is distinct from the other stems, or two pairs of stems in which the stem vowel is the same, and so on.

Similar to gender distinction in nouns and adjectives (Chapter 7), gender distinction in verbs (Section 9.2), and irregularities in the consonants of verb stems (Section 9.1.3), the phenomenon of irregularities in verb-stem vowels can largely be understood through regular sound changes. (In this section, a basic knowledge of these sound changes is assumed. For an overview, see Section 7.2 and especially Table 7.2.)

<sup>3</sup>For verbs whose stems have more than one vowel – e.g. *rīwīxt*, *PRS rīwāz-*, it is always the final vowel which is the stem vowel. The preceding vowels are the result of either a frozen pre-verbal element or epenthesis in what was originally a stem-initial consonant cluster, as discussed in Section 9.1.1.

<sup>4</sup>The discussion in this section leaves aside the issue of gender-distinguishing past and perfect stems. For these verbs, the vowel of the past feminine/plural form generally differs from that of the perfect feminine/plural form. This phenomenon is examined in Section 9.2.

The remainder of this section looks at the historical processes leading to irregularities in the vowels of each stem type. Section 9.1.4.1 first looks at non-third singular present stems, followed by third-singular present stems (Section 9.1.4.2), past (and perfect) stems (Section 9.1.4.3); and finally common patterns in infinitive stems (Section 9.1.4.4). The information in these subsections is rather complete (read also *dense*), and hence the reader wishing for a simpler overview is directed to the summary in Section 9.1.4.5, where the most common vowel patterns in irregular verbs are outlined.

#### 9.1.4.1 Historical tendencies in (non-3<sub>SG</sub>) present stems.

Certain historical developments have made some stem vowels more common in present stems, with the most common being long *ā*, short *a*, and short *i*.

The modern Shughni stem vowel *ā* is often the result of an ancient Iranian stem with the vowel *\*a* in *a*-umlaut position, where this phonetic position was created by the *\*ā* of inflectional endings such as first-person singular *\*-āmi* and first-person plural *\*āmahi* (e.g. Dodykhudoeva 1988: 20). Modern Shughni present stems with this vowel include, for instance, *vār-* ‘bring’, *xār-* ‘eat’, and *zidār-* ‘sweep’, among several others. A short *a* in a present stem is sometimes the reflex of *\*u*, rather than *\*a*, in *a*-umlaut position, as in *viraŷ-* (< *\*bruš-a-*) and *kaŷ-* ‘slaughter’ (< *\*kuš-a-*). In other cases, it is another reflex of a stem *\*a*, often before two unvoiced consonants (e.g. *biđaŷc-* ‘close’ < *\*upa-dab-sa-*).

The stem vowel *i* is often the result of a short stem vowel having undergone fronting, either from the influence of a surrounding palatal consonant *\*č*, as in *ti-* ‘go’ < *\*tač-a-*, or else as a result of a following suffix *\*-ya-*, which marked many intransitive verbs (e.g. *pi-* ‘rot’ < *\*pu-ya-*). Long *\*ī* in non-third-singular present stems is somewhat more rare and is generally the result of a stem-final *\*a*, which in neutral position naturally developed into Shughni *ī*, as in *naŷŷis-* ‘pass’ < *\*nir-ga-ta-*. Examples of verbs whose present stems contain these vowels are given in Table 9.5 along with their reconstructed historical stems, which are taken from Dodykhudoeva (1988).

Vowels which are found more rarely in non-third-singular present stems of irregular are likewise generally the result of regular sound developments. The vowel *e* is most often the reflex of the cluster *\*ai*, as in *dives* ‘show’ < *\*ati-dais-a-*. The vowel *o* in present stems is often the reflex of Old Iranian *\*ā*, as in *boz-* ‘send’ < *\*apa-āz-a-*. The vowel *ê*, although most common in the present stems of regular causative-type verbs such as *firêptow* ‘deliver’, is

**Table 9.5:** Common vowels in (non-3sg) present stems.

STEM VOWEL	PRS STEM	HISTORICALLY	GLOSS
<i>ā</i>	<i>anjāv</i>	* <i>han-kap-a-</i>	grab
	<i>vār</i>	* <i>bar-a-</i>	bring; carry
	<i>xār</i>	* <i>xwar-a-</i>	eat
<i>a</i>	<i>kaŷ</i>	* <i>kuš-a-</i>	slaughter
	<i>virāŷ</i>	* <i>bruš-a-</i>	break
	<i>ġikar</i>	* <i>skar-a-</i>	search
<i>i</i>	<i>andīdz</i>	* <i>ham-tač-a(ya)-</i>	get up
	<i>pī</i>	* <i>pu-ya-</i>	rot
	<i>zini</i>	* <i>sna-ya-</i>	wash

also found in some irregular verbs where the original \**ā* was fronted, presumably either due to the influence of a following suffix \*-*aya*, as in *rinēs-* ‘forget’ < \**fra-nās-aya*, or this same factor in combination with a surrounding \**č*, as in *vidēdz-* ‘irrigate’ < \**abi-tač-aya-*. The vowel \**ū* is uncommon in present stems, although it is found in some instances where the cluster \**aw*/\**au* is found in a closed syllable, such as in *ōūdz-* ‘milk’ < \**dauŷ-a-*. (See Table B.4 in Appendix B for a complete list of irregular verbs listed according to their present-stem vowel.)

#### 9.1.4.2 Deviations in the vowels of third-singular present stems.

For many irregular verbs, the present stem of the third-singular conjugation differs from that of the non-third-singular conjugations. This may be the result of one of three factors. First, some third-singular present stems show the effects of *i*-umlaut as a result of the historical third-singular ending \*-*ti*, whose final vowel *i* placed the stem vowel in this position. Thus, in irregular verbs whose non-third-singular present-stem vowel is *ā* or *a*, the third-singular stem typically has the vowel *ī*, as in *anjīvd*, *vīrt* ‘brings’, and *xīrt* ‘eats’ (non-3sg present stems *anjāv-*, *vār-*, and *xār-*, respectively). In those irregular verbs whose present-stem vowel is *o* (< \**ā*), the third-singular stem is often *ē*, which is the standard reflex of \**ā* in *i*-umlaut position (e.g. *yōs-* ‘take; carry’ > *yēst* ‘takes; carries’).

Second, the third-singular present vowel in some cases has developed differently due to its being in a position before two coda consonants. For some verbs, this results in a short stem vowel *i* lengthening to *ī*, as in PRS *wīn-* > PRS.3SG *wīnt* ‘sees’. For other verbs, this position has led to long *ā*, and in some cases short *a*, becoming long *o*

(e.g. PRS *riwāz-* > PRS.3SG *riwozd* ‘flies away’). The resulting long *o* has changed further to *û* when followed by a nasal, as in PRS *šānd-* > (*šont*) > *šûnt* ‘laughs’.

And finally, a third-singular stem vowel may differ from the non-third singular stem vowel due to the regular contraction of *āw* > *û* when followed by a coda consonant. This occurs, for instance, in the verbs *sāw-* > (*sāwd*) > *sûd* ‘goes’ and *nāw-* > (*nāwd*) > *nûd* ‘cries’. Examples of irregular third-singular stem vowels are given in Table 9.6. A more complete list of verbs exhibiting this phenomenon can be found in Table B.5 in Appendix B.

**Table 9.6:** Common vowels in (non-3SG) present stems.

ALTERNATION	PRS(NON-3SG) STEM	PRS.3SG STEM	GLOSS
<i>ā</i> > <i>ī</i>	<i>vār-</i> <i>xār-</i>	<i>vīrt</i> <i>xīrt</i>	bring eat
<i>i</i> > <i>ī</i>	<i>andīdz-</i> <i>wīn-</i>	<i>andīzd</i> <i>wīnt</i>	get up see
<i>ā</i> > <i>o</i>	<i>bāf-</i> <i>riwāz-</i>	<i>boft</i> <i>riwozd</i>	work; function fly away
<i>āw</i> > <i>û</i>	<i>nāw-</i> <i>sāw-</i>	<i>nûd</i> <i>sûd</i>	cry go; become

### 9.1.4.3 Common patterns in past and perfect stem vowels.

In irregular verbs, the stem vowel of past and perfect stems is most often *û* or *u*. This reflects the Old Iranian vowel *\*a* or *\*u* in neutral position before two consonants, as modern Shughni past and perfect forms historically contain the suffixes *\*-ta-* and *\*ta-ka-*, respectively, where the short vowel *a* created neutral position, and the addition of the past *t/d* and perfect *č/j* (< *\*k*) resulted in the two (or three) syllable-final consonants. Long *û* is typically the result of *\*a* in neutral position before two consonants. This is the case for the verbs *anjûvd* ‘grabbed’, *vûd* ‘brought’, and *xûd* ‘ate’ (cf. PRS *anjāv-*, *xār-*, and *vār-*, respectively). Short *u* in past and perfect stems is in many cases the reflex of *\*u*, as in *vud* ‘was’ (< *\*bu-ta-* and *sut* ‘went; became’ (< *\*č(y)u-ta-*). However, in a number of verbs where we would expect long *û*, we instead find short *û*, such as *arrux̣t* ‘reach/jump up’ (< *\*fra-ras-ta-*) and *riwux̣t* ‘fly away’ (< *\*fra-waş-ta-*).



Other vowels commonly occurring in past stems are long  $\bar{i}$  and  $o$ . The former occurs, on the one hand, where a verb's stem is historically vowel final and  $*a$  in neutral position before a single consonant naturally results in  $\bar{i}$ , as in  $z\bar{i}d$  'kill' (<  $*\bar{j}a-ta-$ ) and  $na\check{y}\bar{j}\bar{i}d$  'pass' (<  $*nir-ga-ta-$ ), and on the other, in later formations in which the past stem has been leveled on analogy with the infinitive stem. This is the case, for instance, with  $f\bar{i}r\bar{i}pt$  'arrived' and  $b\bar{i}ft$  'work; function'. The vowel  $\bar{i}$  in past stems is sometimes also the result of a verbal noun with the suffix  $*-ti-$ , which was typically used for infinitive stems but was also sometimes used for past stems (Sokolova 1967; 1973). The vowel  $o$ , for its part, is the reflex of a stem  $*\bar{a}$ , such as in  $zino\bar{d}$  'washed'. (<  $sn\bar{a}-ta$ ) and  $rimo\bar{d}$  'commanded' (<  $*fra-m\bar{a}-ta-$ ).

Examples with past and perfect stems with the most common vowels, as discussed here, are given in Table 9.7. A full list of irregular past stems organized by stem vowel is given in Table B.6 in Appendix B.

**Table 9.7:** Common vowels in past and perfect stems.

<u>VOWEL</u>	<u>PST STEM</u>	<u>PRF STEM</u>	<u>GLOSS</u>
$\bar{i}$	$na\check{y}\bar{j}\bar{i}d$ $z\bar{i}d$	$na\check{y}\bar{j}\bar{i}o\check{j}$ $z\bar{i}o\check{j}$	pass kill
$o$	$biro\check{x}t$ $zo\check{x}t$	$biro\check{x}\check{c}$ $zo\check{x}\check{c}$	drink take
$u$	$ku\check{x}t$ $riwu\check{x}t$	$ku\check{x}\check{c}$ $riwu\check{x}\check{c}$	slaughter fly away
$\bar{u}$	$v\bar{u}d$ $x\bar{u}d$	$v\bar{u}\check{y}\check{j}$ $x\bar{u}\check{y}\check{j}$	bring eat

#### 9.1.4.4 Common patterns in infinitive stem vowels

Infinitive stems of irregular verbs most commonly contain the stem vowels  $\bar{i}$ ,  $i$ , or  $\acute{e}$ , where these vowels are the result of a vowel in  $i$ -umlaut position, as infinitive stems are almost always the reflexes of Iranian verbal nouns derived by the suffix  $*-ti-$ .

The vowel  $\bar{i}$  occurs commonly with irregular verbs whose present stem contains  $\bar{a}$  or  $i$ , and whose past stem contains  $\bar{u}$ , as in the infinitives  $x\bar{i}d$  'eat' (cf. PRS  $x\bar{a}r$ , PST  $x\bar{u}d$ ) and  $t\bar{i}d$  'go; walk' (cf. PRS  $t\bar{i}$ , PST  $t\bar{u}yd$ ). This is generally the

result of *\*a* in *i*-umlaut position. A short *i* commonly occurs in the infinitive stems of verbs whose stem vowel was historically *\*u*, such as in *kiġt* ‘slaughter’ (< *\*kuš-ti-*). However, many verbs with *\*a* contain short *i* where we would expect long *ī* in their infinitive stems, such as *riwiġtow* (< *\*fra-waz-ti-*).

The infinitive stem vowel *ê* in irregular verbs occurs most commonly in verbs which historically had the stem vowel *\*ā*, for which *i*-umlaut position has naturally caused the reflex *ê*. This is the case, for instance, for the verbs *ġēvd* ‘sleep’ (< *\*xwāp-ti-*) and *zinêd* ‘wash’ (< *\*snā-ti-*). Correspondingly, the stem vowel of these verbs’ present stems is generally *o*, which is the canonical reflex of *\*ā* in neutral and *a*-umlaut position: cf. *ġofc-* and *żoz-*. Again, however, the correlation of an infinitive-stem *ê* to a present-stem *o* is not perfect. In verbs such as *zēġt* ‘take’ – PRS. *zēz-* – the present-tense *ê* may be the result of leveling on analogy with the infinitive stem.

A list of infinitive stems with the most common vowels, as discussed here, is provided in Table 9.8. A full list of the infinitive stems of irregular verbs is given in Table B.7 in Appendix B.

**Table 9.8:** Common vowels in infinitive stems.

STEM VOWEL	INF STEM	HISTORICALLY	GLOSS
<i>ī</i>	<i>anjīvd</i>	<i>*han-kap-ti-</i>	grab
	<i>vīd</i>	<i>*bar-ti-</i>	bring; carry
	<i>xīd</i>	<i>*xwar-ti-</i>	eat
<i>i</i>	<i>kiġt</i>	<i>*kuš-ti-</i>	slaughter
	<i>niyiġt</i>	<i>*ni-guš-ti-</i>	listen
	<i>riwiġt</i>	<i>*fra-waz-ti-</i>	fly away
<i>ê</i>	<i>rimêd</i>	<i>*fra-mā-ti-</i>	command
	<i>ġēvd</i>	<i>*xwāp-ti-</i>	sleep
	<i>zēġt</i>	<i>*gāš-ti-</i>	run

#### 9.1.4.5 Vowel patterns in irregular verbs: Summary

This section has drawn upon historical linguistic research to examine patterns in the stem vowels of modern Shughni irregular verbs. Modern Shughni irregular verbs may have different vowels in any or all of the following four stems: (i) present (non-third-singular); (ii) present third-singular; (iii) past/perfect; and (iv) infinitive. In this summary, the most common three to four patterns found in each stem type are described and then displayed with

an example in Table 9.9.

Present-stem vowels of irregular verbs are commonly *ā*, *a*, or *i*. The former two vowels are the result of a stem vowel undergoing *a*-umlaut, which took place due to the long *\*ā* of Old Iranian agreement endings. The latter vowel is occasionally due to the palatalizing effect of a stem-adjacent *\*č*, but is more often the result of *i*-umlaut caused by a following suffix *\*-ya-*.

The present stem of third-singular forms often differs from that of other person-number combinations. Here, the third-singular vowel has often either undergone *i*-umlaut due to a following agreement suffix *\*-ti*, or else has undergone lengthening due to a following sonorant, often *n*. In some instances, the vowel *ā* does not undergo *i*-umlaut, but rather lengthens to *o*, which becomes *ū* before nasals.

The stem vowel of past and perfect stems in irregular verbs is most often *ū*, *u*, or *o*. The former two vowels are the reflex of *\*a* and *\*u* in neutral position before two consonants, a position which often came about from the addition of a past suffix *\*-ta-* or perfect suffix *\*-ta-ka-* to a consonant-final stem. The vowel *o*, for its part, is the reflex of ancient Iranian *\*ā* in all positions.

Finally, infinitive stems, which descend from verbal nouns with the suffix *\*-ti-*, most often have *i*-like vowels due to the *i*-umlaut caused by this suffix. Here, the vowel *ī* generally comes from *\*a*, *i* generally from *\*u* or *\*a*, and *ē* from *\*ā*.

Table 9.9 shows four vowels which illustrate the patterns described above.

**Table 9.9:** Irregular stem vowels: Summary in four examples.

	<u>PRS</u>	<u>3SG.PRS</u>	<u>PST</u>	<u>INF</u>	<u>GLOSS</u>
<i>*a</i> , PRS <i>a</i> -umlaut	<i>vāṛ-</i>	<i>vīrt</i>	<i>vūd</i>	<i>vīd</i>	bring; carry
<i>*a</i> , PRS <i>i</i> -umlaut	<i>andīdž-</i>	<i>andīzd</i>	<i>andūyd</i>	<i>andīd</i>	get up
<i>*u</i> , PRS <i>a</i> -umlaut	<i>kaṣ-</i>	<i>kīṣd</i>	<i>kuṣt</i>	<i>kīṣt</i>	slaughter
<i>*ā</i> , PRS <i>i</i> -umlaut	<i>ṣīci-</i>	<i>ṣīced</i>	<i>ṣīcod</i>	<i>ṣīcēd</i>	freeze (intr.)

### 9.1.5 Verb-stem shortening

In addition to irregular verbs whose stems differ from one another due to stem-consonant and stem-vowel irregularities, as discussed in Sections 9.1.3 and 9.1.4, respectively, a number of verb stems in Shughni may also undergo processes of *stem-shortening*. In shortened verb stems, one or more sounds of the full stem is dropped, sometimes along with an additional change, such as the shortening of the stem vowel or the blending of the stem with an inflectional affix. In the latter case, it is sometimes unclear where the verb stem stops and the inflection marker begins – see the first-singular present form *xām* (< *xār-um*) ‘I eat’ in Table 9.10 below. Past, perfect, and present verb stems can all undergo this type of change, but shortening is especially common in present stems, including imperative stems, which derive from them. Initial examples of stem shortening are given in Table 9.10, which illustrates this process for the verb *xīdow* ‘eat’. Shortened stems in this table are bolded.

**Table 9.10:** Stem-shortening in the verb *xīdow* ‘eat’.

	<u>FULL STEM</u>	<u>SHORTENED STEM</u>	<u>GLOSS</u>
PRESENT (1SG)	<i>xār-um</i>	<b><i>x-ām</i></b> or <b><i>xā-m</i></b>	‘I eat’
IMPERATIVE	<i>xār</i>	<b><i>xa</i></b>	‘eat!’
PAST	<i>xūd</i>	<b><i>xu</i></b>	‘ate’
PERFECT	<i>xūŷŷ</i>	<b><i>xūŷ</i></b>	‘(has) eaten’

Whether or not a verb stem is allowed to shorten is apparently a function of at least two factors: (i) its frequency of use and (ii) its phonological shape. Thus, on the one hand, it is typically the most commonly used Shughni verbs whose stems are shortened, and on the other, stems of a certain phonological shape are more likely to be shortened than verbs of other phonological shapes. As a result, the stems of certain common verbs are not shortened, ostensibly because of an unfavorable phonological shape, while certain less common stems are allowed to shorten, ostensibly because of a favorable phonological shape. These issues are addressed in the sections which follow, which are organized along the lines of stem type. Stem-shortening is first examined in (non-imperative) present stems (Section 9.1.5.1), followed by imperatives (Section 9.1.5.2), and finally past and perfect stems (Section 9.1.5.3).

### 9.1.5.1 Present-stem shortening

The shortening of (non-imperative) present stems falls into two types, both of which involve the blending of a present stem with an agreement suffix. The first type, which affects first-singular and second- and third-plural present inflected forms, takes a stem with an ending and inflectional suffix of shape  $C\bar{V}C-VN$  (e.g. *xār-um* ‘I eat’) and turns it into  $CVN$  (*xām*), where the final nasal in the shortened form matches that of the inflectional ending (*m* or *n*). The long vowel of the shortened stem, for its part, is either the same as in the original, long-form of the stem – in the case of first-singular shortened forms such as *xām* < *xār-um* – or else it matches the vowel of the inflectional ending – in the case of second- and third-person plural shortened forms such as *xen* < *xār-en* ‘they eat’. This is schematized in (206):

(206) **Shortening of present stems and agreement suffixes: Type 1**

	<u>LONG STEM</u>	<u>AGR. SFX.</u>		<u>SHORT FORM</u>	<u>EXAMPLE</u>
1SG	$C\bar{V}_1C-$	$-V_2N_1$	>	$C\bar{V}_1N_1$	<i>xār-um</i> > <i>xām</i>
2PL, 3PL	$C\bar{V}_1C-$	$-\bar{V}_2N_1$	>	$C\bar{V}_2N_1$	<i>xār-en</i> > <i>xen</i>

For verbs in which this type of present-stem shortening takes place, most commonly the first-singular form, second-plural, and third-plural forms are all affected, although one verb – *lūvdow* ‘say’ – only allows for the shortening of the first-singular form (*lūm* < *lūv-um*), while the second- and third-plural forms require the long stem (*lūvet* and *lūven*, respectively). Other person-number combinations, namely 2SG, 3SG, and 1PL, are never affected by this type of shortening.

Examples of verbs which undergo this type of shortening are given in Table 9.11. The majority of verbs which participate in this type of present-stem shortening belong to the most commonly used verbs in Shughni, but less commonly used verbs, such as *niṣpīdow* ‘step on’, also participate. Other common verbs like *kin-* ‘do’, which because of their frequency might also be expected to shorten, apparently do not shorten because their phonological shape is not conducive to this process. The present stem *kin-*, for instance, lacks the long vowel present in other stems which undergo this type of shortening.

**Table 9.11:** Verbs which undergo present-stem shortening: Type 1.

<u>GLOSS</u>	<u>LONG STEM</u>		<u>1SG.PRS</u>	<u>2PL.PRS</u>	<u>3PL.PRS</u>
eat	<i>xār-</i>	>	<i>xām</i>	<i>xet</i>	<i>xen</i>
bring	<i>vār-</i>	>	<i>vām</i>	<i>vet</i>	<i>ven</i>
go; become	<i>sāw-</i>	>	<i>sām</i>	<i>set</i>	<i>sen</i>
hit; give	<i>ḍāḍ-</i>	>	<i>ḍām</i>	<i>ḍet</i>	<i>ḍen</i>
sell	<i>parḍāḍ-</i>	>	<i>parḍām</i>	<i>parḍet</i>	<i>parḍen</i>
step on	<i>niṣpār-</i>	>	<i>niṣpām</i>	<i>niṣpet</i>	<i>niṣpen</i>
say	<i>lūv-</i>	>	<i>lūm</i>	<i>lūvet*</i>	<i>lūven*</i>
take	<i>zēz-</i>	>	<i>zēm</i>	<i>zet</i>	<i>zen</i>

The second type of present-stem shortening is also a blend of sorts, but in this case only the first-singular form is affected. Here, verbs whose present stem ends in a short *-i-* end up with a long *ī* in their first-person singular forms, and the palatal glide *y* and short *u* of the inflectional suffix *-(y)um* are deleted. For instance, the present stem *tī-* ‘go’ has the first-singular form *tīm* ‘I go’, rather than *tī-yum*. This second type of present-stem shortening is schematized in (207):

(207) **Shortening of present stems and agreement suffixes: Type 2**

	<u>STEM</u>	<u>AGR. SFX.</u>		<u>SHORT FORM</u>	<u>EXAMPLE</u>
1SG	(...)Ci-	-(y)um	>	(...)Cīm	<i>tī-yum</i> > <i>tīm</i>

This process affects all verbs whose present stem ends in a short *-i-*. Examples are given in 9.12.

**Table 9.12:** Verbs which undergo present-stem shortening: Type 2.

<u>GLOSS</u>	<u>PRS STEM</u>		<u>1SG.PRS</u>
go; walk	<i>tī-</i>	>	<i>tīm</i>
be	<i>vī-</i>	>	<i>vīm</i>
fall	<i>ḍī-</i>	>	<i>ḍīm</i>
be able to	<i>vārḍī-</i>	>	<i>vārḍīm</i>
wash	<i>zini-</i>	>	<i>ziniīm</i>
find	<i>viri-</i>	>	<i>viriīm</i>

Note that for present stems which can undergo either type of shortening discussed here, the shortening is generally obligatory, or at the very least, much more common than a lack of shortening. Thus, for instance, in the case of the verb *xīdow* ‘eat’, the shortened form *xām* is much more frequently used than the longer form *xār-um*, which for some speakers may even sound ungrammatical.

### 9.1.5.2 Imperative shortening

The shortening of imperatives is also generally obligatory and affects virtually all the same verbs affected by the first type of shortening discussed above, with the addition of the verb *čīdow* ‘do’. The shortening of second-singular imperative stems, which are the equivalent of a bare (non-inflected) present stem, does not constitute a blend, but rather a truncation. When a second-singular imperative is shortened, the final consonant is dropped and the stem vowel becomes its short-vowel equivalent ( $\bar{a} > a$ ;  $\bar{e} > i$ ;  $\bar{u} > u$ ). Shortened second-plural imperatives, however, are identical to their corresponding shortened present form. This is schematized in (208).

#### (208) Shortening of imperatives

	<u>PRS. STEM</u>	<u>LONG IMP.</u>		<u>SHORT IMP</u>	<u>EXAMPLE</u>
2SG.IMP	<b>C<math>\bar{V}_1</math>C-</b>	<b>C<math>\bar{V}_1</math>C</b>	>	<b>CV<sub>1</sub></b>	<i>xār</i> > <i>xa</i>
2PL	<b>C<math>\bar{V}_1</math>C-</b>	<b>C<math>\bar{V}_1</math>C-et</b>	>	<b>C-et</b>	<i>xār-en</i> > <i>xet</i>

Examples of shortened imperatives are given in Table 9.13.

### 9.1.5.3 Past and perfect-stem shortening

The shortening of past and perfect stems differs from present-stem and imperative shortening in that it is far less common and typically optional. Past-stem shortening involves the truncation a handful stems ending in *-ūd*. Here, the final *-d* is dropped and the long vowel  $\bar{u}$  becomes short *u*. The examples given in 9.14, to my knowledge, constitute an exhaustive of verbs which undergo this type of shortening.

**Table 9.13:** Verbs which undergo imperative shortening.

<u>GLOSS</u>	<u>PRS. STEM</u>		<u>2SG.IMP</u>	<u>2PL.IMP</u>
eat	<i>xār-</i>	>	<i>xa</i>	<i>xet</i>
bring	<i>vār-</i>	>	<i>va</i>	<i>vet</i>
go; become	<i>sāw-</i>	>	<i>sa</i>	<i>set</i>
hit; give	<i>ḏāḏ-</i>	>	<i>ḏa</i>	<i>ḏet</i>
sell	<i>parḏāḏ-</i>	>	<i>parḏa</i>	<i>parḏet</i>
step on	<i>niḵpār-</i>	>	<i>niḵpa</i>	<i>niḵpet</i>
say	<i>lūv-</i>	>	<i>lu</i>	<i>lūvet*</i>
take	<i>zēz-</i>	>	<i>zi</i>	<i>zet</i>
do	<i>kin-</i>	>	<i>ki</i>	<i>kinet*</i>

**Table 9.14:** Shortening of past stems.

<u>GLOSS</u>	<u>LONG PST STEM</u>		<u>SHORT PST STEM</u>
eat	<i>xūd</i>	>	<i>xu</i>
do	<i>čūd</i>	>	<i>ču</i>
bring	<i>vūd</i>	>	<i>vu</i>
sweep	<i>zidūd</i>	>	<i>zidu</i>

The shortening of perfect stems, for its part, does not involve the dropping of a stem-final consonant, as in past-stem shortening, but rather the dropping of a stem-internal consonant. This process affects only those perfect stems which have a *ḏ* or *ḡ* preceding their stem-final consonant *ḡ*. Examples are given in Table 9.15.

**Table 9.15:** Shortening of perfect stems.

<u>GLOSS</u>	<u>LONG PRF STEM</u>		<u>SHORT PRF STEM</u>
eat	<i>xūḡḡ</i>	>	<i>xūḡ</i>
do	<i>čūḡḡ</i>	>	<i>čūḡ</i>
bring	<i>vūḡḡ</i>	>	<i>vūḡ</i>
give; hit	<i>ḏoḡḡ</i>	>	<i>ḏoḡ</i>
die.M	<i>mūḡḡ</i>	>	<i>mūḡ</i>
step on	<i>niḵpūḡḡ</i>	>	<i>niḵpūḡ</i>

In sum, Shughni verbs may be irregular due to a variety of factors. Many verbs show deviations in the consonants



or vowels found in their stems. These phenomena can be understood from a historical linguistic perspective. The shortening of verb stems, which can affect all types of stems except infinitive stems, is more easily understood from a synchronic perspective. In particular, this type of shortening occurs with Shughni verbs which are both commonly used and have favorable phonological shapes which allow for shortening to happen. The following section of this chapter turns to another phenomenon which might be considered a kind of irregularity in verb stems, namely the expression of gender in certain verbs.

## 9.2 Gender in verbs

A subset of Shughni verbs – roughly *unaccusative* verbs in the sense of (Perlmutter 1978) – exhibit gender distinction in their past and/or perfect stems (see Section 4.4.2 on the distinction between unergative and unaccusative verbs in Shughni). The expression of gender on Shughni verbs is similar to the expression of gender on nominals in that it is done primarily via stem-internal vowel alternations, rather than suffixation. However, in perfect stems, the final consonant is also implicated, with feminine perfect stems always containing a final alveolar affricate and masculine perfect stems always containing a final palato-alveolar affricate.

The formal means by which gender is expressed in modern Shughni verbs, including both vowel and consonant alternations, are the result of the same regular sound changes in the vowel and consonant system which shaped gender distinction in nominals. As such, these changes are discussed only briefly in this section, and the reader is directed to Section 7.2 for a more detailed account of these developments.

The remainder of this section is organized as follows. Section 9.2.1 discusses the types of verbs which exhibit gender distinction in their past and perfect stems and how this fits into the larger picture of unaccusativity in Shughni. Section 9.2.2 then turns to a short overview of historical developments which shaped past and perfect stems in Shughni. Section 9.2.3 provides details on the formal means for distinguishing gender in Shughni verbs – that is, the vowel and consonant alternations used in gender-distinguishing past and perfect stems. And finally, Section 9.2.4 introduces a further set of verbs (and adverbs) which have been known to distinguish gender in Shughni, namely onomatopoeic words.

### 9.2.1 Which verbs distinguish gender?

Gender-distinguishing verbs in Shughni generally bear both the semantic properties typical of unaccusative verbs cross-linguistically as well as the language-specific morphosyntactic properties of unaccusative verbs within Shughni. From a semantic perspective, gender-distinguishing verbs in Shughni primarily express directed motion (e.g. *go*, *come*, *arrive*) and other actions or states for which the subject is less agentive (e.g. *be*, *become*, *sleep*, etc.). In this way, the class of gender-distinguishing verbs in Shughni is similar to the class of gender-distinguishing verbs in other languages, notably French and Italian, which are also generally understood to be unaccusative (see Kobayashi 2022 and references therein). From a morphosyntactic perspective, gender-distinguishing verbs invariably display another hallmark sign of unaccusative verbs in Shughni: the lack of a third-singular agreement clitic in the past and perfect.

However, whereas all gender-distinguishing verbs in Shughni lack a third-singular past-tense clitic, not all verbs lacking this clitic distinguish gender. Thus, for instance, the verbs *fir̥ptow* ‘arrive’ and *yattow* ‘come’ behave as unaccusatives in that they do not call for a third-singular clitic with their past stems, but neither have gender distinguishing past or perfect stems (cf. *fir̥pt* (PST) / *fir̥pč* (PRF) and *yat* (PST) / *yaθč* (PRF)). We can say, therefore, that only a subset, albeit the majority of unaccusative verbs in Shughni distinguish gender in their past and perfect stems.

### 9.2.2 Historical considerations in gender-distinguishing verbs

Similar to the expression of grammatical gender on nouns and adjectives, discussed in Section 7.3, inflection for gender in verbs is not realized via suffixes, as it is in many Indo-European languages, but rather by means of a stem-internal vowel alternation (e.g. *vad* ‘was.F’ and *vud* ‘was.M’), which is accompanied in perfect stems by a stem-final consonant alternation – a final alveolar affricate *c* or *dz* for feminine perfect stems (e.g. *vic* ‘be.PR.F’) and a palato-alveolar affricate *č* or *ǰ* for masculine stems (e.g. *vuǰ* ‘be.PR.F.M’). In many cases, as in *vuǰ*, an interdental or velar fricative precedes the final affricate in the masculine perfect stem; this does not occur in feminine stems.

From a historical point of view, the same developments in the Shughni vowel system that were relevant for gender in nouns and adjectives are also at play in Shughni verbs (see Section 7.2, especially Table 7.2). Here, however, the history of changes specific to Shughni past and perfect stems helps explain the vowel patterns we see in these

forms, as well as the consonant alternations of perfect stems.

Both past and perfect verb stems in modern Shughni descend from Iranian participles with the suffixes *\*-ta* (M) or *\*-tā* (F), with the perfect containing an additional suffix *\*-aka* (M) and *\*-(a)či* (F). In past stems, the masculine ending containing short *\*a* left vowels in neutral position with respect to umlaut, whereas the feminine ending containing long *ā* created an *a*-umlaut condition. In the perfect, then, the stem vowel in masculine forms was likewise in neutral position, but the stem vowel in feminine forms was in *i*-umlaut condition. Hence, in Shughni gender-distinguishing verbs, masculine forms often have the same stem vowel in both past and perfect stems. However, feminine past stems often have one vowel in the past stem and a separate vowel in the perfect stem, and in general the former is an *a*-like vowel, while the latter is an *i*-like vowel.

Regarding the consonant alternation in perfect stems, the affricate in both feminine and masculine perfect stems is the descendant, modulo regular sound changes, of the consonant in the ancient perfect stem (*č/-j* < *\*-aka*, *-c/-dz* < *\*-ači*). The fricative before the affricate, when it appears, is the descendant of the stem-final consonant, most often *θ/ð* (< *\*-t*), but also *š/ž* (for ancient stems containing final *\*-g* – c.f. Shughni *mūšž* ‘died.PERF.M’). (See, for instance, Sokolova 1967 and Karamshoev 1978 for more on the relevant historical processes and their reflexes in different languages of the Shughni-Rushani subgroup.) These specific changes have led to the current situation for the expression of gender in modern Shughni, which is the topic of the following subsection.

### 9.2.3 Vowel and consonant alternations in past and perfect stems

Here, vowel alternations in gender-distinguishing past stems are examined in Section 9.2.3.1, and vowel and consonant alternations in gender-distinguishing perfect stems are examined in Section 9.2.3.2.

#### 9.2.3.1 Gender in past stems

There are three vowel alternations used in gender-distinguishing past stems:

- (i) *u* ~ *a* – e.g. *sut* (M) ~ *sat* (F) ‘go.PST’
- (ii) *i* ~ *ā* – e.g. *siftid* (M) ~ *sifād* (F) ‘rise.PST’
- (iii) *ū* ~ *o* – e.g. *nūst* (M) ~ *nost* (F) ‘go.PST’

The first type occurs in verbs which historically contained Iranian *\*u*, *\*ū*. In masculine forms, this vowel was in neutral position, which gave Shughni *u*, and in feminine forms it was in *a*-umlaut position, which gave Shughni *a*. The second two alternations involve the stem vowel *\*a*, with the difference being the presence of one or two consonants following this vowel. In the second type, where there was a single following consonant, neutral position gave the vowel *ī* for masculine forms, and *a*-umlaut position gave the vowel *ā* for feminine forms. In the third type, neutral position gave *ū* for masculine forms, and *a*-umlaut position gave *o* for feminine forms. These vowel alternations are summarized in Table 9.16.

### 9.2.3.2 Gender in perfect stems

The most salient difference in the vowels of perfect stems, compared to past stems, lies in the quality of the feminine vowel. In particular, the feminine vowels in past stems are all either low vowels *a* or *ā*, or else the back vowel *o*. In perfect stems, however, all feminine vowels are front vowels – *i*, *ī*, or *ē*. This situation, as mentioned above, is linked to the notion that the stem vowel in past-stem feminine forms was in *a*-umlaut position, which gave rise to *a*-like vowels, while in feminine perfect stems it was in *i*-umlaut position, which gave rise to *i*-like vowels. The following three vowel alternations are found in perfect stems:

- (i) *u* ~ *i* – e.g. *puđj* (M) ~ *pic* (F) ‘rot.PST’
- (ii) *ū* ~ *ī* – e.g. *ricuūsč* (M) ~ *ricīsč* (F) ‘rise.PST’
- (iii) *o* ~ *ē* – e.g. *xicodj* (M) ~ *xicēc* (F) ‘freeze.PST’

The first alternation involves the ancient vowel *\*u*, *\*ū*. In neutral position, with the masculine suffix *\*-aka*, we get Shughni *u*, and in *i*-umlaut position, with the feminine suffix *\*(a)či-*, we get *i*. The second type involves the vowel *\*a* before two consonants, which in neutral position results in Shughni *ū*, used in masculine stems, and in *i*-umlaut form gives *ī*, used in feminine stems. The third type involves the vowel *\*ā*, which in neutral position gives rise to masculine *o* and in *i*-umlaut position gives *ē*. These alternations are summarized in Table 9.17.

Note that in a few verbs, namely those corresponding to alternation (ii) in past stems – *ī* ~ *ā* – the stem vowel is identical in masculine and feminine perfect stems. These perfect forms both contain the vowel *ī* but are still distinguished by their final consonant. This is the case for the verbs *načjīdow* ‘pass’ (*načjīđj* ~ *načjīc*); *sifīdow* ‘rise’ (*sifīđj* ~ *sifīc*); and *načfīdow* ‘fall out’ (*načfīđj* ~ *načfīc*). The identical vowel in these stems is due to the fact that these verbs contained the stem vowel *\*a*, which in neutral and *i*-umlaut position resulted in a single

**Table 9.16:** Gender-distinguishing pairs of past stems.

ALTERNATION (M~F)	HISTORICAL STEM VOWEL	POSITION	EXAMPLE (M~F)	GLOSS	ETYMOLOGY (M~F)
<b>u~a</b>	<b>*u, *ū</b>	M: Neutral Position F: <i>a</i> -Umlaut Position	<i>v<u>u</u>d</i> ~ <i>v<u>a</u>d</i> <i>p<u>u</u>d</i> ~ <i>p<u>a</u>d</i> <i>θ<u>u</u>d</i> ~ <i>θ<u>a</u>d</i> <i>s<u>u</u>t</i> ~ <i>s<u>a</u>t</i>	be.PST rot.PST burn.PST be.PST	* <i>būta-</i> ~ <i>būtā-</i> * <i>puta-</i> ~ * <i>putā-</i> * <i>θuta-</i> ~ * <i>θutā-</i> cf. Av. <i>šuta-</i> ~ * <i>šutā</i>
<b>ī~ā</b>	<b>*a</b>	M: Neutral Position F: <i>a</i> -Umlaut Position	<i>naǰǰ<u>ī</u>d</i> ~ <i>naǰǰ<u>ā</u>d</i> <i>naǰǰ<u>ī</u>d</i> ~ <i>naǰǰ<u>ā</u>d</i> <i>sif<u>ī</u>d</i> ~ <i>sif<u>ā</u>d</i>	pass.PST fall out.PST rise.PST	* <i>niž-gata-</i> ~ <i>niž-gatā-</i> * <i>niš-fata-</i> ~ <i>niš-fatā-</i> * <i>us-patta-</i> ~ <i>us-pattā-</i>
<b>ū~o</b>	<b>*a</b>	M: Neutral Position (_2C) F: <i>a</i> -Umlaut Position (_2C)	<i>nūst</i> ~ <i>nost</i> <i>tūyd</i> ~ <i>toyd</i> <i>ricūst</i> ~ <i>ricost</i>	sit.PST leave.PST flee.PST	* <i>ni-hasta-</i> ~ <i>ni-hastā-</i> * <i>taxta-</i> ~ <i>taxtā-</i> * <i>us-rasta-</i> ~ * <i>us-rastā-</i>

Shughni vowel  $\bar{i}$ . This same convergence of the vowel  $*a$  into  $\bar{i}$  in neutral and  $i$ -umlaut positions is the reason why certain non-gender-distinguishing nouns in Shughni (e.g. *šīg* ‘calf’) have gender-distinguishing counterparts in other Shughni-Rushani varieties. Unsurprisingly, the Shughni perfect stems which have  $\bar{i}$  in both masculine and feminine also have separate vowels in their counterparts in other varieties (e.g. Rushani *nawžoj* ~ *nawžec* ‘pass.PRF’ and *sifoj* and *sifec* ‘rise.PRF’).

### 9.2.4 Gender distinction in onomatopoeic verbs

In addition to the expression of gender in canonically unaccusative verbs such as those discussed above, gender distinction is also expressed in certain verbs and adverbs which describe the sound an object or person makes.<sup>5</sup> Gender distinction in onomatopoeic words, like in verbs, adjectives and nouns, is done by stem vowel alternations. In the vast majority of cases, the masculine vowel is  $u$ , and the feminine vowel is  $a$ .

Importantly, however, gender-distinguishing verbs of the onomatopoeic type differ from typical unaccusatives in that they distinguish gender not only in their past and perfect stems, but also in their present stem and infinitive. Gender in onomatopoeic words fundamentally involves a stem with a vowel alternation, such as *kur-* ~ *kar-* ‘to rummage’ (see example 209), which can be turned into either a verb or an adverb. In the formation of adverbs, the stem either takes on a suffix  $-a(s)t$  – most commonly  $-ast$  – or else it is reduplicated. Thus, we can have either *kurast* (M) / *karast* (F) or *kur-kur* (M) / *kar-kar* (F). In the formation of verbs, the stem takes on the endings typical of various verb stems, depending on the grammatical context in which it is used. The paradigm of forms for three gender-distinguishing onomatopoeic stems – *kurtow* ~ *kartow* ‘rummage’; *duntow* ~ *dangtow* ‘bang’; and *buydow* ~ *baydow* ‘buzz’ – are provided below.

---

<sup>5</sup>These verbs have been said to be onomatopoeic insofar as they mimic sounds (e.g. Karamshoev 1978), although for some such as *kurtow* ‘rummage’, the connection between the phonological form of the verb and the action it denotes is less clear. I nonetheless use the term *onomatopoeic* to describe these verbs.

**Table 9.17:** Gender-distinguishing pairs of perfect stems.

ALTERNATION (M~F)	HISTORICAL STEM VOWEL	POSITION	EXAMPLE (M~F)	GLOSS	ETYMOLOGY (M~F)
<b>u~i</b>	<b>*u, *ū</b>	M: Neutral Position F: <i>i</i> -Umlaut Position	<i>vuōj</i> ~ <i>viċ</i> <i>puōj</i> ~ <i>piċ</i> <i>θuōj</i> ~ <i>θiċ</i> <i>suōj</i> ~ <i>siċ</i>	be.PR rot.PR burn.PR go.PR	* <i>būta-ka</i> ~ <i>būta-ċi</i> * <i>puta-ka</i> ~ * <i>puta-ċi</i> * <i>θuta-ka</i> ~ * <i>θuta-ċi</i> cf. Av. <i>šuta-ka</i> ~ * <i>šuta-ċi</i>
<b>ū~ī</b>	<b>*a</b>	M: Neutral Position (_2C) F: <i>i</i> -Umlaut Position (_2C)	<i>nūsċ</i> ~ <i>nīsc</i> <i>tūyĵ</i> ~ <i>tīc</i> <i>ricūsċ</i> ~ <i>ricīsc</i>	sit.PR go.PR flee.PR	* <i>hi-hasta-ka</i> ~ <i>ni-hasta-ċi</i> * <i>taxta-ka</i> ~ <i>taxta-ċi</i> * <i>us-patta-ka</i> ~ <i>us-patta-ċi</i>
<b>o~ê</b>	<b>*ā</b>	M: Neutral Position F: <i>i</i> -Umlaut Position	<i>xicōōj</i> ~ <i>xicēc</i> <i>ōōj</i> ~ <i>ōēc</i>	freeze.PR fall/hit.PR	* <i>ščāta-ka</i> ~ <i>ščāta-ċi</i> * <i>dāta-ka</i> ~ <i>dāta-ċi</i>

**Table 9.18:** Gender-distinguishing onomatopoeic verbs.

	MASC.	FEM.	MASC.	FEM.	MASC.	FEM.
GLOSS	‘to rummage’		‘to (make a) bang’		‘to buzz’	
<b>Verbs:</b>						
INF.	<i>kurtow</i>	<i>kartow</i>	<i>dungtow</i>	<i>dangtow</i>	<i>buydow</i>	<i>baydow</i>
PRS.	<i>kur-</i>	<i>kar-</i>	<i>dung-</i>	<i>dang-</i>	<i>buy-</i>	<i>bay-</i>
PST.	<i>kurt</i>	<i>kart</i>	<i>dungt</i>	<i>dangt</i>	<i>buyd</i>	<i>bayd</i>
PRF.	<i>kurč</i>	<i>karč</i>	<i>dungč</i>	<i>dangč</i>	<i>buyj</i>	<i>bayj</i>
IMPER.	<i>kur</i>	<i>kar</i>	<i>dung</i>	<i>dang</i>	<i>buy</i>	<i>bay</i>
<b>Adverbs:</b>						
SUFFIX	<i>kurast</i>	<i>karast</i>	<i>dungast</i>	<i>dangast</i>	<i>buyast</i>	<i>bayast</i>
REDUP.	<i>kur-hur</i>	<i>kar-kar</i>	<i>dung-dung</i>	<i>dang-dang</i>	<i>buy-buy</i>	<i>bay-bay</i>

Examples of gender-distinguishing onomatopoeic verbs in use are given in (209)–(211). The sentences in (209) involve the present stems of *kurtow* ~ *kartow*, where the vowel *u* in (209a) represents masculine and the vowel *a* in (209b) represents feminine. Agreement in person and number – in addition to gender – takes place between the subject and the verb. In the subsequent two examples, agreement is shown between a subject and an onomatopoeic adverb, which functions here as the non-verbal component of a complex verb. In these, too, the vowel *u* is found in the masculine stem, while the vowel *a* is found in the feminine stem.

- (209) a. Yu            yĩdā tar kūxni    dis        kur-t.  
 DEM.DIR.M boy in kitchen so.much rummage.PRS.M-3SG  
 ‘That boy is rummaging so much in the kitchen.’ (MASCULINE STEM: *kur-*)
- b. Yā            yāc tar kūxni    dis        kar-t.  
 DEM.DIR.F girl in kitchen so.much rummage.PRS.F-3SG  
 ‘That girl is rummaging so much in the kitchen.’ (FEMININE STEM: *car-*)
- (210) a. Yu            kolokolčik=i di    čūd    dung-ast.  
 DEM.DIR.M bell=3SG    MIR do.PST bang.M-ADV  
 ‘That bell all of a sudden made a banging (sound).’ (MASC. ADV.: *dungast*)
- b. Yā            dek=i    dang-ast    čūd.  
 DEM.DIR.F pot=3SG bang.F-ADV do.PST  
 ‘That pot made a banging (sound).’ (FEM. ADV.: *dangast*)



- (211) a. Xīr=i    buy-ast    čūd.  
 sun=3SG buzz.M-ADV do.PST  
 ‘The sun buzzed.’ (MASC. ADV.: *buyast*)
- b. Yā        tilifūn=i    bay-ast    čūd.  
 DEM.DIR.F phone=3SG buzz.F-ADV do.PST  
 ‘The telephone buzzed.’ (FEM. ADV.: *bayast*)

Examples of gender-distinguishing onomatopoeic verbs in their infinitive forms are shown in (212). Here, the infinitive verb agrees in gender with the entity whose action is described by the onomatopoeic word, even if this entity is not the grammatical subject of the larger clause. Hence, in (212a), the verb agrees with *yīdā* ‘boy’, a masculine noun, and in (212b), the verb agrees with *yāc*, a feminine noun, regardless of the gender of the subject noun *wuz* ‘I’.

- (212) a. Wuz=um wi        yīdā qoyaz qati    curt-ow        ǰud.  
 I=1SG    DEM.OBL.M boy paper with rustle.INF.M-NMZ hear.PST  
 ‘I heard that boy rustling with the paper.’ (MASC. INF. STEM: *curtow*)
- b. Wuz=um wam        yāc qoyaz qati    cart-ow        ǰud.  
 I=1SG    DEM.OBL.F girl paper with rustle.INF.F-NMZ hear.PST  
 ‘I heard that girl rustling with the paper.’ (MASC. INF. STEM: *cartow*)

When used as imperatives (i.e. as bare present stems), onomatopoeic words index the gender of the entity being addressed. This is shown in (213), where the noun *xīr* ‘sun’ in (213a) is masculine and agrees with the verb *buy-*, and the noun *civīnc* ‘bee’ in (213a) is feminine and agrees with the verb *bay-*.

- (213) a. A    xīr, bās    buy!  
 VOC sun, enough buzz.M.2SG.IMP  
 ‘Hey sun, stop buzzing!’ (MASC. IMPER. *buy*)
- b. A    civīnc, bās    bay!  
 VOC bee,    enough buzz.F.2SG.IMP  
 ‘Hey bee, stop buzzing!’ (FEM. IMPER. *bay*)

Grammatical gender in onomatopoeic words has been mentioned in just a handful of previous works on the Pamir languages. The most extensive treatment of this phenomenon in Shughni and the other languages of the Shughni-Rushani group is that of Karamshoev (1978: 166-190), who indicates that among Iranian languages, and perhaps among Indo-European languages more generally, gender distinction in onomatopoeic verbs and adverbs is

unique to the Shughni-Rushani group. Nonetheless, little is understood about its origins or development, although Karamshoev speculates that given its significant resemblance to the expression of gender in other verbs, it could have well arisen via analogy with gender distinction in canonical unaccusative verbs.

Much remains to be uncovered about gender agreement in onomatopoeic words as they are used presently in the language. First, it is unclear how commonly this type of gender agreement occurs. My consultants have indicated, as does Karamshoev himself, that although gender agreement of this kind holds for a handful of verbs, it is less steadfast in others. For instance, the verb *curtow* ‘rustle’ has a feminine counterpart *cartow* which sounds degraded to speakers, whereas with the verbs listed above gender agreement is obligatory. Moreover, in the case of *curtow*, it is not apparent whether the feminine form was once acceptable and is now falling out of use, or whether it is a new development taking hold in the language. This issue, among others, should be addressed in future investigation.

### 9.3 Recent and ongoing changes in Shughni verb stems

Having seen the historical changes which have led to the current state of affairs in the Shughni verbal system, we now turn to changes which are ongoing and active in the language, as well as those which have taken place in very recent times. These fall into two broad types of changes: (i) borrowings (primarily from Russian) and (ii) leveling of verb-stem paradigms. This section briefly discusses each type of change and then offers a more in-depth look at a specific instance of leveling which has affected perfect stems, namely the leveling of plural perfect stems and the absorption of their function by feminine perfect stems.

#### 9.3.1 Borrowings in the Shughni verb system

In addition to a large number of native verb stems – i.e. stems which have descended directly from its ancestor languages of the ancient Iranian and Proto-Indo-European periods – Shughni possesses many borrowed verbs. Older borrowed verbs typically come from Tajik, or else from Arabic via Tajik, and may be either simplex or complex, and when simplex are most often regular.

Borrowed simplex verbs include *fāmtow* ‘know; understand’ (cf. Tj. фаҳмидан, *fahmīdan*, ultimately from Arabic); *fortow* ‘be desirous to; want’ (cf. Tj. форидаан, *foridan*), and *moltow* ‘rub; massage’ (cf. Tj. молидаан, *molī-*

*dan*). Complex verbs containing a borrowed element include *kuxiř čīdow* ‘try’ (cf. Tj. кӯшиш кардан, *kūšiř kardan*) and *baxiř čīdow* ‘forgive’ (cf. Tj. баҳшиш кардан, *baxšiř kardan* or баҳшидан, *baxšīdan*). Note that in the latter two verbs, the Tajik sound *ř* exhibits its regular correspondence to Shughni *č* when preceded by a vowel, indicating that these words were borrowed relatively long ago and have been in the language long enough to undergo sound change.

More recent borrowings into the Shughni verb system are by and large from Russian. Unlike borrowings from Tajik and Arabic, Russian borrowings in the Shughni verbal system are exclusively complex verbs built on the light verb *čīdow* together with a Russian infinitive. Examples include *poprobovat čīdow* ‘try’ (< Ru. попробовать, *poprobovat*’) and *meřat čīdow* ‘bother; hinder’ (< Ru. мешать, *meřat*’). In some cases, a Russian borrowing has a counterpart construction which is either a native Shughni construction or an older borrowing. Thus, for instance, the Russian borrowing *panimat čīdow* ‘understand’ (Ru. < понимать, *ponimat*’) has the counterpart *fāmtow*, ultimately from Arabic; the Russian borrowing *staratsa čīdow* ‘try; attempt’ (< Ru. стараться, *starat’sya*) has the counterpart *kuxiř čīdow*, an older Tajik borrowing; and the Russian borrowing *ždat čīdow* ‘wait’ (< Ru. ждать, *ždat*’) has the counterpart *pi . . . nīstow*, which is built on a prepositional phrase with *pi* ‘up to’ and is a native Shughni construction.

### 9.3.2 Leveling in verb-stem paradigms

Besides the more salient phenomenon of borrowings, the Shughni verb system is also undergoing changes due to the leveling of verb-stem paradigms. Leveling in Shughni verbs comes in a variety of types and may affect both the consonants and vowels of verb stems. Moreover, leveling may be either internal to a single verb, where one or more of its stems are restructured based on the form of another stem, or it may be external to the verb, where the stems of one verb are restructured on analogy with those of another verb (Dodykhudoeva 1988: 96, 110).

In the internal type, it is most common that the present stem is used as a model for the restructuring of the other stems. This type of leveling has already proceeded fully in some verbs, particularly those whose past and infinitive stems contain a stem-final consonant once found only in present stems. For instance, the past and infinitive stems of the verb *peřctow* contain the reflex of the suffix *\*-sa-*, which is generally found only in present stems. Likewise, the past and infinitive stems of the verb *wīntow* contains a nasal once found only in the present stem. In modern Shughni, leveling of this type has resulted in two parallel past and infinitive stems for a series of verbs, each of

which is acceptable. Examples of such verbs are given in Table 9.19. The irregular, older version of each verb is in the lefthand column, and the regular, leveled version is in the righthand column.

**Table 9.19:** Verbs with parallel irregular and leveled paradigms.

<u>GLOSS</u>	<b>Non-leveled (irregular) paradigm</b>			<b>Leveled (regular) paradigm</b>		
	<u>PRS</u>	<u>PST</u>	<u>INF</u>	<u>PRS</u>	<u>PST</u>	<u>INF</u>
‘milk’	<i>ǫǫdz-</i>	<i>ǫǫyd</i>	<i>ǫǫwd(ow)</i>	<i>ǫǫdz-</i>	<i>ǫǫdzd</i>	<i>ǫǫdzd(ow)</i>
‘irrigate’	<i>vidêdz-</i>	<i>vidǫyd</i>	<i>vidǫwd(ow)</i>	<i>vidêdz-</i>	<i>vidêdzd</i>	<i>vidêdzd(ow)</i>
‘beat’	<i>ǰeb-</i>	<i>ǰǫvd</i>	<i>ǰǫvd(ow)</i>	<i>ǰeb-</i>	<i>ǰept</i>	<i>ǰept(ow)</i>
‘grind’	<i>yǎn-</i>	<i>yǫd</i>	<i>yǫd(ow)</i>	<i>yǎn-</i>	<i>yǎnt</i>	<i>yǎnt(ow)</i>

The external type of leveling, whereby the stems of one verb are restructured based on analogy with those of another verb, is much less common in modern Shughni. It can be seen, however, in certain pairs of non-causative~causative verbs in which the causative stem, but not the non-causative stem, contains the consonant *w*, as in *ǰici-* ‘freeze’ > *ǰicêwdow*. Here, because canonical morphological causative formation of this type is done by substituting the causative vowel *ê* for the present-stem vowel, we would expect instead *ǰicêdow*. The formation with *w* is apparently done on analogy with verbs which have a stem-final *w* in their non-causative present stems, such as *θǎw-* ‘burn (intr.)’ > *θêw-* ‘burn (tr.)’ (Dodykhudoeva 1988: 109).

### 9.3.2.1 Leveling of plural perfect stems

A particularly intriguing instance of leveling in Shughni verb stems, and one which has seemingly come to completion only within the last few decades, concerns the feminine and plural perfect forms of verbs which mark gender in the past tense. In the modern Shughni dialect of Khorugh, a single perfect stem is used to agree with both feminine subjects and plural subjects. This phenomenon is exhibited in examples (214), where the same perfect form *sic* ‘gone.F/PL’ is used with both the singular feminine subject in (214a) and the plural subject in (214b). This form opposes the masculine form, used exclusively with masculine singular subjects, as in (214c).

(214) **Modern Shughni pattern: One perfect form for feminine and plural**a. **Feminine singular perfect: *sic* ‘gone’**

Yā            yāc pi    maktab sic.  
 DEM.DIR.F girl up.to school go.PRF.F  
 ‘That girl has gone to school’

b. **Plural perfect: *sic* ‘gone’**

Wāð            giðā-yen=en pi    maktab sic.  
 DEM.DIR.PL boy-PL=3PL up.to school go.PRF.PL  
 ‘Those boys have gone to school’

c. **Masculine perfect: *suðj* ‘gone’**

Yu            giðā pi    maktab suðj.  
 DEM.DIR.M boy up.to school go.PRF.M  
 ‘That boy has gone to school’

Older publications on the language, however, generally list a third perfect form which is used exclusively with plural subjects and differs in form from both the feminine singular form and masculine singular form. For the verb *sittow* ‘go’, used in the examples in (214), the plural perfect form was *saðj*. Example (215) exhibits the usage of this older form and corresponds to the modern example (214b).

(215) **(Bygone) plural perfect form: *saðj* ‘gone’**

Wāð            giðā-yen=en pi    maktab saðj.  
 DEM.DIR.PL boy-PL=3PL up.to school go.PRF.PL  
 ‘Those boys have gone to school’

(cf. 214b)

In their form, plural perfect forms such as that exhibited in (215) contain the vowel of the feminine past stem but the stem-final consonant(s) of the masculine perfect stem. For instance, the verb *sittow* ‘go’ has feminine past form *sat* and masculine perfect form *suðj*; its plural perfect form is therefore *saðj*. This pattern can be understood from a historical perspective, as the plural perfect form contains the historical suffixes *\*-ta-kā-*, where *\*-kā-* is the plural perfect counterpart to feminine *\*-čī-* and masculine *\*-ka-*. The long *\*ā* in this form created *a*-umlaut position for the stem vowel, just as the feminine participial suffix *\*-tā* created *a*-umlaut position. Hence, the vowels of the feminine past and plural perfect forms are identical. Because the plural perfect suffix historically contained the *\*k*,

as did the masculine singular perfect stem, the stem-final consonants of these two forms are identical. Examples of plural perfect stems, along with their masculine and feminine singular forms, are given in Table 9.20.

**Table 9.20:** Plural perfect forms.

<u>INFINITIVE</u>	<u>GLOSS</u>	<u>MASC.PR</u>	<u>FEM.PR</u>	<u>PL.PR</u>
<i>sittow</i>	go; become	<i>suđj</i>	<i>sic</i>	<b><i>sađj</i></b>
<i>virixtow</i>	break (intr.)	<i>viruxč</i>	<i>virixc</i>	<b><i>viraxč</i></b>
<i>vidow</i>	be	<i>vuđj</i>	<i>vic</i>	<b><i>vađj</i></b>
<i>nīstow</i>	sit	<i>nūš</i>	<i>nīsc</i>	<b><i>nosč</i></b>
<i>ricīstow</i>	flee	<i>ricūsč</i>	<i>ricīsc</i>	<b><i>ricosč</i></b>
<i>tīdow</i>	go; walk	<i>tūyd</i>	<i>tīc</i>	<b><i>toyj</i></b>
<i>naǰjīdow</i>	pass	<i>naǰjīđj</i>	<i>naǰjīsc</i>	<b><i>naǰjāđj</i></b>
<i>naǰfīdow</i>	fall out	<i>naǰiθč</i>	<i>naǰfīc</i>	<b><i>naǰfāđj</i></b>
<i>sifīdow</i>	rise	<i>sifīđj</i>	<i>sifīc</i>	<b><i>sifāđj</i></b>

Independent plural perfect forms were apparently used up until very recently. They are mentioned not only in works on Shughni dating back to the late nineteenth century (e.g. Shaw 1876) and early 20th century (see Zarubin 1960, the data for which was collected in 1914), but also the latter half of the 20th century (Karamshoev 1963a, 1978; Bakhtibekov 1979). However, Dodykhudoeva (1988) points out that discrepancies can be seen in the description of the plural perfect form, and in examples of its usage, even across these works. In particular, all works published in the latter half of the twentieth century indicate, whether explicitly through dictionary entries, or implicitly through linguistic examples, that speakers used the feminine singular perfect form alongside the plural perfect form. However, the results of Dodykhudoeva's own fieldwork from the 1980's, as published in her (1988) monograph on the verbal system, indicate that the plural perfect form had all but fallen out of usage at that time. Indeed, the younger Shughni speakers with whom I have worked have not heard such forms at all.

Recapping briefly, the Shughni verb system is changing during modern times due to two phenomena: borrowing and paradigm leveling. The first phenomenon is much more easily recognized, as Shughni speakers use a wide range of borrowed Russian verbs in everyday life. Paradigm leveling is somewhat more subtle, but certainly no less real. Several examples exist of formerly irregular verbs which are now either completely regular, or else are

used alongside a regular counterpart. The bygone perfect plural form, which has been completely supplanted by the feminine perfect form, provides an example of widespread leveling which has effectively come to completion in the modern language.

This section on recent and ongoing changes is the last of this chapter on verb forms. This chapter and the previous one have dealt in detail with issues in Shughni verb forms and verb-stem paradigms. The following chapter shifts to an investigation into the use of the different verb stems – present, past, perfect, and infinitives. In particular, it examines the way tense, aspect, and mood, as well as evidentiality, are expressed in Shughni.





## Chapter 10

# TAM and Evidentiality

The current state of the modern Shughni verbal system is the result of the complete overhaul of the Old Iranian verbal system, whereby the largely aspect-based system inherited from Proto-Indo-European became the largely tense-based system found in modern Iranian languages (e.g. Dodykhudoeva 1988; Edelman 1975; Haig 2008; Karamshoev 1978; a.o.). This restructuring is at least partially documented through textual attestations of Old and Middle Iranian languages, with many gaps filled in via the reconstruction of unattested forms.

It was seen in Chapters 8 and 9 that modern Shughni present stems are the reflexes of ancient Iranian imperfective stems, past stems are the reflexes of participles in *\*-ta-*, perfect stems of participles in *\*-ka-*, and infinitive stems of verbal nouns in *\*-ti-*. This chapter departs from the historical linguistic approach used in the preceding chapters and explores how the work of expressing nuances in tense, aspect, and mood (TAM), as well as the related concept of evidentiality, is divided among these four verb stems in modern Shughni.

The chapter begins in Section 10.1 where I lay out certain **relevant definitions in the neo-Reichenbachian framework** within which I examine these topics. Section 10.2 looks at the **temporal and aspectual compatibilities** of present (10.2.1), past (10.2.2), and perfect (10.2.3) stems. Section 10.3 then looks at **further aspectual distinctions** realized through the use of infinitive stems in combination with non-verbal elements, especially locative adpositions. And finally, Section 10.4 discusses the **grammatical expression of realis and irrealis mood**, which is a subtle yet important part of Shughni grammar.

## 10.1 Definitions and framework used for TAM

I use working definitions of each category of TAM which are taken primarily from Cover (2015), drawing on additional sources cited there. I take *tense* to be a grammaticalized linguistic form which ‘restricts the temporal location of the time being talked about’. Tense may be past, present, or future. I take *aspect* to be a grammaticalized form which restricts notions of how an eventuality proceeds over time and how it relates to other salient points in the discourse. Basic aspectual distinctions include perfective, imperfective, and prospective. I take *mood* to refer to the expression of speakers’ attitudes towards what is being expressed, including likelihood, desire, necessity, conditionality, and so on. And finally, *evidentiality* refers to the lexical and grammatical means through which speakers convey how they have obtained the information in question.

I employ a neo-Reichenbachian approach to understanding TAM in Shughni (cf. Reichenbach 1947), and I follow a number of authors (especially Klein 1994 and Cover & Tonhauser 2015) in making the distinction between tense and aspect, on the one hand, and *temporal* and *aspectual reference*, on the other. The former terms refer specifically to paradigmatic linguistic forms such as English *-ed*, while the latter two terms refer to the semantic notions they encode, namely how an eventuality is restricted according the time and manner in which it occurs. Thus, we would say that the English suffix *-ed* is a past-tense suffix and that it entails past temporal reference.

This approach allows for the clear separation of meaning and form and for a more accurate description of the semantics a given form is compatible with. Under this approach, I make use of the following time intervals (taken from Cover & Tonhauser 2015: 308-310).<sup>1</sup>

**Utterance Time (UT):** The point in time when a sentence is uttered;

**Evaluation Time (EvT):** The time relative to which a sentence’s temporal reference is evaluated. (EvT is most often the same as UT, but in certain subordinate clauses, as will be seen in Shughni, it may instead be the Eventuality Time of the matrix clause.);

**Eventuality Time (ET):** The time at which the eventuality expressed in an utterance occurs, which may be either a single point in time or a span of time;

---

<sup>1</sup> Note that these labels are distinct from the ones originally used by Reichenbach (1947), who uses S (Speech Time), R (Reference Time), and E (Event Time). These correspond to Utterance Time, Topic Time, and Eventuality Time, respectively.

**Topic Time (TT):** The time being talked about, which may be constrained, for instance, by subordinate clauses and temporal adverbs such as *yesterday*.

The relation between these time points (or intervals) are used to define temporal and aspectual reference. These definitions are given below, and they are then further clarified through an example.

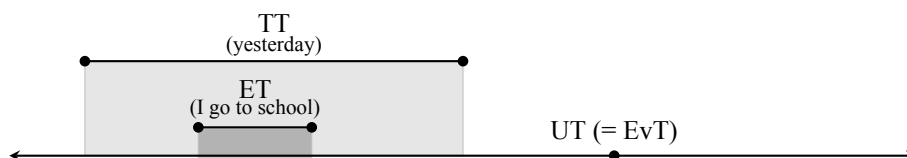
**Temporal reference** is the relation between the Evaluation Time (most often Utterance Time) and the Topic Time. If the Topic Time is temporally located before the Evaluation Time ( $TT < EvT$ ), then the clause has *past* temporal reference. If the Topic Time includes the Evaluation Time ( $EvT \subseteq TT$ ), then the temporal reference is *present*. And if the Topic Time follows the Evaluation Time ( $EvT < TT$ ), then the temporal reference is *future*.

**Aspectual reference** is defined as the relation between Topic Time and Eventuality Time. Two common aspectual references are perfective and imperfective. In the former, the Topic Time includes the Eventuality Time ( $ET \subseteq TT$ ); in other words, the eventuality is entirely within the Topic Time. In imperfective aspect, the Eventuality Time includes the Topic Time ( $TT \subset ET$ ), i.e. the eventuality is not entirely within the Topic Time. A third type of aspectual reference which will be relevant for Shughni is *prospective* aspectual reference, which occurs when the Eventuality Time follows the Topic Time ( $TT < ET$ ). Prospective aspect is encoded, for instance, by English ‘be going to’, as in the sentence ‘I *was going to* eat dinner yesterday’, which has past temporal reference but prospective aspectual reference.

As an illustration of temporal and aspectual relations at work, consider the Shughni sentence in (216).

- (216) Biyor=um      tar maktab sut.  
 yesterday=1SG to school go.PST.M  
 ‘I went to school yesterday.’ ( $TT < UT \rightarrow \text{PST}$ ;  $ET \subset TT \rightarrow \text{PFV}$ )

Here, the Utterance Time (UT) is the time at which the speaker utters the sentence in (216). The Evaluation Time (EvT) in this case is identical to the UT, as the temporal reference for this sentence will be evaluated with respect to the time it is uttered. The Eventuality Time (ET) is the interval in which the speaker went to school, and the Topic Time (TT) is specified by the temporal adverb *biyor* ‘yesterday’. These time intervals are schematized in the timeline in Figure 10.1.

**Figure 10.1:** Relevant time intervals in example (216).

This diagram shows that this sentence has past temporal reference, as the TT, which is the interval represented by the lighter rectangle, is temporally located before the Evaluation Time. It also has perfective aspectual reference. This is because the ET, which is the interval represented by the darker rectangle, is contained by the TT ( $ET \subset TT$ ). That is, the event of the speaker going to the market begins and ends during the Topic Time.

The temporal and aspectual references discussed here, as defined by the respective time intervals, are summarized in Table 10.1.

**Table 10.1:** Summary: Temporal and aspectual references.

Temporal reference	Aspectual reference
PAST: $TT < UT$	PERFECTIVE: $EvT \subseteq TT$
PRESENT: $UT \subseteq TT$	IMPERFECTIVE: $TT \subset ET$
FUTURE: $UT < TT$	PROSPECTIVE: $TT < ET$
TT = Topic Time	ET = Eventuality Time
UT = Utterance Time	EvT = Evaluation Time

In the next section, the framework outlined here is applied to understand the use of Shughni verb stems and their compatibility with other elements that restrict temporal and aspectual reference. Note that this section focuses primarily on the temporal, aspectual, and modal compatibilities of verb stems. Fundamental information on verb stems, including the form of verb stems and basic examples of each type of stem in usage, can be found in Section 8.1.1.

## 10.2 Temporal and aspectual compatibilities of verb stems

As discussed in Section 8.1, Shughni verbs each have four stems – present, past, perfect, and infinitive – each of which exhibit distinct syntactic behavior and are found in distinct, generally non-overlapping grammatical environments. This section looks at the kinds of temporal and aspectual reference compatible with – and entailed by – each of the first three (i.e. finite) stems.

A key concept referred to here is that of *absolute* and *relative tense*. In absolute tense, a clause's temporal reference is always evaluated with respect to the Utterance Time. That is, in absolute tense, the Evaluation Time *is* the Utterance Time. In relative tense, on the other hand, the deictic center relative to which temporal reference is evaluated is not the Utterance Time, but rather some other time. (See Comrie 1985 for an overview of relative vs. absolute tense; for formal approaches to the interpretation of tense in subordinate clauses, see, e.g., Enç 1987; Ogihara 1995; Smirnova 2009, and references therein.) It will be seen below that Shughni present and past stems entail non-past and past tense, respectively, but that this is relative, rather than absolute, and that when EvT is not UT, the time relative to which temporal reference is evaluated is the Eventuality Time of the matrix clause.

In addition to the notion that these stems encode relative tense, it will also be seen that both past and present stems are compatible with perfective and imperfective aspectual reference. Perfect stems, presented in Section 10.2.3, are shown to be used in contexts for which the perfect is commonly used cross-linguistically (for an overview, see Bertrand et al. 2022 and references therein), namely to express experientiality and completed actions which have resulted in states relevant to the Topic Time. Additional uses of perfect stems which spill into the territories of mood and evidentiality will also be mentioned.

### 10.2.1 Present stems

In this subsection, the compatibilities of present stems with respect to temporal reference are first discussed in Section 10.2.1.1. Section 10.2.1.2 then examines the use of present stems with the factual enclitic =*ta*, which is not used with the other three stems and which makes further temporal (i.e. present vs. future) and aspectual (i.e. continuous vs. habitual) distinctions with present stems.

### 10.2.1.1 Fundamental temporal reference of present stems.

Present stems in Shughni are generally compatible with present and future (i.e. non-past) temporal reference and incompatible with past temporal reference. This is shown in the examples in (217). These examples use the temporal adverbs *ik=šič* ‘right now’, *ǰumne* ‘tomorrow’, and *biyor* ‘yesterday’ to constrain the TT of each clause to present, future, and past, respectively. The use of a present stem (bolded) is grammatical with the former two, but not with the latter.

#### (217) Constraints in the temporal reference of present stems

##### a. Present: Grammatical

Wuz ik=šič yi maqola **niviš**-um.  
 I PREC=NOW a article write.PRS-1SG  
 ‘I am writing an article right now.’ (PRS, UT  $\subseteq$  TT)

##### b. Future: Grammatical

Wuz=ta ǰumne yi maqola **niviš**-um.  
 I=FAC tomorrow a article write.PRS-1SG  
 ‘I will write an article tomorrow.’ (FUT, UT < TT)

##### c. Past: Ungrammatical

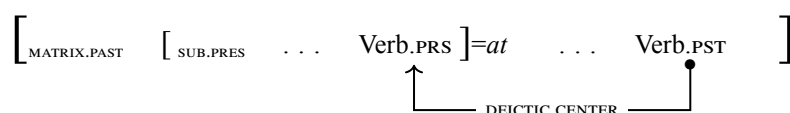
\*Wuz biyor yi maqola **niviš**-um.  
 I yesterday a article write.PRS-1SG  
 Intended: ‘I wrote an article yesterday.’ (PST, TT < UT)

The notion that present stems are incompatible with past temporal reference holds steadfast when they are used in matrix clauses. In certain subordinate clauses, however, this generalization appears to be challenged. In particular, there are at least three types of constructions in which a present stem is used in the subordinate clause, but the temporal reference of the matrix clause, and thus the temporal reference of the entire sentence, is past. In each case, however, the clause containing the present stem has non-past temporal reference with respect to the Eventuality Time of the matrix clause, and hence we can say that Shughni present stems encode relative non-past tense. Each type of construction in which present stems exhibit relative tense is examined below.

The first involves a clause containing a present stem immediately followed another clause containing a past stem, to which it is linked by the conjunction =*at* ‘and; but’. Here, the present stem receives an interpretation corresponding

to the English past progressive, and the eventuality encoded by the past stem in the second clause receives a past interpretation with perfective aspectual reference.<sup>2</sup> This is schematized in (218), and an example is given (219). Note that in the schema in (218), the subordinate (i.e. present-stem clause) clause and the sentence containing both subordinate and matrix clauses are delineated by square brackets, and the temporal reference of each is indicated next to its first square bracket.

(218) **Apparent tense mismatch: Present stem with past interpretation (Schema)**



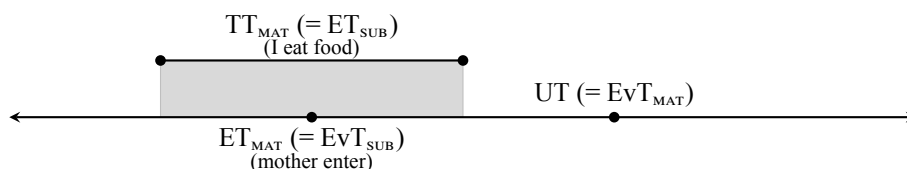
(219) **Apparent tense mismatch: Present stem with past interpretation (Example)**

Wuz(=ta) ʔal xu awqot xām=at mu nān ded<sub>PST</sub>.  
 I(=FAC) still REFL food eat.PRS.1SG=and my mother enter.PST.3SG  
 ‘I was still eating my food when my mother came in.’

In the sentence in (219), which is schematized in Figure 10.2, the TT of the matrix clause, which conveys the eventuality *mother entered*, is specified by the subordinate clause containing the present stem – i.e. the time at which the speaker is eating food. Because the eating took place before the Utterance Time, the matrix clause has past temporal reference ( $TT < UT$ ). The TT of the subordinate clause, which contains a present stem, is specified in matrix clause, i.e. the time at which the speaker’s mother enters. That is,  $TT_{MAT} = ET_{SUB}$ , and  $TT_{SUB} = ET_{MAT}$ .

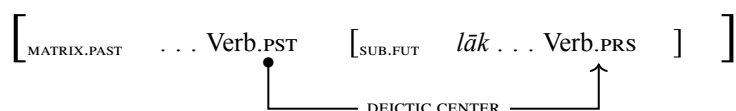
The Topic Time in the subordinate clause, which is denoted by the shaded rectangle, includes its EvT, which is the ET of the matrix clause (i.e. the mother’s entering). Thus, the subordinate clause has present temporal reference relative to the eventuality specified in the matrix clause.

<sup>2</sup>A consultant notes that the use of a past stem in the subordinate clause is equally possible. The use of a present stem in the subordinate clause may be more common in narratives.

**Figure 10.2:** Present stem encoding relative tense.

The second type of construction where present stems exhibit relative non-past temporal reference within a complex clause that has past temporal reference is in purposive clauses built on the connective *lāk* ‘so that’. Here, the matrix verb is not always a past stem, but when it is, the temporal reference of the entire clause is past, and yet the subordinate clause still requires a present stem. This type of construction is schematized in (220), and an example is given in (221). In the example, the eventuality expressed in the subordinate clause with a present stem occurs before the Utterance Time but after the eventuality expressed in the matrix clause. Here, therefore, the verb has future temporal reference relative to the matrix clause.

(220) **Relative tense in purposive clause: Schema**



(221) **Relative tense in purposive clause**

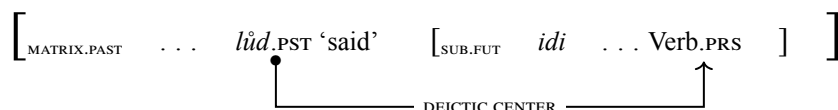
Wam=ām ara mēst tar piro xabar cūd, lāk yā biyor tar sūr yoōd.  
 her=1PL three month in past news do.PST SUBR she yesterday to wedding come.PRS.3SG  
 ‘We informed her three months ago so she would come to the wedding yesterday.’

Lastly, relative tense with present stems can be found in complement clauses such as those with the verb *lūvdow* ‘say’. Here again, the matrix verb is not necessarily a past stem, but when it is, it entails past temporal reference for the matrix clause and for the bigger clause as well. If the Topic Time of the subordinate clause is temporally located after the Eventuality Time of the matrix clause – i.e. when the saying was done – then a present stem is required in the subordinate clause, as schematized in (222). This is the case even when the eventuality expressed in the subordinate clause is temporally located before the Utterance Time, as in the example in (223). Note here the compatibility of the present stem with the temporal adverb *biyor* ‘yesterday’, which stands in contrast to the



example in (217c) above, where the use of this adverb with a present stem in a matrix clause was shown to be illicit.

(222) **Relative tense in complement clause: Schema**



(223) **Relative tense in complement clause**

Mu nān=i      lūd    idi    yā=ta    biyor    tar māš xez    yoðd      (atā yā na-yat).  
 my mother=3SG say.PST COMP she=FAC yesterday to    our location come.PRS.3SG but she NEG-come.PST  
 ‘My mother said that she would come to see us yesterday (but she didn’t come).’

In sum, Shughni present stems are compatible only with non-past temporal reference. For this reason, in matrix clauses, present stems are incompatible with temporal adverbs such as *biyor* ‘yesterday’, which locate the Topic Time before the Utterance Time (= EvT). However, in at least those types of subordinate clauses illustrated above, the temporal reference of present stems is evaluated not with respect to the Utterance Time, but rather with respect to the Eventuality Time of the matrix clause. In these cases, present stems have relative tense and may refer to events which took place before the Utterance Time, a fact which is confirmed by their compatibility with temporal adverbs which constrain the Topic Time to the past. The following subsection examines further temporal and aspectual distinctions with present stems, namely those related to its use with the factual enclitic =*ta*.

**10.2.1.2 Present stems and the factual enclitic =*ta*.**

A key syntactic feature of present stems, and one which sets them apart from past and perfect stems, is their compatibility with the factual enclitic =*ta* (glossed FAC). Because this enclitic is used to realize many of the temporal and aspectual distinctions with present stems, and because it is not used with other types of stems, it is examined here as part of the discussion on the temporal and aspectual compatibilities of present stems.

Although it is mentioned in several publications on the language, pinpointing a single, cohesive semantic function of this enclitic has remained elusive. For instance, Bakhtibekov (1979: 41) characterizes this morpheme as indicating the certainty and permanence of an action; Karamshoev (1988b: 9) adds that it may also be used as an

emotional particle similar to Russian же (*že*) and as an indefinite particle; and Edelman & Dodykhudoeva (2009b: 806-810) opt for the term *factual enclitic* and maintain that it ‘emphasizes the reality or fact of an action’. Here, I adopt the term of the latter authors, but I show below that this term is in many cases not fully satisfying. I note that this morpheme seems to incorporate temporal, aspectual, and modal meanings, and that it remains an important topic for future investigation.

Three general contributions of the enclitic *=ta* can be identified: (i) indication that the predicate is a general fact – that is, a defining feature or habitual action – of the subject; (ii) expression of a (comparatively) high level of certainty that a future action will occur; and (iii) expression of future temporal reference. Somewhat confusingly, however, the contribution of *=ta* in one context may be realized via the lack of *=ta* in another. For instance, *=ta* may show habitual aspect in one sentence where a lack of *=ta* would indicate a continuous reading, whereas in other instances it may show future temporal reference where a lack of *=ta* would indicate a habitual reading. For this reason, when examining the contribution of *=ta*, it is important to compare the semantics of the corresponding sentence without *=ta*. This is precisely the approach I take below.

**Habitual or factual readings of *=ta*** are generally opposed by present continuous readings. This can be clearly seen in sentences which may be responses to the question *čŕ-i=ta* ‘what are you doing?’, which is a common Shughni greeting. Compare, for instance, the sentences in (224). In (224a), *=ta* forces a habitual reading, whereas the lack of *=ta* in (224b) forces a present continuous reading.

(224) **Responses to the question *čŕ-i=ta* ‘what are you doing (/what do you do)?’**

a. **With *=ta*: Habitual reading**

Kor=**ta** kin-um=**at** kino=**ta** čis-um ga.  
work=**FAC** do.PRS-1SG=**and** movie=**FAC** watch.PRS-1SG also  
‘I work and I watch movies. (This is my general routine.)’

b. **Without *=ta*: Present continuous reading**

Kor kin-um=**at** kino čis-um ga.  
work do.PRS-1SG=**and** movie watch.PRS-1SG also  
‘I’m working and watching a movie (at this moment).’

The same distinction is found in other questions with *čŕ* ‘what (action)’. Compare for instance, the examples in (225), where the question with *=ta* in (225a) seeks information about a person’s professional occupation, whereas

the question without *=ta* in (225b) seeks information about what a person is doing at a given moment.<sup>3</sup> For ease of illustration, a context is provided for these examples.

(225) **Questions with and without *=ta***

**Context:** Two colleagues walk into a conference room and see two unfamiliar people who are not from their company. One colleague turns to another and asks about these people.

- a. **With *=ta*: Habitual reading**  
 Wāð=**ta** čīr-en?  
 they=<sub>FAC</sub> what-3<sub>PL</sub>  
 ‘What do they do (professionally)?’
- b. **Without *=ta*: Present continuous reading**  
 Wāð čīr-en?  
 they what-3<sub>PL</sub>  
 ‘What are they doing (right now)?’

The question in (225a) is appropriate when the colleague wishes to know the two strangers’ professional occupation. The question in (225b) is appropriate, for instance, when the two strangers are rummaging through files for no apparent reason.

Note further that the use of *=ta* is obligatory with the adverb *beḡdi* ‘generally; usually’, as in *tu\*(=ta) beḡdi čīr kin-i?* ‘what do you usually do’. The use of *=ta* is also preferred, though perhaps not strictly obligatory, in sentences which describe a property of natural substances or a tradition of a group of people, such as in (226).

(226) ***=ta* used to describe properties or traditions (from (Bakhtibekov 1979: 41))**

- a. **Natural property**  
 ǰāc=**ta** sad girāduš-and worvd.  
 water=<sub>FAC</sub> hundred degrees-LOC boil.PRS.3SG  
 ‘Water boils at 100 degrees.’

---

<sup>3</sup>Interestingly, the habitual vs. continuous distinction shown in (225) does not hold for the questions *čīr-i=ta* ‘what are you (sg.) doing’ or *čīr-et=ta* ‘what are you (pl.) doing’, which inquire about the activities of the addressee(s). The distinction only holds for questions about the activities of a third-person subject (i.e. *čīr(=ta) kiḡt* ‘what is (s)he doing’ or ‘what does (s)he do?’).

Indeed, the questions with a second-singular subject also contain the factual particle *=ta*, but perhaps because they are used so commonly in everyday speech, they do not require an answer with *=ta* indicating a habitual action. A response to the question *čīr-i=ta* ‘what are you (sg.) doing’ is felicitous with or without *=ta*, and with either a habitual or continuous meaning. Outside its use in common questions, nonetheless, the habitual (with *=ta*) vs. present continuous (without *=ta*) distinction occurs with statements and with first- and second-person subjects.

b. **Tradition**

Kicor=**ta** čīd-ard kin-en.  
 water=FAC house-DAT do.PRS-3PL  
 ‘Kicor (a type of oven) is built for houses.’

The factual enclitic =*ta* is used to indicate certainty in many instances where a future plan is discussed, although its use does not necessarily imply full certainty of a future plan, as it is compatible with modal adverbs such as *mumkin* ‘perhaps’. When compared with a future plan expressed without =*ta*, however, a future plan expressed with =*ta* is understood to be more certain. The examples in (227) illustrate.

(227) **Enclitic =*ta* indicating certainty**

**Context:** Maryam is having a bad time at work. She does not enjoy her job and does not agree with the way her bosses run the company. She expresses her intention to leave her current job.

- a. Wuz=**ta** dam mēst tīm.  
 I=FAC DEM.OBL.F month go.PRS.1SG  
 ‘I’m (definitely) going to leave next month.’
- b. Wuz dam mēst tīm.  
 I DEM.OBL.F month go.PRS.1SG  
 ‘I’m going to leave next month.’

Upon hearing the sentence in (227a), which contains =*ta*, a friend would understand that Maryam’s decision to leave is final and that she will almost certainly leave, save some unforeseen circumstance. Upon hearing the sentence in (227b), however, a friend would understand that Maryam is seriously considering the idea of leaving, though she is by no means certain.

**3. A clause containing =*ta* in some instances receives future temporal reference** where a counterpart sentence without =*ta* would receive a habitual reading. Consider, for instance, the example in (228), which will be compared with the example in (229) below.

(228) **Future reading with =*ta***

**Context:** Ahmad has English class on Tuesdays and Thursdays. Today is Tuesday, but he will not go to class because he has an appointment with the dentist. His friend asks whether he will go to the English center today or not.

Nay, wuz#(=**ta**) nur tar anglīsi dars na-sām. Mu-nd zapīs.  
 no I=#(FAC) today to English class NEG-go.PRS.1SG. me-POSS appointment.  
 ‘No, I won’t go to English class today. I have an appointment.’

In the context for the utterance in (228), Ahmad typically goes to English class on Tuesdays, and the fact that he is not going on this particular day is an exception. Hence, the sentence which contains the enclitic =*ta* indicates future reference, rather than habitual reference. Note that the first sentence in (228) is infelicitous without the enclitic =*ta*. Compare this example with that in (229), which shows the opposing habitual reading when =*ta* is not used.

(229) **Habitual reading without =*ta***

**Context:** Ahmad has English class on Tuesdays and Thursdays. Today is Monday, so he will not go to class. His friend is unsure of his schedule and asks whether he will go to the English center today or not.

Nay, wuz#(=**ta**) nur tar anglīsi dars na-sām. Rūziyak-en-and dars nist.  
 no I=#(FAC) today to English class NEG-go.PRS.1SG Monday-PL-LOC class NEG.COP  
 ‘No, I won’t/don’t go to English class today. There is no class on Mondays.’

In (229), the context is such that Ahmad habitually does not attend English class on Mondays. Nonetheless, the use of the factual enclitic here is infelicitous. Note that the distinction between the sentences in (228) and (229) roughly corresponds to the use of *don’t*, which in this context would indicate a habitual action, and *won’t*, which in this context would indicate a one-off action. In the scenario in (228), a response with *don’t* would be infelicitous. In the situation in (229) a response with *won’t* might be used, but it would not alone convey that there is no class on Mondays, as *don’t* would.

To summarize, the factual enclitic =*ta* is used only with present stems and may convey a variety of different semantic nuances. It may call for a factual or habitual reading when describing one’s habits or schedule, or when describing a property of a natural substance or the traditions of a group of people, and it may convey certainty when discussing future plans or that an action will occur in the future. Nonetheless, in some instances =*ta* is used to encode future temporal reference where a corresponding sentence without =*ta* would indicate habitual aspect. Because of these apparent contradictions, the semantic contribution of =*ta* is best understood when sentences with =*ta* are compared with those without. The main semantic contributions of =*ta* are summarized in Table 10.2, which also provides the typical interpretation of a corresponding clause without =*ta*.

**Table 10.2:** Semantic contributions of *=ta*.

	<u>Clause with <i>=ta</i></u>		<u>Clause without <i>=ta</i></u>
	factual; habitual	↔	present continuous
<b>Semantic contribution</b>	certainty	↔	lack of certainty
	future temporal reference	↔	habitual OR present continuous

Clearly, much remains to be understood about the factual enclitic *=ta*, and future research is needed to pinpoint (and unify, if possible) its precise meaning. This morpheme does, however, bear certain similarities with imperfect morphemes of, e.g., West African tenseless languages such as Wolof. For instance, the imperfective morpheme *di* in Wolof is compatible with both habitual and future readings (see Bochnak & Martinović 2018 and references therein). A comparison with morphemes such as these may be a useful starting point for future research on Shughni *=ta*.

### 10.2.2 Past stems

Having seen the compatibilities of present stems regarding temporal and aspectual reference, this subsection turns to past stems. Section 10.2.2.1 looks at temporal reference in past stems, and then Section 10.2.2.2 looks at aspectual reference in past stems.

#### 10.2.2.1 Constraints in the temporal reference of past stems

Past stems are generally used to denote an eventuality with past temporal reference. Past temporal reference is obligatory for eventualities expressed by past stems in matrix clauses. This is shown in the examples in (230). In the first example, the temporal adjective *biyor* ‘yesterday’ restricts the Topic Time to the past and therefore forces past temporal reference. This is compatible with the use of a past stem, and this example is grammatical. In the second example, however, the temporal adjective *ǰumne* tomorrow forces a future temporal reference, which is not compatible with the use of a past stem. This example is ungrammatical.<sup>4</sup>

<sup>4</sup>Forcing a present reading with a past stem is a bit more challenging. The use of a past stem with *ik šič* is allowed if the eventuality is understood to have been completed before the utterance was pronounced. Thus, *wuz=um ik šič ũžin čūd* can be interpreted as ‘I (just) ate

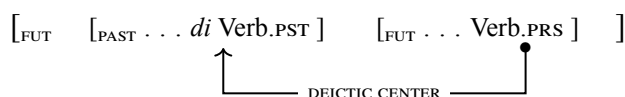
(230) **Temporal compatibility of past stems**a. **Past temporal reference: Grammatical**

Nekrūz=i biyor yi maqola **nivišt**.  
 Nekruz=3SG yesterday a article write.PST  
 ‘Nekruz wrote an article yesterday.’

b. **Future temporal reference: Ungrammatical**

\*Nekrūz=i xūmne yi maqola **nivišt**.  
 Nekruz=3SG tomorrow a article write.PST  
 Intended: ‘Nekruz will write an article tomorrow.’

As with present stems above, there are instances in which past stems, at first glance, seem not to correspond to their fundamental temporal reference. In particular, in some clauses, past stems are used to denote a future occurrence. The use of a past stem for a future event, however, is restricted to a single type of construction where the past stem is in a subordinate clause with the complementizer *di*, and where the matrix clause contains a present stem. This construction is schematized in (231), and examples of this type of construction are given in (232).

(231) **Past stem with (relative) future temporal reference**

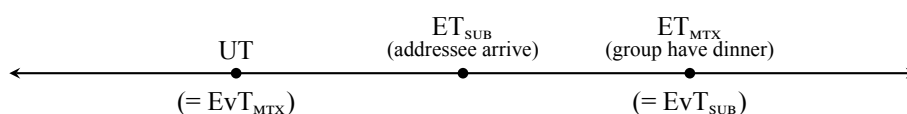
(232) a. Tu=t ar xār di **fīrīpt**, māš=ta qati-yaθ ūžin-ām.  
 you=2SG down.to city SUBR arrive.PST we=FAC with-ADV dinner-1PL  
 ‘When you have arrived to the city, we’ll have dinner together.’

b. Xu awqot=at di **xūd**, qīw mu-rd.  
 REFL food=2SG SUBR eat.PST call.2SG.IMP me-DAT  
 ‘After you have eaten your food, call me.’

Here, the past stem may still be understood to encode relative past tense. In the sentence in (232a), for instance, the Eventuality Time for the matrix clause is specified by the subordinate clause containing the past stem, namely the dinner right now. However, under the definition of present temporal reference above, in which any clause where the Topic Time includes the Utterance Time, we may force a present reading with a sentence like ‘since yesterday and until tomorrow I am working on this article’. In this case, the Shughni equivalent is only grammatical with a present stem and not a past stem. Hence, we can say that past stems in Shughni are incompatible not only with future temporal reference, but also with present temporal reference.

time during which the group has dinner together. This time, which serves as the Evaluation Time for the subordinate clause, occurs after the eventuality denoted in the subordinate clause. Hence, the clause with the past stem has past temporal reference with respect to the matrix clause, which contains a present stem. This is schematized in Figure 10.3

**Figure 10.3:** Past stem encoding relative tense.



We can therefore conclude that past stems, like present stems, entail a specific temporal reference, but that the tense they encode is relative rather than absolute.

### 10.2.2.2 Aspectual compatibilities of past stems

In this subsection it will be seen that although past stems may most often carry a perfective reading, they are also compatible with imperfective aspectual reference, including both habitual and continuous aspects.

Recall that aspectual reference is defined by the relationship between the Eventuality Time and Topic Time of a clause (ET vs. TT). The TT of the first sentence in (216) is denoted by the temporal adverb *biyor* ‘yesterday’. Under perfective readings of past stems, the eventuality denoted by the verb is viewed as having been completed within the Topic Time. Example (233) shows another example of a past stem with perfective aspectual reference.

#### (233) Perfective reading of past stem

Yu=yi biyor **ūžin** **čūd=xu** **yat** pi čīd. (Bād=ām dūs naqli čūd.)  
 he=3SG yesterday dinner do.PST=then come.PST up.to house later=1PL little narrative do.PST  
 ‘Yesterday he ate dinner and then came home. (Then we spoke a bit.)’ (PFV,  $ET \subseteq TT$ )

There are two occurrences in this sentence: the event in which the subject eats dinner and the event in which he comes home. Both events are understood to have occurred within the Topic Time. Hence, the Topic Time contains each eventuality ( $ET \subset TT$ ), and each is understood to have perfective aspectual reference. The second sentence



(in parentheses) is provided to emphasize the perfective nature of the first sentence, as the eventuality described in the second sentence only occurs once those in the first sentence have come to completion. In this sense, the first sentence “advances the narrative”, a feature commonly associated with perfective (but not imperfective) verb forms.

Nonetheless, past stems in Shughni are also compatible with imperfective aspectual reference. The examples below show Shughni past stems used with progressive (234) and habitual aspect (235). Note that in these examples, the use of the (bolded) past stems does not advance the narrative in the same way as the perfective use of a past stem in (233). And moreover, the imperfective reading in these examples is reinforced by the temporal adverb *yal* ‘still’, rather than by any grammatical change in the verb itself.

(234) **Habitual reading of past stem**

Mu bob        parwos   ar māš xez        ca   yat,        yu yal arrüz        tar kor   **sut**.  
 my grandfather last.year to our location SUBR come.PST he still everyday to work go.PST.M  
 ‘When my grandfather came to visit last year, he was still going to work everyday.’ (IPFV, HAB TT  $\subset$  ET)

(235) **Progressive reading of past stem**

Mu oʻxno-yen=en nur   sāraki   ca   yat,        wuz=um yal zavtrak   **čūd**.  
 my friend-PL=3PL today morning SUBR come.PST I=1SG   still breakfast do.PST  
 ‘When my friends arrived this morning, I was still eating breakfast.’ (IPFV, PROG TT  $\subset$  ET)

In each of these sentences, the set of times denoted by the TT is a proper subset of the set of times denoted by the ET. In the first sentence, the TT is specified by the subordinate clause and is the time at which the speaker’s grandfather came to visit. The ET is the interval of time during which the speaker’s grandfather goes to work everyday. Especially if the visit lasts a short time of just a few days, it should be clear that ET contains TT, and because the eventuality denoted by the verb *sut* ‘went’ is in fact made up of multiple subevents, the aspectual reference is habitual (on the distinction between habitual and continuous within the imperfective, see Comrie 1976: 24-30).

In the second sentence, which has a progressive reading, the TT is again indicated by a subordinate clause. Here, it is the point in time at which the speaker’s friends arrived. The ET is the length of time over which the speaker is eating breakfast. Here again, we see that the TT is a proper subset of the ET. And in this case, since the event is viewed as an ongoing process, the aspectual reference is progressive rather than habitual.

Recapping briefly, past stems entail past temporal reference and can thus be viewed as markers of past tense. However, as with present stems, the tense they encode is relative rather than absolute. Moreover, past stems are compatible with both perfective and imperfect aspectual references. The aspectual reading of past stems is not realized by any morphology on the verb, however, but rather by context and the use of adverbs. For instance, in both (234) and (235), the interpretation of the sentence as having imperfective aspect is created through context and the use of the adverb *yal* ‘still’. It will be seen in Section 10.3 that past stems are also compatible with prospective aspect. However, before looking at the prospective aspect, which is built on infinitival constructions, the following subsection turns to perfect stems.

### 10.2.3 Perfect Stems

Three common uses of Shughni perfect stems which are found with the perfect in many other languages are the following: (i) a *resultative* use to denote an event whose completion has resulted in a state which is relevant to the current situation, as in (236); (ii) an *experiential* use to indicate that an eventuality has occurred at some point during history, as shown in (237); and (iii) a *stative* use with unaccusative verbs. An example of the latter is shown in (238), where the perfect stem of the verb *nīstow* ‘sit’ is used to mean ‘staying; residing’. These indicative uses of Shughni perfect stems generally correspond to the most common uses of perfect forms cross-linguistically (e.g., Binnick 1991; Comrie 1976; Inoue 1979; Smith 1997; see Bertrand et al. 2022 for a recent overview). In each of the examples below, a context is provided to further situate the use of the perfect.

#### (236) Perfect: Resultative

**Context:** Maryam and Firuza are at the market and are planning to go eat lunch. They are waiting on Nigora, who is going to meet them at the market before they head to lunch together. Nigora arrives, and Maryam says the following to Firuza.

Nigora **yaθč**.      Toyd=ām!  
 Nigora arrive.PRF. go.PST.PL=1PL  
 ‘Nigora has arrived. Let’s go!’.

#### (237) Perfect: Experientiality

**Context:** Tahmina asks Miskin whether he has been to the Taj Mahal. He answers the following.

Nay, wuz=um ačaθ ar      Indūstūn **na-vuōj**.  
 no, I=1SG at.all down.to India      NEG-be.PRF.M  
 ‘No, I haven’t been to India at all.’

(238) **Perfect: Stativity**

**Context:** Shahlo's parents live in Khorugh but have just arrived in Dushanbe to stay for several days. Her friends ask her where they are staying, and the following is her response.

Wāð=en šič xu qawmiyot=en qati **nīsc**.  
 they=3PL NOW REFL relative-PL with sit.PRF.PL  
 'They are staying with their relatives.'

In addition to canonical indicative uses, perfect stems in Shughni have two further uses: *evidential* and *irrealis*. In their evidential use, the use of a perfect verb stem (vs. a past verb stem) makes inferences about how the speaker has come to know what she is expressing. The evidential use of perfect stems is not uncommon in languages spoken in the same general region as Shughni, particularly those to the west toward Iran and the Caucasus. Comrie (1976: 108) indicates that a perfect-like verb form is used with an 'inferential' function in Persian, Turkish, Bulgarian, Georgian, and Estonian. Tajik also maintains such a use of the perfect. The evidential use of Shughni perfect stems is noted by Edelman & Dodykhudoeva (2009b: 807) and Dodykhudoeva (1988: 8). According to the latter, it has appeared in the language only as a result of contact with Tajik.

Generally, the use of a past stem implies that the information in a given speech act has been witnessed or experienced firsthand by the speaker, whereas the use of a perfect stem, all other things being equal, indicates that the information was obtained indirectly where the speaker does not have direct evidence. This distinction is exhibited in the examples in (239).

(239) a. **Past stem: Direct evidence**

**Context:** Alik and his friend waiting on his father to arrive home. Alik hears a noise outside (direct evidence), and utters the following to his friend.

Mu tāt gumūn-um **yat**.  
 my father think-1SG come.PST  
 'I think my dad has arrived. (I heard a car pull up.)'

b. **Perfect stem: Indirect evidence**

**Context:** Alik and his friend are at the university studying. They are waiting on Alik's father to arrive home before they head home for dinner. Alik's sister calls him to tell him she sees their father's car approaching.

Mu tāt gumūn-um **yaθč**.  
 my father think-1SG come.PERF  
 'I think my dad has arrived. (My sister told me he might have.)'

Importantly, in both examples, the speaker is unsure of whether the event in question (namely the arrival of his father) has actually occurred. The difference, therefore, is not strictly one of modality, as in both cases he acknowledges an equal amount of uncertainty. Instead, the choice of which verb stem to use has to do with the way in which the information was obtained. In this way, Shughni perfect stems are able to convey evidential evidence in the same way as perfect forms are used in Tajik (see Bashir 2006 for an overview of evidentiality in South Asian languages, including Tajik).

Finally, perfect stems are used in both the antecedent and consequent clauses of *counterfactual conditionals* and *desideratives*, where they optionally take the counterfactual suffix *-at* (see the latter part of Section 8.1.1 for a more detailed discussion on this suffix). Examples are given in (240); this topic is examined in Section 10.4 below.

(240) **Perfect stems in counterfactual irrealis**

a. **Counterfactual conditional**

Māš=ām aga tar Amerika az nān **sic(-at)**, māš=ām anglīsi-ti gāp **ōōj(-at)**.  
 we=1PL if in America from mother become.PRF.PL-CF we=1PL English-LOC word hit.PRF-CF  
 ‘If we were born in America, we would speak English.’

b. **Counterfactual desiderative**

Nog yu tar māš dars **vuōj(-at)**.  
 I.wish he in our class be.PRF-CF  
 ‘I wish he were in our class.’

Section 10.4 examines modal uses of Shughni perfect stems in more detail. First, however, the following subsection provides a summary of the temporal and aspectual compatibilities of verb stems, and Section 10.3 looks at further aspectual distinctions built on infinitive stems.

### 10.2.4 Tense and aspect: Interim summary

We have seen in the subsections above that Shughni present and past stems entail non-past and past temporal reference, respectively. The tense encoded in these stems is relative, rather than absolute. This is because in certain complex sentences, a subordinate clause containing a present stem may have past temporal reference with respect to the Utterance Time, but present temporal reference with respect to the eventuality expressed in the matrix

clause. In the same vein, a past stem may have future temporal reference with respect to the Utterance Time, but as long as it has past temporal reference with respect to the eventuality expressed in the matrix clause, against which it is evaluated, the sentence is grammatical. Both present and past stems are compatible with perfective and imperfective aspectual reference, which is often achieved with the help of additional material such as context and adverbs.

Perfect stems, in their turn, are used in many of the same contexts as the perfect in other languages. In particular, Shughni perfect stems often express experientiality, stativity, and eventualities which have ongoing relevance to the Utterance Time. Beyond these standard uses, perfect stems are also used in counterfactual modal constructions and with an evidential function to show that the speaker has only indirect evidence for the information she is relaying.

Before concluding this subsection, a brief note on the names of Shughni verb stems is in order. The choice to use the names *present*, *past*, and *perfect* for the Shughni verb stems examined in this section is largely due to tradition. As far as I am aware, all extant works on the language which address verb stems use these terms (e.g. Bakhtibekov 1979; Karamshoev 1978, 1986; Dodykhudoeva 1988; Edelman & Dodykhudoeva 2009b; Sokolova 1967, 1973; a.o.). These labels align with what we have seen here, given the correspondence of present stems to (relative) present/future temporal reference, the correspondence of past stems to (relative) past temporal reference, and the use of perfect stems in cross-linguistically typical perfect contexts.

However, in recent work on the closely related Pamir language Sarikoli, which also belongs to the Shughni-Rushani group, Palmer (2016) and Kim (2017) opt for a different take on the names of verb stems. Verbs in Sarikoli, like in Shughni, have four fundamental stems which generally correspond to the Shughni stems in etymology and in usage. These authors argue that *imperfective* and *perfective* are more fitting names for the Sarikoli stems which correspond to Shughni present and past stems, respectively. I hope to have shown here that in Shughni, the traditional names for each verb stem are in fact more suitable given the state of affairs in this language. I take no stance on the terminology adopted by the scholars of Sarikoli and simply note that a rigorous comparison of verb stems in Shughni and Sarikoli (along with other Pamir languages) is a compelling line of future research.

### 10.3 Infinitive stems and aspectual reference

So far we have seen how perfective and imperfective (habitual and progressive) aspectual reference is expressed with finite Shughni verb stems. In these cases, aspectual reference is typically realized via a combination of context, subordinate clauses, and adverbs. However, aspectual reference may be specified – or emphasized – in other ways.

This section turns to a discussion of three types of aspectual reference which are overtly expressed through the use of infinitive stems in combination with locative elements. A key feature of all three aspectual distinctions discussed here is that they are compatible with any temporal reference – past, present, or future. Means of expressing progressive aspect via locative suffixes are examined in Section 10.3.1, while the following two sections look at two aspectual distinctions made through the use of pre-verbal elements in combination with a short infinitive form. The inceptive aspect is discussed in Section 10.3.2, and the prospective aspect is examined in Section 10.3.3.

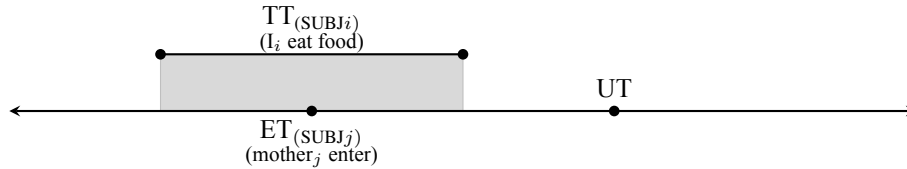
#### 10.3.1 Progressive aspect

The *progressive* aspect, which signifies the ongoing, processual nature of an event, may be indicated with two different locative suffixes: *-and* and *-ti* (cf. Comrie 1976: 24-30 on the distinction between progressive and continuous aspects). These construction types typically serve as a frame for a second eventuality expressed by a finite stem, corresponding to Jespersen's (1931) frame analysis of the progressive.

Although the semantics of progressives with the suffixes *-and* and with *-ti* are similar, they differ syntactically in important ways. First, whereas the suffix *-and* combines with a short infinitive form, such as *xīd* 'eat' in *awqot xīd-and* 'while eating food', the locative suffix *-ti* combines with a long infinitive form, such as *dedow* 'enter' in *dedow-ti* 'while entering'. And second, progressive constructions with *-and* serve as a frame for a second clause with a subject which is necessarily distinct from the one whose progressive action is expressed by the infinitive verb with *-and*. On the other hand, constructions with *-ti* consist of a single clause in which the entity which carries out the eventuality expressed by *-ti* is necessarily the same as the subject of the clause. These distinctions are illustrated by the examples in (241) and (242).

(241) **Locative suffix *-and*: Subjects of each cause are different**

Wuz=um awqot xīd-**and** vud=at mu nān ded.  
 I=1SG food eat.INF-PROG be.PST.M=and my mom enter.PST.3SG  
 ‘I was eating my food when my mother came in.’

(242) **Locative suffix *-ti*: Subjects of each clause are the same**

Yu=yi ded-ow-**ti** mu-rd salūm lūd.  
 he=3SG enter.INF-NMZ-LOC me-DAT hello say.PST  
 ‘He said hello to me as he was entering.’



In both of the sentences above, the Topic Time of the sentence is the interval of time spanned by the action expressed by the infinitival together with the locative element (eating in 241 and entering in 242). And in both cases, the Topic Time includes the Eventuality Time (i.e. when mother enters in 241 and when the subject says hello in 242). Moreover, the eventuality being described by the verb with the locative suffix is a single event, rather than a series of multiple events (i.e. the locative suffix does not encode habitual aspect; see Cover and Tonhauser 2015: 324-325 for a brief illustration of the distinction between the habitual and progressive subtypes of imperfective aspect).

Note further that progressive constructions based on these locative suffixes are compatible with past, present, and future temporal reference. The temporal reference of the clause is governed by the auxiliary verb's stem, and in the case of present stems, by the presence or absence of the factual enclitic *=ta*. This compatibility with any temporal reference is similar of the inceptive and prospective aspects, as discussed below.

### 10.3.2 Inceptive aspect

The inceptive aspect is used to denote the start of an action. In Shughni, the inceptive aspect is expressed periphrastically with the use of either a preverbal element *dar* or *daraw*. The prefixal element *dar* combines only with the short infinitive form of the main verb (i.e. the stem without the infinitival suffix *-ow*), while the particle *daraw* may combine with either the short infinitive form or the full form (i.e. with the suffix *-ow*). The third element of the construction is a light verb, either *sittow* ‘become’ or *ǝǝdow* ‘fall; become’.

Each construction type is schematized below, with the first (*dar*)-type in (243) and the second (*daraw*)-type in (244):

(243) **inceptive aspect with *dar***

. . .      *dar*              VERB.INF              *ǝǝdow/sittow*

(244) **inceptive aspect with *daraw***

. . .      *daraw*              VERB.INF(-ow)              *ǝǝdow/sittow*

Inceptive constructions in Shughni with *dar* and *daraw* are compatible with any temporal reference. The examples in (245) exhibit this compatibility. The inceptive aspect is shown with present temporal reference in (245a), with future temporal reference in (245b), and with past temporal reference in (245c). Note that if a direct object is present, it obligatorily appears in between the inceptive marker *dar* or *daraw* and the infinitive verb. In these examples, the inceptive marker, the verb, and the direct object are all bolded.

(245) **Inceptive aspect and temporal reference**

a. **Present inceptive**

Wāǝ **dar** **garǝā** **niǝempt** sen.  
 they INCEP bread stick.INF become.PRES.3PL  
 ‘They are starting to stick the bread (on the oven walls).’

b. **Future inceptive**

Pi    ǝīd    ca    yad-i,              wuz=ta **dar** **obed** **pêxt**    sām.  
 up.to house SUBR come.PRS-2SG I=FAC INCEP lunch cook.INF become.PRS.1SG  
 ‘When you get home, I’ll start making lunch.’



c. **Past inceptive**

Wað=en **daraw nīwd** sat.  
 they=3PL INCEP cry.INF become.PST.PL  
 ‘They started to cry.’

**10.3.3 Prospective aspect**

The prospective aspect in Shughni expresses the notion that an event is expected to occur shortly after a certain point. Cover and Tonhauser (2015) define this aspectual reference as one in which the Eventuality Time temporally follows the Topic Time (TT < ET). This type of aspect is found, for instance, in the English constructions *be going to* and *be about to*.

In Shughni, prospective aspect is expressed through the combination of a locative preposition *či* followed immediately by a verb’s short infinitive form. The temporal reference of a prospective clause, like the inceptive aspect above, is indicated by an auxiliary verb, which is either the copula (second-position clitics or the verb *vidow*) or else the verb *sittow* ‘become’. This is schematized in (246).

(246) **Prospective clause with locative preposition *či***

. . . **či** VERB.INF (*vidow* / *sittow*)

Examples of the prospective aspect are given in (247), which exhibits the versatility of this aspect with respect to tense. Example (247a) shows an example with a present-tense copula (i.e. second-position clitic), where it is indicated that the subject is in the state of being about to leave. Example (247b) shows a past-tense example, where the use of the copula *vidow* likewise expresses a state, whereas the use of the auxiliary *sittow* ‘become’ expresses an event in which the subject has become on the verge of leaving.

(247) **Prospective aspect with locative *či***

- a. Māš=ām **či** **naǰtīd**.  
 we=1PL PROSP leave.INF  
 ‘We are about to leave.’

- b. Māš=ām **či** **naǰtīd** vad (//sat).  
 we=1PL PROSP leave.INF be.PST.PL (//become.PST.PL)  
 ‘We were about to leave. // We got ready to leave.’

Note, however, that prospective aspectual reference in Shughni, as in English, does not entail that the eventuality in question actually occurs. For instance, any implication that the eventuality occurs can be canceled, as shown in (248), which corresponds to (247b) above.

- (248) Māš=ām **či** **naǰtīd** sat=xu bād yu kasal sut=xu arūd=ām red.  
 we=1PL PROSP leave.INF become.PST.PL=then after he sick become.PST.M=then here=1PL stay.PST  
 ‘We were about to leave, but then he got sick and we stayed here.’

To summarize briefly, this subsection has examined three aspectual references which are commonly expressed in Shughni via the use of locative constructions together with infinitive stems. Progressive aspectual reference may be expressed with locative postpositions *-and* and *-ti*, where the former combines with a short infinitive stem and the latter with a long infinitive form. These constructions differ from one another in that the former provides a frame for the action of another participant, while the latter creates a frame for an action of the same participant.

The inceptive and prospective aspects are expressed by a prepositional or prefixal element in combination with an infinitive stem and an auxiliary verb. The preposition *či* is used for prospective and *dar(aw)* for inceptive. Both of these aspects are compatible with any temporal reference, which is specified by the verb stem of the auxiliary verb used.

The use of adpositional and locative elements across the world’s languages to express nuances in aspectual reference is not uncommon. Shughni follows a cross-linguistic tendency whereby adpositions meaning ‘in’, ‘at’, or ‘on’ are used to mark the progressive aspect (cf. Shughni locative suffixes *-and* and *-ti* for the progressive), while those which denote movement toward the ground are used to denote prospective aspect (cf. the Shughni preposition *či* for the prospective aspect, also used in constructions such as *či ǰac sittow* ‘drown (lit. go toward the water)’). (See Demirdache and Uribe-Etxebarria 2014 and references therein for more on this cross-linguistic tendency.)

## 10.4 Grammatical expressions of mood

Having examined tense and aspect, this section turns to the grammatical expression of mood in Shughni. Grammatical distinctions are made between certain realis and irrealis environments, and in general, the following types of clauses exhibit grammatical characteristics which distinguish them from realis clauses:

- (i) **imperatives**;
- (ii) **subjunctive clauses** (a cover term for purposive and desiderative clauses, as well as clauses with certain modal adverbs);
- (iii) **conditionals**.

These clause types all display at least some behavior in their grammar which both sets them apart from canonical realis clauses and aligns them with other the other types of irrealis environments. They also differ from each other in important ways, however. In what follows, Section 10.4.1 first examines the grammatical phenomena which distinguish realis and irrealis environments in Shughni. Then, Section 10.4.2 discusses the various types of irrealis clauses and how they behave with respect to these phenomena.

### 10.4.1 Grammatical distinctions between realis and irrealis

Shughni does not possess a specific morpheme dedicated to distinguishing realis from irrealis environments. However, irrealis environments in Shughni differ from realis environments via an eclectic mix of four grammatical features: (i) verbal negation; (ii) form of the copula; (iii) use of the suffix *-at* with perfect stems; and (iv) use of bare present stems (for imperative clauses and certain epistemic modals encoding deontic modality). This section presents each of these phenomena in turn.

**Verbal negation in irrealis clauses** is realized via the stressed prefix *mā-* (glossed PROH for *prohibitive*, also used in negative imperatives), whereas in indicative environments verbal negation is realized by the stressed prefix *na-*. This phenomenon was treated in detail in Section 8.2.2, but further examples are provided here in (249). The sentence in (249a) is a realis clause, where the negated verb takes the prefix *na-*. The sentence in (249b) is a desiderative clause marked by the particle *nog*, roughly ‘I hope’. Negation here is with the prefix *mā-*.

(249) **Verbal negation in realis vs. irrealis contexts**a. **Realis verbal negation: *na-***

Davlat=at Bakhmal tar mi maktab **na**-ǰoy-en.  
 Davlat=and Bakhmal at DEM.OBL.M school NEG-study.PRES-3SG  
 ‘Davlat and Bakhmal don’t study at this school.’

b. **Irrealis verbal negation: *mā-***

Nog Davlat=at Bakhmal tar mi maktab **mā**-ǰoy-en.  
 I.hope Davlat=and Bakhmal at DEM.OBL.M school PROH-study.PRS.IRR-3PL  
 ‘I hope Davlat and Bakhmal don’t study at this school.’

**The copula in irrealis environments** is always the verb *vidow*, even in the present tense. This stands in contrast to the present copula of indicative environments, which never contains an overt verb. The examples in (250) illustrate. The first represents an indicative environment where the third-singular copular clitic =*en* is used, while the second represents an irrealis environment where the present stem of the verb *vidow* ‘be’ is used.

(250) **Present copula in realis vs. irrealis contexts**a. **Realis PRS copula: 2nd-position clitics**

Davlat=at Bakhmal=**en** lap boy.  
 Davlat=and Bakhmal=3PL very wealthy  
 ‘Davlat and Bakhmal are very wealthy.’

b. **Irrealis PRS copula: Verb *vidow* ‘be’**

Nog Davlat=at Bakhmal lap-aθ boy **vi-yen**.  
 I.hope Davlat=and Bakhmal very-AUG wealthy be.PRS.IRR-3SG  
 ‘I hope Davlat and Bakhmal are very wealthy.’

**The stressed suffix -at on perfect stems**, which is never found in realis clauses of any kind, is often – albeit optionally – used when these stems occur in irrealis contexts. This is seen in (251). In the indicative clause in (251a), the use of the suffix -at on the perfect stem is ungrammatical. However, in (251b), its use is grammatical, though still optional.

(251) Use of the suffix *-at* on PRF stemsa. **Realis: Suffix *-at* ungrammatical with PRF**

Davlat pi māš čīd **vuðj(\*-at).**

Davlat up.in our house be.PRF.M(\*-at)

‘Davlat was (not) // has (not) been at our house.’

(PST INDICATIVE)

b. **Irrealis: Suffix *-at* optional with PRF**

Nog Davlat ar māš čīd **(mā)-vuðj(-at).**

hope Davlat at our house PROH-be.PST.IRR.M(-at)

‘I hope Davlat was (not) at our house.’

(PST IRREALIS)

Finally, **bare present stems** are used for second-singular imperative forms and in clauses with the epistemic adverb *boyad* ‘must’, whereas present stems must always contain an agreement suffix in realis clauses.

## (252) Present stems in realis vs. imperative contexts

a. **Realis: Present stem + agreement suffix**

Tu tez-aθ **andidz-i.**

you early-AUG get.up-1SG

‘You get up early.’

b. **Irrealis (imperative): Bare present stem**

Tez-aθ **andidz!**

early-AUG get.up.2SG.IMP

‘Get up early!’

These grammatical differences between realis and irrealis clauses are summarized in Table 10.3.

**Table 10.3:** Basic grammatical differences between realis and irrealis.

	<u>REALIS</u>	<u>IRREALIS</u>
Verbal Negation	<b><i>na-</i></b>	<b><i>mā-</i></b>
Present copula	<b>2nd-position clitics</b>	<b>Verb <i>vidow</i></b>
Suffix <i>-at</i> with PRF	<b>illicit</b>	<b>optional</b>
Bare present stem with 2SG	<b>illicit</b>	<b>used with imperatives</b>

### 10.4.2 Types of irrealis clauses

Having seen the grammatical distinctions between realis and irrealis clauses, we now turn to the types of clauses which exhibit irrealis grammatical features. These include **imperatives** (Section 10.4.2.1), **subjunctive clauses**, i.e. desiderative, purposive, and clauses with certain modal adverbs (Section 10.4.2.2, and **conditionals** (Section 10.4.2.3).

#### 10.4.2.1 Imperatives

Imperatives are used to express commands for second-person singular, second-person plural, and first-person plural subjects, although a speaker may direct a command in the second-person singular at herself, in which case the interpretation is that she is telling herself (not) to do something. First- and second-person plural imperatives are formally identical to their indicative counterparts, while second-person singular imperatives call for a distinct verb form, namely a bare (sometimes shortened) present stem.

Examples of affirmative imperatives for each person-number combination are given in (253) with the verb *sittow* ‘go’. Note that in (253a) the second-singular imperative *sa* ‘you (sg.) go!’ is a shortened form of the present stem *sāw-*. The second-plural imperative form *set* ‘you all go!’ in (253b) is likewise shortened, but unlike the second-singular imperative it is identical to its indicative counterpart.

#### (253) Imperatives: Initial examples

##### a. 2<sub>SG</sub> imperative

Tar xu čīd **sa**!  
to REFL house go.2<sub>SG</sub>.IMP  
‘Go to your (sg.) house!’

##### b. 2<sub>PL</sub> imperative

Tar xu čīd **set**!  
to REFL house go.2<sub>PL</sub>.IMP  
‘Go to your (pl.) house!’

##### c. 1<sub>PL</sub> imperative

Tar kino **sāw-ām**!  
to cinema go.IMP-1<sub>PL</sub>  
‘Let’s go to the cinema!’

In addition to the bare present stem used when issuing second-singular commands, imperatives behave distinctly from realis constructions in that (i) a negative imperative always takes the prefix *mā-*, as shown in (254) and (ii) an imperative with the copula always uses the verb *vidow*, even in the present, as shown in (255).

(254) **Negated imperative: prefix *mā-***

Tar kino    ***mā-sāw-ām!***  
 to   cinema PROH-GO.IMP-1PL  
 ‘Let’s not go to the cinema!’

(255) **Present imperative with copula: Verb *vidow***

Bašānd odam   ***vi!***  
 good   person be.2SG.IMP  
 ‘Be a good person!’

As discussed in detail in Section 9.1.5, the phenomenon whereby the final consonant of a present stem is dropped in the second-singular imperative – as in (253a) – occurs with many, but not all Shughni verbs. To recap briefly the discussion on imperative shortening, whether a stem’s final final consonant is dropped depends on both a verb’s form and frequency. In general, verbs with stem shapes CāC drop their final consonant in the imperative (e.g. *vār* > *va!* ‘bring’). However, certain common verbs whose present stem does not have this exact shape may also exhibit shortened imperative forms. The verb *čīdow* ‘do’, which has present stem *kin-* and shortened imperative form *ki*, is unique in that its negative imperative form is further shortened to *mā-k* ‘don’t do’, where *k* < IMP *ki* < PRS *kin-*.

Table 10.4 shows the second-singular imperative forms of six Shughni verbs; note that the first three are part of the group of commonly used verbs which drop their stem-final consonant, whereas the latter three are relatively less common and retain their stem-final consonant. Note also that the verb *sittow* has a truncated second-plural stem (*sāw-et* > *set*), another typical feature of the most common verbs in Shughni.

All imperatives, including second- and first-person plural, as well as negative imperatives, may be used in combination with the imperative form of *vidow* ‘bring’ – *va* (2SG) or *vet* (1PL/2PL) – which appears at the beginning of the imperative clause in the affirmative.<sup>5</sup> The second-plural form *vet* is used with both first-person and second-person plural imperatives. Examples are given in (256).

<sup>5</sup> A similar construction is found in Tajik and other varieties of Persian, in which imperative clauses begin with the verb *āmadan* (Farsi/Dari) or *omadan* (Tajik) ‘come’.

**Table 10.4:** Imperative verb forms.

<u>PRS. STEM</u>	<u>2SG.IMP</u>	<u>NEG.2SG.IMP</u>	<u>GLOSS</u>
<i>kin-</i>	<i>ki</i>	<i>mā-k</i>	do
<i>sāw-</i>	<i>sa</i>	<i>ma-sa</i>	go; become
<i>xār-</i>	<i>xa</i>	<i>mā-xa</i>	eat
<i>andidz-</i>	<i>andidz</i>	<i>mā-andidz</i>	get up
<i>ǰoy-</i>	<i>ǰoy</i>	<i>ma-ǰoy</i>	read; study
<i>birêz-</i>	<i>birêz</i>	<i>ma-birêz</i>	drink

**(256) Imperative constructions with *va* / *vet***

- a. **Va**            tar xu    čīd    **sa!**  
bring.2SG.IMP to   your house go.2SG.IMP  
‘Go to your (sg.) house!’
- b. **Vet**            tar xu    čīd    **set!**  
bring.2PL.IMP to   your house go.2PL.IMP  
‘Go to your (pl.) house!’
- c. **Vet**            tar kino    **mā-sāw-ām**  
bring.2PL.IMP to   cinema PROH-go-1PL  
‘Let’s not go to the cinema!’

Finally, also worthy of note here is the use of indicative verbs in pragmatically imperative contexts, a phenomenon which occurs most commonly in negative commands. The use of an indicative verb in an imperative context does not occur at random; rather, an indicative form of the verb is used in situations where the speaker believes that the addressee intends not to fulfill the command, whereas an imperative form is used in more neutral situations where the speaker has no such belief. To appreciate this difference, compare the examples in (257), which are accompanied by relevant contexts:

**(257) a. Indicative verb as (pragmatically) imperative**

**Context:** A mother has heard her son discuss his intentions of visiting Afghanistan during a trip to a border region. When he leaves for his trip, she says the following to dissuade him.

Tar Awyūnistūn **na-sāw-i!**  
to   Afghanistan NEG-go.PRS-2SG  
‘Don’t go to Afghanistan! (I have reason to believe you might.)’



b. **Corresponding (grammatically) imperative situation**

**Context:** A mother knows her son is going to a region near Afghanistan. She doesn't believe her son as any intention of crossing the border, but wants to voice her concern anyway.

Tar Awyūnistūn **mā-sa!**

to Afghanistan PROH-go.2SG.IMP

'Don't go to Afghanistan! (Though I don't necessarily think you will.)'

**10.4.2.2 Subjunctive clauses**

Shughni makes use of three types of clauses which I give the umbrella label *subjunctive clauses*, as they behave similarly with respect to their grammar and because they correspond semantically to the types of clauses which call for subjunctive verb forms in other Indo-European languages, notably Romance languages. This section examines three major types of subjunctive clauses in Shughni: (i) purposive clauses with the complementizer *lāk*; (ii) desiderative clauses, often with *nog* 'I wish' or the verb *žīwǰdow* 'love'; and (iii) clauses with epistemic adverbs such as *boyad* 'must' and *mumkin* 'perhaps'. These clauses all behave alike in that they require verbs to be negated with the prohibitive marker *mā-*, and they require the verb *vidow*, rather than second-position clitics, to be used as a present copula.

**Purposive clauses** in Shughni are typically marked with a clause-initial complementizer *lāk*. These types of clauses call for a verb in either the present stem, used with present/future temporal reference, or the perfect stem, used with past temporal reference. An example of such a purposive clause is given in (258):

- (258) Māš=ta wev-ard xuŋnūni soz-en lūv-ām, [lāk az xu ziv mā-rinēs-en].  
 we=FAC them-DAT Shughni song-PL sing.PRS-1PL, SUBR from their language PROH-forget.PRS.IRR-3PL  
 'We sing them Shughni songs, so that they don't forget their language.'

- (259) Tu=t boyad wi-rd qīwj, [lāk yu=yi mis fāmč dar borai večer].  
 you=2SG should him-DAT call.PRF SUBR he=3SG also know.PRF about party  
 'You should have called him so he would have known about the party.'

The complementizer *lāk* is also used in subordinate clauses containing an indefinite or imagined subject, as in the sentence 'I am looking for someone who knows Russian', where the subject of the subordinate clause is indefinite and may or may not exist. An example is given in (260); note the use of the present copula *vid*.

(260) **Indefinite subject: Subjunctive**

Uz čorik ʃikar-um, [yu lāk lap biland vi-d].  
 I man search.for-1SG he SUBR very tall be.PRS.IRR-3SG  
 ‘I am looking for a man who is very tall. (Though I don’t have a specific one in mind.)’

Contexts such as the one in (260) exhibit clear grammatical differences from those in which the subject of the subordinate clause is specific. In these cases, the indefinite article *yi* may be used, and the verb (or clitic, in the case of the copula) is in the indicative. An example which stands in contrast to (260) is given in (261). Here, there is no overt form of the verb *vidow* ‘be’, as in the irrealis subordinate clause in (260), and the lack of an overt copular clitic is due to the third-singular subject.

(261) **Specific subject: Indicative – cf. (260)**

Uz (yi) čorik ʃikār-um, [ik=u biland ca].  
 I (a) man search.for-1SG PREC=DEM.DIR.M tall REL  
 ‘I am looking for a tall man. (I know him; he’s around here)’

**Desiderative clauses** in Shughni are generally formed with a clause-initial particle *nog* or *di nay* ‘wish’, roughly ‘I hope’ or ‘I wish’, or with a conjugated form of *ʒīwjdow* ‘love’ or *fortow* ‘want’. In past (counterfactual) desiderative clauses, no overt expression of desire is needed; a verb in its perfect stem (optionally with the suffix *-at*) is enough. Examples of desiderative clauses are given in (262).

(262) **Desideratives with *nog* ‘wish’**

- a. **Nog** kasal mā-sām.  
 I.hope sick PROH-become.PRS.1SG  
 ‘I hope I don’t get sick.’
- b. **Žīwǰ**=um boy=um **vuǰ(-at)**.  
 love.PRF=1SG rich=1SG be.PRF.M-IRR  
 ‘I wish I were rich.’
- c. Wāǰ=en **mā-yaθč(-at)**.  
 they=3PL PROH-come.PRF-CF  
 ‘I wish they hadn’t come.’

In examples like (262a), the indicative may also be used, allowing for a similar nuance as indicative commands (cf. 257). If the indicative is used in (262a) – conveyed through the use of the indicative negation prefix, i.e. *na-sām*

rather than *mā-sām* – then it is implied that the speaker has reason to believe she will become sick. On the other hand, the presence of the irrealis implies a more neutral context.

**Epistemic adverbs** which are often used with grammatically irrealis clauses include *boyad* ‘must; should’ and *mumkin* ‘perhaps’. Examples of these adverbs in irrealis clauses are given in (263) and (264).

- (263) Yā yedard nist; **mumkin** yā ar xu oño xez **vi-d**.  
 she here NEG.COP perhaps she at REFL friend location be.PRS.IRR-3SG  
 ‘She isn’t here; she might be at her friend’s place.’

- (264) Yā=yi mu-rd na-qīwd; yā **boyad** tez-aθ **ǰēvdz-at**.  
 she=3SG me-DAT NEG-call.PST she must early-AUG sleep.PRF-IRR  
 ‘She didn’t call me. She must have gone to sleep early.’

Two phenomena with respect to epistemic adverbs are worthy of note. First, the use of the irrealis is specified in a given lexical item. Thus, whereas both *mumkin* ‘possibly’ and *boyad* require the main verb in their clause to be in the irrealis, the similarly modal word *ǰo* ‘maybe’ does not call for the use of an irrealis verb.

And second, the word *boyad* ‘must; should’ can express both deontic and epistemic modality. Thus, the sentence *yā boyad yamand vid* ‘she must/should be there’ may indicate either: (i) that the subject’s presence in a place is required, e.g. for school or a meeting (deontic), or (ii) that the speaker has made the conclusion that the subject is in a particular location based on information she has gleaned (epistemic).

In a small corner of the grammar, namely with second-person singular subjects, this distinction between epistemic and deontic modality is overtly reflected. Here, when used with deontic modality, the adverb *boyad* requires a verb form equivalent second-singular imperative (i.e. a bare, sometimes shortened present stem). This is shown in (265a). However, when used with epistemic modality, as in (265b), *boyad* calls for a conjugated form of the verb, as in the other subjunctive environments discussed here.

(265) ***Boyad* ‘must’: Distinctions in modality**

a. **Deontic modality: Imperative**

- Tu ɣullā ca sāw-i, tu boyad bašānd odam **vi!**  
 you big SUBR become.PRS-2SG you must good person be.2SG.IRR.PRS.DEO  
 ‘When you grow up, you must be a good person!’

b. **Epistemic modality: Full (non-imperative) verb**

Tu=t fukaθ wev-ard yordām čūd, tu boyad bašānd odam **vi-yi!**  
 you=2SG all them-DAT help do.PST you must good person be.PRS.IRR.EPIS-2SG  
 ‘You helped all of them. You must be a good person!’

### 10.4.2.3 Conditionals

The final irrealis environment to be discussed here is that of conditionals. With respect to their grammar, conditionals in Shughni are in between canonical irrealis clauses and canonical realis clauses. That is, although they correspond to the irrealis mood with respect to their semantics, they display only a subset of the grammatical features typical of irrealis clauses. They behave as irrealis clauses in requiring an overt form of *vidow* as a present-tense copula, as in (266a), and in allowing the perfect forms used in counterfactual conditionals to take the suffix *-at*, as in (266b).

(266) **Conditionals with copula: Irrealis behavior**

- a. Mu nān (aga) tar čīd ca **vi-d**, ku salūm wam-ard lu.  
 my mother (if) at home SUBR be.PRS.IRR-3SG please hello her-DAT say.2SG.IMP  
 ‘If my mom is home, please say hello to her.’ (PRES/FUT CONDITIONAL)
- b. Mu nān (aga) tar čīd ca **vic(-at)**, salūm=um wam-ard **lūvj(-at)**.  
 my mother (if) at home SUBR be.PRF.F-CF hello=1SG her-DAT say.PRF-CF  
 ‘If my mom had been home, I would have said hello to her.’ (PAST CONDITIONAL)

Nonetheless, in all conditional clauses the main verb of both the antecedent and consequent of conditional clauses is negated with the indicative prefix *na-*, rather than the prohibitive *mā-*. This pattern is shown in (267).

(267) **Conditionals with negation: Realis behavior**

- a. Tu (aga) māš qati ar restoran ca **na-yad-i**, tu=ta xu maýdzūnj lāk kin-i.  
 you (if) us with to restaurant SUBR NEG-COME.PRS-2SG you=FAC REFL hungry leave do.PRS-2SG  
 ‘If you don’t come to the restaurant with us, you’ll stay hungry.’
- b. Tu=t (aga) māš qati ar restoran ca yaθč-at, tu=t maýdzūnj **na-vuðj-at**.  
 you=2SG (if) us with to restaurant SUBR come.PRF-2SG you=2SG hungry NEG-be.PRF.M-CF  
 ‘If you had come to the restaurant with us, you wouldn’t be hungry.’

### 10.4.3 Mood: Summary

In summary, irrealis mood in Shughni is expressed through a series of four grammatical features which are not found in indicative contexts. First, second-singular imperative verbs are realized as a bare present stem, whereas in realis contexts second-singular subjects are required to take an agreement suffix. Second, in realis contexts, verbs are negated with the prefix *na-*, while in irrealis contexts they are negated with the prohibitive prefix *mā-*. Third, the present copula in realis contexts is realized by second-position clitics, but in irrealis contexts a conjugated form of the verb *vidow* is used instead. And finally, whereas perfect stems in realis contexts never take the suffix *-at*, but this suffix is optional with irrealis perfect forms.

Environments which call for irrealis grammar include imperatives, subordinate purposive clauses, desideratives, and constructions with certain modal adverbs. Conditionals, for their part, display behavior in between realis and irrealis. On the one hand, they require the verb *vidow* as the present copula and allow perfect stems in counterfactual conditionals to take suffix *-at*. On the other, the main verb of both the antecedent and consequent clauses of conditionals are obligatorily negated with the indicative prefix *na-*. Grammatical distinctions between realis and irrealis moods are summarized in Table 10.5.

**Table 10.5:** Indicative vs. irrealis verb forms.

	Realis	Conditional	Irrealis
Verbal negation:		<i>na-</i>	<i>mā-</i>
Present copula:	2nd-position clitics only	Full verb <i>vidow</i>	
SFX <i>-at</i> with PRF:	Illicit	Optional	

This discussion of grammatical mood in Shughni concludes the chapter on tense, aspect, mood, and evidentiality. This chapter and the previous two chapters have covered topics related primarily to the forms of verb stems, the inflection of verbs, and the way different verb stems, together with other lexical and morphological material, are used to express nuances in temporal and aspectual reference. The following chapter, which draws from much of

the description and analysis in the foregoing chapters, addresses a new set of topics in the Shughni verbal system, namely issues in argument structure and transitivity.

## Chapter 11

# Argument Structure and Transitivity

This chapter, the last dedicated specifically to the Shughni verbal system, explores issues in argument structure and transitivity. The discussion here begins in Section 11.1 on **(passive-like) resultative participial constructions**, which is broken down into an examination of verbal resultative participles in Section 11.1.1, followed by adjectival resultative participles in Section 11.1.2. **Causatives** are then examined in Section 11.2, where morphological causatives are addressed in Section 11.2.1 and then periphrastic causatives in Section 11.2.2.

The discussion on causatives is followed by Section 11.3 on **complex verbs**, which also play an important role in making overt distinctions in transitivity in the language. Shughni possesses a class of **reflexive verbs** which are built on the oblique anaphor *xu* and may be either simplex or complex. These verbs, which to my knowledge have not been addressed in any publication on the language, are the topic of Section 11.4. The language exhibits a separate class of primarily psych verbs which in neutral word order take a non-nominative argument in first position. I refer to these as **oblique-first constructions**, and they are examined in Section 11.5.

Shughni has no dedicated applicative construction, although for a number of intransitive and monotransitive verbs, a benefactive or goal-like argument may be expressed with the addition of the dative suffix *-(a)rd*. However, because there is no special verbal morphology or syntactic construction dedicated for the purpose of benefactives or other types of applicative, I leave the notion of applicative aside for the remainder of this chapter.

## 11.1 Resultative participles and passive-like constructions

Shughni possesses two types of resultative participle, each of which is discussed in turn. **Verbal resultative participles** (with the suffix *-ak*), as well as the passive-like constructions formed with them, are examined first in Section 11.1.1. Then, **adjectival resultative participles** (formed with the perfect stem and suffix *-in*), which may be used in either attributive or predicative function, are examined in Section 11.1.2.

### 11.1.1 Verbal resultative participles: Suffix *-ak*

Verbal resultative participles (VRPs) in Shughni are formed from a verb's perfect stem with the addition of the stressed suffix *-ak* (glossed *v.PTPL*). These participles, like adjectival resultative participles discussed below, are possible with both transitive and unaccusative verbs, but not with unergative verbs. (A verb's ability to be used as a resultative participle was introduced in Section 4.4.2 as one of several diagnostics used to distinguish unaccusative verbs from unergative verbs.) Unlike adjectival resultative participles, however, verbal resultative participles in *-ak* are never inflected for gender; the masculine form is always used even with unaccusative verbs which have gender-distinguishing perfect stems. Examples of VRPs are given in Table 11.1. Note that these participles are possible with both simplex and complex (transitive) verbs; three examples of each type of transitive verb are given in the table. I am unaware of any unaccusative complex verb which has a resultative participle.

**Table 11.1:** Verbal resultative participles from unaccusative and transitive verbs.

	INFINITIVE	GLOSS	PERF. STEM		V. RESULTATIVE PTPL.
			F/PL	M	
UNACCUSATIVE	<i>naǰtīdow</i>	'exit; go out'	<i>naǰtīc</i>	<i>naǰtūyǰ</i>	<i>naǰtūyǰ-ak</i>
	<i>piǰīdow</i>	'ignite'	<i>piǰīc</i>	<i>piǰīdǰ</i>	<i>piǰīdǰ-ak</i>
	<i>θīdow</i>	'burn'	<i>θīc</i>	<i>θuǰǰ</i>	<i>θuǰǰ-ak</i>
	<i>virīǰtow</i>	'break'	<i>virīǰc</i>	<i>viruǰǰ</i>	<i>viruǰǰ-ak</i>
	<i>wizīdow</i>	'go out (of fire, light)'	<i>wizīc</i>	<i>wizuǰǰ</i>	<i>wizuǰǰ-ak</i>
	<i>xāvdow</i>	'go down'		<i>xāvd</i>	<i>xāvdǰ-ak</i>
TRANSITIVE	<i>mīzīdow</i>	'build'		<i>mīzǰ</i>	<i>mīzǰ-ak</i>
	<i>nivīǰtow</i>	'write'		<i>nivīǰc</i>	<i>nivīǰc-ak</i>
	<i>ǰēyīdow</i>	'read'		<i>ǰēyǰ</i>	<i>ǰēyǰ-ak</i>
	<i>bār ǰēdow</i>	'push'		<i>bār ǰoǰǰ</i>	<i>bār ǰoǰǰ-ak</i>
	<i>soz ǰīdow</i>	'fix'		<i>soz ǰūyǰ</i>	<i>soz ǰūyǰ-ak</i>
	<i>xāt ǰēdow</i>	'cross out'		<i>xāt ǰoǰǰ</i>	<i>xāt ǰoǰǰ-ak</i>



VRPs appear with the auxiliary verb *sittow* ‘become’ in a type of construction which invariably has a semantic patient as its subject and expresses a dynamic action. This construction, schematized in (268) below, has previously been labeled the ‘passive’ (e.g. Edelman & Dodykhudoeva 2009b: 800). However, while it bears many similarities with passives, I choose the label “verbal resultative participial construction” (VRPC) because of certain important differences with canonical passives, as discussed below.

(268) **Verbal resultative participial construction: Schema**

- a.  $S_{PAT}$        $V.PRF-ak_{PTPL}$        $sittow_{AUX}$

Examples of VRPCs are given in (269) and (270). The former set of examples contains transitive verbs, while the latter set contains intransitive (unaccusative) verbs.

(269) **Verbal resultative participial constructions with transitives**

- a. Yam=ta      riwêzj-ak      sūd      o?  
 DEM.DIR=FAC fly.CAUS.PRF-V.PTPL become.PRS.3SG PQ  
 ‘Can this (e.g. kite) be flown?’
- b. Yid      xubaθ      žaqč-ak      sut!  
 DEM.DIR PRON.EMPH press.PRF-V.PTPL become.PST.M  
 ‘That (message) got sent all by itself!’

(270) **Verbal resultative participial constructions with unaccusatives**

- a. Yā      řac      nařtüyj-ak      sat.  
 DEM.DIR.F water go.OUT.PRF-V.PTPL become.PST.F  
 ‘The water came out (of some receptacle).’
- b. Wāð      mēx-en=en      zibuðj-ak      sat.  
 DEM.DIR.PL nail-PL=3PL jump.PRF-V.PTPL become.PST.F  
 ‘The nails came out (of the wall).’

In many cases, especially with transitive verbs as in (269), VRPCs resemble passive constructions of other languages such as English and therefore receive a corresponding passive translation. At the same time, however, Shughni VRPCs do not exhibit certain characteristics typically associated with canonical passives. Cross-linguistically, prototypical passive constructions exhibit the following three characteristics (e.g. Givón 1981; Siewierska 1984; Zúñiga and Kittilä 2019; a.o.):

- (i) the agent is either implicit or expressed in an oblique phrase;
- (ii) the patient of the corresponding active clause is *promoted* to grammatical subject;
- (iii) the construction exhibits special morphosyntax (word order and/or morphology) which distinguishes it from its corresponding active construction.

Importantly, these constructions do not appear to involve implicit agents. See first the example in (269b), which indicates that the subject of the verbal participial construction achieved the resulting state ‘all by itself’. Two additional pieces of evidence suggest that agents are generally not implied in Shughni VRPCs. First, my consultants indicate that although an agent may in theory be expressed via the phrase *az X taraf* (lit. ‘from X’s side’), this phrase is almost never used and would only be found in the most formal of speech acts. Moreover, it is clearly modeled off the corresponding Tajik phrase *az taraf-i X*. And second, VRPCs are generally incompatible with purposive clauses (built on the subordinator *lāk*), which do require an implicit or overt agent. This is exhibited in (271):

- (271) Yam mošin soz čūčj-ak sat (?? lāk māš nur ar Dušanbi ti-yām).  
 this car fixed do.PRF-V.PTPL become.PST.F so.that we today to Dushanbe go.PRS-1 PL  
 Intended: ‘This car was fixed so we could go to Dushanbe today.’

It is also not always the case that a patient is “promoted” to grammatical subject. While the patientive subjects of VRPCs built on transitive verbs are objects in corresponding active constructions, VRPCs are also compatible with unaccusative intransitive verbs. In the case of VRPCs with unaccusative verbs, the subject has not been promoted, as it is also the subject of the corresponding “active” construction. Compare, for instance, the examples in (272), in which the subject of the canonical unaccusative construction is the same as its VRPC counterpart:

(272) **Canonical unaccusative construction and VRPC**

a. **Canonical unaccusative construction**

[Yā xāc]<sub>SUBJ</sub> naχtoyd.  
 DEM.DIR.F water come.out.PST.F  
 ‘The water came out.’

b. **Unaccusative VRPC**

[Yā xāc]<sub>SUBJ</sub> naχtūy-ak sat.  
 DEM.DIR.F water go.out.PRF-V.PTPL become.PST.F  
 ‘The water came out (of some receptacle).’

Thus, VRPCs in Shughni are consistent only with one of the three characteristics associated with canonical passives cross-linguistically, namely that they are formed with special morphology (the suffix *-ak* added to a perfect stem). They do not consistently have implied or overtly expressed agents, and they do not consistently promote a patientive argument to grammatical subject.

Furthermore, it can be said that this type of construction is not a particularly salient part of Shughni grammar and is not used in many of the contexts where a passive construction might be used in another language. This fact is likely connected to other aspects of Shughni grammar, notably its status as a topic-prominent, rather than subject-prominent language. In particular, in subject-prominent languages such as English, the information structural notion of topichood is closely correlated with the grammatical notion of subjecthood, and passives are often used to promote a patientive topic to subjecthood. However, in Shughni, where subjecthood is not as closely correlated with topichood, VRPCs are not often used for this purpose. (See e.g. Li and Thompson 1976 and Gundel 1980 on subject- and topic-prominent languages and the connection of the passive to each. See also Section 12.2 on information structure in Shughni.)

Rather than using a VRPC, Shughni speakers are much more likely to indicate topichood via a type of construction which I label the *3pl-passive-like construction*. This construction is similar to a true passive in that no overt agent is expressed, but it differs in that the patient remains in the oblique case. Verbal agreement here is invariably in the third-plural. Examples in (273) illustrate.

(273) **3pl-passive-like construction**

- a. [Xuŋnũni]<sub>TOP-ti=ta</sub> pi Xaray gāp ðen.  
Shughni-LOC=FAC up.in Khorugh word hit.PRS.3PL  
'Shughni is spoken in Khorugh.' (lit. 'In Khorugh (they) speak Shughni.')
- b. [Wi čīd]<sub>TOP=en</sub> sol-i 2020-and mīz].  
DEM.OBL.M house=3PL year-EZ 2020-LOC build.PRF  
'That house was built in 2020.' (lit. '(They) built that house in 2020.')
- c. Yā nāla [wi čorik]<sub>TOP=en</sub> zoxt.  
she say DEM.OBL.M man=3PL take.PST  
'She says that man was arrested.' (lit. 'She says (they) arrested that man.')

Constructions of the type in (273) are common in Shughni. In these examples, the topicalized oblique argument is fronted to clause-initial position but remains in the oblique case. They are clearly not passive *sensu stricto*, as the

patient is not grammatically the subject and there is no distinct morphosyntax. Nonetheless, each is readily translated using the English passive, as they fulfill similar purpose of expressing a transitive event without mentioning an agent.

### 11.1.2 Adjectival passive participle: Suffix *-in*

In addition to verbal resultative participles, which are used with the auxiliary verb *sittow* to express a dynamic action, Shughni has adjectival resultative participles (ARPs), which take the suffix *-in*. ARPs behave syntactically as adjectives and encode a state rather than a process.

One key distinction between the form of VRPs and the form of ARPs has already been mentioned above: ARPs built on unaccusative verbs may inflect to show gender and number agreement with the nouns they modify. This difference is illustrated by the examples in (274) and (275). In both examples, the candidate for agreement is *ǰac* ‘water’, a feminine noun. In (274), however, the VRP contains the masculine perfect form *naǰtūyǰ*, while in (275) the ARP is inflected and appears in its feminine form *naǰtīc*.

(274) **Verbal resultative participle (-ak): No gender or number inflection**

Yā      ǰac      naǰtūyǰ-ak      sat.  
 DEM.DIR.F water go.OUT.PRF-V.PTPL become.PST.F  
 ‘The water came out (of some receptacle).’

(275) **Adjectival Resultative Participle (-in): Gender and number inflection**

Tu=t      wam      naǰtīc-in      ǰac      wīnt      o?  
 you=2SG DEM.OBL.F go.OUT.PRF.F-STAT.PASS water see.PST PQ  
 ‘Did you see the water that came out?’ (lit. ‘did you see the come-out water?’)

Table 11.2 shows further examples of adjectival resultative participles. The first five examples are unaccusative verbs, four of which inflect for gender. The latter group is transitive and do not inflect for gender or number.

**Table 11.2:** Adjectival resultative participles in Shughni.

	INFINITIVE	ADJ. RES. PTPL.		GLOSS
		F/PL	M	
UNACCUSATIVE	<i>naǰǰidow</i>	<i>naǰǰidǰ-in</i>	/ <i>naǰǰic-in</i>	‘crossed; passed’
	<i>tīdow</i>	<i>tīc-in</i>	/ <i>tūyǰ-in</i>	‘gone’
	<i>virix̣tow</i>	<i>virix̣c-in</i>	/ <i>virux̣c-in</i>	‘broken’
	<i>xāvdow</i>	<i>xāvj-in</i>		‘descended’
	<i>zibidow</i>	<i>zibic-in</i>	/ <i>zibuǰ-in</i>	‘jumped’
TRANSITIVE	<i>mīzdow</i>	<i>mīzǰin</i>		‘built’
	<i>riniǰtow</i>	<i>riniǰc-in</i>		‘forgotten’
	<i>xīdow</i>	<i>xūǰ-in</i>		‘eaten’
	<i>tarǰumā čīdow</i>	<i>tarǰumā čūǰ-in</i>		‘translated’
	<i>xāt ōēdow</i>	<i>xāt ōoǰ-in</i>		‘crossed out’

From a syntactic standpoint, ARPs behave as adjectives and are therefore permitted within noun phrases as a modifier of the head noun. Examples are shown in (276).

(276) **Adjectival resultative participles as modifiers within noun phrases**

- a. Wuz=um [xu ǰeyǰ-in (\*ǰeyǰak) kitob-en]<sub>NP</sub> pi universitet yod.  
 I=1SG REFL read.PRF-A.PTPL (\*read.V.PTPL) book-PL to university take.PST  
 ‘I took the books which I read to the university.’
- b. [Yā naw naǰǰic-in (\*naǰǰuyǰak) kino]<sub>NP</sub> fuk-aθ-ard dis xuš yat.  
 DEM.DIR.F new exit.PRF-F-A.PTPL (\*exit.V.PTPL) movie all-AUG-DAT so happy come.PST  
 ‘Everyone really liked that movie that just came out.’

ARPs can also be used predicatively, in which case they appear with the copula (i.e. either second-position clitics alone – in the case of present indicative copular constructions – or else with the verb *vidow*; see Section 8.3). The examples in (277) illustrate.

(277) **Adjectival resultative participles in predicative position**

- a. Yid čīni zinoǰ-in.  
 DEM.DIR bowl wash.PRF-A.PTPL  
 ‘That bowl is (already) washed.’

- b. Wāð      čīd-en=en      ūži      mīzj-in      vad      (\*sat)      o?  
 DEM.DIR.PL house-PL=3PL already build.PRF-A.PTPL be.PST.PL (\*become.PST.PL) PQ  
 ‘Were those houses already built?’

### 11.1.3 Summary: Resultative participles

To summarize briefly, verbal resultative participles in Shughni are built on a verb’s perfect stem (uninflected for gender) together with the suffix *-ak*. These participles are used with the auxiliary verb *sittow* ‘become’ in constructions which express a dynamic action. Adjectival resultative participles, for their part, are also built on a verb’s perfect stem, but are used with the suffix *-in* and are inflected for gender in cases where the (unaccusative) verb in question has gender-distinguishing perfect stems.

There are several aspects of resultative participles which remain open for future investigation. For instance, while resultative participles are possible only with unaccusative and transitive verbs, it is not the case that all unaccusative or transitive verbs may be used as resultative participles. For instance, the unaccusative verbs *firīptow* ‘arrive’ and *yattow* ‘come’ do not have resultative participles, nor do the transitive verbs *tayor čīdow* ‘do; prepare’ or *yēdow* ‘take’. It remains to be seen which criteria, if any, determine whether a given verb can be used as a resultative participle.

Moreover, Shughni also possesses verbal participial constructions which do not have an overt subject. The interpretation of such constructions is similar to the impersonal constructions of German and French (e.g. German *es wurde getanzt*, lit. ‘it was danced’), in which a general (intransitive) event is described (e.g., Perlmutter 1978; Dixon and Aikhenvald 2011; and references within these works). However, whereas impersonal constructions in these languages are built on unergative verbs, such constructions in Shughni are built on unaccusative verbs. Examples of subjectless VRPCs in Shughni are given in (278).

#### (278) Subjectless (unaccusative) VRPCs in Shughni

- a. Azūd      naṣtūyj-ak      na-sūd.  
 from.here go.out.PRF-V.PTPL NEG-become.PRS.3SG  
 ‘One cannot get out of here’ (lit. ‘(it) can’t be exited from here’)
- b. Tarūd      deðj-ak      na-sūd.  
 to.here enter.PRF-V.PTPL NEG-become.PRS.3SG  
 ‘One cannot enter here’ (lit. ‘(it) can’t be entered into here’)

In some instances, it is possible to have a subjectless passive with a transitive verb. An example is given in (279) with the verb *nivištow*. Here, there may or may not be an implied subject such as *arčīzca* ‘something’. This example is felicitous in scenarios where the agent of the action may or may not be understood by both speakers.

(279) **Subjectless (transitive) VRPC in Shughni**

Tu daftar-ti nivišč-ak sut.  
 your notebook-LOC write.PRF-V.PTPL become.PST.M  
 ‘Your notebook got written on.’ (lit. ‘(it) was written on your notebook.’)

Subjectless verbal participial constructions of the kind shown in (278) and (279) are not commonly used in Shughni, but they are nonetheless an intriguing point for future research.

## 11.2 Causative

Canonical causative constructions *increase* the number of core arguments of a verb by one. Shughni makes use of multiple causative-like constructions. For the purposes of the discussion here, causatives in Shughni are divided into morphological causatives (Section 11.2.1) and periphrastic causatives (Section 11.2.2). Section 11.2.3 then discusses the relevant formal and semantic nuances which distinguish each type of causative. Lastly, Section 11.2.4 offers a summary which allows the reader to bypass the details of the bulk of the discussion.

### 11.2.1 Morphological causatives

Although Shughni does not have any fully productive morphological means of deriving causative verbs from non-causative verbs, there are two clearly distinguishable, relatively widespread morphological patterns by which non-causative verbs are related to causative counterparts. These are (i) ablaut, presented in 11.2.1.1 and (ii) suffixation, discussed in Section 11.2.1.2.<sup>1</sup> A final type of morphological causative is discussed in Section 11.2.1.3, namely *labile infinitives*, in which causative~non-causative pairs share an identical infinitive but distinct inflected forms.

---

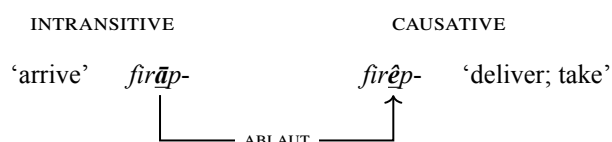
<sup>1</sup> The discussion here does not examine lexical causatives – that is, causative~non-causative pairs which have no clear morphological (or etymological) relation to each other (e.g. Shughni *mīdow* ‘die’, *zīdow* ‘kill’).

The patterns by which morphological causatives are derived in Shughni apply only to intransitive verbs, and with only two exceptions, only to unaccusative verbs (see Section 4.4.2). Hence, if an intransitive verb has a causative counterpart derived by ablaut, suffixation, or stem-vowel alternations in the case of labile infinitives, it is almost certain to display other traits of unaccusativity, including the lack of 3SG past-tense clitic =(y)i and the ability to inflect for gender in its past and/or perfect stems.

### 11.2.1.1 Causative by ablaut

In causatives formed by ablaut, the intransitive stem vowel – typically *a*, *ā*, or *i* – undergoes ablaut to *ê* (before non-nasals) or *e* (before nasals), as schematized in (280). Note that when a verb contains two vowels in its stem, as in this schema, it is the second vowel (i.e. stem vowel) which undergoes ablaut. (See Section on 9.1.1 on monosyllabic and polysyllabic verb stems, and why the second vowel in a polysyllabic stem is always the stem vowel.)

#### (280) Ablaut causative formation



The sentences in (281) exhibit an unaccusative ~ ablaut causative pair in use.

#### (281) Unaccusative ~ causative pair

##### a. Unaccusative: *firāp-* ‘arrive’

Wuz=ta x̣umne pi Maryov **firāp**-um.  
 I=FAC tomorrow to Murghob arrive.PRS-1SG  
 ‘I’ll **arrive** to Murghob tomorrow.’

##### b. Transitive: *firêp-* ‘deliver; take’

Yu=ta mu pi Maryov **firêp**-t.  
 he=FAC me to Murghob deliver.PRS-3SG  
 ‘He will **take** me to Murghob.’

A couple additional notes regarding the form of the ablaut causative are in order. First, in some cases where the stem-final consonant of an intransitive verb is voiceless, the stem-final consonant of its causative counterpart is voiced, as in *niθ-* ‘sit’ > *nêθ-* ‘set’. And second, in at least one causative form, the prefix *ri-*, presumably from



Iranian \**fra-* ‘forth’, appears in addition to ablaut (cf. *nāw-* ‘cry’ > *rinēw* ‘make cry’). This verb is also noteworthy in that, as far as I am aware, it is the only unergative verb (i.e. the only verb compatible with the 3SG past-tense =*i* – see Section 4.4.2) which has a ablaut-causative counterpart. The unergative verb *šānd-* ‘laugh’, which takes a morphological counterpart via suffixation in *šandūn-* ‘cause to laugh’, constitutes the only other exception to the generalization that morphological causatives are formed from unaccusative verbs.

Further examples of intransitive~causative pairs based on ablaut are given in Table 11.3, and examples of two pairs in use are given in (282)–(283).

**Table 11.3:** Intransitive~causative ablaut pairs.

PRES. STEM (INTR.~CAUS.)	GLOSS (INTR.~CAUS.)	INTRANSITIVE INFINITIVE	TRANSITIVE INFINITIVE
<i>fīrāp-</i> ~ <i>fīrēp-</i>	arrive ~ deliver	<i>fīrīptow</i>	<i>fīrēptow</i>
<i>θāw</i> ~ <i>θēw-</i>	burn (INTR.) ~ burn (TR.)	<i>θīdow</i>	<i>θēwdow</i>
<i>nīθ-</i> ~ <i>nēð-</i>	sit ~ set; plant	<i>nīstow</i>	<i>nēðdow</i>
<i>ziban-</i> ~ <i>ziben-</i>	jump ~ make jump	<i>zībidow</i>	<i>zībentow</i>
<i>riwāz-</i> ~ <i>riwēz-</i>	fly away ~ make fly	<i>riwīxtow</i>	<i>riwēzdow</i>
<i>nāw-</i> ~ <i>rinēw-</i>	cry ~ make cry	<i>nīwdow</i>	<i>rinēwdow</i>

(282) **Ablaut causative: *nīθ*~*nēð***

- a. Memūn-en=ta kād-and **nīθ**-en.  
 guest-PL=FAC where-LOC sit.PRS-3PL  
 ‘Where will the guests sit?’ (INTRANSITIVE)
- b. Wāð=ta daraxt kād-and **nēð**-en?  
 they=FAC tree where-LOC sit.CAUS.PRS-3PL  
 ‘Where will they plant that tree?’ (CAUSATIVE)

(283) **Ablaut causative: *θāw*~*θēw***

- a. Yu čīd čīz-ard **θud**?  
 DEM.DIR.M house what-DAT burn.PST.M  
 ‘Why did that house burn?’ (INTRANSITIVE)
- b. Wāð=en čīz-ard wi čīd **θēwd**.  
 they=3PL what-DAT DEM.OBL.M house burn.CAUS.PST  
 ‘Why did they burn that house?’ (CAUSATIVE)

Regarding etymology, intransitive–causative pairs of the ablaut type can be traced back to the productive Iranian causative suffix *\*-aya* (e.g., Sokolova 1967; Dodykhudoeva 1988; Edelman & Dodykhudoeva 2009b). The process by which ablaut came to be the formal means of distinguishing causatives and non-causatives in these pairs is similar to the development of grammatical gender on nouns and adjectives in modern Shughni, as discussed in Sections 7.2 and 7.3. In particular, the causative suffix *\*-aya*, like the bygone grammatical gender suffixes, has itself been lost. However, before it fell out of use, it influenced the stem-vowel through ablaut, and its legacy lives on in causative~non-causative pairs of this kind.

### 11.2.1.2 Causative by suffixation.

The second type of morphological causative to be discussed here is that formed by suffixation. Regular suffixation involves the addition of a suffix *-en/-ûn*, which, as schematized in (284), attaches to present stems and is somewhat more productive than the process of ablaut.<sup>2,3</sup>

#### (284) Suffix-causative formation



The sentences in (285) exhibit an unaccusative ~ suffixal causative pair in use.

<sup>2</sup>The use of the suffixes *-en* and *-ûn* is apparently a newer development in Shughni (Sokolova 1967). This process has perhaps been influenced by similar morphological causatives in Persian. In particular, adding a suffix *-ān* (Tajik *-on*) to present stems is a common and productive way of forming causatives in Persian (e.g. *xar-* ‘eat’ > *xar-ān-* ‘feed; make eat’). The Shughni suffix *-ûn* corresponds to the Tajik causative suffix *-on* through the regular sound alternation where by Shughni *o* becomes *û* before nasals. It is therefore possible that this suffix has been borrowed into Shughni from Tajik. Compare, for instance, Persian *fahm-* ‘know’ > *fahm-ān-* ‘explain’ with Shughni *fām-* ‘know’ > *fām-ûn-* ‘explain’.

<sup>3</sup>The suffixes *-en* and *-ûn* are somewhat more productive than the ablaut process described above, which is virtually unproductive. Nonetheless, even these suffixes are not particularly productive. The present stems to which they can attach seem to be largely preset in the lexicon. They can, however, be used by speakers when playing around with the language and creating new words. For instance, from the intransitive verb *žoz-* ‘run’, a speaker may create the suffixed causative forms *žozen-* or *žozûn-* ‘make run’ while speaking tongue-in-cheek, but a speaker would never create the ablaut form *žêz-*, even when joking (Shahlo Shomansurova, p.c.). This suggests that while speakers may be conscious of the suffixation process by which causatives are formed, they are much less aware of the ablaut process.

## (285) Unaccusative ~ causative pair

a. Unaccusative: *warv-* ‘boil’ (3SG *worvd*)

ǰac=ta tez **worvd**.  
 water=FAC quickly boil.PRS.3SG  
 ‘The water will boil quickly.’

b. Transitive: *warv-en-* ‘(make) boil’

Yu=ta wam ǰac **warv-en-t**.  
 he=FAC DEM.OBL.F water boul.PRS-CAUS-3SG  
 ‘He will boil that water.’

A number of causative verbs are formed by a process which seems to involve suffixation, but which is not as regular as that shown in (284). In these cases, the causative form has a distinguishable suffixal component, but the present stem to which this suffix attaches is altered. For instance, compare *naǰǰis-* ‘pass (INTR.)’ with its causative counterpart *naǰdzimb-* ‘take across’. Here, the original *ǰ* has become *dz*, and the stem-final *-s* is replaced by the cluster *-mb*. A similar phenomenon is observed in the pair *xāfc* ‘descend’ > *xamben-* ‘lower’, where the original cluster *-fc-* is likewise replaced by the cluster *-mb-*. In both cases, and indeed in many causative formed by irregular suffixation, the causative stem ends with *-en*.<sup>4</sup>

Examples of intransitive~causative pairs formed by regular and irregular suffixation are given in Table 11.4. For causative forms which have undergone irregular suffixation, the part of the original stem which remains in the causative form is underlined. For causative forms which undergo regular suffixation, the suffix is bolded. Examples of two of these pairs in use are given in examples (286) and (287).

(286) Causative with irregular suffixation: *naǰǰis*~*naǰdzimb*a. Bād=ām az daryo **naǰǰād**.

later=1PL from river CROSS.PST.PL

‘Then we crossed the river.’

(INTRANSITIVE)

b. Bād=en māš az daryo **naǰdzimpt**.

then=3PL we from river PASS.CAUS.PST

‘Then they took us across the river.’

(CAUSATIVE)

<sup>4</sup>On the etymology of causative verbs formed by irregular suffixation, see Sokolova 1967. For the stem *naǰdzimb-*, for instance, she indicates that *ǰ* has palatalized to *dz* before a front vowel via a regular historical process, and that the addition of *-b* has likely come about via analogy with other present stems containing this cluster, such as *niðemb-* ‘(make) stick’ and *wiremb-* ‘make stand’ (Sokolova 1967: 63).

Table 11.4: Intransitive~causative pairs with suffixation.

Irregular suffixation			
PRES. STEM (INTR.~CAUS.)	GLOSS (INTR.~CAUS.)	INTRANSITIVE INFINITIVE	TRANSITIVE INFINITIVE
<i>naǰfiθ-</i> ~ <i>naǰfen-</i>	fall out ~ pull out	<i>naǰfidow</i>	<i>naǰfentow</i>
<i>xāfc-</i> ~ <i>xāmben-</i>	descend ~ take down	<i>xāvdow</i>	<i>xāmbentow</i>
<i>naǰjīs-</i> ~ <i>naǰdzimb-</i>	pass ~ take across	<i>naǰjīdow</i>	<i>naǰdzimptow</i>
<i>biðāfc-</i> ~ <i>biðemb-</i>	shut ~ shut (TR.)	<i>biðīvdow</i>	<i>biðemptow</i>
<i>wirāfc-</i> ~ <i>wiremb-</i>	stand ~ make stand	<i>wirīvdow</i>	<i>wiremptow</i>

Regular suffixation			
PRES. STEM (INTR.~CAUS.)	GLOSS (INTR.~CAUS.)	INTRANSITIVE INFINITIVE	TRANSITIVE INFINITIVE
<i>warv-</i> ~ <i>warven-</i>	boil ~ make boil	<i>wīrvdow</i>	<i>warventow</i>
<i>fām-</i> ~ <i>fāmūn-</i>	know ~ explain	<i>fāmtow</i>	<i>fāmūntow</i>
<i>šānd-</i> ~ <i>šāndūn-</i>	laugh ~ make laugh	<i>šīntow</i>	<i>šāndūntow</i>
<i>raz-</i> ~ <i>rozen-</i>	flow ~ pour	<i>rixtow</i>	<i>razentow</i>

(287) Causative with regular suffixation: *wārv-*~*warven-*

- a. *ǰac worvd o?*  
 water boil.PRS.3SG PQ  
 ‘Is the water boiling?’ (INTRANSITIVE)
- b. *Tu=t wam ǰac warv-en-t o?*  
 you=2SG DEM.OBL.F water boil-CAUS-PST PQ  
 ‘Did you boil the water?’ (CAUSATIVE)

## 11.2.1.3 Labile infinitives

The final type of morphological causative to be discussed here is that of *labile infinitives*. Under most definitions, labile verbs are those which have a single form which can be either transitive or intransitive, as in English *break* (see Kulikov 2001 and references therein on labile verbs). In Shughni there are a few causative–intransitive pairs of verbs which share an identical infinitive form but differ in their inflected forms. I therefore label this type of verb *labile infinitives*.

Several pairs of verbs in Shughni share an identical infinitive form which may be either transitive (causative) or

intransitive (non-causative). Three of the most common verbs of this type are *virix̣tow* ‘break’; *piðidow* ‘ignite’; and *nêŷdow* ‘stroll; walk’. Despite their shared infinitives, however, most of these pairs of verbs have different present stems for their causative and non-causative variants. Among the three verbs listed above, in two cases, *virix̣tow* and *piðidow*, the non-causative variant also has distinct masculine and feminine/plural forms in its past and/or perfect stems. For *nêŷdow*, the past and perfect stems are identical for the causative and non-causative variants.

The inflection of these verbs is exhibited in Table 11.5. Note that these forms are confirmed both by a native-speaker consultant and by the entries in Karamshoev’s (1988a) dictionary; the citation for the corresponding dictionary entry is given at the bottom of each column.

**Table 11.5:** Labile infinitive verbs.

		<i>virix̣tow</i> ‘break’		<i>piðidow</i> ‘ignite’		<i>nêŷdow</i> ‘walk; stroll’	
		<u>INTR.</u>	<u>TRNS.</u>	<u>INTR.</u>	<u>TRNS.</u>	<u>INTR.</u>	<u>TRNS.</u>
PRS	STEM 3SG	<i>virax̣ŷ-</i> <i>viroŷd</i>	<i>virax̣ŷ-</i> <i>virŷŷd</i>	<i>piðis-</i> <i>piðist</i>	<i>piðin-</i> <i>piðŷnt</i>	<i>noŷŷ-</i> <i>noŷd</i>	<i>nêŷŷ-</i> <i>nêŷd</i>
PST	F/PL M	<i>virax̣ŷt</i> <i>virux̣ŷt</i>	<i>virux̣ŷt</i>	<i>piðid</i>	<i>piðid</i>	<i>nêŷd</i>	<i>nêŷd</i>
PRF	F/PL M	<i>virix̣ŷc</i> <i>virux̣ŷc</i>	<i>virux̣ŷc</i>	<i>piðidŷŷ</i> <i>piðic</i>	<i>piðidŷŷ</i>	<i>nêŷŷŷ</i>	<i>nêŷŷŷ</i>
REFERENCE		(Karamshoev 1988c: 305)		(Karamshoev 1988d: 409)		(Karamshoev 1988d: 304,333)	

Beyond their shared infinitive, labile infinitives are distinct in that their vowel patterns do not resemble those of most derived causatives. In particular, the causative present stems of *virix̣tow* and *piðidow* do not contain the vowel *ê* or *e*, which, as described above, is found in virtually all causative verbs derived through ablaut. Moreover, the intransitive present stem of *nêŷdow* contains the vowel *o*, which is not found in any intransitive present stem which participates in a causative derivation through ablaut. For these reasons, I have decided to treat these verbs separately from what I consider to be *bona fide* ablaut morphological causatives, where the causative vowel is always predictable. A more detailed discussion of this class of verbs can be found in Karamshoev 1978: 150-166.

### 11.2.2 Periphrastic causatives

Whereas morphological causatives constitute a single verb with causative semantics, periphrastic causatives involve additional elements beyond verbal morphology. In this section, I describe three types of periphrastic causatives found in Shughni: (i) **prepositional-phrase causatives**, which are monoclausal and built on a prepositional phrase containing a verbal noun (Section 11.2.2.1); (ii) **coercion causatives**, which are typical biclausal embedding constructions and built on a verb which indicates some type of coercion on the part of the causer (Section 11.2.2.2); and (iii) **locative -*ti* causatives**, which are curious for their lack of causative morphosyntax in the verbal complex.

#### 11.2.2.1 Prepositional-phrase causatives.

This type of causative construction is formed with a prepositional phrase containing a verbal noun and a transitive verb, either *weðdow* ‘throw’ or *vīdow* ‘bring’. There are only three prepositional phrases that may be used in this type of construction: *tar xūðm* ‘to sleep’; *pi nīwd* ‘to crying’; and *pi šūnčak* ‘to laughter’. Despite the fact that they are few in number, these constructions are nonetheless commonly used in everyday Shughni speech. Examples are given in (288)–(290).

(288) **PP causative: *tar xūðm weðdow* ‘make sleep’**

Ar xab=ta yi sūg qatī xu puc-ik **tar xūðm weð-d**.  
 each night=FAC a story with REFL son-DIM to sleep throw-3SG  
 ‘He puts his son to sleep each night with a story.’

(289) **PP causative: *pi nīwd vīdow* ‘make cry’**

Yā kino=yi māš fuk-aθ **pi nīwd vūd**.  
 DEM.DIR.F movie=3SG us all-AUG to crying bring.PST  
 ‘That movie made us all cry.’

(290) **PP causative: *pi šūnčak weðdow* ‘make laugh’**

Wam anekdot-en=ta arwaxt mu **pi šūnčak weð-en**.  
 her joke-PL=FAC every.time me to laughter throw-3PL  
 ‘Her jokes always make me laugh.’

## 11.2.2.2 ‘Coercion’ causatives (in embedding constructions)

The second type of periphrastic causative to be discussed here is that formed on verbs such as *rozi čīdow* ‘convince’, *rimēdow* ‘command’, *majbūr čīdow* ‘force’, and other verbs which indicate coercion, however polite or impolite, of a causer over a causee. This is an embedding construction with the coercion verb in the matrix clause and the verb indicating the caused action in a subordinate clause. The subordinate clause may be either finite or non-finite, with a preference for the former. Finite subordinate clauses typically appear after the ‘command’ verb, while non-finite subordinate clauses typically occur before the ‘command’ verb. Examples of each type are given in (291) and (292); the matrix (coercion) verb is bolded, while the caused action is in square brackets.

(291) Coercion causative: *majbūr čīdow* ‘to force’

- a. Wuz=um wi **majbūr čūd** [awqot xīr-t].  
 I=1sg him force do.PST food eat.PRS-3SG  
 ‘I forced him to eat.’ (FINITE CAUSED ACTION)
- b. Wuz=um wi [awqot xīd] **majbūr čūd**.  
 I=1sg him food eat.INF force do.PST  
 ‘I forced him to eat.’ (NON-FINITE CAUSED ACTION)

(292) Coercion causative: *rimēdow* ‘to order; command’

- a. Ŗo=yi xu askar-en **rimod** [wi čorik vo zen].  
 king=3SG REFL soldier-PL command.PST DEM.OBL.M man again take.PRS.3SG  
 ‘The king commanded his soldiers to get the man again.’ (FINITE CAUSED ACTION)
- b. Ŗo=yi xu askar-en [wi čorik vo zēxt-ow] **rimod**.  
 king=3SG REFL soldier-PL DEM.OBL.M man again take-PURP command.PST  
 ‘The king commanded his soldiers to get the man again.’ (NON-FINITE CAUSED ACTION)

Note that the causing action need not be coercion per se; verbs such as *lūvdow* ‘tell’, *lāk čīdow* ‘let’, and *xoič čīdow* ‘ask’ also participate in this type of construction. A list of verbs which may be used in examples like (291) and (292) are given in Table 11.6. The verb *prosit čīdow* is borrowed from Russian просить (*prosit* ‘to ask’).

**Table 11.6:** Common verbs used in ‘coercion’ causatives.

<u>VERB</u>	<u>GLOSS</u>
<i>majbūr čīdow</i>	‘force’
<i>rimēdow</i>	‘command’
<i>rozi čīdow</i>	‘convince’
<i>lūvdow</i>	‘tell’
<i>lāk čīdow</i>	‘let; allow’
<i>xoiṣ čīdow</i>	‘ask; request’
<i>prosiṣ čīdow</i>	‘ask; request’

### 11.2.2.3 *-ti* causatives

The final type of periphrastic causative construction to be discussed here, and one which does not fall neatly into the morphological–periphrastic dichotomy, is that built on the suffix *-ti*. This suffix is canonically used with locative meaning, such as in *sitūl-ti* ‘on the table’, but as we will see here, it can also have causative and instrumental usages. Importantly, causative constructions with *-ti* are intriguing primarily for their (lack of) clear morphosyntactic indication of causation: the causee is marked with the suffix *-ti*, but the verb indicating the caused action agrees with the causer.

An initial example of this type of construction is shown in (294). The preceding example in (293) is the intransitive, non-causative counterpart of (294) and is provided for comparison. Of importance here is that verbal morphology and agreement are identical in both examples. The verb *žēṣtow* ‘run’ agrees with the subject *yā mollimā* ‘the teacher’ in both (293) and (294), but in the latter, it is not the teacher who is doing the running, but rather the causee *māš* ‘us’, which is marked by the locative suffix *-ti*.

#### (293) Intransitive construction with *žēṣtow* ‘run’

Yā            mollimā=yi    dis            žēṣt  
 DEM.DIR.F teacher.F=3SG so.much run.PST  
 ‘That teacher ran so much.’

#### (294) *ti*-causative construction with *žēṣtow* ‘run’

Yā            mollimā=yi    māš-ti    dis            žēṣt.  
 DEM.DIR.F teacher.F=3SG us-LOC so.much run.PST  
 ‘That teacher made us run so much.’ (lit. ‘The teacher ran on us so much.’)



Without any context, constructions of the kind shown in (294) are ambiguous between a causative and an instrumental reading. Thus, for instance, in (295) below, the subject Davlat may be either forcing Mahram to buy a house (causative reading), or buying a house himself by using Mahram's money (instrumental reading). Similarly, in (296), the subject Shahnoza may be either forcing her friends to write an article (causative reading), or writing an article herself but using her friends knowledge and expertise (instrumental reading).

- (295) Davlat Mahram-**ti** yi čīd    xarīd čūd.  
 Davlat Mahram-**ti** a house buy did  
 ✓ Davlat bought a house using Mahram's money. (INSTRUMENTAL)  
 ✓ Davlat made Mahram buy a house. (CAUSATIVE)
- (296) Shahnoza=yi xu oṣno-yen-**ti** yi maqola nivixt.  
 Shahnoza=3SG REFL friend-PL-**ti** an article wrote  
 ✓ Shahnoza wrote an article using her friends' knowledge or with their help. (INSTRUMENTAL)  
 ✓ Shahnoza made her friends write an article. (CAUSATIVE)

It is worthy of note that the use of the instrumental to mark the causee is not unheard of. This phenomenon is found in a number of languages from a variety of language families, including Hungarian (Kenesei et al. 1998); Kannada (Sridhar 1990); Punjabi (Bhatia 1993); Marathi (Pandharipande 1997); and Wakhi – also a Pamir language (Bashir 2009: 846). Moreover, a locative~instrumental syncretism similar to the one we find in Shughni is also found in Kinyarwanda (Jerro 2017). Ultimately, however, I leave it aside as a feature of Shughni grammar which is particularly ripe for further investigation.

### 11.2.3 Semantics of morphological and periphrastic causatives

Despite the fact that both morphological and periphrastic causatives in Shughni involve causation, there are important differences in their semantics. Morphological causatives, for their part, express more direct causation involving only a single agent (the causer) and a single event subsuming both causing and caused actions. In fact, native speakers may not recognize that the causative verb is regularly derived from the non-causative verb, just as many English speakers may not recognize the causative connection between *fall* and *fell* or *sit* and *set*. By contrast, periphrastic causatives – particularly those of the coercion type – encode a less direct type of causation in which the causee may also be an agent and unity of time and place is not required.

These semantic characteristics can be teased apart through the use of adverbial modification. On the one hand, morphological causatives do not allow causing and caused events to be modified by separate temporal or locative adverbs. Nor do they allow agent-oriented adverbs such as *specialniyaθ* ‘on purpose’ to modify the caused action. Hence, morphological causatives are both monoeventive and monoagentive. On the other hand, periphrastic causatives of the coercion type allow both separate adverbial modification of causing and caused events, as well as agent-oriented adverbs for the caused action. Thus, coercion-type causatives are bieventive and biagentive. (For a formal typology of causative constructions cross-linguistically based on the parameters of number of events and number of agents, see Nie 2020.)

The discrepancy in the agentivity of causee is illustrated in examples (297) and (298). Example (297) shows that in morphological causatives, the use of an instrument is available only for the causer, not the causee. Example (298) shows that in coercion causatives, it is possible to express an instrument used by the causee, indicating the causee is more agentive than in morphological causatives.

(297) **Agentive causee in morphological causative (infelicitous)**

Wuz=um xu ziryot-en lodka qati az daryo naŷdzimpt.

I=1SG REFL child-PL boat with over river pass.CAUS.PST

**Available:** ‘I, using a boat, took my children across the river.’

**Unavailable:** ‘I made my children use a boat to cross the river.’

(298) **Agentive causee in coercion-type causative (felicitous)**

Wuz=um xu ziryot-en maĵbūr čūd [vilka qati awqot xen].

I=1SG REFL child-PL forced do.PST stick with food eat.PRS.3SG

‘I, using a stick, made my children eat food with a fork.’

The discrepancy in eventivity between morphological and coercion causatives is illustrated in examples (299) and (300). In each example, the caused event occurs on Monday. However, only in the coercion-causative example in (300) can the causing event be separated temporally from the caused event.

(299) **Separate causing and caused events in morphological causative (infelicitous)**

Nekruz=i Shahlo ruziyakand ricêθ-ûn-t, (#odixand=i wam-ard tyurma surat-en ca  
 Nekruz=3SG Shahlo Monday flee-CAUS-3SG.PRS Sunday=3SG DEM.OBL.F-DAT prison photo-PL SUBR  
 diviřt).  
 show.PST  
 ‘Nekruz made Shahlo run away on Monday (by showing DEM.OBL.F photos of prison on Sunday).’

(300) **Separate causing and caused events in coercion-type causative (felicitous)**

Nekruz=i Shahlo ruziyakand ricīst-ow majbūr čūd (odixand=i wam-ard tyurma surat-en ca  
 Nekruz=3SG Shahlo Monday flee-PURP force do.PST Sunday=3SG her-DAT prison photo-PL SUBR  
 diviřt).  
 show.PST  
 ‘Nekruz made Shahlo run away on Monday (by showing her photos of prison on Sunday).’

The semantic distinctions between morphological causatives and the coercion-type periphrastic causatives are exhibited in Table 11.7. The notion that Shughni morphological causatives encode more direct causation than their periphrastic causative counterparts is in line with cross-linguistic observations regarding the correspondence between form and semantics (see e.g. Dixon 2000 and Shibatani & Pardeshi 2002). In particular, within a single language, formally compact causative constructions, such as lexical and morphological causatives, tend to signal more direct causation, both temporally and physically, than less compact causative constructions, such as the various types of periphrastic causatives.

**Table 11.7:** Eventivity and agentivity of causative types.

CAUS. TYPE	# OF EVENTS	# OF AGENTS
<b>Morphological</b>	1	1
<b>Periphrastic (coercion)</b>	2	2

**11.2.4 Section summary: Causatives**

Shughni has both morphological causatives and periphrastic causative constructions. Morphological causatives constitute single verbs which are derived from a non-causative, intransitive counterpart. The vast majority of

morphological causatives in Shughni are derived from unaccusative verbs – i.e., in Shughni specifically, those which do not require a third-singular past clitic and generally distinguish gender in their past and perfect stems.

Morphological causatives may be formed either via (i) ablaut of a present-stem vowel (e.g. *firāp-* ‘arrive’ > *firēp-* ‘deliver’) or (ii) adding a suffix *-en* or *-in* to a verb’s present stem (e.g. *warv-* ‘boil (intr.)’ > *warv-en* ‘boil (tr.)’). The former type is the result of an older process, namely umlaut due to the ancient Iranian causative suffix *\*-aya-*. The latter type of formation is generally agreed to be a newer process due in large part to influence from Tajik, which also makes use of causative suffixes. In some cases, an intransitive ~ causative pair may share an infinitive, as in *virixtow* ‘break’. However, in all such verbs, the intransitive and causative members have distinct inflectional patterns. I call these pairs *labile infinitives*.

Periphrastic causative constructions in Shughni can be divided into two primary subtypes: (i) prepositional-phrase causatives of the type *pi nīwd weḍdow* ‘make cry’ (lit. ‘bring to crying’), which consist of a single clause, and (ii) coercion-causatives based on verbs such as *majbūr čīdow* ‘force’, which express the caused action either in an infinitival or finite clausal complement. A third type of causative, in which the causee is marked with the locative suffix *-ti* was also presented. This type of construction is noteworthy for its lack of morphosyntax indicating causation and its compatibility with instrumental readings.

Finally, it was also noted that the different types of causative constructions in Shughni display different semantics, for which morphological causatives and coercion-type causatives were used as a basis for discussion. In comparison with coercion causatives, morphological causatives tend to express more direct causation and less agentivity on the part of the causee. In particular, morphological causatives tend to express only a single event and a single agent (the causer), while coercion-type causatives express two events and two agents. This type of correspondence, in which formal compactness in causative constructions correlates with semantic compactness, has been noted elsewhere cross-linguistically.

## 11.3 Complex verbs

This chapter, and the preceding chapters more generally, have thus far focused primarily on simplex verbs. However, Shughni is like many other Iranian languages in having an extensive system of complex verbs built on a

non-verbal component (NVC), which is uninflected (except for a few gender-distinguishing adjectives used as NVC's) and contributes the core semantic meaning, and a light verb (LV, in the sense of Jespersen 1965: 117), which hosts temporal and aspectual reference, as well as transitivity. (See, e.g., Karimi-Doostan 1999 and Mohammad and Karimi 1992 on the term 'light verb' used in the context of Persian complex verbs.)

Some of the most frequently used verbs in Shughni are complex constructions of this kind. Initial examples are given in (301), where the first example contains the semantically intransitive complex verb *gāp ôêdow* 'speak; converse' (lit. 'hit word'), and the second example contains the transitive complex verb *kud ôêdow* 'scold' (lit. 'give dog').

(301) **Complex verbs: Initial examples**

a. **Intransitive complex verb: *gāp ôêdow* 'speak'**

Māš=ām xu amsoya-yen qati **gāp**<sub>NVC</sub> **ôod**<sub>LV</sub>.  
 we=1PL REFL neighbor-PL with narration do.PST  
 'We talked with our neighbors.'

b. **Transitive complex verb: *kud ôêdow* 'scold'**

Yā mollima=yi wev talaba-yen **kud**<sub>NVC</sub> **ôod**<sub>LV</sub>.  
 DEM.DIR.F teacher=3SG DEM.OBL.PL student-PL dog give.PST  
 'That teacher scolded the students.'

But while complex verbs of the kind in (301) have been studied rather extensively in Persian (e.g., Folli and Harley 2005; Goldberg 2004; Rastorgueva 1953; Shabani-Jadidi 2014; Dabir-Moghaddam 1997; Windfuhr 1979: 113-127), complex verbs in Shughni have been noted only in passing. And although Shughni complex verbs display many similarities to those found in Persian, including many complex verbs borrowed (or calqued) from Tajik, Shughni displays a number of distinctive phenomena in its complex verbs, such as a zero-morpheme allomorph for one light verb in some instances. These issues are examined in detail in what follows.

The following discussion on compound verbs here is divided into three parts: Section 11.3.1 looks at the **structure of compound verbs**, including the types of non-verbal components and light verbs on which they are built; Section 11.3.2 examines the issue of **transitivity in complex verbs** and describes several types of transitivity alternations built on light verbs; and finally Section 11.3.3 looks at unique patterns in the structure of light verbs, namely **light-verb deletion** and **hybrid (simplex~complex) verbs**.

### 11.3.1 Structure of complex verbs

As stated above, complex verbs in Shughni fundamentally consist of a non-verbal component (NVC), which contributes the core semantic meaning of the entire verb, and a light verb (LV). Inflection, including negation and present-tense verbal agreement, are realized on the light verb, with subject agreement in past and perfect stems realized, per usual, as a second-position clitic. This is schematized in (302). The example is given immediately below the schema is the complex verb *kor čīdow*, whose non-verbal component is the noun *kor* ‘work’, and whose light verb is *čīdow* ‘do’.

#### (302) Basic complex verb structure

... NON-VERBAL COMPONENT + (NEG-)LIGHT.VERB(-AGR)  
*kor* ‘work’                      *kin-um* ‘I do’                      → *kor kinum* ‘I work’ (lit. I work do)

There are roughly ten Shughni simplex verbs which may be used as light verbs. All of these may be used as main verbs, although in some instances (e.g. *čīdow* ‘do’), their use as light verbs is much more common. Light verbs are divided into transitive and intransitive based on their usage as simplex verbs. Thus, for instance, the verb *čīdow*, one of the most common transitive light verbs, is used as a simplex to mean ‘do; make; give’ as in *uz=ta tu-rd xac kin-um* ‘I’ll **give/pour** you water’. The verb *sittow*, one of the most commonly used intransitive light verbs, is used as a simplex verb with the meaning ‘go; become’, as in *yā tar maktab sūd* ‘she **goes** to school.’ Note, however, that not all Shughni complex verbs with a transitive light verb are transitive, as in Persian (Platts & Ranking 1911; see also Dabir-Moghaddam 1997: 29). This issue is addressed further in the following section.

There are five light verbs which are by far the most commonly used and which are found in the vast majority of Shughni complex verbs. These are transitive *čīdow* ‘do’, *ōēdow* (pres. stem *ōāō-*) ‘give; hit’, and *weōdow* ‘throw’, and intransitive *sittow* ‘become’ and *ōēdow* (pres. stem *ōi-*) ‘fall; end up’. Less commonly used verbs include *anjīvdow* ‘grab; hold’, as in *kinor anjīvdow* ‘hug’, *lūvdow* ‘say’, as in *soz lūvdow* ‘sing’, *tīždow* ‘pull’, as in *nafas tīždow* ‘breathe’, and *tīdow* ‘go’, as in *pūnd tīdow* ‘walk’. The last two verbs are possibly calques from Tajik *нафас кашидан* *nafas kašīdan* ‘breathe’ and *роҳ рафтан* *roh raftan* ‘walk’, respectively. Shughni light verbs are shown in Table 11.8.

**Table 11.8:** Light verbs in Shughni.

<u>Transitive</u>				<u>Intransitive</u>	
Common:		Uncommon:			
<u>VERB</u>	<u>GLOSS</u>	<u>VERB</u>	<u>GLOSS</u>	<u>VERB</u>	<u>GLOSS</u>
<i>čīdow</i>	do; make	<i>anjīvdow</i>	grab	<i>sittow</i>	become
<i>ōēdow</i> (PRS <i>ōāō-</i> , 3SG <i>ōīd</i> )	give; hit	<i>lūvdow</i>	say	<i>ōēdow</i> (PRS <i>ōi-</i> , 3SG <i>ōed</i> )	fall; end up
<i>weōdow</i>	throw	<i>tīždow</i>	pull	<i>tīdow</i>	‘go; walk’

The non-verbal component in Shughni complex verbs may be an adjective, noun, prepositional phrase, or borrowed verb. Nouns used in Shughni complex verbs, as in Persian, include both native words such as *dārō* ‘pain’, as in *dārō čīdow* ‘hurt’, and borrowed nouns. The latter are often Persian in origin, such as *gāp* ‘word; discussion’ in *gāp ōēdow* ‘speak, or Arabic in origin, as in the verbal noun *arakat* ‘movement’ in *arakat čīdow* ‘move (intr.)’. Certain Russian nouns may also be found in Shughni complex verbs, however, such as *registraciya* ‘registration’ in *registraciya čīdow* ‘register’. The most salient type of borrowed verb used in complex verbs is Russian infinitives, which constitute a productive class in the modern language. Examples of each type of non-verbal component used in Shughni complex verbs are given in Table 11.9.

With respect to the order of elements in complex verbs, neutral word order has the NVC preceding the LV, as in *koncert=ta sar sūd* ‘start’ (lit. the concert will started become). However, the order of the NVC with respect to the LV is variable, and each may be moved to various positions in the sentence in cases where topic or focus call for non-canonical word order. (See section 12.2 on topic, focus, and word order.)

### 11.3.2 Transitivity in complex verbs

Complex verbs, in addition to resultative participles and causatives, as discussed in the preceding subsections, play a large role in making transitive and intransitive distinctions in Shughni. This subsection first examines transitive and intransitive distinctions realized through the alternation of light verbs (Section 11.3.2.1). The discussion then turns in Section 11.3.2.2 to an issue also mentioned above, namely that not all complex verbs built on transitive light verbs are themselves transitive.

**Table 11.9:** Types of non-verbal components in complex verbs.

Adjectives			
ADJECTIVE	GLOSS	COMPOUND VERB	GLOSS
<i>agā</i>	awake	<i>agā čīdow</i>	wake up (tr.)
<i>mot</i>	tired	<i>mot sittow</i>	become tired
<i>werūn</i>	broken	<i>werūn čīdow</i>	break; mess up
Nouns			
NOUN	GLOSS	COMPOUND VERB	GLOSS
<i>arakat</i>	movement	<i>arakat čīdow</i>	move (intr.)
<i>kud</i>	dog	<i>kud ôēdow</i>	scold
<i>nafas</i>	breath	<i>nafas tīždow</i>	breathe
Prepositional phrases			
PREP. PHRASE	GLOSS	COMPOUND VERB	GLOSS
<i>dar qār</i>	into wrath	<i>dar qār ôēdow</i>	get offended
<i>pi nīwd</i>	(up) to crying	<i>pi nīwd weōdow</i>	cause to cry
<i>tar xūōm</i>	to sleep	<i>tar xūōm ôēdow</i>	fall asleep
Russian infinitives			
RUSS. INFINITIVE	GLOSS	COMPOUND VERB	GLOSS
<i>otmenit</i>	cancel	<i>otmenit čīdow</i>	cancel
<i>saxranit</i>	save	<i>saxranit čīdow</i>	save (files)
<i>staratsya</i>	try	<i>staratsya čīdow</i>	try

### 11.3.2.1 Transitive~Intransitive alternations in complex verbs

Within the system of complex verbs in Shughni are several predictable transitivity alternations, wherein there are multiple sets of two parallel complex verbs – one transitive and one intransitive – which are built on consistent sets of light verbs, either *čīdow*<sub>TR</sub>~*sittow*<sub>INTR</sub> or *weōdow*<sub>TR</sub>~*ôēdow*<sub>INTR</sub>. We can identify three clear types of alternations, which are defined by both the light verbs used and the type of NVC.

First, complex verbs with **an adjectival NVC** typically have a transitive~intransitive alternation realized through the verbs *čīdow* ‘do’ (tr.) and *sittow* ‘become’ (intr.), as in *mot čīdow/sittow* ‘make/become tired’. The same



light verbs are also implicated in the second type of alternation, namely with **Russian infinitives NVC's**, whose intransitive counterpart typically has a passive-like interpretation. Thus, for instance, we have the pair *remantīravat čīdow/sittow* ‘remodel / be remodeled’. And third, complex verbs with **prepositional-phrase NVC's** often have transitive~intransitive alternations based on the verbs *weōdow* ‘throw’ (tr.) and *ōēdow* ‘fall; end up’ (intr.). An example is the pair *tar xūōm weōdow/ōēdow* ‘to make sleep / to fall asleep’. Table 11.10 summarizes and provides examples of these transitivity alternations.

**Table 11.10:** Transitivity alternations in complex verbs.

<i>čīdow</i> <sub>TR</sub> ~ <i>sittow</i> <sub>INTR</sub> (Adjectives)			<i>čīdow</i> <sub>TR</sub> ~ <i>sittow</i> <sub>INTR</sub> (Russian Infinitives)			<i>weōdow</i> <sub>TR</sub> ~ <i>ōēdow</i> <sub>INTR</sub> (Prepositional Phrases)		
	VERB	GLOSS		VERB	GLOSS		VERB	GLOSS
TR	<i>agā čīdow</i>	wake up (tr.)		<i>dobavit čīdow</i>	add		<i>ba yoō weōdow</i>	remember
INTR	<i>agā sittow</i>	wake up (intr.)		<i>dobavit sittow</i>	get added		<i>ba yoō ōēdow</i>	remind
TR	<i>qoq čīdow</i>	dry (tr.)		<i>otmenīt čīdow</i>	cancel		<i>dar qār weōdow</i>	offend
INTR	<i>qoq sittow</i>	dry (intr.)		<i>otmenīt sittow</i>	get canceled		<i>dar qār ōēdow</i>	get offended
TR	<i>pur čīdow</i>	fill (tr.)		<i>saxranīt čīdow</i>	save		<i>pi šūnčak weōdow</i>	make laugh
INTR	<i>pur sittow</i>	fill (intr.)		<i>saxranīt sittow</i>	get saved		<i>pi šūnčak ōēdow</i>	start laughing

Alternations of this kind do not exist for transitive complex verbs with nominal NVC's. However, these may have a passive-like counterpart with a verbal resultative participle in *-ak*. For instance, the transitive verb *bispār ōēdow* ‘kick (of an animal)’ does not have the counterpart *bispār sittow* (intended: ‘be kicked’), but does have the resultative form *bispār ōōđjak (sittow)* ‘be kicked’.

### 11.3.2.2 Mismatches in light-verb and complex-verb transitivity

With respect to transitivity, the relationship between light verbs and the complex verbs they form is not always one to one. On the one hand, virtually all complex verbs with the intransitive light verbs *ōēdow* (PRS *ōi-*) ‘fall; end up’ and *sittow* ‘become’ are intransitive (or to be precise, *unaccusative*, as they do not require a third-singular clitic =*i* in their past tense; see Section 4.4.2), and virtually all transitive verbs built on a prepositional phrase and the light verb *weōdow* ‘throw’ are transitive.

On the other hand, complex verbs built on *čīdow* ‘do’ and *ōēdow* (PRS *ōāō-*) ‘give’ may be either transitive or intransitive. For instance, the complex verbs *ōēd weōdow* ‘pit (people) against one another’ (lit. throw fight) and *islo čīdow* ‘correct’ (lit. do correction) are transitive, as shown in (303), while the complex verbs *arakat čīdow* ‘move (intr.)’ (lit. do movement) and *gāp ōēdow* ‘speak’ (lit. hit discussion) are intransitive, as shown in (304).

(303) **Transitive complex verbs with transitive LV’s**

a. *ōēd weōdow* ‘pit (people) against one another’

[Wev oŋno-yen]<sub>SUBJ</sub>=en [wev]<sub>DO</sub> **ōēd weōd**.  
 their friend-PL=3PL them fight throw.PST  
 ‘Their friends pitted them against one another.’

b. *islo čīdow* ‘correct’

Wuz=um wi-rd anglīsi-ti nivišt=xu [yu]<sub>SUBJ</sub>=yi [mu xato-yen]<sub>DO</sub> **islo čūd**.  
 I=1SG him-DAT English-LOC write.PST=and he=3SG my mistake-PL correction do.PST  
 ‘I wrote him in English, and he corrected my mistakes.’

(304) **Intransitive complex verbs with transitive LV’s**

a. *arakat čīdow* ‘to move’

[Māš čimadan-en]<sub>SUBJ</sub>=en pūnd-ti lap **arakat čūd**.  
 our suitcase-PL=3PL road-LOC much movement do.PST  
 ‘Our suitcases moved around a lot on the road’

b. *gāp ōēdow* ‘speak’

[Sohib-i xūna]=yi tu qati **gāp ōōd** o?  
 owner-EZ house=3SG you with word hit.PST PQ  
 ‘Did the landlord speak with you?’

Note that although the intransitive complex verbs in (304) cannot take a direct object, they behave as unergative in taking the third-singular past-tense clitic =(y)i, which is restricted to use with transitive and unergative verbs (cf. 304b, which has a third-singular subject). This is the case for the vast majority of intransitive complex verbs built on transitive light verbs. Curiously, however, a small group of complex verbs with the light verb *čīdow* behave as unaccusative; they neither take a direct object nor allow the third-person past-tense clitic. The common semantic thread between these verbs is a non-agentive (almost always inanimate) subject. The majority of these verbs denote

natural phenomena, such as the rising of the sun or the calming of rain. Examples are given in (305), and a list of these verbs is provided in 11.11.

(305) **Unaccusative complex verbs with *čīdow* ‘do’**

a. ***pal čīdow* ‘rise (of the sun)’**

Xīr(=\*i) ūži pal čūŷj.  
 sun=(\*3SG) already shine do.PRF  
 ‘The sun has already risen.’

b. ***garm čīdow* ‘get warm (of the weather)’**

Taram(=\*i) garm čūd.  
 there=(\*3SG) warm do.PST  
 ‘It got warm there.’

**Table 11.11:** Unaccusative complex verbs with *čīdow* ‘do’.

<u>VERB</u>	<u>GLOSS</u>
<i>dārō čīdow</i>	hurt (of a body part)
<i>garm čīdow</i>	get warm (of a place)
<i>pal čīdow</i>	rise (of the sun)
<i>qaror čīdow</i>	calm (of rain, snow)
<i>šitto čīdow</i>	get cold (of a place)
<i>torik čīdow</i>	get dark (of a place)

### 11.3.3 Other issues in complex verbs

In the last subsection of the discussion on complex verbs, we turn to two issues regarding the form and syntax of these verbs which, to my knowledge, have not been discussed at any length in any publication on the language. These are (i) the shortening of light verbs, examined in Section 11.3.3.1, and (ii) verbs which, with respect to their stem paradigms, display characteristics of both complex and simplex verbs. The latter phenomenon is addressed in Section 11.3.3.2.

### 11.3.3.1 Light-verb shortening: Zero allomorph of *kin-*

Certain complex-verb constructions, namely those which involve the present stem of *čīdow* ‘do’ exhibit a phenomenon reminiscent of other instances verb shortening in Shughni (see Section 9.1.5). Here, the present stem *kin-* is optionally realized as a zero allomorph for all person-number combinations; the third-singular form, which has the suppletive form *kiǰt*, does not have the option to shorten in this way. This process is schematized in (306), with the resulting structure given in (307).<sup>5</sup>

#### (306) Full realization of *kin-*

. . .      NONVERB      ***kin-***AGR

#### (307) Zero allomorph of *kin-*

NONVERB- $\emptyset$ -AGR

The sentences in (308) illustrate this type of shortening. In (308a) the long version of the verb *kor čīdow* retains the present stem *kin-*, while in its shortened counterpart in (308b) the present stem is realized as a zero allomorph, leaving only the non-verbal component and the agreement suffix. For comparison, the entire verb, including the non-verbal component, stem, and agreement, is bolded in each example.

#### (308) Zero allomorph of *kin-*: Examples

##### a. Full stem

Wuz **kor *kin-*um.**  
I    work do.PRS-1SG  
‘I work.’

##### b. Zero allomorph

Wuz **kor  $\emptyset$ -um.**  
I    work do.PRS-1SG  
‘I work.’

The complex verb *kor čīdow* thus has the two paradigms in (309): a paradigm where the overt present stem *kin-* is maintained and a paradigm where it is shortened due to the use of a zero allomorph. Note that the third-singular cell remains unchanged.

<sup>5</sup>This process is noted briefly in the dictionary of Karamshoev (1988a). Many thanks to Prof. Vladimir Plungian for pointing out the elegance of the zero-allomorph analysis.

(309)	a. <u><b>Kor čīdow: Non-shortened.</b></u>	b. <u><b>Kor čīdow: Shortened.</b></u>																								
	<table> <tr> <th></th><th><u>SG</u></th><th><u>PL</u></th></tr> <tr> <td>1</td><td><i>kor kin-um</i></td><td><i>kor kin-ām</i></td></tr> <tr> <td>2</td><td><i>kor kin-i</i></td><td><i>kor kin-et</i></td></tr> <tr> <td>3</td><td><i>kor kiṣt</i></td><td><i>kor kin-en</i></td></tr> </table>		<u>SG</u>	<u>PL</u>	1	<i>kor kin-um</i>	<i>kor kin-ām</i>	2	<i>kor kin-i</i>	<i>kor kin-et</i>	3	<i>kor kiṣt</i>	<i>kor kin-en</i>	<table> <tr> <th></th><th><u>SG</u></th><th><u>PL</u></th></tr> <tr> <td>1</td><td><i>kor-um</i></td><td><i>kor-ām</i></td></tr> <tr> <td>2</td><td><i>kor-i</i></td><td><i>kor-et</i></td></tr> <tr> <td>3</td><td><i>kor kiṣt</i></td><td><i>kor-en</i></td></tr> </table>		<u>SG</u>	<u>PL</u>	1	<i>kor-um</i>	<i>kor-ām</i>	2	<i>kor-i</i>	<i>kor-et</i>	3	<i>kor kiṣt</i>	<i>kor-en</i>
	<u>SG</u>	<u>PL</u>																								
1	<i>kor kin-um</i>	<i>kor kin-ām</i>																								
2	<i>kor kin-i</i>	<i>kor kin-et</i>																								
3	<i>kor kiṣt</i>	<i>kor kin-en</i>																								
	<u>SG</u>	<u>PL</u>																								
1	<i>kor-um</i>	<i>kor-ām</i>																								
2	<i>kor-i</i>	<i>kor-et</i>																								
3	<i>kor kiṣt</i>	<i>kor-en</i>																								

The present stem *kin-* may be realized as a zero allomorph in virtually all complex verbs built on the verb *čīdow* ‘to do’. Thus, for instance, the complex verb *čaq-čaq kin-ām* ‘we chat’ becomes shortened to *čaq-čaq-ām*; the complex verb *qoq kin-ām* ‘we dry (tr.)’ becomes shortened to *qoq-ām*; and so on. However, no zero allomorph exists for the past, perfect, and infinitive stems, and thus verbs with these stems are never shortened in this way. Thus, the following forms are illicit:

PST: *kor čūd* → \**kor-t*

INF: *kor čūṣṣ* → \**kor-č*

INF: *kor čīdow* → \**kortow*

Moreover, this phenomenon is restricted to the present stem of the light verb *čīdow* ‘only’. Other light verbs used in complex verbs do not have zero allomorphs of their present stems. Compare, for instance, the examples in (310), which show the ungrammatical realization of the light verb *ōēdow* (prs. stem *ōāō-*) as a zero allomorph.

(310) **Light verb *ōēdow* ‘give; hit’: No zero allomorph**

a. Tu dis xaṣ **gāp** **ōāō-i**  
 you so loud word hit-2sg  
 ‘You are speaking very loudly.’

b. \*Tu dis xaṣ **gāp-i**  
 you so loud word-2sg  
 Intended: ‘You are speaking very loudly.’

### 11.3.3.2 Hybrid complex verbs

Two common Shughni verbs, identical in form except for their first consonant, exhibit behavior which places them in between simplex and complex verbs. These are the verbs *dāktow* ‘give’, which has an alternative complex infinitive stem *dāk čīdow*, and *lāktow* ‘let; allow’, which has an alternative complex infinitive stem *lāk čīdow*.

These verbs, which I refer to here as *hybrid complex verbs*, exhibit different levels of “hybridness” across their stems. In their present stems, they are nearly fully simplex, with each verb having a simplex present stem – *dāk-* ‘give’ and *lāk-* ‘let; allow’, respectively – which is obligatorily used in all person-number combinations except third-singular. Here, each verb has an alternative form built on the third-singular present stem of the light verb *čīdow* ‘do’, *kišt*, hence *dāk(k)išt* and *lāk(k)išt*. (The fact that these forms are written as single words, rather than separating the non-verbal component from the light verb, as is typically done, reflects both native-speaker intuition and the fact that one *k* tends to be dropped, with the other (re-)syllabifying as the onset of the second syllable.) The present paradigms of each verb are given in (311).

(311) a. *Dāktow* ‘give’: Prs. paradigm

	<u>SG</u>	<u>PL</u>
<u>1</u>	<i>dāk-um</i>	<i>dāk-ām</i>
<u>2</u>	<i>dāk-i</i>	<i>dāk-et</i>
<u>3</u>	<i>dāk-t</i> // <i>dāk(k)išt</i>	<i>dāk-en</i>

b. *Lāktow* ‘let; allow’: Prs. paradigm

	<u>SG</u>	<u>PL</u>
<u>1</u>	<i>lāk-um</i>	<i>lāk-ām</i>
<u>2</u>	<i>lāk-i</i>	<i>lāk-et</i>
<u>3</u>	<i>lāk-t</i> // <i>lāk(k)išt</i>	<i>lāk-en</i>

The examples in (312) and (313) show these verbs in use. Note that in (312), the third-singular form simplex *dāk-t* ‘(she) gives’ has the alternative complex-like form *dāk(k)išt*. In (313), however, the verb *lāk-i* ‘you let’ is in the second singular and therefore does not have a complex alternative form.

(312) Mu mūm=ta māš-ard ūžin **dāk-t** (// **dāk(k)išt**).  
my grandmother=FAC us-DAT dinner give.PRS-3SG (// give.PRS.3SG)  
‘My grandmother will give us dinner.’

(313) Tu xu ziryot-en **lāk-i** (\*// **lāk kin-i**) dis der naštīd-ow.  
you REFL child-PL let.PRS-2SG (\*// let do.PRS-2SG) so late leave.INF-INF  
‘You let your children go out very late.’

In the remainder of their stems, these verbs each have a full set of simplex stems and a full set of complex stems. In the affirmative, each may be used either as a simplex verb – *dāktow* and *lāktow* – or as a complex verb – *dāk čīdow* and *lāk čīdow*. These stems are shown in Table 11.12.

**Table 11.12:** Non-present stems of *dāktow* and *lāktow*.

	‘to give’		‘to let; allow’	
	<u>SIMPLEX</u>	<u>COMPLEX</u>	<u>SIMPLEX</u>	<u>COMPLEX</u>
PST	<i>dākt</i>	<i>dāk čūd</i>	<i>lākt</i>	<i>lāk čūd</i>
PRF	<i>dākč</i>	<i>dāk čūč</i>	<i>lākč</i>	<i>lāk čūč</i>
INF	<i>dākt(-ow)</i>	<i>dāk čūd(-ow)</i>	<i>lākt(-ow)</i>	<i>lāk čūd(-ow)</i>

Finally, with respect to negation, these two verbs always behave as simplex verbs, even when their complex-like stems are used. That is, on the one hand, canonical complex verbs in Shughni are negated by a prefix (indicative *na-* or prohibitive *mā-*) attached directly to the light verb stem. In the case of the complex-like stems *dāk čūdow* and *lāk čūdow*, however, negation obligatorily occurs to the left of the non-verbal components *dāk* and *lāk*. This is the case for both the long form of the third-singular present shown, as shown in (314a) and for the complex-like forms of the past and perfect stems, as shown in (314b).

(314) **Verbal negation with long forms of *dāk čūdow* ‘give’**

- a. Yā=ta mu-rd wam kitob **na-dāk(k)ičt** (// \*dāk na-kičt).  
 she=FAC me-DAT DEM.OBL.F book NEG-give.PRS.3SG (// \*give NEG-do.PRS.3SG)  
 ‘She’s not going to give me the book.’
- b. Yā=yi mu-rd wam kitob **na-dāk-čūd** (// \*dāk na-čūd).  
 she=3SG me-DAT DEM.OBL.F book NEG-give-do.PST (// \*give NEG-do.PRS.3SG)  
 ‘She didn’t give me the book.’

In summary, the verbs *dāktow* // *dāk čūdow* ‘give’ and *lāktow* // *lāk čūdow* ‘let; allow’ appear to be in between canonical simplex and canonical complex verbs in Shughni. Their present stems are nearly fully simplex (with the exception of the third-singular), while the remainder of their stems can be either simplex-like or complex-like. Negation for these verbs, however, is always done as it is with simplex verbs. The diachronic development of these verbs, including their future development, is an intriguing prospect for future investigation.

The topic of hybrid complex verbs concludes the investigation here into complex verbs. It was seen that these verbs play a significant role in indicating alternations in transitivity, with many transitive~intransitive pairs of

complex verbs differing only in the light verb they are built on. The last two sections of this chapter look at two further issues in transitivity and argument structure, namely the class of reflexive verbs (Section 11.4) and oblique-first constructions (Section 11.5), where in neutral word order, the oblique argument of a verb comes before any nominative arguments.

## 11.4 Reflexive verbs

Shughni possesses a class of verbs, which I refer to here as *reflexive verbs*, in which the subject-oriented oblique anaphor *xu* (glossed REFL) appears in direct-object position. These constructions may also be a candidate for *middle voice constructions* (see Inglese 2022), and the most accurate term for them remains open to question. To my knowledge, this class of Shughni verbs is not examined in any prior works on the language. The goal of this section is to provide a brief introduction to this class of verbs and to point out topics which might be of interest for future research.

Initial examples with reflexive verbs, in which the reflexive pronoun *xu* and other parts of the verb are bolded, are given in (315). (For a more thorough discussion on this anaphor, see Section 6.1.4.1.)

### (315) Reflexive verbs: Initial examples

#### a. *xu zār čīdow* – ‘to get injured’

Ar Dušanbi velosiped-ti mā-ti, lāk **zār xu mā-k**.  
 down.in Dushanbe bicycle-LOC PROH-go.2SG.IMP SUBR harm REFL PROH-do.PRS-2SG  
 ‘Don’t ride a bike in Dushanbe, so you don’t get hurt.’

#### b. *xu is čīdow* – ‘to feel (INTR.)’

Gulya pi Pomir ca yat, yā **xu bīdi daraw is čīd** sat.  
 Gulya up.to Pamir SUBR come.PST she REFL better INCEP feeling do.INF become.PST.F  
 ‘When Gulya came to the Pamirs, she began to feel better.’

All reflexive verbs have in common that they are built on transitive light verbs. And most, if not all, reflexive verbs have a transitive counterpart without the oblique anaphor *xu*, as with the two verbs used in examples (315) above. Thus, the reflexive verbs *xu zār čīdow* ‘to injure oneself’ and *xu is čīdow* ‘to feel (INTR.)’ have transitive



counterparts *zār čīdow* ‘to injure’ and *is čīdow* (TR), respectively. Examples with the transitive counterparts of each verb are given in (316).

(316) **Transitive counterparts of reflexive verbs**

a. *zār čīdow* – ‘to injure’, cf. (315a)

Wam daryo quwwat vārðed mardum **zār** **čīd-ow**.  
 DEM.OBL.F river power can.PRS.3SG people injury do.INF-INF  
 ‘The power of the river can injure people.’

b. *is čīdow* – ‘to feel (TR)’, cf. (315b)

Tu=ta padam mardum mêr **is** **kin-i**.  
 you=FAC up.there people kindness feeling do.PRS.-2SG  
 ‘You will feel people’s kindness there.’

Note that the reflexive verbs in (315a) and (315b) are both complex verbs, built on the light verb *čīdow* ‘do’ and non-verbal components *zār* ‘injury; harm’ and *is* ‘feeling’, respectively. As can be seen upon comparison of examples (315a) and (315b), the position of the reflexive component with respect to the other elements of the verb is variable. In the former, *xu* occurs to the left of both the non-verbal component and the verb (*xu – zār – čīdow*), while in the latter it occurs between the two (*is – xu – čīdow*). These two orders of elements, schematized in (317), are the most common and are equally permissible, although a consultant notes that the second order, where the reflexive *xu* occurs between the non-verbal and verbal elements might be preferred in higher registers.

(317) **Order of elements in reflexive verbs**

- |    |                  |                  |      |   |   |
|----|------------------|------------------|------|---|---|
| a. | <b><i>xu</i></b> | NON-VERBAL COMP. | VERB | — | <b><i>xu</i></b> <i>is čīdow</i>        |
| b. | NON-VERBAL COMP. | <b><i>xu</i></b> | VERB | — | <i>is</i> <b><i>xu</i></b> <i>čīdow</i> |

Clauses with reflexive verbs, like other verbal constructions in Shughni, are sensitive to information-structural notions, such as topic and focus, which call for scrambling. The reflexive anaphor *xu* can thus appear in other places in the clause, and in other orders with respect to the verb and non-verbal element. (Section 12.2 looks in detail at the relationship between the order of elements in Shughni clauses and notions of topic and focus.)

Nonetheless, it remains an open question whether the order of elements in reflexive verbs consistently behaves similarly to the order of elements in clauses containing transitive verbs with direct objects.

Regarding their form, not all reflexive verbs in Shughni are complex. Some consist of a simplex verb and the oblique anaphor, as in *xu wīntow* ‘to see oneself’, or of a simplex verb, the oblique anaphor, and an adjective, as in *xu maŷdzūnj/mot lāktow* ‘to leave oneself hungry/tired.’ Examples of these are given (318).

(318) **Simplex reflexive verbs**

a. *xu wīntow* – ‘to see oneself’

Yu pi vannaya sifint=xu tar aynak=i **xu wīnt**.  
 he up.to bathroom go.up.PST=and in mirror=3SG REFL see.PST  
 ‘He went up to the bathroom and saw himself in the mirror.’

b. *xu (ADJ) lāktow* ‘to leave oneself (ADJ)’

Wuz=um nur yičīzaθ na-xūd=xu **maŷdzūnj=um xu lākt**.  
 I=1SG today nothing NEG-eat.PST=and hungry=1SG REFL leave.PST  
 ‘I didn’t eat anything today and left myself hungry.’

Another open question relates to the productivity of reflexive verbs in Shughni. Thus, it is unclear whether all intransitive verbs can, under the right circumstances, be used reflexively with the oblique anaphor *xu*, nor is it clear whether and to what extent intransitive complex verbs built on transitive light verbs, such as *arakat čīdow* ‘move (around)’ may be made reflexive. These are thus two further topics to be investigated.

To conclude this section, an extensive (but by no means exhaustive) list of Shughni reflexive verbs is provided in Table 11.13. Where a reflexive verb has a non-reflexive transitive counterpart, the transitive verb is provided as well. In some instances, the meaning of a transitive verb and its reflexive counterpart differs.

## 11.5 Oblique-first constructions

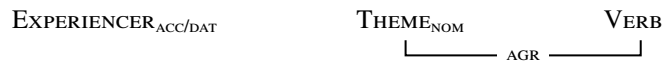
In this final section in the chapter on issues in transitivity an argument structure in the Shughni verbal system, the discussion turns to a group of heterogeneous constructions which I shall refer to collectively as *oblique-first constructions*. Semantically, these are generally predicates denoting psychological states (i.e. psych-verbs) and

**Table 11.13:** Reflexive verbs in Shughni (examples).

Reflexives		Transitives	
VERB	GLOSS	VERB	GLOSS
<i>xu is čīdow</i>	feel (INTR)	<i>is čīdow</i>	feel (TR)
<i>xu registraciya čīdow</i>	register oneself	<i>registraciya čīdow</i>	register (someone)
<i>xu xaǰ čīdow</i>	restrain oneself	<i>xaǰ čīdow</i>	stick (something somewhere)
<i>xu xīp čīdow</i>	put make up on oneself	<i>xīp čīdow</i>	put make up on (someone)
<i>xu zār čīdow</i>	injure oneself	<i>zār čīdow</i>	injure (someone)
<i>xu dām ôêdow</i>	rest	<i>dām ôêdow</i>	blow; fan (a fire)
<i>xu pāst ôêdow</i>	stoop (in one's behavior)	–	–
<i>xu dor tīždow</i>	hang oneself	<i>dor tīždow</i>	hang (someone)
<i>tar zibo xu tīždow</i>	hold oneself back (FIG)	<i>tar zibo tīždow</i>	hold (someone) back (FIG)
<i>xu wīntow</i>	see oneself	<i>wīntow</i>	see
<i>xu mi ôust xu anjīvdow</i>	hold oneself back (FIG)	<i>tar zibo tīždow</i>	hold (someone) back (FIG)
<i>xu wuǰ tar xu čīdow</i>	get oneself together	–	–
<i>xu qīn čīdow</i>	make life difficult (on oneself)	<i>qīn čīdow</i>	make (something) difficult

typically involve a non-agentive, animate experiencer and an inanimate theme or stimulus. Syntactically, their defining characteristic is that in neutral word order, the first element of the clause – here, typically an animate experiencer – is in either the accusative case (i.e. the oblique case without an additional suffix) or the dative case (i.e. the oblique case with an additional suffix *-(a)rd*). The semantic theme or stimulus is in the nominative and appears after the non-nominative argument, and the verb agrees with the nominative argument. This is schematized in (319); an initial example is given in (320).

(319) **Oblique-first construction (neutral word order)**



(320) **Oblique-first construction: *xūðm yêdow* ‘feel sleepy’**

Mu xūðm na-yêst.  
 me sleep NEG-take.PRS.3SG  
 ‘I can’t sleep’ or ‘I’m not sleepy’ (lit. ‘sleep doesn’t take me’)

Shughni oblique-first constructions share many similarities with *impersonal constructions* in other Indo-European languages, such as German *mich fröstelt* or *es fröstelt mich* ‘I’m shivering; I’m feeling cold’, lit. ‘me shivers’ or ‘it shivers me’ (Barðdal 2004) and Russian *mne stydno* ‘I’m ashamed’, lit. ‘to me (it is) shameful’. Similar constructions are also found throughout the Iranian languages, although the precise semantic notions and predicates expressed through oblique-first constructions vary from language to language.<sup>6</sup>

Oblique-first constructions in Shughni have received relatively little attention in the literature. A short, cross-linguistic (pan-Iranian) overview of this type of construction is provided by Joy I. Edelman in her (1974) article on clausal structure in Iranian languages, and a recent (2023) thesis by Aleksandr Sergienko revisits the topic from a generative perspective. This section, for its part, seeks to provide a more thorough description of the range and usage of oblique-first constructions specific to Shughni.

Oblique-first constructions involve a foregrounded, non-nominative experiencer, and a backgrounded nominative theme or stimulus. Moreover, in some cases, it is unclear whether the theme is the true subject. Both of these aspects of Shughni oblique-first constructions align them with the impersonal constructions of other Indo-European languages. However, because precisely what constitutes an impersonal construction is still debated in the literature (see Malchukov & Ogawa 2011 for discussion), I have chosen here to avoid this term here and to use instead the descriptively more precise ‘oblique-first constructions’.

This section is organized as follows. Section 11.5.1 examines oblique-first constructions in which the first argument is in the accusative (i.e. **accusative-first constructions**), while Section 11.5.2 examines those in which the first argument is in the dative (i.e. **dative-first constructions**). Because dative-first constructions in Shughni are more widespread and syntactically diverse than their accusative-first counterparts, the latter section is further broken down into dative-first constructions built on simplex verbs (Section 11.5.2.1), those built on complex verbs (Section 11.5.2.2), and those built on copular constructions (Section 11.5.2.3). Finally, Section 11.5.3 provides a few notes

---

<sup>6</sup>With respect to their semantics, but less with respect to their syntax, Shughni oblique-first constructions also share many similarities with Persian impersonal constructions of the type *sard=am shod* ‘I got cold’ and *xande=ash gereft* ‘he started laughing’. In particular, Shughni and Persian constructions of this type are used in many of the same semantic and pragmatic contexts. However, whereas in Persian (and many other Iranian languages), impersonal constructions are built on pronominal enclitics which are etymologically closely related to Shughni past-tense clitics (e.g. Haig 2008), oblique-first constructions in Shughni are built on full pronominals. (See Thackston 2009: 115-116 for a short introduction to Persian impersonal constructions; the discussion in Sedighi 2005: 139-187, as well as references cited therein, provides a more thorough overview and theoretical treatment of this topic.)

on the intersection of **Shughni oblique-first constructions and notions of subjecthood**. A summary is given in Section 11.5.4. Note that another type of construction which might also be analyzed as oblique first is that of existential-like possessive constructions of the type *mu-nd mošin (yast)* ‘I have a car’, lit. ‘me-LOC car (exists)’. These are not examined here, but rather in Section 8.3.4 as part of an examination of copular constructions.

### 11.5.1 Accusative-first constructions

Only a handful of oblique-first constructions in Shughni call for an experiencer in the accusative case – that is, in the oblique case without an additional suffix. These are summarized in Table 11.14, and examples are given in (321), where the accusative argument is bolded:

**Table 11.14:** Accusative-first constructions in Shughni.

<u>VERB</u>	<u>GLOSS</u>
<i>xūðm yêdow</i>	‘to feel sleepy’
<i>šev čîdow</i>	‘to shiver’
<i>šittow ôêdow</i>	‘to catch a cold’

#### (321) Accusative-first constructions

##### a. *xūðm yêdow* ‘to feel sleepy’

**Mu** xūðm yêst=atā      mu-nd lap kor=xu    šič na-vārðīm      xêvd-ow.  
me sleep take.PRS.3SG=but me-POSS much work=and now NEG-can.PRS.1SG sleep.INF-PURP  
‘I’m sleepy, but I have a lot of work and can’t sleep right now.’

##### b. *šev čîdow* ‘to shake; shiver’

Samīra kasal ca    vad,    **wam**=i    diš šev    čūd.  
Samira sick SUBR be.PST.F her=3SG so shiver do.PST  
‘When Samira was sick, she shivered very badly.’

##### c. *šitto(w) ôêdow* ‘to get a cold’

Yu boyad mev      luq-en    pinīzd,      lāk šittow **wi**    mā-ôîd.  
he must DEM.OBL.PL clothes-PL put.ON.PRS.3SG SUBR cold him PROH-hit.SBJ  
‘He needs to put these clothes on so he won’t get a cold.’

Note that when used with a past or perfect stem, the verbs above behave as transitive in requiring a third-person-singular clitic =*yi*, as can be seen in example (321b). The third-singular clitic is used also when the experiencer is first- or second-person (e.g. *mu-rd*=\*(*i*) *dis šev čūd* ‘I was shivering so much.’). As discussed below with dative-first constructions, the issue of subjecthood in accusative-first clauses, and in Shughni in general, remains up for debate.

### 11.5.2 Dative-first constructions

The majority of experiencer-first constructions in Shughni call for an experiencer in the dative case. These include at least three fundamental types of predicate: (i) those built on a simplex verb (i.e. *fortow* ‘to want, lit. to be desired’); (ii) those built on a complex verb with a non-verbal component (e.g. *qīni čīdow* ‘to be difficult for’); and (iii) copular constructions built on an adjective or noun (e.g. *xuš* (*vidow*) ‘to like, lit. to be pleasant for’ and *nīwjak* (*vidow*) ‘feel like crying’, lit. ‘exist crying to’). Each of these types are examined in turn in the subsections which follow.

#### 11.5.2.1 Dative-first constructions with simplex verbs

Only a single simplex verb in Shughni calls for a dative-first construction. This is the verb *fortow*, often translated as ‘want’, but more accurately translated as ‘be desirous to’. This verb agrees with the semantic theme – i.e. the nominative argument – and never with the dative experiencer. (In general, verbal agreement in Shughni seems to take place only with nominative arguments and never with non-nominative arguments. See Moravcsik 1974 and Bobaljik 2008 on typological generalizations regarding the intersection of agreement and the case-marking of arguments.) The theme, for its part, may be either a concrete object, as in (322) (or even a person, though this carries a sexual connotation) or an infinitival clause, as in (323).

##### (322) *fortow* ‘be desirous’ – (concrete) nominal theme

Wi-rd    disga    birūken na-for-en.  
 him-DAT this.type pants.PL NEG-be.desirous.PRS-3PL  
 ‘He doesn’t want this kind of pants.’

(323) *fortow* ‘be desirous’ – infinitival theme

Wuz=um dis maʃdzūnj, mu-rd šīrčoy birêxt-ow fort.  
 I=1SG so hungry me-DAT milk.tea drink.INF-PURP want.PRS.3SG  
 ‘I’m so hungry; I want to drink some milk tea.’

Like in the accusative-first constructions discussed above, *fortow*, when used in the form of its past or perfect stem, behaves as a transitive verb in always requiring the third-person singular clitic. This is shown in example (324):

- (324) Wi-rd\*(=i) biyor na-fort māš qati ar xār sitt-ow.  
 him-DAT\*(=3SG) yesterday NEG-be.desirous.PST us with to city go.INF-NMZ  
 ‘He didn’t want to go with us to the city yesterday.’

## 11.5.2.2 Dative-first constructions with complex verbs

Another set of dative-first constructions in Shughni are built on complex verbs consisting of a non-verbal component and the light verb *čīdow* ‘do’. These are listed in Table 11.15, and examples are given in (325)–(326).

Table 11.15: Dative-first constructions with complex verbs.

CONSTRUCTION	GLOSS	LITERAL TRANSLATION
<i>šittoyi čīdow</i>	‘feel cold’	‘do cold (to)’
<i>garmi čīdow</i>	‘feel warm’	‘do warm (to)’
<i>qīni čīdow</i>	‘be difficult for’	‘do difficulty to’
<i>xušrūyi čīdow</i>	‘seem beautiful to’	‘do beauty to’

(325) *garmi čīdow* ‘feel hot’

Tar mi čīd ačaθ šamol nist, mu-rd garmi kixt.  
 in DEM.OBL.M house at.all wind COP.NEG 1SG.OBL-DAT warmth do.PRS.3SG  
 ‘There is no wind at all in this house, and I feel hot.’

(326) *qīni čīdow* ‘be difficult (for)’

Tu-rd=en wāð ikzamin sawol-en qīni čūd o?  
 you-DAT=3PL DEM.DIR.PL exam question-PL difficulty do.PST PQ  
 ‘Were those exam questions difficult for you?’

Similar to accusative-first constructions and to the verb *fortow* discussed in Section 11.5.2.1 above, dative-first constructions built on complex verbs with *čīdow* ‘do’ require the third-person singular clitic when used with past and perfect stems. This is shown in (327):

- (327) Mu-rd\*(=i) biyor lap šittoyi čūd=ata nur mu-rd garmi kiǰt!  
 me-DAT\*(=3SG) yesterday very cold do.PST=but today me-DAT warmth do.PRS.3SG  
 ‘Yesterday I felt very cold, but today I feel very hot!’

### 11.5.2.3 Dative-first constructions with the copula

The final type of dative-first construction to be discussed here is that built on the copula – i.e. second-person clitics in the present tense and the verb *vidow* ‘be’ in the past. The non-verbal part of the copular type of dative-first construction involves either a noun – as in *šumak* (*vidow*) ‘feel lazy’, lit. ‘exist laziness (to)’ – or an adjective – as in *xuš* (*vidow*) ‘like’, lit. ‘be pleasant to’. These are listed in Table 11.16, and examples are given in (328)–(330). Note that the last two words *interesni* ‘interesting’ and *obidni* ‘offensive’ are the Russian borrowings интересно ‘interesting’ and обидно ‘offensive’, respectively.

**Table 11.16:** Dative-first constructions in Shughni.

CONSTRUCTION	GLOSS	LITERAL TRANSLATION
<i>šumak</i> ( <i>vidow</i> )	‘feel lazy (to)’	‘exist laziness (to)’
<i>šūnčak</i> ( <i>vidow</i> )	‘feel like laughing’	‘exist laughter (to)’
<i>nīwjak</i> ( <i>vidow</i> )	‘feel like crying’	‘exist crying (to)’
<i>xuš</i> ( <i>vidow</i> )	‘like’	‘be pleasant (to)’
<i>berawot</i> ( <i>vidow</i> )	‘feel uncomfortable (doing)’	‘be uncomfortable (to)’
<i>darkor</i> ( <i>vidow</i> )	‘need’	‘be necessary (to)’
<i>interesni</i> ( <i>vidow</i> )	‘find interesting’	‘be interesting (to)’
<i>obidni</i> ( <i>vidow</i> )	‘find offensive’	‘be offensive (to)’

- (328) *berawot* (*vidow*) ‘be uncomfortable for’

Mu-rd berawot čap őust qati nivišt-ow.  
 me-DAT uncomfortable left hand with write.INF-NMZ  
 ‘It’s uncomfortable for me to write with my left hand.’



- (329)
- šūnčak widow*
- ‘feel the need to laugh’

Wam-ard šūnčak vud=xu yā azam naǰtoyǰ!  
 her-DAT laughter be.PST.M=and she from.there leave.PST.F  
 ‘She was about to laugh and left.’

- (330)
- šumak (widow)*
- ‘feel lazy (to)’

Wuz=ta na-sām tar kīno, mu-rd dis šumak mi garmi-ndi naǰtīd-ow.  
 I=FAC NEG-go.PRS.1SG to theater, me-DAT such laziness this heat-LOC go.out.INF-NMZ  
 ‘I’m not going to go to the theater; I’m so lazy to go out in this heat.’

Certain dative-first constructions which are typically used with the copula – including *xuš (widow)* ‘like’ and *darkor (widow)* ‘need’ – may also be used with the verbs *sittow* ‘become’ or *yattow* ‘come’, instead of the copula, to emphasize a change of state. Thus, for instance, the construction *xuš (widow)* ‘like’ expresses the state of liking, whereas the verb *xuš yattow* emphasizes the moment when liking begins to take place. Compare, for instance, the examples in (331) and (332):

- (331)
- xuš widow*
- ‘to like (state)’

Mu-rd yā kitob xuš vad=ata šič yā mu-rd xuš nist.  
 me-DAT DEM.DIR.F book pleasant be.PST.F=but now DEM.DIR.F me-DAT pleasant NEG.COP  
 ‘I liked that book, but now I don’t like it anymore.’

- (332)
- xuš yattow*
- ‘to like (process)’

Tu-rd az mev kitob-en-and čidūm xuš yat?  
 you-DAT from DEM.OBL.PL book-PL-LOC which pleasant come.PST  
 ‘Which one of these books did you like?’

### 11.5.3 Subjecthood and oblique-first constructions

The final topic to be discussed in connection with oblique-first constructions in Shughni is that of subjecthood. In Shughni, the question arises as to whether the dative arguments of such constructions pattern with subjects or not. Here, two aspects of oblique-first constructions will be pointed out with potential implications for the status of the dative as subject or non-subject.

First, the construction *šitto ôêdow* ‘become cold’, unlike the other accusative-first constructions *xūôm yêdow* ‘feel sleepy’ and *šev čîdow* ‘shiver’, does not have a nominal component, but rather an adjectival component in *šitto*

‘cold’ (cf. the noun *šittoyi* ‘cold(ness)’). This is in fact the only oblique-first construction without a clear candidate for subjecthood, as all others have either a noun in the nominative case or an infinitival clause (see, e.g., the examples in 328–330). And while it may not be immediately clear whether *xūðm* ‘sleep (N)’ and *šev* ‘shiver (N)’ are the true subjects of their respective accusative-first constructions, it is much more difficult to argue that the adjective *šittow* is the true subject of *šittow ðêdow*. This is the most viable candidate for a subjectless impersonal construction in Shughni.

The second point to be made here is that in some instances, the dative argument in dative-first constructions may bear at least one property of subjecthood (besides appearing clause-initially in canonical word order), in that it is sometimes permitted to be co-referential with the subject-oriented oblique anaphor *xu*. This is only permitted (and is in fact optional), however, in cases where the anaphor is embedded within an infinitival clause. Two such examples are given in (333). It is never permitted when *xu* appears directly in the main clause. This is shown in (334). Note that in both (333b) and (334b), the use of a third-person pronoun results in ambiguity, as this pronoun may refer either to the subject of the matrix clause or to a third party:

(333) **Dative-first with infinitival theme – anaphor *xu* co-indexed with dative**

- a. Tu-rd [tu/*xu* tāt qati gap ðêd-ow] *xuš* o?  
 you-DAT your/REFL rather with word hit.INF-NMZ pleasant PQ  
 ‘Do you like speaking with your father?’
- b. Wi<sub>i</sub>-rd [pi wi<sub>i/j</sub>/*xu*<sub>i</sub> čīd sitt-ow] fort.  
 him-DAT to his/REFL house go.INF-NMZ be.desirous.PRS.3SG  
 ‘He wants to go to his house.’

(334) **Dative-first without infinitival theme – co-indexation between dative and anaphor *xu* illicit**

- a. Tu-rd tu/\**xu* čīd *xuš* o?  
 you-DAT your/\*REFL house pleasant PQ  
 ‘Do you like your house?’
- b. Wi<sub>i</sub>-rd wi<sub>i/j</sub>/\**xu*<sub>i</sub> řac fort o?  
 him-DAT his/\*REFL water be.desirous.PRS.3SG PQ  
 ‘Does he want his water?’

Hence, the notion of subject – typically understood in Shughni to be the clause-initial nominative argument which agrees with both transitive and intransitive verbs – requires closer attention here. Subjecthood in Shughni in

general, and especially its intersection with oblique-first constructions, is therefore a topic ripe for future investigation. Previous work on subjecthood in similar oblique-first constructions in other Iranian languages may be a useful starting point (see Jügel & Samvelian 2020 on Persian and Akkuş et al. 2023 on Kurdish).

#### 11.5.4 Summary: Oblique-first constructions

This section has examined oblique-first constructions in Shughni, which generally involve a psych predicate of the type corresponding to English ‘like’ or ‘want’. In these constructions, as the name implies, the non-nominative experiencer occurs precedes any nominative arguments in neutral word order. These constructions bear resemblance to impersonal constructions in other Indo-European languages such as German *mir ist kalt* and Russian *mne xolodno*, both meaning ‘I am cold’.

The non-nominative argument in Shughni oblique-first constructions may be either accusative (i.e. oblique without any additional morphology), as in *mu šev kiŋt* ‘I’m shivering’, or else dative (i.e. oblique with the dative suffix *-(a)rd*), as in *mu-rd xuš* ‘I like it.’ Moreover, oblique-first constructions in Shughni may be built on either verbs (e.g. *mu-rd fort* ‘I want’) or copular constructions with adjectives (e.g. *mu-rd xuš* ‘I like’).

The issue of subjecthood in Shughni oblique-first constructions is of particular interest. The non-nominative argument, although in the oblique case and unable to agree with verbs, does possibly show at least one property of subjects. In particular, it may have the ability to bind the subject-oriented oblique anaphor *xu*, although this is a matter of analysis. Issues regarding subjecthood in this type of clause is thus an important matter for further research.

This section on oblique-first constructions concludes not only this chapter on issues in transitivity and argument structure, but also the series of four chapters dedicated to the Shughni verbal system. The next chapter examines other issues in Shughni syntax, including subordinate clauses, question formation, and morphosyntactic alignment.



## **Part IV**

# **Syntax**



## Chapter 12

# Other Syntactic Phenomena

This chapter examines a number of syntactic phenomena which have not been treated explicitly in any detail thus far. Specifically, it investigates complex clauses, question formation, and morphosyntactic alignment. The chapter is organized as follows.

First, Section 12.1 first examines **complex clauses**. This section provides a detailed look at relative and adverbial clauses, which are both generally formed using subordinating elements such as the morphemes *ca* and *di*. It also examines complement clauses and dependent speech, which optionally take the complementizer *idi*, as well as the coordination of clauses using three different conjunctions =*xu*, =*at*, and =*atā*.

Section 12.3 then looks at various aspects of **question formation** in Shughni, first discussing *wh*-questions and polar questions, before turning to more specialized forms of questions, including echo questions, tag questions, and questions with the emphatic particle *ku*.

Finally, Section 12.4 discusses **morphosyntactic alignment** in the language, including the phenomenon of vestigial ergativity, a split-intransitive alignment pattern found only in one small corner of the past tense. This section provides a particularly detailed discussion of this aspect of Shughni grammar. It draws on previous works on the history of alignment in Iranian and the diversity of alignment systems in modern Iranian languages to provide a thorough historical and typological picture of alignment in Shughni.

## 12.1 Complex clauses

This subsection examines several types of complex clauses in Shughni. The first three sections focus on different types of subordinate clauses, including **relative clauses** (Section 12.1.1), **temporal adverbial clauses** (Section 12.1.2), and **other types of subordinate clauses** (Section 12.1.3), such as the antecedents of conditionals, concessive clauses, and purposive clauses. Then, Section 12.1.4 turns to **complement clauses and dependent speech**, and Section 12.1.5 examines the **coordination of clauses**. Lastly, Section 12.1.6 offers a brief summary of complex clauses.

### 12.1.1 Relative clauses

There are three types of relative clauses found in Shughni: (i) **externally headed**; (ii) **internally headed** (specifically, a subtype of internally headed relative clause known as a *correlative clause*), and (iii) **free (headless) relatives** (specifically, *free-choice* free relatives). Initial examples of each type are given in (335). The head noun in each example is bolded, where applicable, and square brackets are placed around the relative clause.<sup>1</sup>

#### (335) Shughni relative clauses: Initial examples

##### a. Externally headed relative clause

Ik=ā                    **ŷinik**<sub>i</sub> [biyor=at      wam<sub>i</sub> ca wīnt]<sub>RC</sub> yā<sub>i</sub> mis ikūdand kor    kiḫt.  
 PREC=DEM.DIR.F woman yesterday=2SG her    REL see.PST she also here    work do.PRS.3SG  
 ‘The woman you saw yesterday also works here.’

##### b. Internally headed (correlative) relative clause

[Ik=wam                    **ŷinik**<sub>i</sub>=at    biyor      ca wīnt]<sub>RC</sub> yā<sub>i</sub> mis ikūdand kor    kiḫt.  
 PREC=DEM.OBL.F woman=2SG yesterday REL see.PST she also here    work do.PRS.3SG  
 ‘The woman you saw yesterday also works here.’

##### c. Free-choice free relative clause

[Arčīz    wev tar    ḍust ca    ḍod]<sub>RC</sub> zoḫt=en.  
 whatever them into hand REL fall.PST take.PST=3SG  
 ‘They took whatever fell into their hands.’

(Karamshoev 1988a: 135)

<sup>1</sup>Head nouns are bolded in both the Shughni and the free translation throughout this section.



Relative clauses in Shughni most often contain the subordinator *ca* (glossed REL in relative clauses), although in some instances it may be absent (Edelman & Dodykhudoeva 2009b: 811). This subordinator attaches to the left edge of the verb, if present, and otherwise appears at the right edge of the subordinate clause (e.g. in copular constructions with no overt verb; see 338b for such an example).

Although relative clauses are mentioned briefly by Edelman & Dodykhudoeva (2009b), and although many examples are given throughout the dictionaries of Karamshoev (1988a) and Alamshoev & Alamshoev (2020), they have largely been neglected in scholarship on Shughni. Indeed, as far as I am aware, the existence of both externally and internally headed relative clauses in the language has yet to be explicitly noted.

With this in mind, the remainder of this subsection provides a description of each type of relative construction, including their clause-internal structure, the position of the relative clause with respect to the matrix clause, and the use of elements such as resumptive pronouns. Externally headed relative clauses are examined in Section 12.1.1.1, followed by correlative clauses in Section 12.1.1.2 and free relatives in Section 12.1.1.3

### 12.1.1.1 Externally headed relative clauses

Externally headed (EH) relative clauses in Shughni obligatorily follow the head noun. The basic structure of an EH relative construction, together with an example, is given in (336).

#### (336) Externally headed relative construction: Schema

<b>Head</b> $N_i$	<b>[</b>	<b>(. . .)</b>	<b>(PRON<sub>i</sub>)</b>	<b><i>ca</i></b>	<b><i>V</i></b>	<b>]<sub>RC</sub></b>
<i>yā</i> <i>yāc<sub>i</sub></i>	<b>[</b>	<i>biyor=um</i>	<i>wam<sub>i</sub></i>	<i>ca</i>	<i>wīnt</i>	<b>]<sub>RC</sub></b>
that.F girl		yesterday=1SG	her	REL	see.PST	
‘the <b>girl</b> I saw yesterday’						

In EH relative constructions, the relative clause may either immediately follow the head noun, as in the example in (336), or else be extraposed, in which case it appears to the right of the matrix clause. In such cases, the relative clause still follows the head noun, but does not *immediately* follow it. The examples in (337) exhibit these two possible placements of the relative clause.

(337) **Ordering of EH relative clause and matrix clause**a. **RC extraposed**

Ik=wāð      piš-en=en tar kā      sat      [wam      mošin pi bīr=en ca vad]<sub>RC</sub>?  
 PREC=DEM.DIR.F cat-PL=3PL to where go.PST.PL DEM.OBL.F car      to under REL be.PST  
 ‘Where did those **cats** go which were lying under the car?’

b. **RC immediately follows head noun**

Ik=wāð      piš-en [wam      mošin pi bīr=en      ca vad]<sub>RC</sub> (wāð=en) tar kā      sat?  
 PREC=DEM.DIR cat-PL DEM.OBL.F car      to under=3PL REL be.PST.PL (they=3PL) to where go.PST.PL  
 ‘Where did those **cats** which were lying under the car go?’

The sentence in (337a) shows an extraposed relative clause. Here, the head noun and the relative clause are separated by the matrix predicate, namely *tar kā sat* ‘went where’. In (337b), by contrast, the relative clause immediately follows the head noun, and the remaining part of the matrix clause appears after the relative clause.<sup>2</sup>

Note that in EH relative constructions, the head noun appears in the case corresponding to its grammatical function in the *matrix* clause, rather than its grammatical function in the relative clause. This is illustrated in the examples in (338). In these examples, the demonstrative which shows the case of the head noun in the matrix clause is underlined.

(338) a. **Head noun is *subject* in MC: Direct case**

Yā      kitob lap bašānd vad      [qatiyaθ=ām wam      ca xēyd]<sub>RC</sub>.  
 DEM.DIR.F book very good be.PST.F together=1PL DEM.OBL.F REL read.PST  
 ‘The **book** we read together was very good.’

b. **Head noun is *object* in MC: Oblique case**

Ik=wam      kitob=at xēyĵ o [ik=ā      mīz pi bīr ca]<sub>RC</sub>?  
 PREC=DEM.OBL.F book=2SG read.PST PQ PREC=DEM.DIR.F table to under REL  
 ‘Have you read that **book** (which is) under the table?’

In each of the examples in (338), there is a mismatch between the grammatical function of the head noun in the

<sup>2</sup>In cases where the relative clause is not extraposed, as in (337b), a pronoun cross-referencing the head noun is often used in the part of the matrix clause which follows the relative clause (cf. the optional *wāð* ‘they’ in 337b). This phenomenon may be a type of prolepsis (as discussed in Salzmann 2017, e.g.), where the head noun is uttered first, and the predicate which relates information to it is anticipated to occur after the relative clause. I leave a precise analysis of this construction to future investigation.

matrix clause and its grammatical function in the relative clause, and in both instances, the case of the head noun corresponds to its grammatical function in the matrix clause. In (338a), the head noun *kitob* appears in the direct case – as indicated by the direct demonstrative *yā* – because it is the subject of the matrix clause. And in (338b), it appears in the oblique case (cf. the oblique demonstrative *wam*) because it is the direct object in the matrix clause.

The final part of the discussion on externally headed relative clauses concerns the relativization strategies used with different types of relativized arguments. Cross-linguistically, it has been noted that the ability of noun (phrases) to be relativized (i.e. to be the head noun of a relative construction) varies according to the noun's syntactic position within the relative clause. Keenan & Comrie (1977) propose an implicational hierarchy which ranks certain types of arguments as more or less “accessible” to relativization. The hierarchy, schematized in (339), states that noun phrases which are subjects of the relative clause are cross-linguistically the most accessible for relativization, followed by direct objects, indirect objects, objects of adpositions, possessors, and objects of comparison, in that order.<sup>3</sup>

(339) **Noun phrase accessibility hierarchy (Keenan and Comrie 1977: 66)**

SUBJECT > DIR. OBJ. > IND. OBJ. > OBJ. OF ADPOSITION > POSSESSOR > OBJ. OF COMPARISON

Because this hierarchy is implicational, if a given language allows the relativization of noun phrases which are direct objects in the relative clause, then it necessarily also allows the relativization of noun phrases which are subjects in the relative clause. If it allows the relativization of noun phrases which are indirect objects in relative clauses, then it necessarily also allows the relativization of direct objects and subjects, and so on.

Moreover, in a given language, it may be that multiple types of arguments are capable of being relativized, but that different relativization strategies are used with different types of relativized arguments. This is precisely the situation in Shughni, in which all six types of arguments in (339) are capable of being relativized, but there is a slight difference in relativization strategies with respect to head nouns which take the direct case in the relative clause (subject relatives), on the one hand, and those which appear in the oblique case in the relative clause (oblique

---

<sup>3</sup>This relativization hierarchy is related to the idea that subject relatives are more easily processed than object relatives (cf. Gibson 1998; Kwon et al. 2010). However, a challenge for the hierarchy is posed by syntactically ergative languages with extraction restrictions on ergative arguments (see Clemens et al. 2015 for discussion). In such languages, extraction restrictions are imposed on ergative subjects, while absolutive objects are more accessible for extraction by virtue of their unmarked case (in the sense of Bobaljik 2008; see especially fn. 12).

relatives), on the other.<sup>4</sup>

Two types of relativization strategies are used in Shughni: the *gapping strategy*, in which no resumptive pronoun co-referencing the head noun is used in the relative clause, but rather there is a “gap” where the head noun would be), and the *pronoun retention strategy*, in which a resumptive pronoun co-referencing the head noun is found in the relative clause. (On different types of relativization strategies and their cross-linguistic application, see Comrie & Kuteva 2013a; 2013b; 2013c, as well as references therein.)

For subject relatives in Shughni, a resumptive pronoun cross-referencing the head noun is *optional*, and hence both the gapping strategy and the pronoun retention strategy are available. This is shown in the examples in (340). The first of these examples (340a) shows a subject relative with gapping in the relative clause (indicated by the underscore), while the second (340b) shows the use of a resumptive pronoun (underlined) in the relative clause. Both strategies are perfectly viable.

(340) **Relativized noun is subject in relative clause: Gapping**

- a. Wāð      **sūmka-yen**=en tu-nd      o [\_\_\_ az      wam      sitol zibo=yen      ca vad]<sub>RC</sub>?  
 DEM.DIR.PL bag-PL=3PL      you-POSS PQ      from DEM.OBL.F table behind=3PL REL be.PST.PL  
 ‘Are those **bags** that were behind the table yours?’
- b. Wāð      **sūmka-yen**=en tu-nd      o [wāð=en az      wam      sitol zibo      ca vad]<sub>RC</sub>?  
 DEM.DIR.PL bag-PL=3PL      you-POSS PQ they=3PL from DEM.OBL.F table behind REL be.PST.PL  
 ‘Are those **bags** that were behind the table yours?’

For oblique relatives, on the other hand, a resumptive pronoun cross-referencing the head noun is required in the relative clause, regardless of whether the noun’s grammatical function in the relative clause is direct or indirect object, object of an adposition, or possessor. With oblique relatives, therefore, the *pronoun retention strategy* is the only relativization strategy available. Examples showing oblique relatives with various grammatical functions in the relative clause are given in (341). In each of these examples, the use of a resumptive pronoun (underlined) in the relative clause is obligatory.

---

<sup>4</sup>Objects of comparison in Shughni are realized as objects of the adposition *az* ‘from’, as in *az tu biland-di* ‘taller than you’ (lit. from you taller). In Shughni, objects of comparison therefore do not constitute a separate category of argument in the accessibility hierarchy, but rather fall within the ‘object of adposition’ category.

(341) **Relativized noun is non-subject in externally headed relative clause: Pronoun retention**a. **Relativized noun is *direct object* in relative clause**

Wi            **mollim** [ biyor=um        ar maktab \*(wi)            ca wīnt ]<sub>RC</sub> tu wi fām-i.  
 DEM.OBL.M teacher yesterday=1SG at school DEM.SG.OBL.M REL see.PST you him know.PRS-2SG  
 ‘You know the **teacher** that I saw at school yesterday.’

b. **Relativized noun is *indirect object* in relative clause**

Ik=u            **čorik** [nur=ta        \*(wi-rd) mukofot ca dāk-en]<sub>RC</sub> (yu) čāy?  
 PREC=DEM.DIR.M man today=FAC him-DAT award REL give.PRS-3SG he who.DIR  
 ‘Who is the **man** they’re giving an award to today?’

c. **Relativized noun is *object of adposition* in relative clause**

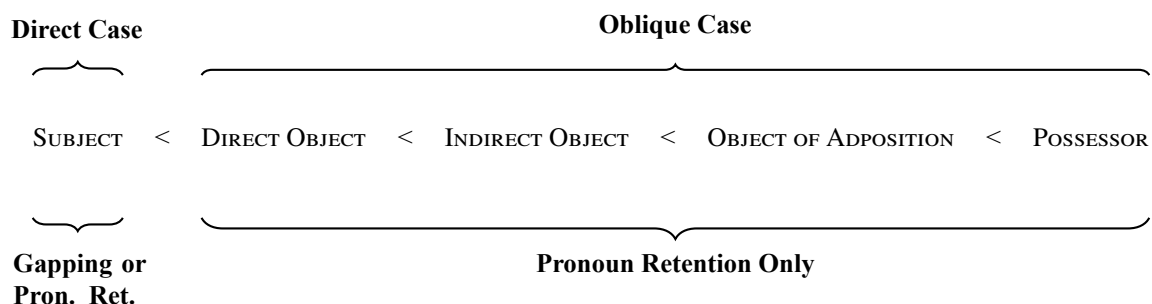
Yu            **mollim** [wuz=um \*(wi) qati gāp ca đod]<sub>RC</sub> (yu) az Dūšanbi.  
 DEM.DIR.M teacher I=1SG him with word REL hit he from Dushanbe  
 ‘The **teacher** I was talking with is from Dushanbe.’

d. **Relativized noun is *possessor* in relative clause**

Wam        **yāc-ak** [\*(wam) nān duxtur ca]<sub>RC</sub> wam qati=yum balad sut.  
 DEM.OBL.F girl-DIM DEM.OBL.F mother doctor REL her with=1SG acquainted become.PST.M  
 ‘I met that **girl** whose mother is a doctor.’

The accessibility of noun phrase types with respect to their syntactic function within the relative clause, as well as the relativization strategy used with each, is summarized in Figure 12.1 (based on the accessibility hierarchy in Keenan and Comrie 1977: 66).

**Figure 12.1:** Relativization strategies in Shughni externally headed relative clauses.



Shughni employs a subtype of internally headed relative clause known as a correlative clause. I follow the definition of correlative clauses given by Comrie & Tuteva (2013a, 2013b), who state that this type of clause cross-linguistically has two defining characteristics. First, the relative clause (i.e. correlative clause), which contains the head noun, obligatorily precedes the matrix clause. And second, an element within the matrix clause co-referencing the relativized noun is obligatory (a resumptive pronoun is used for this purpose in Shughni). This latter characteristic distinguishes correlative clauses from other internally headed relative clauses, in which the head noun is not co-referenced by a resumptive element. (See e.g. Izvorski 1996 and Lipták 2012 on the properties of correlative clauses; see also Bhatt 2015 for an alternative take on the definition of correlative clauses).

(342) **Correlative clause: Schema**

In correlative clauses, unlike externally headed relative clauses, the head noun appears in the case corresponding to its grammatical function in the *relative* clause, rather than its grammatical function in the *matrix* clause. This can be seen in the examples below, which exhibit internally headed relative clauses with a relativized subject (343a), object of preposition (343b), and possessor (343c). Note in these examples the obligatory use of resumptive pronouns (underlined> in the matrix clause.

### a. Relativized subject

[Ik=u                      **čorik**=i   māš qati gāp   ca   ḡod]<sub>RC</sub>, \*(wi) mis fām-i                      o?  
 PREC=DEM.DIR.M man=3SG US   with word REL hit.PST him   also know.PRS-2SG PQ  
 ‘Do you also know the **man** who spoke with us?’

b. **Relativized object of preposition**

[Wuz=um biyor dar borai wi **yiðā** gāp ca ðod]<sub>RC</sub> \*(wi)=yen mis zoxt.  
 I=1SG yesterday about DEM.OBL.M boy word REL hit.PST him=3PL also take.PST  
 ‘They also accepted the **boy** I was talking about yesterday (onto the team).’

c. **Relativized possessor**

[Wam **ŷinik**-and mošīn ca]<sub>RC</sub>, \*(yā) toyd ūži.  
 DEM.OBL.F woman-POSS car REL she leave.PST.PL already  
 ‘The **woman** who has a car left already.’

In (343a), the head noun *čorik* ‘man’ takes the direct demonstrative *yu* because it is the subject of the relative clause (although it is the object of the matrix clause). In examples (343b) and (343c), head nouns *yiðā* and *ŷinik* ‘woman’ are each in the oblique case because they are the object of a preposition and possessor in their respective relative clauses. Note, however, that the latter noun *ŷinik* is the subject in the relative clause, and the resumptive pronoun which co-references it is therefore in the direct case.

From a typological perspective, correlative clauses may not be so common within Iranian, as can be gleaned, for instance, from the grammatical descriptions of various modern Iranian languages in Windfuhr 2009b (although they are commonly used in the Eastern Iranian language Ossetian for relative clauses and other types of subordination – see Belyaev 2014a; Belyaev 2014b; Belyaev & Haug 2014). Nonetheless, other languages spoken in the same region as Shughni also make use of both correlative clauses and externally headed relative clauses, perhaps most notably Indo-Aryan languages such as Hindi (e.g. Bhatt 2003; Lipták 2012), but also the Pamir language Wakhi (Bashir 2009: 849-850).

Regarding their usage, although certain semantic differences have been noted between correlative and externally headed relative clauses in Hindi (see especially Lipták 2012), there is no immediately salient distinction in the use of correlative clauses and externally headed relative clauses in Shughni. My consultants have noted that both internally and externally headed relative clauses are available in the same types of discourse contexts, although more data, preferably naturally occurring speech, is needed to assess the usage of each type of relative clause.

### 12.1.1.3 Free-choice free relative clauses

The final type of relative clause to be discussed here is that of *free relatives* (FR's), also known as *headless* relatives, in which there is no overt head noun within or outside the relative clause (see van Riemsdijk 2017 and Šimik 2021 for recent discussions).

Shughni makes use of a subtype of FR's known as *free-choice* free relatives. Shughni free-choice FR's are semantically similar to those in English built on words of the form *wh-ever* (e.g. *whoever*). Like other types of relative clauses in the language, FR's in Shughni generally contain the subordinator *ca*.

The structure of a free-choice free relative clause is schematized in (344).

(344) **Free-choice free relative clause: Schema**

[	<i>ar-wh</i>	(. . .)	<i>ca</i>	<i>V</i>	]	<sub>RC</sub>
	<i>arčīz</i>	<i>tu-rd</i>	<i>ca</i>	<i>fort</i>		
	whatever	you-DAT	REL	want.PRS.3SG		
	'whatever you want'					

As is common cross-linguistically (Giannakidou & Cheng 2006: 135–136), free-choice FR's in Shughni contain a special *wh*-word which appears inside the relative clause. In Shughni, these *wh*-words are built on the morpheme *ar-* 'each; every' preposed to a full *wh*-word, as in *arcarāng* 'however' (*ar* 'each' + *carāng* 'how'). Bare *wh*-words such as *čīz* 'what' and *carāng* 'how' are not used in free relatives. A full list of *wh*-words used in free-choice FR's is given in Table 12.1.

**Table 12.1:** *Wh*-words used in free-choice free relatives.

<i>arcarāng</i>	'however'
<i>arcūnd</i>	'however much/many'
<i>arčāy</i> (DIR)	'whoever'
<i>arči</i> (OBL)	'whomever'
<i>arčīz</i>	'whatever'
<i>arjoy</i>	'wherever'
<i>ar(ca)waxt</i>	'whenever'

A free relative may have a different syntactic status depending on the *wh*-form used (see van Riemsdijk 2017: 12–22 on the syntactic types of free relatives). In Shughni, FR's may behave as a noun (cf. *arčīz* 'whatever',



*arčāy* ‘whoever’) or as an adverb of various types: a manner adverb (*arcarāng* ‘however’), a temporal adverb (*arwaxt* ‘whenever’), or a locative adverb (*arjoy* ‘wherever’). Examples of each type are given in (345).

(345) **Free relative clauses of different syntactic types**

a. **Nominal type: *arčīz* ‘whatever’**

[Arčīz tu-rd ca fort]<sub>RC</sub> zi.  
 whatever you-DAT REL want.PRS.3SG take.IMPER.2SG  
 ‘Get whatever you want.’

b. **Nominal type: *arčīrd* ‘(to) whomever’**

[Arči-rd ca na-fort]<sub>RC</sub> mā-xīrt!  
 whoever.OBL-DAT REL NEG-want.PRS.3SG PROH-eat.PRS.3SG  
 ‘Whoever doesn’t want it (lit. ‘to whomever it’s not desirable), don’t eat it!’

c. **Manner adverb type: *arcarāng* ‘however’**

[Tama-rd arcarāng ca fort]<sub>RC</sub> māš=ta ik=az wi moz-ām.  
 you.PL-DAT however REL want.PRS we=FAC PREC=from DEM.OBL.M build.PRS-1PL  
 ‘However you want it, that’s how we’ll build it.’

d. **Temporal adverb type: *arcawaxt* ‘whenever’**

[Arwaxt tu-nd waxt ca sūd]<sub>RC</sub> xabar mu ki.  
 whenever you-POSS time REL become.PRS.3SG news me do.IMPER.2SG  
 ‘Let me know whenever you have time.’

e. **Locative adverb type: *arjoy* ‘wherever’**

[Arjoy=ta ca sāw-i]<sub>RC</sub> wuz=ta mis taram sām.  
 wherever=FAC REL go.PRS-2SG I=FAC also to.there go.PRS.1SG  
 ‘I’ll go wherever you go.’

Lastly, note that a pro-form is sometimes used in the matrix clause to cross-reference the entity denoted by the relative clause. In the examples above, pro-forms (underlined) are used to cross-reference the adverbial free relatives in (345c) and (345e), but no pro-forms are used to cross-reference the nominal free relatives in (345a) or (345b), or for the temporal type in (345d).

The precise circumstances under which a pro-form is used are at this point not fully clear, and the rules are apparently different from those which govern the use of resumptive pronouns in externally headed relative clauses. For

instance, whereas in externally headed relative clauses, a resumptive pronoun is required whenever the syntactic function of the head noun in the relative clause calls for the oblique case, a pro-form is not always required when the entity denoted by a free-choice free relative clause would appear in the oblique case in the matrix clause. This can be seen in (345a), where no pro-form is used despite the fact that the entity denoted by the free relative is the direct object in the matrix clause. The rules which underlie the use of pro-forms together with FR's in Shughni are thus a matter for future investigation.

### 12.1.2 Temporal adverbial clauses

Like relative clauses, (finite) temporal adverbial clauses are formed with a subordinator targeting the second-to-last position in the clause, either *ca* or *di* (both of which are glossed here SUBR for 'subordinator'). This subsection is divided into discussions on *finite* temporal adverbial clauses (Section 12.1.2.1), which take one of these subordinators, and *non-finite* temporal adverbial clauses (Section 12.1.2.2), which have no subordinating morpheme and take a verb in the infinitive.

#### 12.1.2.1 Finite temporal adverbial clauses: Subordinators *ca* and *di*

Finite temporal adverbial clauses in Shughni generally indicate simultaneous or sequential actions and are formed with one of the subordinators *ca* or *di*. The subordinators *ca* and *di* are not interchangeable, however. The latter, although compatible only with past stems, is restricted to use in adverbial clauses with future temporal reference. (This discrepancy between form and meaning – i.e. where a past stem is used in a clause with future temporal reference – is discussed in Section 10.2.2 in Chapter 10 on TAM.) Initial examples with the subordinator *di* are given in (346).

#### (346) Temporal adverbial clauses with *di*: Future events

- a. [Mu nān **di** yat]<sub>TAC</sub> māš=ta ar bozor sāw-ām.  
my mother SUBR come.PST we=FAC to market go.PRS-1 PL  
'When my mom comes, we'll go to the market.'
- b. [Tu=t xu tāt xūmne **di** wīnt]<sub>TAC</sub> lu wi-rd.  
you=2SG REFL father tomorrow SUBR see.PST say.2SG.IMP him-DAT  
'When you see your father tomorrow, tell him.'

The subordinator *ca*, for its part, is generally used for temporal adverbial clauses with past or present temporal reference. Examples illustrating these usages are given in (347).

(347) **Temporal adverbial clauses with *ca***

a. **Subordinator *ca* with past temporal reference**

[Tu=t arūd ca yat]<sub>TAC</sub> wuz=um tar magazīn rawūn vud.  
 you to.here SUBR come.PST I=1SG to store going be.PST.M  
 ‘When you came by here, I was on my way to the store.’

b. **Subordinator *ca* with present temporal reference**

[Farrukh ar Tojikistūn ca yođd]<sub>TAC</sub> xoli qurutob xīrt.  
 Farrukh to Tajikistan SUBR come.PRS.3SG only qurutob eat.PRS.3SG  
 ‘When Farrukh comes to Tajikistan, he only eats qurutob.’

Note further that the use of the subordinator *di* is the preferred way of forming adverbial clauses with future temporal reference. If a subordinate clause built on *ca* has present temporal reference, it is interpreted as a temporal adverbial clause with habitual aspect, as in (347b). If, on the other hand, a subordinate clause built on *ca* has future temporal reference (indicated by the factual particle =*ta*), then the clause is interpreted as the antecedent of a conditional.

The distinction between the use of *ca* and *di* with future temporal reference is illustrated via the examples in (348). These two examples are string-identical except for the subordinator. Example (348a), which contains *ca*, is a conditional and is felicitous, for example, in a scenario where the speaker is unsure of whether the addressee will come to Tajikistan. By contrast, example (348b), which contains *di*, is a temporal adverbial clause and is felicitous in a scenario in which the speaker is certain that the addressee will come to Tajikistan.

(348) **Past/future adverbial clauses with subordinators *ca* and *di***

a. **Subordinator *ca*: Antecedent of conditional**

[Tu=ta ar Tojikistūn **ca** yad-i]<sub>TAC</sub> māš=ta qatiyaθ pi Pomir sāw-ām.  
 you=2SG to Tajikistan SUBR come.PRS-2SG we=FAC together to Pamir go.PRS-1PL  
 ‘When/if you come to Tajikistan, we’ll go to the Pamirs together.’

b. **Subordinator *di*: Temporal adverbial clause**

[Tu=t ar Tojikistũn **di** yat]<sub>TAC</sub> māš=ta qatiyaθ pi Pomir sāw-ām.  
 you=2SG to Tajikistan SUBR come.PST we=FAC together to Pamir go.PRS-1PL  
 ‘When you come to Tajikistan, we’ll go to the Pamirs together.’

A summary of the uses of the subordinators *di* and *ca* in temporal adverbial clauses is given in Table 12.2.

**Table 12.2:** Subordinators *ca* and *di* in temporal adverbial clauses.

SUBR.	VERB STEM	TEMPORAL REF.	INTERPRETATION
<i>di</i>	Past	Future	Temporal adverbial clause
<i>ca</i>	Present	Present (habitual)	Temporal adverbial clause
	Past	Past	Temporal adverbial clause
	Present	Future	Antecedent of conditional

**12.1.2.2 Non-finite temporal adverbial clauses**

Non-finite temporal adverbial clauses do not require the use of a subordinator such as *di* or *ca*. Rather, this type of clause is built on a verb in its long infinitive form (i.e. the infinitive stem with the infinitival suffix *-ow*). Infinitival adverbial clauses must contain some additional material indicating when the eventuality expressed in the adverbial clause occurs with respect to that of the matrix clause. Generally, this is done through an adpositional phrase which takes the verb as an object (e.g. *bādi xīdow* ‘after eating’, lit. ‘after to eat’).

If the subject of the infinitival clause is the same as that of the matrix clause, then no material indicating the subject is needed in the infinitival clause. Such examples are given in (349).

**(349) Infinitival temporal adverbial clauses: Subject is the same in adverbial and main clauses**

- a. Wuz=ta [piro az xēvd-ow]<sub>TAC</sub> kitob xoy-um.  
 I=FAC before from sleep.INF-CVB book read.PRS-1SG  
 ‘I read a book before sleeping.’
- b. Wāð=en [bādi xīd-ow]<sub>TAC</sub> xu čīni-yen jām čūd=xu zinod=en wev.  
 they=3PL after eat.INF-CVB REFL bowl-PL together do.PST=and wash.PST=3PL them  
 ‘After eating they gathered their bowls and washed them.’

If, however, the subject of the infinitival clause is different than that of the matrix clause, then a possessive nominal – either an oblique pronoun or a full noun phrase immediately preceding the infinitive verb – is used to indicate the subject of the infinitival clause. Examples of this type are given in (350).

(350) **Infinitival temporal adverbial clauses: Subject is the different in adverbial and main clauses**

- a. [Piro az xīr nīst-ow]<sub>TAC</sub> boyad padam firāp-ām.  
before from sun set.INF-CVB must up.there arrive.PRS-1PL  
'Before the sun sets, we must get there.'
- b. [Bād az tu maktāb tayor čīd-ow]<sub>TAC</sub> māš=ta ukmand-aθ taškīlī kin-ām.  
after from you school completed do.INF-CVB we=FAC certain-ADV party do.PRS-1PL  
'After you finish school, we'll definitely have a party.'

Finally, the examples in (351) show the use of the locative suffix *-ti* in infinitival adverbial clauses, which indicates that the eventualities expressed in the adverbial clause and matrix clause occur simultaneously. Here, the subject of each clause must be the same.

(351) **Infinitival temporal adverbial clauses: Simultaneous actions with locative *-ti***

- a. [Tīd-ow-**ti**=yum] divī qulf ḏod.  
leave.INF-CVB-LOC=1SG door lock hit.PST  
'I locked the door upon leaving.'
- b. [Yatt-ow-**ti**] qīw.  
come.INF-CVB-LOC call.2SG.IMP  
'Call when you're coming.'
- c. [Awqot xīd-ow-**ti**] gāp mā-ḏa.  
food eat.INF-CVB-LOC word PROH-hit.2SG.IMP  
'Don't talk when eating food.'

### 12.1.3 Other subordinate clauses

This subsection examines five further types of subordinate clauses in Shughni beyond relative and temporal adverbial clauses: (i) conditional clauses, (ii) result clauses, (iii) concessive clauses, (iv) causal clauses, and (v) purposive clauses. Each type is examined in turn.

**Conditional clauses**, examined in more detail in Subsection 10.4.2.3, invariably contain the subordinator *ca*. Examples of the two fundamental types of conditional constructions, present/future and counterfactual, are given in (352). Present/future conditionals contain a verb in the present tense. (Recall that the present form of the verb *vidow* ‘be’, which appears in example (352a) below, is only used in irrealis environments such as conditionals.) On the other hand, both the antecedent and the result clauses of counterfactual conditions contain a perfect stem, optionally with the stressed suffix *-at*.

(352) **Conditional clauses with the subordinator *ca***

a. **Present/future conditional**

[Mu nān=ta (aga) ar čīd **ca** vid]<sub>COND</sub> māš=ta qatiyaθ awqot xār-ām.  
 my mom=FAC (if) at home SUBR be.3SG.IRR, we=FAC together food eat.PRS-1PL  
 ‘If my mom is home, we will eat together’

b. **Counterfactual conditional**

[Mu nān (aga) **ca** yaθč(-at)]<sub>COND</sub> māš=am ar bozor sic(-at).  
 my mother (if) SUBR come.PERF-IRR, we=1PL to market go.PERF.PL-IRR  
 ‘If my mother had come, we would have gone to the market.’

**Result clauses**, for their part, indicate the result of an eventuality expressed in the matrix clause. Generally, these clauses correspond to English constructions with the structure *so* (ADJ) *that* (. . .), *so much that* (. . .), or *such* (NOUN) *that* (. . .), as in *it was so cold that our water froze*. In Shughni result clause constructions, the matrix clause contains the adverb *disga* ‘so (much)’, and the result clause itself optionally (albeit often) begins with the complementizer *idi*. Examples are given in (353).

(353) **Result clauses with optional complementizer *idi***

a. Wuz=um disga mot vad [(**idi**) wuz=um xēvd-ow na-vārōod]<sub>RESULT</sub>.  
 I=1SG so tired be.PST.F (COMP) I=1SG sleep-INF-PURP NEG-can.PST  
 ‘I was so tired **that I couldn’t sleep**.’

b. Yu disga xuš vud [(**idi**) az xuši=yi fuk-ard-aθ ik=wam rūz yordam  
 he so happy be.PST COMP from happiness=3SG all-DAT-AUG PREC=DEM.OBL.F day help  
 čūd]<sub>RESULT</sub>.  
 do.PST  
 ‘He was so happy **that out of happiness he helped everyone that day**.’

- c. Mu ðindũn disga dārð čūd [(idi) wuz=um yičīzaθ xīd-ow na-vārðod]<sub>RESULT</sub>.  
 my tooth so.much pain do.PST COMP I=1SG nothing eat.INF-PURP NEG-can.PST  
 ‘My tooth hurt so much **that I couldn’t eat anything.**’

**Concessive clauses** express an eventuality in spite of which the eventuality in the matrix clause holds. Such constructions are generally the equivalent of clauses containing *although*, *even though*, or *even if* in English. Concessive clauses in Shughni are formed with *mis* ‘also’ and the subordinator *ca*. Examples are given in (354).

(354) **Concessive clauses with *mis* ‘also’ and the subordinator *ca***

- a. [Wuz=um Tojiki pīndz sol mis **ca** xēyd]<sub>CONCESS</sub> Tojiki-ti zūr-aθ gāp na-ðām.  
 I=1SG Tajik five year also SUBR study.PST Tajik-LOC great-ADV word NEG-hit.PRS.1SG  
 ‘**Even though I studied Tajik for five years**, I don’t speak Tajik very well.’
- b. [Tu-nd pūl mis **ca** vid]<sub>CONCESS</sub> vārði-yi xuš no-vid-ow.  
 you-POSS money also SUBR be.SUBJ.3SG can.PRS-2SG happy NEG-be.INF-NMZ  
 ‘**Even if you have money**, you not be happy.’
- c. [Yā=ta mošin mis xarīd **ca** kixt]<sub>CONCESS</sub> wam-ja=ta pūl mis rist.  
 she=FAC car also purchase SUBR do.PRS.3SG her-POSS=FAC money also remain.PRS.3SG  
 ‘**Even if she buys a car**, she will still have money.’

**Causal clauses** give the cause for the eventuality expressed in the matrix clause. Causal clauses in Shughni generally follow the matrix clause and contain the subordinator *dūndjāt* ‘because’ (likely etymologically composed of the two morphemes *dūnd-jāt* ‘so.much-for’). Clauses of this type also optionally take the complementizer *idi*, which follows *dūndjāt*.

(355) **Causal clauses with *dūndjāt* ‘because’ and optional complementizer (*idi*)**

- a. Tu-rd=ta yal vodītelski pravā na-dāk-en [dūndjāt (idi) tu=t yal haždā  
 you-DAT=FAC still driving license NEG-give.PRS-3PL because COMP you=2SG still eighteen  
 solā nist]<sub>CAUS</sub>.  
 years.old NEG.COP  
 ‘They still won’t give you a driver’s license **because you’re still not 18 years old.**’
- b. Yu=yi fuk xu examin-en bašānd-aθ siport [dūndjāt (idi) yu=yi fukwaxt  
 DEM.DIR.M=3SG all REFL exam-PL well-ADV give.PST because COMP DEM.DIR.M=3SG always  
 xu dars zūr-aθ xēyd]<sub>CAUS</sub>.  
 REFL class great-ADV study.PST  
 ‘He did well on all his tests **because he always studied well.**’

- c. Fanā odam-en pi xu maqsad na-firāp-en [dūndjat (idi) wev-and bašānd  
 many person-PL up.to REFL goal NEG-attive.PRS-3PL because COMP them-POSS good  
 motivaciya nist]<sub>CAUS.</sub>  
 motivation NEG.COP  
 ‘Many people don’t reach their goals **because they don’t have good motivation.**’

Finally, **purposive clauses** express the incentive behind the action expressed in the main clause. In Shughni, purposive clauses always follow the main clause and are built on the subordinator *lāk*, which comes either at the very beginning of the subordinate clause or after the subject (see the sentence in 356c). No other subordinators, including *ca*, are used in purposive clauses. Examples are given in (356).

(356) **Purposive clauses with the subordinator *lāk***

- a. Wuz=um xu yax-ard mam tov wam wazifā qati yordam na-čūd [lāk xubaθ  
 I=1SG REFL sister-DAT this time her task with help NEG-do.PRS.3SG SUBR PRON.EMPH  
 uxmānd sūd]<sub>PURP.</sub>  
 accustomed become.PRS.3SG  
 ‘I didn’t help my sister with her homework this time **so that she could learn herself.**’
- b. Xu zūntik-en zet ca naṣtiy-et [lāk xist vo mā-set]<sub>PURP.</sub>  
 REFL umbrella-PL take.2PL.IMP SUBR leave.PRS SUBR wet again PROH-become.2PL.PRS  
 ‘Take your umbrellas if you go out **so you don’t get wet again.**’
- c. Wāð=en taksī-ti sat [wāð lāk der mā-ken]<sub>PURP.</sub>  
 they=3PL taxi-LOC go.PST.PL they SUBR late PROH-do.PRS.3PL  
 ‘They took a taxi **so they wouldn’t be late.**’

#### 12.1.4 Complement clauses and dependent speech

*Complement clauses* are clauses which function as the argument of a predicate (Noonan 2007: 52; see also Schmidtke-Bode 2014; Schmidtke-Bode & Diessel 2017). For the purposes of Shughni, we focus here on object complement clauses – i.e. those which are the object of a predicate, often verbs related to speech or thought, such as *say*, *hear*, or *think*.

Complement clauses in Shughni, unlike the subordinate clauses examined above, do not contain the complementizer *ca*. Instead, they optionally take the complementizer *idi*, which, if present, appears at the beginning of the



complement clause. Note that this complementizer does not count for the purposes of second-position clitic placement in the complement clause; that is, any second position clitic must appear at the right edge of the syntactic phrase which follows *idi*.

Examples of sentences containing complement clauses are given in (357). Each of these examples contains a second-position clitic, either the factual enclitic *=ta* or the second-singular past-tense subject clitic *=(a)t*, and in each case the clitic appears at the right edge of the syntactic phrase following *idi*.

(357) **Examples of sentences containing complement clauses**

- a. Daler=i lūd [(**idi**) tu=ta na-yad-i]<sub>cc</sub>.  
 Daler=3SG say.PST COMP YOU=FAC NEG-COME.PRS-2SG  
 ‘Daler said (that) you aren’t coming.’
- b. Wuz=ta ar meθ x̣in-um [(**idi**) tu=t dis aqlidor]<sub>cc</sub>.  
 I=FAC every day hear.PRS-1SG COMP you=2SG so smart  
 ‘I hear everyday (that) you are so smart.’
- c. Māš fukaθ xoyiṣ kin-ām [(**idi**) tu=ta ar Toj̣ikistūn ris-i]<sub>cc</sub>.  
 we all wish do.PRS-1PL COMP YOU=FAC in Tajikistan stay.PRS-2SG  
 ‘We all hope (that) you’ll stay in Tajikistan.’

A (non-exhaustive) list of verbs which commonly take complement clauses in Shughni is given in Table 12.3.

**Table 12.3:** Verbs which take complement clauses with the complementizer *idi*.

<u>VERB</u>	<u>GLOSS</u>
<i>lūvdow</i>	‘say’
<i>fāmtow</i>	‘know’
<i>ṣ̣idow</i>	‘hear’
<i>gumūn č̣idow</i>	‘assume; reckon’
<i>f̣ikri č̣idow</i>	‘think’
<i>xoyiṣ č̣idow</i>	‘wish; hope’
<i>panimat č̣idow</i>	‘understand’

As seen in (357a), **dependent (or reported) speech** may be introduced by the verb *lūvdow*. However, it may also be indicated by the tenseless quotative particle *nāla* (glossed QUOT), which appears before the dependent speech.

This particle may either be used alone or in combination with various verbs of communication such as *lūvdow* ‘say’, *pēxtow* ‘ask’, and *qīwdow* ‘call’.

Examples of dependent speech with the particle *nāla* are given in (358). If no verb is present together with the quotative particle, as in (358a), then agreement markers such as the subject clitics used in the past tense, are illicit (hence the notion that *nāla* is tenseless). Thus, even if the temporal reference of the sentence in (358a) is past, the use of the subject clitic *⇒yi* is ungrammatical. On the other hand, if an overt verb is present in combination with *nāla*, as in (358b), then the use of agreement markers such as subject clitics is required.

(358) **Dependent speech with the particle *nāla***

- a. Yu(*⇒yi*) **nāla** wuz=*um* tu žīwǵ.  
he=3SG QUOT I=1SG you love.PERF  
‘He said/says he loved/loves me.’
- b. Yu=*yi* **qīwd** **nāla** wuz=*um* tu žīwǵ.  
he=3SG call.PST QUOT I=1SG you love.PERF  
‘He called and said he loved me.’

An interesting question regards the types of constructions used for conveying reported speech, and particularly how direct vs. indirect quotations are construed and how often each type is used. I adopt here the interpretation of these notions provided by Li (1986: 30-33). His definition is as follows: in *direct* speech, first- and second-pronouns within the quotation are co-referential with the reported speaker and addressee, respectively, while in *indirect* speech, first- and second-person pronouns within the quotation are co-referential with the reporter-speaker and the addressee of the present utterance, respectively. Compare, for instance, the following examples from English:

(359) **Direct speech:** John<sub>i</sub> said to Mary<sub>j</sub>, “You<sub>j</sub> are a great musician.”

(360) **Indirect speech:** John<sub>i</sub> told Mary<sub>j</sub> that [you<sub>\*j</sub> were a great musician].

In the example (359), which contains direct speech, the second-person pronoun in the subordinate clause containing dependent is co-referential with the addressee of the *reported* utterance, rather than the current utterance. In the indirect speech in (360), on the other hand, the second-person pronoun is co-indexed not with the addressee of the reported utterance, but rather with the addressee of the *current* utterance. Note further that in English, the use of the complementizer *that* is incompatible with direct speech. The use of complementizers with (different types

of) reported speech is a point of cross-linguistic variation (e.g., Deal 2020), but as we will see below, Shughni is similar to English in that the use of the complementizer *idi* forces an indirect-speech reading.

Reported speech has not been examined in any detail in the existing literature on Shughni, although a few examples are given by Edelman & Dodykhudoeva (2009b: 816), who claim that “constructions of indirect narration are practically not used”. While I cannot comment here on the relative frequency with which direct and indirect quotations are used in the language, the results of my fieldwork indicate that both direct and indirect reported speech are possible. Consider the ambiguous example in (361).

(361) **Ambiguity between direct and indirect speech**

John<sub>i</sub> mu-rd<sub>j</sub> nāla [wuz<sub>i/j</sub>=um dis xušrūy].  
 John me-DAT QUOT I=1SG so beautiful  
 Direct speech interpretation: ‘John told me he was beautiful.’  
 Indirect speech interpretation: ‘John told me I was beautiful.’

The example in (361), without any further context, is ambiguous between a direct speech reading and an indirect speech reading. In other words, *indexical shift* (see Deal 2020) may or may not occur. On a direct speech reading (i.e. one in which indexical shift has occurred), the first-person pronoun *wuz* in the dependent clause containing the reported speech is co-indexed with the reported speaker, *John*. On the indirect speech reading (i.e. where indexical shift has not occurred), the same first-person pronoun *wuz* is co-indexed with the speaker reporter, which in this instance is also indicated in the matrix clause with the dative *mu-rd* ‘to me’.

Not all instances of reported speech in Shughni are ambiguous in this way, however. A few aspects of reported-speech constructions appear to modulate whether indirect- and/or direct-speech interpretations are available. First, when no overt addressee is present in the matrix clause, the default interpretation for constructions with *nāla* is apparently direct speech. This is shown in the example in (362).

(362) John<sub>i</sub> nāla [wuz<sub>i/?j</sub>=um dis xušrūy].  
 John QUOT I=1SG so beautiful  
 ✓ Direct speech interpretation: ‘John said he was beautiful.’  
 ✗ Indirect speech interpretation: ‘John said I was beautiful.’

Importantly, the indirect-speech interpretation is unavailable for (362) but is available for (361), and the only difference between these two examples is that an overt addressee is present in the matrix clause of (361) but absent

in that of (362). Nonetheless, whether and to what extent the presence of an overt addressee truly modulates the availability of direct and indirect speech readings will require further investigation.

Secondly, an indirect-speech interpretation can be forced through the use of certain grammatical means. For instance, as mentioned above, the use of the complementizer *idi* in reported speech is only compatible with indirect speech. This is shown in (363). Compare this example to that in (361), which lacks *idi*, and in which both an indirect- and direct-speech interpretation are available.

(363) **Complementizer *idi* present: Indirect speech reading only**

- John mu-rd nāla **idi** [wuz=um dis xušrūy].  
 John me-DAT QUOT COMP I=1SG so beautiful  
 ✗ Direct speech interpretation: ‘John told me he was beautiful.’  
 ✓ Indirect speech interpretation: ‘John told me I was beautiful.’

Two further ways in which an indirect-speech interpretation can be forced are (i) through the use of *guyo* ‘seemingly; apparently’ in the dependent clause containing the reported speech and (ii) the use of the subordinator *ca* in the dependent clause. Examples of each are given in (364) and (365), respectively.

- (364) John nāla [wuz=um **guyo** dis xušrūy].  
 John QUOT I=1SG apparently so beautiful  
 ✗ Direct speech interpretation: ‘John told me he was beautiful.’  
 ✓ Indirect speech interpretation: ‘John told me I was beautiful.’

- (365) John nāla [wuz=um dis xušrūy **ca**].  
 John QUOT I=1SG so beautiful SUBR  
 ✗ Direct speech interpretation: ‘John told me he was beautiful.’  
 ✓ Indirect speech interpretation: ‘John told me I was beautiful.’

Although the sentences in both (364) and (365) are compatible only with an indirect-speech reading, a key difference regards the speaker’s attitude to the proposition expressed in the reported speech. In the former, which contains the adverb *guyo* ‘apparently’, the speaker has doubts about or disagrees with the proposition, while in the latter, the speaker agrees with the proposition expressed by dependent speech with the subordinator *ca*. Therefore, in addition to forcing an indirect-speech reading, the use of these two elements in clauses containing reported speech also evoke nuances in the speaker’s attitude toward the proposition. The precise mechanisms whereby these two elements simultaneously force indirect speech readings and add modal nuances are a matter for further research.

Finally, note that dependent speech in Shughni does not exhibit coordination of tenses (sometimes also referred to as *sequence of tenses*; see von Stechow 1995 and Ogiwara 1996). That is, the tense of the verb in the clause containing the reported speech is not dependent on the tense in the matrix clause. Rather, the tense of the verb in the reported-speech clause is as it was in the original (reported) utterance. This is exhibited (366).

(366) **Dependent speech as direct speech in Shughni**

Karīm=i lūd (idi) [tu boyad ðu sol kor ki].  
 Karim=3SG say.PST (COMP) you must two year work do.PRES.IRR.DEO.2SG  
 ‘Karim said (that) I/you had to work for two years.’

In this example, the verb in the main clause is in the past tense, but the verb in the subordinate (dependent speech) clause is in the present, as it was in the original utterance being reproduced. This is the case whether the quotation is interpreted as direct speech or indirect speech (as would necessarily be the case with the use of the complementizer *idi*).

### 12.1.5 Coordination of clauses

Independent finite clauses in Shughni are coordinated with three different conjunctions, all of which behave as enclitics and attach at the right edge of the final word of the first clause. These are =*xu*, =*at*, and =*atā*. Although their usage is rather similar, there are certain nuances in meaning that come with each. The conjunction =*xu* is used to express consecutive actions – i.e. actions which do not occur simultaneously. This conjunction is often used when listing a sequence of events, as shown in example (367).

(367) **Conjunction =*xu*, consecutive actions**

[Wuz=um ʃarvo xūd]=**xu** [toyd=um].  
 I=1SG soup eat.PST=and leave.PST.F=1SG  
 ‘I ate soup and (then) left.’

The conjunction =*at*, on the other hand, generally implies that the two actions or states expressed by the predicates of each clause occur simultaneously. In this sense, two clauses linked by =*at* often receive a progressive type of interpretation, as exhibited in example (368).

(368) **Conjunction =*at*, simultaneous actions**

[Wuz=um kofe biroxt]=**at** [toyd=um].  
 I=1SG coffee drink.PST=and go.PST.F=1SG  
 ‘I was drinking coffee and walking (at the same time).’

Finally, the conjunction =*atā* tends to imply a kind of contrast or, in some cases, a logical contradiction, between the first and second clauses, and for this reason may often find a more appropriate translation as ‘but’. An example is given in (369).

(369) **Conjunction =*atā*, contrast or contradiction**

[Mu nān tar America sat]=**atā** [wuz=um ar Tojikistūn red].  
 my mother to America go.PST.F=but I=1SG in Tajikistan stay.PST  
 ‘My mother went to America, but I stayed in Tajikistan.’

The three conjunctions are summarized in Table 12.4.

**Table 12.4:** Coordinating conjunctions in Shughni.

FORM	USED WITH
= <i>xu</i>	consecutive actions
= <i>at</i>	simultaneous actions
= <i>atā</i>	contrasts; contradictions

### 12.1.6 Complex clauses: Summary

This section has examined various types of complex clauses in Shughni. The language possesses a number of morphemes which serve to indicate the type of complex clause in which they are used, including at least three subordinators and one complementizer.

The morpheme *ca*, which targets the second-position of the subordinate clause and immediately precedes a finite verb (if present), is perhaps the most versatile subordinator in the language. It is used obligatorily in relative clauses, (non-future) temporal adverbial clauses, and a number of other types of subordinate clauses, including conditional and concessive clauses.

Besides *ca*, two other subordinators are used in the language, albeit with relatively restricted usage. The morpheme *di* is found only in future temporal adverbial clauses and, like *ca*, also targets the left edge of the finite verb in the subordinate clause. The morpheme *lāk*, for its part, appears as the first or second element in purposive clauses.

Unlike subordinators, the complementizer *idi* is generally optional in all the constructions where it is found. This morpheme targets the left edge of complement clauses, including dependent speech, which may also be introduced via the quotative particle *nāla*. The morpheme *idi* is also found in at least one type of subordinate clause, namely result clauses.

Subordinators and complementizers in Shughni are summarized in Table 12.5.

**Table 12.5:** Shughni complementizers.

<u>FORM</u>	<u>STATUS</u>	<u>POSITION</u>	<u>USAGE</u>
<i>idi</i>	COMP	left edge of clause (always optional)	complement, causal and result clauses
<i>ca</i>	SUBR	left edge of fin. verb (if present)	relative & temporal adverbial clauses; conditional, concessive clauses
<i>di</i>	SUBR	left edge of fin. verb (if present)	future temporal adverbial clauses (with past stems)
<i>lāk</i>	SUBR	first or second element of clause	purposive clauses
<i>nāla</i>	QUOT	before dependent speech of clause	direct speech

Finally, it was seen that Shughni makes use of three different coordinating elements, all of which behave as morphophonological clitics and attach to the right edge of the first clause. The conjunction =*xu* is used with consecutive actions, =*at* with simultaneous actions, and =*atā* with clauses which express contrast or logical contradiction.

## 12.2 Information structure and clause-level word order

Neutral word order in Shughni is Subject–Object–Verb (SOV), although, as we will see throughout this section, word order is flexible and scrambling is common. In neutral word order, indirect objects generally precede direct objects, temporal adverbs and adverbs of frequency occur between the subject and verb, and adverbs of manner appear between any objects and the verb. This gives the order in (370).

### (370) Neutral (SOV) word order in Shughni

**Subject** – ADV.FREQ – ADV.TEMP – **Ind. Object** – **Dir. Object** – ADV.MANNER – **Verb**

SOV word order is used in discourse-neutral contexts, such as in response to the question *Čiz gāp?* ‘What happened?’. An example of a sentence uttered in a discourse-neutral context is given in (371). Here, no particular element of the sentence is focused, as the entire sentence provides new information. In this context, any word order besides SOV would be infelicitous.

### (371) SOV word order in discourse-neutral context

**Context:** The speaker and addressee are sitting at the addressee’s house when a group of men approaches. The speaker, not anticipating the arrival of the men, asks the addressee what is happening.

[Māð      čorik-en]<sub>S</sub>=ta nur    [mu mošīn]<sub>O</sub> [xarīd-en]<sub>V</sub>.  
 DEM.DIR.PL man-PL=FAC    today my    car            buy.PRS-3PL  
 ‘These men are going to buy my car today.’

Deviations from SOV word order in Shughni are common, however, and are generally connected to the expression of nuances in *information structure* – that is, the way the information encoded in an utterance is partitioned and packaged in order to facilitate communication between the speaker and hearer within a given discourse context. (See Aissen, to appear, for a recent survey on information structure; Skopeteas et al. 2006 provides an extensive questionnaire designed for elicitation. The Shughni data in this section were elicited using methods described in both of these sources.)

This section looks at the ways in which two information structural notions – topic and focus – interact with word order in Shughni. It will be seen below that both topicalized and focused elements commonly, but not always,



appear in clause-initial position, which may result in non-canonical (scrambled) word order, particularly in the case of topicalized and focused objects. Further, there are clear differences in Shughni with respect to the behavior of two subtypes of each of these information structural notions: contrastive and non-contrastive topics exhibit important syntactic distinctions from one another, as do contrastive and non-contrastive foci. In what follows, each of these information structural notions, as well as their (non-)contrastive subtypes, will be examined in turn. Section 12.2.1 looks at **topic** and Section 12.2.2 at **focus**. A **summary** is provided in Section 12.2.3.

Note that word order in Shughni could easily constitute the topic of an entire dissertation (for a book-length, corpus-based work on word order in Persian, a language related to Shughni which also displays basic SOV order, see Roberts 2009). This section, however, limits itself primarily to an investigation of how information structural considerations affect the ordering of *core verbal arguments*, including subjects, direct objects, and indirect objects, although the placement of adverbs is mentioned when relevant. Future studies on word order in Shughni might explore the placement of other elements such as prepositional phrases in more detail. These studies could benefit greatly from an extensive corpus of naturally occurring Shughni speech, which, to my knowledge, does not yet exist.

### 12.2.1 Expression of topic in Shughni

The *topic* (glossed TOP) of an utterance may be thought of as the entity it is about. The notion of topic stands in contrast to that of *comment*, which refers to that which is said about the topic (for a more formal definition of topic and comment, see Gundel 1988: 210). In a given utterance, the topic-comment distinction may or may not correspond to that of subject-predicate, and as will be discussed below, the degree of correlation between subjecthood and topichood varies from language to language.

As an initial example of topic, consider the English sentence *As for Anna, she will arrive tomorrow*. Here, the topic of the sentence, *Anna*, has been placed inside a special syntactic phrase *as for* . . . , which in English is dedicated to the expression of topics. The comment of this utterance, for its part, is the portion which follows the topic and relays information about it – in this case, that the topic’s referent will be arriving the following day. This is schematized in (372).

(372) As for [Anna]<sub>TOP</sub>, [she will arrive tomorrow]<sub>COMMENT</sub>.

Topics are not always packaged as explicitly as in the example above, however. Cross-linguistically, a typological distinction is drawn between *subject-prominent* and *topic-prominent* languages (see, e.g., Li & Thompson 1976 and Gundel 1980). In the former type, which includes English, the topic of a clause is often simply its subject, and subject-promoting operations such as passivization are commonly used to maintain the subject-topic correspondence.

On the other hand, in *topic-prominent* languages, such as Mandarin (e.g., Li & Thompson 1976; Lapolla 2009) and Persian (Rasekhi 2018: 27), topichood is not so closely correlated with subjecthood. In this type of language, other means such as scrambling and specialized morphology earmarked for topics – as in the Japanese suffix *-wa*; cf. Vermeulen 2013 and references therein – are used to encode topichood.

Shughni displays many of the characteristics commonly associated with topic-prominent languages. In particular, although the language lacks specialized morphology dedicated to marking topics, topicalization is commonly indicated through word order, with topics tending to appear at the left edge of the clause, followed by the comment.<sup>5</sup>

Constituents of various grammatical functions, including subjects, objects, and adverbs, are capable of being topicalized via movement to clause-initial position. When direct and indirect objects are topicalized in this way, the subject, if present, occurs later in the clause, resulting in scrambled O(S)V word order. Examples of scrambling due to a topicalized object are shown in (373), with a topicalized direct object (373a) and a topicalized indirect object (373b) both shown. For each example, a context is provided, and a first sentence is given to establish that the fronted element in the second sentence is the topic.

(373) **Topicalized objects: O(S)V word order**

a. **Fronted Direct-Object Topic**

**Context:** The speaker, a tour guide, is showing a group of tourists around the city of Dushanbe. They stop at a restaurant serving *qurutob*, one of the national dishes of Tajikistan. The speaker brings the group's attention to the dish and says the following:

---

<sup>5</sup>Despite this lack of dedicated topic morphology, it is worth pointing out that because topics most often occur in first position, and because many sentences in Shughni contain second-position clitics, such as the factual enclitic *=ta* and past-tense subject clitics, the topic is often separated from the rest of the clause by a second-position clitic. This means that there is often overt morphology separating topic and comment, even if its primary purpose is not to mark topic. A similar point is made for topics in Guarani by Estigarribia (2020: 283), which also uses second-position clitics but no specialized topic morphology. The relationship between second-position clitics and topicalization in Shughni is an intriguing topic for future research.

(Yam qurutob.) [Mi]<sub>TOP</sub>=ta (māš) beḫdi be-gūḫt-aθ xār-ām.  
 DEM.DIR qurutob DEM.DIR.M=FAC we mostly without-meat-ADV eat.PRS-1PL  
 ‘This is qurutob. We usually eat it without meat.’

**b. Fronted Indirect-Object Topic**

**Context:** The speaker and addressee are sitting on a bench at school. They see a friend of theirs, Davlatjon, walk by proudly carrying a piece of paper in his hand. The speaker says to the addressee:

(Yida Davlatjūn.) [Di]<sub>TOP</sub>-rd=i nur direktur iftixonūmā dākt.  
 PRSV.MED Davlatjon him-DAT=3SG today director certificate give.PST  
 ‘There’s Davlatjon. The director gave him a certificate today.’

Another feature of Shughni commonly ascribed to topic-prominent languages is that passive-like constructions (Section 11.1) are not a prominent feature of the grammar and not a common means of placing or keeping topics at the front of a clause. Nonetheless, continuing topics which are semantically patients commonly appear in passive-like constructions of the kind in (374) (cf. example (273) and the typological discussion in Section 11.1). Here, there is no overt or implied subject, and the subject displays third-plural agreement. The topicalized object remains in the oblique case and appears at the front of the clause. Note that the examples in (374a) and (374b) are felicitous in the exact same contexts as those given for (373a) and (373b), respectively.

**(374) Topicalization through 3PL-passive constructions**

**a. Direct-Object Topic in 3PL-passive construction (compare 373a)**

(Yam qurutob.) [Mi]<sub>TOP</sub>=ta beḫdi be-gūḫt-aθ xen.  
 DEM.DIR qurutob DEM.OBL.M=FAC mostly without-meat-ADV eat.PRS.3PL  
 ‘This is qurutob. It’s usually eaten without meat.’

**b. Indirect-Object Topic in 3PL-passive construction (compare 373b)**

(Yida Davlatjūn.) [Di]<sub>TOP</sub>-rd=en nur iftixonūmā dākt.  
 PRSV.MED Davlatjon him-DAT=3PL today certificate give.PST  
 ‘There’s Davlatjon. He was given a certificate today.’

A final point to be addressed regarding topichood in Shughni concerns an important distinction between topics which are *contrastive*, i.e. those which stand out among a handful of contextually salient alternatives, and those which are *non-contrastive*, for which no such set of contextually salient alternatives exists. **Non-contrastive topics**, on the one hand, are eligible to undergo right-dislocation, as illustrated in the examples in (375). Here, the

comment precedes the topic, and in the case of subject topics, the result is scrambled (VS) word order. Like the examples above, a context is provided for each example in (375) in order to establish the right-dislocated NP as the topic. Right-dislocation is most often used when the comment is uttered emphatically.

(375) **Right dislocation of a continuing topic**

- a. **Context:** Karim and Alex are searching through a streaming service trying to find a movie to watch. As Alex flips through the options, he notices a movie that looks good. He suggests this one to Karim by saying *ik mam kino čisām o?* ‘Shall we watch this movie?’. Karim responds:

Ūn, [lap bašānd]<sub>COMM</sub> [yid]<sub>TOP</sub>!  
yes very good DEM.DIR  
‘Yes, it’s really good, that one!’

- b. **Context:** Zuhro and Mumina are having lunch and reminiscing on good times they had in high school. They are talking about their favorite teachers, and Zuhro brings up one of her favorites, saying *Mullimā Mavjūdā dis zūr vad. Wam-and=en wam dars-en fukwaxt interesni vad.* ‘Teacher Mavjuda was so great. Her classes were always very interesting.’ Mumina agrees and responds:

Ūn, [lap zūr vad]<sub>COMM</sub> [yā]<sub>TOP</sub>!  
yes, very great be.PST.F she  
‘Yes, she was very great indeed!’

**Contrastive topics** (glossed CTR.TOP), on the other hand, are ineligible to undergo right dislocation and *invariably* appear in clause-initial position. An initial example of contrastive topics in Shughni is given in (376). Here, the two contrasted topics, *wāḍ tarelka-yen* ‘those plates’ and *wāḍ čīni-yen* ‘those bowls’, are in separate clauses (delineated by large square brackets).

(376) **Contrastive topics: Initial example (adapted from Aissen to appear: 36)**

**Context:** After a party, Shams and his father are cleaning up the kitchen. Shams sees that a group of dishes which were in the living room are not there anymore. He asks his dad *Wāḍ virek-en=en kādand?* ‘Where are those dishes?’ His father responds:

- a. [ [Wāḍ tarelka-yen]<sub>CTR.TOP=en</sub> ar rakovīna=yat ] [ [wev čīni-yen]<sub>CTR.TOP=um</sub> stol-ti ribūd. ]  
DEM.DIR.PL plate-PL=3PL in sink=and DEM.OBL.PL bowl-PL=1SG table-LOC put.PST  
‘The plates are in the sink, and the bowls I put on the table.’
- b. # [ [Wāḍ tarelka-yen=en]<sub>CTR.TOP</sub> ar rakovīna=yat ] [ wuz=um [wev čīni-yen]<sub>CTR.TOP</sub> stol-ti ribūd. ]  
DEM.DIR.PL plate-PL=3PL in sink=and I=1SG DEM.OBL.PL bowl-PL=3PL table-LOC put.PST

The ordering in (376a), where each contrastive topic appears first in its respective clause, is obligatory. This is despite the fact that the first topic is the grammatical subject, while the second topic is grammatically the direct object. The alternative order in (376b), in which an overt non-topical subject is placed in first position of the second clause, is infelicitous.

Two further examples containing contrastive topics are given in (377) and (378). These examples illustrate the importance of context in establishing which piece of information is the topic in a given sentence. The same noun phrases which serve as contrastive topics in (377) are the focus in (378), and vice versa. That which determines which noun phrases serve as topics, and thus which must appear first in their respective clauses, is the context.

(377) **Contrastive indirect-object topics**

**Context:** Omina and her family are moving abroad and are unable to take many of their belongings with them. A few days before they move, two of Omina's friends, John and Sayora, come to the house to say goodbye. Omina plans to give each person a few things. Her mom asks her what she will give to each person, and Omina responds:

- a. [ [John-ard]<sub>CTR.TOP=ta</sub> [wev dek-en]<sub>FOC</sub> dāk-um=at ] [ [Sayora-yard]<sub>CTR.TOP=ta</sub> [xu kitob-en]<sub>FOC</sub> dāk-um ].  
       John-DAT=FAC DEM.OBL.PL pot-PL give-1SG=and Sayora-DAT=1SG REFL book-PL give-1SG  
       'To John I will give the pots and to Sayora I will give my books.'
- b. # [ [Wev dek-en]<sub>FOC=ta</sub> [John-ard]<sub>CTR.TOP</sub> dāk-um=at ] [ [xu kitob-en]<sub>FOC=ta</sub> [Sayora-yard]<sub>CTR.TOP</sub> dāk-um ].  
       DEM.OBL.PL pot-PL=FOC John-DAT give-1SG=and REFL book-PL=FAC Sayora-DAT give-1SG

(378) **Contrastive direct-object topics**

**Context:** Omina and her family are moving abroad and are unable to take many of their belongings with them. Omina's mom needs to get rid of some pots, while Omina needs to get rid of some of her books. Her mom suggests Omina give these things away to some of her friends. Omina agrees with the idea, and her mom asks her who she will give each item to.

- a. [ [Wev dek-en]<sub>CTR.TOP=ta</sub> [John-ard]<sub>FOC</sub> dāk-um=at ] [ [xu kitob-en]<sub>CTR.TOP=ta</sub> [Sayora-yard]<sub>FOC</sub> dāk-um ].  
       DEM.OBL.PL pot-PL=FOC John-DAT give-1SG=and REFL book-PL=FAC Sayora-DAT give-1SG  
       'The pots I'll give to John, and my books I'll give to Sayora.'
- b. # [ [John-ard]<sub>FOC=ta</sub> [wev dek-en]<sub>CTR.TOP</sub> dāk-um=at ] [ [Sayora-yard]<sub>FOC=ta</sub> [xu kitob-en]<sub>CTR.TOP</sub> dāk-um ].  
       John-DAT=FAC DEM.OBL.PL pot-PL give-1SG=and Sayora-DAT=1SG REFL book-PL give-1SG

In each context, the speaker, Omina, is giving some items to two of her friends. In (377), the identity of the friends is already within the information shared by speaker and addressee (i.e. within the *common ground*; see Aissen, to appear: 3), whereas the identity of the items given to each friend is not. Hence, the topics of the clauses in this example are noun phrases denoting each of the friends; these noun phrases obligatorily occur before the focused information.

In (378) the situation is reversed. Here, it is the identity of the items given to Omina's friends which are part of the common ground in (378), and the identity of the friends is not yet part of the common ground. Here, therefore,

the noun phrases denoting the items are the topics of their respective clauses.

Ultimately, it has been seen in this section that the notion of topic affects clause-level word order in Shughni in that topics, regardless of their grammatical function, tend to appear at the beginning of the clause. Topicalized direct and indirect objects routinely appear first in the clause and are followed by the subject, if present. Furthermore, whereas non-contrastive topics are able to undergo right dislocation and appear after the comment, contrastive topics must appear clause-initially. In the next subsection, it will be seen that the notion of focus interacts with word order in similar ways as topic, and that as with contrastive and non-contrastive topic, there are important differences between the behavior of contrastive and non-contrastive foci.

### 12.2.2 Expression of focus in Shughni

*Focus* (glossed FOC) refers to the part of an utterance which contributes new information to the discourse context. An utterance's focus stands in contrast to old or presupposed information, often referred to as the *background*. As an initial illustration of focus, consider the exchange in (379). Note that context is crucial for determining which element(s) of an utterance are focused. *Wh*-questions are particularly useful in this regard, and for many examples, including the initial example here, a *wh*-question is part of the context.

#### (379) Focus: Initial example

**Context:** There was an English competition at a school in Khorog, the winner of which won a trip to Dushanbe to compete at the national level. A few months after the national competition occurred, one student, not even aware of who participated in the competition, asked another student who went to Dushanbe (with the question in 379a). The other student's answer is in (381b).

- a. Čāy        ar Dūšanbi    sut?  
     who.DIR to Dushanbe go.PST.M  
     'Who went to Dushanbe?'
- b. [NEKRŪZ]<sub>FOC</sub> ar Dūšanbi    sut.  
     Nekruz        to Dushanbe go.PST.M  
     '**Nekruz** went to Dushanbe.'

In this example, the fact that someone went to Dushanbe is (pre-)supposed in the *wh*-question, and this information can thus be considered the background in Speaker B's response. The missing piece of information being requested by Speaker A is the identity of the person who went, and the portion of the answer which fills in this information gap, namely *Nekrūz*, is the focus.

As with the expression of topic, Shughni does not possess special morphology dedicated to marking focus. Rather, focus is marked through a combination of prosodic and syntactic factors. My observations indicate that all focused constituents receive sentence-level prosodic prominence, although the phonetic nature of this prominence remains to be investigated. In discourse-neutral contexts – i.e. broad focus, as discussed below – where no particular element of the clause is focused, sentence-level stress occurs on the pre-verbal constituent. Note that small caps are used in the sentence in (381b) and throughout the rest of the discussion on information structure to indicate sentence-level stress in focused constituents.<sup>6</sup>

From a syntactic perspective, there is a clear difference in the behavior of contrastive and non-contrastive foci. Whereas non-contrastive foci obligatorily remain in-situ, contrastively focused elements are optionally fronted to clause-initial position. The behavior of both types of foci is elaborated upon below. Non-contrastive focus is examined in Section 12.2.2.1 and contrastive focus in Section 12.2.2.2.

### 12.2.2.1 Non-contrastive focus

The example in (379) exhibits non-contrastive focus (also called *information focus*), in which the focused information does not stand in contrast to a small set of contextually salient alternatives. In Shughni narrow focus (i.e. when focus is on an element which is smaller than the entire clause), elements which are focused non-contrastively must remain in-situ.

In the example in (379), focus is narrow, as only the subject of the clause *Nekrūz* contributes new information to the discourse context. Further examples of narrow focus are provided in (380) and (381). In the former the direct object is focused, and in the latter the temporal adverb is focused. For each example, a *wh*-question is provided to establish which piece of information in the answer is focused.

#### (380) Narrow Focus: Direct object focused

- a. Nekrūz=i čīz xūd? (*What did Nekruz eat?*)

---

<sup>6</sup>An examination of the precise acoustic features which make up sentence-level stress in Shughni is beyond the scope of this dissertation. However, see Taheri-Ardali & Xu 2012 and Taheri-Ardali et al. 2014 for an examination of the acoustic properties of sentence-level stress in Persian, which seems to display many of the same features as in Shughni.

- b. Nekrüz=i [LAPŠĀ]<sub>FOC</sub> xūd.  
 Nekruz=3SG noodles eat.PST  
 ‘Nekruz ate **noodles**.’

c. **Infelicitous word order for 380b: OSV**

- # [LAPŠĀ]<sub>FOC</sub>=yi Nekrüz xūd.  
 noodle=3SG Nekruz eat.PST

(381) **Narrow Focus: Temporal adverb focused**

- a. Nekrüz **cawaxt** ar Dūšanbi tūyd? (*When did Nekruz go to Dushanbe?*)

- b. Nekrüz [BIYOR]<sub>FOC</sub> ar Dūšanbi tūyd.  
 Nekruz yesterday to Dushanbe go.PST.M  
 ‘Nekruz went to Dushanbe **yesterday**.’

c. **Infelicitous word order for 380b: Temporal adverb precedes subject**

- # [BIYOR]<sub>FOC</sub> Nekrüz ar Dūšanbi tūyd.  
 yesterday Nekruz to Dushanbe leave.PST.M

In (380), in which the direct object is focused, prosodic prominence is as it would be in a discourse-neutral (broad-focus) context, as the direct object is immediately pre-verbal. In (381), however, it has been retracted to a marked position: it is on the temporal adverb *biyor* ‘yesterday’, which is not-preverbal. Moreover, as shown by the third example (c) of both (380) and (381), non-canonical word order in these information-focus contexts is infelicitous. Although the subject in both examples may be dropped, the focused element cannot be fronted if the subject is overtly realized.

Non-canonical word order is likewise infelicitous in broad focus, an example of which was already given in (370) in the initial presentation of neutral word order. Another example of broad focus is provided in (382). A context is provided which leads up to the question meaning *What happened?* in (382a). The answer in (382b) contains neutral word order (SUBJECT – ADVERB – VERB) with stress on the immediately pre-verbal constituent, in this case the prepositional phrase [ar Dūšanbi] ‘to Dushanbe’. The sentences in (382c), which contain infelicitous word orders for this context, illustrate the idea that deviations from basic SV word order are prohibited in broad focus.<sup>7</sup>

<sup>7</sup>In some clauses which appear to have broad focus, the verb may precede the adverb, as in *Nekrüz tūyd ar Dūšanbi* ‘Nekrüz WENT to Dushanbe’. Here, the stress occurs on the verb. It is unclear what factors determine the ordering of verb and adverb in these cases. However, the same phenomenon occurs in Persian (e.g. Roberts 2009), and this is an important point to be examined in a more in-depth study on Shughni word order.



(382) **Broad Focus**

**Context:** Inam and Ali are walking by Nekruz's house, and they see several of his family members gathered outside talking. They walk by often and don't typically see such a scene, so Inam asks Ali:

- a. Čīz gāp? (*What happened?*)
- b. [Nekrūz ar Dūšanbi tūyd.]<sub>FOC</sub>  
Nekruz to Dushanbe go.PST.M  
'Nekruz left for Dushanbe.'
- c. **Infelicitous word orders in this context (examples):**
  - # Ar Dūšanbi tūyd Nekrūz. (Adv. – Verb – Subj.)
  - # Tūyd Nekrūz ar Dūšanbi. (Verb – Subj. – Adv.)

12.2.2.2 **Contrastive focus**

In contrastive focus, like in information focus, sentence-level stress invariably occurs on the focused piece of information. However, unlike in information focus, contrastively focused elements are optionally fronted. In this sense, Shughni displays important similarities to Persian, in which scrambled objects may also appear either in-situ or clause-initially (Karimi 2003; Rasekhi 2018).

The examples in (383) offer an initial look at contrastive focus (specifically, *corrective focus*) in Shughni. A context is provided, after which the example in (383a) provides the initial utterance, a portion of which (namely, the temporal adverb *biyor* 'yesterday') will subsequently be corrected. The following two examples exhibit the two possible word orders in corrective focus: the in-situ variant (383b) and the variant in which the temporal adverb is moved to clause-initial position (383c). The two sentences in (383d), for their part, show infelicitous word orders for this context, in which the contrastively focused element is neither in-situ nor in fronted position.

(383) **Corrective focus: Adverb focused**

**Context:** Sattor and Mirzo are telling a story to Karim. Both Sattor and Mirzo are familiar with the story, which is about Andrey. However, as Sattor begins to tell the story (with the sentence in 383a), he mixes up a few details and is subsequently corrected by Mirzo (with either one of the sentences in 383b or 383c).

- a. **Initial Utterance**  
Andrey=i biyor xu mošin parōd.  
Andrey=3SG yesterday REFL car sell.PST  
'Andrey sold his car yesterday.'

b. **Corrective focus in-situ (✓)**

Nay, Andrey=i [AĤĪB]<sub>FOC</sub> xu mošin parðod.  
 No, Andrey day.before.yesterday REFL car sell.PST  
 ‘No, Andrey sold his car **the day before yesterday**.’

c. **Corrective focus with movement (✓)**

Nay, [AĤĪB]<sub>FOC</sub>=i Andrey xu mošin parðod.  
 No, day.before.yesterday=3SG Andrey REFL car sell.PST  
 ‘No, **the day before yesterday** Andrey sold his car.’

d. **Infelicitous word orders for this context (X)**

Nay, parðod=i Andrey [AĤĪB]<sub>FOC</sub> xu mošin. (V – S – Adv – O)  
 Nay, xu mošin=i Andrey [AĤĪB]<sub>FOC</sub> parðod. (O – S – Adv – O)

The examples in (384) exhibit corrective focus on other elements of the utterance in (383a). The same context given above applies here. Example (384a) shows corrective focus on the subject, and examples (384b) and (384c) show focus on the direct object (without and with movement, respectively).

(384) **Corrective focus: Subject, direct-object focus**

**Context:** Same as (383).

a. **Corrective focus: Subject focused**

Nay, [KARĪM]<sub>FOC</sub>=i biyor xu mošin parðod.  
 no Karim=3SG yesterday REFL car sell.PST  
 ‘No, **Karim** sold his car yesterday.’

b. **Corrective focus: Direct object focused (no movement)**

Nay, Andrey=i biyor [XU MOTOCĪC]<sub>FOC</sub> parðod.  
 no Andrey=3SG yesterday REFL motorcycle sell.PST  
 ‘No, Andrey sold his **motorcycle** yesterday.’

c. **Corrective focus: Direct object focused (with movement)**

Nay, [XU MOTOCĪC]<sub>FOC</sub>=i Andrey biyor parðod.  
 no REFL motorcycle=3SG Andrey yesterday sell.PST  
 ‘No, it was his **motorcycle** Andrey sold yesterday.’

Note that the order SOV in (384a), which shows subject focus, is the only available word order, as the subject appears in clause-initial position in neutral word order. For the focusing of the direct object, either the in-situ variant (384b) or the fronted variant (384c) are felicitous.

The fronting of a verb is generally reserved for instances of corrective focus in which the polarity of the clause is being corrected – i.e. when the speaker indicates that an action did in fact occur (or didn’t in fact occur), contrary to

information previously introduced into the discourse context. Examples are given in (385) and (386). The context provided in (385) serves both sets of examples.

(385) **Corrective focus on (affirmative) polarity: Verb fronted**

**Context (for both 385 and 386):** A group of five friends has just had dinner at a restaurant. They decide to have each person pay individually – i.e. only for what they ordered. One of them takes the bill and starts making the calculations. As she does so, she says certain statements out loud. A friend disagrees and makes the following corrections.

a. **Initial utterance**

...=at Gulnoza=yi pitsā xūd.  
 ...=and Gulnoza=3SG pizza eat.PST  
 ‘... and Gulnoza ate pizza.’

b. **Corrective focus (polarity): Verb fronted**

[NA-XŪD]<sub>FOC</sub>=i Gulnozā pitsā!  
 NEG-eat.PST=3SG Gulnoza pizza  
 ‘Gulnoza **didn’t** eat pizza.’

(386) **Corrective focus on (negative) polarity: Verb fronted**

a. **Initial utterance**

...=at Mahram=i kampot na-biroxt.  
 ...=and Mahram=3SG compote NEG-drink.PST  
 ‘... and Mahram didn’t drink compote.’

b. **Corrective focus (polarity): Verb fronted**

[BIROXT]<sub>FOC</sub>=i Mahram kampot!  
 drink.PST=3SG Mahram compote  
 ‘Mahram **did** drink compote!’

Finally, note that if a clause contains both a contrastive topic and a contrastively focused constituent, the contrastive topic must precede the contrastively focused element. This order, which is common cross-linguistically (Aissen (to appear: 23), is schematized in (387), and an example is given in (388).

(387) **Order of contrastive topic and contrastive focus**

[ [ CONTRASTIVE TOPIC ] [ CONTRASTIVE FOCUS ] . . . ]

(388) **Contrastive topic and contrastive focus in the same clause**

**Context** (same as 383): Sattor and Mirzo are telling a story to Karim. Both Sattor and Mirzo are both familiar with the story, which is about Andrey. However, as Sattor begins to tell the story, he mixes up a few details and is subsequently corrected by Mirzo. Sattor utters the sentence in (388a) to begin the story, but he is then corrected by Sattor with the sentence in (383b).

a. **Initial utterance**

Andrey=i xu mošīn biyor parðod.  
 Andrey=3SG REFL car yesterday sell.PST  
 ‘Andrey sold his car yesterday.’

b. **Contrastive topic and contrastive focus together**

Nay, [xu mošīn]<sub>CTR.TOP</sub>=i [Aǰīb]<sub>CTR.FOC</sub> parðod=atā [xu motocīc]<sub>CTR.TOP</sub>=i [BIYOR]<sub>CTR.FOC</sub>  
 no REFL car=3SG day.before.yesterday sell.PST=and REFL motorcycle-3SG yesterday  
 parðod  
 sell.PST  
 ‘No, he sold his car the day before yesterday. His bike he sold yesterday.’

### 12.2.3 Information structure and clause-level word order: Summary

Basic word order in Shughni is SOV, with temporal and frequency adverbs appearing between the subject and object, and manner adverbs appearing between object and verb. Although word order in the language is flexible, non-canonical (i.e. non-SOV) orders do not occur at random, but rather are intimately linked to the expression of information structural notions including topic and focus.

Topics in Shughni tend to occur in clause-initial position. Arguments in the direct case (i.e. subjects) and arguments in the oblique case (i.e. objects), as well as adverbs, are all eligible to appear in clause-initial position when topicalized. However, whereas non-contrastive topics, such as a topic which continues throughout the narration a series of events, may also appear in clause-final position (via right dislocation), contrastive topics obligatorily occur clause-initially.

A similar discrepancy occurs for different types of focused elements in Shughni. In particular, non-contrastively focused elements (i.e. the focused element in instances of information focus) must occur in-situ. On the other hand, contrastively focused elements, such as in corrective focus, may either occur in-situ or in clause-initial position. All focused elements in the language, however, receive primary sentence stress, which targets the preverbal position in broad focus (i.e. when no particular element of the clause is focused).

Finally, a cross-linguistic tendency regarding the ordering of topics and focused elements was noted. Specifically, in cases where there is both a contrastive topic and a contrastively focused constituent in the same clause, in Shughni, as in most other languages, the contrastive topic must appear before the contrastively focused element.

## 12.3 Questions

This section examines general question formation and question-related phenomena. The first two subsections deal with the fundamentals of question formation in Shughni: Section 12.3.1 looks at ***wh*-questions**, and Section 12.3.2 examines **polar questions**. The final two subsections turn to special types of questions in Shughni which, to my knowledge, have not been examined in any detail in the previous literature on the language. Specifically, Section 12.3.3 discusses **echo questions**, in which different grammatical strategies are used for *wh*-echo questions and polar echo questions. Section 12.3.4 then looks at three particles commonly used in questions, including two **tag-question particles** and the **emphatic particle *ku***. Section 12.3.5 offers a brief summary.

### 12.3.1 *Wh*-questions

Shughni is a *wh-in-situ* language. Or more precisely, under neutral conditions, *wh*-words appear in the same position in the sentence as their non-*wh* counterparts. The examples in (389) exhibit *wh*-questions with neutral word (i.e. *wh*-in-situ) word order:

(389) **Neutral word order in *wh*-questions**

- a. Čāy    biyor    tar kor    sut?  
    who.DIR yesterday to work go.PST.M  
    ‘Who went to work yesterday?’
- b. Tu=t    cawaxt tar kor    sut?  
    you=2SG when to work go.PST.M  
    ‘When did you go to work?’
- c. Tu=t    biyor    tar k̄ā    sut?  
    you=2SG yesterday to where go.PST.M  
    ‘Where did you go yesterday?’

However, word order in *wh*-questions, like in statements, may be scrambled for the purpose of expressing nuances in topic and focus. For instance, contrastive topics in *wh*-questions are invariably fronted, just as they are in statements. Examples of fronted contrastive topics resulting in scrambled word order in *wh*-questions are shown in (390). Note that each example in (390) corresponds to an example with neutral word order in (389).

(390) **Scrambling in *wh*-questions: Contrastive topics**

- a. **Context:** I know who went to work *today* (CTR.TOP), but I need to know who went *yesterday* (CTR.TOP).

[Biyor]<sub>CTR.TOP</sub> tar kor čāy sut?  
 yesterday to work who.DIR go.PST.M  
 ‘Yesterday who went to work?’

(cf. neutral order in 389a)

- b. **Context:** I know who when you went to *school* (CTR.TOP), but I need to know when you went to *work* (CTR.TOP).

[Tar kor]<sub>CTR.TOP</sub> =at (tu) cawaxt sut?  
 to work=2SG (you) when go.PST.M  
 ‘When did you go to work?’

(cf. neutral order in 389b)

Note further that in cases of scrambling due to contrastive topics, the *wh*-word tends to appear in pre-verbal position, as in the examples in (389). This may be related to the fact that this is the default position for focus, and *wh*-words in this position naturally receive sentence-level prosodic prominence, although this should be examined in future research.

In addition to scrambling due to the fronting of contrastive topics, word order in *wh*-questions may be scrambled due to the fronting of the *wh*-word itself. This is often done in scenarios in which the addressee has not understood the *wh*-question the first time it was uttered. This may be interpreted as a type of contrastive focus, because, as discussed in Section 12.2.2.2, only *contrastive* foci are eligible to be fronted. Nonetheless, this idea, like the placement of *wh*-words in questions with contrastive topics, is also an important point for future investigation.

(391) **Fronted *wh*-word in *wh*-question**

**Context:** The speaker has already uttered the question in (391) with neutral word order once (as in 389c), and the addressee did not understand.

[Tar kã]<sub>FOC</sub> =yat biyor sut?  
 to where=2SG yesterday go.PST.M  
 ‘(I said) where did you go yesterday?’

(cf. neutral order in 389c)

Shughni allows the use of multiple *wh*-words in a single question. As in English, the use of multiple *wh*-words in a question elicits a “pair list” response. Examples of questions with two *wh*-words are given in (392a) and (392b).

(392) *Wh*-questions with two *wh*-words

- a. **Context:** Sattor threw a party last night. When telling his co-workers about the next day, he notes that people either arrived very late or very early. One of his co-workers is curious and asks which people arrived late and which arrived early.

**Question:**

Čāy cawaxt yat?

who when arrive.PST

‘Who arrived when?’

**Ex. Answer:**

Maryam=at Ali=yen tez-aθ yat. Davlat=at Shahlo=yen der-aθ yat.

Maryam=and Ali=3PL early-AUG come.PST Davlat=and Shahlo=3PL late-AUG come.PST

‘Maryam and Ali arrived early; Davlat and Shahlo arrived late.’

- b. **Context:** Muyassar is preparing dinner for some friends. They day before, while they are all in class, she gathers them together and tells them she will prepare two dishes: a vegetarian dish and a meat stew. She asks what her friends plan to eat so she can buy the right amount of ingredients.

**Question:**

Čāy=ta čiz xīr-t?

who.DIR=FAC what eat.PRS-3SG

‘Who’s going to eat what?’

**Ex. Answer:**

Maryam=at Ali=ta salat xen. Davlat=at Shahlo=ta šarvo xen.

Maryam=and Ali=FFAC salad eat.PRS.3PL Davlat=and Shahlo=FAC soup eat.PRS.3PL

‘Maryam and Ali will eat salad; Davlat and Shahlo will eat soup.’

Moreover, it is possible to have *wh*-questions with more than two *wh*-words. Such questions still elicit a pair list as a response, as in (393).

(393) *Wh*-question with three *wh*-words

**Context:** Zuhro runs a food distribution company. Recently she’s been getting complaints from her customers and needs to figure out why. She wants to check her records to make sure they are correct. During a meeting with her employees, she asks them who delivered which items to which place.

Čāy=i čiz tar kā firêpt?

who.DIR=3SG what to where deliver.PST

‘Who delivered what where?’

**Ex. Answer:** Maryam delivered chicken to Somoni Street; Ali delivered rice to the market; Davlat delivered fruit to Rudaki Plaza; etc.

### 12.3.2 Polar questions

Polar questions in Shughni are formed with the polar-question particle *(y)o* (glossed PQ), which is used at the end of a polar question. The allomorph *yo*, with an initial palatal glide, occurs when following a vowel. Examples with this particle are given in (394):

(394) **Polar-question particle *o***

- a. Awqot=at taylor čūd o?  
 food=2SG prepare do.PST PQ  
 ‘Did you finish making the food?’
- b. Wam qawmiyot mis yamand zindagi-yen o?  
 her relatives also there live-3PL PQ  
 ‘Do her relatives live there too?’

This same particle is also used in embedded questions, as in (395):

(395) **Polar-question particle *o* in embedded clause**

- Wāð=en mu peχct, tu=ta sāw-i yo?  
 they=3PL me ask.PST you=FAC go.PRS-2SG PQ  
 ‘They asked me whether you will go.’

Note that *(y)o* is also used to mean ‘or’, and in both matrix questions and embedded questions, an equally felicitous way to mark a polar question is to use *(y)o nay* ‘or no’ in the place of the interrogative particle *o*:

(396) **Polar questions with *(y)o nay* ‘or no’**

- a. Awqot=at taylor čūd o nay?  
 food=2SG prepare do.PST or no  
 ‘Did you finish making the food?’
- b. Wāð=en mu peχct, tu=ta sāw-i yo nay?  
 they=3PL me ask.PST you=FAC go.PRS-2SG or no  
 ‘They asked me whether you will go.’



### 12.3.3 Echo questions

An echo question is a type of question in which the speaker acknowledges that the information being sought has already been given previously in the discourse. Echo questions may be used either to ask for clarification or to signal surprise or disbelief, though these functions are not necessarily mutually exclusive. In most cases, an echo question asks for clarification of, or expresses surprise at, some piece of information in the utterance immediately preceding.

In English, echo questions have two special grammatical properties: stress and word order. In *wh*-echo questions, the *wh*-word is left in-situ and receives stress, as in *You ate **what** for dinner?* (cf. *What did you eat for dinner?*). The distinctive syntactic, semantic, and pragmatic properties of echo questions have been investigated, for instance, by Artstein (2002), Reis (2017), and Beck & Reis (2018). The reader is directed to these works, and references within them, for further discussion of echo questions in English and other languages.

Echo questions can be divided into *wh*-echo questions, which elicit a content response, and polar echo questions, which elicit a *yes/no* response. This distinction has morphosyntactic implications in Shughni, where echo questions are marked with question-final particles, and the type of echo question, polar or *wh*-, governs the question-final particle to be used. In *wh*-echo questions, the question-final particle *ik* is used, while in polar echo questions, the question-final particle *a* is used. Each type of echo question is discussed in turn below.

#### 12.3.3.1 *Wh*-echo questions: *ik*.

*Wh*-echo questions in Shughni are marked with the question-final particle *ik*. If an echo question consists of a single *wh*-word, the particle appears directly on this word, as in (397a). However, if the question contains content beyond the *wh*-word, then the particle appears at the end of the question, regardless of the category of the final word, as in (397b).

#### (397) *Wh*-echo questions: Initial examples

- a. Čȩz ik?  
    what ECWH  
    ‘What was that (again)?’

- b. Čīz=at lūd **ik**?  
 what=2SG say.PST ECWH  
 ‘What did you say (again)?’

Furthermore, an echo question may consist of the entire echoed utterance, as exhibited in (398), or may only contain a single *wh*-word or phrase in place of the piece of information being clarified, as in (399). These examples represent possible echo-question responses to the proposition *Mu tāt biyor ar Dūšanbe sut* ‘My father went to Dushanbe yesterday’.

- (398) Tu tāt biyor tar kā sut **ik**?  
 your father yesterday to where go.PST.M ECWH  
 ‘Where did your father go yesterday (again)?’

- (399) Tar kā **yik**?  
 to where ECWH  
 ‘To where (again)?’

In many cases, *wh*-echo questions, like polar echo questions discussed below, *echo* the preceding utterance about which they are inquiring. That is, they restate the utterance and insert a *wh*-word in the place of the unheard (or surprising) information. A *wh*-echo question may follow utterances of various types, including both affirmative and negative propositions, and both polar and *wh*-questions. Table 12.6 illustrates the use of *wh*-echo questions in response to a proposition, polar question, and *wh*-question. In the example column, the piece of information to be asked about in the *wh*-echo response is underlined. In both the example and the gloss columns, the *wh*-word corresponding to that piece of information is bolded.

Note that although the last *wh*-echo question in Table 12.6 contains two *wh*-words, the speaker only intends for the second *wh*-word – *kādand* ‘where’ – to be responded to. This is because the first *wh*-word – *čāy* ‘who’ – was in the initial utterance, and *kādand* has replaced the unheard piece of information. This is in contrast to non-echo *wh*-questions which contain two *wh*-words, which, as discussed in Section 12.3.1, elicit a pair list response.

**Table 12.6:** Example utterances and *wh*-echo responses.

Utterance type	Example	Gloss
PROPOSITION WH-ECHO	<i>Sohiba biyor tar kor na-sat.</i> <i>Čīr=i biyor na-čūd ik?</i>	‘Sohiba didn’t go to work yesterday’ ‘(She) didn’t do <b>what</b> yesterday?’
POLAR Q WH-ECHO	<i>Sohiba biyor tar kor vad o?</i> <i>Čāy biyor tar kor vad ik?</i>	‘Was Sohiba at work yesterday?’ ‘Was <b>who</b> at work yesterday?’
WH-Q WH-ECHO	<i>Čāy biyor tar kor vud?</i> <i>Čāy biyor kādand vud ik?</i>	‘Who was at work yesterday?’ ‘Who was <b>where</b> yesterday?’

The following conversation illustrates a typical conversation with *wh*-echo questions.<sup>8</sup>

**Speaker 1:** *Wuz=um biyor tar kor na-sat* ‘I didn’t go to work yesterday’

**Speaker 2:** *Čīz ik?* ‘What was that (again)?’

**Speaker 1:** *Tar kor=um lūm na-sat biyor* ‘I said I didn’t go to work yesterday.’

**Speaker 2:** *Čīr=at na-čūd ik?* ‘(I still didn’t hear you.) You didn’t do what yesterday?’

In the first usage of *ik* in this conversation (line 2), Speaker 2 asks for clarification because she has not properly heard what Speaker 1 has said. In its second usage (line 4), Speaker 2 has understood that Speaker 1 is saying she hasn’t done something, but she hasn’t understood what. Thus, the *wh*-word *čīr* ‘what (action)’ is used, and *ik* appears at the end of the question.

### 12.3.3.2 Polar echo questions: Dubitative particle *a*.

Polar echo questions are marked by the dubitative particle *a* (glossed DUB), which appears question-finally. Like *wh*-echo questions, polar echo questions may follow both propositions and questions. Unlike *wh*-echo questions, however, the speaker supplies what they believe to have heard and elicit a *yes/no* response for clarification. Table 12.7 exhibits polar echo questions in response to various types of utterances. In the examples in this table, the information being questioned is the entire utterance. In other words, there is no specific word or phrase being inquired about, and hence intonation is neutral.

<sup>8</sup>Many thanks to Shahlo Shomansurova for supplying this conversation.

**Table 12.7:** Example utterances and *wh*-echo responses.

Utterance type	Example	Gloss
PROPOSITION	<i>Sohiba tar kor na-vad.</i>	‘Sohiba wasn’t at work.’
WH-ECHO	<i>Sohiba tar kor na-vad <u>a</u>?</i>	‘(You said) Sohiba wasn’t at work, (right)?’
POLAR Q	<i>Sohiba tar kor vad o?</i>	‘Was Sohiba at work?’
WH-ECHO	<i>Sohiba tar kor vad <u>a</u>?</i>	‘(You asked) whether Sohiba was at work, (right)?’
WH-Q	<i>Čāy tar kor vud?</i>	‘Who was at work?’
WH-ECHO	<i>Čāy tar kor vud <u>a</u>?</i>	‘(You asked) who was at work, (right)?’

In addition to asking clarification about an entire utterances, as in the examples in Table 12.7, polar echo questions may also ask specifically about a word or phrase in the initial utterance. In these cases, the focus on this piece of information is typically indicated by stress and by fronting the information in question. Polar echo questions which contain a fronted, stressed piece of information may carry pragmatic implications regarding the reason why this information is brought into question. Examples of this type of polar echo question, along with an example implication they may carry, are given in Table 12.8. The initial utterance is given at the top of the table, and the unfocused polar echo question is given immediately below for comparison with the focused polar echo questions that follow. Note further that, in accordance with the convention established in the discussion on focus above, the initial (focused) constituent in polar echo questions has been written in small caps to indicate that it contains sentence-level prosodic prominence.

Finally, note that in addition to polar echo questions, the question particle *a* is used statements where the speaker is unsure of which element of a list – among multiple options – is correct. This use is exhibited in (400). Here, *a* is more readily analyzable as a dubitative suffix (glossed DUB) rather than a polar echo question particle.

- (400) Yā        kitob na-fām-um        rūšt vad    **a**,    safed **a**,    rozovi **ya**.  
 DEM.DIR.F book NEG-know.PRS-1SG red be.PST.F DUB white DUB pink DUB  
 ‘I don’t know whether that book was red, white, or pink.’

**Table 12.8:** Focused polar echo questions.

Initial utterance	
UTTERANCE	<i>Tu-rd nur tar kino sittow fort-o?</i>
GLOSS	‘Do you want to go to the movies today?’
Unfocused polar echo	
POLAR ECHO	<i>Mu-rd nur tar kino sittow fort-a?</i> (neutral intonation)
GLOSS	‘Do I want to go to the movies today?’
IMPLICATION	I’m not sure that’s what you said.
Focused polar echo	
POLAR ECHO	<u>MU-RD</u> <i>nur tar kino sittow fort a?</i>
GLOSS	‘Do <b>I</b> want to go to the movies today?’
EX. IMPLICATION	I thought you wanted to go with someone else.
POLAR ECHO	<u>NUR</u> <i>mu-rd tar kino sittow fort a?</i>
GLOSS	‘Do I want to go to the movies <b>today?</b> ’
EX. IMPLICATION	I thought you wanted to go another day.
POLAR ECHO	<u>TAR KINO SITOW</u> <i>nur mu-rd fort a?</i>
GLOSS	‘Do I want <b>to go to to the movies</b> today?’
EX. IMPLICATION	I thought you wanted to do something else today.
POLAR ECHO	<u>FORT</u> <i>mu-rd nur tar kino sittow a?</i>
GLOSS	‘Do I <b>want</b> to go to the movies today?’
EX. IMPLICATION	I have to, but I don’t want to.

### 12.3.4 Other question particles

This final subsection of the discussion on questions in Shughni turns to three question particles used to indicate different pragmatic functions. These are the neutral tag question particle *ani* (Section 12.3.4.1), the incredulous tag question particle *nā* (Section 12.3.4.2), and the emphatic question particle *ku* (Section 12.3.4.3).

### 12.3.4.1 Neutral tag questions: Particle *ani*

Neutral tag questions in Shughni are formed with the utterance-final particle *ani*. This particle immediately follows a proposition, indicating that the speaker believes what she has said is true and is seeking confirmation. Examples are given in (401). Note that this particle may attach to either an affirmative proposition, as in (401a), or to a negative proposition, as in (401b).

#### (401) Incredulous tag question particle *ani*

- a. Tu=t pi Xaray mis vud **ani**?  
 you=2SG up.in Khorugh also be.PST.M TAG  
 ‘You were in Khorugh, too, weren’t you?’
- b. Tu=t pi Xaray na-vud **ani**?  
 you=1SG up.in Khorugh NEG-be.PST.M TAG  
 ‘You weren’t in Khorugh, were you?’

### 12.3.4.2 Incredulous tag questions: Particle *nā*

A similar, but more expressive question particle is *nā*, which also appears at the end of a proposition. This particle expresses disbelief on the part of the speaker that the proposition is true. Similar to *ani*, this particle may be used request confirmation from the listener. An example of such usage is given in (402).

#### (402) Incredulous tag question particle *nā*

**Context:** Michael is telling Zahro about his time in Tajikistan. He tells her about a story that occurred when he was in Khorugh, but Zahro was under the impression that Michael had not visited Khorugh.

- Tu=t pi Xaray mis vud **nā**?  
 you=2SG up.in Khorugh also be.PST.M INCTAG  
 ‘You were (really) in Khorugh?’

Unlike, *ani*, however, the particle *nā* may also be used rhetorically to signal an incongruity or contradiction in the information that the speaker has. An example of the rhetorical use of *nā* is shown in (403).

#### (403) Rhetorical use of *nā*

**Context:** Shahlo tells her parents that she is planning to sign up for an English class. But Shahlo is already excellent at English and her parents know this, so her mother poses this question rhetorically.

Tu anglīsi na-fām-i      **nā**?  
 you English NEG-know-2SG INCTAG  
 ‘As if you didn’t know English . . .’

### 12.3.4.3 Emphatic particle *ku* in interrogatives

The third and final particle to be discussed here is the emphatic discourse particle *ku*, which is commonly used not only in questions, but also used in commands and propositions, where it has at least two distinct functions: to imply a sense of bewilderment or perplexity or to signal a heightened relevance or urgency of the information being solicited. The first type of usage is exhibited in the examples in (404); an example of the latter usage is shown in (405). Note that *ku* alone cannot be used to signal that a question is being asked. Hence, when used in questions, it must be accompanied by another interrogative element – either the polar interrogative particle *o* or a *wh*-word.

#### (404) Emphatic interrogative particle *ku*: Bewilderment

- a. A potxō, tu-nd      **ku**      tu      rāng čīz-ard      zīrd?  
 voc king, you-POSS EMPH your color what-DAT yellow?  
 ‘King, why in the world are you so pale?’ (Karamshoev 1988d:141)
- b. Gīgurd kād-and      **ku**      virī-m?  
 Match where-LOC EMPH find.PRS-1SG  
 ‘Where on earth can I find a match?’ (Karamshoev 1988d:141)

#### (405) Emphatic particle *ku*: Urgency

Wuz=ta **ku**      sāraki      cow      sām      o?  
 I=FAC EMPH in.the.morning harvest go.PRS 1SG PQ  
 ‘Am I going to do the harvest in the morning?’ (Karamshoev 1988d:141)

With respect to its position within the clause, *ku* is more flexible in questions than in commands. In fact, whereas *ku* in commands appears almost exclusively as the first element of the clause, in questions it almost never has this position. In questions, *ku* seems to gravitate toward second position, as in (404a) and (405), but it may also appear naturally in any other position except first position. In (404b), for instance, it appears in third position, immediately following the *wh*-word *kād-and* ‘where’, and it may also appear in final position, as in (406):

(406) **Final emphatic *ku* in a question**

Gīgurd kād-and virīm **ku**?  
 match where-LOC find.PRS.1SG EMPH  
 ‘Where on earth can I find a match?’

One exception where we may find emphatic *ku* in clause-initial position are embedded polar questions, as in (407). Note, however, that even in an embedded question, *ku* cannot signal a question on its own, but rather must still be accompanied a question-forming element, in this case the phrase *yo nay* ‘or not’:

- (407) Fikri=yi čūd didi, **ku** ar wi xār sū-d yō nay.  
 thought=3SG did COMP, EMPH to that city go.PRS-3SG or not  
 ‘He thought about whether or not to go to that city.’ (Karamshoev 1988d:141)

**12.3.5 Questions: Summary**

This section has examined question formation in Shughni. Fundamental takeaways include the notion that Shughni is a *wh-in-situ* language which allows scrambling in some instances of question formation. Specifically, scrambling may occur for reasons of contrastive topicalization, in which an element other than the *wh*-word is fronted, or else due to the fronting of the *wh*-word itself, which is commonly done when the addressee has not heard the question the first time it was uttered.

Polar questions, for their part, are formed with the interrogative particle (*y*)o, which appears at the end of a polar question. This particle also has the meaning ‘or’, and an alternative means for forming a polar question is through the sequence (*y*)o *nāy* ‘or no’ placed at the end of the question.

The final two subsections looked at further types of questions in Shughni. First, it was seen that echo questions, which are used commonly used to clarify information the speaker has misheard, are formed with either with the question-final particle *ik*, in the case of echo *wh*-questions, or with the question final particle *a*, in the case of polar echo questions. Three further particles were then examined, namely the neutral tag question particle *ani*, the incredulous tag question particle *nā*, and the emphatic particle *ku*, all of which have extensive usage in questions.



## 12.4 Morphosyntactic alignment and vestigial ergativity

Morphosyntactic alignment is a vast topic within the context of Indo-Iranian languages generally, and within the Iranian subgroup specifically. A number of works are dedicated to understanding the history of morphosyntactic alignment in Iranian, particularly the changes which took place between Old and Middle Iranian and led to the development of an ergative-like system in the Iranian past tense (see, e.g., Haig 2008; Jügel 2012, and references in these works). Other studies are dedicated to investigating the kinds of morphosyntactic alignment systems found in modern Iranian, which include a diverse set ranging from nominative-accusative to ergative, as well as many hybrid systems falling somewhere in between (e.g., Edelman 1974 on Iranian generally; Farrell 1995 and Korn 2009 on Balochi; Haig 1998, 2004, and Jügel 2009 on Kurdish; Payne 1989; Wendtland 2009; and Sergienko 2023 on Pamir languages; and Stump & Hippisley 2011 and Parker 2020 on Shughni specifically).

Shughni and other Pamir languages are likewise diverse in their alignment systems. One language, Rushani, is even said to have a double oblique construction, in which both subjects and objects are marked in the oblique case in the past tense (see Faizov 1966 for relevant data). Shughni, for its part, has nearly reached accusativity throughout its alignment, but still displays a vestige of a more robust ergative system in one small corner of its past tense grammar. This section is dedicated to providing a complete picture of alignment in Shughni.

The remainder of the section is organized as follows. Section 12.4.1 provides an **introduction to different types of morphosyntactic alignment and the terminology** employed to describe them. Next, Section 12.4.2 provides an **overview of alignment in Iranian languages**, including a look at the historical factors which led to the development of non-accusative alignment, as well as the diversity of alignment systems across modern Iranian languages. Then, Section 12.4.3 examines **morphosyntactic alignment in Shughni** and presents the phenomenon of vestigial ergativity. And finally, Section 12.4.4 summarizes and sets the Shughni data within the larger picture of alignment in Iranian.

### 12.4.1 Morphosyntactic alignment: Overview and terminology

*Morphosyntactic alignment* (or simply *alignment*) refers to the patterning of different types of verbal arguments with respect to morphological phenomena, often agreement or case marking, and syntactic phenomena, often

the (in)ability of arguments to participate in certain syntactic operations. The description of morphosyntactic alignment systems tends to be based on at least three different types of verbal arguments: subjects of intransitive verbs; subjects of transitive verbs; and direct objects of transitive verbs. I follow, e.g., Comrie (1978) and Dixon (1979) in using the labels S, A, and O, respectively, for each type of argument. (See these same publications, as well as a revised version of the latter, Dixon 1994, for overviews of morphosyntactic alignment.)

The two most fundamental types of alignment systems, both of which are relevant in the Iranian context, are *nominative-accusative* (also simply *accusative*) and *ergative-absolutive* (also simply *ergative*). In an accusative system, transitive subjects (A) and intransitive subjects (S) pattern alike with respect to some morphological phenomenon, to the exclusion of objects of transitive verbs (O), which pattern differently. In a canonical nominative-accusative system, both S and A take nominative (or unmarked) case and agree with verbs in the same way —i.e. via the same paradigm of agreement morphemes—while O is marked in the accusative case and either does not agree with verbs, or agrees via a separate set of morphemes. In an ergative system, on the other hand, it is intransitive subjects (S) and transitive objects (O) which pattern alike, to the exclusion of transitive subjects (A). In a canonical ergative-absolutive system, S and O take absolutive (or unmarked) case and agree with verbs, while A is marked in the ergative case and either does not agree with verbs or agrees by using a different set of morphemes.

Accusative and ergative alignment are schematized in Figure 12.2 below, where the circled arguments pattern together with respect to case-marking, agreement, or some syntactic phenomenon (e.g. extraction). Examples from languages which exhibit each type of alignment are given below.

**Figure 12.2:** Accusative and ergative alignment

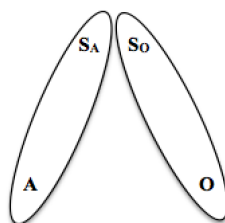


Languages which display nominative-accusative alignment are abundant and include most European languages (with the notable exception of Basque). Ergative alignment is cross-linguistically somewhat less common than accusative alignment, but is still represented in many languages from a wide range of geographic and genealogical backgrounds. The examples in (408) (from Hale 1983, as presented in Deal 2015) illustrate ergative alignment in Warlpiri, spoken in Australia. Example (408a) shows an example of a transitive sentence, in which the transitive subject, *ngarra* ‘man’, appears with the ergative marker *-ngku*, while the object, *wawirri* ‘kangaroo’, is unmarked. The sentences in (408b) and (408c) show that subjects of intransitive verbs, here *kurdu* ‘child’, pattern like direct objects, but unlike transitive subjects, in that they do not take any overt case morphology.

(408) **Ergative alignment in Warlpiri**

- a. **ngarra-ngku**<sub>A</sub> ka **wawirri**<sub>O</sub> panti-rni  
man-ERG AUX kangaroo spear-NONPAST  
‘The man is spearing the kangaroo.’
- b. **kurdu**<sub>S</sub> ka wanka-mi  
child AUX wanka-mi  
‘The child is speaking.’
- c. **kurdu**<sub>S</sub> kapi wanti-mi  
child AUX fall-NONPAST  
‘The child will fall.’

A third type of alignment system, which I will refer to as *split-intransitive* (but which is also sometimes called *split-S* or *active-stative*), is particularly relevant for Shughni and other Iranian languages. Languages with split-intransitive alignment vary with respect to where in the grammar the split is made, but in the type of split intransitivity relevant for Shughni, a distinction is made between two types of intransitive subjects: *unergative* and *unaccusative* (in the sense of Perlmutter 1978). Unergative subjects, roughly agentive intransitive subjects, pattern alike with respect to some morphological phenomenon, while unaccusative subjects, roughly non-agentive intransitive subjects, pattern differently with respect to the same morphological phenomenon (see Section 4.4.2 for more on the distinction between unaccusative and unergative verbs in Shughni). I follow Dixon (1994) in using the label  $S_A$  for unergative subjects and the label  $S_O$  for unaccusative subjects. In the type of split-intransitive alignment found in one corner of Shughni grammar, unergative subjects behave identically to transitive subjects, while unaccusative subjects behave identically to direct objects. This type of split-intransitive pattern is schematized in Figure 12.3 below:

**Figure 12.3:** Split-intransitive alignment

Split-intransitive alignment is often considered to be related to ergativity (Comrie 1978; Dixon 1994: 70-83; cf. also Deal 2015, who refers to this phenomenon as the *argument-structure property*). Initial examples exhibiting split-S alignment in the Caucasian language Batsbi are given in (409) (taken from Harris 2010: 210-211). These examples illustrate the distinct behavior of the first-singular subject when it is the subject of the unaccusative verb meaning ‘hang’, and when it is the subject of the unergative verb meaning ‘undress’. In the former, it takes the nominative form *so* and agrees with the verb via the morpheme *-sõ*, while in the latter, it takes the ergative form *as* and agrees with the verb via the morpheme *-as*.

(409) **Split-intransitive alignment in Batsbi (Caucasian)**

a. **Unaccusative subject: Nominative case**

(so) xe-n-mak qac'-u-sõ  
 I tree-DAT-on hang-PRES-I  
 ‘I am hanging in a tree.’

b. **Unergative subject: Ergative case**

(as) dah y-apx-yail-n-as  
 1SG.ERG PV CM-undress-AUX-AOR-1SG.ERG  
 ‘I got undressed.’

Finally, note that in the literature on morphosyntactic alignment, an important distinction is drawn between *morphological ergativity*, on the one hand, and *syntactic ergativity*, on the other. The former encompasses ergativity as it relates to purely morphological phenomena, while the latter refers to ergativity in the context of syntactic phenomena such as the movement of arguments. Morphological ergativity is more common than syntactic ergativity, and it can be said that syntactic ergativity is found only in a subset of morphologically ergative languages.

Nonetheless, there is an ongoing debate regarding the interconnectedness of these two types of ergativity and how to best capture them within a formal framework (see Deal 2015 and Coon & Parker 2019 for an overview of the typology of ergativity and the different formal methods used to capture its various shades). For the purposes of alignment and for Iranian languages in general, we can restrict ourselves to alignment based on morphological phenomena only, as no Iranian language has been shown to display syntactic ergativity (Haig 2008: 8).<sup>9</sup>

### 12.4.2 History and diversity of morphosyntactic alignment in the Iranian past tense

The presence of non-canonical alignment systems is perhaps one of the most salient features of the Indo-Iranian branch of the Indo-European language family. Several languages in both the Iranian and Indo-Aryan subbranches possess split alignment systems, in which one portion of the grammar – generally the present tense or imperfective aspect – is accusative, while another portion of the grammar – generally the past tense or perfective aspect—exhibits some flavor of ergativity. In the Iranian branch, the split tends to be based on tense rather than aspect, with the present tense displaying accusative alignment and the past tense displaying non-canonical alignment (e.g., Anderson 1992; Haig 2008).

The fact that split alignment systems are so pervasive throughout the Iranian branch, coupled with the fact that the split is uniformly tense-based, suggests that there is a common source from which non-accusative alignment in these languages developed. This section first discusses the process by which past-tense alignment in Iranian languages is likely to have developed from accusative to non-accusative (i.e. some flavor of ergative alignment), a discussion which is divided into a presentation of alignment in Old Iranian (Section 12.4.2.1) followed by a discussion of the construction from which ergativity likely developed (Section 12.4.2.2). The last subsection (Section 12.4.2.3) looks at the diversity in alignment systems found in modern Iranian. The historical outline provided here highlights a number of important features in the grammar of older Iranian languages, the diachronic evolution of which is crucial to understanding several aspects of modern Iranian languages, including, but not limited to, their morphosyntactic alignment.

---

<sup>9</sup>One possible exception to this generalization is found in a variety of Kurdish (see Haig 1998).

### 12.4.2.1 Alignment in Old Iranian

As discussed in Section 2.2.1 of the Introduction, the term Old Iranian does not refer to a single language or dialect, but rather to the languages of a broad time period, spanning some two thousand years and ending sometime in the first millennium BC, during which many Iranian languages were spoken simultaneously. Ancient texts and inscriptions provide evidence for the existence of multiple languages and dialects within Old Iranian, including Old Persian and Avestan, varieties of which were likely separated by over a thousand years.

Old Persian and Avestan shared a number of linguistic features which offer insight into the changes which must have taken place in the millennia to come. Two features of Old Iranian grammar are particularly important for the purposes of understanding alignment change. First, Old Iranian languages inherited from Proto-Indo-European a robust case system in which nominal forms inflected for at least six cases (e.g., Brandenstein 1964). However, by the time of Old Iranian, the PIE dative and genitive cases had collapsed into a single case, the Old Iranian genitive. In turn, the Old Iranian genitive subsumed the functions of the dative and formed a kind of ‘wastebasket’ case, appearing on nouns in a diverse range of syntactic functions and semantic roles.

And secondly, pronouns in Old Iranian could be expressed in either a full (strong) form or enclitic (weak) form. Clitic pronouns were realized as second-position (Wackernagel) clitics, attaching to the right edge of the first constituent in the clause, including conjunctions and complementizers. Compare the the strong form of the first-singular genitive pronoun *manā* in (410a) to its enclitic form *=mai* in (410b):

#### (410) Old Iranian full and enclitic 1sg genitive pronouns (Kent 1953)

- a. *imā dahyāva tyā manā patiyāša*  
 these province.PL which 1SG.GEN come.PST.3PL  
 ‘These (are) the provinces which came unto **me**.’
- b. *ava=mai visam ucāram āha*  
 that=1SG.GEN all successful be.PST.3SG  
 ‘all that was successful for **me**’

Furthermore, the past-tense morphosyntactic alignment of Old Iranian languages was fundamentally nominative-accusative. This is illustrated in the Old Persian examples in (411). Crucially, the same nominative pronoun *adam* ‘I’ is used whether it is a transitive subject, as in (411a), or an intransitive subject, as in (411b):

(411) **Nominative accusative alignment in Old Persian (Kent 1953)**

- a. pasāva      **adam** kāram      frāišayam      Bābirum  
 thereupon I      army.ACC send.PST.1SG to.Babylon  
 ‘Thereupon I sent an army to Babylon’
- b. **adam** xšāyaθiya abavam  
 I      king      become.PST.1SG  
 ‘I became king.’

Hence, the development of non-accusative alignment in the past tense of Iranian languages likely occurred after the time of Old Iranian; nonetheless, it is likely that the linguistic features and grammatical constructions through which alignment shift proceeded were already present and prevalent in Old Iranian. The relevant features and constructions, as well as the diachronic process which brought about ergativity, are the topic of the next subsection.

**12.4.2.2 The *manā kartam* construction and the development of ergativity**

It is well known that a common route from accusative to ergative alignment is through the reanalysis of a passive construction as active (e.g. Estival & Myhill 1988), and many scholars believe that the ultimate source of alignment change in Iranian languages was a passive, or passive-like, construction (e.g. Bynon 1979, 1980; Payne 1980; Bubenik 1989; a.o.). However, the notion that ergativity in Iranian descends from a passive construction is not universally accepted. Benveniste (1952), for instance, argues that the construction from which ergativity developed was possessive in nature, and more recently, Haig (2008) proposes that it was based on ‘indirect participation’ but was not formally a passive.

The construction at hand is known as the *manā kartam* construction and receives its name from the combination of the first-singular genitive pronoun *manā* and the neuter participle *kartam* ‘done’, which are often found together in this type of construction. The *manā kartam* construction calls for a semantically agent-like argument expressed in the genitive case, a semantically patient-like argument in the nominative case, and a participial verb form which agrees with the (nominative) patient. The structure of the *manā kartam* construction is schematized in (412), and examples are given in (413). Note that the agentive genitive argument can either be expressed as a full pronoun, as in (413a), or a clitic, as in (413b) and (413c).

(412) **The *manā kartam* construction: Schema**

**Agent**.GEN + **Patient**.NOM + **Verb**.PTCPL

(413) **The *manā kartam* construction of Old Persian: Examples (Kent 1953)**

- a. ima tya **manā kartam** pasāva yaθā xšāyaθiya abavam  
 that which 1SG.GEN do.PTCPL after when king become.PST.1SG  
 ‘This (is) that (which) was done by me after (I) became king.’
- b. utā=**maiy** aniyaçiy vaiy astiy **kartam**  
 and=1SG.GEN much else COP.PRES.3SG do.PTCPL  
 ‘And much else was done by me.’
- c. avaθā=**šām** hamaranam **kartam**  
 thus=3PL.GEN battle do.PTCPL  
 ‘thus by them battle was done’

At least three aspects of this construction provide the basis upon which speakers could have reanalyzed it as active. First, in many instances, including the examples above, the genitive nominal is both high on the animacy hierarchy and a speech act participant (i.e. first or second person), both qualities typical of canonical subjects cross-linguistically (Haspelmath 2001). Second, the nominative argument in this construction is generally not only semantically patient-like; it also tends to be neither animate nor a speech act participant, qualities which are typical of canonical objects. And third, when the agentive (genitive) argument is expressed as a clitic, it almost always occurs before the patient-like (nominative) argument, which gives the order AGENT–PATIENT–VERB. Taken together, these facts indicate that the arguments of this type of construction were in a position from which reanalysis would be highly feasible.<sup>10</sup>

Thus, we can imagine a scenario where the genitive nominals in the *manā kartam* construction, due to their *semantically* subject-like properties, begin to take on the *syntactic* properties of subjects. At the same time, the nominative arguments, due to their semantically object-like (or patient-like) properties, begin to take on the syntactic properties of objects, and the verb, originally a tenseless participle, begins to take on the properties of a finite

<sup>10</sup>Haig (2008: 75) notes that establishing subjecthood in languages which are no longer spoken is notoriously difficult, as subjecthood tests often rely heavily on negative data, which, for obvious reasons, are not possible to obtain in the case of Old Iranian languages. Of course, the dearth of data from these languages also contributes to the difficulty of determining if and at what point the genitive arguments became subjects.



verb, including the ability to agree with objects not just in gender and number, but also in person. This process is outlined in Table 12.4 below, modeled off that of Haig (2008: 34).

**Figure 12.4:** (Hypothesized) step-by-step development of ergative alignment in Iranian

STAGE	CHANGE	EXAMPLE
1	A passive or passive-like construction built on a participle acquires the ability to take an agent expressed in the oblique.	[Battle] <sub>NOM</sub> was done. → [Battle] <sub>NOM</sub> was done [ <i>by-them</i> ] <sub>OBL</sub> .
2	The agent phrase, presumably due to its tendency to be both <i>agentive</i> and <i>animate</i> , is reanalyzed as the syntactic subject and begins to take on language-specific subject properties.	[Battle] <sub>NOM</sub> was done [by-them] <sub>OBL</sub> . → [Battle] <sub>NOM</sub> was done [by-them] <sub>SUBJ/OBL</sub> .
3	At the same time as the oblique phrase is gaining subject properties, the nominative is losing subject properties and begins to behave as a syntactic object.	[Battle] <sub>NOM</sub> was done [by-them] <sub>OBL/SUBJ</sub> . → [Battle] <sub>NOM/OBJ</sub> was done [by-them] <sub>OBL/SUBJ</sub> .
4	The participial verb begins to take on the properties of a finite verb (e.g. it no longer relies on the support of an auxiliary and can now agree with the object in person/number).	[Battle] <sub>NOM/OBJ</sub> was done [by-them] <sub>OBL/SUBJ</sub> . → [Battle] <sub>NOM/OBJ</sub> [did] <sub>fin</sub> [by-them] <sub>OBL/SUBJ</sub> .
5	The oblique agent-phrase is fronted, in line with its new-found subject properties.	[Battle] <sub>NOM/OBJ</sub> [did] <sub>fin</sub> [by-them] <sub>OBL/SUBJ</sub> . → [By-them] <sub>OBL/SUBJ</sub> [did] <sub>fin</sub> [battle] <sub>NOM/OBJ</sub> .
6	<i>By-them battle did</i> is now the unmarked way of saying ‘they did battle’. Note that the agent phrase is marked in the oblique and has all subject properties, while the object is marked in the nominative and has all object properties. The verb continues to agree with the object.	The past tense is now ergative in its agreement and case-marking. Compare the transitive example above with the hypothetical intransitive example below:  [They] <sub>NOM/SUBJ</sub> left.

At the same time as the putative reanalysis described above took place, Iranian languages would have been undergoing another important change, namely the decay of the once complex case system. By the time of Middle Iranian, the case system had been leveled from at least six cases to only two, which I will refer to as direct and oblique. The direct case was the reflex of the Old Iranian nominative case and, after reanalysis, was used to mark

all subjects in the present tense, and objects and non-agentive subjects in the past tense. The oblique case, for its part, was used to mark present-tense objects, past-tense agentive subjects, objects of prepositions, among other non-subject functions. The changes in the Iranian case system are summarized in Table 12.9:

**Table 12.9:** Development of case in Iranian languages.

<u>P. Indo-European</u>		<u>Old Iranian</u>		<u>Middle Iranian</u>
NOM	→	NOM	→	DIR
GEN	→	GEN	→	OBL
DAT	→		→	OBL
ACC	→	ACC	→	OBL
...	→	...	→	OBL

In the end, whether or not it was the *manā kartam* construction which was the original source of alignment change in the Iranian past tense, it is clear that something in these languages set in motion a considerable and sweeping change in past-tense morphosyntactic alignment. The effects of this alignment change are still felt in many modern Iranian languages, and their manifestations are the topic of the next subsection.

### 12.4.2.3 Reflexes of ergativity in modern Iranian

Although it is generally agreed upon that non-canonical alignment in the past tense of Iranian languages developed from a single origin—the reanalysis of the *manā kartam* construction—modern Iranian languages show striking diversity in the types of alignment systems they display. As Haig (2008) notes, it is difficult to classify the alignment systems of many modern Iranian languages as accusative or ergative, as many of these languages display hybrid types of alignment that fail to conform to the properties of either canonical accusativity or canonical ergativity.

In this subsection, I discuss alignment in three modern Iranian languages: **Persian** (Section 12.4.2.3.1), which displays an accusative system, **Pashto** (Section 12.4.2.3.2), which displays a robustly ergative (split-intransitive) system, and **Davani** (Section 12.4.2.3.3), which exhibits a hybrid system. This collection of languages illustrates

alignment at various points on the accusative-ergative spectrum. The description of these languages will set the stage for the discussion of Shughni alignment in Section 12.4.3 and aid in understanding where the Shughni pattern falls along the spectrum.

**12.4.2.3.1 Persian – Accusative.** We begin with Persian, which displays a nominative-accusative system in both agreement and case-marking. Recall that in Old Iranian, pronominal forms were expressed in different cases via suppletion. In Persian, these distinctions have been lost entirely, such that in modern Persian each pronoun has only one form. However, Persian has developed an innovative accusative case marker in *rā* (Tajik Cyrillic -po), which is used to mark only specific direct objects, and hence the language displays Differential Object Marking (DOM) (Bossong 1985, 1991).

As the examples in (414) show, the Persian past tense, like the present tense, is nominative-accusative in case-marking and agreement. These examples use the first-person pronoun, which shares the same form *man* as both an intransitive and transitive subject (414a and 414b, respectively), but takes the accusative marker *-rā* as the direct object of a transitive verb (414c). Note also that verbal agreement occurs with this pronoun when it is S and A, as in the former two examples, but not when it is a direct object. Thus, in (414c), agreement is with the second-singular subject pronoun *to* ‘you’.

(414) **Past-tense alignment in Persian**

a. **1sg subject of intransitive (S): *man***

[**Man**]<sub>S</sub> be maktab raft-**am**.  
I.NOM to school go.PST-1SG  
‘I went to school.’

b. **1sg subject of transitive (A): *man***

[**Man**]<sub>A</sub> to-*rā* did-**am**.  
I.NOM you-ACC see.PST-1SG  
‘I saw you.’

c. **1sg object of transitive (O): *man-rā***

To [man-*rā*]<sub>O</sub> did-**ī**.  
I.NOM you-ACC see.PST-2SG  
‘You saw me.’

The pronominal clitics used to expressed the agent in the *manā kartam* construction have survived into modern Persian and have retained a number of their original oblique functions, including possessor, object of a preposition, and direct object. Importantly, however, they are not used for past-tense agentive subjects, as past-tense subjects are not oblique in Persian. Moreover, they do not pattern as second-position clitics, but rather attach to (or near) the head which governs them. Examples of the Persian pronominal clitics are given in (415):

(415) **Pronominal clitics in Persian**

a. **Pronominal clitic as possessor**

Pedar=am dar Irān zendegī kard.  
 father=1SG in Iran life did.3SG  
 ‘My father lived in Iran.’

b. **Pronominal clitic as object of preposition**

Man az=at hich chiz na-khwāst-am.  
 I from=2SG.CLTC no thing NEG-want-1SG  
 ‘I don’t want anything from you.’

c. **Pronominal clitic as direct object**

Man diruz did-am=aš.  
 I yesterday see.PST-1SG=3SG.CLTC  
 ‘I saw him/her yesterday.’

Thus, although Persian does not exhibit an ergative past tense, oblique pronominal clitics have retained a significant usage in the language. Past-tense alignment in Persian is summarized in Table 12.10 below:

**Table 12.10:** Morphosyntactic alignment in Persian

ARG. TYPE	CASE-MARKING	AGREE (SFX)?	CAN BE CLITIC?
<b>A</b>	NOM	✓	✗
<b>S</b>	NOM	✓	✗
<b>O</b>	ACC - (if specific)	✗	✓

Persian is the only language discussed in this section whose past-tense alignment is identical to its present-tense alignment. It is accusative in both agreement and case-marking. However, only specific direct objects are marked with the overt accusative marker *rā* (DOM). Direct objects, unlike both transitive and intransitive subjects, may be expressed as a clitic.

**12.4.2.3.2 Pashto – Ergative.** If Persian represents the accusative end of the alignment spectrum in Iranian languages, Pashto represents the ergative end. Pashto retains a distinction between direct and oblique cases on both pronouns and common nouns. All past-tense agentive subjects in Pashto are marked in the oblique case (i.e. *ergative* in an ergative-absolutive system), while direct objects in the past tense are marked in the direct case (i.e. *absolutive*). Verbs in the Pashto past tense agree with objects, rather than subjects, in person, number, and occasionally gender.

This ergative pattern is exhibited in the examples in (416). The sentences in (416a) and (416b) show that transitive and unergative subjects, respectively, appear in the oblique case in the past tense. Examples (416c) and (416d) show that unaccusative subjects and direct objects, respectively, appear in the direct case in the past tense. Note also that agreement consistently occurs with the direct argument, whether it is an unaccusative subject or direct object.

(416) **Split-intransitive alignment in the Pashto past tense**

a. **Transitive subject: Oblique (Roberts 2000: 20)**

[Maa]<sub>A</sub> Mina pe baagh kee we lid-**a**.  
 1SG.OBL Mina.DIR.F at garden in PFV SAW-FEM.3SG  
 ‘I saw Meena in the garden.’

b. **Unergative subject: Oblique (Roberts 2000: 23)**

[Maa]<sub>SA</sub> khand-él(e).  
 1SG.OBL laugh-MASC.3PL(PST)  
 ‘I was laughing.’

c. **Unaccusative subject: Direct (David 2014: 161)**

[Zə]<sub>SO</sub> lə tsawk-əy jəg-é-**dəm**.  
 1SG.DIR from chair-F.OBL tall-become-CONT.PAST.1SG.AGR  
 ‘I was getting up from the chair.’

d. **Transitive object: Direct (Roberts 2000: 20)**

Minee [ze]<sub>O</sub> pe baagh kee we lid-**em**.  
 Meena.OBL 1SG.DIR at garden in PFV SAW-1SG  
 ‘Meena saw me in the garden.’

Moreover, like Persian, oblique pronominal clitics enjoy a robust usage in Pashto, but unlike Persian they can also

be used to express past-tense agentive subjects (A and S<sub>A</sub>) in Pashto. The Pashto pronominal clitics possess this function in addition to other oblique functions, such as present-tense direct objects. Examples of Pashto pronominal clitics are given in (417). In these examples, the same oblique pronominal clitic is used to express a past-tense transitive subject (example 417a) and present-tense object (example 417b). This is precisely the pattern we would have expected after the reanalysis of the *manā kartam* construction:

(417) **Pronominal clitics in Pashto (Roberts 2000: 35)**

a. **Pronominal clitic as *past-tense* transitive subject (A)**

Khaaysta kawél-em      [dee]<sub>A</sub>.  
 Beautiful do.PST.IPFV-1SG 2SG.CLTC  
 ‘You were making me beautiful.’

b. **Pronominal clitic as *present-tense* direct object subject (O)**

Khaaysta kaw-em      [dee]<sub>O</sub>.  
 Beautiful do.PRS.IPFV-1SG 2SG.CLTC  
 ‘I am making you beautiful.’

A summary of alignment in the Pashto past tense is given in Table 12.11 below:

**Table 12.11:** Morphosyntactic alignment in the Pashto past tense.

ARG. TYPE	CASE-MARKING	AGREE (SFX)?	CAN BE CLITIC?
<b>A</b>	OBL	✗	✓
<b>S<sub>A</sub></b>	OBL	✗	✓
<b>S<sub>O</sub></b>	DIR	✓	✗
<b>O</b>	DIR	✓	✗

Pashto thus displays split-intransitive alignment in its past tense. Transitive and unergative subjects appear in the oblique case and do not agree with verbs. Unaccusative subjects and direct objects appear in the direct case and agree with verbs in person, number, and gender. Moreover, transitive subjects may be expressed as second-position clitics in the past tense, while intransitive subjects and direct objects may not.

**12.4.2.3.3 Davani – Hybrid** Having established the two ends of the spectrum in Persian (accusative) and Pashto (ergative), we now look at Davani, a language whose past-tense alignment seems to fall in between, displaying elements of both accusativity and ergativity. Davani is an endangered Western Iranian language spoken

in southern Iran. According to Moghaddam (2016), this language lacks all morphological case; hence, all clues toward its morphosyntactic alignment come from head-marking on the predicate.

Davani, like Shughni, makes use of both suffixes and morphophonological clitics to index the person and number values of subjects, with the crucial difference that in Davani, verbal suffixes are not restricted to the present tense. This aspect of the language's grammar is discussed in greater detail below. The paradigms of suffixes and clitics are given in Tables 12.12 and 12.13, respectively.<sup>11</sup>

**Table 12.12:** Davani verbal suffixes

	<u>SINGULAR</u>	<u>PLURAL</u>
1	-ε	-u
2	-ε	-i
3	-t (prs.) / -∅ (pst.)	-εn

**Table 12.13:** Davani subject clitics

	<u>SINGULAR</u>	<u>PLURAL</u>
1	=m	=mu
2	=t	=tu
3	=f	=fu

Verbal suffixes in Davani are used to cross-reference all present-tense subjects and no present-tense objects. Hence, present-tense agreement in the language is strictly accusative. The past-tense is more complex, however. Past-tense agentive subjects (A and S<sub>A</sub>) obligatorily trigger a second-position clitic. The examples in (418) illustrate the use of subject clitics to index the person and number features of past-tense transitive and unergative subjects, respectively.<sup>12,13</sup>

(418) **Davani subject clitics with past-tense agentive subjects**

a. **Past transitive subject (A), (Moghaddam 2016: 83)**

[To]<sub>A</sub>=t una xa.  
 you=2SG they ate  
 'You ate them.'

<sup>11</sup>Moghaddam (2016) labels the paradigms of these clitics and suffixes Set A and Set B, respectively. However, because the discussion here involves multiple languages which have similar series of morphemes, I forgo the use of these arbitrary labels and, for the sake of simplicity and clarity, maintain the labels 'clitic' and 'suffix' for all languages discussed, including Davani.

<sup>12</sup>Note that the subject in (418b) has been dropped. Davani, like many Iranian languages, is pro-drop.

<sup>13</sup>Moghaddam analyzes constructions like (418b) as transitive, as they are built on the transitive light verb *ke*; thus, a literal translation for (418b) would be 'they did laughing'. However, for the purposes of the discussion here I label the subject of this construction as S<sub>A</sub> (unergative), as the construction is clearly semantically intransitive.

b. **Past unergative subject (S<sub>A</sub>) (Moghaddam 2016: 134)**

Xænd-æ=ʃu                      kɛ.  
 Laugh.PST-a-3<sub>PL</sub>.SETA did  
 ‘They laughed.’

Past-tense unaccusative subjects, on the other hand, trigger the same suffixes used for present-tense agreement, as shown in (419):

(419) **Past unaccusative subject (S<sub>O</sub>), (Moghaddam 2016: 31); cf. 418a)**

[To]<sub>SO</sub> ʃɛð-ɛ.  
 you    go.PAST-2<sub>SG</sub>  
 ‘You went.’

Hence, there is already a significant clue pointing toward an ergative (or split-intransitive) system for the Davani past tense in that some subjects behave differently than others with respect to agreement. To complete the ergative picture, however, we would expect objects in Davani to pattern like unaccusative subjects. In fact, only *some* objects in the Davani past tense behave like unaccusative subjects in being co-referenced by a verbal agreement suffix. The relevant data are shown in (420). In order to agree with a past-tense verb, an object must be human, specific, and highly affected, as is the object *Hasan-o Ali* ‘Hasan and Ali’ in (420a), which is agreed with via the verbal suffix *-en*. If any of these semantic properties are not present, an object cannot pattern with a verbal suffix. This is the case in (420b), where the object *sev-gæl* ‘apples’ lacks the human requirement and thus does not appear with the verbal suffix *-en*. We are therefore dealing here with a type of Differential Object Marking (DOM – see Kalin 2018 and references therein for an overview), but one which is restricted to the past tense:

(420) **Davani DOM in the past tense**

a. **Object is human, specific, and highly affected → Suffix, (Moghaddam 2016: 33)**

Hæsan-o æli=mu. zeð-ɛn  
 Hasan-and Ali=1<sub>PL</sub> hit.PST-3<sub>PL</sub>  
 ‘We hit Hasan and Ali.’

b. **Object is *not* human, specific, and highly affected → No suffix, (Moghaddam 2016: 64)**

Sev-gæl-ku-ʃu                      xa.  
 apple-PL-DEF-3<sub>PL</sub>.CLTC eat.PST  
 ‘They ate the apples.’



The Davani past tense thus displays a type of differential object marking whereby objects with the three necessary semantic requirements pattern like unaccusative subjects. In this sense, the language displays a split-intransitive system with differential object marking, in which unaccusative subjects and objects pattern alike only under certain circumstances. That is, if an object possesses the necessary semantic requirements to agree with a verb, then agreement follows a split-intransitive pattern. Otherwise, it follows a kind of tri-partite system in which transitive/unergative subjects, unaccusative subjects, and objects all pattern differently with respect to how (and whether) they are cross-referenced morphologically. The Davani pattern is summarized in Table 12.14 below:

**Table 12.14:** Davani past-tense alignment.

ARG. TYPE	CASE-MARKING	SUFFIX?	CLITIC?
<b>A</b>	—	✗	✓
<b>S<sub>A</sub></b>	—	✗	✓
<b>S<sub>O</sub></b>	—	✓	✗
<b>O</b>	—	✓ - (if spec., hum., affect)	✗

Importantly, DOM is restricted to the past-tense in Davani, contrary to Persian, where it is found throughout the language. Moreover, while differentially marked objects in Persian highlight an underlyingly accusative system, differentially marked objects in Davani seem to highlight an underlyingly ergative system. That is, when the necessary features are present for objects to be morphologically marked, the resulting alignment pattern is accusative in Persian, but ergative in Davani. This latter fact is represented in Moghaddam’s (2016) formal analysis of this pattern.

Together with Pashto, Davani and Persian provide a representative picture of the diverse alignment types in Iranian languages. In the following section, we will see yet another type of system—that of Shughni—which is unique in its own right.

### 12.4.3 Morphosyntactic alignment in Shughni

We saw above that Iranian languages, despite all descending from an ancestor with an ergative past tense, differ drastically in the extent to which they have maintained ergative alignment in their past-tense grammars. While some display a nominative-accusative system (e.g. Persian), others possess a relatively robust ergative system (e.g. Pashto). Still others exhibit hybrid systems that can be described neither as accusative nor ergative (e.g.

Davani). In this section, I show that Shughni shows an ergative system which is so restricted in the extent to which it occurs as to be deemed *vestigial ergativity* (a term borrowed from Stump & Hoppisley 2011; cf. also the term *decay of ergativity* in Shughni and other Pamir languages, as in Payne 1980). The ergative pattern in Shughni occurs only in one exceedingly small corner of its grammar, while the rest of the language is accusative in both agreement and case marking.

### 12.4.3.1 Case-marking alignment in Shughni

Shughni has preserved the two-way distinction between direct and oblique cases, although overt case distinction is only found in a subset of pronouns and determiners in the language. The paradigm of Shughni personal pronouns (first and second person), along with distal demonstratives often used pronominally, is given below in Table 12.15 (adapted from Table 6.1 in Section 6.1.1 on personal pronouns). Cells of pronouns which have distinct forms for direct and oblique case (first-singular and demonstratives) are shaded:<sup>14</sup>

**Table 12.15:** Shughni pronouns.

	SINGULAR		PLURAL	
	DIR (F/M)	OBL (F/M)	DIR (F/M)	OBL(F/M)
1	<i>wuz</i>	<i>mu</i>	<i>māš</i>	
2	<i>tu</i>		<i>tama</i>	
3	<i>yā/ya</i>	<i>wam/wi</i>	<i>wāō</i>	<i>wew</i>

Direct case is used for subjects of all types (A, S<sub>A</sub>, and S<sub>O</sub>), while the oblique case is used elsewhere (direct objects, indirect objects, objects of prepositions, etc.). Hence, with respect to its case-marking alignment, Shughni shows a strictly nominative-accusative pattern in both the present and the past tenses, without exception.

The examples below illustrate this pattern for the first-singular pronoun *wuz* (DIR) / *mu* (OBL). The sentences in (421) show nominative-accusative case alignment in the present tense, while the sentences in (422) show nominative-accusative case marking in the past tense. In both sets of sentences, the first-singular pronoun takes

<sup>14</sup>The third-singular pronouns provided in this table are equivalent to the distal demonstrative forms. Demonstratives are used in the function of third-person pronouns.

the direct form *wuz* as a transitive subject (421a and 422a), unergative subject (421b and 422b), and unaccusative subject (421c and 422c). The pronoun takes the oblique form *mu* as a direct object in both the present and past tenses, as seen in examples (421d) and (422d).

(421) **Present-tense accusative alignment**

a. **Present-tense A: Direct case**

[Wuz]<sub>A</sub> wi                      čorik win-um.  
I                      DEM.OBL.M man see.PRS-1SG  
'I see that man.'

b. **Present-tense S<sub>A</sub>: Direct case**

[Wuz]<sub>S<sub>A</sub></sub> nur    tar xār žoz-um.  
I                      today to city run.PRS-1SG  
'I am running to the city today.'

c. **Present-tense S<sub>O</sub>: Direct case**

[Wuz]<sub>S<sub>O</sub></sub> nur    tar xār firap-um.  
I                      today to city arrive.PRS-1SG  
'I am arriving to the city today.'

d. **Present-tense O: Oblique case**

Yu [mu]<sub>O</sub> wīnt.  
he me see.PRS.3SG  
'He sees me.'

(422) **Past-tense accusative alignment**

a. **Past-tense A: Direct case**

[Wuz]<sub>A</sub>=um wi                      čorik wīnt.  
I=1SG                      DEM.OBL.M man see.PST  
'I saw that man.'

b. **Past-tense S<sub>A</sub>: Direct case**

[Wuz]<sub>S<sub>A</sub></sub>=um biyor    tar xār žêxt.  
I=1SG                      yesterday to city run.PST  
'I ran to the city yesterday.'

c. **Past-tense S<sub>O</sub>: Direct case**

[Wuz]<sub>S<sub>O</sub></sub>=um biyor    tar xār firīpt.  
I=1SG                      yesterday to city arrive.PST  
'I arrived to the city yesterday.'

d. **Past-tense O: Oblique case**

Yu=yi [mu]<sub>O</sub> wīnt.  
he=3SG me see.PST  
'He saw me.'

The nominative-accusative case-marking pattern shown in these examples is found throughout the language. In the following subsection, we will see that head-marking (i.e. the indexing of subjects morphologically on or near the verb) generally shows a similar pattern.

### 12.4.3.2 Alignment in agreement-like phenomena

As discussed in Section 8.2, Shughni makes use of two types of morphemes which agree with subjects in person and number. With present stems, subjects are cross-referenced by *suffixes*, which appear directly on the verb stem. In the past tense, however, subjects are cross-referenced by *morphophonological clitics*, which occur in

second position and attach to whichever syntactic constituent is in first position, excluding complementizers and conjunctions. As a refresher, this phenomenon is shown with the following present- and past-tense examples:

(423) **Present agreement suffixes and past-tense clitics**

a. **Present stem: 1PL agreement suffix *-ām***

Māš tama-rd qīw-ām.  
 we you.PL-DAT call.PRS-1PL.SFX  
 ‘We call you (pl.).’

b. **Past stem: 1PL subject clitic *=ām***

Māš=ām tama-rd qīwd.  
 we=1PL.CLTC you.PL-DAT call.PST  
 ‘We called you (pl.).’

This difference in distribution is the most salient discrepancy between the verbal agreement-like morphemes used in the present tense and those used in the past tense. In form, they are nearly identical, differing only in the second- and third-person singular cells. As a reference for the remainder of this section, the paradigms of present agreement suffixes and past subject clitics are provided in Tables 12.16 and 12.17, respectively. Note that the alternation in the third-singular cell of past-tense clitics between the overt clitic *=i* and  $\emptyset$  is a crucial one and is the topic of Section 12.4.3.3 below.

**Table 12.16:** Shughni present suffixes

	SINGULAR	PLURAL
1	<b>-um</b>	<b>-ām</b>
2	<b>-i</b>	<b>-et</b>
3	<b>-t/d</b>	<b>-en</b>

**Table 12.17:** Shughni past clitics

	SINGULAR	PLURAL
1	<b>=um</b>	<b>=ām</b>
2	<b>=(a)t</b>	<b>=et</b>
3	<b>=i / <math>\emptyset</math></b>	<b>=en</b>

Regarding the alignment of these morphemes, present-tense suffixes show a fully accusative pattern, cross-referencing all subjects and absolutely no objects. Past-tense clitics also display accusative alignment, with the exception of the third-person singular, to be discussed in section 12.4.3.3. In all other person-number combinations, alignment is clearly accusative.

The accusative pattern of suffixes and clitics is shown in the examples below, again using the first-singular pronoun *wuz* (DIR) / *mu* (OBL). In both the present and past tenses, an agreeing morpheme co-refers with transitive subjects

(424a and 425a), unergative subjects (424b and 425b), and unaccusative subjects (424c and 425c). In both the present and past tenses, however, it is always ungrammatical for an agreeing suffix or clitic to be co-indexed with direct objects, as shown in examples (424d) and (425d), respectively.

(424) **Accusative alignment in present suffixes**

a. **Present-tense A: Suffix required**

Wuz wi win\*(-um).  
I=1SG him see.PRS-1SG  
'I see him.'

b. **Present-tense S<sub>A</sub>: Suffix required**

Wuz taram žoz(\*-um).  
I=1SG to.there run.PRS-1SG  
'I run there.'

c. **Present-tense S<sub>O</sub>: Suffix required**

Wuz wirafc\*(-um).  
I stand.PRS-1SG  
'I stand up.'

d. **Present-tense O: Suffix ungrammatical**

Tu mu win-i (/\*win-um).  
you me see.PRS-2SG (/\*see.PRS-1SG)  
'You see me.'

(425) **Accusative alignment in past (non-3sg) clitics**

a. **Past-tense A: Clitic required**

Wuz=um wi wĩnt.  
I=1SG him see.PST  
'I saw him.'

b. **Past-tense S<sub>A</sub>: Clitic required**

Wuz=um taram žêxt.  
I=1SG to.there run.PST  
'I ran there.'

c. **Past-tense S<sub>O</sub>: Clitic required**

Wuz=um tũyd.  
I=1SG leave.PST.MASC  
'I left.'

d. **Past-tense O: Clitic ungrammatical**

Tu(=t)=(\*-um) mu wĩnt.  
you=(2SG)=(\*1SG) me see.PST  
'You saw me.'

Thus, in this regard, Shughni displays an accusative pattern with respect to the phenomenon of both present suffixes and past clitics. Note further that unlike any of the languages discussed in section 12.4.2.3 – namely Persian, Pashto, and Davani – Shughni does not use clitics to cross-reference non-subject arguments (e.g. possessors, direct/indirect objects, etc.).

At least superficially, then, it appears that past-tense clitics, despite their distinct distribution and form, are fulfilling the same grammatical role as present-tense suffixes. One might even argue that the clitic/suffix distinction plays a crucial role in distinguishing the past tense from the present tense. Nonetheless, there is a subtle yet crucial distinction between the patterning of suffixes and clitics in Shughni. This distinction is found only for third-person singular arguments and is the topic of the next subsection.

### 12.4.3.3 Vestigial Ergativity

We saw above that Shughni is nominative-accusative throughout its case marking and present-tense agreement suffixes. With present verb stems, all subject types – transitive, unergative, and unaccusative – pattern with the same suffix, while it is ungrammatical to have a suffix co-referencing the person and number values of the object. The same pattern holds for past-tense subjects of all person-number combinations *except* third-singular, and it is precisely in this cell that Shughni displays a vestige of its prior stage of ergativity.

In the third-singular cell, only a subset of past-tense subjects pattern with the overt clitic  $=(\text{y})i$ . Subjects of transitive and unergative verbs trigger the clitic, while subjects of unaccusative verbs do not trigger a clitic. This pattern is shown in the examples below, where the transitive subject in (426a) and the unergative subject in (426b) each trigger the clitic  $=(\text{y})i$ , while it is ungrammatical to have this clitic with the unaccusative subject in (426c) or with the direct object in (426d):

#### (426) Vestigial ergativity in Shughni: The data

##### a. Third-singular transitive subject: Clitic obligatory

Yā=**yi** xu nān-ard qīwd.  
 she=3SG REFL mom-DAT call.PST  
 ‘She called her mom.’

##### b. Third-singular unergative subject: Clitic obligatory

Yā=**yi** tar xu čīd žêxt.  
 she=3SG to REFL house run.PST  
 ‘She ran to her house.’

##### c. Third-singular unaccusative subject: Clitic ungrammatical

Yā(**\*=yi**) toyd.  
 she(**\*=3SG**) leave.PST.FEM  
 ‘She left.’

##### d. Third-singular transitive object: Clitic ungrammatical

Tu=t(**\*=i**) wam wīnt.  
 you=2SG(**\*=3sg**) DEM.OBL.F see.PST  
 ‘You saw her.’

Therefore, unaccusative third-singular subjects in Shughni behave like objects in that they do not pattern with a

clitic. In this sense, the language displays a split-intransitive pattern with third-singular subjects in the past tense, where subjects of transitive and unergative verbs behave alike in triggering a clitic, to the exclusion of subjects of unaccusative verbs and objects, which trigger no clitic.

To recap the system of morphosyntactic alignment found in Shughni past-tense clitics, consider first Table 12.18 below, which displays past-tense alignment for non-3sg arguments:

**Table 12.18:** Shughni past-tense alignment (non-3sg).

ARG. TYPE	CASE-MARKING	OVERT CLITIC?
<b>A</b>	DIR	✓
<b>S<sub>A</sub></b>	DIR	✓
<b>S<sub>O</sub></b>	DIR	✓
<b>O</b>	OBL	✗

Now consider Table 12.19, which displays the alignment system found for third-singular subjects in the past-tense:

**Table 12.19:** Shughni past-tense alignment (3sg).

ARG. TYPE	CASE-MARKING	OVERT CLITIC?
<b>A</b>	DIR	✓
<b>S<sub>A</sub></b>	DIR	✓
<b>S<sub>O</sub></b>	DIR	✗
<b>O</b>	OBL	✗

To summarize, for all subjects except third-singular, Shughni is accusative with respect to both case-marking and past-tense clitics. In the case of third-singular subjects, the language still shows an accusative pattern with respect to case-marking, but shows a split-intransitive pattern with respect to the patterning of past-tense clitics. In this sense, the split is found in a domain even smaller than the third-person singular cell of past-tense subjects; it is restricted to only a single phenomenon (rather than, say, both case and agreement). It appears that Shughni is as close as possible to showing accusative alignment throughout its past tense without fully being there.

**12.4.3.3.1 A typological perspective.** Despite its restricted presence in the language, the Shughni split-intransitive pattern is found precisely in the corner of the grammar that we would predict, given what is known about the typology of split-ergativity. Cross-linguistically, in languages which display split-ergativity based on some property (or properties) of the arguments in question, inanimate arguments tend to be ergative before ani-

mate arguments, non-human arguments before human arguments, third-person before speech-act participant (i.e. first- and second-person), etc. (McGregor 2009; see also Dixon 1972 for a person-based in Dyirbal and Valenzuela 2000 for such a split in Yaminawa). This same pattern holds for the Shughni data, where the split is based not only on person, but also on number. With respect to person, it is the least-marked (third-person) cell where the non-canonical pattern is found, and the same is true of number, as the split-intransitive pattern is found only in the third-*singular* cell and not in the third-*plural*.

The restricted nature of ergativity in the Shughni past tense can be appreciated in Table 12.20, which shows which arguments trigger overt clitics. Cells which trigger overt clitics are in gray; note that it is only the third-singular cell which exhibits a split:

**Table 12.20:** Shughni past-tense subjects and overt clitics.

	SING	PLUR
1		
2		
3	TRANS/UNERG	
	UNACC	

#### 12.4.4 Morphosyntactic alignment in Shughni: Summary

In this section, we have seen that Iranian languages display striking diversity in their past-tense morphosyntactic alignment. This diversity is rooted in the notion that at some point in the history of Iranian, specifically some time after the Old Iranian period, an ergative pattern developed in the past tense from a passive or passive-like construction. This pattern has left its mark, albeit to differing degrees, in the modern Iranian languages.

Shughni, for its part, displays nominative-accusative alignment in its case-marking system and present-tense agreement suffixes throughout the language, without exception. In its past-tense system of agreement, which is carried out via second-position clitics rather than suffixes, the language is strictly nominative-accusative everywhere except in the third-person singular, where it displays a split-intransitive pattern. That is, past-tense arguments of all person-number combinations except third-singular, all subjects – whether transitive, unergative, or unaccusative – require a second-position clitic co-indexing their person and number values. For third-singular past-tense subjects,



however, only transitive and unergative subjects, to the exclusion of unaccusative subjects, call for the third-person second-position clitic  $=(\nu)i$ . In this way, past-tense third-person singular unaccusative subjects behave like direct objects in triggering no agreement-like morpheme.

A split-intransitive pattern restricted to such a small corner of the language's grammar may be uncommon cross-linguistically, but it nonetheless follows tendencies regarding the kinds of arguments implicated in the ergative pattern in languages with split alignment. In particular, in Shughni, it is the type of argument least likely to be human or animate, namely third-singular, which displays the non-nominative pattern. The same pattern is found in other split-ergative languages throughout the world. Nonetheless, one aspect of the Shughni split is unattested, to my knowledge. Specifically, the Shughni split is based not only on person, but also on number. This is because whereas third-person *singular* arguments show a non-nominative pattern, third-person *plural* arguments fall within the accusative side of the language.

Within the context of Iranian languages, Shughni is like Davani and many other languages in exhibiting a hybrid alignment system which falls somewhere between accusativity and ergativity. We saw that some Iranian languages, notably Persian, are fully accusative in their case and agreement alignment, save some differential object marking. Others, such as Pashto, are virtually fully ergative in their past tense. Indeed, the past tenses of Iranian languages present many interesting phenomena within the realm of morphosyntactic alignment. Even within the Pamir languages the diversity of alignment systems is rather vast. Ultimately, alignment in Iranian, and especially within Eastern Iranian, is a topic which is still understudied, and it is my hope that this presentation of alignment in Shughni might contribute to its investigation.



## Chapter 13

# Conclusion

This dissertation has provided a description of a wide variety of topics in the Shughni language, from historical linguistic issues to the sound system, and from topics in nominals and verbs to issues in syntax and information structure. It has summarized, integrated, and built upon the work of previous authors, most notably those who worked and published during the Soviet era. It is my hope, therefore, that this thesis will serve as a comprehensive source on Shughni grammar, one which scholars working in-depth on specific topics might be able to use as a reference. Similarly, I believe the presentation of novel data in the language might provide ideas and inspiration for future studies. With this in mind, throughout the thesis, I have pointed out areas of the language where more work is needed to fill in gaps in our knowledge and understanding.

The goal of this conclusion is twofold. First, it will summarize the main points made in each part of the thesis, and second, it will discuss the topics which merit further investigation. In many cases I will provide more insight than in previous chapters into the puzzles which remain to be solved. This chapter might therefore serve as an idea bank for future studies on Shughni.

**Part I** of this thesis provided preliminary information on Shughni, including demographic and geographic background as well as the genetic position of Shughni within Indo-European. It was seen that a common misconception in certain publications on Shughni is that the Pamir languages are a cohesive genetic subgroup of languages within Iranian. In reality, although they share many distinctive phonological and grammatical features, these languages

have not been demonstrated to constitute a separate genetic unit. Therefore, the Pamir languages are for now more readily described as a Sprachbund. Moreover, while they generally display the phonological and grammatical features ascribed to Eastern Iranian – such as the spirantization of certain word-initial stops and a more conservative inflectional system than those typically found in Western Iranian – even the Eastern Iranian group itself has not yet been demonstrated to possess shared innovations which set it apart from all Western Iranian languages. And although significant work on genetic relations within Iranian has already been carried out, a more thorough investigation of Shughni and other under-documented Pamir languages might help further our understanding. In particular, a more comprehensive documentation of these languages might lead to new insights regarding genetic relations both within the Pamir group and within Eastern Iranian more broadly.

In Chapter 3 on phonology and orthography, it was seen that Shughni possesses a sound system which is similar to the other languages of the Shughni-Rushani group. However, a number of issues in the language's vowel system remain ripe for investigation. It was seen, for instance, that the language's ten vowel phonemes are canonically thought of as belonging to multiple overlapping classes: short vs. long, low vs. non-low, and non-low rounded vs. non-low unrounded. Novel morphophonological evidence was provided to support the notion that each long vowel corresponds to a single short vowel. In particular, it was shown that the vowel *ũ* shares a phonological correspondence with short *u*, despite the fact that short *u* is never as close in quality to *ũ* as it is to the other non-low, rounded vowels. Evidence for this correspondence came from the phenomenon of verb-stem shortening, in which the original long vowel in shortened stems always becomes its corresponding short vowel. We saw that the shortened form of the imperative *lũv* 'say!' is *lu*, just as, for instance *čũd* 'did' shortens to *ču*. This is, to my knowledge, the first piece of morphophonological evidence demonstrating the phonological correspondence between groups of short and long vowels. Future studies should seek to examine these correspondences in more detail with both phonetic and phonological data.

In the same vein, it was noted that the vowel *ũ* seems to be somewhat of an outlier among the non-low rounded group with respect to its articulatory and acoustic properties. Previous authors have consistently referred to this group as back vowels. And while some have noted that the vowel in question is produced farther forward than the other vowels of the group, the present study provided acoustic data indicating that this vowel has such a forward quality as to be considered a central or even front vowel. Further investigations into Shughni phonetics and phonology might examine the articulatory and acoustic properties of this vowel in more detail. If indeed it proves to be a front or central vowel, then the group of long vowels to which it belongs would be united only by 'non-low' and

‘unrounded features’ (and not by a ‘back’ feature, as has been claimed in previous studies).

Moreover, Shughni prosody remains virtually wide open for investigation. Very few studies on Shughni have provided any insight into the language’s intonational system, and to my knowledge, none have provided acoustic data to support their claims. Similarly, although patterns of lexical stress have been described in previous studies, including the present thesis, the acoustic properties of stress in Shughni have yet to be examined. Prosody thus remains something of a blank slate for future studies.

**Part II** of the dissertation turned to an examination of various issues in nominals. In Chapter 5, a lengthy discussion was provided on the syntactic properties of the reflexive pronoun *xu*, which behaves as other oblique pronouns in being able to function as direct or indirect object, object of a preposition, or possessor. A promising line of future investigation lies in the intersection of this anaphor with the notion of subjecthood in Shughni. In particular, it was shown that this pronoun is readily analyzable as being an oblique subject-oriented anaphor, insofar as subjects in Shughni are generally considered to be direct-case, agreement controlling arguments. That is, the pronoun *xu* must be both in an oblique position (a property which excludes it from being a coordinated component of a subject noun phrases, as in ‘he and *his* father’) and co-referential with the direct-case, agreement-controlling argument in its clause. Nonetheless, it was also seen that this anaphor is acceptable in non-finite complement clauses of dative-first constructions, where it is co-indexed not with a nominal argument, but rather with a dative experiencer. This fact raises questions not only about the status of *xu* as a subject-oriented anaphor, but also about the notion of subjecthood in Shughni more broadly, and even about the nature of dative-first constructions and non-finite complement clauses. This is therefore a potentially fruitful area for future investigation which may shed light on various aspects of the language.

Demonstrative pronouns, examined in Chapter 6, are another area which contains many open questions despite an in-depth treatment in this thesis. In particular, this dissertation provided a rigorous description of the triple deictic system of Shughni as it surfaces in the language’s demonstrative pronouns. It was shown that the medial demonstrative is fundamentally addressee-oriented in that it is used to refer to entities within the personal space of the addressee. However, the medial form was also found to be felicitous in instances where its referent is within the shared interactional space of the speaker and addressee, but not clearly closer to either speaker or addressee. Moreover, it was noted that the choice between medial and distal forms depends not only on notions which commonly mediate the choice of such forms cross-linguistically, such as whether the entity in question is

visible or not, but also on additional factors such as whether the speaker perceives the addressee's attention to be already on the referent or not. Thus, many questions remain regarding the use of medial and distal demonstrative forms and the factors which determine the use of one or the other, particularly concerning their interaction with information structural notions of topic and focus.

Chapter 7 provided a detailed description of grammatical gender in nominals, including its morphological expression on a subset of nouns, as well as on certain adjectives which exhibit gender concord. Moreover, the system whereby gender is assigned to Shughni nouns was seen to be rather distinctive within the Indo-European context. The most reliable predictor of a noun's gender is not its phonological shape, as in many Indo-European languages, but rather its meaning. In this sense, the Shughni system of gender assignment resembles that of South Asian languages such as Burushaski and the Dravidian languages. This semantic-based system is thought to be a shared areal feature of these languages, and the extent to which Shughni shares this feature with other languages remains a topic for future investigation. In addition, more work is needed to understand the interplay of the various semantic factors which contribute to gender assignment, including thematic categorization, abstractness vs. concreteness, individual vs. mass, and part vs. whole. The influx of Russian loanwords into Shughni may well prove to be a useful tool in this regard. Gender assignment in Russian loanwords has already revealed that although thematic categorization does much of the work, the notions of part vs. whole are much more active than previously thought. The existence of many apparent exceptions to our current understanding of gender assignment also provides an intriguing topic for subsequent studies.

**Part III** of this thesis looked at the Shughni verbal system, including, but not limited to, the form and usage of verb stems, regular and irregular verbs, issues in tense, aspect, and mood, passive-like and causative constructions, and complex verbs. A particularly important area in Shughni verbs is the notion of transitivity, which was introduced in Chapter 4 and appeared again and again throughout the thesis in the treatment of a variety of topics in the language's verbal system. It was seen that Shughni verbs can broadly be categorized into three classes: (i) a transitive class whose verbs which takes direct objects and require the third-singular clitic  $=\langle y \rangle i$  with past and perfect stems; (ii) an unergative class whose verbs do not take direct objects but which nonetheless require the same clitic; and (iii) an unaccusative class whose verbs do not take a direct object and are ungrammatical with the third-singular past-tense clitic. Verbs in the unaccusative class differ from those in the other two classes in other ways as well, notably in their ability to agree with their subjects in gender and number, the availability of morphological causatives formed from their verb stems, and the existence of gender-distinguishing adjectival resultative participles based on these

verbs (where such participles are impossible with unergative verbs). Although many aspects of the way these verb classes show up in Shughni grammar were described in the thesis, much remains to be understood. For instance, the content of each class has not been properly documented, and it is unknown what factors – semantic, historical linguistic, or otherwise – play a role in the categorization of intransitive verbs into unergative and unaccusative. Similar classes of verbs exist in many Iranian languages and in non-Indo-European languages spoken in the region, including Burushaski. The comparison of the content of these verb classes across languages may be a useful endeavor in this regard.

In addition, there are a number of open questions regarding the form of verbs. Chapter 9 presented the various ways in which Shughni verbs may be irregular, including through variation in the consonants and vowels of each stem. Although this phenomenon has been relatively well documented from a historical perspective, the ongoing changes taking place in the verbal system have generally not been investigated. Future studies might examine the phenomenon of verb-stem shortening, which appears to be optional for some verbs but obligatory for others. A similarly intriguing topic is that of hybrid complex verbs which exhibit inflectional characteristics of both simplex and complex verbs. Not only does the behavior of these verbs appear to have changed in recent times, but it also appears to vary dialectally. Moreover, preliminary observations indicate that this class of verbs might be considered a continuum, with some verbs mostly exhibiting properties of complex verbs and others exhibiting mostly properties of simplex verbs. All of these issues remain open for investigation.

Many un(der)-investigated phenomena in the area of TAM were examined in Chapter 10. The temporal and aspectual compatibilities of each of the three finite verb stems – present, past, and perfect – were described in detail. Moreover, the use of infinitive stems in the formation of aspectual constructions such as the progressive, inceptive, and prospective, was also presented. Nonetheless, our understanding of several areas of TAM remains less than satisfactory. One such area is the use of the factual enclitic *=ta*, which appears to have the ability to express future temporal reference, progressive aspect, and habitual aspect. This enclitic is only compatible with present stems, but a precise understanding of its semantic contribution remains a topic for future investigation. In particular, seeming contradictions in which it appears to encode habitual aspect in some examples but progressive in others pose interesting puzzles. It was shown that a potentially useful approach is to compare the meaning and pragmatic felicitousness of minimal examples with and without *=ta*.

Lastly, **Part IV**, which consists only of Chapter 12, examined several phenomena in Shughni syntax, including

various types of complex clauses – notably relative clauses and temporal adverbial clauses – as well as information structure, question formation, and morphosyntactic alignment. In the first part of the chapter, it was shown that Shughni has three basic types of relative clauses: externally headed relative clauses, internally headed (correlative) relative clauses, and free-choice free relative clauses. Relative clauses in Shughni had not been investigated in any detail to this point, and the existence of correlative clauses in Shughni had not been pointed out. Nonetheless, several aspects of relative clauses in Shughni remain to be investigated. For instance, although relative clauses generally contain the subordinator *ca*, which targets the right of the clause immediately before the verb, this subordinator is not obligatory. It is not yet understood how often and under what circumstances this subordinator is used. Thus, the grammatical and stylistic factors which modulate its usage in relative clauses remain to be studied. Moreover, as far as the data from this study indicate, correlative clauses and externally headed relative clauses are equally possible and may be used in the same grammatical and pragmatic contexts. However, future studies, particularly those based on naturally occurring speech, might shed light on the factors affecting which type of relative clause is used.

Similar to relative clauses, information structure in Shughni had not been investigated in any previous study. It was seen that Shughni exhibits many characteristics of a topic-prominent language. For instance, unlike in subject-prominent languages, passive-like constructions in Shughni do not have widespread usage. In cases where a passive would be used in English to promote a topic which is semantically an object to the status of grammatical subject, Shughni instead uses scrambling. Topicalized semantic objects are allowed to remain grammatical objects and appear in sentence-initial position. It was also demonstrated that whereas non-contrastive topics commonly (but not always) appear in sentence-initial position, contrastive topics are required to be the first element in the sentence. Elements which are non-contrastively focused, on the other hand, generally do not appear in sentence-initial position, while contrastively focused parts of a sentence are allowed to be fronted. Because this is the first study on the information structural notions of topic and focus in Shughni, the claims made here would ideally be checked against a larger set of naturally occurring data (for instance, through a corpus of spoken Shughni). Moreover, the interface of intonation and information structure has not been investigated here, although it was noted that focused elements generally receive sentence-level prosodic prominence. The acoustic nature of this prominence, and the way it varies between contrastively and non-contrastively focused elements remains to be seen.



# Appendices



# Appendix A

## Nouns

**Table A.1:** Predominantly masculine thematic categories.

Place names		Sicknesses		Footwear	
WORD	GLOSS	WORD	GLOSS	WORD	GLOSS
<i>Xaray</i>	Khorog	<i>dārđ</i>	pain; illness	<i>būt</i>	boot
<i>Riǰūn</i>	Rushan	<i>bođ</i>	eczema	<i>ǰirīb</i>	(large) sock
<i>Rošorv</i>	Roshorv	<i>xaraǰ</i>	scabies	<i>ǰirībak</i>	sock
<i>Xuǰand</i>	Khujand	<i>kêxak</i>	cough	<i>kafǰ</i>	(wooden) shoe
<i>Baǰūw</i>	Bajuw	<i>piršak</i>	sneezing	<i>krasovka</i>	gym shoe
<i>Dušanbi</i>	Dushanbe	<i>girīp</i>	flu	<i>tāpka</i>	slippers
<i>Maskow</i>	Moscow	<i>tiburkulos</i>	tuberculosis	<i>šilopak</i>	sandal

Milk products		Vegetables		Infinitives and gerunds	
<i>ǰūvd</i>	milk	<i>piyoz</i>	onion	<i>zidāridz</i>	sweeping
<i>đūy</i>	doogh (drink)	<i>bodrīng</i>	cucumber	<i>čêridz</i>	harvesting
<i>marūb</i>	cream	<i>kilo</i>	pumpkin	<i>žêǰtow</i>	running
<i>qurūt</i>	(salty) yogurt ball	<i>kartuškā</i>	potato	<i>nivištow</i>	writing
<i>tuǰp</i>	(unsalted) yogurt ball	<i>zardak</i>	carrot	<i>kitob ǰêyd</i>	book reading
<i>limanāt</i>	lemonade	<i>kapūst</i>	cabbage	<i>xāt nivišt</i>	letter writing
<i>yogurt</i>	yogurt	<i>baqlažān</i>	eggplant	<i>mot sit</i>	becoming tired

Table A.2: Predominantly feminine thematic categories.

Bodies of water		Literary and artistic items		Trees	
WORD	GLOSS	WORD	GLOSS	WORD	GLOSS
<i>daryo</i>	river	<i>kitob</i>	book	<i>daraxt</i>	tree
<i>qūl</i>	lake	<i>sūg</i>	tale	<i>mūn</i>	apple tree
<i>bār</i>	sea	<i>čistūn</i>	riddle	<i>tūð</i>	mulberry tree
<i>šarvidoʃ</i>	mountain stream	<i>ramān</i>	novel	<i>sīzd</i>	silverberry tree
<i>wêð</i>	canal	<i>pesā</i>	(theatrical) play	<i>wed</i>	willow
<i>vadapad</i>	(large) waterfall	<i>kino</i>	movie	<i>yūz</i>	walnut tree
<i>riʃūv</i>	(small) waterfall	<i>xāt</i>	letter	<i>arar</i>	poplar

---

Musical instruments		Drinks (except milk products)		Fruits	
<i>tor</i>	tar	<i>šac</i>	water	<i>mewā</i>	fruit
<i>dutor</i>	dutar	<i>sok</i>	juice	<i>mūn</i>	apple
<i>setor</i>	setar	<i>kampot</i>	compote	<i>tūð</i>	mulberry
<i>dāf</i>	tambourine	<i>čoy</i>	tea	<i>sīzd</i>	silverberry
<i>gitār</i>	guitar	<i>kofe</i>	coffee	<i>kayūn</i>	pear
<i>sikirīpka</i>	violin	<i>šarob</i>	alcohol	<i>apilsīn</i>	orange
<i>piyanīna</i>	piano	<i>vīno</i>	wine	<i>limūn</i>	lemon

---

Technology and transport		Containers and tableware		Bedding items	
<i>mošīn</i>	car	<i>toθč</i>	bowl	<i>lef</i>	blanket
<i>tilīfūn</i>	telephone	<i>dek</i>	pot	<i>adyāl</i>	blanket
<i>aftobus</i>	bus	<i>čib</i>	spoon	<i>daroškā</i>	carpet
<i>poyīzd</i>	train	<i>čêd</i>	knife	<i>qoilīn</i>	rug
<i>taksī</i>	taxi	<i>xaladīlnik</i>	refrigerator	<i>madrās</i>	mattress
<i>rādiyo</i>	radio	<i>čimadān</i>	suitcase	<i>bolaxč</i>	small blanket
<i>tilivīzor</i>	television	<i>čalak</i>	bucket	<i>namad</i>	wool carpet

Table A.3: Split thematic categories.

Body parts		Buildings and building parts		Plants	
WORD	GLOSS	WORD	GLOSS	WORD	GLOSS
FEMININE		FEMININE		FEMININE	
<i>angiġt</i>	finger	<i>magazġn</i>	store	<i>ġar</i>	dogrose
<i>zġv</i>	tongue	<i>lafkġ</i>	store	<i>ġġš</i>	burdock
<i>yġv</i>	mouth	<i>sġtan</i>	column		
<i>noxġn</i>	nail	<i>xidorġ</i>	mill		
<i>noy</i>	throat	<i>kġxni</i>	kitchen		
<i>mġk</i>	neck	<i>wġs</i>	ceiling beam		
<i>šand</i>	lip	<i>dġlġdz</i>	vestibule (Pamiri house feature)		
MASCULINE		MASCULINE		MASCULINE	
<i>cem</i>	eye	<i>ċġd</i>	house	<i>woġ</i>	grass
<i>dġm</i>	back	<i>yġġd</i>	stable	<i>cġġm</i>	wormwood
<i>kġl</i>	head	<i>dġšġd</i>	roof	<i>šġġ</i>	thorn
<i>nusk</i>	chin	<i>divi</i>	door	<i>šitorġk</i>	rhubarb
<i>sġvd</i>	shoulder	<i>spġlni</i>	bedroom	<i>zamborġy</i>	mushroom
<i>pġc</i>	face	<i>bġnya</i>	bathroom	<i>amoġak</i>	ephedra
<i>ben</i>	palm	<i>rġz</i>	opening for light (Pamiri house feature)	<i>yorġ</i>	alfalfa

**Table A.4:** Examples of (masculine) abstract nouns.

Shughni			Tajik		
<u>WORD</u>	<u>GLOSS</u>		<u>WORD</u>	<u>GLOSS</u>	<u>TAJIK</u>
<i>žtɯɟaɕ</i>	love		<i>mehr</i>	kindness	мехр
<i>ʁoʃ</i>	fear		<i>barobari</i>	equality	баробарӣ
<i>xũðm</i>	sleep		<i>unar</i>	talent	хунар
<i>zaqaɕ</i>	agitation		<i>dũniš</i>	knowledge	дониш
<i>sũzaɕ</i>	pain		<i>dũsti</i>	friendship	дустӣ
<i>šittoyi</i>	coldness		<i>kor</i>	work	кор
<i>moɟdz</i>	hunger		<i>umeð</i>	hope	умед

Arabic			Russian		
<u>WORD</u>	<u>GLOSS</u>	<u>ARABIC</u>	<u>WORD</u>	<u>GLOSS</u>	<u>RUSSIAN</u>
<i>išq</i>	love	عَشَقَ	<i>sipicialnust</i>	specialty	специальность
<i>fikri</i>	thought	فَكَرَ	<i>uspex</i>	success	успех
<i>aql</i>	mind	عَقْلَ	<i>zavisimust</i>	addiction	зависимость
<i>maqsad</i>	objective	مَقْصَدَ	<i>paviden</i>	behavior	поведение
<i>toqat</i>	patience	طَاقَةَ	<i>izmena</i>	betrayal	измена
<i>quwwat</i>	power	قُوَّةَ	<i>sobstvenost</i>	trait	собственность
<i>wijdũn</i>	conscience	وَجْدَانَ	<i>pirava</i>	right	право

## **Appendix B**

### **Verb stems**

**Table B.1:** Etymology of agreement suffixes and past-tense clitics (cf. Dodykhudoeva 1988: 6-8).

<u>Present suffixes</u>				<u>Past clitics</u>			
	<u>FORM</u>	<u>ETYMOLOGY</u>	<u>NOTES</u>		<u>FORM</u>	<u>ETYMOLOGY</u>	<u>NOTES</u>
1SG	<b>-um</b>	*-ami			<b>=um</b>	COP *ahmi	via contamination with CLTC *mai and 1SG PRS SFX
2SG	<b>-i</b>	*-ahi			<b>=(a)t</b>	CLTC *tai	
3SG	<b>-t/d</b>	*-ati	early loss of thematic vowel *a; creates i-umlaut position		<b>=i</b>	CLTC *hai OR DEM *i	
1PL	<b>-ām</b>	*-amahi			<b>=ām</b>	COP *hmahi	via contamination with PRS SFX
2PL	<b>-et</b>	*-aita	or *ata, and contamination with 3PL -en		<b>=et</b>	COP *=sta	via contamination with PRS SFX
3PL	<b>-en</b>	*-anti			<b>=en</b>	COP *anti	



## B.1 Consonant irregularities in verb stems

Table B.2: Consonant alternations in verb stems due to regular sound changes.

SOUND CHANGE	PRS. STEM	HISTORICALLY	PST STEM	HISTORICALLY	GLOSS
*s, š, z > ǰ, _t	<i>aboz-</i>	* <i>apa-āz-a-</i>	<i>abēǰt</i>	<i>apa-āš-ti</i>	swallow
	<i>birēz-</i>	* <i>apa-āz-a-</i>	<i>biroǰt</i>	<i>upari-āš-ta-</i>	send
	<i>boz-</i>	* <i>upa-rāz-aya-</i>	<i>boǰt</i>	<i>upari-āš-ti-</i>	drink
	<i>divez-</i>	* <i>ati-dais-a-</i>	<i>diviǰt</i>	<i>de-diš-ta-</i>	show
	<i>mez-</i>	* <i>maiz-a-</i>	<i>miǰt</i>	<i>miš-ta-</i>	urinate
	<i>raz-</i>	* <i>raz-a-</i>	<i>riǰt</i>	<i>riš-ta-</i>	pour
	<i>rinēs-</i>	* <i>fra-nās-aya-</i>	<i>rinūǰt</i>	<i>fra-nas-ta-</i>	forget
	<i>riwāz-</i>	* <i>fra-waz-a-</i>	<i>riwiǰt</i>	<i>fra-waš-ta-</i>	fly away
*rt, rd > ǰ, _t	<i>žoz-</i>	* <i>gāz-a-</i>	<i>žēǰt</i>	<i>gāš-ti-</i>	run
	<i>rarō-</i>	* <i>fra-rd-</i>	<i>ruǰt</i>	<i>fra-rd-ta</i>	take apart
	<i>šarō-</i>	* <i>xard-a-</i>	<i>šuǰt</i>	<i>xard-ta</i>	defecate
	<i>tidarō-</i>	* <i>ati-tard-a-</i>	<i>tiduǰt</i>	<i>ti-tard-ta</i>	rip
*w, r > ∅, _t	<i>zidrō-</i>	* <i>us-tard-a-</i>	<i>ziduǰt</i>	<i>us-tard-ta</i>	unravel
	<i>sāw-</i>	* <i>čyaw-a-</i>	<i>sut</i>	<i>č(y)u-ta</i>	go; become
	<i>sirāw-</i>	* <i>us-raw-a-</i>	<i>sirud</i>	<i>us-raw-ta</i>	separate
	<i>θāw-</i>	* <i>θaw-a-</i>	<i>θud</i>	<i>θaw-ta</i>	burn
	<i>wizāw-</i>	* <i>wi-zaw-a-</i>	<i>wizud</i>	<i>wi-zaw-ta</i>	go out (fire)
	<i>mār-</i>	* <i>mar-a-</i>	<i>mūd</i>	<i>mar-ta</i>	die
	<i>niǰpār-</i>	* <i>nir-par-a-</i>	<i>niǰpūd</i>	<i>nir-par-ta</i>	step on
	<i>vār-</i>	* <i>bar-a-</i>	<i>vūd</i>	<i>bar-ta</i>	bring
	<i>xār-</i>	* <i>xwar-a-</i>	<i>xūd</i>	<i>xwar-ta</i>	eat
	<i>ǰikar-</i>	* <i>xwar-a-</i>	<i>ǰikūd</i>	<i>xwar-ta</i>	search
	<i>zidār-</i>	* <i>us-tar-a-</i>	<i>zidūd</i>	<i>us-tar-ta</i>	sweep

Table B.3: Stem-consonant voicing alternations.

<u>ALTERNATION</u>	<u>PRS STEM</u>	<u>PST STEM</u>	<u>GLOSS</u>
<i>b~p</i>	<i>tilā<u>b</u>-</i>	<i>tilā<u>p</u>t</i>	ask for
<i>t~d</i>	<i>šā<u>n</u>d-</i> <i>yā<u>d</u>-</i>	<i>šī<u>n</u>t</i> <i>yā<u>t</u>(t)</i>	laugh come
<i>ǰ~ǰ̣</i>	<i>ka<u>ǰ</u>-</i> <i>ni<u>ǰ</u>u<u>ǰ̣</u>-</i> <i>virā<u>ǰ</u>-</i> <i>žiira<u>ǰ̣</u>-</i>	<i>ku<u>ǰ̣</u>t</i> <i>ni<u>ǰ</u>u<u>ǰ̣</u>t</i> <i>viru<u>ǰ̣</u>t</i> <i>ži<u>ǰ̣</u>ru<u>ǰ̣</u>t</i>	slaughter listen break bite; sting

## B.2 Vowel irregularities in verb stems

Table B.4: Present (non-3sg) stem vowels in irregular verbs.

STEM VOWEL	PRS STEM	GLOSS	HISTORICALLY	PRS STEM	GLOSS	HISTORICALLY
<i>a</i>	<i>angaxc-</i>	become stuck	* <i>ham-kux-sa-</i>	<i>tarō-</i>	take apart	* <i>tard-a-</i>
	<i>biðafc-</i>	close	* <i>upa-dab-sa-</i>	<i>viraŷ-</i>	break	* <i>bruš-a-</i>
	<i>cirafc-</i>	sting; burn	* <i>us-raf-sa-</i>	<i>xičand-</i>	cut	* <i>skɾnt-a-</i>
	<i>kaŷ-</i>	slaughter	* <i>kuš-a-</i>	<i>xiḱafc-</i>	bloom	* <i>škaf-sa-</i>
	<i>niðafc-</i>	stick	* <i>ni-daf-sa-</i>	<i>xiḱar-</i>	search	* <i>skar-a-</i>
	<i>nixarθ-</i>	collapse	* <i>nir-χrt-sa-</i>	<i>yad-</i>	come	* <i>yad-a-</i>
	<i>raz-</i>	pour	* <i>raz-a-</i>	<i>ziban-</i>	jump	* <i>us-ba-na-</i>
	<i>šarō-</i>	defecate	* <i>xard-a-</i>	<i>zidarō-</i>	tear	* <i>us-tard-a-</i>
<i>ā</i>	<i>siṭafc-</i>	fry (intr.)	* <i>us-taf-sa-</i>	<i>žiraŷ-</i>	bite; sting	* <i>gruš-a-</i>
	<i>anjāv-</i>	grab	* <i>ham-kap-a-</i>	<i>tār-</i>	clean up	* <i>tar-a-</i>
	<i>bāf-</i>	be able to	* <i>upa-af-a-</i>	<i>tāž-</i>	pull	* <i>tag-a-</i>
	<i>čān-</i>	dig	* <i>kan-a-</i>	<i>θāw-</i>	burn	* <i>θāw-a-</i>
	<i>firāp-</i>	reach	* <i>fra-ap-a-</i>	<i>vār-</i>	bring	* <i>bar-a-</i>
	<i>nāw-</i>	cry	* <i>naw-a-</i>	<i>wāf-</i>	weave	* <i>tag-a-</i>
	<i>niḱpār-</i>	step on	* <i>ni-spar-a-</i>	<i>wārv-</i>	boil (intr.)	* <i>warb-a-</i>
	<i>nāw-</i>	cry	* <i>naw-a-</i>	<i>wirāfc-</i>	stand	* <i>wi-rab-s-ya-</i>
	<i>riwāz-</i>	fly away	* <i>fra-waz-a-</i>	<i>wizāw-</i>	go out (fire)	* <i>wi-zaw-a-</i>
	<i>sāw-</i>	go	* <i>čyaw-a-</i>	<i>xār-</i>	eat	* <i>xwar-a-</i>
	<i>sirāw-</i>	separate	* <i>us-raw-a-</i>	<i>xičāf-</i>	burst	* <i>(s)kaf-a-</i>
	<i>sifān-</i>	rise	* <i>us-fan-a-</i>	<i>yān-</i>	grind	* <i>ar-na-</i>
	<i>šānd-</i>	laugh	* <i>xand-a-</i>			
<i>e</i>	<i>dives-</i>	show	* <i>ati-dais-a-</i>	<i>mez-</i>	urinate	* <i>maiz-a-</i>
	<i>xeb-</i>	beat	* <i>xšaip-a-</i>			
<i>ê</i>	<i>birêz-</i>	drink	* <i>upa-rāz-aya-</i>	<i>rinês-</i>	forget	* <i>fra-nās-aya-</i>
	<i>vidêdz-</i>	irrigate	* <i>abi-tāč-aya-</i>	<i>zêz-</i>	take	* <i>zāz-aya-</i>
<i>i</i>	<i>andidz-</i>	become stuck	* <i>ham-tač-a(ya)-</i>	<i>rimi-</i>	command	* <i>fra-maya-</i>
	<i>čis-</i>	watch	* <i>kas-a-</i>	<i>ris-</i>	stay	* <i>rič-ya-</i>
	<i>ði-</i>	fall	* <i>da-ya-</i>	<i>ti-</i>	go; wak	* <i>tač-a(ya)-</i>
	<i>naḱti-</i>	leave	* <i>niš-tač-a(ya)-</i>	<i>vi-</i>	be	* <i>buya-</i>
	<i>niθ-</i>	sit	* <i>ni-had-s-ya-</i>	<i>vir-i-</i>	find	* <i>abi-ar-aya-</i>

	<i>pi-</i>	rot	<i>*pu-ya-</i>	<i>ǰici-</i>	freeze (intr.)	<i>*straya-</i>
	<i>piðin-</i>	ignite (tr.)	<i>*pati-di-na-</i>	<i>ǰin-</i>	hear	<i>*sɣ-nau-</i>
	<i>piðis-</i>	ignite (intr.)	<i>*pati-di-sa-</i>	<i>zi-</i>	give birth	<i>*za-ya-</i>
	<i>pinidz-</i>	wear	<i>*pati-muč-a-</i>	<i>zini-</i>	wash	<i>*sna-ya-</i>
<i>ī</i>	<i>naǰǰis-</i>	pass	<i>*ham-kux-sa-</i>	<i>vīnd-</i>	connect	<i>*band-aya-</i>
	<i>parǰiv-</i>	grab	<i>*pari-kap-aya-</i>	<i>zīn-</i>	kill	<i>*jan-(a)ya-</i>
	<i>ricīθ-</i>	close	<i>*fra-θruθ-ya-</i>			
<i>o</i>	<i>aboz-</i>	swallow	<i>*apa-āz-a-</i>	<i>noǰ-</i>	wander	<i>*nārž-a-</i>
	<i>boz-</i>	send	<i>*apa-āz-a-</i>	<i>ǰofc-</i>	sleep	<i>*xwaf-sa-</i>
	<i>moz-</i>	build	<i>*māz-a-</i>	<i>ǰoy-</i>	read	<i>*srāy-aya-</i>
	<i>niwoz-</i>	play (instrument)	<i>*ni-wāz-a-</i>	<i>žoz-</i>	run	<i>*gāz-a-</i>
<i>ū</i>	<i>ðūdz-</i>	milk	<i>*dauǰ-a-</i>	<i>wizūn-</i>	know	<i>*awi-zān-a-</i>
	<i>niyūǰ-</i>	listen	<i>*ni-gauš-a-</i>			

Table B.5: Irregular third-singular present stems.

ALT.	PRS	PRS.3SG	GLOSS	ALT.	PRS	PRS.3SG	GLOSS
$\bar{a} > \bar{i}$	<i>anĵāv-</i>	<i>anĵīvd</i>	grab	$a > \bar{i}$	<i>kaŷ-</i>	<i>kīŷd</i>	slaughter
	<i>ḍāḍ-</i>	<i>ḍīd</i>	give		<i>šarḍ-</i>	<i>šīrōd</i>	defecate
	<i>niṣpār-</i>	<i>niṣpīrt</i>	step on		<i>tidarḍ-</i>	<i>tidīrōd</i>	tear
	<i>siṣān-</i>	<i>siṣīnt</i>	rise		<i>viṣaŷ-</i>	<i>viṣīŷd</i>	break (tr.)
	<i>vār-</i>	<i>vīrt</i>	bring		<i>ṣīkar-</i>	<i>ṣīkīrt</i>	search
	<i>xār-</i>	<i>xīrt</i>	eat		<i>ziban-</i>	<i>zībīnt</i>	jump
	<i>yān-</i>	<i>yīŷd</i>	grind		<i>žiraŷ-</i>	<i>žirīŷd</i>	bite; sting
$i > \bar{i}$	<i>zidār-</i>	<i>zidīrt</i>	sweep	$\bar{a} > o$			
	<i>andidz-</i>	<i>andīzd</i>	get up		<i>bāf-</i>	<i>boft</i>	be able to
	<i>naṣti-</i>	<i>naṣtīzd</i>	step on		<i>fīrāp-</i>	<i>fīropt</i>	reach
	<i>niṯ-</i>	<i>nīṯt</i>	rise		<i>riwāz-</i>	<i>riwozd</i>	fly away
	<i>piḍin-</i>	<i>piḍīnt</i>	ignite (tr.)		<i>tāž-</i>	<i>tožd</i>	pull
	<i>pinidz-</i>	<i>pinīzd</i>	wear		<i>wāf-</i>	<i>wofft</i>	weave
	<i>ti-</i>	<i>tīzd</i>	go; walk		<i>wīrāfc-</i>	<i>wīroft</i>	stand
$a > o$	<i>win-</i>	<i>wīnt</i>	see	$i > o$	<i>wārv-</i>	<i>worvd</i>	boil (intr.)
	<i>ṣin-</i>	<i>ṣīnt</i>	hear		<i>ṣīcāf-</i>	<i>ṣīcoft</i>	explode
	<i>nixarṯ-</i>	<i>nixorṯt</i>	collapse				
	<i>tarḍ-</i>	<i>torōd</i>	take apart		<i>čis-</i>	<i>čost</i>	watch
	<i>viṣaŷ-</i>	<i>viroŷd</i>	break (intr.)	$\bar{a} > \bar{u}$			
$\bar{a}w > \bar{u}$	<i>yad-</i>	<i>yoōd</i>	come		<i>šānd-</i>	<i>šūnt</i>	laugh
	<i>zidarḍ-</i>	<i>zidorōd</i>	unravel	$i > e$	<i>čān-</i>	<i>čūnt</i>	dig
	<i>nāw-</i>	<i>nūd</i>	cry		<i>ḍi-</i>	<i>ḍed</i>	fall
	<i>sāw-</i>	<i>wūd</i>	go; become		<i>pi-</i>	<i>ped</i>	rot
	<i>sirāw-</i>	<i>sirūd</i>	separate		<i>rimi-</i>	<i>rīmed</i>	command
$o > \hat{e}$	<i>ṯāw-</i>	<i>ṯūd</i>	burn		<i>vārḍi-</i>	<i>vārōed</i>	be able to
	<i>wizāw-</i>	<i>wizūd</i>	go out (of a fire)		<i>viṣi-</i>	<i>vīred</i>	find
$o > \hat{e}$	<i>aboz-</i>	<i>abēzd</i>	swallow		<i>ṣīci-</i>	<i>ṣīced</i>	freeze (intr.)
	<i>boz-</i>	<i>bēzd</i>	send		<i>zi-</i>	<i>zed</i>	give birth
	<i>yos-</i>	<i>yēst</i>	take		<i>zini-</i>	<i>zīned</i>	wash

Table B.6: Past stem vowels in irregular verbs.

STEM VOWEL	PRS STEM	GLOSS	HISTORICALLY	PRS STEM	GLOSS	HISTORICALLY
<b>a</b>	<i>yat</i>	come	* <i>yad-ta-?</i>			
<b>e</b>	<i>red</i>	stay	* <i>raik-ta-</i>	<i>wizent</i>	know	* <i>awi-zān-ti-?</i>
<b>ê</b>	<i>nêŷd</i> <i>niwêzd</i> <i>wêxt</i>	wander play (instrument) fall	* <i>wāš-ti</i> * <i>ni-wāz-ti-</i> * <i>wāš-ti-</i>	<i>ġêvd</i> <i>ġêyd</i> <i>žêxt</i>	sleep read run	* <i>xwāp-ti-?</i> * <i>srāy-ti-?</i> * <i>gāš-ti-</i>
<b>i</b>	<i>angixt</i> <i>divixt</i> <i>mixt</i>	become stuck show; seem urinate	* <i>han-kux-ti-</i> * <i>de-diš-ta-</i> * <i>miš-ta-</i>	<i>piðid</i> <i>rix̣t</i>	ignite (tr.) pour	* <i>pati-di-ta-?</i> * <i>riš-ta-?</i>
<b>ī</b>	<i>bīft</i> <i>čīnt</i> <i>fīrīpt</i> <i>mīzd</i> <i>naŷīd</i> <i>nīwd</i> <i>sīftd</i> <i>šīnt</i>	be able to dig reach build pass cry rise laugh	* <i>upa-aŷ-ti</i> * <i>kan-ti-</i> * <i>fra-ap-ti-</i> * <i>maz-ti-</i> * <i>nir-ga-ta-</i> * <i>naw-ti-</i> * <i>us-fa-ti-</i> * <i>xand-ta-</i>	<i>tīzd</i> <i>wīft</i> <i>wīnt</i> <i>wīrvd</i> <i>ġičīft</i> <i>ġīvd</i> <i>zīd</i> <i>žīvd</i>	pull weave see boil (intr.) burst beat kill spin (yarn)	* <i>taž-ti-</i> * <i>waf-ti-</i> * <i>win-ta-</i> * <i>warb-ti-</i> * <i>škaŷ-ti-</i> * <i>xšip-ta-</i> * <i>ga-ta-</i> * <i>žeb-ti-</i>
<b>o</b>	<i>aboxt</i> <i>boxt</i> <i>biroxt</i> <i>ðod</i> <i>parðod</i> <i>rimod</i>	swallow send drink give sell command	* <i>apa-āš-ta</i> * <i>apa-āš-ta-</i> * <i>upari-āš-ta-</i> * <i>dā-ta-</i> * <i>pari-dā-ta-</i> * <i>fra-mā-ta-</i>	<i>ġicod</i> <i>yod</i> <i>zinod</i> <i>zod</i> <i>zoxt</i>	freeze (intr.) take wash give birth take	* <i>strā-ta-</i> * <i>yā-ta-</i> * <i>snā-ta-</i> * <i>zā-ta-</i> * <i>zāš-ta-</i>
<b>u</b>	<i>kuxt</i> <i>niyuxt</i> <i>nixuxt</i> <i>pud</i> <i>riwuxt</i> <i>sirud</i> <i>sut</i> <i>šuxt</i> <i>tuxt</i> <i>θud</i>	get up listen collapse rot fly away separate go; become defecate tear burn	* <i>kuš-ta</i> * <i>ni-guš-ta-</i> * <i>ni-ḳt-ta-</i> * <i>pu-ta-</i> * <i>fra-wāš-ta-</i> * <i>us-ru-ta-</i> * <i>č(y)u-ta-</i> * <i>xard-ta-</i> * <i>tard-ta-</i> * <i>θu-ta-</i>	<i>viruxt</i> <i>vud</i> <i>wizud</i> <i>ġičuxt</i> <i>ġikuft</i> <i>ġud</i> <i>zibud</i> <i>ziduxt</i> <i>žiruxt</i>	break be go out (fire) cut blossom hear jump unravel bite; sting	* <i>bruš-ta-</i> * <i>bu-ta-</i> * <i>awi-zu-ta-</i> * <i>skart-ta-</i> * <i>skup-ta-</i> * <i>sru-ta-</i> * <i>us-bu-ta-</i> * <i>us-tard-ta-</i> * <i>gruš-ta-</i>

<i>ū</i>	<i>andūyd-</i>	get up	<i>*han-tak-ta-</i>	<i>rinūxt</i>	forget	<i>*fra-nas-ta-</i>
	<i>anjūvd</i>	grab	<i>*han-kab-ta-</i>	<i>rūpt</i>	sweep; clear	<i>*raub-ta-</i>
	<i>biđūvd</i>	close	<i>*upa-dab-ta-</i>	<i>sitūvd</i>	fry (intr.)	<i>*us-tab-ta-</i>
	<i>cirūvd</i>	sting; burn	<i>*us-rap-ta-</i>	<i>tūyd</i>	go; walk	<i>*tak-ta-</i>
	<i>čūd</i>	do	<i>*kar-ta-</i>	<i>vidūyd</i>	connect	<i>*bas-ta-</i>
	<i>čūxt</i>	watch	<i>*pati-muk-ta-</i>	<i>virū(y)d</i>	find	<i>*abi-ar-ta-</i>
	<i>ōūyd</i>	wear	<i>*pati-muk-ta-</i>	<i>vūd</i>	bring; carry	<i>*bar-ta-</i>
	<i>naxtūyd</i>	leave	<i>*niš-tak-ta-</i>	<i>vūst</i>	connect	<i>*bas-ta-</i>
	<i>niđūvd</i>	stick	<i>*ni-dab-ta-</i>	<i>wirūvd</i>	stand	<i>*awi-rab-ta-</i>
	<i>ničpūd</i>	step on	<i>*ni-spr-ta-</i>	<i>xūd</i>	eat	<i>*xwar-ta-</i>
	<i>nūst</i>	sit	<i>*ni-had-ta-</i>	<i>čikūd</i>	search	<i>*skḡ-ta-</i>
	<i>parjūvd</i>	grab	<i>*pari-gab-ta-</i>	<i>yūd</i>	grind	<i>*ar-ta-</i>
	<i>pinūyd</i>	wear	<i>*pati-muk-ta-</i>	<i>zidūd</i>	sweep	<i>*us-tar-ta-</i>
	<i>ricūst</i>	flee	<i>*frat-rad-ta-</i>			

Table B.7: Infinitive stems of irregular verbs.

STEM VOWEL	PRS STEM	GLOSS	HISTORICALLY	PRS STEM	GLOSS	HISTORICALLY
<b>a</b>	<i>yat</i>	come	—			
<b>e</b>	<i>red</i>	stay	* <i>raik-ti</i> -?	<i>wizent</i>	know	* <i>awi-zan-ti</i> -
<b>ê</b>	<i>abêxt</i>	swallow	* <i>apa-āš-ti</i> -	<i>wêxt</i>	fall	* <i>wāš-ti</i> -
	<i>bêxt</i>	send	—	<i>ġêvd</i>	sleep	* <i>xwāp-ti</i> -
	<i>birêxt</i>	drink	* <i>upari-āš-ti</i> -	<i>ġêyd</i>	read	* <i>srāy-ti</i> -
	<i>ôêd</i>	give / fall	* <i>dā-ti</i> -	<i>ġicêd</i>	freeze	* <i>strā-ti</i> -
	<i>nêŷd</i>	wander	* <i>nārž-ti</i> -	<i>yêd</i>	take; carry	* <i>yā-ti</i> -
	<i>niwêzd</i>	play (instrument)	* <i>ni-wāz-ti</i> -	<i>zêd</i>	give birth	* <i>zā-ti</i> -
	<i>parôêd</i>	sell	* <i>pari-dā-ti</i> -	<i>zêxt</i>	take	* <i>zāš-ti</i> -
	<i>rimêd</i>	command	* <i>fra-mā-ti</i> -	<i>žêxt</i>	run	* <i>gāš-ti</i> -
<b>i</b>	<i>virêd</i>	find	* <i>abi-ar-ti</i> -	<i>zinêd</i>	wash	* <i>snā-ti</i> -
	<i>angixt</i>	become stuck	* <i>han-kux-ti</i> -	<i>sixt</i>	forget	* <i>xard-ti</i> -
	<i>arrixt</i>	jump up	* <i>upa-af-ti</i> -	<i>tixt</i>	take apart	* <i>tard-ti</i> -
	<i>biðid</i>	close (intr.)	* <i>fra-pas-ti</i> -	<i>θid</i>	burn	* <i>θu-ti</i> -
	<i>divixt</i>	show; seem	* <i>di-ðes-ti</i> -	<i>vid</i>	be	* <i>bu-ti</i> -
	<i>kixt</i>	slaughter	* <i>kuš-ti</i> -	<i>virixt</i>	break	* <i>bruš-ti</i> -
	<i>mixt</i>	urinate	* <i>miš-ti</i> -	<i>wizid</i>	go out (fire)	* <i>awi-zu-ti</i> -
	<i>niyixt</i>	listen	* <i>ni-guš-ti</i> -	<i>ġiċixt</i>	cut	* <i>skart-ti</i> -
	<i>nixixt</i>	collapse	* <i>ni-xɣt-ti</i> -	<i>ġid</i>	hear	* <i>sru-ti</i> -
	<i>pid</i>	rot	* <i>pu-ti</i> -	<i>ġikift</i>	bloom	* <i>skup-ti</i> -
	<i>piðid</i>	ignite	* <i>pati-di-ti</i> -	<i>zibid</i>	jump	* <i>us-bu-ti</i> -
	<i>riwixt</i>	fly away	* <i>fra-was-ti</i> -	<i>zidiġt</i>	tear (intr.)	* <i>us-tard-ti</i> -
	<i>rixt</i>	pour	* <i>riš-ti</i> -	<i>žiriġt</i>	bite; sting	* <i>gruš-ti</i> -
	<i>sit(t)</i>	go; become	* <i>č(y)u-ti</i> -			
	<i>andīd</i>	get up	* <i>han-tak-ti</i> -	<i>rinīġt</i>	forget	* <i>fra-nas-ti</i> -
	<i>anġīvd</i>	grab	* <i>upa-dab-sa</i> -	<i>sifīd</i>	rise	* <i>us-fa-ti</i> -
<b>ī</b>	<i>bīft</i>	be able to	* <i>upa-af-ti</i> -	<i>siīvd</i>	fry (intr.)	* <i>us-taf-ti</i> -
	<i>cirīvd</i>	sting; burn	* <i>us-raf-sa</i> -	<i>šīnt</i>	laugh	* <i>xand-ti</i> -
	<i>čīd</i>	do	* <i>kar-ti</i> -	<i>tīd</i>	go; walk	* <i>tak-ti</i> -
	<i>čīnt</i>	dig	* <i>kan-ti</i> -	<i>tīžd</i>	pull	* <i>tāž-ti</i> -
	<i>čīxt</i>	watch	* <i>kas-ti</i> -	<i>vidīwd</i>	irrigate	* <i>abi-tak-ti</i> -
	<i>ðīwd</i>	milk	—	<i>vīd</i>	bring	* <i>bar-ti</i> -
	<i>fīrīpt</i>	reach	* <i>fra-ap-ti</i> -	<i>vīst</i>	connect	* <i>bad-ti</i> -
	<i>mīd</i>	die	* <i>mar-ti</i> -	<i>wirīvd</i>	stand	* <i>awi-rab-ti</i> -
	<i>mīzd</i>	build	* <i>maz-ti</i> -	<i>wīft</i>	weave	* <i>waf-ti</i> -
	<i>naġġīd</i>	pass	* <i>nir-ga-ti</i> -	<i>wīrvd</i>	boil	* <i>warb-ti</i> -



	<i>naǰtīd</i>	leave	<i>*niš-tak-ti-</i>	<i>xīd</i>	eat	<i>*xwar-ti-</i>
	<i>niðtīvd</i>	stick (intr.)	<i>*ni-dab-ti-</i>	<i>ǰičtīft</i>	burst	<i>*škaf-ti-</i>
	<i>niǰpīd</i>	step on	<i>*ni-spar-ti-</i>	<i>ǰikīd</i>	search	<i>*skar-ti-</i>
	<i>nīst</i>	sit	<i>*ni-had-ti-</i>	<i>ǰtīvd</i>	beat	<i>*xšip-ti-</i>
	<i>nīwd</i>	cry	<i>*naw-ti-</i>	<i>yīd</i>	grind	<i>*ar-ti-</i>
	<i>parjīvd</i>	take away	<i>*pari-kap-ti-</i>	<i>zīdīd</i>	sweep	<i>*us-tar-ti-</i>
	<i>pinīwd</i>	wear	<i>*pati-muk-ti-</i>	<i>zīd</i>	kill	<i>*ga-ti-</i>
	<i>ricīst</i>	flee	<i>*frat-rad-ti-</i>			
<i>ū</i>	<i>rūpt</i>	sweep (snow)	—			



# Bibliography

ABAEV, VASILY IVANOVICH, AND HAROLD WALTER BAILEY. 1985. Alans. *Encyclopedia Iranica* .

ABBI, ANVITA. 1985. Reduplicative structures: A phenomenon of the South Asian Linguistic Area. *Oceanic Linguistics Special Publications* 159–171.

AISSIN, JUDITH. to appear. Documenting topic and focus. *Language Documentation & Conservation* Special Issue, edited by Peter Jenks and Lev Michael.

AKKUŞ, FARUK, MOHAMMED SALIH, AND DAVID EMBICK. 2023. Case and the syntax of argument indexation: An analysis of Sorani Kurdish. Ms. UMass Amherst and University of Pennsylvania.

ALAMSHOEV, MUQBILSHO. 1994. *Система местоимений шугнанского языка (в сравнении с другими языками шугнано-рушанской группы памирских языков [The pronominal system of the Shughni language (in comparison with other languages of the Shughni-Rushani group of Pamir languages)]*. Dushanbe: Pomir.

ALAMSHOEV, MUQBILSHO. 2015. *Дастури Таълимии Забони Шугнонӣ [Pedagogical Grammar of the Shughni Language]*. Dushanbe: Akademiya Ilmhoyi Jumhuriyi Tojikiston [Academy of Sciences of Tajikistan].

ALAMSHOEV, MUQBILSHO MIRZOVAFOEVICH, AND SHERVONSHO MUQBILSHOEVICH ALAMSHOEV. 2020. *Шугнано-таджикско-русско-английский словарь в трех томах [Shughni-Tajik-Russian-English dictionary in three volumes]*. Dushanbe: Irfon.

ANDERSON, STEPHEN R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

- ANDERSON, STEPHEN R., AND EDWARD L. KEENAN. 1985. Deixis. In *Language typology and syntactic description: Grammatical categories and the lexicon*, ed. Timothy Shopen, 259–308. Cambridge: Cambridge University Press.
- ARLUND, PAMELA S. 2006. An acoustic, historical, and developmental analysis of Sarikol Tajik diphthongs. Doctoral Dissertation, University of Texas at Arlington, Arlington, TX.
- ARMAND, ELENA E. 2022. Глаголы ‘лить(ся)’ и ‘сыпать(ся)’ в шугнанском языке [Verbs meaning ‘pour’ (transitive and intransitive) in the Shughni language]. *Rodnoj yazyk: lingvisticheskij zhurnal* 73–108.
- ARTSTEIN, RON. 2002. A focus semantics for echo questions. In *Workshop on Information Structure in Context*, ed. Ágnes Bende-Farkas and Arndt Riester, 98–107. Stuttgart: IMS, University of Stuttgart.
- BAKHTIBEKOV, TUPCHI. 1979. *Грамматикаи забони шугнонӣ* [*Grammar of the Shughni language*]. Dushanbe: Donish.
- BAO, MINGZHEN. 2013. Acoustic evidences for glottalization of word-initial vowels in Shughni. *The Linguistics Journal* 7:250–272.
- BARIE, AMANDA. 2009. Exploring cleft sentences and other aspects of Shughni syntax. Master’s thesis, University of Kentucky, Lexington, KY.
- BARULIN, ALEKSANDR. 1980. The category of number in pronouns. In *Issledovanija v oblasti grammatiki i tipologii jazykov*, ed. Vladislav Mitrofanovich Andrjushchenko. Moscow: MGU.
- BARÐDAL, JÓHANNA. 2004. The semantics of the impersonal construction in Icelandic, German, and Faroese: Beyond thematic roles. In *Focus on Germanic Typology*, ed. Werner Abraham, 105–137. Berlin: Akademie Verlag.
- BASHIR, ELENA. 2006. Evidentiality in South Asian languages. In *Proceedings of the lfg06 conference*, ed. Miriam Butt and Tracy Holloway King. Stanford, CA: CSLI Publications, Stanford University.
- BASHIR, ELENA. 2007. Dardic. In *The Indo-Aryan Languages*, ed. Danesh Jain and George Cardona, 818–894. London: Routledge.

- BASHIR, ELENA. 2009. Wakhi. In *The Iranian Languages*, ed. Gernot Windfuhr, Routledge Language Family Series, 825–862. New York, NY: Routledge.
- BECK, SIGRID, AND MARGA REIS. 2018. On the form and interpretation of echo wh-questions. *Journal of Semantics* 35:369–408.
- BEEKES, ROBERT. 1988. *A grammar of Gatha-Avestan*. Leiden: Brill.
- BELYAEV, OLEG. 2014a. Anaphora in Ossetic correlatives and the typology of clause combining. In *On Diversity and Complexity of Languages Spoken in Europe and North and Central Asia*, ed. Pirkko Suihkonen and Lindsay J. Whaley, 275–310. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- BELYAEV, OLEG, AND DAG HAUG. 2014. Pronominal coreference in Ossetic correlatives and the syntax-semantics interface. In *Proceedings of the LFG14 Conference*, ed. Miriam Butt and Tracy Holloway King. Stanford, CA: CSLI Publications, Stanford University.
- BELYAEV, OLEG IGOREVICH. 2014b. Коррелятивная конструкция в осетинском языке в типологическом освещении [Correlative constructions in Ossetic from a typological perspective]. Doctoral Dissertation, Lomonsov Moscow State University, Moscow.
- BENVENISTE, ÉMILE. 1952. *La construction passive du parfait transitif [The passive construction of the transitive perfect]*. Paris: C. Klincksieck.
- BERTRAND, ANNE, YURIKA AONUKI, SIHWEI CHEN, HENRY DAVIS, JOASH GABARAGE, LAURA GRIFFIN, MARIANNE HUIJSMANS, LISA MATTHEWSON, DANIEL RESINGER, AND HOTZE RULLMANN. 2022. Nobody's perfect. *Languages* 7:148.
- BEYER, DANIELA, AND SIMONE BECK. 2011. A linguistic assessment of the Munji language of Afghanistan. *Language Documentation & Conservation* 6:38–103.
- BHATIA, TEJ K. 1993. *Punjabi*. London: Routledge.
- BHATT, RAJESH. 2003. Locality in correlatives. *Natural Language and Linguistic Theory* 21:485–541.

- BHATT, RAJESH. 2015. Relative clauses and correlatives. In *Syntax – theory and analysis*, 708–749. Berlin, München, Boston: De Gruyter Mouton.
- BIELMEIER, ROLAND. 1989. Sarmatisch, Alansich, Jassisch, und Altossetisch [Sarmatian, Alanian, Jassic, and Old Ossetian]. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 236–245. Wiesbaden: Reichert.
- BINNICK, ROBERT. 1991. *Time and the verb*. Oxford: Oxford University Press.
- BOBALJIK, JONATHAN DAVID. 2008. Where’s Phi? Agreement as a postsyntactic operation. In *Phi-theory: Phi features across interfaces and modules*, ed. Daniel Harbour, David Adger, and Susana Béjar, 295–328. Oxford: Oxford University Press.
- BOCHNAK, M. RYAN, AND MARTINA MARTINOVIĆ. 2018. Modal height and modal flavor: The case of Wolof *di*. In *Proceedings of sinn und bedeutung 22*, ed. Uli Sauerland and Stephanie Solt, volume 1, 223–240. Berlin: ZAS.
- BOLANDER, BROOK. 2016. English language policy as ideology in multilingual Khorog, Tajikistan. In *Discursive approaches to language policy*, ed. Elisabeth Barakos and Johann W. Unger, 253–274. London: Palgrave Macmillan.
- BOSSONG, GEORG. 1985. *Differentielle Objektmarkierung in den Neuiranischen Sprachen [Differential object marking in the New Iranian languages]*. Tübingen: Gunter Narr Verlag.
- BOSSONG, GEORG. 1991. Differential object marking in Romance and beyond. In *New analyses in Romance linguistics: Selected papers from the XVIII linguistic symposium on Romance languages*, ed. Dieter Wanner and Douglas Kibbee, number 69 in Current Issues in Linguistic Theory, 143–170. Amsterdam: John Benjamins.
- BOWERN, CLAIRE. 2008. *Linguistic fieldwork: A practical guide*. New York, NY: Palgrave Macmillan.
- BRANDENSTEIN, WILHELM. 1964. *Handbuch des Altpersischen [Handbook of Old Persian]*. Wiesbaden: O. Harrasowitz.
- BUBENIK, VIT. 1989. An interpretation of split ergativity in Indo-Iranian languages. *Diachronica* 6:181–212.
- BYNON, THEODORA. 1979. The ergative construction in Kurdish. *Bulletin of the School of Oriental and African Studies* 42:211–224.

- BYNON, THEODORA. 1980. From passive to active in Kurdish via the ergative construction. In *Papers from the Fourth International Conference on Historical Linguistics*, ed. Elizabeth Traugott, Rebecca Labrum, and Susan Shepherd. Amsterdam: John Benjamins.
- CANTERA, ALBERTO. 2017. The phonology of Iranian. In *Handbook of Comparative and Historical Indo-European Linguistics*, ed. Jared Klein, Brian Joseph, and Matthias Fritz, volume 481–503. De Gruyter Mouton.
- CHISTIAKOVA, DARIA G. 2022. Синтаксические и семантические свойства китики *i* в шугнанском языке [Syntactic and semantic features of the clitic *i* in Shughni]. *Literaturovedenie. Yazykoznanie. Kulturologiya* 5:67–91.
- CLEMENS, LAUREN EBY, JESSICA COON, PEDRO MATEO PEDRO, ADAM MILTON MORGAN, MARIA POLINSKY, GABRIELLE TANDET, AND MATTHEW WAGERS. 2015. Ergativity and the complexity of extraction: a view from Mayan. *Natural Language & Linguistic Theory* 33:417–467, 469–470.
- COMRIE, BERNARD. 1976. *Aspect*. Cambridge: Cambridge University Press.
- COMRIE, BERNARD. 1978. Ergativity. In *Syntactic typology: Studies in the phenomenology of language*, ed. Winfred P. Lehmann, 329–394. Austin: University of Texas Press.
- COMRIE, BERNARD. 1985. *Tense*. Cambridge: Cambridge University Press.
- COMRIE, BERNARD, AND TANIA KUTEVA. 2013a. Relativization on obliques. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Zenodo.
- COMRIE, BERNARD, AND TANIA KUTEVA. 2013b. Relativization on subjects. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Zenodo.
- COMRIE, BERNARD, AND TANIA KUTEVA. 2013c. Relativization strategies. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Zenodo.
- COON, JESSICA, AND CLINT PARKER. 2019. Case interactions in syntax. In *Oxford Research Encyclopedia of Linguistics*. New York: Oxford University Press.

- CORBETT, GREVILLE G. 1982. Gender in Russian: An account of gender specification and its relation to declension. *Russian Linguistics* 6:197–232.
- CORBETT, GREVILLE G. 1991. *Gender*. Cambridge: Cambridge University Press.
- CORBETT, GREVILLE G. 2000. *Number*. Cambridge: Cambridge University Press.
- CORBETT, GREVILLE G. 2014. Gender typology. In *The Expression of Gender*, ed. Greville G. Corbett, 87–130. Berlin: de Gruyter.
- COVER, REBECCA T. 2015. Semantic Fieldwork on TAM. In *Methodologies in semantic fieldwork*, ed. M. Ryan Bochnak and Lisa Matthewson, 233–268. New York: Oxford University Press.
- COVER, REBECCA T., AND JUDITH TONHAUSER. 2015. Theories of meaning in the field: Temporal and aspectual reference. In *Methodologies in semantic fieldwork*, ed. M. Ryan Bochnak and Lisa Matthewson, 306–350. New York: Oxford University Press.
- DABIR-MOGHADDAM, MOHAMMAD. 1997. Compound verbs in Persian. *Studies in the Linguistic Sciences* 27:25–59.
- DAHL, ÖSTEN. 2000. Animacy and the notion of semantic gender. In *Gender in Grammar and Cognition, Part 1: Approaches to Gender*, ed. Barbara Unterbeck, 99–116. Berlin: de Gruyter.
- DANIEL, MICHAEL, AND EDITH MORAVCSIK. 2013. The associative plural. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- DAVID, ANNE. 2014. *Descriptive grammar of Pashto and its dialects*, volume 1 of *Mouton-CASL Grammar Series I*. Berlin: Mouton de Gruyter.
- DEAL, AMY ROSE. 2015. Ergativity. In *Syntax: Theory and Analysis: An International Handbook*, ed. Artemis Alexiadou and Tibor Kiss, 654–707. Berlin: Mouton de Gruyter.
- DEAL, AMY ROSE. 2020. *A theory of indexical shift: Meaning, grammar, and cross-linguistic variation*. Cambridge, MA: MIT Press.



- DEGENER, ALMUTH. 2003. The Nuristani languages. In *Indo-Iranian Languages and Peoples*, ed. Nicholas Sims-Williams, 103–117. Oxford: Oxford University Press.
- DEMIRDACHE, HAMIDA, AND MYRIAM URIBE-ETXEBARRIA. 2014. Aspect and temporal anaphora. *Natural Language & Linguistic Theory* 32:855–895.
- DIESSEL, HOLGER. 1999. *Demonstratives: Form, function, and grammaticalization*. Amsterdam and Philadelphia: John Benjamins.
- DEN DIKKEN, MARCEL, AND THERESA O'NEILL. 2017. Copular constructions in syntax. In *Oxford research encyclopedia of linguistics*, ed. Mark Aronoff. Oxford: Oxford University Press.
- DIXON, ROBERT M. W. 1972. *The Dyirbal language of north Queensland*. Cambridge Studies in Linguistics 9. Cambridge: Cambridge University Press.
- DIXON, ROBERT M. W. 1979. Ergativity. *Language* 55:59–138.
- DIXON, ROBERT M. W. 1994. *Ergativity*. Cambridge Studies in Linguistics 9. Cambridge: Cambridge University Press.
- DIXON, ROBERT M. W. 2000. A typology of causatives. In *Changing valency*, ed. Robert M. W. Dixon and Alexandra Aikhenvald, 30–83. Cambridge: Cambridge University Press.
- DIXON, ROBERT M. W., AND ALEXANDRA AIKHENVALD. 2011. *Language at large: Essays on syntax and semantics*, chapter 2: A Typology of Argument-Determined Constructions, 44–85. Leiden: Brill.
- DODYKHUDOEV, RAKHIM KHALILOVICH. 1962. *Материалы по исторической фонетике шугнанского языка [Materials on the historical phonetics of the Shughni language]*. Dushanbe: V. I. Lenin Tajik National University.
- DODYKHUDOEV, RAKHIM KHALILOVICH. 1963. *Историческая фонетика шугнанского языка (Консонантизм) [Historical phonetics of the Shughni language (Consonantism)]*. Doctoral Dissertation, Leningrad State University, Leningrad.
- DODYKHUDOEV, RAKHIM KHALILOVICH. 1964a. Об отражении древнеиранской *rt* в шугнанском языке [On the reflex of Old Iranian *rt* in the Shughni language]. *Iranskaya Filologiya* 57–61.

- DODYKHUDOEV, RAKHIM KHALILOVICH. 1964b. Об отражении некоторых древнеиранских гласных в шугнанском языке [On the reflex of some Old Iranian vowels in Shughni]. *Iranskaya Filologiya*.
- DODYKHUDOEVA, LEILA R. 1988. *Шугнанский глагол в историческом освещении [The Shughni verb in a historical light]*. Dushanbe: Donish.
- DODYKHUDOEVA, LEILA R. 2003. Шугнанский язык [The Shughni language]. In *Yazyki rossiskoy federacii i sosednikh gosudarstv. entsiklopediya iii.*, 435–446. Moscow: Nauka.
- DODYKHUDOEVA, LEILA R., AND VALDIMIR B. IVANOV. 2008. Language immersion in endangered Pamiri communities: Elicitation strategies in tutorial work. *Endangered Languages and Language Learning* 101–108.
- DOST-MOHAMMAD, NOOR ALI, YUSUF AMIRI, AND MOHSEN HAFEZIAN. 2011. *Alifboyi zaboni shughni [Alphabet of the Shughni language]*. Montreal, QC, Canada: Sohravardi Foundation for Iranian Studies.
- DURKIN-MEISTERERNST, DESMOND. 2009. Khwarezmian. In *The Iranian Languages*, ed. Gernot Windfuhr, 336–376. London: Routledge.
- EDELMAN, JOY I. 1966. *Язгулямский язык [The Yazgulami language]*. Moscow: Nauka.
- EDELMAN, JOY I. 1971. *Язгулямско-русский словарь [Yazgulami-Russian dictionary]*. Moscow: Nauka.
- EDELMAN, JOY I. 1974. О конструкциях предложения в иранских языках [On clausal constructions in Iranian languages]. *Voprosy Yazykoznanija* 23–33.
- EDELMAN, JOY I. 1975. Категория наклонения [The category of mood]. In *Opyt istoriko-tipologicheskogo issledovaniya iranskikh yazykov, tom 2: Evolyuciya grammaticheskikh kategorii [essays on the historical-typological study of Iranian languages, vol. 2: Evolution of grammatical categories]*, ed. Vera Sergeevna Rastorgueva, 412–450. Moscow: Nauka.
- EDELMAN, JOY I. 1976. К истории язгулямский и шугнано-рушанских указательных местоимений [On the history of Yazghulami and Shughni-Rushani demonstrative pronouns]. In *Iranskoe yazykoznanie. istorija, etimologija, tipologija: k 75-letiju prof. V.I. Abaeva*, ed. Vera Sergeevna Rastorgueva, 85–96. Nauka.

- EDELMAN, JOY I. 1980a. History of the consonant systems of the North-Pamir languages. *Indo-Iranian Journal* 22:287–310.
- EDELMAN, JOY I. 1980b. К субстратному наследию центральноазиатского языкового союза [On the substratal heritage of the Central Asian Sprachbund]. *Voprosy Yazykoznanii* 5:21–32.
- EDELMAN, JOY I. 1986. *Сравнительная грамматика восточноиранских языков: Фонология* [Comparative grammar of Eastern Iranian languages: Phonology]. Nauka.
- EDELMAN, JOY I. 1987. Шугнано-рушанская группа языков [The Shughni-Rushani group of languages]. In *Osnovy iranskogo yazykoznanii* [iii/2]: *Novoiranskie yazyki: Vostochnaya gruppa* [Foundations of Iranian linguistics: New Iranian languages: Eastern group], ed. Vera Sergeevna Rastorgueva, 236–306. Moscow: Nauka.
- EDELMAN, JOY I. 1990. *Сравнительная грамматика восточноиранских языков: Морфология. Элементы синтаксиса* [Comparative grammar of Eastern Iranian languages: Morphology and elements of syntax]. Moscow: Nauka.
- EDELMAN, JOY I. 2000. Шугнано-рушанская группа [The Shughni-Rushani group]. In *Yazyki mira: Iranskie yazyki, tom 3: Vostochnie iranskie yazyki* [languages of the world: Iranian languages, vol. 3: Eastern Iranian languages], ed. Andrej A. Kibrik, 225–242. Moscow: Nauka.
- EDELMAN, JOY I. 2009. *Сравнительная грамматика восточноиранских языков: Лексика* [Comparative grammar of Eastern Iranian languages: Lexicon]. Moscow: Vostochnaya Literatura.
- EDELMAN, JOY I., AND LEILA R. DODYKHUOEVA. 2009a. The Pamir languages. In *The Iranian Languages*, ed. Gernot Windfuhr, Routledge Language Family Series, 773–786. New York, NY: Routledge.
- EDELMAN, JOY I., AND LEILA R. DODYKHUOEVA. 2009b. Shughni. In *The Iranian Languages*, ed. Gernot Windfuhr, Routledge Language Family Series, 825–862. New York, NY: Routledge.
- EDELMAN, JOY I., AND SHODIKHON PARVONAEVICH YUSUFBEKOV. 2000. Шугнанский язык [The Shughni language]. In *Yazyki mira iii*, 225–242. Moscow: Indrik.
- EMMERICK, RONALD E. 1989. Khotanese and Tumshuqese. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 204–229. Wiesbaden: Reichert.

- EMMERICK, RONALD E. 2009. Khotanese and Tumshuqese. In *The Iranian Languages*, ed. Gernot Windfuhr, 377–415. London: Routledge.
- ENÇ, MÜRVET. 1987. Anchoring conditions for tense. *Linguistic Inquiry* 18:633–657.
- ESTIGARRIBIA, BRUNO. 2020. *A grammar of Paraguayan Guaraní*. London: UCL Press.
- ESTIVAL, DOMINIQUE, AND JOHN MYHILL. 1988. Formal and functional aspects of the development from passive to ergative systems. In *Passive and voice*, ed. Masayoshi Shibatani, Typological Studies in Language 16, 441–524. Amsterdam: John Benjamins.
- FAIZOV, M. 1966. *Язык рушанцев советского памира [The language of the Rushanis of the Soviet Pamir]*. Dushanbe: Akademia nauk Tadzhikskoi SSSR, Institut yazika i literatury.
- FARRELL, TIMOTHY. 1995. Fading ergativity? A study of ergativity in Balochi. In *Subject, voice, and ergativity: Selected essays*, ed. David C. Bennett, Theodora Bynon, and B. George Hewitt, 218–243. London: School of Oriental and African Studies.
- FILLMORE, CHARLES J. 1982. Towards a descriptive framework for spatial deixis. In *Speech, place, and action: Studies in deixis and related topics*, ed. Robert J. Jarvella and Wolfgang Klein, 31–59. New York: John Wiley & Sons.
- FOLLI, RAFFAELLA, AND HEIDI HARLEY. 2005. Determinants of event type in Persian complex predicates. *Lingua* 115:1365–1401.
- FREEZE, RAY. 1992. Existentials and other locatives. *Language* 68:553–595.
- GAO, ERQIANG. 1963. Tajikeyu gaikuang [Overview of the Tajik language]. *Zhongguoyuwen* 2:161–175.
- GAO, ERQIANG. 1985. *Tajikeyu jianzhi [Outline of the Tajik language]*. Beijing: Nationalities Publishing House.
- GAUTHIOT, ROBERT, AND IVAN ZARUBIN. 1914. Отчет профессора Р. Гаутио и И. Зарубина по командировке на Памир летом 1914 [Report of R. Gauthiot's and I. Zarubin's trip to the Pamir in 1914]. *Izvestiya russkogo komiteta dlya izucheniya Sredney i Vostochnoy Azii*.

- GEBHARDT, LEWIS. 2009. Classifiers, plural, and definiteness in Persian. In *Aspects of Iranian linguistics*, ed. Simin Karimi, Vida Samiian, and Donald Stilo, 33–55. Newcastle: Simin Karimi and Vida Samiian and Donald Stilo.
- GEIGER, WILHELM. 1887. Zur Kenntniß der Pamir-Dialekte [On the Pamir dialects]. *Kuhns Zietschrift (KZ)* 28:208–210.
- GEIGER, WILHELM, AND ERNST KUHN (EDS.). 1895. *Grundriß der iranischen Philologie [Outline of Iranic philology]* 2 vols.. Straßburg: Karl J. Trübner.
- GHANIABADI, SAEED, JILA GHOMESHI, AND NIMA SADAT-TEHRANI. 2006. Reduplication in Persian: A morphological doubling approach. In *Proceedings of the 2006 annual conference of the Canadian Linguistic Association*, ed. Milica Radisic and Claire Gurski, 1–15.
- GIANNAKIDOU, ANASTASIA, AND LISA LAI-SHEN CHENG. 2006. (In)definiteness, Polarity, and the Role of wh-morphology in Free Choice. *Journal of Semantics* 23:135–186.
- GIBSON, EDWARD. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76.
- GIVÓN, THOMAS. 1981. Typology and functional domains. *Studies in Language* 5:163–193.
- GNOLI, GHERARDO. 1987. Avestan geography. *Encyclopedia Iranica*.
- GOLDBERG, ADELE. 2004. Words by default: The Persian complex predicate construction. In *Mismatches*, ed. Elaine Francis and Laura Michaelis. Amsterdam: Benjamins.
- GRENOBLE, LENORE, AND MATTHEW RILEY. 1996. The role of deictics in discourse coherence: French *voici/voilà* and russian *vot/von*. *Journal of Pragmatics* 25:819–838.
- GRIERSON, GEORGE ABRAHAM. 1920. *Ishkashimi, Zebaki, and Yazghulami: An account of three Eranian dialects*. London: Royal Asiatic Society.
- GRIERSON, GEORGE ABRAHAM. 1921. *Linguistic survey of India 10: Specimens of the languages of the Eranian family*. Calcutta: Government Press.

- GRJUNBERG, ALEKSANDR L. 1972. *Языки восточного Гиндукуша. Мунджанский язык. Тексты, словарь, грамматический очерк* [*The languages of the eastern Hindukush. Munji. Texts, dictionary, grammatical sketch*]. Leningrad: Nauka.
- GRJUNBERG, ALEKSANDR L., AND IVAN M. STEBLIN-KAMENSKII. 1976. *Ваханский язык. Тексты, словарь, грамматический очерк* [*The Wakhi language: Texts, dictionary, grammatical sketch*]. Moscow: Moscow.
- GUNDEL, JEANETTE K. 1980. Zero-anaphora in Russian: A case of topic-prominence? In *Proceedings of the Parasession on Anaphora*, ed. Jody Kreiman and Almerindo Ojeda, 139–146. Chicago Linguistics Society.
- GUNDEL, JEANETTE K. 1988. Universals of topic-comment structure. In *Studies in syntactic typology*, ed. Michael Hammond, Edith A. Moravcsik, and Jessica Wirth, 209–239. Amsterdam and Philadelphia: John Benjamins.
- HAIDARI, KHAIR MOHAMMAD. 2004. *Alifboyi zaboni shughni* [*Alphabet of the Shughni language*]. Kabul: Academy of Sciences of Afghanistan, Human Sciences Section, Center of Languages and Literature.
- HAIG, GEOFFREY. 1998. On the interaction of morphological and syntactic ergativity: Lessons from Kurdish. *Lingua* 105:149–173.
- HAIG, GEOFFREY. 2004. Alignment in Kurdish: A diachronic perspective. Unpublished Habilitationsschrift: Philosophische Fakultät der Christian-Albrechts-Universität zu Kiel.
- HAIG, GEOFFREY. 2008. *Alignment change in Iranian languages: A construction grammar approach*. Berlin: Walter de Gruyter.
- HALE, KEN. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory* 5–48.
- HALLIDAY, MICHAEL A. K., AND RUQAIYA HASSAN. 1976. *Cohesion in English*. London: Longman.
- HARRIS, ALICE C. 2010. Origins of differential unaccusative/unergative case marking: Implications for inateness. In *Hypothesis a/hypothesis b: Linguistic explorations in honor of David M. Perlmutter*, ed. Donna B. Gerdts, John Moore, and Maria Polinsky, 203–220. MIT Press.

- HASPELMATH, MARTIN. 2001. Non-canonical marking of core arguments in European languages. In *Non-canonical marking of subjects and objects*, ed. Alexandra Y. Aikhenvald, Robert M. W. Dixon, and Masayuki Onishi. Amsterdam and Philadelphia: John Benjamins.
- HENGVELD, KEES. 1992. *Non-verbal predication*. Berlin and New York: Mouton de Gruyter.
- HIMMELMANN, NIKOLAUS. 1996. Studies in anaphora. In *Typological studies in language*. Amsterdam: John Benjamins.
- HIMMELMANN, NIKOLAUS. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur [Deixis, article, and noun phrase: On the emergence of syntactic structure]*. Tübingen: Niemeyer.
- HIPPISLEY, ANDREW, AND GREGORY STUMP. 2011. Head & dependent marking and the Pamiri verb: A defaults-based account in network morphology. Paper presented at Explorations into Syntactic Government and Subcategorization. Cambridge.
- HOFFMAN, KARL. 1987. Avestan language. *Encyclopedia Iranica*.
- HUMBACH, HELMUT. 1989. Choresmian. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 193–203. Wiesbaden: Reichert.
- INGLESE, GUGLIELMO. 2022. Towards a typology of middle voice systems. *Linguistic Typology* 26:489–531.
- INOUE, KYOKO. 1979. An analysis of the English present perfect. *Linguistics* 17:561–589.
- IVANOV, DMITRIY L'VOVICH. 1884. Путешествие на Памир [Travels to the Pamir]. *Izvestiya Imperatorskogo Russkogo Geograficheskogo Obshchestva* 20:247–248.
- IVANOVA, VIOLA. 2023. Система гласных шугнанского языка [Vowel system of the Shughni language]. Ms. Higher School of Economics, Moscow.
- IZVORSKI, ROUMYANA. 1996. The syntax and semantics of correlative proforms. *North East Linguistics Society* 26.
- JAIN, DANESH, AND GEORGE CARDONA (EDS.). 2007. *The Indo-Aryan languages*. London and New York: Routledge.

- JAKOBSON, ROMAN. 1957. Shifters, verbal categories, and the Russian verb. In *Selected writings*, volume 2, 130–147. The Hague: Mouton.
- JERRO, KYLE. 2017. The causative-instrumental syncretism. *Journal of Linguistics* 53:751–788.
- JESPERSEN, OTTO. 1931. *A modern English grammar on historical principles, part iv*. Copenhagen: Munksgard.
- JESPERSEN, OTTO. 1965. *A modern English grammar on historical principles, part vi, morphology*. London: George Allen and Unwin Ltd.
- JÜGEL, THOMAS. 2009. Ergative remnants in Sorani Kurdish. *Orientalia Suecana* 58:142–158.
- JÜGEL, THOMAS. 2012. Die Entwicklung der Ergativkonstruktion im Alt- und Mitteliranischen: Eine korpusbasierte Untersuchung zu Kasus, Kongruenz und Satzbau [The development of the ergative construction in Old and Middle Iranian: A corpus-based approach]. Doctoral Dissertation, Goethe University Frankfurt, Frankfurt.
- JÜGEL, THOMAS, AND POLLET SAMVELIAN. 2020. Topic agreement, experiencer constructions, and the weight of clitics. In *Selected current studies in Iranian linguistics (current issues in linguistic theory)*, ed. Richard Larson, 137–153. New York: John Benjamins.
- KALIN, LAURA. 2018. Licensing and differential object marking: The view from Neo-Aramaic. *Syntax* 21:112–159.
- KAPOVIĆ, MATE, ANNA GIACALONE RAMAT, AND PAOLO RAMAT (EDS.). 2017. *The Indo-European languages*. Routledge Language Family Series. London: Routledge.
- KARAMKHUOEV, NODIR. 1973. *Бартагский язык [The Bartangi language]*. Dushanbe: Donish.
- KARAMSHOEV, DODKHUDO. 1962a. Имена в баджувском диалекте шугнанского языка [Nominals in the Bajuwi dialect of the Shughni language]. *Izvestiya AN Tadzhijskoy SSR, Otdelnie Obshchestvennykh Nauk* 17–35.
- KARAMSHOEV, DODKHUDO. 1962b. Отношение баджувского диалекта к шугнанскому языку [Relation of the Bajuwi dialect to the Shughni language]. *Iranskaya Filologiya*.
- KARAMSHOEV, DODKHUDO. 1963a. *Баджувский диалект шугнанского языка [The Bajuwi dialect of the Shughni language]*. Dushanbe: Izdatel'stvo Akademii Nauk Tadzhijskogo SSR.



- KARAMSHOEV, DODKHUDO. 1963b. О некоторых изменениях гласных в баджувском диалекте шугнанского языка [On some vowel changes in the Bajuwi dialect of the Shughni language]. *Iranskaya Filologiya* 62–72.
- KARAMSHOEV, DODKHUDO. 1966. О диалектном членении шугнанского языка [On the dialectal division of the Shughni language]. *Iranskaya Filologiya* 172–174.
- KARAMSHOEV, DODKHUDO. 1986. Категория рода в памирских языках (шугнаноруханская группа), выпуск 2 [*The category of gender in Pamir languages (Shughni-Rushani group), 2nd part*]. Dushanbe: Donish.
- KARAMSHOEV, DODKHUDO. 1988a. Шугнанско-русский словарь в трех томах [*Shughni-Russian Dictionary in Three Volumes*]. Leningrad: Nauka.
- KARAMSHOEV, DUDKHUDO. 1978. Категория рода в памирских языках (шугнано-рушанская группа), выпуск 1 [*The category of gender in Pamir languages (Shughni-Rushani group), 1st part*]. Dushanbe: Donish.
- KARAMSHOEV, DUDKHUDO. 1988b. Шугнанско-русский словарь в трех томах: Том 3 [*Shughni-Russian Dictionary in Three Volumes: Vol. 3*]. Leningrad: Nauka.
- KARAMSHOEV, DUDKHUDO. 1988c. Шугнанско-русский словарь в трех томах: Том 1 [*Shughni-Russian Dictionary in Three Volumes: Vol. 1*]. Leningrad: Nauka.
- KARAMSHOEV, DUDKHUDO. 1988d. Шугнанско-русский словарь в трех томах: Том 2 [*Shughni-Russian Dictionary in Three Volumes: Vol. 2*]. Leningrad: Nauka.
- KARAMSHOEV, DUDKHUDO, MUQBILSHO ALAMSHOEV, AND SH. ALAMASHOEV. 2019. *Xuǧnūni ziv Grammatika* [*Grammar of the Shughni Language*]. Dushanbe: Akademiya Ilmhoyi Jumhuriyi Tojikiston [Academy of Sciences of Tajikistan].
- KARIMI, SIMIN. 2003. On object specificity, scrambling, and scrambling in Persian. In *Word order and scrambling*, ed. Simin Karimi, 91–124. Malden, MA: Blackwell Publishing.
- KARIMI-DOOSTAN, GHOLAMHOSSEIN. 1999. Light verb constructions in Persian. Doctoral Dissertation, Essex University, Essex, England.
- KEANE, ELINOR. 2001. Echo words in Tamil. Doctoral Dissertation, Oxford University, Oxford.

- KEENAN, EDWARD L., AND BERNARD COMRIE. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63–99.
- KELLENS, JEAN. 1989. Avestique [Avestan]. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt. Wiesbaden: Reichert.
- KENESEI, ISTVÁN, ROBERT M. VAGO, AND ANNA FENYVESI. 1998. *Hungarian*. London: Routledge.
- KENT, ROLAND. 1953. *Old Persian*. New Haven, CT: American Oriental Society, second, revised edition edition.
- KIBRIK, ALEKSANDR EVGEN'EVIC, S. V. KODZASOV, I.P. OLOVIANNIKOVA, AND D.S. SAMEDOV. 1977. *Опыт структурного описания архинского языка, i: Лексика, фонетика [Essays on the structural description of the archi language, i: Lexicon, phonetics]*. Moscow: Izdatel'stvo Moskovskogo universiteta.
- KIBRIK, ANDREJ A. (ED.). 1997. *Языки мира: Иранские языки, Том 1: Юго-западные иранские языки [Languages of the world: Iranian languages, vol. 1: Southwestern Iranian languages]*. Moscow: Nauka.
- KIBRIK, ANDREJ A. (ED.). 1997-2000. *Языки мира: Иранские языки в 3 томах [Languages of the world: Iranian languages, 3 vols.жс]*. Moscow: Nauka.
- KIBRIK, ANDREJ A. (ED.). 1999. *Языки мира: Иранские языки, Том 2: Северо-западные иранские языки [Languages of the world: Iranian languages, vol. 2: Northwestern Iranian languages]*. Moscow: Nauka.
- KIBRIK, ANDREJ A. (ED.). 2000. *Языки мира: Иранские языки, Том 3: Восточные иранские языки [Languages of the world: Iranian languages, vol. 3: Eastern Iranian languages]*. Moscow: Nauka.
- KIM, DEBORAH. 2014. Subordination in Sarikoli. Master's thesis, University of North Dakota, Grand Forks, ND.
- KIM, DEBORAH. 2015. The Sarikoli reflexive pronoun. *Open Linguistics* 788–799.
- KIM, DEBORAH. 2017. Topics in syntax of Sarikoli. Doctoral Dissertation, Leiden University, Leiden.
- KLEIN, WOLFGANG. 1994. *Time in language*. New York: Routledge.
- KOBAYASHI, AKIKO. 2022. Past participle agreement in French and Italian: A two-Agree analysis. *Glossa: A Journal of General Linguistics* 7.

- KORN, AGNES. 2009. Marking of arguments in Balochi ergative and mixed constructions. In *Aspects of iranian linguistics*, ed. Simin Karimi, Vida Samiian, and Donald Stilo, 249–276. Newcastle: Cambridge Scholars Publishing.
- KRAMER, RUTH. 2020. Grammatical gender: A close look at gender assignment across languages. *Annual Review of Linguistics* 6:45–66.
- KULIKOV, LEONID. 2001. Causatives. In *Language typology and language universals*, ed. Martin Haspelmath, Ekkehard Konig, Wulf Oesterreicher, and Wolfgang Raible, volume 2 of *Handbooks of Linguistics and Communication Science*, 886–898. De Gruyter Mouton.
- KULIKOV, LEONID. 2017. Indo-Iranians and their origins. In *The Indo-European languages*, ed. Matéć, 205–213. Routledge.
- KUNO, SUSUMU. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.
- KURBANOV, H. 1976. *Ровшорский язык [The Roshorvi language]*. Dushanbe: Donish.
- KWON, NAYOUNG, YOONHYOUNG LEE, PETER C. GORDON, ROBERT KLUENDER, AND MARIA POLINSKY. 2010. Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of pre-nominal relative clauses in Korean. *Language* 86:546–582.
- LAPOLLA, RANDY J. 2009. Chinese as a topic-comment language. In *Studies of chinese linguistics: Functional approaches*, ed. Janet Zhiqun Xing, 9–22. Hong Kong: Hong Kong University Press.
- LAZARD, GILBERT. 1989. Le persan [Persian]. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 263–293. Wiesbaden: Riechert.
- LENTZ, WOLFGANG. 1933. *Pamir-Dialekte i: Materialien zur Kenntnis der Schugni-Gruppe [Pamir dialects 1: Materials on the Shughni group]*. Göttingen: Vandenhoeck & Ruprecht.
- LEVINSON, STEVEN C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- LI, CHARLES N. 1986. Direct and indirect speech: A functional study. In *Direct and indirect speech*, ed. Florian Coulmas, 29–45. Berlin, New York, Amsterdam: Mouton de Gruyter.

- LI, CHARLES N., AND SANDRA A. THOMPSON. 1976. Subject and topic: A new typology of language. In *Subject and topic*, 458–489. New York: Academic Press.
- LIPTÁK, ANIKÓ. 2012. Correlative topicalization. *Acta Linguistica Hungarica* 59:245–302.
- LORIMER, DAVID L. R. 1935–1938. *The Burushaski Language I: Introduction and Grammar; II: Texts and Translation; III: Vocabularies and Index*. Instituttet for sammenlignende kulturforskning.
- LORIMER, DAVID ROCKHART ROBINSON. 1958. *The Wakhi language*. London: School of Oriental and African Studies, University of London.
- LYONS, CHRISTOPHER. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- LYONS, JOHN. 1977. *Semantics*. 2 vols. Cambridge: Cambridge University Press.
- MACKENZIE, DAVID NEIL. 1991. Chorasmia iii: The Chorasmian language. *Encyclopedia Iranica*.
- MADDIESON, IAN, CAROLINE L. SMITH, AND NICOLA BESSELL. 2001. Aspects of the Phonetics of Tlingit. *Anthropological Linguistics* 43:135–176.
- MAKAROV, YURIJ, AND MAKSIM MELENCHENKO. 2021. Pamir Languages (pamiri.online), Moscow: School of Linguistics, HSE University. URL <https://pamiri.online/>.
- MAKAROV, YURIJ, MAKSIM MELENCHENKO, AND DMITRY NOVOKSHANOV. 2020. Digital resources for the Shughni language. In *Proceedings of the EURALI Workshop 2020*, 61–64.
- MAKAROV, YURIJ, AND VLADIMIR PLUNGIAN. 2023. К описанию шугнанской морфонологии: предварительные замечания [towards the description of Shughni morphophonology: Preliminary remarks]. In *”Вперёд и вверх по лестнице звучащей” Сборник статей к 80-летию Ольги Фёдоровны Кривновой [”Forward and up on the stairs of the sounding.” Essays for the 80th birthday of Olga Fyodorovna Krivnova]*, 178–193. Moscow: Buki Vedi.
- MALCHUKOV, ANDREJ, AND AKIO OGAWA. 2011. Towards a typology of impersonal constructions: A semantic map approach. In *Impersonal constructions: A cross-linguistic perspective*, ed. Andrej Malchukov and Anna Siewierska, 19–56. Amsterdam and Philadelphia: John Benjamins.

- MARTÍNEZ, JAVIER, AND MICHEL DE VAAN. 2014. *Introduction to Avestan*. Leiden and Boston: Brill.
- MATTHEWSON, LISA. 2004. On the methodologies of semantic fieldwork. *International Journal of American Linguistics* 369–415.
- MCGREGOR, WILLIAM BERNARD. 2009. Typology of ergativity. *Language and Linguistics Compass* 3:480–508.
- MILSARK, GARY. 1974. Existential sentences in English. Doctoral Dissertation, MIT, Cambridge, MA.
- MILSARK, GARY. 1977. Toward an explanation of certain peculiarities in the existential construction in English. *Linguistic Analysis* 3:1–30.
- MOGHADDAM, SAFIEH. 2016. Split Ergativity in Davani. Doctoral Dissertation, University of Toronto, Toronto.
- MOHAMMAD, JAN, AND SIMIN KARIMI. 1992. Light verbs are taking over: Complex verbs in Persian. In *Proceedings of WECOL*, volume 5.
- MORAVCSIK, EDITH A. 1974. Object-verb agreement. *Working Papers on Language Universals* 15:25–140.
- MORAVCSIK, EDITH A. 2003. A semantic analysis of associative plurals. *Studies in Language* 27:469–503.
- MORGENSTIERNE, GEORG. 1926. *Report on a linguistic mission to Afghanistan*. Oslo: Aschehoug.
- MORGENSTIERNE, GEORG. 1938a. *Indo-Iranian Frontier Languages*. Oslo: H. Aschehoug.
- MORGENSTIERNE, GEORG. 1938b. *Indo-Iranian Frontier Languages: Iranian Pamir languages, Yidgha-Munji, Sanglechi-Ishkashimi, and Wakhi*, volume 2. Oslo: H. Aschehoug.
- MORGENSTIERNE, GEORG. 1974. *Etymological Vocabulary of the Shughni Group*. Wiesbaden: Reichert.
- MUELLER, KATJA. 2014. Deixis in Shughni: Grammatical and semantic considerations. Doctoral Dissertation, University of North Dakota, Grand Forks, ND.
- MUELLER, KATJA, ELISABETH ABBESS, DANIEL PAUL, CALVIN TIESSEN, AND GABRIELA TIESSEN. 2010a. Ishkashimi: A father's language. In *Studies in Languages of Tajikistan*, ed. John M. Clifton. Dallas, TX: SIL International.

- MUELLER, KATJA, ELISABETH ABBESS, DANIEL PAUL, CALVIN TIESSEN, AND GABRIELA TIESSEN. 2010b. Language in community-oriented and contact-oriented domains: The case of the Shughni of Tajikistan. In *Studies in languages of tajikistan*, ed. John M. Clifton, 151–185. Dallas, TX: SIL International.
- MUNSHI, SADAF. 2006. Jammu and Kashmir Burushaski. Doctoral Dissertation, University of Texas at Austin, Austin, TX.
- MUNSHI, SADAF. 2019. *Srinagar Burushaski: A descriptive and comparative account with analyzed texts*. Leiden and Boston: Brill.
- NAWATA, TETSUO. 1979. *Shughni*. Asian and African Grammatical Manual, No. 17. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University.
- NIE, YINING. 2020. Licensing arguments. Doctoral Dissertation, New York University, New York, NY.
- NOONAN, MICHAEL. 2007. Complementation. In *Language typology and syntactic description, vol. ii: Complex constructions*, ed. Timothy Shopen, 52–150. Cambridge: Cambridge University Press, 2 edition.
- NORRIS, MARK. 2014. A theory of nominal concord. Doctoral Dissertation, University of California at Santa Cruz, Santa Cruz, CA.
- NORRIS, MARK. 2019. A typological perspective on nominal concord. In *Proceedings of the Linguistic Society of America*, volume 4, 1–15.
- OGIHARA, TOSHIYUKI. 1995. The semantics of tense in embedded clauses. *Linguistic Inquiry* 26:663–679.
- OGIHARA, TOSHIYUKI. 1996. *Tense, attitudes, and scope*. Dordrecht: Kluwer Academic Publishers.
- OLSON, KAREN. 2017. Shughni phonology statement. Technical report, SIL International, Dallas, TX.
- ORANSKIĬ, IOSEF MIKHAILOVICH. 1975. *Die Neuiranischen Sprachen der Sowjetunion [The New Iranian Languages of the Soviet Union]*. The Hague: Mouton.
- PAKHALINA, TATIANA NIKOLAEVNA. 1959. *Ишканимский язык [The Ishkashimi language]*. Moscow: Izdatel'stvo Akademii Nauk SSSR.

- PAKHALINA, TATIANA NIKOLAEVNA. 1966. *Сарокольский язык [The Sarikoli language]*. Nauka.
- PAKHALINA, TATIANA NIKOLAEVNA. 1969. *Памирские языки [The Pamir languages]*. Moscow: Nauka.
- PAKHALINA, TATIANA NIKOLAEVNA. 1975. *Ваханский язык [The Wakhi language]*. Moscow: Nauka.
- PAKHALINA, TATIANA NIKOLAEVNA. 1983. *Исследование по сравнительно-исторической фонетике памирских языков [Study on the comparative-historical phonetics of the Pamir languages]*. Moscow: Nauka.
- PAKHALINA, TATIANA NIKOLAEVNA. 1989. *Сравнительно-историческая морфология памирских языков [Comparative-historical morphology of the Pamir languages]*. Moscow: Nauka.
- PALMER, TIMOTHY. 2016. Verbal morphology and grammatical aspect in Sarikoli. Master's thesis, University of North Dakota, Grand Forks, ND.
- PANDHARIPANDE, RAJESHWARI V. 1997. *Marathi*. London: Routledge.
- PARKER, CLINTON. 2020. Vestigial ergativity in Shughni: At the intersection of alignment, clitic doubling, and feature-driven movement. *Glossa: A Journal of General Linguistics* 5.
- PAYNE, JOHN. R. 1980. Pamir languages. In *Compendium linguarum Iranicum*, ed. Rüdiger Schmitt, 417–444. Reichert.
- PAYNE, JOHN. R. 1989. The decay of ergativity in Pamir languages. *Lingua* 51:147–186.
- PAYNE, THOMAS E. 1997. *Describing morphosyntax*. Cambridge: Cambridge University Press.
- PERLMUTTER, DAVID. 1978. Impersonal passives and the unaccusative hypothesis. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, ed. Jeri J. Jaeger, 157–189. Berkeley, CA: Berkeley Linguistics Society.
- PIREIKO, LIYA ALEKSANDROVNA. 1975. Категория лица, категория залога [category of person and category of voice]. In *Опыт историко-типологического исследования иранских языков, Том ii [Essays in historical-typological research of Iranian languages, Volume ii]*, 250–336. Moscow: Nauka.

- PLATTS, JOHN T., AND GEORGE S. A. RANKING. 1911. *A grammar of the Persian language*. Oxford: Clarendon Press.
- PUSTET, REGINA. 2003. *Copulas: Universals in the categorization of the lexicon*. Oxford studies in typology and linguistic theory. Author: Oxford University Press.
- RASEKHI, VAHIDEH. 2018. Ellipsis and information structure: Evidence from Persian. Doctoral Dissertation, Stony Brook University, Stony Brook, NY.
- RASTORGUEVA, VERA SERGEEVNA. 1953. Краткий очерк грамматики персидского языка [A short sketch of the grammar of Persian]. In *Персидско-русский словарь [Persian-Russian dictionary]*, ed. B. V. Miller, 615–668. Gosudarstvennoe izdatel'stvo inostrannykh i nacional'nykh slovarej, 2nd revised enlarged edition edition.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1975a. *Опыт историко-типологического исследования иранских языков, Том 1: Фонология, эволюция морфологического типа [Essays on the historical-typological study of Iranian languages, vol. 1 phonology, evolution of morphological type]*. Moscow: Nauka.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1975b. *Опыт историко-типологического исследования иранских языков, Том 2: Эволюция грамматических категорий [Essays on the historical-typological study of Iranian languages, vol. 2: Evolution of grammatical categories]*. Moscow: Nauka.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1975c. *Опыт историко-типологического исследования иранских языков в 2 томах [Essays on the historical-typological study of Iranian languages], 2 vols..* Moscow: Nauka.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1979. *Основы иранского языкознания [i]: Древнеиранские языки [Foundations of Iranian linguistics: Old Iranian]*. Moscow: Nauka.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1979–1997. *Основы иранского языкознания в 6 томах [Fundamentals of Iranian linguistics], 6 vols..* Moscow: Nauka.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1981. *Основы иранского языкознания [ii]: Среднеиранские Языки [Foundations of Iranian linguistics: Middle Iranian]*. Moscow: Nauka.



- RASTORGUEVA, VERA SERGEEVNA (ED.). 1982. *Основы иранского языкознания [iii]: Новоиранские Языки: Западная группа, прикаспийские языки* [Foundations of Iranian linguistics: New Iranian languages: Western group, Caspian languages]. Moscow: Nauka.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1987. *Основы иранского языкознания [iii/2]: Новоиранские языки: Восточная группа* [Foundations of Iranian linguistics: New Iranian languages: Eastern group]. Moscow: Nauka.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1991. *Основы иранского языкознания [iv/1]: Новоиранские языки: Северо-западная группа, i* [Foundations of Iranian linguistics: New Iranian languages: Northwestern group i]. Moscow: Nauka.
- RASTORGUEVA, VERA SERGEEVNA (ED.). 1997. *Основы иранского языкознания [iv/2]: Новоиранские языки: Северо-западная группа, ii* [Foundations of Iranian linguistics: New Iranian languages: Northwestern group ii]. Moscow: Nauka.
- REICHENBACH, HANS. 1947. *Elements of symbolic logic*. Berkeley, CA: University of California Press.
- REIS, MARGA. 2017. On the analysis of echo questions. *Florida Linguistics Papers* 4:1–24.
- ROBERTS, JOHN R. 2009. *A study of Persian discourse structure*. Uppsala: Acta Universitatis Upsaliensis.
- ROBERTS, TAYLOR. 2000. Clitics and agreement. Doctoral Dissertation, MIT, Cambridge, MA.
- SALEMANN, CARL HERMANN. 1895. Шугнанский словарь Д.Л. Иванова [The Shughni dictionary of D. L. Ivanov]. *Vostochnye Zametki* 269–320.
- SALEMANN, CARL HERMANN. 1898. Отчет о поездке в Среднюю Азию в 1897 г. [Report on a trip to Central Asia in 1897]. *Izvestiya Imperatorskoy AN* 8:vi–viii.
- SALZMANN, MARTIN. 2017. Prolepsis. In *The Wiley Blackwell Companion to Syntax*, ed. Martin Everaert and Henk C. van Riemsdijk. Hoboken, NJ: John Wiley & Sons, Second edition.
- SCHEGLOFF, EMANUEL A., AND HARVEY SACKS. 1973. Opening up closings. *Semiotica* 7:289–327.

- SCHMIDTKE-BODE, KARSTEN. 2014. Complement clauses and complementation systems: A cross-linguistic study of grammatical organization. Doctoral Dissertation, Friedrich-Schiller-Universität Jena, Jena.
- SCHMIDTKE-BODE, KARSTEN, AND HOLGER DIESSEL. 2017. Cross-linguistic patterns in the structure, function, and position of (object) complement clauses. *Linguistics* 55:1–38.
- SCHMITT, RÜDIGER. 1989a. Andere altiranische dialekte [Other Old Iranian dialects]. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt. Wiesbaden: Reichert.
- SCHMITT, RÜDIGER. 1989b. *Compendium Linguarum Iranicum*. Wiesbaden: Reichert.
- SCHMITT, RÜDIGER. 2018. Scythian language. *Encyclopedia Iranica*.
- SCHRIFFIN, DEBORAH. 1987. *Discourse markers*. Cambridge: Cambridge University Press.
- SEDIGHI, ANOUSHA. 2005. Subject-predicative agreement restrictions in Persian. Doctoral Dissertation, University of Ottawa, Ottawa.
- SERGIENKO, ALEXANDER. 2023. Oblique subjects in Pamir languages. Master's thesis, National Research University, Higher School of Economics, Moscow.
- SHABANI-JADIDI, POUNEH. 2014. *Processing compound verbs in Persian: A psycholinguistic approach to complex predicates*. Leiden: Leiden University Press.
- SHAW, ROBERT. 1876. On the Ghalchah languages (Wakhi and Sarikoli). *Journal of the Asiatic Society of Bengal* 45:139–278.
- SHAW, ROBERT. 1877. On the Shighni (ghalchah) dialect. *Journal of the Asiatic Society of Bengal* 46:97–126.
- SHIBATANI, MASAYOSHI, AND PRASHANT PARDESHI. 2002. The causative continuum. In *Grammar of causation and interpersonal manipulation*, ed. Masayoshi Shibatani, 85–126. Amsterdam: John Benjamins.
- SIEWIERSKA, ANNA. 1984. *The passive: A comparative linguistic analysis*. London and Dover, NH: Croom Helm.
- SIMS-WILLIAMS, NICHOLAS. 1988. Bactrian language. *Encyclopedia Iranica*.

- SIMS-WILLIAMS, NICHOLAS. 1989a. Bactrian. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 230–235. Wiesbaden: Reichert.
- SIMS-WILLIAMS, NICHOLAS. 1989b. Sogdian. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 173–192. Wiesbaden: Reichert.
- SIMS-WILLIAMS, NICHOLAS. 1996. East Iranian Languages. *Encyclopedia Iranica* VII 649–652.
- SIMS-WILLIAMS, NICHOLAS. 2017. Iranian. In *The indo-european languages*, ed. Mate Kapović, 263–286. London: Routledge.
- SKJÆRVØ, PRODS OKTOR. 1989. Yidgha and Munji. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 411–416. Wiesbaden: Reichert.
- SKJÆRVØ, PRODS OKTOR. 2009a. Middle West Iranian. In *The Iranian Languages*, ed. Gernot Windfuhr, 196–278. London: Routledge.
- SKJÆRVØ, PRODS OKTOR. 2009b. Old Iranian. In *The Iranian Languages*, ed. Gernot Windfuhr, 43–195. London: Routledge.
- SKOPETEAS, STAVROS, INES FIEDLER, SAMANTHA HELLMUTH, ANNE SCHWARZ, RUBEN STOEL, GISBERT FANSELOW, CAROLINE FÉRY, AND MANFRED KRIFKA. 2006. *Questionnaire on information structure: Reference manual*. Universitätsverlag Potsdam, Potsdam.
- SKÖLD, HANNES. 1936. *Materialen zu den iranischen Pamirsprachen [Materials on the Iranian Pamir languages]*. Lund: Gleerup.
- SMIRNOVA, ANASTASIA. 2009. A case against ‘defective’ tense in the Bulgarian subjunctive. In *Proceedings of Formal Approaches to Slavic Languages (FASL) 17: The Yale Meeting*. Ann Arbor, MI: Michigan Slavic Publications.
- SMITH, CARLOTA S. 1997. *The parameter of aspect*. Dordrecht: Springer.
- SOHN, HO-MIN. 1994. *Korean*. London: Routledge.

- SOKOLOVA, VALENTINA S. 1953. *Очерки по фонетике иранских языков [Sketches of the phonetic systems of Iranian languages]*. Moscow-Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- SOKOLOVA, VALENTINA S. 1966. Шугнано-рушанская языковая группа [The Shughni-Rushani group of languages]. In *Yazyki narodov sssr [languages of the peoples of the ussr]*, ed. Viktor Vladimirovich Vinogradov. Nauka.
- SOKOLOVA, VALENTINA S. 1967. *Генетические отношения язгулямского языка и шугнанской языковой группы [Genetic relations between the Yazghulami language and the Shughni group]*. Leningrad: Nauka.
- SOKOLOVA, VALENTINA S. 1973. *Генетические отношения мунджанского языка и шугнано-язгулямской языковой группы [Genetic relations of the Munji language and the Shughni-Yazghulami group]*. Moscow: Nauka.
- SRIDHAR, SHIKARIPUR. 1990. *Kannada*. London: Routledge.
- STASSEN, LEON. 1997. *Intransitive predication*. Oxford: Clarendon.
- STEBLIN-KAMENSKIĬ, IVAN M. 1998. Eškāšemī. *Encyclopedia Iranica*.
- STUMP, GREGORY, AND ANDREW HIPPISEY. 2011. Valence sensitivity in Pamirian past tense inflection: A realizational analysis. In *Topics in iranian linguistics*, ed. Agnes Korn, Pollet Samvelian, and Geoffrey Haig, 103–115. Wiesbaden: Reichert.
- SUNDERMANN, WERNER. 1989a. Mittelpersisch [Middle Persian]. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 138–164. Wiesbaden: Reichert.
- SUNDERMANN, WERNER. 1989b. Parthisch [Parthian]. In *The Iranian Languages*, ed. Gernot Windfuhr, 114–137. London: Routledge.
- SUNDERMANN, WERNER. 1989c. Westmitteliranische sprachen [Western Middle Iranian languages]. In *Compendium Linguarum Iranicum*, ed. Rüdiger Schmitt, 106–113. Wiesbaden: Reichert.
- TAHERI-ARDALI, MORTAZA, HAMED RAHMANI, AND XU YI. 2014. The perception of prosodic focus in Persian. In *7th International Conference on Speech Prosody*. Dublin.

- TAHERI-ARDALI, MORTAZA, AND YI XU. 2012. Phonetic realization of prosodic focus in Persian. In *Proceedings of Speech Prosody 2012*, 326–329. Shanghai.
- TAHERKHANI, NEDA. 2019. A syntactic analysis of motion predicates in Southern Tati (Takestani dialect). Doctoral Dissertation, Purdue University, West Lafayette, IN.
- THACKSTON, WHEELER. 2009. *An introduction to Persian*. Bethesda, MD: Ibex Publishers, 4 edition.
- TOMASCHEK, WILHELM. 1880. Centralasiatische Studien 2: Die Pamir-Dialekte [Central Asian studies 2: The Pamir Dialects]. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien, philosophisch-historische Klasse* 96.
- TUCKER, G. RICHARD, WALLACE E. LAMBERT, AND ANDRÉ A. RIGAUT. 1977. *The French Speaker's Skill with Grammatical Gender: An Example of Rule-Governed Behavior*. The Hague: Mouton.
- VALENZUELA, PILAR M. 2000. Ergatividad escindida en wariapano, yaminawa y shipibo-konibo [Split ergativity in Wariapano, Yaminawa, and Shipibo-Conibo]. In *Essays on indigenous languages of lowland South America: Contributions to the 49th International Congress of Americanists in Quito*. Quito.
- VAN RIEMSDIJK, HANK C. 2017. Free relatives. In *The Wiley Blackwell Companion to Syntax*, ed. Martin Everaert and Henk C. van Riemsdijk, 1–46. Hoboken, NJ: John Wiley & Sons, Second edition.
- VERMEULEN, REIKO. 2013. On the position of topics in Japanese. *The Linguistic Review* 30:117–159.
- VON STECHOW, ARNIM. 1995. On the proper treatment of tense. In *Proceedings of salt 5*, ed. Mandy Simons and Teresa Galloway, 362–386. Ithaca, NY: Cornell University.
- ŠIMIK, RADEK. 2021. Free relatives. In *The Semantics Companion*, ed. Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, and Thomas Ede Zimmermann. Wiley-Blackwell.
- WENDTLAND, ANTJE. 2009. On ergativity in the Pamir languages. In *Aspects of iranian linguistics*, ed. Simin Karimi, Vida Samiian, and Donald Stilo, 419–434. Newcastle: Cambridge Scholars Publishing.

- WILKINS, DAVID P. 2018. The demonstrative questionnaire: “THIS” and “THAT” in comparative perspective. In *Demonstratives in cross-linguistic perspective*, ed. Stephen C. Levinson, Sarah Cutfield, Michael J. Dunn, N.J. Enfield, and Sérgio Meira, 43–71. Cambridge: Cambridge University Press.
- WILLIAMSON, PAUL. 2017. Munji. *Journal of the International Phonetic Association* 47:213–218.
- WINDFUHR, GERNOT. 1979. *Persian grammar: History and state of its study*. The Hague: Mouton.
- WINDFUHR, GERNOT. 2009a. Dialectology and topics. In *The Iranian Languages*, ed. Gernot Windfuhr, 5–42. London: Routledge.
- WINDFUHR, GERNOT, AND JOHN. R. PERRY. 2009. Persian and Tajik. In *The Iranian Languages*, ed. Gernot Windfuhr, 416–544. London: Routledge.
- WINDFUHR, GERNOT (ED.). 2009b. *The Iranian Languages*. London: Routledge.
- YOSHIDA, YUTAKA. 2009. Sogdian. In *The Iranian Languages*, ed. Gernot Windfuhr, 279–335. London: Routledge.
- YOSHIDA, YUTAKA. 2016. Sogdian language i: Description. *Encyclopedia Iranica*.
- YUSUFBEKOV, SHODIKHON PARVONAEVICH. 2000. Сангличский язык в синхронном и историческом освещении [The Sanglechi language from a synchronic and historical perspective]. Doctoral Dissertation, Russian Academy of Sciences, Institute of Linguistics.
- YUSUFBEKOV, SHODIKHON PARVONAEVICH, AND LEILA R. DODYKHUOEVA. 2008. Сангличский язык [The Sanglechi language]. In *Osnovy iranskogo yazykoznaniya, vii: Sredneiranskie i novoiranskie yazyki [fundamentals of Iranian linguistics, vol. 7: Middle Iranian and New Iranian languages]*, 110–234. Vostochnaya Literatura.
- ZARUBIN, IVAN. 1926. Образец припамирской народной поэзии [Sample of folk poetry from the Pamir]. *Izvestiya AN SSSR, 6 seriya* 12:1165–1170.
- ZARUBIN, IVAN. 1927. Одна орошорская сказка [A Roshorvi story]. *Vostochnye Zapiski* 1:297–318.
- ZARUBIN, IVAN. 1930. *Орошорские тексты и словарь [Roshorvi texts and dictionary]*. Leningrad: Pamir expedition of 1928.

- ZARUBIN, IVAN. 1937. *Бартагские и рушанские тексты и словарь [Bartangi and Rushani texts and dictionary]*. Moscow-Leningrad: Akademiya Nauk SSSR.
- ZARUBIN, IVAN. 1960. *Шугнанские тексты и словарь [Shughni texts and dictionary]*. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- ZÚÑIGA, FERNANDO, AND SEPPO KITILÄ. 2019. *Grammatical voice*. Cambridge: Cambridge University Press.
- ZWICKY, ARNOLD. 1977. *On clitics*. Bloomington, IN: Indiana University Linguistics Club.
- ZWICKY, ARNOLD, AND GEOFFREY PULLUM. 1983. Cliticization vs. inflection: English n't. *Language* 59:502–513.