

Anne Boyle David

Descriptive Grammar of Pashto and its Dialects

Mouton-CASL Grammar Series



Editors

Anne Boyle David

Claudia M. Brugman

Thomas J. Connors

Amalia Gnanadesikan

Volume 1

Anne Boyle David

Descriptive Grammar of Pashto and its Dialects

Edited by
Claudia M. Brugman

DE GRUYTER
MOUTON

Funding/Support: This material is based upon work supported, in whole or in part, with funding from the United States Government. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of Maryland, College Park and/or any agency or entity of the United States Government. Nothing in this report is intended to be and shall not be treated or construed as an endorsement or recommendation by the University of Maryland, United States Government, or the authors of the product, process, or service that is the subject of this report. No one may use any information contained or based on this report in advertisements or promotional materials related to any company product, process, or service or in support of other commercial purposes. The Contracting Officer's Representative for this project is John Walker, Government Technical Director at CASL, (301) 226-8912, jwalker@casl.umd.edu.

ISBN 978-1-61451-303-2
e-ISBN 978-1-61451-231-8

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliografische Information der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>.

© 2014 University of Maryland. All rights reserved.

Cover photo: A Khan Mamdi carpet from Pakistan. Image graciously provided by Bunyaad Oriental Rugs, committed to producing hand knotted Oriental rugs made by fairly paid adults (<http://rugs.tenthousandvillages.com>).

Printing: Hubert & Co. GmbH & Co. KG, Göttingen

♻️ Printed on acid-free paper

Printed in Germany

www.degruyter.com

To my teacher, Eric P. Hamp

Foreword

It is remarkable that, in this age of unprecedented global communication and interaction, the majority of the world's languages are as yet not adequately described. Without basic grammars and dictionaries, these languages and their communities of speakers are in a real sense inaccessible to the rest of the world. This state of affairs is antithetical to today's interconnected global mindset.

This series, undertaken as a critical part of the mission of the University of Maryland Center for Advanced Study of Language (CASL), is directed at remedying this problem. One goal of CASL's research is to provide detailed, coherent descriptions of languages that are little studied or for which descriptions are not available in English. Even where grammars for these languages do exist, in many instances they are decades out of date or limited in scope or detail.

While the criticality of linguistic descriptions is indisputable, the painstaking work of producing grammars for neglected and under-resourced languages is often insufficiently appreciated by scholars and graduate students more enamored of the latest theoretical advances and debates. Yet, without the foundation of accurate descriptions of real languages, theoretical work would have no meaning. Moreover, without professionally produced linguistic descriptions, technologically sophisticated tools such as those for automated translation and speech-to-text conversion are impossible. Such research requires time-consuming labor, meticulous description, and rigorous analysis.

It is hoped that this series will contribute, however modestly, to the ultimate goal of making every language of the world available to scholars, students, and language lovers of all kinds. I would like to take this opportunity to salute the linguists at CASL and around the world who subscribe to this vision as their life's work. It is truly a noble endeavor.

Richard D. Brecht
Founding Executive Director
University of Maryland Center for Advanced Study of Language

Series Editors' Preface

This series arose out of research conducted on several under-described languages at the University of Maryland Center for Advanced Study of Language. In commencing our work, we were surprised at how many of the world's major languages lack accessible descriptive resources such as reference grammars and bilingual dictionaries. Among the ongoing projects at the Center is the development of such resources for various under-described languages. This series of grammars presents some of the linguistic description we have undertaken to fill such gaps.

The languages covered by the series represent a broad range of language families and typological phenomena. They are spoken in areas of international significance, some in regions associated with political, social, or environmental instability. Providing resources for these languages is therefore of particular importance.

However, these circumstances often make it difficult to conduct intensive, in-country fieldwork. In cases where such fieldwork was impractical, the authors of that grammar have relied on close working relationships with native speakers, and, where possible, corpora of naturalistic speech and text. The conditions for data-gathering—and hence our approach to it—vary with the particular situation.

We found the descriptive state of each language in the series to be different from that of the others: in some cases, much work had been done, but had never been collected into a single overview; in other cases, virtually no materials in English existed. Similarly, the availability of source material in the target language varies widely: in some cases, literacy and media are very sparse, while for other communities plentiful written texts exist. The authors have worked with the available resources to provide descriptions as comprehensive as these materials, the native speaker consultants, and their own corpora allow.

One of our goals is for these grammars to reach a broad audience. For that reason the authors have worked to make the volumes accessible by providing extensive exemplification and theoretically neutral descriptions oriented to language learners as well as to linguists. All grammars in the series, furthermore, include the native orthography, accompanied where relevant by Romanization. While they are not intended as pedagogical grammars, we realize that in many cases they will supply that role as well.

Each of the grammars is presented as a springboard to further research, which for every language continues to be warranted. We hope that our empirical work will provide a base for theoretical, comparative, computational, and pedagogical developments in the future. We look forward to the publication of many such works.

Claudia M. Brugman
Thomas J. Conners
Anne Boyle David
Amalia E. Gnanadesikan

Preface

Pashto is a challenging language to study and describe, for several reasons. Its location in areas of rugged terrain, at the heart of a historical crossroad for traders, invaders, and migrating peoples, has led to alternating cycles of isolation and upheaval in the various Pashto-speaking regions. Furthermore, Pashto, a member of the Iranian language group itself, has undergone longterm influence from the many neighboring Indo-Aryan languages. Centuries of political turmoil, demographic shifts, and complex contact situations have contributed to significant dialectal variation. In addition, the current political situation makes in-country fieldwork highly problematic.

This grammar builds on the considerable previous work of many scholars, among them J.G. Lorimer, Georg Morgenstierne, Herbert Penzl, D.N. MacKenzie, D.A. Shafeev, Manfred Lorenz, Wilma Heston, Daniel Septfonds, Habibullah Tegey and Barbara Robson, Taylor Roberts, Farooq Babrakzai, Naseer Hoonar Pashtoon and Zeeya A. Pashtoon, and David Pate.¹ Without their diligent scholarship we would not have been able to begin this task. Among the features our grammar adds to the corpus of Pashto research are some new analyses of previously described data and coverage of all the regional dialects in a single volume, along with a detailed exposition of the dialectal situation, data presented in both native orthography and transcription, and finally, a formal grammar which can be used to feed a morphological parser, available online for download to purchasers of this volume.

The data for this grammar come from a wide range of printed resources, complemented by naturalistic corpora and work with native speaker consultants. We provide extensive examples and full paradigms, complete with full interlinearization of the example sentences: a native script line, a phonemic transcription, a morpheme-by-morpheme gloss line, and a free translation. Although native orthography is frequently omitted from descriptive grammars, it is particularly useful not only to the language expert but also to the language learner.

In our description we have attempted to be theory-neutral without being simplistic. Any abstract description of a language is necessarily informed by theory at some level. We aim to be theoretically informed in as broad a way as possible, such that the descriptions and explications contained within this grammar will be of use not only to descriptive linguists, but also to those from a variety of theoretic backgrounds. However, our primary loyalty is to the language being described and not to a particular theoretic approach to Language.

A descriptive grammar is never really finished. Two areas in particular that we wish we could devote more time to are syntax and prosody, although the description

¹ We have been made aware of a substantial literature on Pashto written in Russian (Lutz Rzehak, p.c.), including Lebedev 1996, Lebedev 2003, and Grjunberg 1987; we have unfortunately been able to consult Grjunberg only briefly and Lebedev not at all.

of Pashto syntax provided here is more detailed than previous overviews available in English and benefits from analyses of individual phenomena made by other scholars. There remains much work to be done on Pashto, and we view this volume as a springboard for scholars to continue working on this fascinating language in all its varieties.

Many people have helped in the creation of this book. The authors would like to thank all our colleagues at the University of Maryland Center for the Advanced Study of Language for their support—in particular, CASL's Executive Director, Amy Weinberg, and our founding Executive Director, Richard Brecht—as well as Pashto language experts in the United States Government. Our colleague and patron David Cox, who is much missed since his retirement, also deserves special mention for his assistance in promoting the idea of this series and for his constant, infectious enthusiasm for the enterprise of language description. Individuals who have had a part in producing this manuscript or advising our research include Nikki Adams, Farooq Babrakzai, Evelyn Browne, Katherine Burk, James Caron, Thomas Conners, Amalia Gnanadesikan, Wilma Heston, Mohammed Shahab Khan, Craig Kopris, Nathanael Lynn, Michael Marlo, Zeeya Pashtoon, Tristan Purvis, Shawna Rafalko, Paul Rodrigues, and Tamara Wehmeir. We also thank the generous native speakers of Pashto who worked with us patiently, and without whom this volume would not exist.

And finally, I would like to thank my co-authors for devoting themselves so diligently to this sometimes frustrating, always stimulating project, for being willing to disagree with me when I was wrong, and for never losing their senses of humor. I especially thank Claudia Brugman for her conscientious editing of this entire book and Sarah Goodman for her tireless work finding and making sense of naturally occurring example sentences. All of these people have worked with scrupulous care to ensure that as few mistakes and misstatements crept into this book as possible; those that remain fall to me alone.

Anne Boyle David

Contents

Foreword — vii

Series Editors' Preface — ix

Preface — xi

1 About this Grammar — 1

- 1.1 Scope of this book — 1
- 1.2 Orthography — 1
- 1.3 Tables and examples — 2
- 1.4 Abbreviations and symbols — 4

2 The Pashto Language — 7

- 2.1 Background — 7
- 2.2 Population of speakers — 7
 - 2.2.1 History and classification — 8
 - 2.2.2 Dialectal variation — 8

3 Phonology and Orthography — 9

- 3.1 Phonetics and phonology — 9
 - 3.1.1 Consonants — 9
 - 3.1.1.1 Inventory — 9
 - 3.1.1.2 “Elegant” consonants — 9
 - 3.1.2 Vowels — 10
 - 3.1.2.1 Inventory — 10
 - 3.1.2.2 Vowel transcription — 11
 - 3.1.2.3 “Elegant” vowels — 13
 - 3.1.2.4 Middle dialect vowels — 13
 - 3.1.3 Stress — 15
- 3.2 Orthography — 15
 - 3.2.1 The script — 15
 - 3.2.1.1 Letters unique to Pashto — 16
 - 3.2.1.2 Tables of letters and numerals — 18
 - 3.2.1.3 Representation of vowels — 25
 - 3.2.1.3.1 Word-initial vowels — 25
 - 3.2.1.3.2 Word-internal vowels — 26

- 3.2.1.3.3 Word-final vowels — 28
- 3.2.2 Rationale for transcription system — 28
- 3.2.3 Orthographic variation — 29

- 4 Pashto Dialects — 31**
 - 4.1 Introduction — 31
 - 4.2 Characterizing Pashto dialects — 31
 - 4.2.1 Dialect marking in this work — 34
 - 4.2.1.1 Dialect marking in tables — 34
 - 4.2.1.2 Dialect marking in interlinear examples — 34
 - 4.3 The four dialects of General Pashto — 35
 - 4.3.1 International differences — 36
 - 4.4 The Middle dialects — 37
 - 4.4.1 Middle dialect vowels — 37
 - 4.4.2 Middle dialect consonants — 39
 - 4.5 Other approaches — 40
 - 4.5.1 Two dialects — 40
 - 4.5.2 Three dialects — 42

- 5 Nouns — 45**
 - 5.1 Inflection — 45
 - 5.1.1 Gender — 45
 - 5.1.2 Number — 46
 - 5.1.3 Case — 47
 - 5.1.3.1 Direct case — 48
 - 5.1.3.2 Oblique case — 48
 - 5.1.3.3 Ablative case — 48
 - 5.1.3.4 Vocative case — 49
 - 5.1.3.5 Split ergativity — 50
 - 5.1.4 Animacy — 50
 - 5.2 Inflectional affixation — 51
 - 5.2.1 Introduction — 51
 - 5.2.2 Stem allomorphy and other morphophonemic alternations — 52
 - 5.2.3 Class I — 52
 - 5.2.3.1 Overview — 52
 - 5.2.3.2 Class I masculine nouns — 55
 - 5.2.3.2.1 Masculine animate nouns in General Pashto — 55
 - 5.2.3.2.2 Masculine animate nouns in Waziri — 61
 - 5.2.3.2.3 Masculine inanimate nouns in General Pashto and Waziri — 62
 - 5.2.3.3 Class I feminine nouns in General Pashto and Waziri — 67
 - 5.2.3.3.1 General Pashto Class I feminine animate nouns — 67
 - 5.2.3.3.2 General Pashto Class I feminine inanimate nouns — 69

- 5.2.3.3.3 Waziri Class I feminine nouns — 71
- 5.2.4 Class II — 73
 - 5.2.4.1 Overview — 73
 - 5.2.4.2 General Pashto Class IIa — 75
 - 5.2.4.3 General Pashto Class IIb — 79
- 5.2.5 Class III — 80
 - 5.2.5.1 Overview — 80
 - 5.2.5.1.1 Subclassification of Class III in General Pashto — 80
 - 5.2.5.2 Class IIIa — 84
 - 5.2.5.2.1 Masculine Class IIIa nouns — 84
 - 5.2.5.2.2 Feminine Class IIIa nouns — 85
 - 5.2.5.3 Class IIIb — 87
 - 5.2.5.3.1 Masculine Class IIIb nouns — 87
 - 5.2.5.3.2 Feminine Class IIIb nouns — 88
 - 5.2.5.4 Class III in Waziri — 89
- 5.2.6 Irregular nouns and irregular patterns in General Pashto — 92
 - 5.2.6.1 Kinship terms — 93
 - 5.2.6.2 Arabic borrowings — 95
- 5.3 Inflection and agreement of conjoined nouns — 97
- 5.4 Derivational morphology and loanwords — 98
 - 5.4.1 Derivational morphology of nouns — 98
 - 5.4.1.1 Nouns derived with suffixes — 98
 - 5.4.1.2 Compounds — 101
 - 5.4.2 Reduplication of nouns — 101
 - 5.4.3 Loanwords — 102
- 6 Adjectives and Other Noun Modifiers — 103**
 - 6.1 Introduction — 103
 - 6.2 Inflectional classes in General Pashto and Waziri — 103
 - 6.2.1 Inflectional classes of General Pashto adjectives — 103
 - 6.2.1.1 General Pashto Class I — 104
 - 6.2.1.1.1 Case-marking suffixes — 104
 - 6.2.1.1.2 Stem allomorphy — 104
 - 6.2.1.1.3 Class I forms with stem allomorphy — 105
 - 6.2.1.1.4 Class I forms without stem allomorphy — 107
 - 6.2.1.1.5 Animacy in Class I adjectives — 108
 - 6.2.1.2 General Pashto Class II — 109
 - 6.2.1.2.1 Case-marking suffixes — 109
 - 6.2.1.2.2 Stem allomorphy — 110
 - 6.2.1.2.3 Class II forms with stem allomorphy — 111
 - 6.2.1.2.4 Class II forms without stem allomorphy — 115
 - 6.2.1.3 General Pashto Class III — 116

6.2.1.3.1	Class IIIa —	116
6.2.1.3.2	Class IIIb —	118
6.2.1.4	General Pashto Class IV (non-declining adjectives) —	119
6.2.2	Inflectional classes of Waziri adjectives —	121
6.2.2.1	Waziri Class I —	121
6.2.2.2	Waziri Class II —	122
6.2.2.3	Waziri Class III —	122
6.3	Determiners and definiteness —	124
6.3.1	Demonstrative determiners —	124
6.3.2	The indefinite determiner <i>يو</i> /yaw/ ‘one’ —	130
6.4	Non-numerical noun quantifiers —	130
6.4.1	The quantifier <i>ټول</i> /ʈol-/ ‘all’ —	130
6.4.2	The quantifier <i>هر</i> /har/, /ar/ ‘every’ —	131
6.4.3	The quantifier <i>هيڅ</i> /hets/ ‘none’ —	132
6.5	Number names —	132
6.5.1	Cardinal numbers in Pashto —	133
6.5.1.1	Morphology —	133
6.5.1.2	Inventory —	137
6.5.2	Ordinal numbers in General Pashto and Waziri —	144
6.5.3	Reduplication of number names —	146
6.6	Interrogative adjectives —	147
6.7	Inflection of conjoined adjectives —	149
6.8	Derivation of adjectives —	150
6.8.1	Derivational suffixes —	150
6.8.1.1	Some Class I derivational suffixes —	150
6.8.1.2	Some Class IIIa derivational suffixes —	152
6.8.1.3	Some Class IV (non-declining) derivational suffixes —	152
6.8.2	Negators —	152
6.8.3	Compound adjectives —	153
6.8.4	Reduplication of adjectives —	153
6.9	Usage —	154
6.9.1	Attributive and predicative adjectives —	154
6.9.2	Zero-derivation of nouns from adjectives —	155
6.9.3	Comparatives and superlatives —	156
6.9.4	Adjectives as adverbs —	156
7	Pronouns —	157
7.1	Overview —	157
7.2	Strong personal pronouns —	157
7.2.1	Forms in General Pashto and Middle dialects —	157
7.2.2	Usage —	161
7.2.3	Strong possessive pronouns —	163

- 7.3 Weak personal pronouns — 164
- 7.3.1 Forms — 164
- 7.3.2 Usage — 166
- 7.3.2.1 Occurrence restrictions — 166
- 7.3.2.2 Possessive constructions — 168
- 7.4 Deictoids: را /rā/, در /dər/, and وړ /wər/ — 169
- 7.4.1 Oblique pronominal clitics — 170
- 7.4.2 Directional verbal clitics — 171
- 7.4.3 Deictic prefixes — 172
- 7.5 Demonstratives — 172
- 7.6 Interrogative pronouns — 176
- 7.7 Indefinite pronouns — 179
- 7.8 Relative pronouns — 181
- 7.9 Expressions of coreference — 181
- 7.10 Reciprocal pronouns — 184
- 7.11 Other pro-forms — 184

- 8 Verbs — 185**
- 8.1 Overview — 185
- 8.1.1 Properties of verbs — 185
- 8.1.2 Classifying verbs — 186
- 8.2 Verb components — 187
- 8.2.1 Structure of the verb — 187
- 8.2.2 Personal suffixes — 191
- 8.2.3 The infinitive — 194
- 8.2.4 Simplex and complex verbs — 196
- 8.2.4.1 Overview — 196
- 8.2.4.2 *a*-initial verbs — 196
- 8.2.4.3 Prefixed verbs — 198
- 8.2.4.4 Denominal verb constructions — 200
- 8.2.5 Conjugation classes — 203
- 8.2.5.1 Overview of conjugation classes — 203
- 8.2.5.2 First conjugation class in General Pashto — 204
- 8.2.5.3 First conjugation class in Middle dialects — 207
- 8.2.5.4 Second conjugation — 209
- 8.2.5.5 Third conjugation — 210
- 8.2.5.5.1 Forming the aorist in third conjugation verbs — 210
- 8.2.5.5.2 A special case of third conjugation verbs: infinitive/past participle + کېدل /kedāl/ — 212
- 8.2.6 Stem classes and the four bases — 212
- 8.2.6.1 The four verb bases — 214
- 8.2.6.2 Weak verbs (one stem) — 215

- 8.2.6.3 Strong verbs (more than one stem) — 217
- 8.2.6.3.1 Strong verbs with two stems — 218
- 8.2.6.3.2 Strong verbs with three or four stems — 219
- 8.2.6.3.3 List of strong verbs — 224
- 8.2.7 The causative morpheme — 227
- 8.2.8 The auxiliary *to be* and the verbalizers كبدل /kedâl/ and كول /kawâl/ — 228
- 8.2.8.1 Forms of *to be* — 229
- 8.2.8.2 Forms of كبدل /kedâl/ ‘to become’ — 234
- 8.2.8.3 Forms of كول /kawâl/ ‘to make; to do’ — 239
- 8.2.9 Participles — 246
- 8.2.9.1 Present participle — 246
- 8.2.9.2 Past participle — 247
- 8.2.9.3 Irregularities among past participles — 249
- 8.3 Simple verb constructions — 249
- 8.3.1 Present continuous — 249
- 8.3.2 Present aorist — 252
- 8.3.3 Past continuous — 254
- 8.3.4 Past aorist — 256
- 8.3.5 Continuous imperative — 258
- 8.3.6 Aorist imperative — 260
- 8.3.7 Continuous optative — 261
- 8.3.8 Aorist optative — 263
- 8.4 Compound verb constructions — 265
- 8.4.1 Perfect constructions — 265
- 8.4.1.1 Present perfect — 265
- 8.4.1.2 Past perfect — 267
- 8.4.2 Potential constructions — 267
- 8.4.2.1 Present potential — 268
- 8.4.2.2 Past potential — 268
- 8.5 Verb usage — 269
- 8.5.1 Uses of the verb *to be* — 269
- 8.5.1.1 *to be* as a copula — 269
- 8.5.1.2 *to be* as an auxiliary verb — 270
- 8.5.2 Simple verb constructions — 270
- 8.5.2.1 Present continuous — 270
- 8.5.2.1.1 Negation of present tense verbs — 272
- 8.5.2.1.2 Present continuous for expressing future events — 273
- 8.5.2.2 Present aorist — 274
- 8.5.2.2.1 Expressing the future with present aorist plus به /bə/ — 275
- 8.5.2.2.2 Other uses of the present aorist — 275
- 8.5.2.3 Past continuous — 279

- 8.5.2.4 Past aorist — 281
- 8.5.2.5 Imperative — 283
- 8.5.3 Compound constructions: perfect — 287
 - 8.5.3.1 Present perfect — 287
 - 8.5.3.2 Past perfect — 290
 - 8.5.3.3 Negation of perfect tenses — 291
- 8.5.4 Compound constructions: potential — 293
 - 8.5.4.1 Expressing potential present events — 293
 - 8.5.4.2 Expressing potential past events — 294
 - 8.5.4.3 Expressing potential future events — 295
 - 8.5.4.4 Negative — 298
- 8.5.5 Infinitives — 299
 - 8.5.5.1 Infinitives as nouns — 299
 - 8.5.5.2 The periphrastic passive — 300
- 8.5.6 Present participles — 303

- 9 Adpositions — 305**
 - 9.1 Overview — 305
 - 9.2 Adpositions and case assignment — 306
 - 9.2.1 Assignment of oblique case — 307
 - 9.2.2 Assignment of ablative case — 308
 - 9.2.3 Assignment of direct case — 309
 - 9.2.4 Mixed case-marking inside objects of adpositions — 311
 - 9.3 Prepositions — 312
 - 9.3.1 The prepositions /də/, /de/, /ye/, /e/ ‘of’ — 312
 - 9.3.1.1 The General Pashto preposition د /də/ — 312
 - 9.3.1.2 Complex adpositions using د /də/ — 313
 - 9.3.1.3 The Middle dialect prepositions /de/, /ye/, /e/ — 314
 - 9.3.1.4 Middle dialect complex adpositions using /ye/ — 315
 - 9.3.2 The General Pashto preposition له /lə/ ‘from’ — 316
 - 9.3.3 The preposition بې /be/ ‘without’ — 317
 - 9.3.4 The prepositions په /pə/, پر /pər/ — 317
 - 9.3.4.1 The locational په /pə/, پر /pər/ ‘on’ — 318
 - 9.3.4.2 The instrumental په /pə/ ‘with, by means of’ — 319
 - 9.3.4.3 The temporal په /pə/, پر /pər/ ‘at, on’ — 320
 - 9.3.4.4 With aspectual verbs — 322
 - 9.3.5 The preposition تر /tər/ ‘up to’ — 323
 - 9.3.6 The preposition لکه /leka/ ‘like’ — 324
 - 9.4 Postpositions — 325
 - 9.4.1 Overview — 325
 - 9.4.2 The postposition ته /ta/ ‘to, for’ — 326
 - 9.4.3 The postposition سره /səra/ ‘with’ — 327

- 9.4.4 The postposition *ضدی* /zidi/ ‘against’ — 327
- 9.4.5 The postposition *وړاندې* /wṛānde/ ‘before’ — 328
- 9.4.6 The postposition *غوندې* /ɣunde/ ‘like’ — 328
- 9.4.7 The postposition *باندي* /bānde/ — 328
- 9.4.8 Some additional postpositions in Middle dialects — 329
- 9.5 Circumpositions — 329
- 9.5.1 Circumpositions with *د* /də/ — 334
- 9.5.1.1 General Pashto simple circumpositions with *د* /də/ — 334
- 9.5.1.2 Middle dialect simple circumpositions with /ye/, /e/ — 336
- 9.5.1.3 Complex circumpositions with *د* /də/ — 337
- 9.5.2 General Pashto circumpositions with *له* /lə/ — 346
- 9.5.2.1 GP simple circumpositions with *له* /lə/ — 346
- 9.5.2.2 GP complex circumpositions with *له . . . نه* /lə...na/ — 350
- 9.5.3 Circumpositions with *په* /pə/ — 353
- 9.5.4 Circumpositions with *تر* /tər/ — 357
- 9.5.5 A Middle dialect circumposition with *و* /wə/ — 359
- 9.6 Coalesced adpositional phrases — 360
- 9.6.1 Pro-adpositional phrases — 360
- 9.6.2 The adpositional phrase *کړه* /kara/ ‘at the home of’ — 361
- 9.7 Adposition usage — 362
- 9.7.1 *نه* /na/ vs. ablative case-marking on object — 363
- 9.7.2 *د* /də/ and strong pronoun objects — 363
- 9.7.3 *په* /pə/ and *په . . . باندې* /pə... bānde/ used in a causative construction — 363
- 9.7.4 Omission of pronoun objects of adpositions — 364
- 9.7.5 Postpositions with oblique pronominal clitics — 365
- 10 Other Lexical Elements — 367**
- 10.1 Particles — 367
- 10.1.1 The existential particle *شته* /šta/ — 367
- 10.1.2 Modal particles — 369
- 10.1.2.1 The modal clitic *به* /bə/ — 369
- 10.1.2.2 The modal particles *دې* /de/ and *باید* /bāyad/ — 372
- 10.1.2.3 The modal particle *ښايي* /šāyi/ ‘may; must’ — 373
- 10.1.2.4 The optative particle *کاشکې* /kāške/ — 373
- 10.1.3 Affirmation particles — 374
- 10.1.4 The emphatic clitic *خو* /xo/ — 375
- 10.1.5 Vocative particles — 375
- 10.1.6 Interjections — 376
- 10.2 Adverbs — 377
- 10.2.1 Adverbs of time — 378

- 10.2.2 Adverbs of place — 380
- 10.2.3 Adverbs of manner — 382
- 10.2.3.1 The adverb سره /səra/ — 385
- 10.2.4 Adverbs of degree — 386
- 10.2.5 Adverbs borrowed from Arabic — 387
- 10.2.6 Adverbial interrogatives — 388
- 10.2.7 Adjectives as adverbs — 391
- 10.2.8 Reduplication of adverbs — 392
- 10.3 Reduplication — 393
- 10.3.1 Full (morphological) reduplication — 393
- 10.3.2 Partial reduplication: echo words — 393

- 11 Syntax — 399**
- 11.1 Overview — 399
- 11.2 Phrasal syntax — 399
- 11.2.1 Noun phrases — 399
- 11.2.2 Adpositional phrases — 400
- 11.2.3 Verb phrases — 401
- 11.2.3.1 Light verb constructions — 401
- 11.2.3.2 Elements in the verbal group — 403
- 11.2.3.2.1 The verbal group in General Pashto — 403
- 11.2.3.2.2 The verbal group in Middle dialects — 405
- 11.2.3.2.3 Negative placement in the aorist verb phrase — 406
- 11.3 Main clause sentence types — 410
- 11.3.1 Declaratives — 411
- 11.3.1.1 Order of elements in declaratives — 411
- 11.3.1.2 Order of elements in ditransitive main clauses — 413
- 11.3.1.3 Locative alternation — 413
- 11.3.1.4 Adpositional phrases with oblique pronominal clitics — 414
- 11.3.1.5 Passive clauses — 414
- 11.3.2 Interrogatives — 417
- 11.3.2.1 Yes-or-no questions with the particle آيا /āyā/ — 417
- 11.3.2.2 Information questions with interrogative pronouns — 418
- 11.3.2.3 Affirmation questions with the particle که نه /kə na/ — 419
- 11.3.3 Imperatives — 419
- 11.3.3.1 The imperative verb form — 419
- 11.3.3.2 The negative imperative particle مه /mā/ — 420
- 11.3.4 Generic and existential sentences with شته /šta/ — 421
- 11.3.5 Other principles of word order in main clauses — 421
- 11.3.5.1 Weak pronouns — 421
- 11.3.5.2 Particles — 424
- 11.3.5.3 Adpositional phrases in main clauses — 424

11.3.5.4	Interpretation of adpositional phrases headed by د /də/	424
11.4	Some subordinate clause types	425
11.4.1	Relative clauses	425
11.4.2	Noun complement clauses	428
11.4.3	Verb complement clauses	429
11.4.3.1	Reported speech	432
11.4.4	Subordinate clauses as modifiers	434
11.4.4.1	Subordinate clauses with time reference	434
11.4.4.2	Conditional and counterfactual clauses with که /kə/ ‘if’	435
11.4.4.3	Subordinate clauses with خُکِه /dʒəka/ ‘because’	436
11.4.4.4	Subordinate clauses expressing result	438
11.4.4.5	Subordinate clauses expressing reason	439
11.4.4.6	Subordinate clauses expressing purpose	439
11.4.4.7	Subordinate clauses of concession	440
11.5	Periphrastic causatives	440
11.6	Conjunction	442
11.6.1	Coordinating conjunctions	443
11.6.1.1	او /aw/ ‘and’	443
11.6.1.2	یا /yā/ ‘or’	444
11.6.1.3	خو /xo/ ‘but’	445
11.6.1.4	بیا (او) /(aw) byā/ ‘then’	446
11.6.2	Correlative conjunctions	446
11.7	Principles of case-marking and agreement	448
11.7.1	Tense-based case-marking and split ergativity	448
11.7.2	Agreement of conjoined items	450
11.7.3	Concordant adverbs	450
11.7.4	Case-marking patterns of verbs of sensation or preference	450
11.7.4.1	Four denominal verbs of sensation	450
11.7.4.2	Denominal خو نښېږ /xwaxeg-/ ‘like, enjoy’	451
11.7.4.3	Three more expressions of preference	452
11.7.5	An unergative or middle voice construction	453
A	Structure of this Grammar	455
A.1	Overview	455
A.2	Audience	456
A.3	More on uses of this grammar	457
A.3.1	The grammar as a basis for computational tools	457
A.3.1.1	Building a parser and generator	458
A.3.2	The grammar as a description	460
A.4	Spell correction	461
A.5	Grammar adaptation	462
A.5.1	Manual grammar building	462
A.5.2	Automated grammar adaptation	463

A.6	Formatting the grammar for viewing —	464
B	Sources of Pashto Data —	467
B.1	Sources of interlinear examples taken from the web —	467
B.2	List of web pages mined for language data —	474
	Bibliography —	477
	Index —	487

List of Figures

4.1	Pashto dialects —	32
4.2	Waziri metaphony —	39

List of Tables

1.1	Class I, stem alternation: سپک /spək/ ‘light’ —	3
1.2	Present continuous of کول /kawâl/ ‘to make, to do’ —	4
3.1	Pashto consonants —	9
3.2	“Elegant,” or formal, consonants —	10
3.3	Pashto vowels —	10
3.4	Pashto diphthongs —	11
3.5	Comparison among vowel transcription systems —	12
3.6	Comparison among M vowel transcription systems —	14
3.7	Pashto retroflex letters with panḍak —	16
3.8	Pashto letters with dot below and dot above —	17
3.9	Pashto affricate letters based on ح —	17
3.10	Pashto letters based on ى —	17
3.11	Pashto alphabet —	19
3.12	Positional variants of letters —	22

- 3.13 Pashto (Eastern Arabic) numerals — 24
- 3.14 Short vowel diacritics — 25
- 3.15 Word-initial vowels — 26
- 3.16 Pashto initial digraphs — 26
- 3.17 Word-internal vowels — 27
- 3.18 Word-final vowels — 28
- 4.1 Phonological variation among major Pashto dialects — 33
- 4.2 Correspondence between /i/ and /e/ in closed-class words — 33
- 4.3 Four dialects of General Pashto — 35
- 4.4 Tribal and geographical associations by dialect — 36
- 4.5 Waziri metaphony — 38
- 4.6 Middle dialect consonantal deviation from GP — 40
- 4.7 Alternative dialect codes — 40
- 4.8 Names for components of a two-dialect analysis of Pashto — 42
- 4.9 Eastern and Peshawar dialects compared — 43
- 4.10 Three-dialect approach — 44
- 4.11 Pronunciation of ځ — 44
- 5.1 GP Class I Masc. noun suffixes — 53
- 5.2 Middle dialect Class I Masc. noun suffixes — 53
- 5.3 GP Class I Fem. noun suffixes — 54
- 5.4 Middle dialect Class I Fem. noun suffixes — 54
- 5.5 GP Class I Masc. animate—morphophonemic alternations — 56
- 5.6 GP Class I Masc. animate: وېښته /wextá/ ‘hair’ — 56
- 5.7 GP Class I Masc. animate: بنده /bandá/ ‘slave’ — 57
- 5.8 GP Class I Masc. animate: ډاکو /ḍākú/ ‘bandit’ — 57
- 5.9 GP Class I Masc. animate: بازو /bāzú/ ‘arm’ — 58
- 5.10 GP Class I Masc. animate: میرزا /mirzā/ ‘clerk’ — 58
- 5.11 GP Class I Masc. animate: ماما /māmā/ ‘maternal uncle’ — 59
- 5.12 GP Class I Masc. animate: درزي /darzī/ ‘tailor’ — 59
- 5.13 GP Class I Masc. animate: پلندر /plandár/ ‘stepfather’ — 60
- 5.14 GP Class I Masc. animate: پیل /pil/ ‘elephant’ — 60
- 5.15 Waziri Class I Masc. animate: /šāgárd/ ‘student [male]’ — 61
- 5.16 Waziri Class I Masc. animate: /mirzā/ ‘clerk’ — 61
- 5.17 Waziri Class I Masc. animate: /kazī/ ‘judge’ — 62
- 5.18 Class I Masc. inanimate—stem allomorphy — 63
- 5.19 GP Class I Masc. inanimate: غر /ɣar/ ‘mountain’ — 63
- 5.20 GP Class I Masc. inanimate: دفتر /daftár/ ‘office’ — 64
- 5.21 GP Class I Masc. inanimate: غوړ /ɣwəg/ ‘ear’ — 65
- 5.22 GP Class I Masc. inanimate: پسه /psə/ ‘sheep’ — 65
- 5.23 Waziri Class I Masc. inanimate: /taɣár/ ‘rug’ — 65

- 5.24 Waziri Class I Masc. inanimate: /daftár/ ‘office’ — 66
- 5.25 Waziri Class I Masc. inanimate: /pəsâ/ ‘sheep’ — 66
- 5.26 Waziri Class I Masc. inanimate: /ɣar/ ‘mountain’ — 66
- 5.27 GP Class I Fem. animate—stem allomorphy — 67
- 5.28 GP Class I Fem. animate: خوآنبی /xwāxe/ ‘mother-in-law’ — 68
- 5.29 GP Class I Fem. animate/inanimate: بیزو /bizó/ ‘monkey’ — 68
- 5.30 GP Class I Fem. animate/inanimate: برېښنا /brexnā/ ‘lightning’ — 69
- 5.31 GP Class I Fem. inanimate—stem allomorphy — 69
- 5.32 Class I Fem. inanimate: اسپه /āspa/ ‘mare’ — 70
- 5.33 Class I Fem. inanimate: میاشت /miāšt/ ‘month’ — 70
- 5.34 Class I Fem. inanimate: رڼا /raṅā/ ‘light’ — 71
- 5.35 Waziri Class I Fem., consonant-final: /wradz/ ‘day’ — 72
- 5.36 Waziri Class I Fem., unstressed-vowel-final: /jǎrga/ ‘council’ — 72
- 5.37 Waziri Class I Fem., stressed-vowel-final: /žandā/ ‘flag’ — 72
- 5.38 GP Class IIa noun suffixes — 74
- 5.39 GP Class IIb noun suffixes — 74
- 5.40 GP Class IIa—stem allomorphy — 75
- 5.41 GP Class IIa inanimate: پالیز /pālez/ ‘kitchen garden’ — 76
- 5.42 GP Class IIa inanimate: پښتون /pəxtún/ ‘Pashtun’ — 76
- 5.43 GP Class IIa inanimate: تنور /tanúr/ ‘oven’ — 77
- 5.44 GP Class IIa animate/inanimate: غل /ɣal/ ‘thief’ — 78
- 5.45 GP Class IIa animate: مېلمه /melmá/ ‘guest’ — 78
- 5.46 GP Class IIb—stem allomorphy — 79
- 5.47 GP Class IIb: دښمن /duxmán/ ‘enemy’ — 80
- 5.48 GP Class IIIa Masc. noun suffixes — 81
- 5.49 GP Class IIIa Fem. noun suffixes — 82
- 5.50 GP Class IIIb Masc. noun suffixes — 83
- 5.51 GP Class IIIb Fem. noun suffixes — 83
- 5.52 GP Class IIIa Masc. animate: سپی /spây/ ‘dog [male]’ — 84
- 5.53 GP Class IIIa Masc. inanimate: گډی /gađáy/ ‘feast’ — 85
- 5.54 GP Class IIIa Fem. animate: سپی /spây/ ‘dog [female]’ — 86
- 5.55 GP Class IIIa Fem. animate: کورنی /koranây/ ‘family’ — 86
- 5.56 GP Class IIIa Fem. inanimate: خالاکي /čālāki/ ‘trickiness’ — 87
- 5.57 GP Class IIIb Masc.: ملگری /malgóray/ ‘friend [male]’ — 88
- 5.58 GP Class IIIb Fem.: ملگری /malgóre/ ‘friend [female]’ — 89
- 5.59 Middle dialect Class IIIa Masc. noun suffixes — 90
- 5.60 Middle dialect Class IIIb Masc. noun suffixes — 90
- 5.61 Middle dialect Class III Fem. noun suffixes: /-ây/ — 90
- 5.62 Middle dialect Class III Fem. noun suffixes: /-o/ — 91
- 5.63 Middle dialect Class III Fem. noun suffixes: /-yé/ — 91
- 5.64 Waziri Class IIIa Masc. animate: /xusây/ ‘calf’ — 91

- 5.65 Waziri Class III Masc. inanimate: /patáy/ ‘star’ — 91
- 5.66 Waziri Class III Fem.: /šaráy/ ‘woolen jacket’ — 91
- 5.67 Waziri Class IIIa Fem. inanimate: /xamto/ ‘cloth’ — 92
- 5.68 Waziri Class IIIa Fem. inanimate: /gutyé/ ‘ring’ — 92
- 5.69 Irregular Fem. kinship noun: مور /mor/ ‘mother’ — 93
- 5.70 Irregular Masc. kinship noun: وراره /wrārâ/ ‘brother's son’ — 94
- 5.71 Irregular Masc. kinship noun: زوی /zoy/ ‘son’ — 94
- 5.72 Irregular Fem. kinship noun: لور /lur/ ‘daughter’ — 95
- 5.73 Masc. Arabic loanword: مجاهد /mujāhíd/ ‘fighter’ — 96
- 5.74 Masc. Arabic loanword: موضوع /mawzô/ ‘topic’ — 96
- 5.75 Derived noun suffixes — 98
-
- 6.1 GP Class I adjective suffixes — 104
- 6.2 GP Class I, stem alternation: سپک /spæk/ ‘light’ — 106
- 6.3 GP Class I, حور /wár/ alternation: زرهور /zřawár/ ‘brave’ — 106
- 6.4 GP Class I, حجن /jân/ alternation: غمجن /ɣamjân/ ‘sad’ — 107
- 6.5 GP Class I, consonant-final adjective, no alternation: پلواند /palwând/ ‘fat’ — 108
- 6.6 GP Class II adjective suffixes — 109
- 6.7 GP Class II, -â-final adjective: تیره /terâ/ ‘sharp’ — 111
- 6.8 GP Class II, back vowel lowering: پوخ /pox/ ‘cooked, ripe’ — 112
- 6.9 GP Class II, back vowel lowering: روند /rund/ ‘blind’ — 112
- 6.10 GP Class II, back vowel breaking: تود /tod/ ‘hot’; stem = /tawd/ — 113
- 6.11 GP Class II, back vowel breaking: خور /xoř/ ‘sweet’; stem = /xwag/ or /xwaz/ — 114
- 6.12 GP Class II, Syncope II: سور /sur/ ‘red’ — 114
- 6.13 GP Class II, Syncope II and epenthesis: تريخ /trix/ ‘bitter’ — 115
- 6.14 GP Class II, consonant-final adjective, no stem allomorphy: اوم /um/ ‘raw, green’ — 116
- 6.15 GP Class IIIa adjective suffixes — 117
- 6.16 GP Class IIIa: زلمی /zalmáy/ ‘young’ — 117
- 6.17 GP Class IIIb adjective suffixes — 118
- 6.18 GP Class IIIb: سوی /sâway/ ‘burnt’ — 119
- 6.19 GP Class IV: بنايسنه /xâyistâ/ ‘pretty’ — 120
- 6.20 GP Class IV: يوازي /yawâzi/ ‘alone’ — 120
- 6.21 GP Class IV: هوسا /hosâ/ ‘comfortable’ — 121
- 6.22 Waziri Class I adjective suffixes — 122
- 6.23 Waziri Class II adjective suffixes — 122
- 6.24 Waziri Class III adjective suffixes — 123
- 6.25 Waziri Class III adjective with Fem. suffix /-ay/: /léwanay/ ‘mad’ — 123
- 6.26 Waziri Class III adjective with Fem. suffix /-yé/: /meranay/ ‘matri-lineally related’ — 123

- 6.27 Proximal demonstrative دا /dā/ — 124
- 6.28 Proximal demonstrative دغه /dāya/ — 126
- 6.29 Medial demonstratives — 128
- 6.30 Distal demonstratives — 129
- 6.31 GP يو /yaw/ ‘one’ — 133
- 6.32 Waziri /yaw/ ‘one’ — 134
- 6.33 GP دوه /dwa/ ‘two’ — 134
- 6.34 Waziri /dwa/ ‘two’ — 134
- 6.35 GP درې /dre/ ‘three’ — 135
- 6.36 GP شلور /tsalor/ ‘four’ — 135
- 6.37 Waziri plural forms of declinable number names — 137
- 6.38 GP numerals and number names — 138
- 6.39 Waziri number names that differ from GP forms — 144
- 6.40 GP Class I: لسم /lasám/ ‘tenth’ — 145
- 6.41 Waziri irregular ordinal number names — 146
- 6.42 GP interrogative adjectives — 147
- 6.43 Waziri interrogative adjectives — 147
-
- 7.1 GP strong pronouns, 1st and 2nd person — 158
- 7.2 Middle dialect strong pronouns, 1st and 2nd person — 158
- 7.3 GP strong pronouns, 3rd person — 158
- 7.4 Middle dialect strong pronouns, 3rd person — 159
- 7.5 Distal 3rd person pronoun هغه /hayá/ — 159
- 7.6 Possessive pronouns — 164
- 7.7 GP weak pronouns — 165
- 7.8 Middle dialect weak pronouns — 165
- 7.9 Distribution pattern for weak pronouns — 166
- 7.10 Oblique pronominal clitics — 171
- 7.11 GP proximal demonstrative دا /dā/ — 173
- 7.12 Middle dialect proximal demonstrative دا /dā/ — 173
- 7.13 GP proximal demonstrative دغه /dayá/ — 174
- 7.14 Waziri proximal demonstrative /daya/ — 174
- 7.15 Dzadrani proximal demonstrative /daya/ — 174
- 7.16 GP medial demonstrative هغه /hay-â/ — 175
- 7.17 Waziri medial demonstrative /ayə/ — 175
- 7.18 Dzadrani medial demonstrative /ayə/ — 175
- 7.19 GP distal demonstrative هوغه /huy-â/ — 176
- 7.20 GP human interrogative pronoun شوک /tsok/ — 177
- 7.21 Middle dialect human interrogative pronoun /tsok/ — 177
-
- 8.1 Structure of weak (one stem) verbs — 188
- 8.2 Structure of strong (more than one stem) verbs — 189

- 8.3 Stem shapes — 190
- 8.4 GP verbs: personal suffixes — 192
- 8.5 Waziri verbs: personal suffixes — 193
- 8.6 Dzadrani verbs: personal suffixes — 193
- 8.7 Infinitives — 195
- 8.8 Deictic verb prefixes — 198
- 8.9 Non-productive verb prefixes — 199
- 8.10 Examples of prefixed verbs — 200
- 8.11 Behavior of denominal verbs — 202
- 8.12 Examples of denominal verbs — 203
- 8.13 GP first conjugation verbs: present tense stems and aorist bases — 205
- 8.14 GP *a*-initial verbs (first conjugation): aorist bases — 206
- 8.15 Waziri verbs: forms corresponding to GP *a*-initial verbs — 208
- 8.16 Dzadrani verbs: forms corresponding to GP *a*-initial verbs — 208
- 8.17 Second conjugation verbs: aorist bases (present tense) — 209
- 8.18 Contracted third conjugation verbs: aorist stems (present tense) — 211
- 8.19 Verbs and their stems: strong and weak — 213
- 8.20 Weak verb bases: first conjugation (transitive) — 216
- 8.21 Weak verb bases: first conjugation (intransitive) — 216
- 8.22 Weak verb bases: second conjugation — 217
- 8.23 Strong verb bases: first conjugation (two stems) — 218
- 8.24 Strong suppletive verb, first conjugation (two stems) — 219
- 8.25 Strong verb bases: second conjugation (two stems) — 219
- 8.26 Strong verb bases: كَوَل /kawál/ ‘to do’ — 220
- 8.27 Strong verb bases: وِرَل /wɪrəl/ ‘to carry’ — 221
- 8.28 Strong verb bases: بِيَوَل /biwál/ ‘to lead away’ — 221
- 8.29 Strong verb bases: اِيَسْبُوَدَل /ixodál/ ‘to put’ — 221
- 8.30 Strong verb bases: كِبَدَل /kedál/ ‘to become’ — 222
- 8.31 Strong verb bases: تَلَل /tlál/ ‘to go’ — 222
- 8.32 Strong verb bases: رَاتَلَل /rātlál/ ‘to come’ — 222
- 8.33 Waziri strong verb bases: /kawál/ ‘to do’ — 223
- 8.34 Dzadrani strong verb bases: /kawál/ ‘to do’ — 223
- 8.35 Waziri strong verb bases: /kedál/ ‘to become’ — 223
- 8.36 Dzadrani strong verb bases: /kedál/ ‘to become’ — 224
- 8.37 Strong verbs — 224
- 8.38 Causative verbs — 228
- 8.39 GP present continuous of *to be* — 230
- 8.40 Waziri present continuous of *to be* — 230
- 8.41 Dzadrani present continuous of *to be* — 231
- 8.42 GP present aorist of *to be* (= present continuous except in 3rd person) — 231

- 8.43 Waziri present aorist of *to be* (= present continuous except in 3rd person) — 231
- 8.44 Dzadrani present aorist of *to be* — 232
- 8.45 GP past of *to be* — 232
- 8.46 Waziri past of *to be* — 233
- 8.47 Dzadrani past of *to be* — 233
- 8.48 Imperative of *to be* — 233
- 8.49 GP present continuous of كېدل /kedâl/ ‘to become’ — 234
- 8.50 Waziri present continuous of /kedâl/ ‘to become’ — 234
- 8.51 Dzadrani present continuous of /kedâl/ ‘to become’ — 235
- 8.52 GP present aorist of كېدل /kedâl/ ‘to become’ — 235
- 8.53 Waziri present aorist of /kedâl/ ‘to become’ — 235
- 8.54 Dzadrani present aorist of /kedâl/ ‘to become’ — 236
- 8.55 GP past continuous of كېدل /kedâl/ ‘to become’ — 237
- 8.56 Waziri past continuous of /kedâl/ ‘to become’ — 237
- 8.57 Dzadrani past continuous of /kedâl/ ‘to become’ — 238
- 8.58 GP past aorist of كېدل /kedâl/ ‘to become’ — 238
- 8.59 Dzadrani past aorist of /kedâl/ ‘to become’ — 239
- 8.60 GP present continuous of كول /kawâl/ ‘to make; to do’ — 239
- 8.61 Waziri present continuous of /kawâl/ ‘to make; to do’ — 240
- 8.62 Dzadrani present continuous of /kawâl/ ‘to make; to do’ — 240
- 8.63 GP present aorist of كول /kawâl/ ‘to make; to do’ — 241
- 8.64 Waziri present aorist of /kawâl/ ‘to make; to do’ — 241
- 8.65 Dzadrani present aorist of /kawâl/ ‘to make; to do’ — 242
- 8.66 GP past continuous of كول /kawâl/ ‘to make; to do’ — 242
- 8.67 Waziri past continuous of /kawâl/ ‘to make; to do’ — 243
- 8.68 Dzadrani past continuous of /kawâl/ ‘to make; to do’ — 243
- 8.69 GP past aorist of كول /kawâl/ ‘to make; to do’ — 244
- 8.70 Waziri past aorist of /kawâl/ ‘to make; to do’ — 245
- 8.71 Dzadrani past aorist of /kawâl/ ‘to make; to do’ — 245
- 8.72 Present participles — 246
- 8.73 Present participle: declension — 247
- 8.74 Past participles — 248
- 8.75 Past participle: declension — 248
- 8.76 Past participles built on aorist bases — 249
- 8.77 Present continuous, first conjugation (intransitive) — 250
- 8.78 Present continuous, first conjugation (transitive) — 250
- 8.79 Present continuous, second conjugation — 251
- 8.80 Present continuous, third conjugation — 251
- 8.81 Present aorist, first conjugation (intransitive) — 252
- 8.82 Present aorist, first conjugation (transitive) — 252
- 8.83 Present aorist, second conjugation — 253

- 8.84 Present aorist, third conjugation — 253
- 8.85 Past continuous, first conjugation (intransitive) — 254
- 8.86 Past continuous, first conjugation (transitive) — 255
- 8.87 Past continuous, second conjugation — 255
- 8.88 Past continuous, third conjugation — 256
- 8.89 Past aorist, first conjugation (intransitive) — 256
- 8.90 Past aorist, first conjugation (transitive) — 257
- 8.91 Past aorist, second conjugation — 257
- 8.92 Past aorist, third conjugation — 258
- 8.93 Continuous imperative, first conjugation — 259
- 8.94 Continuous imperative, first conjugation (negative) — 259
- 8.95 Continuous imperative, second conjugation — 259
- 8.96 Continuous imperative, second conjugation (negative) — 259
- 8.97 Continuous imperative, third conjugation — 260
- 8.98 Continuous imperative, third conjugation, negative — 260
- 8.99 Aorist imperative, first conjugation — 260
- 8.100 Aorist imperative, second conjugation — 261
- 8.101 Aorist imperative, third conjugation — 261
- 8.102 Continuous optative forms — 262
- 8.103 Aorist optative forms — 264
- 8.104 Present perfect, first and second conjugations — 265
- 8.105 Present perfect, third conjugation — 266
- 8.106 Past perfect — 267
- 8.107 Present potential — 268
- 8.108 Past potential — 268
- 9.1 Some GP simple circumpositions — 331
- 9.2 Some Middle Dialect circumpositions in contrast with GP — 333
- 10.1 Some adverbs of time — 379
- 10.2 Some adverbs of place — 381
- 10.3 Some adverbs of manner — 384
- 10.4 Some adverbs of degree — 386
- 10.5 Arabic adverbs in Pashto — 388
- 10.6 Some other interrogative words — 389
- 10.7 Some doublets and their base stems — 397
- 11.1 Element ordering in negative future constructions — 407
- 11.2 Negative placement — 409
- 11.3 GP additional interrogative adverbs — 418
- 11.4 Case-marking pattern for nouns — 449
- 11.5 Case-marking pattern for human interrogative pronouns — 449
- 11.6 Case-marking pattern for strong pronouns — 449

1 About this Grammar

1.1 Scope of this book

This grammar covers the four standard dialects and the Middle dialects—including Waziri—of modern Pashto, with greatest emphasis on morphology. Morphological and some lexical features specific to the Middle dialects are given their own sections, corresponding to their counterparts in **General Pashto** (the set of dialects that exist in contrast to the Middle dialects, as described in Chapter 4). Where Middle dialects (abbreviated as “M”) and General Pashto (“GP”) are not known to differ—for example, in their syntax—only one description is given. If no Middle-specific form of a particular feature is described, we presume that it conforms to General Pashto. Chapter 4 contains a more complete description of the differences among Pashto dialects.

In order to describe the wide range of Pashto dialects with precision, we consulted a variety of sources, including previously published grammars, publicly available data on the internet, and consultations with native speakers. Pashto, as a living language spoken in a politically volatile region, is constantly evolving, and our data collection strategy enabled us to include examples from a wide swath of dialectal situations.

Native speakers from several dialect areas were consulted on issues ranging from basic pronunciation to the appropriate use of complex syntactic constructions. After providing detailed personal language background information, each speaker was recorded pronouncing single word examples from prescribed lists. Over the course of several sessions, the speaking tasks increased in complexity. Speakers were asked to provide complete paradigms and examples of usage. We were able to consult speakers about phenomena that were insufficiently or inconsistently described in the literature, asking for grammatical judgments about specific examples and more open-ended commentary on general issues.

1.2 Orthography

Except for the fact that written Pashto always uses the Perso-Arabic script, Pashto writing varies significantly according to a number of factors such as region, influence of other languages, and so on, and many words may have multiple widely-accepted spellings. Additionally, apparent word boundaries can vary as well. For instance, some writers orthographically treat forms we assert to be free forms as if they are bound forms (e.g. by joining a preposition with the word it governs), or, conversely, treat forms we believe to be bound forms as if they are free (e.g. by separating an aorist prefix from its stem).

We present all authentic written examples as we found them, with their original spellings and word boundaries; however, we represent words as consistently as possible from one transcription and morpheme-by-morpheme gloss to the next, which may

result in examples in which a single word in the Pashto script is represented as multiple words in the gloss, or in which multiple words in the Pashto script are represented as a single word in the gloss. Where the authentic text represents variance from what we believe to be standard conventions of spelling or word boundaries, a standardized version of the Pashto script is presented in an accompanying footnote.

Authentic spoken examples are represented in the dialect of the speaker and this dialect information is indicated in subscript, when known. A complete table of the transcription schema used in this book is given in Table 3.11.

Where we have cited examples from other scholarly works, we have retained as much information as the original example provides. (If such an example does not contain script, we have not added it, unless we were certain of their transcription system.) In some instances, we have adapted the Romanization system or the morpheme glossing used by the author in order to elucidate the point at hand.

1.3 Tables and examples

Table titles are marked, where relevant, with information about the dialect(s) concerned. Where there is no dialect information in the table title, the contents of that table are presumed to hold for all dialects (General Pashto and Middle dialects).

To more easily represent widespread syncretism in Pashto grammatical forms, the tables in this grammar use an empty cell to represent a form which is identical to the form above it (or, if there is no form above it, or if the form above it is separated with a horizontal line, identical to the form to its left). Forms whose existence is uncertain are represented with a question mark; this is more likely in Middle dialects than in General Pashto. The complete absence of a form is denoted by a dash in the cell. Where different dialect forms exist, these are shown vertically, with the initial of the dialect name in subscript after the form. Where no alternate forms are given, the form in the table should be taken as applying to all dialects.

Table 1.1 and Table 1.2 are examples showing these table layouts.

In Table 1.1,

- The masculine singular oblique form is the same as the masculine singular direct;
- The masculine singular vocative form is the same as the masculine singular ablative;
- The masculine plural ablative and vocative, the feminine singular ablative and vocative, and the feminine plural ablative and vocative forms are the same as the masculine plural oblique, the feminine singular oblique, and the feminine plural oblique, respectively; and
- The masculine plural direct form is the same as the masculine singular direct.

In Table 1.2,

- The third person feminine singular form is the same as the third person masculine singular form, and
- Both the third person plural forms are the same as the third person singular form.

Some of these forms are spelled the same but pronounced differently from one dialect to the next, while other forms have different spellings as well as different pronunciations.

ə~o~e	Masculine		Feminine	
	Singular	Plural	Singular	Plural
Direct	سپک spak ^E spək ^W		سپکه spāk-a ^E spāk-a ^W	سپکې spāk-e ^E سپکي spāk-i ^W
Oblique		سپکو spāk-o ^E	سپکې spāk-e ^E	سپکو spāk-o ^E
Ablative	سپکه spāk-a ^E spāk-a ^W	spāk-o ^W spók-o ^W	سپکي spāk-i ^W	spāk-o ^W spók-o ^W
Vocative				

Table 1.1: Class I, stem alternation: سپک /spək/ ‘light’

Examples of Pashto words appear occasionally in text, with the Pashto script followed by the transcription in phonemic slashes and the gloss in single quotation marks: سپک /spək/ ‘light’.

Examples of phrases and complete sentences appear in numbered four-line interlinear examples, with the Pashto script in the first line, the transcription in the second, the morpheme-by-morpheme gloss in the third, and the translation of the sentence in the last line. 1.1 is a sample of an interlinear example.

- (1.1) *ښوونځی جوړونه، انسان جوړونه ده!*
xuwəndz-ay joɾawəna-∅ insān-∅ joɾawəna-∅
 school-M.DIR building-F.DIR people-M.DIR building-F.DIR
da
 be.CONT.PRS.3SG.F
 ‘Building schools means growing our youth!’

کول kawól 'to make, to do'		Singular	Plural
1st		کوم kawám	کوو kawú
2nd		کوي kawé	کوي kawáy
3rd	M	کوي kawí	
	F		

Table 1.2: Present continuous of کول /kawól/ 'to make, to do'

When an interlinear example is used to illustrate a particular word or morphological concept, the term appears in bold type. The transcription and morpheme gloss are always bolded in such instances, and the free translation may be bolded if the English words clearly correspond to the Pashto. For ease of reading, the Pashto script is never bolded.

1.4 Abbreviations and symbols

Where possible, morpheme glosses in this grammar follow the Leipzig Glossing Rules,¹ a set of formatting conventions widely adopted in the linguistics community.

Commonly used abbreviations in this grammar include the following:

*: non-existent or unacceptable form

~: variation in forms (within or across dialects)

[]: non-overt element

∅: zero morpheme

1: first person pronominal clitic

1DVC: first person directional verbal clitic

2: second person pronominal clitic

2DVC: second person directional verbal clitic

3: third person pronominal clitic

¹ <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>

3DVC: third person directional verbal clitic
 ABL: ablative
 ADJ: adjective
 ADJZ: adjectivizer
 ANIM: animate
 AOR: aorist
 C: consonant
 COMMIT: comitative
 COMP: complementizer
 CONT: continuous
 DIR: direct
 DZA: Dzadrani (dialect)
 E: Eastern (dialects)
 ECHO: echo word
 EMPH: emphatic
 EXT: existential particle
 F: feminine
 GP: General Pashto
 INF: infinitive
 IMP: imperative
 INSTR: instrumental
 LVC: light verb construction
 M: (as diacritic) Middle (dialect)
 M: (in glosses) masculine
 N: noun
 NE: Northeastern (dialect)
 NEC: modal of necessity or obligation
 NEG: negative particle
 NMLZ: nominalizer
 NW: Northwestern (dialect)
 OBL: oblique
 OPT : optative
 PL: plural
 PNG: person-number-gender
 POSS: possessive
 PRS: non-past
 PTCP: participle
 PST: past
 Q: question particle
 SE: Southeastern (dialect)
 SG: singular
 STR: strong

SW: Southwestern (dialect)

V: vowel

VOC: vocative

W: Western (dialects)

WAZ: Waziri (dialect)

WAZ-L: Waziri (dialect); example comes from Lorimer

WK: weak

WOULD: modal of irrealis or habitual events

Anne Boyle David

2 The Pashto Language

2.1 Background

Pashto is considered to be the second-most important Iranian language, after Persian, given its widespread use in Afghanistan and Pakistan, its official status in Afghanistan, and its long literary tradition dating back to or before the 16th century (MacKenzie 1992; Skjærvø 1989). This section provides an overview of the Pashto language, including demographic information, linguistic classification, dialectal variation, and available linguistic resources.

2.2 Population of speakers

Pashto is an Indo-Iranian language of the East Iranian branch, spoken primarily in Afghanistan and Pakistan. In Afghanistan it is predominantly spoken in the eastern (Nangarhar, Laghman, Kunar), central (Kabul, Logar, Wardak), southeastern (Ghazni, Khost, Paktiya, Paktika), southwestern (Kandahar, Helmand, Uruzgan, Zabul), and western (Herat, Farah) regions. In Pakistan it is primarily spoken in the Khyber Pakhtunkhwa (formerly North-West Frontier Province); the Federally Administered Tribal Areas (including Waziristan); and in northeastern Balochistan, including the city of Quetta. Communities of Pashto-speaking migrant laborers from Afghanistan and Pakistan are reportedly found in Iran and the United Arab Emirates, and small populations of Pashto speakers have been documented in Tajikistan (Lewis, 2009).

Babrackzai (1999) notes that most estimates of the number of Pashto speakers in available linguistic studies and encyclopedic descriptions are quite rough. Until recently, figures have tended to be outdated or otherwise underestimated, ranging from 13 million (a figure widely quoted from Penzl 1955; see e.g. Lockwood 1972; Lorenz 1982; MacKenzie 1987; MacKenzie 1992) to above 20 million (Inozemtsev 2001; Tegey & Robson 1996). Austin (2008) sets the range at 30–50 million speakers; this appears to be a fairly sound figure when compared to the estimate of 41.3 million speakers that can be derived from the World Factbook's population projections and ethnic percentage figures: 42% of Afghanistan's 33.6 million inhabitants and 15.42% of Pakistan's total population of 176 million (Directorate of Intelligence 2009a; Directorate of Intelligence 2009b).¹ At the upper end of this range, Ethnologue tentatively reports 49.5

¹ Though itself dominated by neighboring languages such as Dari and Urdu in some spheres, Pashto is a relatively dominant language, so it is reasonable to expect that Pashtun ethnicity corresponds roughly with Pashto language acquisition. Furthermore, although there may be some attrition through marriage, immersion in non-Pashto speaking urban settings, etc., the numbers are probably more than offset by speakers of other languages who adopt Pashto as a second language.

million speakers of Pashto in all countries (Lewis, 2009). On the lower end, a posthumously updated version of MacKenzie's description in *The World's Major Languages* places the number at about 25 million (MacKenzie 2009, cf. MacKenzie 1987).

2.2.1 History and classification

Pashto belongs to the East Iranian branch of the Indo-Iranian language family, meaning it is closely related to other Iranian languages such as the Pamir languages (also in the East Iranian branch) and Persian, Kurdish, and Balochi (in the West Iranian branch). It is somewhat less closely related to the neighboring Nuristani languages and to Indo-Aryan languages such as Urdu, Punjabi, and Sindhi; and more distantly related to other Indo-European languages. Beyond the classification of Pashto as an East Iranian language, there are competing views and descriptions regarding the exact subclassification of Pashto and the degree of affinity between Pashto and certain related languages, especiallyOrmuri and Parachi.

2.2.2 Dialectal variation

Pashto can be divided into numerous dialects, or varieties, as described in more detail in Chapter 4. Dialectal differences in Pashto are primarily phonological in nature. The Southwest (Kandahar) dialect makes use of retroflex fricatives /ʂ/ and /ʐ/ and of the dental affricates /ts/ and /dʒ/. In other dialects, the retroflex fricatives, and sometimes also the dental affricates, are replaced with other phonemes. For example, Southwest /ʂ/ is elsewhere pronounced as /š/, /ç/, or /x/, depending on dialect. One result of this dialectal variation is the use of both *Pakhto* and *Pashto* as Romanized spellings of the language's name. Another result is the assignment of several alternative values to the pronunciation of the relevant Pashto letters, shown in Table 3.11 in Section 3.2.1.2. Dialectal variation in morphology has also been documented, especially for the dialects classified as Middle (M) in Section 4.4, which differ more from other dialects of Pashto than the other dialects do among themselves (Hallberg 1992, Lorimer 1902).

Attrition is more likely in Pakistan, where Pashto is a less dominant language, in which case the World Factbook's estimated 8% Pashto language use would be a more conservative indicator than the 15.42% ethnic population.

Corey Miller and Anne Boyle David

3 Phonology and Orthography

3.1 Phonetics and phonology

3.1.1 Consonants

3.1.1.1 Inventory

Pashto has consonants at seven possible places of articulation, as shown in Table 3.1. Consonants not used by all speakers are parenthesized. The representation of these segments in the written language is discussed in Section 3.2.

	Labial	Dental	Palato- alveolar	Retroflex	Palatal	Velar	Glottal
Stops	p b	t d		ʈ ɖ		k ɡ	
Affricates		(ts ɖʒ)	č j				
Fricatives	(f)	s z	š (ž)	(ʂ ʐ)	(ç ʝ)	x ɣ	h
Nasals	m	n		ŋ			
Lateral		l					
Approximants	w	r	y	ɻ			

Table 3.1: Pashto consonants

3.1.1.2 “Elegant” consonants

Depending on how closely speakers wish to approximate Arabic sounds not otherwise present in Pashto, they may use what are sometimes called “elegant” consonants. That is, when speaking in what they regard as a formal situation, many educated speakers will approximate the Arabic pronunciations of those consonants; these are listed in Table 3.2.

Penzl (1955) did not seem to think that the voiced pharyngeal fricative was ever pronounced in Pashto; however, Tegey & Robson (1996) reported 41 years later that it does occur in some speakers’ formal speech. It is possible that this discrepancy is due

Arabic consonant	Educated, formal Pashto	Uneducated or informal Pashto
ء <i>glottal stop</i>	ʔ	(not pronounced)
ع <i>voiced pharyngeal fricative</i>	ʕ or ʔ	(not pronounced)
ح <i>voiceless pharyngeal fricative</i>	ħ	h (or not pronounced)
ق <i>voiceless uvular stop</i>	q	k

Table 3.2: “Elegant,” or formal, consonants

to an ongoing linguistic change under Arabic influence or to a difference between the dialects being reported on. Elfenbein (1997: 742) notes that /h/ is often dropped and is only retained in some dialects as an “elegance”.

3.1.2 Vowels

3.1.2.1 Inventory

Most dialects of Pashto have seven vowels and seven diphthongs (Heston, 1992), as shown in Table 3.3 and Table 3.4.

	Front	Central	Back
High	i		u
Mid	e	ə	o
Low	a		ā

Table 3.3: Pashto vowels

	Front	Central	Back
High			uy
Mid		əy	oy
Low	ay āy aw āw		

Table 3.4: Pashto diphthongs

3.1.2.2 Vowel transcription

For the sake of comparison with other sources, we provide the corresponding transcriptions of the vowel sounds in Table 3.5. Note that Tegey and Robson's distinction between both /i/ and /ɪ/ (row 1) and /u/ and /ʊ/ (row 2) may correlate with the presence or absence of **ي** and **و** respectively in the spelling of the word (see Section 3.1.2.3). Penzl, who describes the Kandahari dialect, makes a distinction between /ey/ for **ئ** and /ei/ for **ئ**, and does not include /oy/ in his inventory.

This grammar	Example	IPA	Penzl	Tegey and Robson
i	امتحان imtihān 'test'	i	i	i, ɪ
u	اردو urdu 'army'	u	u	u, ʊ
e	یہ ye 'you are'	e	ee	e
ə	زہ zə 'I'	ə	e	ə
o	اور or, wor 'fire'	o	oo	o
a	لس las 'ten'	a	a	a
ā	کال kāl 'year'	ɑ	aa	ā
uy	دوی duy 'they'	uj	uy	uy
əy	لکی lakəy 'tail'	əj	ey, ei	əy
oy	زوی zoy 'son'	oj	uy	oy
ay	سری saṛay 'man'	aj	ay	ay
āy	ڇای ɟāy 'place'	ɑj	aay	āy
aw	یو yaw 'one'	aw	aw	aw
āw	واورہ wāwra 'snow'	ɑw	aaw	āw

Table 3.5: Comparison among vowel transcription systems

3.1.2.3 “Elegant” vowels

Persian or Arabic loanwords that contain an etymologically long /ī/ or /ū/ (spelled with *ی* *mārufa ye* and *و* *wāw*, respectively) are sometimes pronounced with long vowels by educated speakers, or speakers of the Northeastern dialect (Section 4.3). Like some consonants in Arabic loanwords, these may be called “elegant” vowels. Due to variation in pronunciation, the short vowels /i/ and /u/ are now sometimes represented in Pashto writing with *ی* and *و*, despite the usual convention of not representing short vowels in Arabic scripts.

3.1.2.4 Middle dialect vowels

The vowel inventory for the Middle dialects described in Chapter 4 differs slightly from that described in Section 3.1.2.1 and Section 3.1.2.2. One issue is the phenomenon described as “Waziri Metaphony” in Section 4.4. In Table 3.6, we compare three treatments of the M vowel system, by Septfonds (1994), Lorimer (1902), and Hallberg (1992). While the objects of study in each of those works are slightly different, they all fall under the M rubric as described in Chapter 4. In order to compare these treatments to each other, we provide example words representing each class of sounds. For these, we have provided Pashto script reflecting the General Pashto pronunciation; however, to our knowledge, there exists no systematic representation of M pronunciation using Pashto script.

There does not seem to be a reliable contrast in M between /o/ and /u/. Lorimer and Hallberg distinguish between /a/ and /ā/, whereas Septfonds does not. Septfonds distinguishes between /əy/ and /ay/, whereas Lorimer does not. As can be seen in Table 3.6, Lorimer makes some additional distinctions compared to the others. For example, Lorimer uses è as a variant of his é before /r/. Lorimer’s use of ü (presumably IPA /y/) as a variant of /i/, and ö (presumably IPA /ø/) as a variant of /e/ may reveal that there existed “midway” points between the GP vowels and their metaphony targets (see Section 4.4). We have provided two examples of the /e/ vowel: one that shares an /e/ with other dialects of Pashto, and another that represents the metaphony target corresponding with GP /o/. Lorimer also employs a sequence of two symbols, aṅ, to indicate /ã/.

This grammar	Example	IPA	Septfonds	Lorimer	Hallberg
i	لور lir 'daughter'	i	i	i, ī, ü	i
e	دېر dər 'many', مور mer 'mother'	e	e	é, è, ö	e
ə	زە zə 'I'	ə	ə	e, ê, u	ʌ
o	پلار plor 'father'	o	o	o	o
a	سر sar 'head'	a	a	a	ʌ
ā	مياشت myāšt 'month'	ɑ	a	ā	ɑ
ā	وتېم wətapā 'I close'	ā		aŋ	
əy	لمسى lmasəy 'granddaughter'	əj	əy	ai	
ay	سړى saray 'man'	aj	ay	ai	ai, ʌi
aw	پلو palaw 'side'	aw	aw	au	

Table 3.6: Comparison among M vowel transcription systems

3.1.3 Stress

Word stress in Pashto is assigned lexically. In our transcription of Pashto, we indicate stressed syllables by putting an acute accent mark over the vowel portion: سړی /saráy/ ‘man’, ژبه /žǎba/ ‘tongue’. Some words differ only by the placement of stress, as with the pronouns هغه /haḡá/ ‘he/she/it’ vs. هغه /háḡə/ ‘that’, or continuous versus aorist forms among second and third conjugation verbs (Section 8.2.5.4 and Section 8.2.5.5). In this grammar we generally only indicate stress in paradigms and on particular lexical items in example sentences where its placement sheds light on meaning, or when relevant to the discussion .

3.2 Orthography

The languages of the world vary along a scale of the ease with which one may deduce pronunciation from orthography, or the spelling from the pronunciation. This is known as *orthographic depth* (Sproat, 2000: 6). On a spectrum of depth, Spanish may be considered fairly shallow, or easy to pronounce, and Chinese would be considered especially deep. Pashto would fall somewhere near English on such a scale, though the difficulties of deducing pronunciation from spelling (and vice versa) are somewhat different, as will be described below.¹

3.2.1 The script

Pashto is written primarily in the Perso-Arabic script, which includes modifications made for Persian, as well as additions specific to Pashto. Like those scripts, and unlike, for example, the Devanagari script in which Hindi is written, the Pashto script is in theory an *abjad* (Daniels & Bright, 1996), in which the letters represent only the consonants of the language. However, like the Arabic and Persian scripts, the Pashto alphabet deviates from the strict definition of an abjad in that some letters can also be used to indicate vowel sounds. For example, the letter *alif*, which is used to write both long /ā/ and an initial glottal stop in Arabic, is also used in Pashto to represent the vowel /ā/. The letter *wāw*, which writes both /w/ and /ū/ in Arabic, can be used to represent /w/, /u/, or /o/ in Pashto; and *ye*, which writes both /y/ and /ī/ in Arabic, can be used in Pashto for /y/, /i/, and even (with modification to the letter), /e/ and several other sounds, which can be seen in Table 3.10.

As with other Arabic-based writing systems, the Pashto script is written from right to left, but numerals are written left to right. The script is cursive, and letters may have

¹ More on issues that arise from a deep orthography, in particular variations in spelling, can be found at Appendix A, Section A.4.

up to four allographs: the independent form, which is unconnected to other letters; the initial form, connected only on the left; the medial form, connected on both sides; and the final form, connected only on the right. (In most word processors, Unicode rendering automatically displays the proper positional form.) Some letters do not connect to the letter to their left; in these cases the initial and independent forms are the same, as are the medial and final forms. (See the tables of letters in Section 3.2.1.2.)

3.2.1.1 Letters unique to Pashto

Due to its particular consonant and vowel inventory, Pashto has innovated several letters not present in either Arabic or Persian. To represent the retroflexes /ɖ ɳ ʀ ʈ/, Pashto employs a diacritic known as a پنډک /paṇḍak/ or غړونډی /ɣaɾwandaɣ/ ‘ring’ on the letters used for the dentals, as shown in Table 3.7.

Letter	Sound
ډ	ɖ
ڳ	ɳ
ڀ	ʀ
ټ	ʈ

Table 3.7: Pashto retroflex letters with paṇḍak

Pashto has innovated two letters employing a dot above and below ږ and ښ. These letters represent retroflex sounds in the Kandahar (Southwest) dialect, but other sounds elsewhere, as displayed in Table 3.8. Further pronunciations of these letters, particularly in the Northwest dialects, are discussed in Chapter 4.

The two letters in Table 3.9 are based on ځ and ښ and represent the affricates /tʃ/ and /dʒ/ in some dialects, but in others they have been simplified to the fricatives /s/ and /z/, as discussed in Chapter 4.

Finally, Pashto employs a set of five letters based on ښ, which have a range of phonetic and morphological values, as expressed in Table 3.10. Note that in much text encountered on the internet and elsewhere, ښ or ښ may be used in place of the other forms, without affecting the expected pronunciation.

Letter	Sound
ږ	ʒ Kandahar (Southwest), ž Quetta (Southeast), g Peshawar (Northeast)
ښ	ʃ Kandahar (Southwest), š Quetta (Southeast), x Eastern

Table 3.8: Pashto letters with dot below and dot above

Letter	Sound
څ	ts (Southwest, Southeast), s (Northwest, Northeast)
ځ	ɟ (Southwest, Southeast), z (Northwest, Northeast)

Table 3.9: Pashto affricate letters based on ح

Letter	Sound	Function/comments
ی	ay	word-finally
	ā	word-finally in some words derived from Arabic, such as حتی hatā ‘even’. This is known as <i>alif maqsura</i> .
ي	i, y	
ې	e	
ی	əy	word-finally in feminine nouns and adjectives
ئ	əy	word-finally in verbal forms
	not pronounced or ?	word-medially in some Perso-Arabic borrowings such as مسئله masala ‘problem’

Table 3.10: Pashto letters based on ی

3.2.1.2 Tables of letters and numerals

The full list of Pashto letters is given in Table 3.11. Sources differ on some of the names for Pashto letters, or give more than one. These differences are due in part to regional variation in how the letters are pronounced and in part to variation in how the Pashto has been Romanized. To aid in letter identification, we have listed a variety of possible Romanized names. Where practical, we have indicated the most common pronunciations of the letters; where there is further variation we refer to the relevant sections. The pronunciations listed in parentheses are the formal pronunciations described in Section 3.1.1.2. Note that some of the symbols listed here (e.g. $\bar{\text{آ}}$ *alif mad* and ۀ *he-hamza*) are not considered separate letters from their counterparts without diacritics (e.g. آ *alif* and ه *he*); we list them separately here for ease of reference. Letter names that end in a vowel or diphthong are grammatically feminine, while letter names that end in a consonant are grammatically masculine.

Following are some remarks on particular letters. $\bar{\text{آ}}$ *aleftanwin* appears word-finally in adverbs derived from Arabic, e.g. تقريباً /taqribā/ ‘approximately’. The letter و *wāw* is also transcribed as /v/ in some sources, such as Shafeev (1964). The letter ؤ *wāw-hamza* appears word-medially in Perso-Arabic borrowings such as سؤال /swāl, suwāl/ ‘question’. When appearing word-finally, the letter ه *he* may represent a vowel such as /a/ or /ə/. The letter ی *ye* is also commonly used when referring to the letter *ye* in general, as in a section of a dictionary. It thus can be used to encompass ی *mārufa ye*. In this usage, it would be called simply *ye*. The name *mārufa ye* for the letter ی means ‘known ye’, reflecting the fact that its sound, /i/, was known in Arabic. The name *majhula ye* for the letter ی means ‘unknown ye’, reflecting the fact that its sound, /e/, was unknown in Arabic. The names for the letter ی include the words for ‘feminine’ (*ṣaḍāina tānis, muanasa*) and ‘heavy’ (*saqila*). The names for the letter ئ include *kəṛwāla* and *feli*, meaning ‘verbal’, *tazkir* meaning ‘masculine’, and *saqila* meaning ‘heavy’.

Some of the letters are used chiefly in Arabic loanwords, or in loanwords based on Arabic. The ح *halwa he*, for example, appears in Arabic loanwords which are spelled with this letter in Arabic, and in borrowings from Persian in which one of the morphemes is Arabic: پرحرارَت /purharārat/ ‘ardent, emotional’, which is composed of a Persian prefix پُر /pur/ ‘full’ and an Arabic root حرارَت /harārat/ ‘heat’. The other /h/ sound, ه *he*, is used in native Pashto words, as well as in Arabic and Persian loanwords which are spelled with this letter. Representing /s/, the letters ث *se* and ص *sād* are found in the Arabic/Persian component of the vocabulary, while س *sin* is found in the native component as well. Representing /t/, the letter ط *tā* is found in the Arabic/Persian component of the vocabulary, while ت *te* is also found in the native component. Finally, representing /z/, the letters ذ *zāl*, ض *zād*, and ظ *zā* are found in the Arabic/Persian component of the vocabulary, while ج *ze* is found in the native component as well.

Table 3.11: Pashto alphabet

Unicode	Pashto script	Name	IPA	Penzl	Tegey and Robson	This grammar
U+0627	ا	alif	see Section 3.2.1.3			
U+0622	آ	alif mad	ɑ	aa	ā	ā
U+0627 U+064B	أ	alif tanwin	an	an	an	an
U+0628	ب	be	b	b	b	b
U+067E	پ	pe	p	p	p	p
U+062A	ت	te	t̪	t	t	t
U+067C	ټ	ṭe	t̪	tt	ṭ	ṭ
U+062B	ث	se	s	s	s	s
U+062C	ج	jim	dʒ	dʒh	j	j
U+0686	چ	če/čim	tʃ	tsh	ch	č
U+0681	غ	dʒe/dʒim/ze	dʒ/z	dʒ	dʒ	dʒ/z
U+0685	ځ	ts/s tse/tsim/se	ts/s	ts	ts	ts/s
U+062D	ح	he/halwa he	∅/h/(h̄)	∅/h/(h̄)	∅/h/(h̄)	∅/h/(h̄)
U+062E	خ	xe	x	kh	kh	x
U+062F	د	dāl	d̪	d	d	d
U+0689	ډ	dāl	d̪	dd	d̪	d̪
U+0630	ذ	zāl	z	z̪	z	z
U+0631	ر	re	r	r	r	r
U+0693	ړ	ṛe	ɽ	rr	ṛ	ṛ
U+0632	ز	ze	z	z	z	z
U+0698	ژ	že	ʒ/dʒ/z	zh	zh/j/z	ž/j/z
U+0696	ږ	že/ge	ʒ/ʒ̪/g	zz	g	ʒ/ž̪/g
U+0633	س	sin	s	s	s	s
U+0634	ش	šin	ʃ	sh	sh	š

Table 3.11: (continued)

Unicode	Pashto script	Name	IPA	Penzl	Tegey and Robson	This grammar
U+069A	ځ	ṣin/xin	ʃ/x	ss	x	ʃ/x
U+0635	ص	sād/swād/ sawāt/swā	s	ṣ	s	s
U+0636	ض	zād/zwād/ zawāt/zwā	z	ṣ	z	z
U+0637	ط	tā/twe/ txwe/ toy/twā	ṭ	ṭ	t	t
U+0638	ظ	zā/zwe/ zywe/ zoy/zwā	z	z̄	z	z
U+0639	ع	ʔayn/ayn	(ʔ)/(ʕ)	ʔ	ʔ	see Section 3.2.1.3 and Section 3.1.1.2
U+063A	غ	ɣayn	ɣ	gh	gh	ɣ
U+0641	ف	fe	f	f	f	f
U+0642	ق	qāf/qaf	k/(q)	q	q	k/(q)
U+06A9	ک	kāf/kaf	k	k	k	k
U+06AB	گ	gāf	g	g	g	g
U+0644	ل	lām	l	l	l	l
U+0645	م	mīm	m	m	m	m
U+0646	ن	nun	n	n	n	n
U+06BC	ڼ	ṇun	ɳ	nn	ṇ	ṇ
U+0648	و	wāw	w/o/u	w	w/u/o	w/u/o
U+0624	ؤ	wāw- hamza	∅/w			∅/w
U+0647	ه	he/hā	h	h	h	h

Table 3.11: (continued)

Unicode	Pashto script	Name	IPA	Penzl	Tegey and Robson	This grammar
U+06C0	ۀ	he(/hā)- hamza/ zwarakai	ə	e	ə	ə (see Section 3.2.3)
U+06CC	ی	ye/yā/ mulayana ye (/yā)/prata ye (/yā)	aj	ay	ay	ay (see Table 3.10)
U+064A	ي	saxtə ye (/yā) /klaka ye (/yā) /mārufa ye(/yā)	i/y	i/y	i/y	i/y (see Section 3.2.1.3)
U+06D0	ې	pasta ye (/yā) /majhula ye (/yā)	e	ee	e	e
U+06CD	ی	şəðzina ye (/yā) /saqila/ de tānis saqila ye (/yā) /mua- nasa saqila ye (/yā)	əj	ei	əy	əy
U+0626	ئ	kərwāla ye (/yā) /ye (/yā)- hamza/ feli ye (/yā) /de tazkir saqila ye (/yā/)	əj	ey	əy	əy

Positional variants for each letter are shown in Table 3.12. In keeping with the directionality of the script, the initial-position form is on the right. In addition to the positional variants of individual letters, the script includes an obligatory digraph, لآ , used for the sequence of *lām + alif*.

The Eastern Arabic numerals, which are used in Pashto, are given in Table 3.13. Numeral forms of compound numbers are given along with the word-form number names in Table 6.38.

Table 3.12: Positional variants of letters

Independent form	Final position	Medial position	Initial position
ا	ا	ا	ا
ب	ب	ب	ب
پ	پ	پ	پ
ت	ت	ت	ت
ټ	ټ	ټ	ټ
ث	ث	ث	ث
ج	ج	ج	ج
چ	چ	چ	چ
خ	خ	خ	خ
څ	څ	څ	څ
ح	ح	ح	ح
خ	خ	خ	خ
د	د	د	د
ډ	ډ	ډ	ډ
ذ	ذ	ذ	ذ
ر	ر	ر	ر
ړ	ړ	ړ	ړ

Table 3.12: (continued)

Independent form	Final position	Medial position	Initial position
ز	ز	ز	ز
ژ	ژ	ژ	ژ
ب	ب	ب	ب
س	س	س	س
ش	ش	ش	ش
بش	بش	بش	بش
ص	ص	ص	ص
ض	ض	ض	ض
ط	ط	ط	ط
ظ	ظ	ظ	ظ
ع	ع	ع	ع
غ	غ	غ	غ
ف	ف	ف	ف
ق	ق	ق	ق
ک	ک	ک	ک
گ	گ	گ	گ
ل	ل	ل	ل
م	م	م	م
ن	ن	ن	ن
و	و	و	و
ه	ه	ه or ه	ه
ی	ی		