TOWARDS PROTO-PERSIAN: AN OPTIMALITY THEORETIC HISTORICAL RECONSTRUCTION

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By

Daniel A. Rees, M.S.

Washington, DC
August 13, 2008

Copyright 2008 by Daniel A. Rees
All rights reserved

TOWARDS PROTO-PERSIAN: AN OPTIMALITY THEORETIC HISTORICAL RECONSTRUCTION

Daniel A. Rees, M.S.

Thesis Advisor: Shaligram Shukla, Ph.D.

ABSTRACT

The intent of the present course of inquiry is to examine historical Persian phonology in light of advances in Optimality Theory (OT) as applied to language change and variation, and in doing so to present a historical reconstruction of Proto-Modern-Persian (henceforth Proto-Persian). The phonemic inventories and synchronic phonological processes observed in the modern dialects of Persian bear witness to a number of changes accounted for within the framework of OT. Taking into account previous literature on pre-Modern Persian diachronic change, colloquial speech from various speakers of the three principal Modern Persian dialects of Iranian[1], Dari and Tajik, is analyzed and a comparative reconstruction is generated, positing proto-forms as predecessors to attested phonemes based on patterns of predictability in the data. OT unifies the phonological account of both variation and change by tying all processes to the physiological notions of ease of articulation and parsing, bolstering the claim that sound change at the lexical level is based on gesture economy at the articulatory level. As a direct result of this reconstruction, it is quantifiably demonstrated that the Afghan dialect of Persian, Dari, is not only the most conservative of the three major dialects, but it is merely a slightly modified version of what will be labeled Proto-Persian.

[1] Please see I.A, "A Note on Terminology," concerning use of the terms "Persian," "Iranian," "Farsi," etc.

TABLE OF CONTENTS

I. Introduction..... 1

I.A A Note on Terminology..... 1

I.B Review of Existing Literature..... 2

I.C Background on Historical Indo-Iranian.....6

I.D Modern Persian Dialectology.....9

I.E Introduction to Optimality Theory..... 11

I.F Methods of Data Collection..... 15

I.G IPA and Abbreviations Used.....18

I.G.1 IPA: Vowels..... 19

I.G.2 IPA: Consonants.....20

I.G.3 Abbreviations..... 21

I.H Constraints Used.....21

I.H.1 Maximization and Faithfulness Constraints..... 21

I.H.2 ID/DEP and Faithfulness Constraints.....22

I.H.3 Markedness Constraints..... 22

I.H.4 Miscellaneous Well-formedness Constraints..... 23

II. From Old Persian to Pahlavi to Proto-Persian.....26

II.A Old Persian.....26

II.A.1 Old Persian Vowels.....27

II.A.2 Old Persian Consonants..... 28

II.B Old Persian to Middle Persian.....30

II.B.1 Old Persian to Middle Persian: Vowels.....30

II.B.1.a Low Vowels /a/ and /ā/.....31

II.B.1.b High Front Vowels /i/ and /ī/..... 31

II.B.1.c High Back Vowels /u/ and /ū/..... 32

II.B.1.d Diphthongs /ai/ and /au/ 32

II.B.1.e */ṛ/ 33

II.B.2 Old Persian to Middle Persian: Consonants 35

II.B.2.a /p/, /t/, /ʧ/, /k/ 35

II.B.2.b /b/, /d/ (/ð/), /ʤ/, /g/ (/ɣ/) 36

II.B.2.c /f/, /θ/ (/ç/[2]), /x/ 37

II.B.2.d /s/, /z/, /ʃ/, /h/ 38

II.B.2.e /m/, /n/ 38

II.B.2.f /j/, /r/, /l/, /υ/ 39

II.C From Middle Persian to Proto-Modern-Persian 40

II.C.1 From Middle Persian to Proto-Modern-Persian: Vowels 40

II.C.2 From Middle Persian to Proto-Modern-Persian: Consonants 45

II.C.2.a /p/, /t/, /ʧ/, /k/ 46

II.C.2.b /b/, /d/, /ʤ/, /g/ 47

II.C.2.c /f/, /x/ 48

II.C.2.d /ð/ /ɣ/ 48

II.C.2.e /ʃ/ /ç/ /s/ /h/ /z/ 49

II.C.2.f /m/, /n/ 50

II.C.2.g /j/, /r/, /l/, /υ/ 50

III. From Proto-Persian to Dari 52

III.A Dari Phonology: Introduction 52

III.B Dari Vocalic Phonemic Inventory 53

[2] /ç/ and /θ/ shall be treated together as the former is merely /θr/ with elided /r/.

III.C Dari Vowels Illustrated..... 56
III.C.1 High Front Vowels: [i].....56
III.C.2 High Back Vowels: [u], [ʊ].....57
III.C.3 Mid Front Vowels: [e], [ɛ], ([ɪ]), ([ə])..... 58
III.C.4 Mid Back Vowel: [o]..... 60
III.C.5 Low Vowels: [a], [ɑ], ([ɔ]).....64
III.C.6 Diphthongs: [aw], [aj], [ɑj], [uj]..... 65
III.D Dari Vowels and Diphthongs: Overview..... 72
III.D.1 Vowel Overview..... 72
III.D.2 Diphthong Overview.....74
III.E Dari Consonants: Introduction..... 76
III.E.1 Dari Aspiration.....78
III.E.2 Dari Phonology: h-drop..... 79
III.E.3 Dari Phonology: ʔ-drop.....80
III.E.4 Dari Phonology: h-drop, ʔ-drop and Compensatory Lengthening.....80
III.E.5 Dari Phonology: r-drop.....83
III.E.6 Dari Phonology: Definite Article Particle “را” (/rɑ/).....84
III.E.7 Dari Phonology: n-drop..... 85
III.E.8 Dari Phonology: /ʤ/ > [j].....86
III.E.9 Dari Phonology: /f/ and /b/ > [w].....86
III.E.10 Dari Phonology: Verbal Morphology and Possessive Pronouns..... 87
III.E.11 Dari Phonology: Third Person Copula “است” (/ast/).....88
III.E.12 Dari Phonology: Phonotactics.....89
III.E.12 Dari Phonology: Other Phonological Phenomena..... 90
III.E.13.a Dari Phonology: Vowel Lenition..... 91
III.E.13.b Dari Phonology: Word-final Consonant Shifts..... 92

III.E.13.c Dari Phonology: Metathesis..... 93
III.E.13.d Dari Phonology: s-drop..... 93

IV. From Proto-Persian to Iranian Persian..... 94
IV.A Iranian Phonology: Introduction..... 94
IV.B Iranian Vocalic Phonemic Inventory.....95
IV.C Iranian Vowels Illustrated..... 96
IV.C.1 High Front Vowel: [i]..... 96
IV.C.2 High Back Vowel: [u].....99
IV.C.3 Mid Front Vowel: [ɛ]..... 100
IV.C.4 Mid Back Vowel: [o]..... 104
IV.C.5 Low Vowels: [æ], [ɑ].....107
IV.C.6 Diphthongs: [ej], [ɑj]..... 111

IV.D Iranian Vowels and Diphthongs: Overview..... 115
IV.D.1 Vowel Overview..... 115
IV.D.2 Diphthong Overview.....117
IV.E Iranian Consonants: Introduction..... 119
IV.E.1 Iranian Aspiration..... 120
IV.E.2 Iranian Phonology: Palatalization..... 121
IV.E.3 Iranian Phonology: h-drop and ʔ-drop.....123
IV.E.4 Iranian Phonology: n-drop..... 125
IV.E.5 Iranian Phonology: r-drop.....125
IV.E.6 Iranian Phonology: Definite Article Particle “را” (/rɑ/).....126
IV.E.7 Iranian Phonology: Phonotactics..... 127
IV.E.8 Iranian Phonology: /q/ > [ɢ]..... 128

IV.E.9 Iranian Phonology: Verbal Morphology and Possessive Pronouns... 129

IV.E.10 Iranian Phonology: Third Person Copula “است” (/æst/)..... 129

IV.E.11 Iranian Phonology: “یک” (/jɛk/)..... 130

IV.E.12 Iranian Phonology: “گذاشتن” (/go-zɑʃ-tænˈ/)..... 130

V. From Proto-Persian to Tajik..... 132

V.A Tajik Phonology: Introduction..... 132

V.B Tajik Vocalic Phonemic Inventory..... 133

V.C Tajik Vowels Illustrated..... 135

V.C.1 High Front Vowel: [i]..... 135

V.C.2 High Back Vowels: [u]; [ʊ]..... 138

V.C.3 Mid High Vowel: [ʉ]..... 139

V.C.4 Mid Front Vowel: [e]..... 141

V.C.5 Mid Back Vowel: [o]..... 143

V.C.6 Low Vowel: [a]..... 144

V.C.7 Diphthongs: [aj], [oj], [ʉj]..... 144

V.D Tajik Vowels and Diphthongs: Overview..... 148

V.D.1 Vowel Overview..... 148

V.D.2 Diphthong Overview..... 150

V.E Tajik Consonants: Introduction..... 152

V.E.1 Tajik Phonology: Aspiration..... 153

V.E.2 Tajik Phonology: Palatalization..... 154

V.E.3 Tajik Phonology: h-drop and ʔ-drop..... 157

V.E.4 Tajik Phonology: n-drop..... 158

V.E.5 Tajik Phonology: r-drop..... 160

V.E.6 Tajik Phonology: Definite Article Particle “po” (/ro/)..... 161

V.E.7 Tajik Phonology: Verbal Morphology and Possessive Pronouns.....161
V.E.8 Tajik Phonology: Third Person Copula “аст” (/ast/)..... 162
V.E.9 Tajik Phonology: Phonotactics..... 162
V.E.10 Tajik Phonology- Dark /l/ (/ɫ/).....163

VI. OT Analysis.....165
VI.A OT Analysis: Introduction.....165
VI.A.1 OT Analysis: Introduction to OT Sound Change..... 165
VI.A.2 OT Analysis: OT Sound Change and the Initial State..... 166
VI.A.3 OT Analysis: Length Matters.....168
VI.A.4 OT Analysis: From Middle Persian to Proto-Persian..... 175
VI.B From Proto-Persian to Dari: Vowels..... 177
VI.C From Proto-Persian to Iranian: Vowels..... 179
VI.C.1 *-ATR +High.....181
VI.C.2 *+ATR –High..... 182
VI.C.3 Max-Low (-ATR +Front) and Max-Low (+ATR +Back)..... 184
VI.D From Proto-Persian to Tajik: Vowels.....185
VI.D.1 *+Low +ATR.....187
VI.D.2 *+ATR –High and *-ATR +High.....188
VI.E Symmetry and Conspiracy in PP Markedness Constraints..... 189
VI.F Diphthongs..... 191
VI.F.1 Diphthongs: From PP to Dari..... 194
VI.F.2 Diphthongs: From PP to Iranian..... 197
VI.F.3 Diphthongs: From PP to Tajik.....201
VI.F.4 Diphthong Overview.....202
VI.G Other Vowel Issues: Introduction.....202
VI.G.1 Iranian /ō/.....203
VI.G.1.a /ō/: MP /ōj/ → Dari /ō/ or /ūj/; Iranian /ū/; Tajik /ū(j)/ or /ʉ(j)/. 205

VI.G.1.b /ō/: MP /ūh/ → Dari and Iranian /ō/..... 208

VI.G.1.c /ō/: MP /ō/ → Dari and Iranian /ō/..... 209

VI.G.1.d /ō/: MP /ō/ → Dari /ū/ and Iranian /ō/..... 211

VI.G.2 Iranian /ɛ/..... 212

VI.G.2.a Iranian Word-Final /ɛ/..... 214

VI.G.2.b Iranian [ɛ] and Adjacent Palatals.....215

VI.G.3 Iranian Pre-Nasal /ɑ/ → [u].....218

VI.G.4 The Status of Tajik /ʉ/.....222

VI.H Proto-Persian Consonants: Introduction.....222

VI.H.1 Proto-Persian Consonants: Aspiration..... 225

VI.H.2 Proto-Persian Consonants: h/ʔ-drop..... 225

VI.H.2.a h/ʔ-drop and Compensatory Lengthening (Dari)..... 227

VI.H.2.b Exceptions to Compensatory Lengthening (Dari).....228

VI.H.2.c Compensatory Lengthening in Non-/ah/, /aʔ/ Cases (Dari)..... 230

VI.H.2.d /h/ and /ʔ/ in Verbal Morphology (Dari)..... 231

VI.H.2.e Dari /h/ → [j] and /ʔ/ → [j].....231

VI.H.2.f Dari /h/ → [w] and /ʔ/ → [w]..... 233

VI.H.2.g h/ʔ-drop and Compensatory Lengthening (Iranian)..... 235

VI.H.3 Proto-Persian Consonants: r-drop 238

VI.H.3.a Proto-Persian Consonants: Onset Cluster r-drop (Dari).....240

VI.H.4 Proto-Persian Consonants: n-drop..... 241

VI.H.4.a The Problem of “آن” (/ɑn/) (Dari, Iranian, Tajik)..... 243

VI.H.4.b n-drop Within Lexical Items and Phrases (Dari).....245
VI.H.4.b.i Dari Plural Demonstrative Pronouns- Formal Register..... 246
VI.H.4.b.ii Dari Plural Demonstrative Pronouns- Slightly Less Formal Register..... 247
VI.H.4.b.iii Dari Plural Demonstrative Pronouns- Colloquial Register.. 248
VI.H.4.b.iv Dari Plural Demonstrative Pronouns- Most Colloquial Register.....250
VI.H.4.b.v Tajik/Iranian Plural Demonstrative Pronouns.....251
VI.H.5 Proto-Persian Consonants: /ʤ/ > [j] (Dari).....252
VI.H.6 Proto-Persian Consonants: /f/ and /b/ > [w] (Dari).....253
VI.H.6.a Proto-Persian Consonants: /b/ > [w] / __]$_\sigma$ (Dari).....254
VI.H.6.b Proto-Persian Consonants: /-ɑb/ > [-aw] / __# (Dari)..... 255
VI.H.6.c Proto-Persian Consonants: /b/ > [w] / __C]$_\sigma$ (Dari)..... 256
VI.H.6.d Proto-Persian Consonants: /b/ > [w] / C$_\sigma$ -[__ (Dari).....257
VI.H.6.e Proto-Persian Consonants: /f/ > [w] / __]$_\sigma$ C[+Velar] (Dari).... 258
VI.H.7 Proto-Persian Consonants: Palatalization (Iranian, Tajik).....259
VI.H.7.a Proto-Persian Consonants: Iranian Palatalization..... 260
VI.H.7.b Proto-Persian Consonants: Tajik Palatalization..... 261
VI.H.8 Proto-Persian Consonants: /q/ > [ɢ] (Iranian)..... 264
VI.H.8.a /q/ > [ɢ]: The Reanalysis Approach (Iranian)..... 265
VI.H.8.b /q/ > [ɢ]: The Convergence Approach (Iranian)..... 269
VI.H.9 Proto-Persian Consonants: Dark /l/ (/ɫ/) (Tajik)..... 271
VI.J Proto-Persian Phono-Morphological Interplay (Dari, Iranian, Tajik).... 273
VI.J.1 Verbal Morphology and Possessive Pronouns.....274
VI.J.1.a Verbal Morphology: /d#/ → ∅ (Dari, Iranian, Tajik)..... 282

VI.J.1.b Verbal Morphology: /d#/ → /n#/ (Dari, Iranian, Tajik)..... 284

VI.J.1.c Present Perfect Enclitic Phonology (Dari, Iranian, and Tajik)... 286

VI.J.1.c.i Present Perfect Enclitic Phonology (Dari and Tajik)..... 286

VI.J.1.c.ii Present Perfect Enclitic Phonology (Iranian)..... 290

VI.J.1.d Verb Stem Internal Reduction (Dari, Iranian, Tajik)..... 292

VI.J.1.d.i Verb Stem Internal Reduction: /ha/ → ∅..... 294

VI.J.1.d.ii Verb Stem Internal Reduction: /aʊ/ → ∅..... 300

VI.J.1.d.iii Verb Stem Internal Reduction: /ū/ → ∅.....302

VI.J.1.d.iv Verb Stem Internal Reduction: /nī/ → ∅.....303

VI.J.1.d.v Verb Stem Internal Reduction: C# → ∅.....304

VI.J.1.d.vi Verb Stem Internal Reduction: Compound Verbs.....305

VI.J.1.e Verbal Affix Phonology (Dari, Iranian, Tajik).....306

VI.J.1.e.i Verbal Affix Phonology: The Negative Proclitic..... 308

VI.J.1.e.ii Verbal Affix Phonology: The Imperfective Prefix.....314

VI.J.1.e.iii Verbal Affix Phonology: The Subjunctive Prefix..... 319

VI.J.1.f Stress Heirarchy of Verbal Morphology (Dari, Iranian, Tajik)..324

VI.J.2 Proto-Persian Definite Article "را" (/rɑ/) (Dari, Iranian, Tajik)..... 326

VI.J.2.a Definite Article "را" (/rɑ/) (Dari)..... 326

VI.J.2.b Definite Article "را" (/rɑ/) (Iranian)..... 328

VI.J.2.c Proto-Persian Consonants: Definite Article "را" (/ro/) (Tajik)...329

VI.J.2.d Proto-Persian Consonants: n-drop and "را" (/rɑ/) Interaction.... 329

VI.J.3 Proto-Persian Consonants: 3rd Person Singular Copula "است" (/ast/)

(Dari, Iranian, Tajik)..... 331

VI.J.3.a 3rd Person Singular Copula "است" (/ast/) in Dari and Tajik.....335

VI.J.3.b 3rd Person Singular Copula "است" (/æst/) in Iranian.....336

VI.J.3.c 3rd Person Singular Copula “аст” (/ast/) in Tajik..... 339

VI.J.4 The Ezafe and Enclitic Vowel Lowering (Dari, Iranian and Tajik)..... 341

VI.J.4.a The Ezafe and Enclitic Vowel Lowering (Dari and Tajik)..... 344

VI.J.4.b Borrowed Enclitic Vowel Lowering (Dari, Iranian and Tajik).. 345

VI.K Proto-Persian Phonotactics (Dari, Iranian, Tajik)..... 348

VI.L Conclusion.....352

VII. Implications for Use-Based Approach to Language Change..... 353

VII.A Introduction..... 353

VII.B /n/-drop..... 355

VII.C /r/-deletion..... 357

VII.D /aʕ/ → [ɑ].....361

VII.D.1 /aʕ/ → [ɑ]: Frequency..... 361

VII.D.2 /aʕ/ - [ɑ]: Register..... 364

VII.D.3 /aʕ/-[ɑ]: The Lexicon..... 366

VII.D.3.a /aʕ/-[ɑ]: The Register Counterpart.....367

VII.D.3.b /aʕ/-[ɑ]: Register Environment.....368

VII.E Conclusions.....371

VIII. Appendix I: Proto-Persian Verb Stems..... 373

VIII.A -/d/ Past Stems.....376

VIII.B -/ft/ Past Stems.....382

VIII.C -/st/ Past Stems..... 383

VIII.D -/ʃt/ Past Stems.....385

VIII.E -/xt/ Past Stems.....386

VIII.F Suppletive Forms.....387

IX. Appendix II: Field Methodology.....388
IX.A Questions for Minimal Pair Ellicitation and Segment Clarification..... 388
IX.B Questions for Narrative Ellicitation.....393

X. Bibliography.....396
X.A Historical Persian and Persian Dialectology..... 396
X.B Kabuli Persian..... 498
X.C Tehrani Persian.....400
X.D Dushanbe Persian..... 402
X.E Fieldwork Methodology.....403
X.F Optimality Theory and Use-Based Phonology.....404
X.G Miscellaneous.....406

I. Introduction

I.A A Note on Terminology

From the beginning a note must be made on terminology. Nomenclature pertinent to the area of Persian languages seems an unending source of contention (cf. Spooner, 1994), and so rather than justify any of my own terminological norms, I will state them, and may the toes finding themselves stepped on forgive me. This essay and forthcoming dissertation shall dispense with the term 'Farsi,' as this is simply the Persian word for 'Persian,' and is furthermore offensive to many Iranians when coined in English. The unfortunate result of the current objection to this term is that there remains a great deal of confusion at this point. The preferred nomenclature, both for the ethnicity and the language, of those whose national roots are in modern day Iran is 'Persian,' but then how does one differentiate between the dialect of Iran and that of other Persians? Surely the Second World War has ruined the term 'Aryan' forever, and so I will avoid that as well. I am forced to avoid ambiguity by choosing the lesser of two evils.

'Persian' will not be used to refer to a citizen of the Islamic Republic of Iran nor to any particular language, but rather to the language family and corresponding ethnic affiliation which spans throughout ancient Persia, and includes the entirety of its offspring. Thus the Persian language family shall incorporate such distinct languages as Pashto, Kurdish, Baluchi, and dozens of others of like descent. "Modern Persian" shall refer to the conglomerate of the Iranian, Dari, and Tajik dialects, because reference to the Tehrani dialect as 'Persian' discredits the ethnic Persians of Afghanistan, Turkey, Iraq, Tajikistan, and other reaches of the former Soviet Union and even Iran outside of the

capital. The Persian language that serves as the official language of Iran will thus be referred to as 'Iranian,' that of Afghanistan (barring Pashto) will be referred to as 'Dari,' and that of Tajikistan, 'Tajik.' More specifically, in this dissertation, 'Iranian' shall refer to the standard dialect of Tehran, 'Dari' shall refer to the standard dialect of Kabul, and 'Tajik' shall refer to the Dushanbe standard dialect. The term 'Indo-Iranian' shall be kept as the normative reference to the branch of Indo-European which gave birth to the Persian and Indic languages (cf. tree chart in section I.C). The term 'Aryan' will be avoided entirely as it confuses the distinctions laid out above.

I.B Review of Existing Literature:

There has been some, although not a great deal, written about the dialectology of modern Persian, and some research at the University of Moscow is still devoted to Iranian Persian, Dari and Tajik, but none of it seeks to unify the dialects of Modern Persian into a Proto-Modern-Persian reconstruction. To my knowledge, the only straightforward comparison of the principal dialects is in Wei (1962). This however is not a scientific work, but rather a didactic handbook aimed at pointing out some systematic differences between Iranian, Afghan, and Tajik Persian so that students of one or the other might have a frame of reference. The data are useful but more of a textbook aid for students of Iranian Persian than an overview of synchronic variation. A very brief synopsis of Modern Persian dialectology can be found in Windfuhr (1990) at the end of a rather useful overview of Iranian (to which he consistently refers canonically as 'Persian'). This even briefly posits a set of "Early Modern Persian" vowels, but with no justifications for them, nor any mention of consonants or semivowels.

Other works such as Nawid (1994) and Ivanov (1995) deal with certain specific differences between two or three of the dialects. The former is an informative look at some, but by no means all of the differences between Iranian Persian and Afghan Dari. It may be viewed as a good start. Ivanov (1995) is also a useful work dealing with comparative word boundary issues in Dari, Iranian, and Tajik.

Others such as Sokolova (1953); Shakibi (1980, 1988, 1989); and Christensen (1930) deal with much more variant Persian languages such as Kurdish and Baluchi, and treat Dari, Iranian and Tajik as one Persian dialect, ascribing to it the qualities of the Iranian variety. Sokolova (1953) is a deft coverage of Baluchi, Kurdish, Ossete, Yaghnobi, and other lesser Persian languages not relevant to the present course of inquiry. Shakibi (1980, 1988, 1989) cover more of the Northwest Persian Languages spoken in and around Iran, but not the broader dialect of Modern Persian. Oranskij (1998) makes very brief mention of Dari, Iranian, and Tajik as distinctive in a few features, but is mainly devoted to more distinct varieties.

Yet another body of scholarship, exemplified by Matini (1994) and Spooner (1994) examine the issue of whether or not Dari, Iranian, and Tajik are dialects, languages, or merely accents. This is a topic of interest as it may lend itself to the study of how attitudes can drive language change. One issue concerning these works is that they tend to be written primarily as social commentaries objecting to nomenclature and bemoaning the state of political posturing in terms of language. They do however bring to light factors leading to the divergence between dialects.

The remainder of literature on Iranian, Dari and Tajik covers the particular varieties individually. Iranian is covered extensively, as it is considered to be the standard for Modern Persian. Lazard (1957, 1992) is a concise and informative overview of Iranian Persian covering the phonology and morpho-syntax of the variety as is Mace (2003), but the latter delves slightly deeper into the modern spoken form. McLendon (1960) is a segmental phonemic analysis derived from native informants and serves as a descriptive look at Iranian phonemes. Windfuhr (1997) is brief and concise, and is a thorough guide to the phonology of Iranian.

Dari follows Iranian in number of works written on the topic, with a handful of local, Soviet, and western grammars. Bogdanov (1930) is a collection of notes written over a four-year period in which the author resided in Kabul. It is written from the standpoint of a non-native Iranian speaker, and since Iranian is the frame of reference, it is somewhat comparative in nature. Dorofeeva (1960) is to my knowledge, the authoritative grammar of Dari, and like most Soviet scholarship on Central Asian linguistics, is skillfully crafted. Farhadi (1955) contains something of a list of phonological, morphological and syntactic phenomena, and some comparisons with Iranian. Henderson (1972) is a well wrought phonology of Dari from a structuralist point of view. While it seems there is a healthy body of work on Dari, the scholarship is outdated, and many of the points raised have fallen into obsolescence. Furthermore, there is very little comparative analysis.

Tajik was, for quite some time, covered almost exclusively by Soviet scholarship, treated only marginally by Western scholars, for example Lazard (1956), which is a brief

look at the peculiarities of Tajik with Iranian as a frame of reference. Sokolova (1949), in typical Soviet fashion, is a meticulously thorough look at the Tajik language from articulatory phonetics to syntax. Rastorgueva (1963) is more succinct in nature, but equally meticulous and accessible. Lately there appears to be something of a return to Tajik in scholarship as Ido (2005) and Perry (2005) prove, the latter of which is a significant achievement in a descriptive grammar.

Finally there is a small body of work related to the history of Iranian and Indo-Iranian, for example, Beekes (1997) and Gray (1965) but all of these treat the Iranian Persian variant as the canonical Modern Persian form. I have found no historical/diachronic research carried out to date on the synchronic variation within modern Persian.

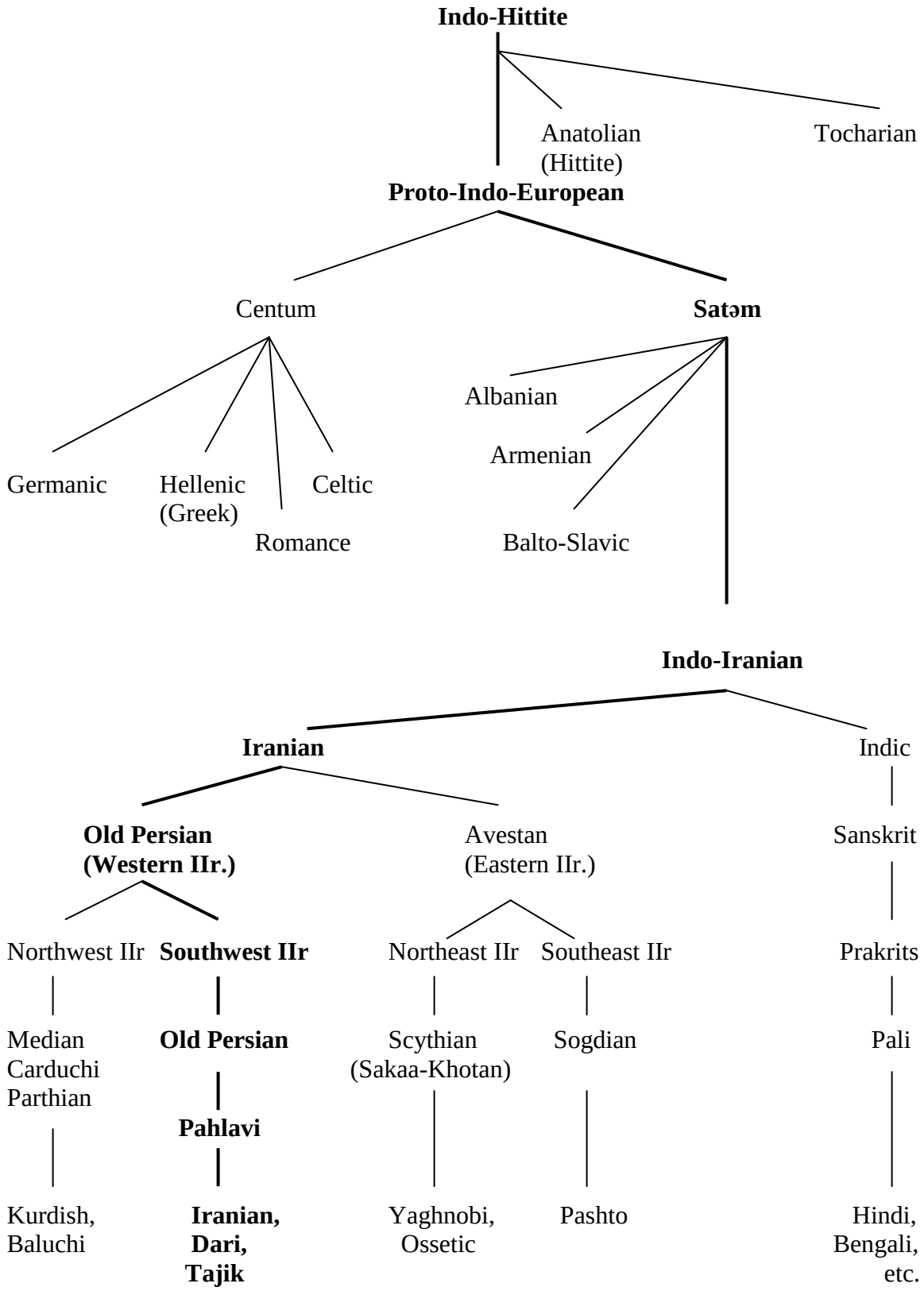
The scholarship as a whole provides a significant framework for the work at hand, but does not allow us to draw definite conclusions as to the developmental processes involved in the formation of the Modern Persian dialects. A comparative probe of natural language samples backed by past historical pursuits will provide ample fodder for a reconstruction of Proto-Persian.

I.C Background on Historical Indo-Iranian

Old Persian and Avestan are the two attested ancestor tongues of the Modern Persian branch of the Indo-European Family. Traditionally, Indo-European languages were divided into Eastern and Western along the lines of the phonetic realization of the Proto-Indo-European (PIE) phoneme /ǩ/ (or /ḱ/) either as a sibilant for eastern familial branches and a velar[3] for western branches. This is demonstrated by the distinction in the lexical item for 'hundred' in Latin ('centum') and in the Avestan ('satəm'). The discovery of Hittite, an eastern centum language, leads us to the conclusion that the eastern/western distinction is not valid, and the assessment in the chart on the following page is more accurate. Therefore, the Anatolian (Hittite) family and the Tocharian families are thought to have split from the parent tongue before a Centum/Satəm distinction developed. Then a five way split occurred between the Germanic, Hellenic, Celtic, and Italic families, and the Satəm Group. From this point the Satəm languages split into four distinct groups, Albanian, Armenian, Balto-Slavic, and Indo-Iranian, and the latter two each divided two ways producing Baltic, Slavic, Indic, and Iranian.

The following page contains a tree that illustrates the course of diachronic change that took place in the development of Modern Persian, tracing the lineage of Modern Persian in bold:

[3] Germanic will realize this sound as [h], a phenomenon that will be characterized by Grimm's Law.


Indo-Hittite
Anatolian
(Hittite)
Tocharian
Proto-Indo-European
Centum
Satəm
Germanic
Hellenic
(Greek)
Celtic
Romance
Albanian
Armenian
Balto-Slavic
Indo-Iranian
Iranian
Indic
Old Persian
(Western IIr.)
Avestan
(Eastern IIr.)
Sanskrit
Northwest IIr
Southwest IIr
Northeast IIr
Southeast IIr
Prakrits
Median
Carduchi
Parthian
Old Persian
Scythian
(Sakaa-Khotan)
Sogdian
Pali
Pahlavi
Kurdish,
Baluchi
Iranian,
Dari,
Tajik
Yaghnobi,
Ossetic
Pashto
Hindi,
Bengali,
etc.

The following is a more complete list of the modern languages within the Persian family, further emphasizing the justification for not referring to the official language of Iran or its natives as simply Persian:

Southwestern	Northwestern	Southeastern	Northeastern
- Iranian	- Kurdish	- Pashto	- Ossete
- Dari	- Talishi	- Yazgulami	- Yaghnobi
- Tajik	- Baluchi	- Shughni	- Pamiri
- Luri and Bakhtiari	- Gilaki	- Roshani	
- Kumzari	- Mazandarani	- Bartangi	
- Tati	- Zaza	- Oroshori	
- Shirazi dialects:	- Gurani	- Sarikoli	
· Kazerun	- Bashkardi	- Ishkashmi	
· Sivand	- Parachi	- Sanglechi	
· Lar	- Ormuri	- Zebaki	
	- Semnani dialects	- Wakhi	
	- Tat dialects:	- Munji	
	· Tabrizi Tat	- Yidgha	
	· Zanjani Tat		
	· Qazvini Tat		
	· Savehi Tat		
	- Vafsi / Ashtiyani		
	- Central Iranian:		
	· Kashani		
	· Esfahani		
	· Yazdi		
	· Kermani		
	· Dashte-Kaviri		

Modern Persian, being a member of the Southwestern Indo-Iranian subgroup, is the direct descendent of Old Persian (600BC-300AD), with intermediary attestations in Pahlavi (300AD-1100AD). It is important to note that Avestan was the language of the Mazdayasnians or Zoroastrians, whose texts date from around 600 BC, and Old Persian is mainly attested in the cuneiform inscriptions of the Achaemenian dynasty found in Persepolis, Naqš-i-Rustam, Murghab, Susa, Hamadan, Elvend, Van, and the Suez Canal,

dating from around 520 BC to around 300 BC. Middle Iranian, also known as Pahlavi, refers to the languages of the Persians spoken from around 300 BC to around 900 AD. The oldest Pahlavi dialect is that of the Arsacid dynasty, which ruled from 250 BC until 226 AD, when the Sasanian dynasty took over. The Sasanids ruled until the Islamic conquest of 652 AD. In the few centuries following the conquest there were numerous manuscripts from the eighth and ninth centuries found at Turfan, which include text in Arsacid and Sasanian as well as Sogdian and Eastern Iranian. Aside from the four Pahlavi dialects mentioned above, there exists a dialect of the Zoroastrians referred to as 'Book-Pahlavi,' including all of the Middle Iranian Pahlavi religious writings and translations of the original Avestan Zoroastrian texts. The most significant Pahlavi contribution to Modern Persian was its tendency to allow borrowing from Semitic languages. The extensive literature in Pahlavi will provide copious evidence to aid in the comparative reconstruction of Proto-Persian, but it is important to note that, aside from brief mention, the geneology presently under consideration will only pertain to the diachronic changes between Pahlavi and the modern descendent dialects.

I.D Modern Persian Dialectology

The principle distinctions between Iranian, Dari, and Tajik are phonological and lexical, and are so minimal that the three are almost entirely mutually intelligible with some effort. Syntactic differences are also plentiful, but will only be considered as they affect the respective phonologies. The interdialectal variation in terms of lexical stock is due to borrowing and semantic shift, and many of these influences have altered the phonetic and phonemic inventories, as well as affected phonological processes within the

dialects. The phonologies themselves have developed separately due to geopolitical isolation, but the distiction remains gradient, which is to say that a northern Afghan Dari speaker and a southern Tajik speaker will tend to have very similar phonologies yet retain marked lexical variation. The same is to be found with an eastern Iranian and a western Afghan. Below is an example of lexical variation, and the parenthetical following each example refers to the origin of borrowed lexical items if not original Persian[4] :

‘colonel’

Iranian:	سرهنگ، کولونل	[sær-hæng^{l}] [ku-lu-nɛl^{l}] (Persian, Fr)
Dari:	دگروال	[da-gar-wɑl^{l}] (Pashto)
Tajik:	полковник	[pol-kov^{l}-nik] (Russian)

‘train, railroad’

Iranian:	قطار، ترن، راه ای آهن	[ɢa-tɑr^{l}] [t^{ɛ}-rɛn^{l}] [rɑ^{l}-hɛ#ɑ-hæn^{l}] (Ar, Fr, Persian)
Dari:	خط ریل، راه ای آهن	[xa^{l}-tɛ#rel^{l}] [rɑj-ɑan^{l}] (English, Persian)
Tajik:	поезд, роҳи оҳан	[po-jezd] [ro^{l}-hi#o-han^{l}] (Russian, Persian)

‘momma’

Iranian:	مامان	[mɑ-mɑn^{l}] (French)
Dari:	مادر	[mɑ-dar^{l}] (Persian)
Tajik:	оча, модар	[o-tʃa^{l}] [mo-dar^{l}] (Uzbek, Persian)

These examples illustrate the lexical stock from which each language chiefly borrows, Iranian from French, Dari from English and Pashto, and Tajik from Russian and Uzbek. All of the above have borrowed heavily from Arabic before the time of their deviation into the present dialects, and naturally they all retain original Persian lexemes. This generalization is of course an oversimplification, as each language borrows from a number of lexical stocks, but it will serve to explain various phonological borrowings

[4] Note that Iranian Persian and Dari use a modified form of Arabic script, while Tajik uses a modified Cyrillic. Please refer to section I.G “IPA” and I.H “Abbreviations” for standards of notation.

such as Tajik palatalization and use of [ɬ], both from Russian, and Dari and Iranian differing over such uses as [ʒ] vs. [ʤ] in “جاپان” (Dari: [ʤɑ-pɑnˈ] ‘Japan’) and “ژاپون” (Iranian: [ʒɑ-punˈ] ‘Japan’).

I.E Introduction to Optimality Theory

The purpose of employing Optimality Theory is not that it will aid in the reconstruction but rather, it will provide an anchor of justification for the reconstruction, grounding it in physiological fact. At the outset, I mentioned that nomenclature is a source of some argument within the Persian community. One desired impact of this work will be to shed light on the issue of what Modern Persian should be considered, whether one language with three names or three sister languages. I intend to provide a conclusion that allows the scholar to form a more well founded opinion on the above question by viewing these languages as the product of concisely explained diachronic change brought about by reranking of constraints within Optimality Theory. I make no apologies however about the fact that I take exception with using language to expound upon theory. The intent of this research is to describe language using theory, and the reason I have chosen Optimality Theory is two-fold. Firstly, it works. Secondly, I intend to unify Persian by means of simply quantifying a theoretical set of language changes responsible for the resultant variation. Using Optimality Theory, one strikingly similar set of constraints will be shown to conspire distinctly in each language to produce rather conflicting results. That is, varied interplay of the same featural markedness and

faithfulness constraints will lead to opposing changes both within each language, and between the three.

I will provide a short but detailed explanation of the proposed OT model. OT presupposes a finite, universal set of markedness constraints on language grounded in ease of articulation and/or ease of parsing, and attested in cross-linguistic evidence. For example, "$*_{\sigma}$[[+lat][+nas]" (That is, a lateral – nasal cluster is marked in syllable onset position). Not only is onset coarticulation of these sounds physiologically difficult, this sequence is rarely attested in language, if at all. Hence, this constraint is sound. A second type of constraint, faithfulness constraint, simply states that an input segment (phonemic or underlying level) will be realized at the output (phonetic or surface level). Phonological processes occur as universal faithfulness and markedness constraints compete in a language-specific hierarchy. Thus as long as a faithfulness constraint on lateral – nasal onset clusters (Ident-IO ([+lat][+nas])) outranks the markedness constraint "$*_{\sigma}$[[+lat][+nas]" in a specific language's hierarchy of constraints, then that language will allow the marked construction. However, if "$*_{\sigma}$[[+lat][+nas]" outranks that faithfulness constraint, all lateral – nasal onset clusters will somehow be eliminated. Crucially, this allows for violability of constraints, thus permitting one language to allow shifts that are prohibited in another, but maintaining a universal set of constraints within the theory. The particulars of the abovementioned elimination process will be discussed presently.

My theoretical model assumes that language change begins as the promotion of markedness constraint(s) over faithfulness constraint(s), and that the respective ordering of constraints weakens due to pressures of the constraint grounding being either ease of

articulation or ease of parsing. There is then a period of optionality, and subsequent gradual preference for the raised constraint, and eventual reanalysis of the sound due to Lexicon Optimization on the part of the generation born after the preference is set. The language change is finalized with demotion of the markedness constraint and possibly the promotion of a faithfulness constraint, restoring the highly ranked faithfulness constraint to preserve the new sound.

Below is the chart employed by the theory to express the above detailed hierarchy:

/δ/ → [α]

Input δ	ID-IO (η)	ID-IO (θ)	*ι	*κ
☞α				*
β		*!	*	*
γ	*!	*		*
δ			*!	*
ε		*!	*	*
ζ	*!			

In the above chart, the first column contains an input (or underlying) segment /δ/ and an infinite number of potential outputs including the input segment itself, but only /α/ through /ζ/ are shown. In practice, it is not necessary to include each potential output, but rather only those that must be ruled out. For instance, in an analysis of the MP to Tajik change /a/ → /o/, it is necessary to include [o], [ɔ], [ə], [æ], [∅] and even [a] as possible outputs, but clearly not necessary to include /a/ → [fɪˈ-lə-bəkːs], as this alternation is highly unlikely. The first row then contains four ranked constraints in the following order: ID-IO (η), ID-IO (θ), *ι, and *κ. The first two constraints are faithfulness constraints mandating that the output must match the input for the features η and θ, and the second two are

markedness constraints declaring the segments (or features) ι and κ marked. The output [α] is indicated by '☞' and is determined by the fact that β through ζ all violate higher ranking constraints than α, indicated by * and *!, the latter being terminal violations. Thus even though α violates *κ, it is still optimal, as *κ is outranked by the other relevant constraints.

To exemplify this, an example from Iranian (a propos) will be used to illustrate how a divergent constraint re-ranking can account for the following changes between MP and Iranian: /e/ > /i/ and /o/ > /u/. More simply put: [+mid] > [+high], or *ō/*ē ('no +Long/+ATR Mid vowels). Assuming that ATR is a process that is associated with more closed vowels, it is fair to state that a +ATR vowel should be maximal in height. Thus, I propose that this vowel shift occurs due to a condition on ATR, obeying the following constraint:

*[+ATR –High] ("Maximize the height of a +ATR vowel." The constraint is grounded in ease of articulation, and the physiological fact that ATR is a height-proportional feature.)[5]

e > i (condition on ATR)

MP *e	ID-IO(ATR)	ID-IO(Back)	*+ATR -Hg	ID-IO(Hg)
☞i				*
o		*!	*	*
ʊ	*!	*		*
e			*!	*
ɔ		*!	*	*
ɪ	*!			

[5] This constraint was proposed by Calabrese (1988) before OT was postulated. See Kenstowicz (1994, p. 529).

o > u (condition on ATR)

PIE * o	ID-IO(ATR)	ID-IO(Back)	*+ATR -High	ID-IO(Hg)
☞ u				*
i		*!	*	*
ɪ	*!	*		*
o			*!	*
ɔ		*!	*	*
ʊ	*!			

In both analyses, the –ATR segments are eliminated (*!) by the highest constraint ID-IO(ATR), the segments mismatched for backness are eliminated by the second constraint ID-IO(Back), the input segments are eliminated by the markedness constraint *+ATR –High, and [u] is left as the only possible outcome. These assertions will only stand if they can be seen not to affect the remainder of the vowel inventory of Iranian. Rather than drag this out longer than it has to be, suffice it to say that this is only a small sample of the OT analysis of Persian, and so far, the model fits.

I.F Methods of Data Collection

Data collection was the first major undertaking in this study, and it represented an effort to collect sufficient narrative from a great enough number of native speakers to compile three separate phonologies, Dari, Iranian and Tajik. The first step was to determine which native informants to use. Iranians and Afghans are plentiful in almost any metropolitan area due to the mass exodus of political asylees since the 1979 Soviet invasion of Afghanistan and Islamic Revolution in Iran. After identifying several pockets of Iranians, Afghans and Tajiks living in the Washington DC and Baltimore areas,

willing members were asked to contact other members of their communities to soliticit further participation.

In all, over a hundred hours of narrative were recorded from around fifty informants spanning the gamit of the three nationalities, two sexes, and three age groups, 18-35, 35-50, and 50+, and then set lists of questions (cf. appendix II, section IX.A) were answered in which minimal pairs were imbedded. This was done using a small lavalier microphone clipped to the informants' lapel. After reviewing the narratives, I chose the recordings of nine informants from which to extract hundreds of vowel and consonant samples. The nine were chosen as their speech was the most representative dialectal usages, and the recordings were the most clear. These were then analyzed and an average formant freqency contour was determined for each vowel in each dialect. The first and second frequencies of each vowel from the aforementioned hundreds of samples were charted and are displayed at the end of each vowel section in chapters III through V. The vowel portions themselves include spectrograms from individual speakers that match, as closely as possible, mean formant contours of the entire group of over fifty. Note that the abovementioned age groups were not sought in an effort to determine intergenerational variation, but rather to broaden the net in terms of sample quality.

Sample quality, however is affected by more than demographics. With proper equipment, it is fairly simple to control for sound, but eliciting a non-performative, "natural" sample from native speakers of Tehrani Persian, Kabuli Dari, and Dushanbe Tajik was difficult and yet imperative for a proper comparison. Performative speech is inevitable due in part to my non-native command of the language, which potentially

causes speakers to clarify their speech for my benefit, and also in part to the formal nature of an interview setting (cf. “the observer’s paradox” (Labov, 1986, p. 30)). This was compounded by the fact that Iranian and Dari are still written with classical, essentially Middle Persian (Pahlavi) orthography, and hence speakers will still tend in their speech to converge on the written language in formal settings. In essence, the sentence "آن مرد من را دیده است." (‘that man has seen me’), chosen arbitrarily, if read as written would be [ɑn mard man rɑ diˈ-da ast] in both Dari and Iranian, but colloquial renderings would be as follows:

Iranian: [un mærd mæˈ-nu di-dɛːˈ]

Dari: [u mard maˈ-ra di-daˈ]

Tajik: [vaj mard maˈ-na di-da-giˈ]

Another obstacle to collecting quality samples is the fact that my informants have all lived for at least some period outside of their native land, often adding new foreign elements to the sample. In light of this, my goal was to collect data of sufficient quality and quantity for an encompassing overview, providing all of the phonetic and phonological data necessary for analysis. To that end, I elicited narratives of personal experience, fostered group interaction, and sought topics which deemphasized the presence of recording equipment. When certain structures were still lacking from the corpus of “natural” speech, I resorted to more structured elicitation methods. Aside from this, I conducted numerous informal interviews without recording equipment taking mental notes on grammatical forms observed.

To avoid having the data converge on the written form, I devised topics of discussion that were intrinsically personal to the informants designed to eliminate the observer paradox. Labov (1986) cites "death and danger of death, sex, and moral indignation" (p. 33) as foci of interest that span many speech communities. Milroy (2003) contends that foci of interest cannot be universalized because " not all topics work for all speakers." (p. 61) Sex is not a comfortable option for any Muslim, but fear of death and moral indignation are a way of life for inhabitants of the region under consideration, so I directed discussions around the Soviet Union, the Islamic Revolution, the West, and the Taliban, as well as several items of particular historical note in Tajikistan. All of these are topics which relate to death and evoke moral indignation, but are at the same time close to the heart of the Persians under examination. With little exception, the narratives were minimally performative. Appendix II, section IX.B is a list of the questions that I employed in directing narratives.

I.G IPA and Abbreviations Used

Throughout the course of this dissertation, I will naturally use copious lexical items from Iranian, Dari, and Tajik. I will establish from the outset my norms of anotation and justifaction for them. For Dari and Iranian lexemes, I will provide the lexical item in original Perso-Arabic script and for Tajik, Cyrillic. That will invariably be followed by a transcription in International Phonetic Alphabet (IPA), and an English gloss surrounded by single quotes. The reason for choosing IPA over APA (American Phonetic Alphabet) and for providing each lexical item in its native script lies in the hope that this research may prove a universally useful tool in both eastern and western

linguistic scholarship. Below are the IPA symbols used throughout the paper. The default language for examples below will be given in American English (in *italics*, first column) and formal Dari, unless otherwise indicated. I attempted to find homophonous lexical items in English and Dari for each of the below examples:

I.G.1 IPA: Vowels

High:

[i] high front tense unround	*shear*	شیر [ʃir] ‘milk’
[ɪ] high front lax unround	*dill*	دل [dɪl] ‘heart’
[ʉ] high central lax round	*Küche* (German)	кӯхна (Tj) [kʉh-naˈ] ‘old’
[u] high back tense round	*hoot*	حوت [hut] ‘Pisces’
[ʊ] high back lax round	*put*	پت [pʊt] ‘hidden’

Mid:

[e] mid front tense unround	*fail*	فعل [fel] ‘verb’
[ɛ] mid front lax unround	*dell*	دل [dɛl] ‘heart’[6]
[o] mid back tense round	*rose*	روز [roz] ‘day’
[ɔ] mid back lax round	*hall*	حال [hɔl] (var.) ‘state’
[ə] mid central unround	*fun*	فن [fən] (var.) ‘science’

Low:

[æ] low front unround	*man*	من (Ir) [mæn] ‘I’
[a] low central unround	*rod*	رد [rad] ‘denial’
[ɑ] low back tense round	*hall*	حال [hɑl] ‘state’

I.G.2 IPA: Consonants

Labial:

[p] voiceless bilabial stop	*pull*	پل [pʊl] ‘bridge’
[b] voiced bilabial stop	*boot*	بود [budˤ] ‘was’
[f] voiceless labiodental fricative	*fun*	فن [fən] ‘science’
[v] voiced labiodental fricative	*valley*	ولی (Ir) [væ-liˈ] ‘but’

[6] Note that “دل” is transcribed both as [dɪl] and [dɛl]. These vowels are unpredictable free variations.

[m] voiced bilabial nasal	*mall*	مال [mɑl] ‘item’
[w] voiced bilabial approximate	*wee*	وی [wi] ‘oh!’

Dental / Alveo-palatal / Post-Alveolar:

[θ] voiceless dental fricative	*thigh*	*n/a*
[ð] voiced dental fricative	*thy*	*n/a*
[t] voiceless alveolar stop	*tie*	طی [taj] ‘during’
[d] voiced alveolar stop	*dude*	دود [dud] ‘smoke’
[s] voiceless alveolar fricative	*sin*	سن [sɪn] ‘age’
[z] voiced alveolar fricative	*zoot*	زود [zudˤ] ‘soon’
[n] voiced alveolar nasal	*now*	نو [naw] ‘new’
[r] voiced alveolar trill	*raw*	راه [rɑ] ‘way’
[l] voiced lateral approximate	*lab*	لب [lab] ‘lip’
[ʧ] voiceless alveolar affricate	*chaw*	چاه [ʧɑ] ‘well’
[ʤ] voiced alveolar affricate	*jam*	جمع [ʤam] ‘together’
[ʃ] voiceless postalveolar fricative	*shah*	شاه [ʃɑ] ‘king’
[ʒ] voiced postalveolar fricative	*janvier* (Fr)	ژانویه (Ir) [ʒɑn-vjɛˈ] ‘January’
[j] voiced palatal approximate	*yak*	یک [jak] ‘one’

Velo-uvular / Glottal:

[k] voiceless velar stop	*car*	کر [kar] ‘deaf’
[g] voiced velar stop	*gap*	گپ [gap] ‘issue’
[x] voiceless velar fricative		خود [xʊd] ‘self’
[ɣ] voiced velar fricative		غم [ɣam] ‘sorrow’
[q] voiceless uvular stop		قند [qand] ‘sweet
[ɢ] voiced uvular stop		قدر (Ir) [ɢædɾˈ] ‘quality’
[χ] voiceless uvular fricative	*Bach* (Gm)	وقت [waχt] ‘time’
[ʁ] voiced uvular fricative	*Arles* (Fr)	عقل (Ir) [aʁl] ‘intellect’
[ʕ] voiced pharyngeal fricative	عین (Ar) [ʕajn] ‘ʕ’	*n/a*
[ħ] voiceless pharyngeal fricative	حال (Ar) [ħæl] ‘state’	*n/a*
[ʔ] glottal stop		ادعا [ʔɛd-ʔɑˈ] (Fs) ‘claim’
[h] voiceless glottal fricative	*high*	هی [haj] ‘hey!’

I.G.3 Abbreviations

Ar	Arabic
Dr	Dari
En	English
Fr	French
Gm	German
IA	Indo-Aryan
IE	Indo-European
IIr	Indo-Iranian
Ir	Iranian Persian
MP	Middle Persian
P	Persian
Pa	Pashto
Pah	Pahlavi
PIE	Proto-Indo-European
PP	Proto-Persian
Ru	Russian
Tj	Tajik

I.H Constraints Used

Various constraints are proposed over the course of this dissertation, and below each is listed with its prose description. If a constraint has been proposed by another scholar, it will be followed by a citation.

I.H.1 Maximization and Faithfulness Constraints

MAX-μ-IO	("Ellipsis is prohibited.")
MAX-IO[αLong αATR]	("If input segment is αLong, output will be αATR.")
MAX-ATR +Low +Back	("Maximize Backness of Low +ATR segments.")
MAX-Low (-ATR +Front)	("Maximize Frontness of –ATR inputs.")
MAX-Low (+ATR +Back)	("Maximize Backness of +ATR inputs.")
MAX Height]#	("Maximize the height of word-final vowels.")
MAX-$_{LEX}$-IO	("Ellipsis in lexical items is prohibited.")

MAX(High)V[+Long -Strs]# ("Maximize height of word-final unstressed long vowels.")

MAX-$_{MORPH}$-IO ("Ellipsis of whole morphemes is prohibited.")

MAX-$_{LONG}$-IO ("Ellipsis of long vowels is prohibited.")

I.H.2 ID/DEP and Faithfulness Constraints

DEP-μ-IO ("Epenthesis is prohibited.")

ID-IO (α) ("Segments in the output match the input for feature α")

ID-$_{BACK}$-IO ("Output back vowels match their input segments.")

ID-IO (#[C) ("Word-initial inputs are represented in the output.")

I.H.3 Markedness Constraints

*-ATR +High ("-ATR High vowels are prohibited.")

*+ATR -High: ("+ATR –High vowels are prohibited.")

*+Back –Round ("Unrounded back vowels are prohibited.")

*Low +ATR ("Low +ATR segments are prohibited.")

*+Long –Low ("Long vowels must be open.")

*[υ] ("The segment [υ] is prohibited.")

*[w] ("The segment [w] is prohibited.")

*Fricative [+voice] ("Voiced Fricatives are prohibited.")

*C[+labial -obstruent] ("Labial glides and approximates are prohibited.")

*Diphthong ("Diphthongs are prohibited.")

*V[+High] [h]]$_{\sigma}$ ("Vowels preceding glottal spirants must lower.")

*C]$_{\sigma}$ ("Closed syllables are prohibited.")

$*_{\sigma}$[CCV	("Complex syllable onsets are prohibited.")
*+Back –Round]#	("Unround Back segments are prohibited word-finally.")
*VV	("Adjacent vowels are prohibited.")
*C[+Fricative +Voice]	("Voiced fricatives are prohibited.")
$*^{j}$i	("Palatals are prohibited before high front vowels.")
*[q]	("Voiceless uvular stops are prohibited.")
*V-[+Uvular]-V	("Intervocalic uvular segments are prohibited.")
*[+Uvular]#	("Word-final uvular segments are prohibited.")
*V-[-Voice]-V	("Intervocalic voiceless segments are prohibited.")
*Dorsal Fricative	("Dorsal fricatives are prohibited.")
*Velar [+voice]	("Voiced Velars are prohibited.")
*+V_nC_nC[oral stop]#	("Morpheme final oral stops are prohibited.")
*Asp [+voice]	("Voiced Aspirates are prohibited.")
*C+C…	("Cons. clusters across morpheme boundaries prohibited.")
*V[+Long -Stress]#	("Word-final unstressed long vowels are prohibited.")
*CCVC	("Onset consonant clusters are prohibited.")
*CVCC[+syl]	("Syllablic consonants are prohibited in syllable codas.)"

<u>I.H.4 Miscellaneous Well-formedness Constraints</u>

MAXIMAL DISTINCTION	("Diphthong components must be maximally distinct.")
MINIMAL DISTANCE	("Minimal auditory distance between contrasting forms") Miglio (1998: 9)
C^{h}V	("Word-initial simple onset voiceless plosives aspirate.")

V $C_1C_2C_3$	("Coda clusters must increase in sonority.")
HARMONY CC (velar)	("Adjacent consonants must match feature velar.")
REDUCE (-Stress)	("Reduce unstressed segments first.")
DHH (High): $ [αhigh] [αhigh] $	("Diphthong components converge for height.")
DBH (Back): $ [βback] [βback] $	("Diphthong components converge for backness.")
$C[+pal]_n$ [MAX Height] $C[+pal]_n$	("Maximize height of vowels between palatals.")
PITCH HARMONY (NASAL)	("Minimize frequency of vowel before a nasal.")
PALATAL HARMONY	("Palatalization should occur between segments when the tongue position between is palatal.")
REANALYSIS α →β	("When most allophones of phoneme α are identical to phoneme β, reanalyze α as β.")
ANALOGY (Plural verb stem)	("Morph prohibited segments togrammatically adjacent segments that do not cause ambiguity.")
VERB STEM REDUCE (1-15)	("Ellipsis occurs in descending order of sonority.[7]")
VSR (1)	("Low lax vowels in frequent verbs are ellided.")
VSR (2)	("Low tense vowels in frequent verbs are elided.")
VSR (3)	("Mid lax vowels in frequent verbs are elided.")
VSR (4)	("Mid tense vowels in frequent verbs are elided.")

[7] The Sonority Sequencing Principle traditionally considers sonority to the exists in the following ascending order (from most to least sonorant):
Low lax vowels
Low tense vowels
Mid lax vowels
Mid tense vowels
High lax vowels
High tense vowels
Semivowels
Flaps
Laterals
Nasals
Voiced fricatives
Voiceless fricatives
Voiced plosives
Voiceless plosives
Complex plosives

VSR (5)	(“High lax vowels in frequent verbs are elided.”)
VSR (6)	(“High tense vowels in frequent verbs are elided.”)
VSR (7)	(“Semivowels in frequent verbs are elided.”)
VSR (8)	(“Flaps in frequent verbs are elided.”)
VSR (9)	(“Laterals in frequent verbs are elided.”)
VSR (10)	(“Nasals in frequent verbs are elided.”)
VSR (11)	(“Voiced fricatives in frequent verbs are elided.”)
VSR (12)	(“Voiceless fricatives in frequent verbs are elided.”)
VSR (13)	(“Voiced plosives in frequent verbs are elided.”)
VSR (14)	(“Voiceless plosives in frequent verbs are elided.”)
VSR (15)	(“Complex plosives in frequent verbs are elided.”)

DIPHTHONG HEIGHT HARMONY

$$\$ \begin{bmatrix} \alpha\text{high} & \alpha\text{high} \\ \beta\text{back} & \beta\text{back} \end{bmatrix} \$$$

McMahon (2003: 137)

FRONT VOWEL PALATALIZATION (“Front vowels take onset palatalization.”)

HARMONY (Short Prefix) (“Prefix short vowels harmonize with verb stems.”)

II From Old Persian to Pahlavi to Proto-Persian

This dissertation will not assess language developments that occurred before the time of attestation of Old Persian, thus Old Persian shall be presented, and this chapter will trace the developments from Old Persian into Middle and in fact Proto-Modern-Persian. The chapter seeks to introduce the foundation of Persian and nothing more. A full analysis of the phonologies of Dari, Iranian and Tajik will be presented in chapters III, IV, and V respectively, and an Optimality Threoretic model for the shifts will be advanced in chapter VI, but for the sake of this overview, the phonologies and shifts are merely introducted. Section II.A will present the phonemic inventory of Old Persian as attested in the cuneiform writings of Ancient Persia. Section II.B will present the shifts from OP into Middle Persian, and Section II.C will constitute a reconstruction of Proto-Persian, being Proto-Modern-Persian or the intermediary step between Middle Persian and three Modern Persian variants currently under consideration.

II.A Old Persian

As with many historical linguistic pursuits, the inadequacy of the graphology precludes a straightforward analysis of Old Persian and Pahlavi. As Beekes (1997) points out, “The reading intended by the syllabic script can often be established only through comparison with the related languages.” (p.3) Thus, this portion is to be considered little more than an introduction to Persian’s place in the Indo-European family of languages. Below are the Old Persian graphical representations allowed by the cuneiform graphological system:

/ă/, /a/		/i/	/u/	
/pa/		/ta/, /tu/	/ʧa/	/ka/, /ku/
/ba/		/da/, /di/, /du/	/ʤa/, /ʤi/	/ga/, /gu/
/fa/		/θa/		/xa/
/ʃa/	/ça/[8]	/sa/	/za/	/ha/
/ma/, /mi/, /mu/			/na/, /nu/	
/ja/	/ra/, /ru/	/la/	/ʋa/, /ʋi/	

II.A.1 Old Persian Vowels

Because of the relatively small corpus and incomplete information, it can be assumed that the alphabet, like most, failed to capture the entirety of the Old Persian (OP) phonemic inventory, but it can safely be asserted in light of the above data and extant scholarship, that the Old Persian phonemic inventory contained the following six vowels and two diphthongs:

/ī/	/ū/	
/i/	/u/	
	/ā/	
	/a/	

Diphthongs: /ai/; /au/

[8] The phoneme represented by /ç/ is the historical reflux of /θr/, so at the time of OP, it is unclear what this sound actually was. In Modern Persian the /r/ has elided and the remaining /θ/ → /s/.

Below are lexical items exemplifying each vowel and diphthong of the OP vocalic phonemic inventory:

/a /	OP patar	‘father’
/ā/	OP dāta	‘law’
/u/	OP θuxra	‘red’
/ū/	OP būta	‘was’
/i/	OP ʧiθra	‘face’
/ī/	OP ʤīʋa	‘life’
/ai/	OP raiçaj-	‘to pour’
/au/	OP gauʃa	‘ear’

II.A.2 Old Persian Consonants

OP had the following consonantal phonemic inventory, of which a great deal remained in MP. Exemplary lexical items have been provided:

- Voiceless Stops/Affricates: /p/ /t/ /ʧ/ /k/

OP patar	‘father’
OP tava	‘you’
OP ʧaxra	‘wheel’
OP kadha	‘when’

- Voiced Stops/Affricates: /b/ /d/ /ʤ/ /g/

OP barta	‘borne’
OP dāta	‘law’
OP ʤīʋa	‘life’
OP gauʃa	‘ear’

- Voiced Aspirates: /ɣ/ /ð/[9]

OP raughna	‘oil’
OP madhu	‘wine’

- Voiceless Fricatives: /f/ /θ/ /x/

OP farzaiñti	‘offspring’
OP θuxra	‘red’
OP xri-	‘to buy’

- Sibilants: /ʃ/ /ç/ /z/ /h/

OP ʃātō	‘joyful’
OP çpaita	‘white’
OP zānu	‘knee’
OP haruva	‘all’

- Nasals: /m/ /n/

OP mātar	‘mother’
OP nava	‘new’

[9] It appears from the evidence that both /ɣ/ and /g/ were represented as /g/ and both /d/ and /ð/ were represented as /d/ in the cuneiform alphabet. Again, it is unclear what the actual forms of /ð/ and /ɣ/ were, save that they were originally /dh/ and /gh/ in their pre-OP form, and /ð/ → /z/ in PP and /ɣ/ remained. It is likely that Proto-Persian retained the phoneme /ð/ even after adoption of the Arabic script, as “گذاردن” [gu-zɑr-danˈ] (‘to put’) and “گزاردن” [gu-zɑr-danˈ] (‘to measure’) illustrate. These homophones vary in orthography only in the second letter “ذ” [zɑl] in the first and “ز” [ze] in the second. In Arabic, these graphemes still represent /ð/ and /z/ respectively, but they have converged as /z/ in Persian, for reasons much the same as the [zə] Shibboleth of the [zə] French accent in English.

- Semivowels and Liquids: /j/ /r/ /ʋ/

OP jātu	'witch'
OP raiçaj-	'to pour'
OP ʋāta	'wind'

II.B Old Persian to Middle Persian

Middle Persian (MP), or Pahlavi (Pah, the terms will be used interchangeably) will be seen to retain the same vowel inventory, but add mid vowels /ē/ and /ō/ and long diphthongs /āi/, /āu/ and /ūj/. As for the consonant inventory, there will be a number of shifts of allophonic convergence and split, but the only change to the phonemic inventory will be loss of the interdental fricative /θ/, which merges with /ç/. Again, the exact status of /ç/ is unclear save that in pre-OP it was definitively /θr/ and in PP it has definitively morphed into /s/. The sections that follow provide the sound shifts from OP into MP.

II.B.1 Old Persian to Middle Persian: Vowels

The vowel system in Old Persian essentially remains intact today with some minor shifts. As graphemic evidence attests, Old Persian has lost the Indo-European vocalic consonant */ṛ/, and that will be dealt with in section II.B.1.e. The remaining portions of section II.B.1 will deal with each of these vowels in the following order: /a/ and /ā/, /i/ and /ī/, /u/ and /ū/, /ai/ and /au/, and finally PIE */ṛ/.

II.B.1.a Low Vowels /a/ and /ā/

/a/ and /ā/ remained in Pahlavi, as is evidenced by the following:

OP martija	Pah.	/mart/	‘man’
OP hapta	Pah.	/haft/	‘seven’
OP dāna	Pah.	/dā-nak/	‘seed’
OP υāta	Pah.	/υāt/	‘wind’

There are, however, other cases which underwent phonologically conditioned change, but these appear to have occurred in verb stems undergoing inflectional change, and those alternations remain as synchronic inflectional ablauts:

OP barta	PP “برده”	/bur-da/	‘carried’
OP fra-mā-na	PP “فرمودن”	/far-mū-dan/	‘to order’

As mentioned, both of these instances exhibit the same shift synchronically in the Modern Persian dialects:

PP “بردم” /bur-dam/ ‘I carried’ → PP “می برم” /mē-ba-ram/ ‘I carry’

PP “فرمودم” /far-mū-dam/ ‘I ordered’ → PP “می فرمایم” /mē-far-mā-jam/ ‘I order’

II.B.1.b High Front Vowels /i/ and /ī/

The vowels /i/ and /ī/ also remained steady as exemplified by the following tokens:

OP ʧiθra	Pah.	/ʧihar/	‘face’
OP xʃīra	Pah.	/ʃīr/	‘milk’

II.B.1.c High Back Vowels /u/ and /ū/

Finally are the vowels /u/ and /ū/, and these also remain in PP:

OP çuxta	Pah.	/çux-ta/	‘burned’
OP θuxra	Pah.	/çuxr/	‘red’
OP būta	Pah.	/būʈ/	‘was’
OP dūra	Pah.	/dūr/	‘far’

II.B.1.d Diphthongs /ai/ and /au/

The diphthongs /ai/ and /au/ shifted to mid monophthongs in the exact fashion that it will be shown Iranian Persian to shift in later iterations of the monophthongs /ai/ and /au/. That is, /ai/ → /ē/ and /au/ → /ō/. This explains how an inventory of six vowels will shift to eight vowels in Proto-Persian:

OP rauʈʃah	Pah.	/rōʤ/	‘day’
OP gauʃa	Pah.	/gōʃ/	‘ear’
OP raiçaj-	Dari “ریختن”	/rēx-tan/	‘to pour’
OP çpaita	Dari “سفید”	/sa-fēd/	‘white’

But note that in some cases, /ai/ → /ī/, not /ē/:

OP υain-	Pah.	/υī-naʃn/	‘sight’

There appears to be no phonological conditioning for this exception.

One apparent paradox in this is the fact that the diphthongs /ai/ and /au/ will be shown to exist in Proto-Persian and even in fact remain in present day Dari:

“کی” /kaj/ ‘when’
“نو” /naw/ ‘new’

These diphthongs are in fact not original diphthongs but rather compensatory relics of earlier ellipses:

OP kadha	→	Dari “کی”	/kaj/	‘when’
OP naʋa	→	Dari “نو”	/naw/	‘new’

In the first instance, intervocalic /dh/ has weakened to palatal semivowel /j/, and in both cases, the final vowel is elided.

<u>II.B.1.e */ṛ/</u>

Evidence from Sanskrit and even from synchronic variation in Old Persian indicates that Persian retains historical remnants of PIE */ṛ/. The reason this bears mention at this point is to introduce the root cause of alternations like the following in Modern Persian:

“کردم” /kar-dam/ ‘I did; made’ → “کنم” /ku-nam/ ‘I do; make’

This has been labeled a suppletive form, but in light of historical data, that is a false claim. Historical data, to include the Modern English lexical item ‘create,’ indicates that the original pre-Old Persian Indo-Iranian verb root for this verb was */kṛ/, and the changes happened accordingly:

*/kṛ/ → */ku/ / +nau-mi (1st pers sg pres) → /ku-nau-mi/ → “کنم” /ku-nam/ ‘I do’

*/kṛ/ → */kar/ / +ta (pres participle) → /kar-ta/ → “کرده” /kar-da/ ‘done’

Note that the starred items above are Proto-Indo-Iranian. This will now explain the phenomenon addressed briefly in section II.B.1.a:

PP “بردم” /bur-dam/ ‘I carried’ → PP “می برم” /mē-ba-ram/ ‘I carry’

PP “فرمودم” /far-mū-dam/ ‘I ordered’ → PP “می فرمایم” /mē-far-mā-jam/ ‘I order’

These are now viewed as originally Indo-Iranian */bṛ-ta/ and */fra-mṝ-ta/, and the changes occurred due to the grammar’s intolerance for the phoneme */ṛ/.[10]

The alternation of PIIr */ṛ/ → */u/ is also evidenced by comparing the Sanskrit word for back, /pṛʃṭha/ and the Modern Persian “پشت” /puʃt/. This shift is overtly natural as speech pathologists will attest that speakers unable to articulate /ɹ/ will almost invariably render it /ʊ/. In fact, all of Modern British ‘received pronunciation’ renders the American English /bɹd/ as /bʊd/ in precisely the fashion mentioned. The ‘naturalness’ of this shift cries out for OT’s sort of universal constraints based on ease of articulation and parsing.

[10] This is an extension of the suggestion made in Darmesteter (1883, p. 102).

II.B.2 Old Persian to Middle Persian: Consonants

The Old Persian to Middle Persian shift saw little wholesale change in the vowel system, and this will be seen not to be the case with the consonants.

II.B.2.a /p/, /t/, /ʧ/, /k/

/p/ → /p/ / #__; __#
/p/ → /b/ / V__V
/p/ → /f/ / __ /t/

OP paça	Pah.	/paç/	‘before’
OP āp	Pah.	/āp/	‘water’
OP rapati	Pah.	*/rabad/	‘goes’
OP hapta	Pah.	/haft/	‘seven’

/t/ → /t/ / #__; __#
/t/ → /ð/ / V__V

OP tanu	Pah.	/ta-nu/	‘body’
OP raut(a)	Pah.	/rōt/	‘river’
OP υi-tar	Pah.	/υi-ðā-ran/	‘to pass’
OP pati-kāra	Pah.	/pað-kār/	‘battle’

/ʧ/ → /ʧ/ / #__
/ʧ/ → /ʤ/ / V__V; __#

OP ʧiθra	Pah.	/ʧihar/	‘face’
OP rautʃah	Pah.	/rōʤ/	‘day’

/k/ → /k/ / #__

/k/ → /g/ / __#; V__V; C[+voice] __ V

OP kaufa	Pah.	/kōf/	‘mountain’
OP υehrka	Pah.	/gurg/	‘wolf’

<u>II.B.2.b /b/, /d/ (/ð/), /ʤ/, /g/ (/ɣ/)</u>

/b/ → /b/ / #__

/b/ → /f/ / V__ V

OP bandaka	Pah.	/ban-dak/	‘slave’
OP nābi	Pah.	/nā-fak/	‘navel’

/d/ → /d/

OP dūra	Pah.	/dūr/	‘far’

/ð/ → /j/ / V__V

/ð/ → /ð/ / elsewhere

OP gauð	Pah.	/gūz/	‘hide; cover’
OP pāða	Pah.	/pāj/	‘foot’

/ʤ/ → /z/ / #__

/ʤ/ → /ʤ/ / elsewhere

OP ʤīυ	Pah.	/zī-υa-stan/	‘to live’
OP taiʤa	Pah.	/tēʤ/	‘fast’

/g/ → /g/ / #__

/ɣ/ → /ɣ/ / V__ V

OP gauʃa	Pah.	/ga-ʃa/	‘ear’
OP drauga	Pah.	/drōɣ/	‘lie’
OP mereɣa ‘bird’	Pah.	/miriɣ/	‘chicken’

<u>II.B.2.c /f/, /θ/ (/ç/[11]), /x/</u>

/f/ → /f/

OP framāna	Pah.	/far-mān/	‘order’
OP kaufa	Pah.	/kōf/	‘mountain’

/ç/; /θ/ → /ç/ / elsewhere

→ /h/ / V__ V; V __ [r]

OP θuxra	Pah.	/çuxr/	‘red’
OP raç-	Pah.	/ra-çī-tan/	‘to arrive’
OP maçja	Pah.	/mā-hīk/	‘fish’

Note that /θ/ as borrowed from Arabic also shifts to /s/ in MP and PP as seen with “ثلث” from Arabic /θʊlθ/ → /suls/ ‘third (fraction).’

/x/ → /x/

→ ∅ / __ [ʃ]

[11] /ç/ and /θ/ shall be treated together as the former is merely /θr/ with elided /r/.

OP xãnja	Pah.	/xān/	‘source; noble man’
OP xʃīra	Pah.	/ʃīr/	‘milk’

<u>II.B.2.d /s/, /z/, /ʃ/, /h/</u>

/z/ → /z/

OP zem-	Pah.	/zem-/	‘earth’

/ʃ/ → /ʃ/

OP ʃajāti	Pah.	/ʃā-tīh/	‘joy’
OP gauʃa	Pah.	/gōʃ/	‘ear’

/h/ → /h/ / #__

/h/ → ∅ / #__

/h/ → /x/ / #__ u; ʊ[12]

OP haruʊa	Pah.	/har/	‘all; each’
OP handāma	Pah.	/an-dām/	‘limb; organ’
OP hʊap	Pah.	/xuf-tan/	‘sleep’

<u>II.B.2.e /m/, /n/</u>

/m/ → /m/

OP madhu	Pah.	/maj/	‘wine’
OP kāma	Pah.	/kām/	‘desire’

[12] In cases of /hʊ/ → /x/, the /ʊ/ often remains in orthographic convention:
“خواندن” */xʊɑn-dan/ → /xɑn-dan/ ‘to read’
“خویش” */xʊēʃ/ → /xēʃ/ ‘oneself’

/n/ → /n/

OP naʋa	Pah.	/nōk/	‘new’
OP framāna	Pah.	/far-mān/	‘order’

<u>II.B.2.f /j/, /r/, /l/, /ʋ/</u>

/r/ → /r/

OP rautʃah	Pah.	/rōz/	‘day’
OP ʋāra	Pah.	/ʋā-rān/	‘rain’

/j/ → /ʤ/ / #__

OP jātu	Pah.	/ʤā-tūk/	‘witch’

This is not surprising, and even happens in modern Puerto Rican Spanish: “lleno” /jeˈ-no/ → /ʤeˈ-no/ ‘full.’

/ʋ/ → /ʋ/

/ʋ/ → /g/ / #__ (sometimes)

OP ʋafra	Pah.	/ʋafr/	‘snow’
OP ʋīra	Pah.	/ʋīr/	‘male’
OP ʋehrka	Pah.	/gurg/	‘wolf’

II.C From Middle Persian to Proto-Modern-Persian

At this point in the analysis, the Middle Persian phonemic inventory has been established as such:

OP:		Pah:
/p/ /t/ /ʧ/ /k/	→	/p/ /t/ /ʧ/ /k/
/b/ /d/ /ʤ/ /g/	→	/b/ /d/ /ʤ/ /g/
/f/ /θ/ /x/	→	/f/ /x/
/ð/ /ɣ/	→	/ð/ /ɣ/
/ʃ/ /ç/ /s/ /h/ /z/	→	/ʃ/ /ç/ /s/ /h/ /z/
/m/ /n/	→	/m/ /n/
/j/ /r/ /l/ /ʋ/	→	/j/ /r/ /l/ /ʋ/
/ī/ /i/	→	/ī/ /i/
/ū/ /u/	→	/ū/ /u/
/ā/ /a/	→	/ā/ /a/
/ai/	→	/ē/
/au/	→	/ō/

II.C.1 From Middle Persian to Proto-Modern-Persian: Vowels

The Middle Persian vocalic phonemic inventory is thus established as an eight vowel system:

/ī/ /i/ /ū/ /u/

/ē/ /ō/

/ā/ /a/

The phonologies of Dari, Iranian and Tajik, to be fully analyzed in Chapters III through V of this dissertation, are as follows:

Dari:	Iranian:	Tajik:
/i/ /u/	/i/ /u/	/i/ /ʉ/ /u/
/ʊ/		/ʊ/
/e/ /o/	/ɛ/ /o/	/e/ /o/
/ɛ/		
/a/ /ɑ/	/æ/ /ɑ/	/a/

Drawing on examples from the above section, the following eight lexical items contain all of the Middle Persian vocalic inventory:

Pahlavi:	Modern Persian:	Dari:	Iranian:	Tajik:	Gloss:
/haft/	“هفت” / “хафт”	/aft/	/hæft/	/haft/	‘seven’
/ʊā-rān/	“باران” / “борон”	/bɑ-rɑn/	/bɑ-rɑn/	/bo-ron/	‘rain’
/ʃīr/	“شیر” / “шир”	/ʃir/	/ʃir/	/ʃir/	‘milk’
/ʧihar/	“چهره” / “чехра”	/ʧɛ-ra/	/ʧɛh-rɛ/	/ʧi-ra/	‘face’
/dūr/	“دور” / “дур”	/dur/	/dur/	/dur/	‘far’
/çuxr/	“سرخ” / “сурх”	/sʊrx/	/sorx/	/sʊrx/	‘red’
/rōz/	“روز” / “руз”	/roz/	/ruz/	/rʉz/	‘day’
/tēʤ/	“تیز” / “тез”	/tɛz/	/tiz/	/tʲɛz/	‘fast’

With the exception of the Tajik high front vowels and the Iranian high back vowels, the Modern Persian high vowels realize the long/short distinction as a tense/lax distinction, and thus the reconstruction begins as such:

Pah:		Proto-Persian:	Modern Persian:	
/ī/	→	/i/	/i/	(Dari)
			/i/	(Iranian)
			/i/	(Tajik)
/i/	→	/ɪ/	/ɛ/	(Dari)
			/ɛ/	(Iranian)
			/i/	(Tajik)

/ū/ → /u/ → /u/ (Dari)
/u/ (Iranian)
/u/ (Tajik)

/u/ → /ʊ/ → /ʊ/ (Dari)
/o/ (Iranian)
/ʊ/ (Tajik)

With the exception of the Tajik low long vowel, the low vowels are all realized as /a/ and /ɑ/, and so the reconstruct continues:

Pah:		Proto-Persian:		Modern Persian:	
/ā/	→	/ɑ/	→	/ɑ/	(Dari)
				/ɑ/	(Iranian)
				/o/	(Tajik)
/a/	→	/a/	→	/a/	(Dari)
				/a/	(Iranian)
				/a/	(Tajik)

The final two MP vowels /ē/ and /ō/ will be reconstructed in Proto-Persian as /e/ and /o/, because in the former case, two of the three Modern Persian dialects manifest /e/, and the third dialect, Iranian, merges /e/ and /i/. In the latter case, Iranian again will merge /o/ with an existing vowel /u/, and the Tajik vowel /ʉ/ is a highly marked segment:

Pah:		Proto-Persian:	Modern Persian:	
/ē/	→	/e/	/e/	(Dari)
			/i/	(Iranian)
			/e/	(Tajik)
/ō/	→	/o/	/o/	(Dari)
			/u/	(Iranian)
			/ʉ/	(Tajik)

The resultant reconstructions of the eight lexical items above are as follows:

OP hapta	→	Pah.	/haft/	→	PP “هفت” /haft/ ‘seven’
OP υāra	→	Pah.	/υā-rān/	→	PP “باران” /bɑ-rɑn/ ‘rain’
OP xʃīra	→	Pah.	/ʃīr/	→	PP “شیر” /ʃir/ ‘milk’
OP ʧiθra	→	Pah.	/ʧihar/	→	PP “چهره” /ʧɪh-ra/ ‘face’
OP dūra	→	Pah.	/dūr/	→	PP “دور” /dur/ ‘far’
OP θuxra	→	Pah.	/çuxr/	→	PP “سرخ” /sʊrx/
OP rauʧah	→	Pah.	/rōz/	→	PP “روز” /roz/ ‘day’
OP taiʤa	→	Pah.	/tēʤ/	→	PP “تیز” /tez/ ‘fast’

The dipthongs of Middle Persian are either the result of consonant lenition to semivowel, /ð/ → /j/ or consonant + /υ/. Thus, the five diphthongs of Middle Persian are /aj/, /aw/, /āj/, /ōj/, and /āj/:

Pahlavi:	Modern Persian:	Dari:	Iranian:	Tajik:	Gloss:
Pah. /pāj/	“پای” / “пой”	/pɑj/	/pɑj/	/poj/	‘foot’
Pah. /maj/	“می” / “май”	/maj/	/meʲ/	/maj/	‘wine’
Pah. /gāʊ/	“گاو” / “гов”	/gɑw/	/gɑv/	/gov/	‘cow’
Pah. /ʤaʊ/	“جو” / “чав”	/ʤaw/	/ʤo/	/ʤav/	‘barley’
Pah. /mōj/	“موی” / “муй”	/moj/	/mu/	/mʉj/	‘hair’

Thus, we observe that Iranian levels the short diphthongs and elides the palatal glide in the mid back diphthong. Futhermore, this clearly demonstrates that Dari will realize the labial element as the labio-velar /w/, and Iranian and Tajik will realize /v/. Thus the below reconstructed diphthongal system reflects the vocalic reconstruction of Proto-Persian along with the reconstruction of the two semivocalic diphthonal constituents as /j/ and /ʊ/:

Pah:		Proto-Persian:	Modern Persian:	
/āj/	→	/ɑj/	/ɑj/	(Dari)
			/ɑj/	(Iranian)
			/oj/	(Tajik)
/aj/	→	/aj/	/aj/	(Dari)
			/e/	(Iranian)
			/aj/	(Tajik)
/āʊ/	→	/ɑʊ/	/ɑw/	(Dari)
			/ɑv/	(Iranian)
			/ov/	(Tajik)
/aʊ/	→	/aʊ/	/aw/	(Dari)
			/o/	(Iranian)
			/av/	(Tajik)

/ōj/ → /oj/ → /oj/ (Dari)
→ /u/ (Iranian)
→ /ʉj/ (Tajik)

The resulting vocalic and diphthongal reconstruction of Proto-Persian is as follows:

/i/			/u/	/ɑj/
/ɪ/			/ʊ/	/aj/
	/e/		/o/	/ɑʊ/
				/aʊ/
		/a/	/ɑ/	/oj/

II.C.2 From Middle Persian to Proto-Modern-Persian: Consonants

MP borrows the phonemes /q/ and /ʔ/ from Arabic, and those enter PP as set phonemes. The exact origin of /ʒ/ is unknown as none of the few lexical items containing this segment are attested in earlier literature. The rest of the consonantal inventory remains the same between Pahlavi and all of the Modern Persian dialects save two shifts: /ç/ → /s/ and /ð/ → /z/. As established above, the consonant inventory of Middle Persian is as follows:

/p/ /t/ /tʃ/ /k/
/b/ /d/ /dʒ/ /g/
/f/ /x/
/ð/ /ɣ/
/ʃ/ /ç/ /s/ /h/ /z/
/m/ /n/
/j/ /r/ /l/ /ʊ/

II.C.2.a /p/, /t/, /ʧ/, /k/

/p/ → /p/

OP çpaita	Pah. /spēd/	PP “سفید” /sa-fẽd/ ‘white’
OP paça	Pah. /paç/	PP “پس” /pas/ ‘before’
OP āp	Pah. /āp/	PP “آب” /āb/ ‘water’

/t/ → /d/ / __#; V__V

/t/ / elsewhere

OP tanu	Pah. /ta-nu/	PP “تن” /tan/ ‘body’
OP raut(a)	Pah. /rōt/	PP “رود” /rōd/ ‘river’
OP martija	Pah. /mart/	PP “مرد” /mard/ ‘man’

/ʧ/ → /ʧ/

OP ʧiθra	Pah. /ʧihar/	PP “چهره” /ʧih-ra/ ‘face’

/k/ → /k/

OP kadha	Pah. /kaj/	PP “کی” /kaj/ ‘when’
OP kaufa	Pah. /kōf/	PP “کوه” /kōh/ ‘mountain’

II.C.2.b /b/ /d/ /ʤ/ /g/

/b/ → /ʋ/ / V__V

/b/ / elsewhere

OP bandaka	Pah. /ban-dak/	PP “بنده” /ban-da/ ‘slave’
OP rapati	Pah. */rabad/	PP “رود” /ra-ʋad/ ‘goes’

/d/ → /d/

OP dāna	Pah. /dā-nak/	PP “دانه” /dā-na/ ‘seed’
OP ʃātō	Pah. /ʃādīh/	PP “شاد” /ʃād/ ‘joy
OP mātar	Pah. /mādar/	PP “مادر” /mā-dar/ ‘mother’
OP çpaita	Pah. /spēd/	PP “سفید” /sa-fẽd/ ‘white’
OP dāta	Pah. dād	PP “داد” /dād/ ‘law’

/ʤ/ → /z/ / __#

/ʤ/ / elsewhere

OP ʤātu	Pah. /ʤādūg/	PP “جادو” /ʤā-dū/ ‘witch’
OP rautʃah	Pah. /rōʤ/	PP “روز” /rōz/ ‘day’
OP taiʤa	Pah. /tēʤ/	PP “تیز” /tēz/ ‘fast’

/g/ → /g/

OP gāθa	Pah. /gāç/	PP “گاه” /gāh/ ‘place’
OP ʋehrka	Pah. /gurg/	PP “گرگ” /gurg/ ‘wolf’

II.C.2.c /f/ /x/

/f/ → /h/ / __#

→ /f/ / elsewhere

OP farzaiñti	Pah. /frazand/	PP “فرزند” /far-zand/ ‘son’
OP hapta	Pah. /haft/	PP “هفت” /haft/ ‘seven’
OP nābi	Pah. /nā-fak/	PP “ناف” /nāf/ ‘navel’
OP kaufa	Pah. /kōf/	PP “کوه” /kōh/ ‘mountain’

/x/ → /x/

OP xãnja	Pah. * /xān/	PP “خان” /xān/ ‘source; noble man’
OP xri-	Pah. /xrīdan/	PP “خریدن” /xa-rī-dan/ ‘to buy’
OP çuxta	Pah. /çux-ta/	PP “سوخته” /sux-ta/ ‘burned’

II.C.2.d /ð/ /ɣ/

/ð/ → /j/ / __#; __Stop[-voice]

→ /z/ / elsewhere

OP υi-tar	Pah. /υi-ðā-ran/	PP “گذاردن” /gu-zār-dan/ ‘to pass’
OP pati-kāra	Pah. /pað-kār/	PP “پیکار” /paj-kār/ ‘battle’

/ɣ/ → /ɣ/

OP rauɣna	Pah. /rōɣn/	PP “روغن” /rō-ɣan/ ‘oil’
OP drauga	Pah. */drōɣ/	PP “دروغ” /drōɣ/ ‘lie’

<u>II.C.2.e /ʃ/ /ç/ /s/ /h/ /z/</u>

/ʃ/ → /ʃ/

OP ʃajāti	Pah. /ʃā-tīh/	PP “شادی” /ʃā-dī/ ‘joy’
OP xʃīra	Pah. /ʃīr/	PP “شیر” /ʃīr/ ‘milk’
OP ʃātō	Pah. /ʃādīh/	PP “شاد” /ʃād/ ‘joy
OP gauʃa	Pah. /gōʃ/	PP “گوش” /gōʃ/ ‘ear’

/ç/ → /s/

OP çuxta	Pah. /çux-ta/	PP “سوخته” /sux-ta/ ‘burned’
OP çpaita	Pah. /çpēd/	PP “سفید” /sa-fēd/ ‘white’
OP raç-	Pah. /ra-çī-tan/	PP “رسیدن” /ra-sī-dan/ ‘to arrive’
OP paça	Pah. /paç/	PP “پس” /pas/ ‘before’

/h/ → /h/

OP hapta	Pah. /haft/	PP “هفت” /haft/ ‘seven’
OP ʧiθra	Pah. /ʧihar/	PP “چهره” /ʧih-ra/ ‘face’
OP maçja	Pah. /mā-hīk/	PP “ماهی” /mā-hī/ ‘fish’

/z/ → /z/

OP zānu	Pah. /zānūg/	PP “زانو” /zā-nū/ ‘knee’
OP zem-	Pah. /zem-/	PP “زم” /zim-/ ‘earth’
OP farzaiñti	Pah. /frazand/	PP “فرزند” /far-zand/ ‘son’
OP ʤīʋ	Pah. /zī-ʋa-stan/	PP “زیستن” /zī-stan/ ‘to live’

II.C.2.f /m/ /n/

/m/ → /m/

OP mātar	Pah. /mādar/	PP “مادر” /mā-dar/ ‘mother’
OP maçja	Pah. /mā-hīk/	PP “ماهی” /mā-hī/ ‘fish’
OP martija	Pah. /mart/	PP “مرد” /mard/ ‘man’
OP zem-	Pah. /zem-/	PP “زم” /zim-/ ‘earth’

/n/ → /n/

OP rauɣna	Pah. /rōɣn/	PP “روغن” /rō-ɣan/ ‘oil’
OP zānu	Pah. /zānūg/	PP “زانو” /zā-nū/ ‘knee’
OP raiçaj-	Pah. /rēxtan/	PP “ریختن” /rēx-tan/ ‘to pour’
OP dāna	Pah. /dā-nak/	PP “دانه” /dā-na/ ‘seed’
OP xãnja	Pah. */xān/	PP “خان” /xān/ ‘source; noble man’

II.C.2.g /j/ /r/ /l/ /υ/

/j/ → /j/

OP pāða	Pah. /pāj/	PP “پای” /pāj/ ‘foot’

/r/ → /r/

OP rauɣna	Pah. /rōɣn/	PP “روغن” /rō-ɣan/ ‘oil’
OP raiçaj-	Pah. /rēxtan/	PP “ریختن” /rēx-tan/ ‘to pour’
OP martija	Pah. /mart/	PP “مرد” /mard/ ‘man’
OP xʃīra	Pah. /ʃīr/	PP “شیر” /ʃīr/ ‘milk’
OP dūra	Pah. /dūr/	PP “دور” /dūr/ ‘far’

/l/ → /l/

?	Pah./larzīdan/	PP “لرزیدن” /lar-zi-dan/ ‘tremble’

?	Pah. /lab/	PP “لب” /lab/ ‘lip’
?	Pah. /bālāi/	PP “بالا” /bɑlɑ/ ‘up; tall’

/ʋ/

OP ʋāta	Pah. /ʋāt/	PP “باد” /bād/ ‘wind’
OP ʋāra	Pah. /ʋā-rān/	PP “باران” /bā-rān/ ‘rain’
OP ʋafra	Pah. /ʋafr/	PP “برف” /barf/ ‘snow’
?	Pah. /parʋast/	PP “پروست” /par-ʋast/ ‘closed’
?	Pah. /gāʋ/	PP “گاو” /gɑʋ/ ‘cow’

So the Proto-Persian consonant inventory is as follows:

/p/ /t/ /ʧ/ /k/
/b/ /d/ /ʤ/ /g/
/f/ /x/
/ɣ/
/ʃ/ /s/ /h/ /z/
/m/ /n/
/j/ /r/ /l/ /ʋ/

III. From Proto-Persian to Dari

III.A Dari Phonology: Introduction

The Dari phonemic inventory is the most conservative of the three dialects under consideration, and consequently it displays the largest and most distinctive inventory, retaining distinct segments that have become homophonous in the other two dialects. It is interesting to note however that Dari has the greatest degree of allophonic and synchronic variation of the three dialects. Thus, assuming that it is in fact the most conservative variety, it is a logical starting point for description.

Section III.B will introduce the vowel system of Dari, noting allophonic and free variation and justifying the inventory as proposed. Sections III.C.1 – III.C.6 will treat the individual vowels and diphthongs providing spectrogram analyses and formant frequency contours for each of the vowels. Section III.D will incorporate all of the data from III.C as well as multiple other vowel samplings into two charts mapping the first and second formant frequencies for vowels and diphthongs. Section III.E will introduce consonants, III.E.1 - III.E.12 will cover consonant phonology as particular to Dari, but also providing comparisons with the other dialects, and III.E.13 will cover a few other phenomena.

III.B Dari Vocalic Phonemic Inventory

	Front	Central	Back
High:	/i/		/u/ /ʊ/
Mid:	/e/ /ɛ/		/o/
Low:		/a/	/ɑ/

Diphthongs: /aj/, /aw/, /ɑj/, /ɑw/ and /uj/.

Two phones [ə] and [ɪ] must also be added to the inventory as variants, [ə] as a variation of both /a/ and /ɛ/, and [ɪ] also as a variation of /ɛ/. The surface representation of /ɛ/ as [ɪ] or [ɛ] is not phonologically conditioned and occurs both as a result of dialectal variation and randomly within particular intradialectal varieties and ideolects. The allophonic variation of /a/ as [a] or [ə] appears to be conditioned by stress, [a] being the only phonological realization of stressed /a/, and [ə] occurring in random alternation with [a] only in unstressed positions. The /ɛ/ to [ə] alternation is only found in a few instances. Thus lexical items such as "هندوستان" [hɛn-du-stɑnˈ] 'India', "سند" [sa-nadˈ] 'document' and "فکر" [fɛˈ-kar] 'thought' will phonetically realize alternate forms as [hɪn-du-stɑnˈ], [sə-nadˈ], and either [fɪˈ-kar] or [fəˈ-kar] respectively with no semantic effect.

It is assumed with respect to data from Middle Persian (covered in section II) that there are three sets of historical long/short vocalic pairs as well as two mid long vowels without short counterparts remaining in Dari:

MP:		Dari:
/ī/ - /i/	→	/i/ - /ɛ/
/ū/ - /u/	→	/u/ - /ʊ/
/ā/ - /a/	→	/ɑ/ - /a/
/ē/	→	/e/
/ō/	→	/o/

This will be explored further in Section VI within an Optimality Theoretic framework, but for now, suffice it to say that the historical long/short distinction has been replaced with a qualitative rather than quantitative distinction. That is, the below set of minimal pairs illustrates that the distinction is height and ±ATR:

/i/	"سیر"	[sir]	'garlic'
/ɛ/	"سر"	[sɛr]	'secret'
/e/	"سیر"	[ser]	'satiated'
/u/	"پول"	[pul]	'money'
/ʊ/	"پل"	[pʊl]	'bridge'

/u/	“طول”	[tul]	‘length’
/o/	“تول”	[tol]	‘to weigh’
/a/	“خر”	[xar]	‘donkey’
/ɑ/	“خار”	[xɑr]	‘thorn’

The realization of “long” /a/ as /ɑ/ and “short” /i/ as /ɛ/ must also be justified as these are distinctive in position as well as ATR: the former in backness and the latter height. For both, I will rely on Arabic loanwords for a justification, and for the former I will also cite post-ellipsis effects. First, both the Arabic front low and front high vowels are distinctive for quantitative length, as exemplified by the following minimal pairs in Arabic:

“عَدّ”	[ʕædd]	‘counted’	“عِد”	[ʕid]	‘promise!’
“عادّ”	[ʕæːdd]	‘came back’	“عيد”	[ʕiːd]	‘celebration’

Each instantiation of Arabic [æː] is realized in Dari as [ɑ] as is each instantiation of [æ] realized as [a]. Additionally, Arabic short front high vowels are realized as [ɛ]:

“كِتاب”	[ki-tæːbˈ] → [kɛ-tɑbˈ]	‘book’	“صليب”	[sˤæ-liːb] → [sa-lib]	‘cross’
“حَقّ”	[ħæq] → [haq]	‘right’	“فكر”	[fi-kær] → [fɛ-kar]	‘think’

Second, Dari (to the exclusion of the other Modern Persian dialects) converts /a/ + /a/ constructions to [ɑ] in cases of intervocalic ellipsis, indicating that the phonology of Dari interprets the "long" reflux of /a/ as [ɑ]:

"بعد" /baʔd/ → *[baːd] → [bɑd] 'after'[13]

"شهر" /ʃahr/ → *[ʃaːr] → [ʃɑr] 'city'

III.C Dari Vowels Illustrated

For each vowel, I will provide examples of the vowel's occurrence in initial, medial and final position where available in the lexicon. Then I will choose another lexical item from which to derive a spectrogram analysis of each vowel followed by a chart of the vowel's first four formant frequencies. For cases in which the vowel exhibits a variety of formant frequency permutations (as with the aforementioned [ɪ] / [ɛ] alternation in /ɛ/), I will provide multiple examples illustrating the variance.

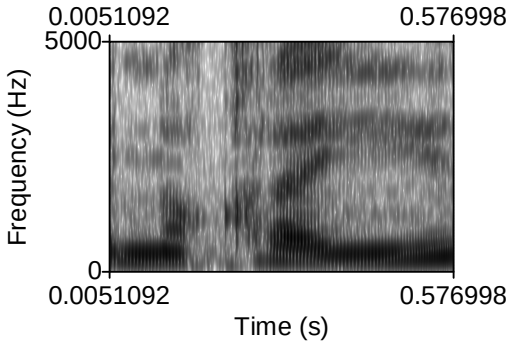
III.C.1 High Front Vowel: [i]

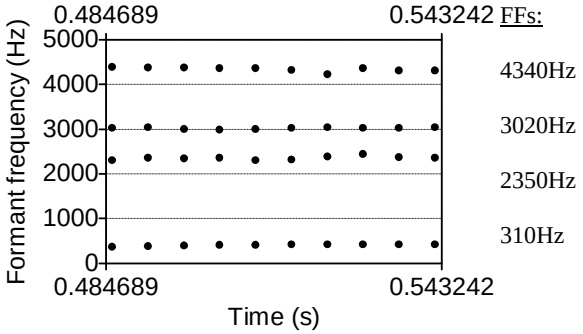
The Dari high front +ATR vowel is stable with a first formant frequency varying very little (300 to 390 Hz) and a second formant frequency varying widely (2050 to 2720 Hz), indicating that the height of the vowel is predicable but its frontness varies.

[i] "این" [in] 'this' "شیر" [ʃir] 'milk' "فنی" [fa-niˈ] 'scientific'

[13] It bears noting that this alternation occurs almost without exception in colloquial speech despite the fact that the resultant [bɑd] is homophonous with "باد" [bɑd] meaning 'wind.'

- example: “نقرئی” [nʊ-qra-jiˈ] ‘silver’:



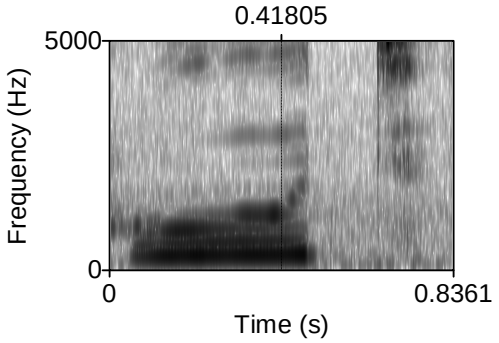


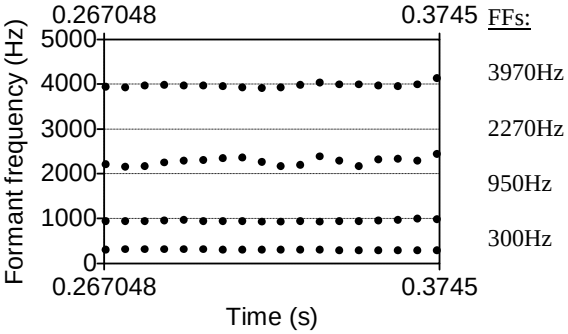
III.C.2 High Back Vowels: [u], [ʊ]

The high back +ATR vowel is stable and less varied than its front counterpart with a first formant ranging from 290 to 430 Hz and the second from 770 to 1070 Hz.

[u] “وجود” [u-ʤudˈ] ‘existance’ “بوت” [but] ‘shoe’ “پتو” [pa-tuˈ] ‘blanket’

- example: “حوت” [hut] ‘Pisces’:



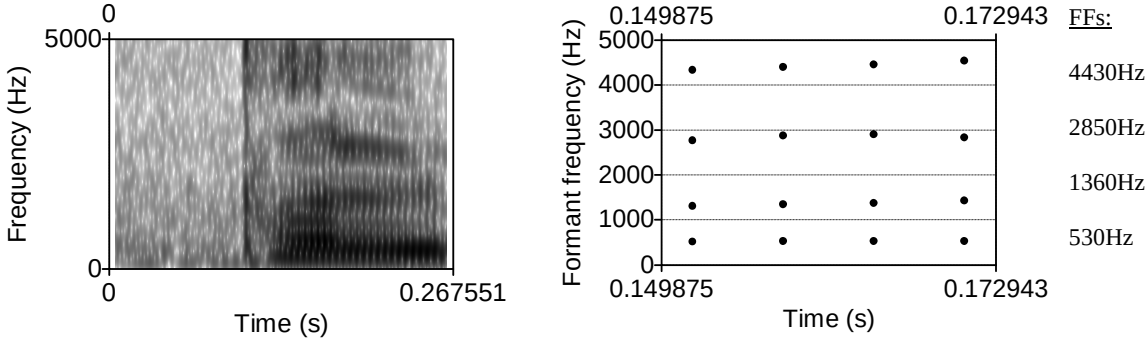


The high back –ATR vowel [ʊ] is more variable displaying a height from a first formant frequency of 450 to 650 Hz and a backness of 1000 to 1400 Hz:

[ʊ] “عهده” [ʊh-daˈ] ‘responsibility’ “پت” [pʊt] ‘hidden’ n/a[14]

[14] The –Low –ATR vowels cannot occur word-finally in Dari.

- example: “پل” [pʊl] ‘bridge’:

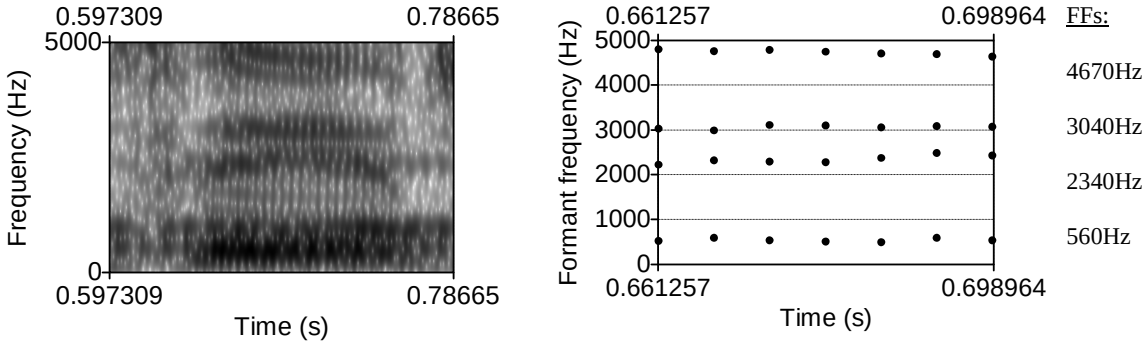


III.C.3 Mid Front Vowels: [e], [ɛ], ([ɪ]), ([ə])

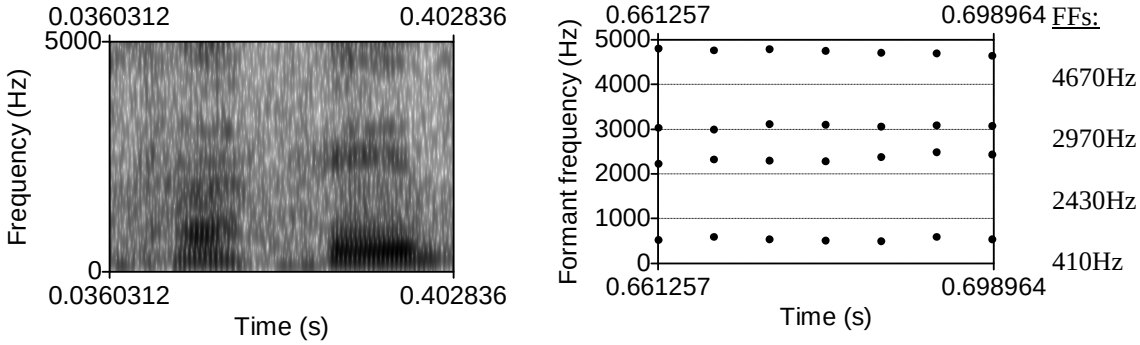
The mid front +ATR Vowel [e] is fairly stable with a first formant frequency ranging from 370 to 520 Hz and a second formant frequency from 2120 to 2500 Hz, placing it exactly between [i] and [ɛ] in height.

[e]	“عيد” [ejd] ‘celebration’	“شير” [ʃer] ‘lion’	“مانع” [mɑ-neˈ] ‘prohibited’

- example: “سير” [ser] ‘satiated’:



- example: “سفید” [sa-fedˈ] ‘white’:



The Dari mid front –ATR vowel varies widely in position, ranging in height from a first formant frequency of 450 to 670 Hz and in frontness from a second formant frequency of 1620 to 2650 Hz.

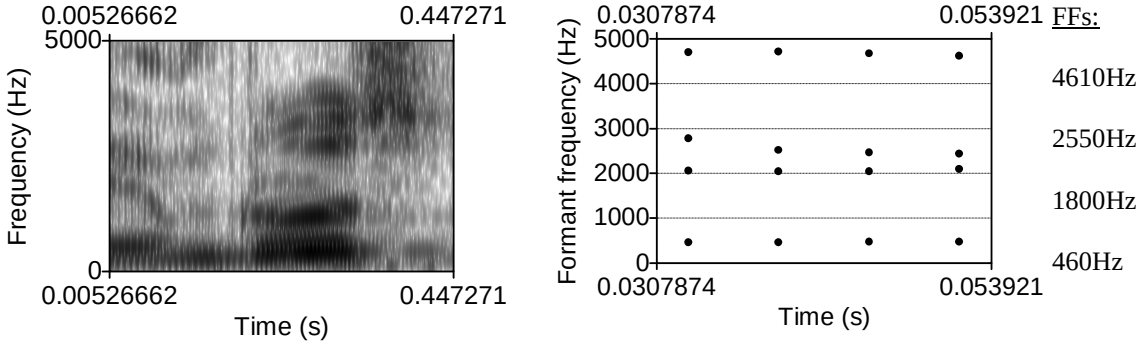
[ɛ] “اقبال” [ɛq-bɑlˈ] ‘luck’ “سن” [sɛn] ‘age’ n/a[15]

In a judgment test using English lexemes, subjects (all with a non-native command of English) were asked to identify which lexical item was produced, ‘hid’ or ‘head,’ and the subjects displayed an inability to confidently determine, indicating that Afghan Persians found difficulty in differentiating [ɪ] from [ɛ]:

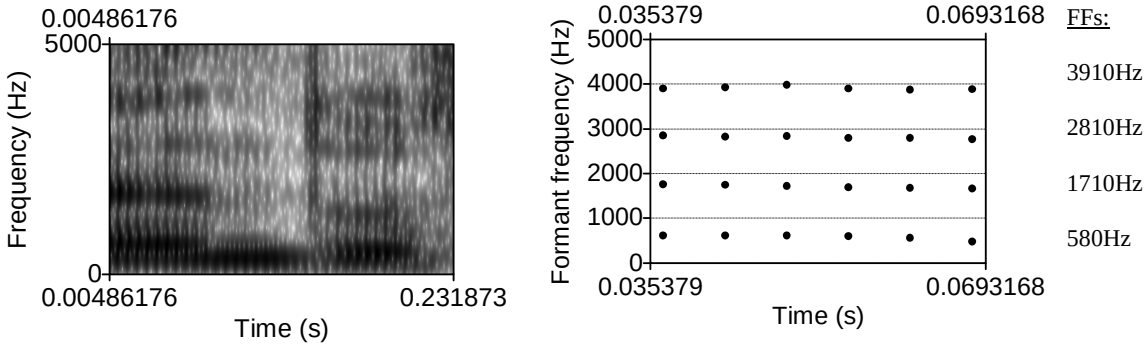
[ɛ], which is the normative realization of /ɛ/ can be seen below:

[15] This section will treat [ɪ] and [ɛ] as unpredictable allophonic variants of /ɛ/, but certain examples inexplicably tend toward [ɛ], such as “امروز” [ɛmˈ-roz] ‘today’ and “ادعا” [ɛd-ʔɔˈ] ‘claim,’ while others tend towards [ɪ], such as “سن” [sɪn] ‘today’ and “اقبال” [ɪq-bɑlˈ] ‘luck.’

- example: “امروز” [εmˈ-roz] ‘today’:



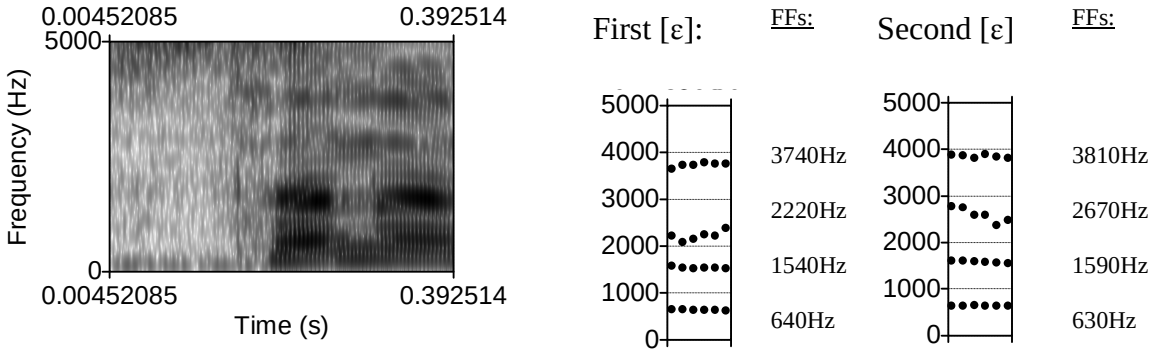
- example: “هندوستان” [hεn-du-stɑnˈ] ‘India’:



A morphosyntactic particle [ɛ] (called “ezafe”) serves as an enclitic epended postpositionally to nouns or noun phrases followed by a modifier. This particle shows a similar variation in frontness:

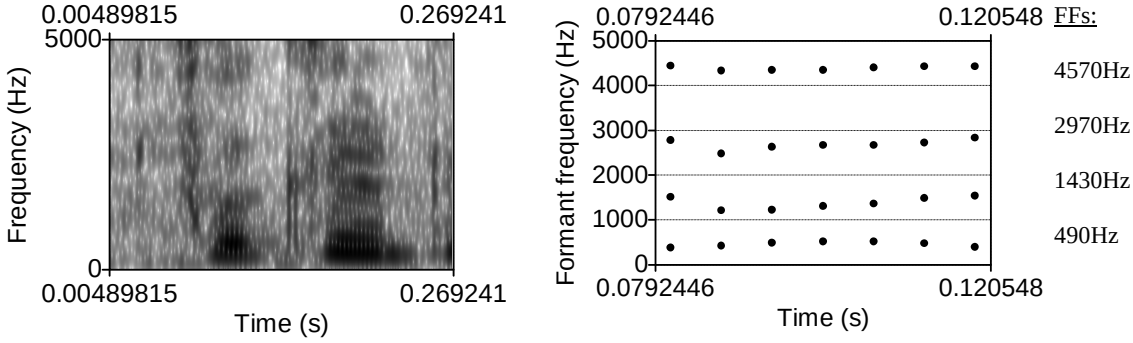
“دوست من”	/dost#ɛ#man/ → [dos-tɛ#man] ‘friend-e + I’	‘my friend’
“دوست من”	/dost#ɛ#xob/ → [dos-tɛ#xob] ‘friend-e + good’	‘good friend’

- example: “سن”[sɛˈ-n#ɛ] ‘age of’:

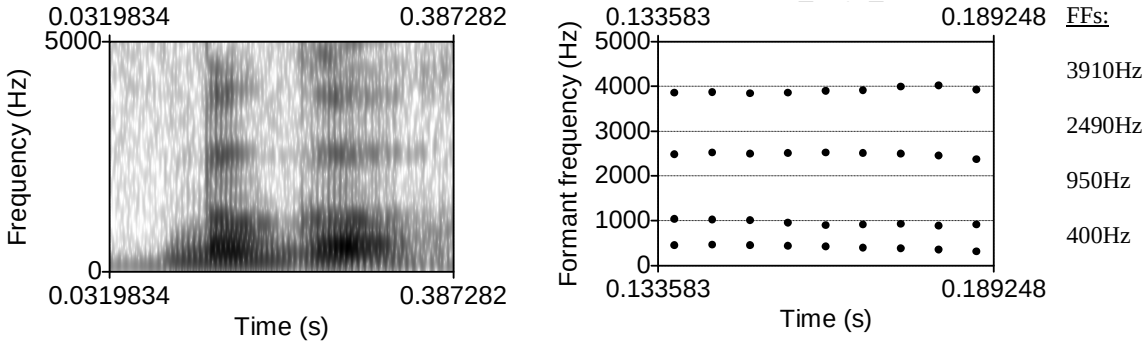


In a few cases, the phoneme /ɛ/ becomes [ə], but only in colloquial speech and not predictably. The lexical item “فکر” is [fɛˈkar] ‘thought’ becomes [fɛkr̩] or further reduced as [fɛk] in Iranian Persian, and sometimes [fəˈ-kər] in Dari. The /ɛ/ - [ə] alternation is also observed in the subjunctive verbal prefix /bɛ-/ which is rendered [bə-] in spoken Dari.

- example: “فکر” [fə-kərˈ] ‘thought’:



- example: “بگو” [bəˈ-go] ‘say!’:



III.C.4 Mid Back Vowel: [o]

The mid back vowel appears to be in flux in Dari. In Iranian, the long mid back vowel /ō/ has fully converged with [u], whereas in Dari, the distinction remains:

Mod Pers:	Dari:	Iranian:	Gloss:
“طول”	/tul/	/tul/	‘length’
“توپ”	/top/	/tup/	‘ball; cannon’

However, monosyllabic Dari lexemes containing /o/ in the word-final position converge with [u] in colloquial Dari speech:

Mod Pers:	Formal Dari:	Colloquial Dari:	Gloss:
“رو”	/ro/	[ruj]	‘face’
“مو”	/mo/	[muj]	‘hair’
“خو”	/xo/	[xuj][16]	‘habit; nature’

[16] As a side note, and merely for the sake of interest, this lexical item is homophonous with a vulgar Russian token “хуй” [xuj], which is a pejorative term for ‘penis,’ and hence the lexical item has fallen out of use in Tajik due to years of diglossia in Tajikstan. The same will be true of the word “هم” [ham] meaning ‘also,’ which elides the initial [h] in both Iranian and Dari, but not in Tajik, as the resultant lexical item [am] is vulgar in Uzbek, bearing the meaning of the female counterpart to the abovementioned taboo token. Thus Tajik must in fact content with triglossia in its development.

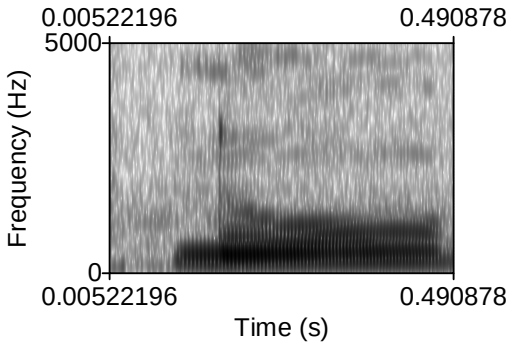
“بو”	/bo/	[buj]	‘smell’
“جو”	/ʤo/	[ʤuj]	‘stream’
“سو”	/so/	[suj]	‘direction’

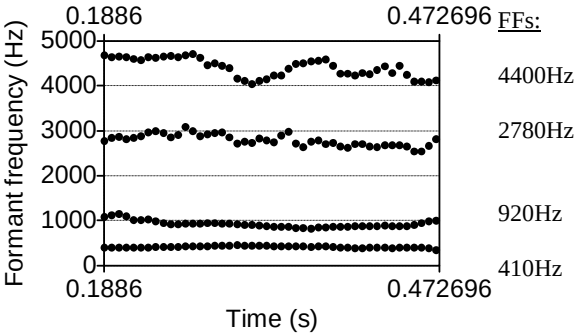
In formal use, this vowel is [o], but in common parlance the vowel converts to [u], and an epenthetic semivowel is added[17]. There is one exception to the abovementioned epenthesis, and that is found in the lexeme “او” [o] ‘he; she,’ which is simply rendered [u] in informal register. Additionally, it bears noting that lexical items “پتو” [pa-tuˈ] ‘blanket’ and “تو” [tu] ‘you-familiar’ do not take an epenthetic semivowel, and polysyllabic items such as “موضوع” [maw-zoˈ] neither take the epenthetic semivowel nor raise the vowel. This phenomenon will be further analyzed in section VI.

The vowel [o] is stable in all other instances with a first formant frequency ranging from 400 to 490 Hz and a second from 880 to 1300 Hz:

[o] “او” [o] ‘he; she’ “روز” [roz] ‘day’ “نرو!” [naˈ-ro] ‘don’t go!’

- example: “نه” [no] ‘nine’:





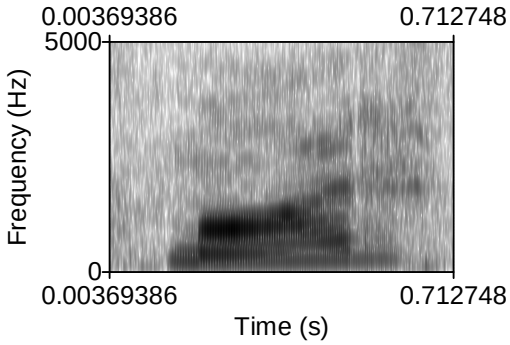
[17] Note that this also occurs with other monosyllabic, vowel-final tokens such as “جا” [ʤɑ] which becomes [ʤɑj] in spoken Dari.

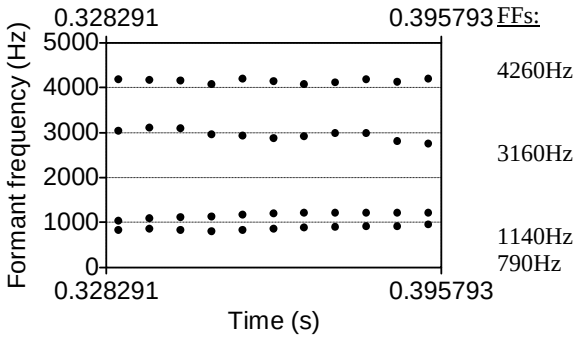
III.C.5 Low Vowels: [a], [ɑ], ([ɔ]) [18]

Because the “long” low vowel differs from its “short” counterpart in backness and ATR, it is more straightforward and obviously distinct. The long low vowel has a first formant frequency range of 450 to 950 Hz and a second formant frequency range of 1010 to 1400 Hz:

[ɑ] “عادی” [ɑ-diˈ] ‘principal’ “مال” [mɑl] ‘hidden’ “حتی” [ha-tɑˈ] ‘until’

- example: “مار” [mɑr] ‘snake’:



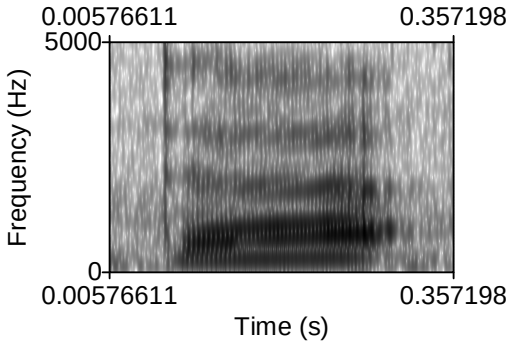


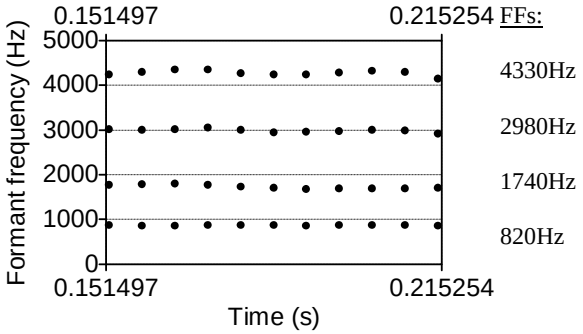
The low mid vowel [a] has a first formant frequency range from 570 to 890 Hz and a second formant frequency range from 1600 to 1900 Hz:

[a] “از” [az] ‘from’ “سر” [sar] ‘head’ “کرده” [kar-daˈ] ‘done’

[18] Note that in this instance, “low vowels” includes the mid vowel [ɔ] as an allophone of [ɑ] and as the “long” counterpart to [a].

- example: “ده” [da] ‘ten’:



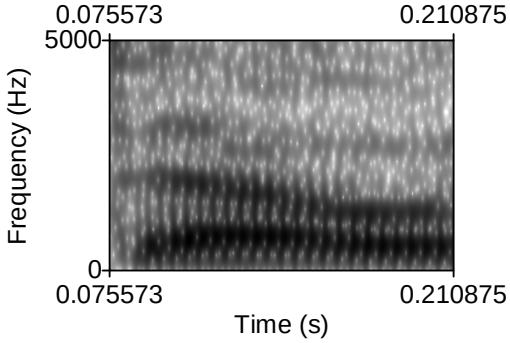


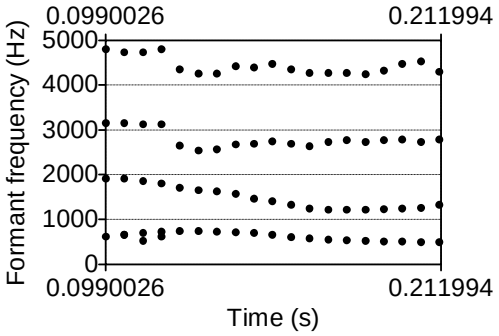
III.C.6 Diphthongs: [aw], [aj], [ɑj], [uj]

Dari is the only dialect of the three to have retained the diphthong /aw/, as this was typically converted to [av] in Tajik and [o] in Iranian. It is assumed to be original because of evidence from Arabic loanwords:

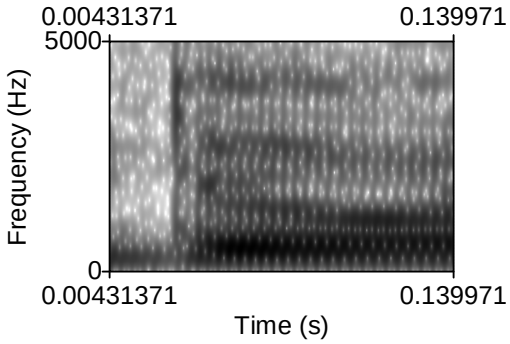
Arabic “دَوْلَت” [dæw-læt^{l}]: Dari: [daw-lat^{l}] Iranian: [do-læt^{l}] Tajik: [dav-lat^{l}]

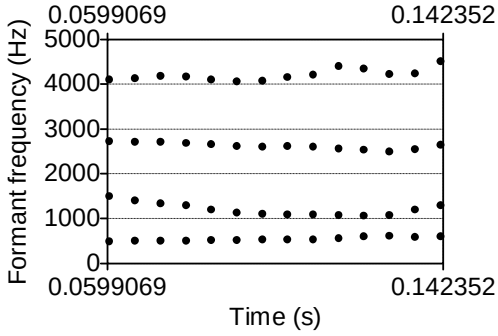
- example: [daw] from “دولتی” [daw-la-ti^{l}] ‘governmental’ Dari:



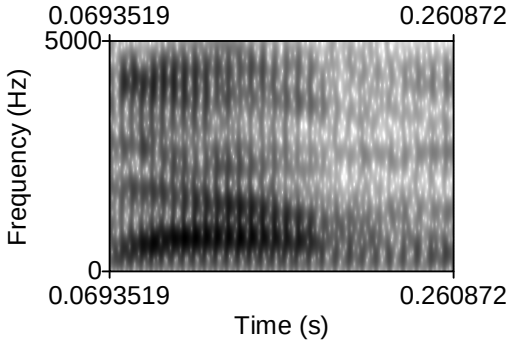


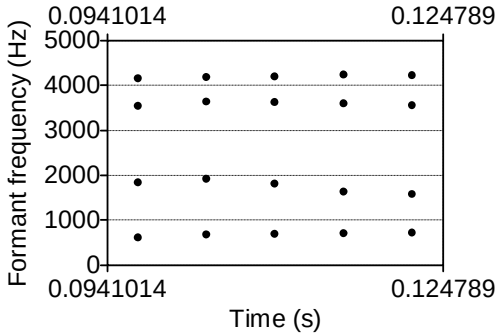
- example: [do] from “دولتی” [do-læ-tiˈ] ‘governmental’ in Iranian Persian:





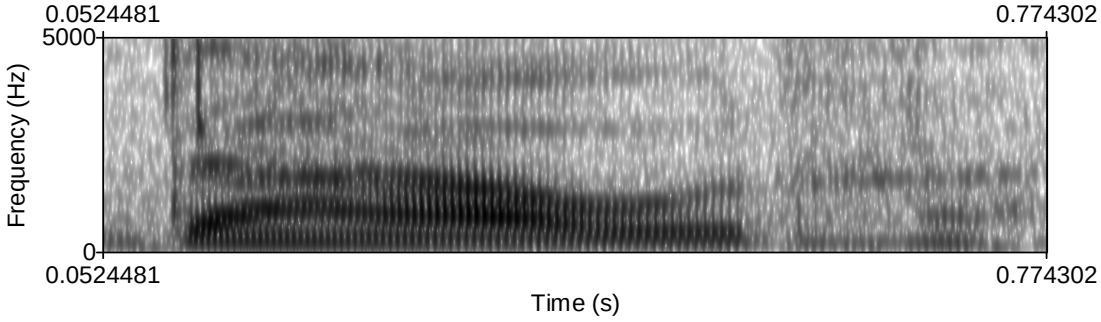
- example: [dav] from “دولتی” [dav-la-tiˈ] ‘governmental’ Tajik:

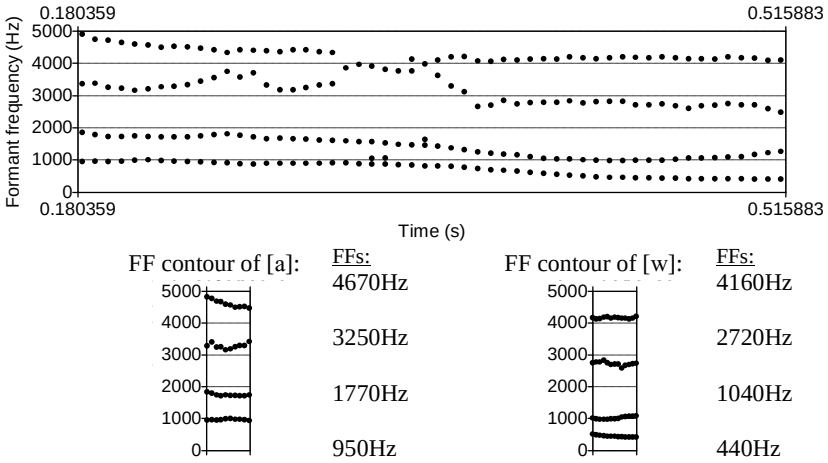




[aw] “اولاد” [aw-lɑdˈ] ‘children’ “فوج” [fawʤ] ‘army’ “نو” [naw] ‘now’

- example: [aw] from “دور” [dawr] ‘turn’:





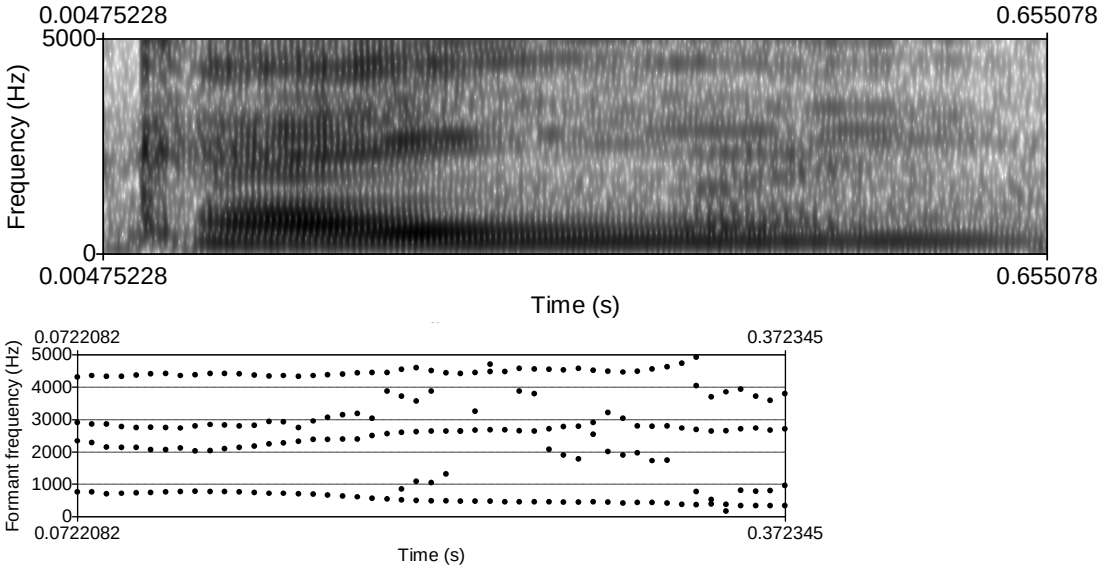
The first contour is nearly identical to the Dari [a], and the second to the Dari [u].

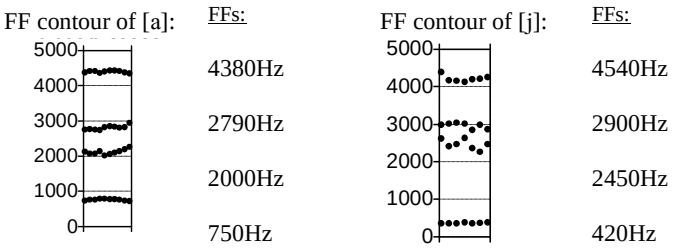
The diphthong /aj/ is preserved in Tajik but lost in Iranian as well:

Dari: [ɣajr] Iranian: [ɢer] Tajik: [ɣajr] ‘un-’

[aj] “ایماق” [aj-mɑqˈ] ‘now’ “قید” [qajd] ‘limit’ “پی” [paj] ‘gristle’

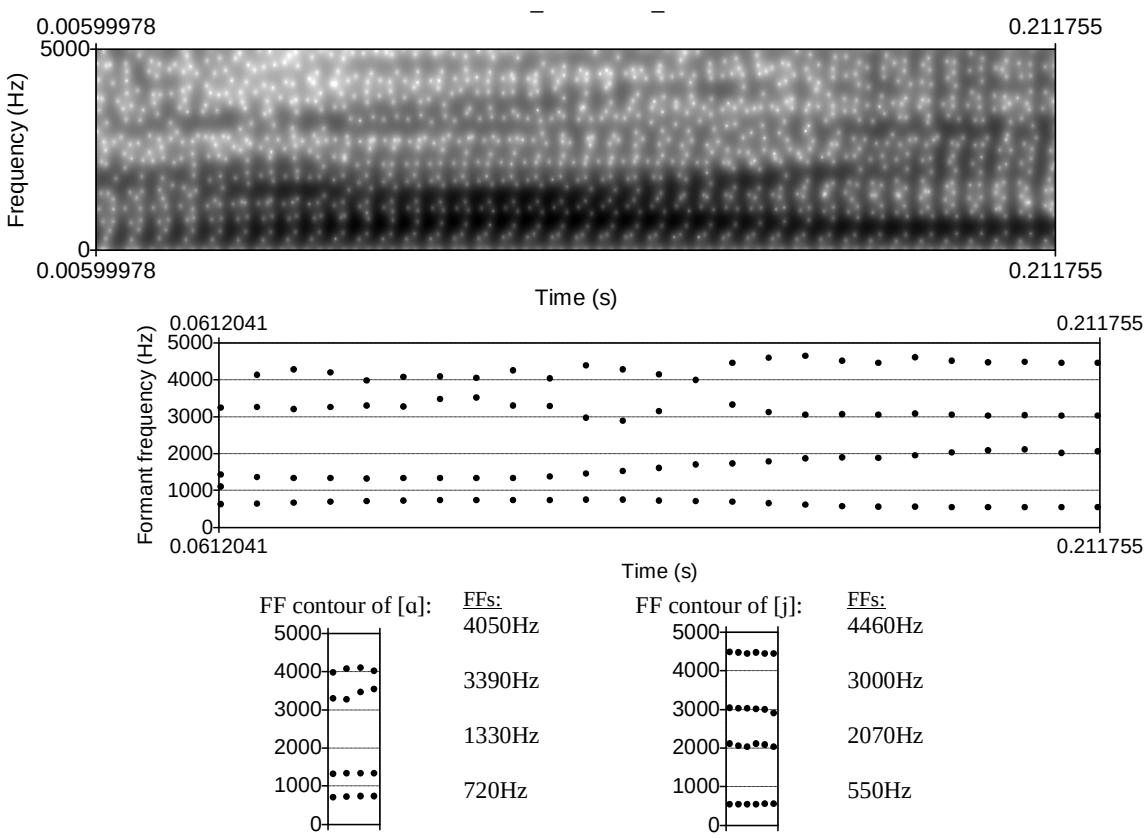
- example: [aj] from “کی” [kaj] ‘when’:





[ɑj] “آینه” [ɑj-naˈ] ‘mirror’ “موبایل” [mo-bɑjlˈ] ‘mobile’ “رآی” [rɑj] ‘vote’

- example: [ɑj] from “رأی” [rɑj] ‘vote’:



The fourth Dari diphthong is [ɑw], which in colloquial speech has converged with [aw]. This convergence is made evident by the remnant distinction in Iranian and Tajik:

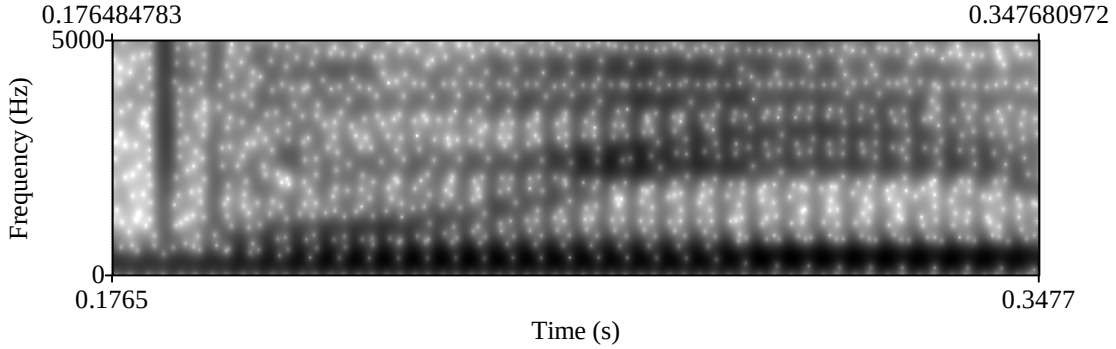
Mod Pers:	Dari:	Iranian:	Tajik:	Gloss:
“نو”	[naw]	[no]	[nav]	‘new’
“گاو”	[gaw]	[gɑv]	[gov]	‘cow’

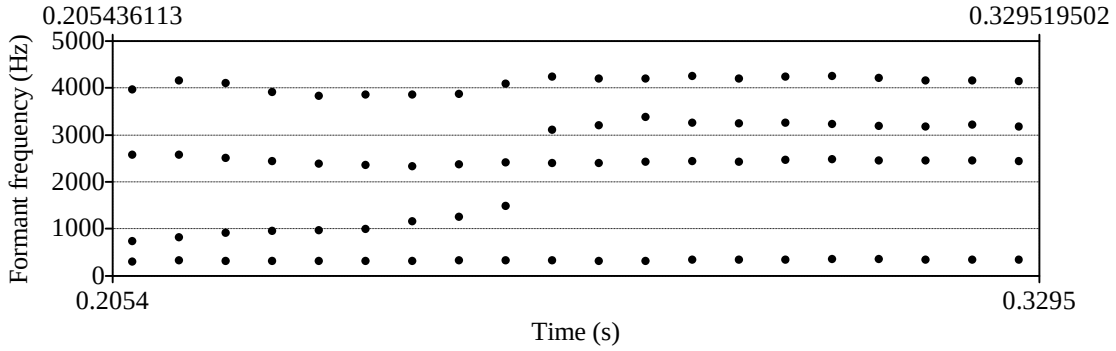
For a spectrogram analysis of this diphthong, please see the example for [aw] above, as there is no distinction in colloquial Dari.

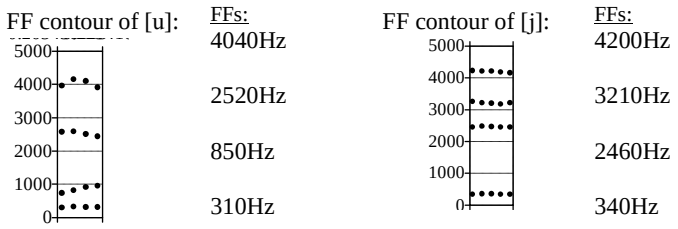
The fifth and final diphthong of Dari is [uj] that occurs as a variant of the word-final segment /o/:

[uj] n/a “سویقسد” [suj-qasdˠ] ‘evil intent’ “بوی” [buj] ‘smell; scent’

- example: [uj] from “بوی” [buj] ‘smell’:

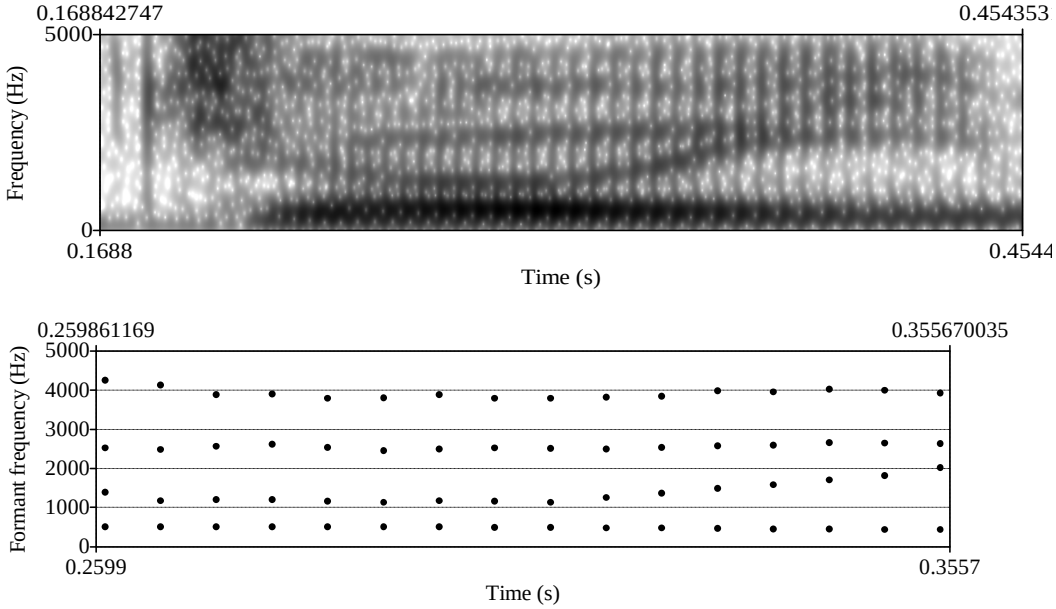




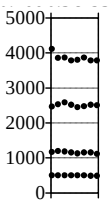


This exists in variation with [oj], but this is rare, as the normative formal [-o] in these lexical items realizes a colloquial form [-uj], and [oj] is somewhat of a combination of two competing forms. This phenomenon will be covered at length in section VI.G.1.a, but for now suffice it to say that [oj] is a possible variation on the normative colloquial [uj]:

- example: [oj] from “جوی” [ʤoj] ‘ditch; creek’:

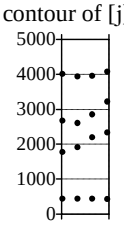


FF contour of [o]:



FFs:
3840Hz
2560Hz
1170Hz
500Hz

FF contour of [j]:

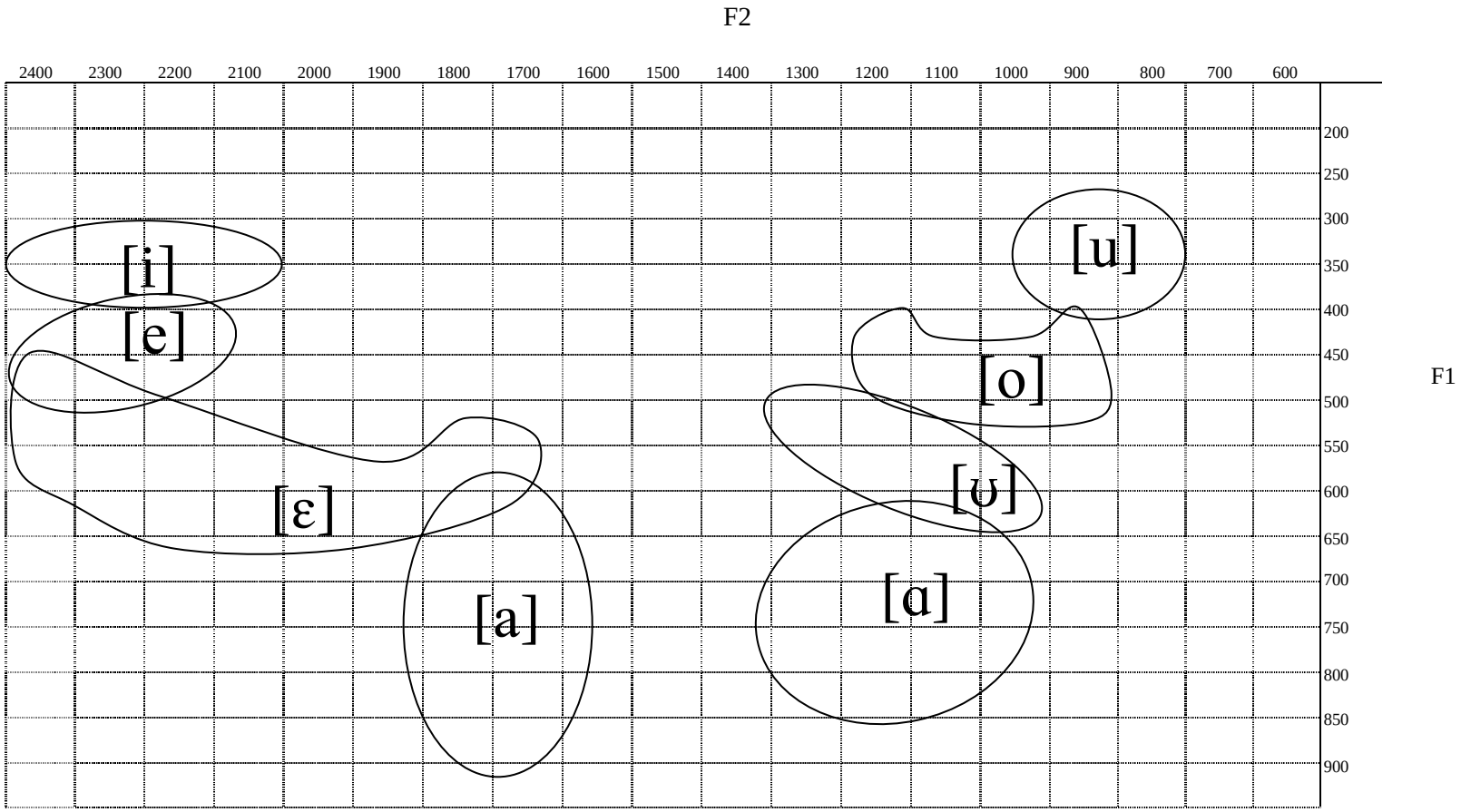


FFs:
3980Hz
2800Hz
2100Hz
430Hz

III.D Dari Vowels and Diphthongs: Overview

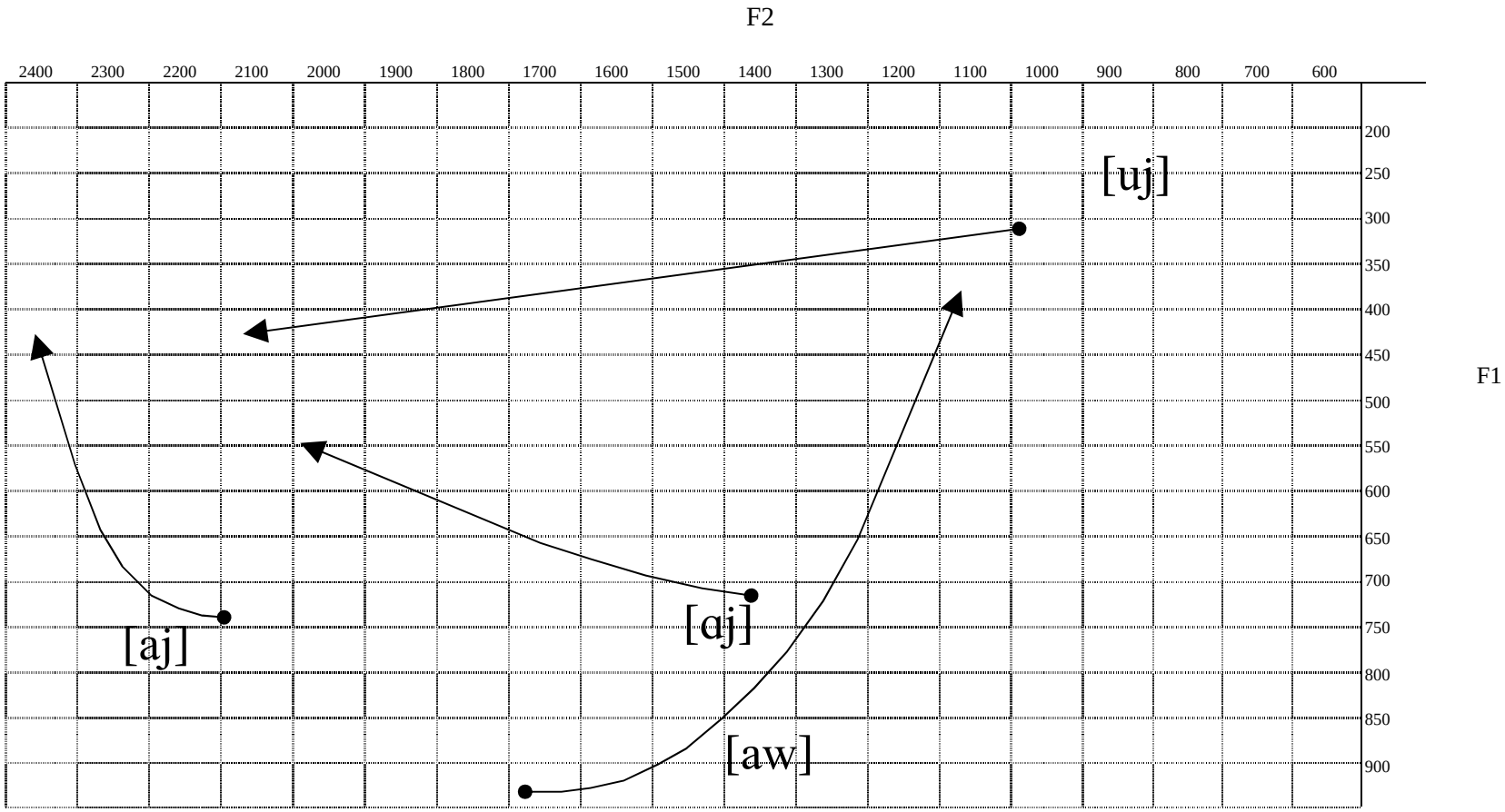
III.D.1 Vowel Overview

V	F1:	F2:	V	F1:	F2:
[i]	300	2050	[u]	290	770
	390	2720		430	1170
[ɛ]	450	1620	[ʊ]	450	1000
	670	2650		650	1400
[e]	370	2120	[o]	400	880
	520	2500		490	1300
[a]	570	1600	[ɑ]	450	1010
	890	1900		950	1400


F2
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
F1
[i]
[e]
[ɛ]
[a]
[u]
[o]
[ʊ]
[ɑ]

III.D.2 Diphthong Overview

V_1	F1:	F2:		V_2	F1:	F2:
[aj]	[a]	750	2120	[j]	420	2450
[aw]	[a]	950	1770	[w]	440	1040
[ɑj]	[ɑ]	720	1330	[j]	550	2070
[uj]	[u]	310	850	[j]	430	2100


F2
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
F1
[uj]
[aj]
[ɑj]
[aw]

III.E Dari Consonants: Introduction

The consonant system of Modern Persian varies relatively little between the three principal dialects, and as Dari is the first of the three under consideration, I will note elements of the consonant phonologies of all three dialects and further explore them as this paper progresses. Thus, mention will be made of the processes particular to Iranian Persian and Tajik, but they will not be analyzed until each dialect is fully dealt with. The following is the consonantal phonemic inventory of Dari:

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glottal
Nasal	m		n						
Plosive	p b	f	t d				k g	q	ʔ
Fricative				s z	ʃ ʒ		x ɣ		h
Affricate					ʧ ʤ				
Approximate	w					j			
Trill			r						
Lateral			l						

The starred phonemes in the below list are those that systematically differ in the three dialects. Dari and Tajik are more similar to each other than Iranian is to the other two dialects in its phonetic realization of consonants:

- The Dari phoneme /w/ is realized as /v/ in both Tajik and Iranian with little exception.

- The Dari and Iranian Persian /l/ is optionally rendered [ɫ] in Tajik.

- /ɣ/ and /q/ have merged in Iranian as [ɢ] (with allomorphs [ɣ] and [χ]).

- /ʔ/ and /h/ are widely elided in Dari and Tajik, but generally remain in Iranian Persian aside from some phonologically conditioned ellipses in colloquial Iranian.

Below are examples of each of the Dari consonants in word-initial, word-medial, and word-final positions:

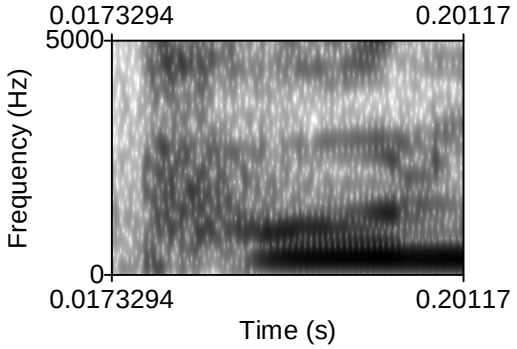
[b]	“بی” [be] ‘without’	“صباح” [sa-bɑ] ‘tomorrow’	“خوب” [xub] ‘good’
[p]	“پا” [pɑ] ‘foot’	“چپه” [ʧa-paˈ] ‘flip’	“چاپ” [ʧɑp] ‘print’
[f]	“فارسی” [fɑr-siˈ] ‘Persian’	“صفحه” [saf-haˈ] ‘page’	“صاف” [sɑf] ‘clear’
[m]	“من” [man] ‘I’	“عمه” [a-maˈ] ‘aunt (pat.)’	“گم” [gʊm] ‘lost’
[w]*	“وی” [waj] ‘he; she’	“عوض” [a-waz] ‘changed’	“دلو” [dalw] ‘Aquarius’
[d]	“دست” [dast] ‘hand’	“ادعاء” [ɛd-ʔɑˈ] ‘claim’	“رد” [rad] ‘deny’
[t]	“طی” [taj] ‘during’	“حتی” [ha-tɑˈ] ‘even’	“بوت” [but] ‘shoe’
[z]	“ضن” [zan] ‘doubt’	“مزه” [ma-zaˈ] ‘taste’	“میز” [mez] ‘table’
[s]	“سند” [sa-nadˈ] ‘document’	“خنثی” [xʊn-sɑˈ] ‘neutral’	“کس” [kas] ‘person’
[n]	“نه” [no] ‘nine’	“تنها” [tan-hɑˈ] ‘only’	“تن” [tan] ‘body’
[r]	“روز” [roz] ‘day’	“قرار” [qa-rɑrˈ] ‘place’	“کنار” [kɛ-nɑrˈ] ‘side’
[l]*	“لطف” [lʊtf] ‘favor’	“گلاب” [gʊ-lɑbˈ] ‘rose’	“گل” [gʊl] ‘flower’
[ʤ]	“جای” [ʤaj] ‘place’	“کجا” [ku-ʤɑˈ] ‘where?’	“حاج” [hɑʤ] ‘Hajj’
[ʧ]	“چای” [ʧɑj] ‘tea’	“بچه” [ba-ʧaˈ] ‘child’	“لوچ” [luʧ] ‘naked’
[ʒ]	“ژاله” [ʒɑ-laˈ] ‘frost’	“ویژه” [wi-ʒaˈ] ‘special’	“ماساژ” [mɑ-sɑʒˈ] ‘massage’
[ʃ]	“شد” [ʃʊdˠ] ‘became’	“آشک” [ɑ-ʃakˈ] ‘awshak’	“موش” [muʃ] ‘mouse’
[j]	“یک” [jak] ‘one’	“آیا” [ɑ-jɑˈ] ‘whether’	“رآی” [rɑj] ‘vote’
[g]	“گفت” [gʊft] ‘said’	“دیگر” [di-garˈ] ‘other’	“سگ” [sag] ‘dog’
[k]	“کرد” [kardˠ] ‘did’	“چکر” [ʧa-karˈ] ‘crash’	“پشک” [pɛ-ʃækˈ] ‘cat’
[ɣ]*	“غیر” [ɣajr] ‘un-’	“باغبان” [bɑɣ-bɑnˈ] ‘gardener’	“دروغ” [dʶruɣˈ] ‘lie’
[x]	“خیر” [xajr] ‘fine’	“آخر” [ɑ-xərˈ] ‘final’	“یخ” [jax] ‘ice’
[q]*	“قندول” [qan-dulˈ] ‘cute’	“نقش” [naqʃ] ‘sketch’	“حق” [haq] ‘right’
[h]*	“حرف” [harf] ‘letter’	“بهار” [ba-hɑrˈ] ‘spring’	“تشریح” [taʃ-rihˈ] ‘describe’

[ʔ]*[19] “آن” [ʔɑn] ‘that’ “ادعا” [ɛd-ʔɑˈ] ‘claim’ n/a

III.E.1 Dari Aspiration

The following spectrogram analyses of Dari voiceless plosives demonstrate the phonological aspiration common to all dialects of Modern Persian. The distribution of aspiration in Persian matches that of English and is not contrastive with its unaspirated counterpart. In the first and last examples, note that the duration of the aspiration burst is almost as long as the vowel itself.

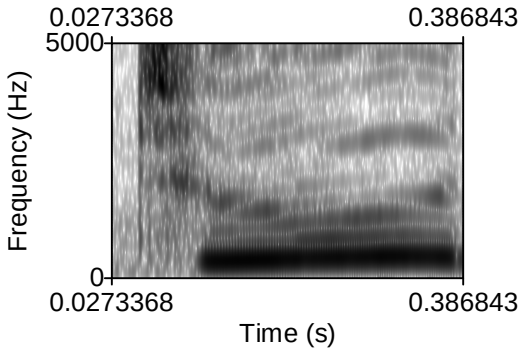
- example: “پول” [pʰul] ‘money’



[p | ʰ | u | l]

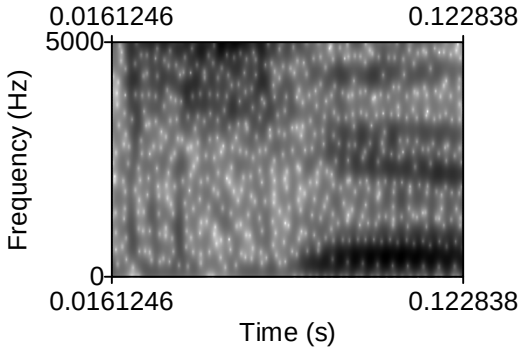
[19] The glottal stop is not contrastive in pre-vocalic word-initial position, and only ever pronounced in formal register. It is not a native Persian occurrence, but rather a reinterpretation of [ʔ] and [ʕ] in original Arabic tokens.

- example: “تو” [t^hu] ‘you’



[t | ^{h} | u]

- example: “کی” [k^hi] ‘who’



[k | ^{h} | i]

III.E.2 Dari Phonology: h-drop

Colloquial Dari, unlike its two dialectal counterparts, does not phonologically realize [h], and in many cases this affects the surrounding phonemes[20]:

[20] Phonological processes associated with this ellipsis are the same as those which occur when the glottal stop is elided, and hence they will be covered together in section III.5.D.

	Formal:	Spoken:	Gloss:
“همه”	/haˈ-ma/	[aˈ-ma]	‘all’
“بهار”	/ba-hɑrˈ/	[ba-ɑrˈ]	‘spring’
“صبح”	/sʊbḥ/	[sʊb]	‘morning’

III.E.3 Dari Phonology: ʔ-drop

Many Arabic loanwords contain /ʔ/ and /ʕ/ in their original forms, and these have merged in formal Dari (as well as Iranian and Tajik) as /ʔ/. These are also elided in spoken Dari with the same effect on the surrounding phonemes as ellipsis of [h] (see section III.E.4):

Arabic:		Formal Dari:	Spoken Dari:	Gloss:
“انعکاس”	[æn-ʕæ-kæːsːˈ]	/an-ʔa-kɑsˈ/	[a-na-kɑsˈ]	‘reaction’
“هیئت”	[hæj-ʔætˈ]	/haj-ʔatˈ/	[a-jatˈ]	‘delegation’

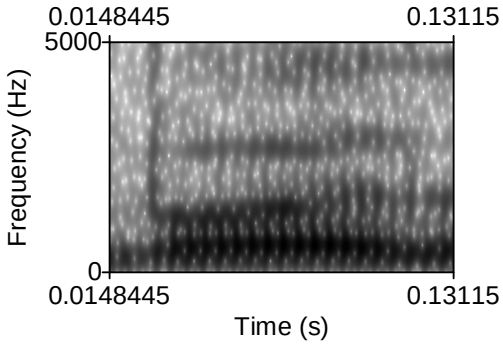
III.E.4 Dari Phonology: h-drop, ʔ-drop and Compensatory Lengthening

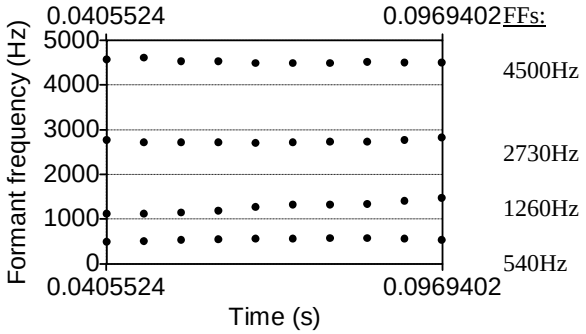
One phenomenon unique to Dari is sound shift accompanying compensatory lengthening associated with the ellipsis of [h] and [ʔ] in colloquial speech. Take for example the token illustrated in section III.B:

		Formal Dari:	Spoken Dari:	Gloss:
“بعد”	/baʔd/ →	*[baːd] → [bad]	[bɑd]	‘after’

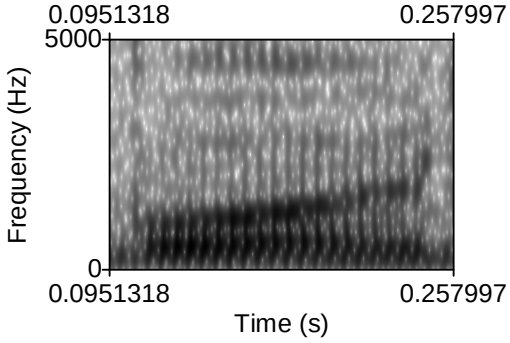
Note in the below examples that there is no appreciable distinction in vowel quality between the Dari long low vowel [ɑ] and the lengthened reflux of /aʔ/ → [aa] → [ɑ]:

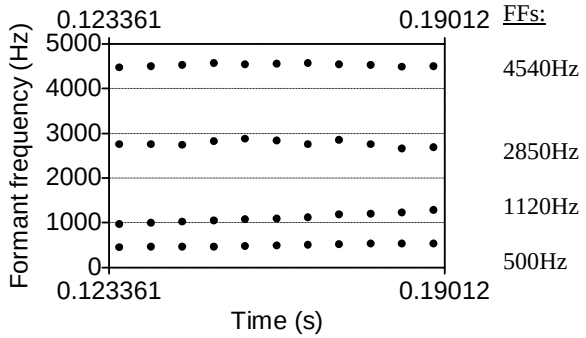
- example: “بعد” [bɑd] ‘after’:





- example: “باد” [bɑd] ‘wind’:





The same can be said of other lexical items that elide /h/, such as:

		Formal Dari:	Spoken Dari:	Gloss:
“قهر”	/qahr/ →	*[qaːr] → [qar]	[qɑr]	‘angry’
“فهمیدن”	/fah-mi-danˈ/ →	[fa-mi-danˈ]	[fɑ-mi-danˈ]	‘understand’

Two homophonous exceptions to this alternation are:

	Formal Dari:	Spoken Dari:	Gloss:
“بهر”	/bahr/ → *[baːr]	[bar]	‘for; to’
“بحر”	/bahr/ → *[baːr]	[bar]	‘sea’

The alternation does not occur when the elided glottal /h/ or /ʔ/ is surrounded by different segments, such as one case in which /ɑʔa/ reduces to [ɑ]:

	Formal Dari:	Spoken Dari:	Gloss:
“ساعت”	/sɑ-ʔatˈ/ →	[sɑ-atˈ] → *optionally* [sɑt]	‘hour; watch’

In other instances the final reduction does not occur both with /ɑʔa/ and /aʔɑ/:

	Formal Dari:	Spoken Dari:	Gloss:
“آهن”	/ɑ-hanˈ/	[ɑ-anˈ]	‘iron’
“بهار”	/ba-hɑrˈ/	[ba-ɑrˈ]	‘spring’

In two examples of verb stem morphology, the ellipsis of [h] results in different phonological effects, semivowel epenthesis in one and devoicing of an adjacent consonant in another:

	Formal Dari:	Spoken Dari:	Gloss:
“می خواهم”	/meˈ-xɑ-ham/ → *[meˈ-xɑ-am] →	[meˈ-xɑ-jam]	‘I want’
“می دهم”	/meˈ-da-ham/ → [meˈ-dʰam] → *[meˈ-dam] →	[meˈ-tam]	‘I have’

There is no phonological conditioning on the above section, but it appears that the availability might display frequency effects (This will be addressed further in section VII on frequency and register effects).

III.E.5 Dari Phonology: r-drop

Dari allows for the optional ellipsis of word-final /r/ in certain frequently used lexical items:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“دیگر”	/di-garˈ/	[di-gaˈ]	‘other’[21]
“در”	/dar/	[da]	‘in’
“اگر”	/a-garˈ/	[aˈ-ga]	‘if’
“اینطور”	/inˈ-tawr/	[iˈ-to][22]	‘like this’
“آنطور”	/ɑnˈ-tawr/	[uˈ-to]	‘like that’
“کردن”	/kar-danˈ/	[ka-danˈ]	‘to do’
“مگر”	/ma-garˈ/	[ma-gaˈ]	‘but not’
“خیر”	/xajr/	[xaj]	‘okay; well’
“سیر”	/ser/	[sej]	‘satiated’

[21] To be considered later are the cases in which only one member of a homophonous pairs allows the ellipsis of a phoneme: “دیگر” [di-garˈ] ‘other’ ([di-gaˈ]) vs. “دیگر” [di-garˈ] ‘afternoon’ (*[di-gaˈ]) and “در” [dar] ‘in’ ([da]), but “در” [dar] ‘door; opening’ (*[da]).

[22] Clearly more is at play here than merely r-drop, but for now, the ellipsis of [r] is the focal point of the analysis.

In one case, an instantiation of /r/ that is part of an onset consonant cluster is elided, and the high vowel before it becomes a semivowel:

	Formal Dari:	Spoken Dari:	Gloss:
"برادر"	/bˈ-rɑ-darˈ/ or /brɑ-darˈ/	[bjɑ-darˈ]	'brother'

III.E.6 Dari Phonology: Definite Article Particle "را" ([rɑ])

The Dari lexicon contains an enclitic [rɑ] that appends postpositionally to direct objects and direct object phrases:

"من آن مرد را دیدم."	/man ɑn mard-rɑ diˈ-dam/ 'I that man-rɑ saw-1p-sing'	'I saw that man.'
"من آن مرد بلند را دیدم."	/man ɑn mard-e bland-rɑ diˈ-dam/ 'I that man-e tall-rɑ saw-1p-sing'	'I saw that tall man.'
"من یک مرد دیدم."	/man jak mard diˈ-dam/[23] 'I one man saw-1p-sing'	'I saw a man.'

The vowel of the enclitic is reduced to [a] in colloquial speech, and the consonant elides following a consonant:

	Formal Dari:	Spoken Dari:	Gloss:
"تو را دید."	/tu#rɑ#did/	[tuˈ-ra#did]	'he/she saw you.'
"سگ را دید."	/sag#rɑ#did/	[saˈ-ga#did]	'he/she saw the dog.'

[23] Note in this example that the enclitic is absent as the direct object is indefinite.

III.E.7 Dari Phonology: n-drop

As with /r/, Dari allows for ellipsis of word-final /n/ in certain frequently used lexical items:

	Formal Dari:	Spoken Dari:	Gloss:
"من"	/man/	[ma]	'I'
"این"	/in/	[i]	'this'
"آن"	/ɑn/	[u][24]	'that'
"کن"	/kʊn/	[ko]	'do!'
"درون"	/da-runˈ/	[da-ruˈ]	'within'
"بیرون"	/bi-runˈ/	[bi-ruˈ]	'outside'

In one case, it is observed that the h-drop and n-drop conspire exhibiting compensatory lengthening in a similar fashion to that described in section III.E.4:

"من هم" /man#ham/ → [man#am] →*[ma#am] → *[maːm] → [mɑm] 'I also'

The plural demonstrative pronouns also exhibit effects of both h-drop and n-drop:

"اینها" /in-hɑˈ/ → [in-ɑˈ] → [i-jɑˈ] → [jɑ] 'they- here'

"آنها" /ɑn-hɑˈ/ → *[ɑ-nɑˈ] → [u-nɑˈ] → [u-wɑˈ] → [wɑ] 'they- there'

[24] In this example and the next, a vowel change has taken place along with the ellipsis. The pre-nasal raising of the low back vowel [ɑ] will be dealt with extensively in chapter VI, as this lexical item seems to be the only token in which this occurs in Dari, while in Iranian [ɑ] → [u]/__[+nas] is almost a wholesale shift.

And those same pronouns as part of locative adverbs ‘here’ and ‘there’ also reduce, as does the lexeme “جا” [ʤɑ] ‘place:’

“اینجا”	/inˈ-ʤɑ/ → [inˈ-ʤa] → [iˈ-ʤa]	‘here’
“آنجا”	/ɑnˈ-ʤɑ/ → [unˈ-ʤɑ] → [unˈ-ʤa] → [uˈ-ʤa]	‘there’

III.E.8 Dari Phonology: /ʤ/ > [j]

There are two lexical items in Dari in which /ʤ/ optionally shift to [j] in casual speech:

	Formal Dari:	Spoken Dari:	Gloss:
“پنج”	/panʤ/	[pajn]	‘five’
“مجلس”	/maʤ-lɛsˈ/	[maj-lɛsˈ]	‘meeting’

III.E.9 Dari Phonology: /f/ and /b/ > [w]

Dari phonology allows an optional alternation of certain instances of the phoneme /b/ and some and pre-velar instances of /f/ as [w].

	Formal Dari:	Spoken Dari:	Gloss:
“دیشب”	/diˈ-ʃab/	[diˈ-ʃaw]	‘last night’
“خواب	/xɑb/	[xɑw] → [xaw]	‘sleep; dream’
“خوب”	/xob/	[xo]	‘good’
“آب”	/ɑb/	[ɑw] → [aw]	‘water’

“آفتاب”	/ɑf-tɑb/	[ɑf-taw]	‘sky’
“ماهتاب”	/mɑh-tɑb/	[mɑh-taw]	‘moon’
“شربا”	/ʃʊr-bɑˈ/	[ʃʊr-wɑˈ]	‘soup’
“سیب”	/seb/	[sew]	‘apple’
“سبز”	/sabz/	[sawz]	‘green’
“قبظ”	/qabz/	[qawz]	‘constipated’
“افغانستان”	/af-ɣɑ-nɛ-stɑnˈ/	[aw-ɣɑ-nɛ-stɑnˈ]	‘Afghanistan’
“افگار”	/af-gɑr/	[aw-gɑr]	‘hurt; wounded’

There is one other related shift that predates Proto-Persian as evidenced by the fact that the lexical items affected retain the variant as an optional pronunciation in all dialects, and that is /b/ → /f/ / __ [+sibilant]. Both lexical items are represented as underlying lexical forms because they have been reanalyzed as variants:

“افزار”	/ab-zɑrˈ/	/af-zɑrˈ/	‘equipment; tools’

III.E.10 Dari Phonology: Verbal Morphology and Possessive Pronouns

Another area of phonological reduction in Dari is that of verbal morphology. The normative third person present tense endings in Dari verbs end in /d/ which is elided in colloquial speech:

/me	+	kʊ	+	nad/	→	[me-kʊ-na]
Prog.		‘do’		3p. sing.		‘he/she does’

/me	+	kʊ	+	nand/	→	[me-kʊ-nan]
Prog.		‘do’		3p. pl.		‘they do’

A separate phenomenon is shared by the present perfect verbal morphology and the possessive pronoun enclitics:

/kar	-	da	+	am/	→	[ka-daːmˈ] → [ka-demˈ]
‘do-past stem’			+	‘I am’		‘I have done’

The explanation for the above vowel raising will be dealt with later, but this phenomenon also occurs in the case of singular possessive pronouns:

/ba	-	ʧaˈ	#	am/	→	[ba-ʧaːmˈ] → [ba-ʧemˈ]
‘child’			+	‘my’		‘my child’
/ba	-	ʧaˈ	#	at/	→	[ba-ʧaːtˈ] → [ba-ʧetˈ]
‘child’			+	‘your-fam.’		‘your child’
/ba	-	ʧaˈ	#	aʃ/	→	[ba-ʧaːʃˈ] → [ba-ʧeʃˈ]
‘child’			+	‘his/her’		‘his/her child’

This appears to be related to the ezafe effect on vowel-final lexical items noted in section III.C.3.

<u>III.E.11 Dari Phonology: Third Person Copula “است” ([ast])</u>

The third person singular conjugation of the verb to be, “است” /ast/ or “هست” /hast/[25] also undergoes reduction but to varying degrees, and in some cases (unpredictable) undergoes epenthesis:

[25] These are historically distinct verbs in Persian, but because of Dari’s wholesale loss of [h], they have fallen together as one, and the distinction is obsolete.

	Formal Dari:		Reduced:	Spoken Dari:	Gloss:
"هست/است"	/hast/ / /ast/	→	[as], [s], or [a][26]	[asˈ-ta][27]	'is'

This will be returned to in section VI.J.3.

III.E.12 Dari Phonology: Phonotactics

Dari is unique among the three dialects under discussion in that it optionally permits onset consonant clusters, but prohibits consonant + syllabic consonant in the coda:

CCVC:

	Iranian:	Tajik:	Dari:	Gloss:
"بلند"	[bo-lændˈ]	[ba-łandˈ]	[bland] or [bə-landˈ]	'tall'
"برنج"	[bɛ-rɛnʤˈ]	[bˈ-rinʤˈ]	[brɛnʤ] or [bə-rɛnʤˈ]	'rice'
"شکم"	[ʃɛ-kæmˈ]	[ʃˈ-kamˈ]	[ʃkam] or [ʃə-kamˈ]	'belly'

*CVCC̣:

	Iranian:	Tajik:	Dari:	Gloss:
"چشم"	[ʧɛʃm]	[ʧaˈ-ʃəm]	[ʧʃəm] or [ʧɛˈʃəm]	'eye'
"فکر"	[fɛkr̩]	[fiˈ-kər]	[fɛˈ-kər] or [fəˈ-kər]	'think'
"شکر"	[ʃokr̩]	[ʃuˈ-kər]	[ʃʊˈ-kər] or [ʃuˈ-kər]	'thank'

[26] This occurs less frequently than [asˈ-ta], but quite frequently among speakers of the Hazaragi dialect of Dari, a dialect of central Afghan Shiites. This will not be dealt with in detail here, as it is the same process as occurs normatively in spoken Iranian.
[27] See section VI.J.3 for further explanation of this analogous epenthesis.

“طفل”	[tɛfl̩]	[tiˈ-fəl]	[tɛˈ-fəl]	‘child’
“قطل”	[ɢætl̩]	[qaˈ-təl]	[qaˈ-təl]	‘murder’
“قدر”	[ɢædr̩] or [ɢæd]	[qa-darˈ]	[qaˈ-dar] or [qa][28]	‘amount’
“شغل”	[ʃoɣl̩]	[ʃʊˈ-ɣəl]	[ʃʊˈ-ɣəl]	‘occupation’
“شکل”	[ʃækl̩]	[ʃaˈ-kəl]	[ʃaˈ-kəl]	‘form’
“قهر”	[ɢæhr̩]	[qaˈ-hər] or [qaːr]	[qaˈ-hər] or [qɑr]	‘angry’
“شهر”	[ʃæhr̩]	[ʃaˈ-hər] or [ʃaːr]	[ʃaˈ-hər] or [ʃɑr]	‘city’

III.E.13 Dari Phonology: Other Phonological Phenomena

This section will address two other cases of shifts and lenition particular to Dari phonology that, while significant, do not merit individual treatment. Section III.E.13.a will treat lenition of stressed long vowels /ī/ and /ū/ when a derivative affix renders them unstressed. Furthermore, it will deal with ellipsis of short vowels due to the addition of verbal prefixes. Section III.E.13.b will deal with word-final consonant changes of which the first is syllable related and the second is strictly phonological. Section III.E.13.c will cover random cases of metathesis that occur according to the sonority sequencing principle. Finally, section III.E.13.d will deal with one case of s-drop.

[28] This reduction only occurs in combination with demonstrative pronouns “آن” and “این,” so:

“آنقدر” [ɑnˈ-qa-dar] → [unˈ-qa-dar] → [uˈ-qa-dar] → [uˈ-qa] ‘that much’

“اینقدر” [inˈ-qa-dar] → [iˈ-qa-dar] → [ɛˈ-qa-dar] → [ɛˈ-qa] ‘this much’

III.E.13.a Dari Phonology: Vowel Lenition

The first case of lenition under consideration in this section involves shortening of long non-low vowels when derivational morphology causes the aforementioned segments to be destressed:

“تاریک” /tɑ-rikˈ/ ‘dark’ → “تاریکی” /tɑ-rɛ-kiˈ/ ‘dark’

“شیر” /ʃir/ ‘milk’ → “شیرین” /ʃɛ-rinˈ/ ‘sweet’

“شیرین” /ʃɛ-rinˈ/ ‘sweet’ → “شیرینی” /ʃɛr-niˈ/ ‘sweetness’

“دور” /dur/ ‘far’ → “دوربین” /dʊr-binˈ/ ‘binoculars’

“شور” /ʃor/ ‘salty’ → “شوربا” /ʃʊr-bɑˈ/ ‘soup’

This phenomenon is almost trivial as the phonological process is so natural. It remains part of this analysis, as it does not occur in the case of the long low vowel:

“کباب” /ka-bɑbˈ/ ‘kabob’ → “کبابی” /ka-bɑ-biˈ/ ‘kabob vendor’

“آب” /ɑb/ ‘water’ → “آبی” /ɑ-biˈ/ ‘blue’

Fortunately for this analysis, it will be shown in Chapter VI that OT can account for the shift without causing the low vowels to participate.

The second case of lenition involves ellipsis of short vowels in the unstressed first syllable of polysyllabic verb stems when the aspect prefix is affixed. This is not particular to Dari, but occurs in all three dialects:

Infinitive:	Gloss:	Present tense:	Gloss:
“گزاشتن” /gᵘ-zɑʃ-tanˈ/	‘to pass’	“می گزارم” /megˈ-zɑ-ram/	‘I pass’
“شناختن” /ʃᵉ-nɑx-tanˈ/	‘to know’	“می شناسم” /meʃˈ-nɑ-sam/	‘I know’

This will be returned to in chapter VI, but for the sake of this analysis, history will demonstrate that the balance of such cases are not deletions in the inflected form, but rather epentheses in the stem form. Even the second example below bears this out in the gloss which is a direct congate: /ʃn/ ↔ /kn/.

III.E.13.b Dari Phonology: Word-final Consonant Shifts

Word-final voiced consonants devoice. This is straightforward and so cross-linguistically attested that it shall itself be considered a well-formedness constraint in the OT chapter:

	Formal Dari:	Spoken Dari:	Gloss:
“آزاد”	/ɑ-zɑdˈ/	[ɑ-zɑtˈ] or [ɑ-zɑdˀˈ]	‘free’
“سگ”	/sag/	[sak] or [sagˀ]	‘dog’

Other word-final consonant shifts have occurred that are not wholesale and hence not phonologically conditioned. This shift, /r/ → /l/, among other incomplete shifts, will raise questions regarding change that must bear the caveat ‘sometimes:’

	Formal Dari:	Spoken Dari:	Gloss:
“دیوار”	/de-wɑrˈ/	[de-wɑlˈ]	‘wall’
“گرسیر”	/garm#ser/	[garm#sel]	‘Garmser (region)’

This second example may prove elucidating as the region is in Helmand Province, in the heart of Afghan Pashtunistan. The Pashto language retains both the flapped or trilled /r/, as Dari but is also possessed of the retroflex /ɽ/. This retroflex in loanwords is rendered /l/ in Dari such as in "کیلکین" [kil-kinˈ] 'window' borrowed from Pashto "کیړکین" [kiɽ-kinˈ] 'window.' It is possible that this shift /r/ → /l/ is analogous with /ɽ/ → /l/.

III.E.13.c Dari Phonology: Metathesis

In a few cases, Dari will metathesize word final consonant clusters whose orders violate the sonority sequencing principle such as "قلف" [qʊlf] 'lock' from Arabic "قفل" [qʊfl]. One case that is particularly interesting is the Arabic lexical item "لنعت" [lan-ʕatˈ] which is rendered [naʔ-lat] → [nɑ-lat] in Dari. Because this is not systematic, it will not be addressed beyond this mention of it.

III.E.13.d Dari Phonology: s-drop

There is one case of ellipsis that does not belong in other sections and that is /s/ → ∅ / __ /ʤ/ in the following:

	Formal Dari:	Spoken Dari:	Gloss:
"مسجید"	/mas-ʤidˈ/	[ma-ʤidˈ]	'mosque'

IV. From Proto-Persian to Iranian Persian

IV.A Iranian Phonology: Introduction

The Iranian phonemic inventory is less conservative than Dari (and in point fact less conservative than Tajik as well), with fewer phonemes due to several instances of convergence that have not occurred in Dari:

/ū/, /ō/ > /u/

/ī/, /ē/ > /i/ (Although Iranian retains [e], an alternation of the diphthong [ej] from original /aj/.)

/u/, /aw/ > /o/

/q/, /ɣ/ > /G/ ([ɣ] remains only as an allophone.)

This chapter will mirror the Dari Phonology section in that Section IV.B will introduce the Iranian Persian vowel system, noting the various allophonic and free variations, and Sections IV.C.1 – IV.C.6 will provide an analysis of the individual vowels and diphthongs with spectrograms and formant frequency contours for each of the vowels. Section IV.D will chart the first and second formant frequencies for vowels and diphthongs, and Section IV.E will provide an analysis of the consonant phonology of Iranian.

IV.B Iranian Vocalic Phonemic Inventory

	Front	Central	Back
High:	/i/		/u/
Mid:	/ɛ/		/o/
Low:	/æ/		/ɑ/
Diphthongs:	/ɛj/; /ɑj/		

The phones [ʊ] and [ə] may be added to the inventory as variations of /o/ and /ɛ/ respectively, and /ɑ/ may additionally be rendered [ɔ]. None of these alternations are phonologically conditioned, but rather, they reflect the broadness in articulatory position of these phonemes, which is made evident by the chart in section IV.D.1.

It is assumed with respect to data from Middle Persian (covered in section II) that there only remain three of the original five sets of historical long/short vocalic pairs in Iranian:

MP:		Iranian:
/ī/ - /i/, /ē/	→	/i/ - /ɛ/
/ū/ - /u/, /ō/	→	/u/ - /o/
/ā/ - /a/	→	/ɑ/ - /æ/

This along with the other two dialects will be further analyzed in chapter VI within an Optimality Theoretic framework, but at present, it should be noted that the

Iranian distinctions appear to be in height rather than solely ATR as illustrated by the below set of minimal pairs:

/i/	"سیر"	[sir]	'garlic'
/ɛ/	"سر"	[sɛr]	'secret'
/u/	"دور"	[dur]	'far'
/o/	"دور"	[dor]	'turn'
/ɑ/	"دار"	[dɑr]	'possessed of'
/æ/	"در"	[dær]	'in; on'

IV.C Iranian Vowels Illustrated

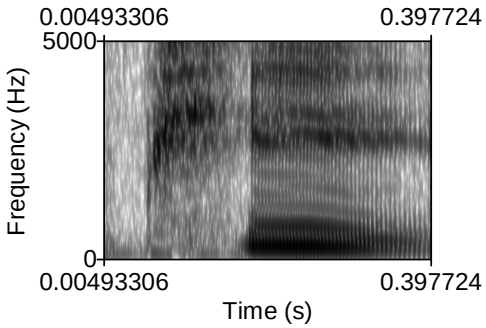
As with the Dari vowels sections, examples will be provided of each vowel in initial, medial and final position where available in the lexicon. Then spectrogram analyses and formant contour charts of each vowel will be added.

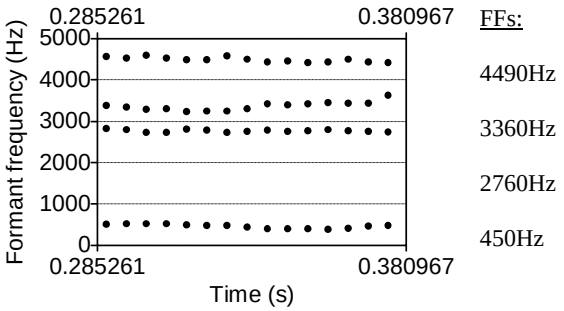
IV.C.1 High Front Vowel: [i]

The Iranian high front vowel [i] varies very little in height and backness with a first formant frequency range of 390 to 530 Hz and a second formant frequency range of 2400 to 2740 Hz. It is comparatively more fronted than the Dari [i] in that the Dari second formant frequency ranges from 2050 to 2720 Hz:

[i] “ایتالیا” [i-tɑ-ljɑˈ] ‘Italy’ “دیر” [dir] ‘late’ “آبی” [ɑ-biˈ] ‘blue’

- example: “پی” [pʰi] ‘lard’:

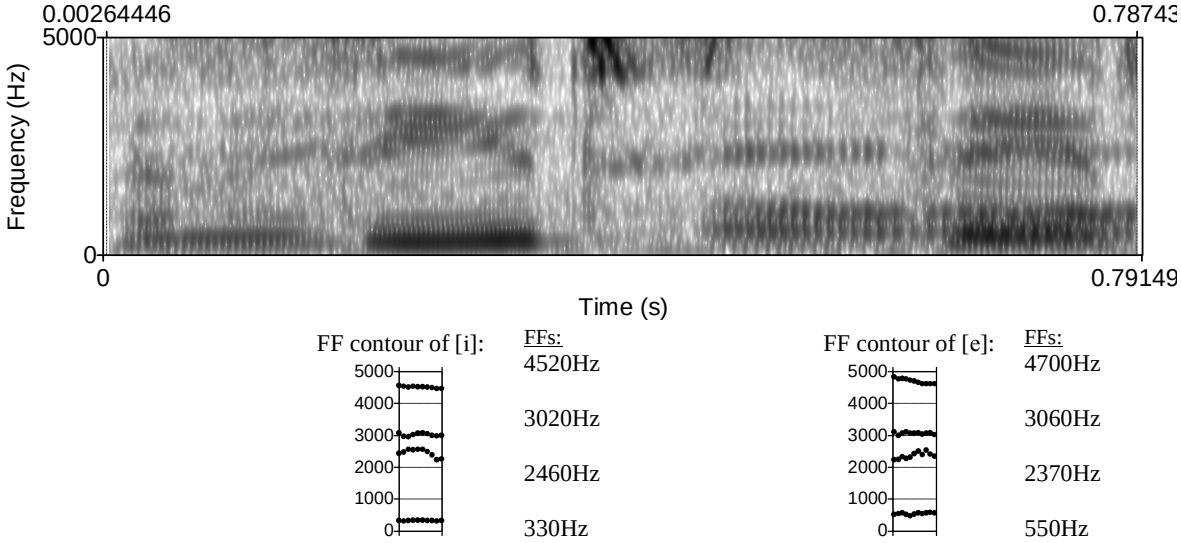




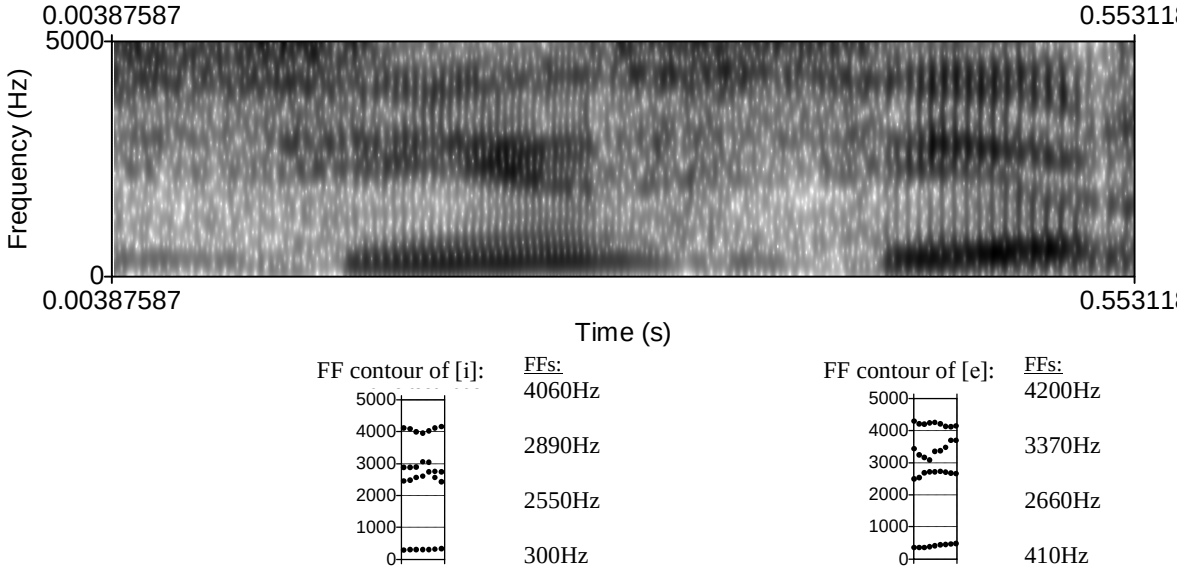
This will be addressed later in section IV.E.1, but the example above is a classic example of Iranian Persian aspiration. Note the intense burst of noise before the vowel in the spectrogram.

In order to determine whether Iranian shares the contrastive status of high front and mid front +ATR vowels [i] and [e], I composed a sentence with embedded minimal pairs [sir] ‘garlic’ and [ser] ‘satiated:’ “من از تورشی سیر سیر شدم.” /man#az#tur-ʃi#sir#ser#ʃʊ-dam/ ‘I have had my fill of pickled garlic.’ As the spectrograms show, [i] and [e] have come together in Iranian as [i], and have remained distinct in Dari and Tajik:

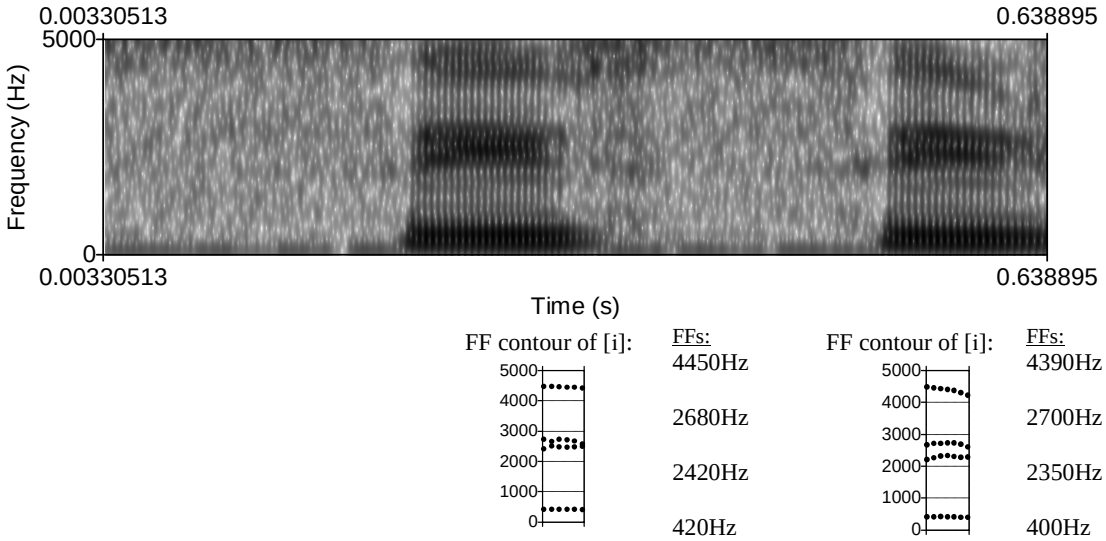
- example: [sir#ser] Afghan female:



- example: [sirser] Tajik male:



- example: [sirsir] Iranian female:

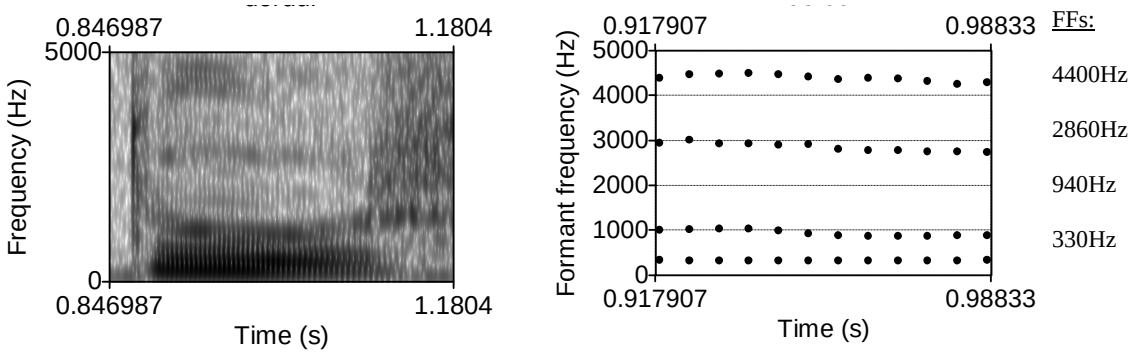


IV.C.2 High Back Vowel: [u]

The Iranian Persian high back vowel [u] ranges in height from a first formant frequency of 300 to 450 Hz and in backness from a second formant frequency of 680 to 1200 Hz.

[u] “وجود” [vu-ʤudˈ] ‘existence’ “دور” [dur] ‘far’ “بو” [bu] ‘scent’

- example: “دور” [dur] ‘far’:



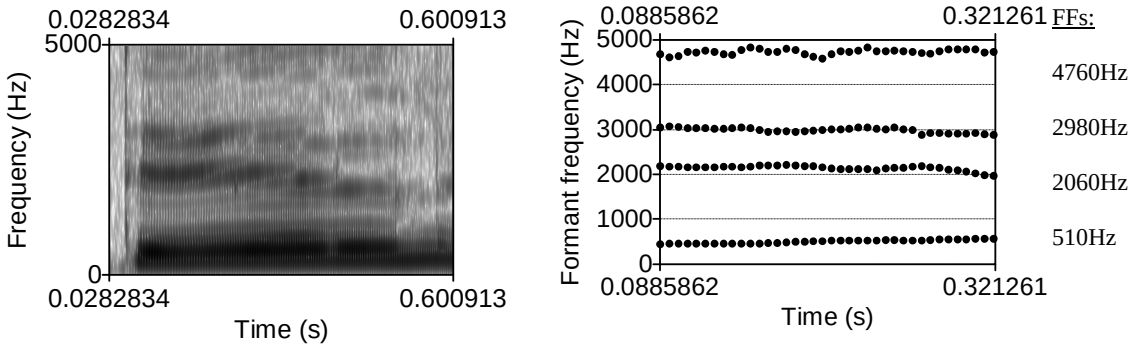
IV.C.3 Mid Front Vowel: [ɛ]

The mid front vowel [ɛ] is comparatively more back than its Dari counterpart with a first formant frequency range of 440 to 610 Hz and a second of 1420 to 2270 Hz:

[ɛ] “انقلاب” [ɛn-ɢˈ-lɑbˈ] ‘revolution’ “سن” [sɛn] ‘age’ “کرده” [kʲær-dɛˈ] ‘done’

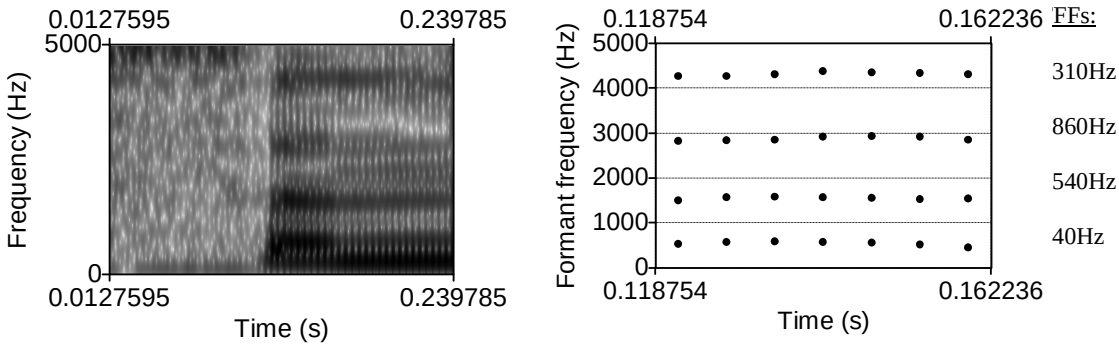
The Iranian [ɛ] occurs in a number of predictable environments, and for that reason I will include examples from several different morpho-syntactic elements. First is the vowel [ɛ] as it occurs as the historical Persian high front short vowel:

- example: “دل” [dɛl] ‘heart’:



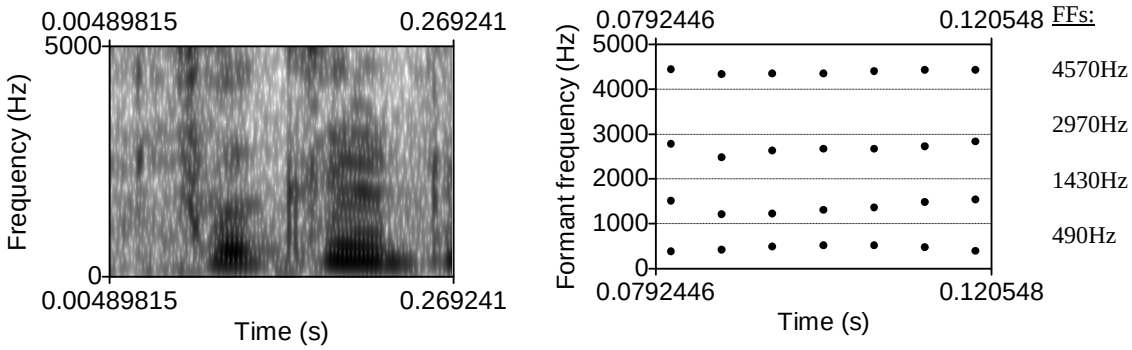
The below example, however demonstrates that /ɛ/ realizes variance as far back as [ə] in Iranian:

- example: “سن” /sɛn/ → [sən] ‘age’:



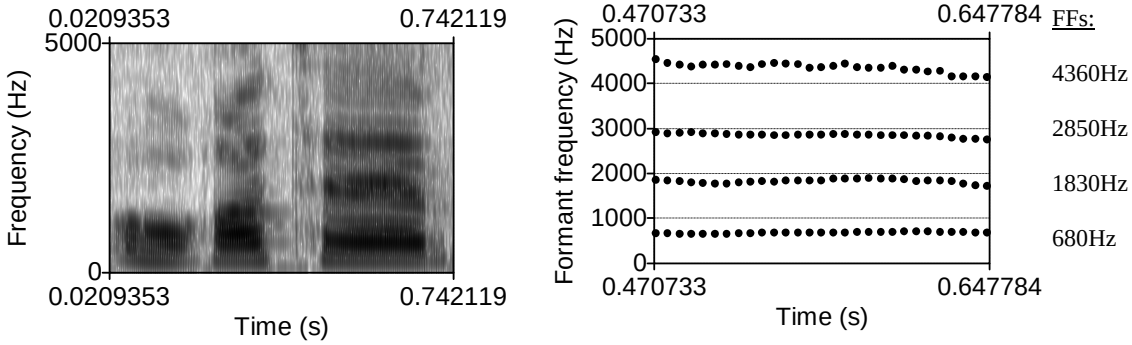
This variation between Iranian [ɛ] and [ə] is also noted in Dari in cases such as /fɛˈ-kər/ which can optionally be realized as [fəˈ-kər]:

- example (Dari): “فکر” /fɛˈ-kər/ → [fəˈ-kər]‘thought’:



The second case is that of the ezafe [ɛ], as explained in section III.C.3, which involves epenthesis of the vowel [ɛ] between nouns or noun phrases and their modifiers:

- example: “اوقت” [oʷ-ɣɑˈ-tɛ] ‘periods of’:



There is little if any difference between the quality of the ezafe [ɛ] and that of its non-grammatical counterpart.

The third case, and second systematic instance of the vowel [ɛ] is found word-finally in cases that were historically /-a/, and remain so in Dari and Tajik. This includes multiple nouns and adjectives, the perfect participial form of verbs, and the third person singular ending of present tense verbs:

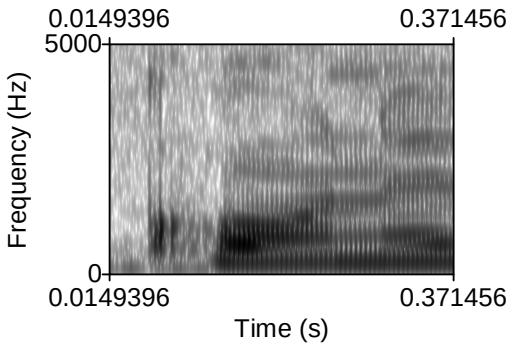
	Mod. Pers:	Iranian:	Dari:	Tajik:	Gloss:
Noun:	“خانه”	[xu-nɛˈ]	[xɑ-naˈ]	[xo-naˈ]	‘house’
Adj.:	“کهنه”	[koh-nɛˈ]	[kʊ-naˈ]	[kʉ-naˈ]	‘old’
Participle:	“کرده”	[kʲær-dɛˈ]	[ka-daˈ]	[ka(r)-daˈ]	‘done’
3pers Sing:	“می کند”	[miˈ-ko-nɛ]	[meˈ-kʊ-na]	[mʲeˈ-kʊ-na]	‘does’

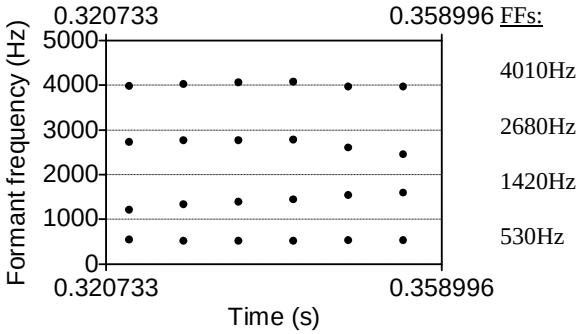
Along with historical evidence, this final example proves the originality of the sound as /-a/, as the formal register third person singular present tense ending of verbs is /-ad/, but in

all three dialects the final consonant is elided, leaving the vowel /-a/, save Iranian, which shifts this vowel to [-ɛ].

The Iranian word-final vowel [-ɛ] appears to be the most back of the instances of [ɛ], but this is likely the result of the phonological effects of the vowel appearing word-finally:

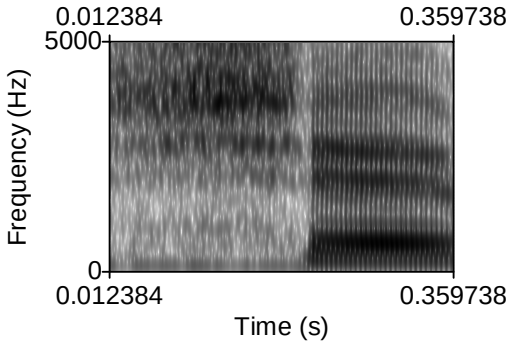
- example: “کهنه” [koh-nɛˈ] ‘old’:

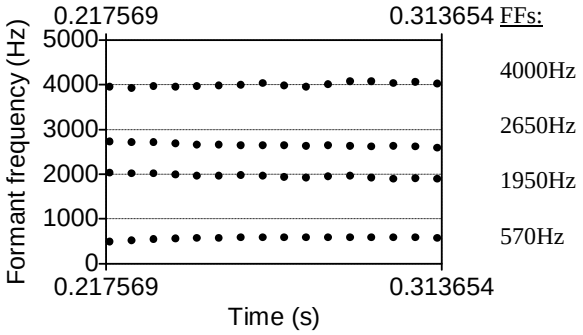




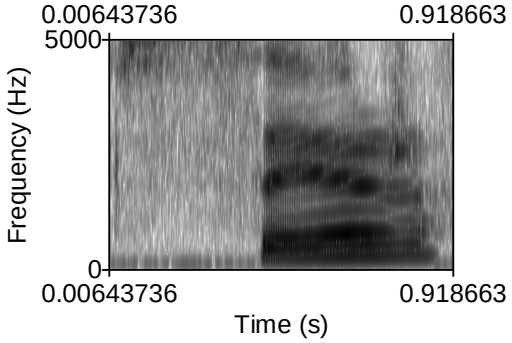
The fourth and final instance of Iranian [ɛ] is the realization of the Persian front mid long vowel, which is [e] in Dari and Tajik, but becomes [ɛ] in Iranian:

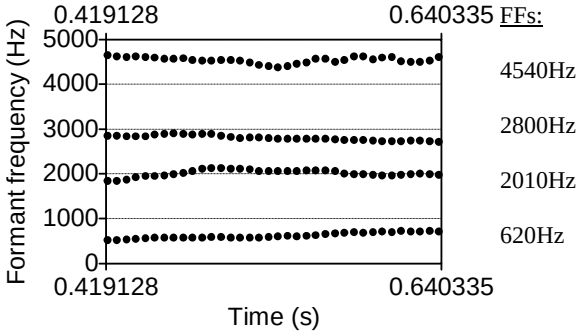
- example: “شعر” [ʃɛʔr] ‘poetry’:





- example: “سر” [sɛr] ‘secret’:





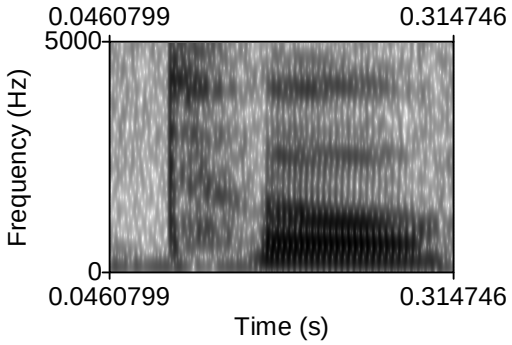
The above two examples were chosen because they are distinct in Dari, in which they are [ʃer] and [sɛr] and in Tajik, in which they are [ʃer] and [sir].

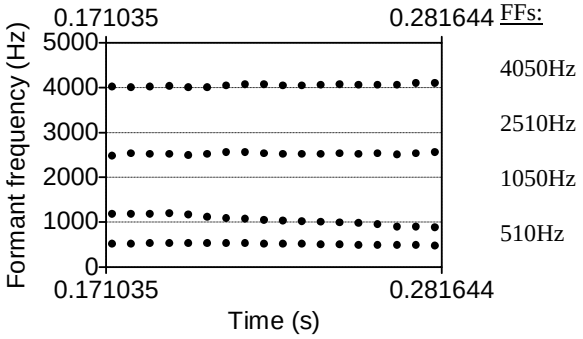
IV.C.4 Mid Back Vowel: [o]

The Iranian mid back vowel [o] has a first formant frequency range of 380 to 660 Hz and a second formant frequency range of 870 1200 Hz:

[o] “عهده” [oh-daˈ] ‘responsibility’ “پل” [pol] ‘bridge’ “چلو” [tʃɛ-loˈ] ‘bridge’

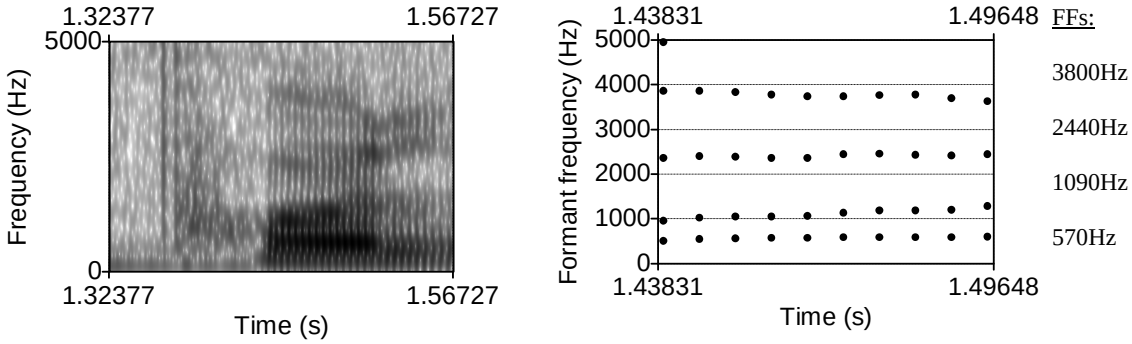
- example: “تو” [to] ‘you’:



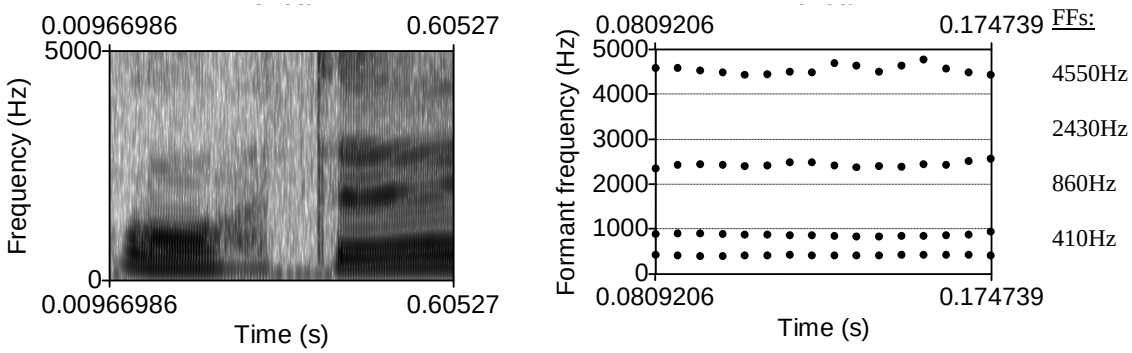


The above is an example somewhat in the middle of the possible positions for [o], but the forth formant can be higher or lower as evidenced by the following:

- example: “پل” [pol] ‘money’:

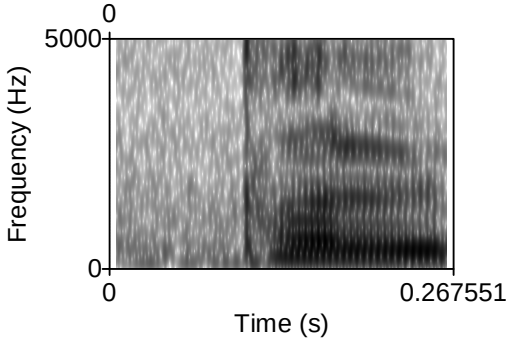


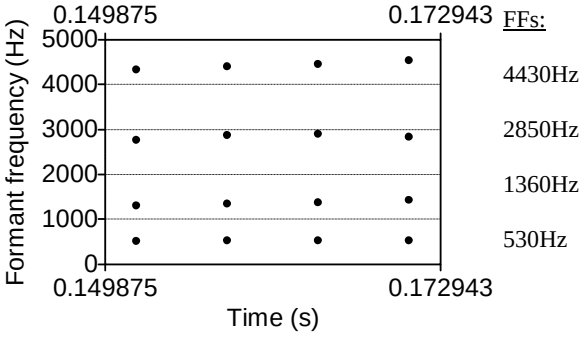
- example: “عهده” [oh-dɛˈ] ‘responsibility’:



Again, note that the quality of the Dari counterpart (the historical high back short vowel) in [pʊl] ‘bridge’ is significantly more fronted than the Iranian:

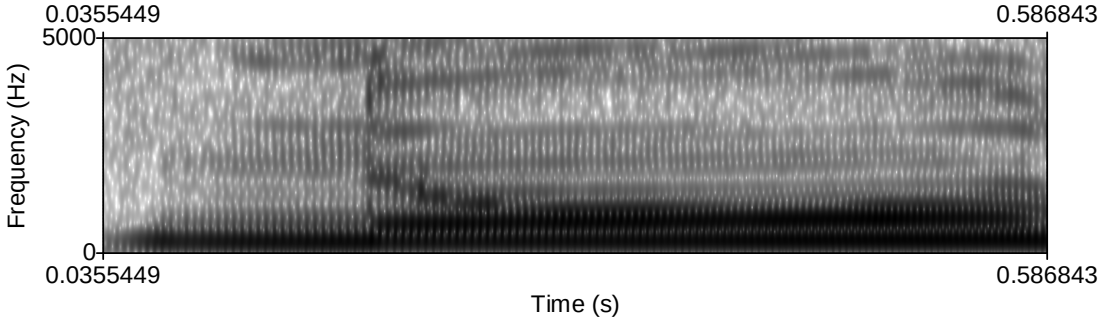
- example: “پل”[pʊl] ‘bridge’:

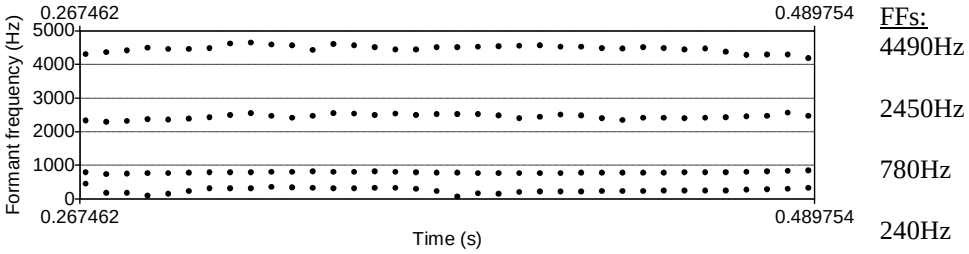




Certain sources have claimed that the Iranian [o] derived from original diphthong [aw] (see Dari section III.C.6) is the diphthong [ow], but this is inconsistent with the data obtained in this research, in which the vowel is essentially identical to the above vowel [o]:

- example: “نو” [no] ‘new’:





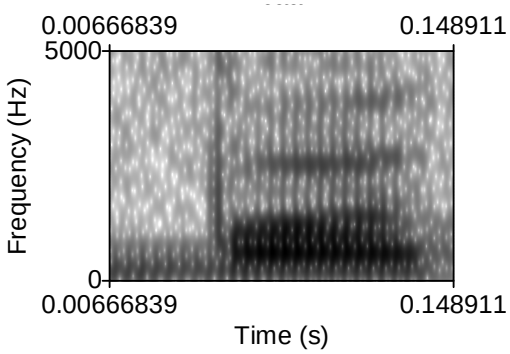
As is evident from the spectrogram and formant contour, the vowel is steady throughout articulation, and despite the fact that the first formant is considerably lower, the vowel is the same.

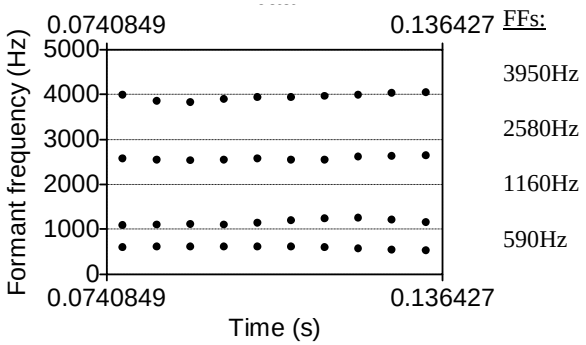
IV.C.5 Low Vowels: [æ], [ɑ]

The Iranian low back vowel varies widely in backness, bordering on free variation with [a]. The first formant frequency varies from 640 to 850 Hz, and the second varies from 1000 to 1500 Hz:

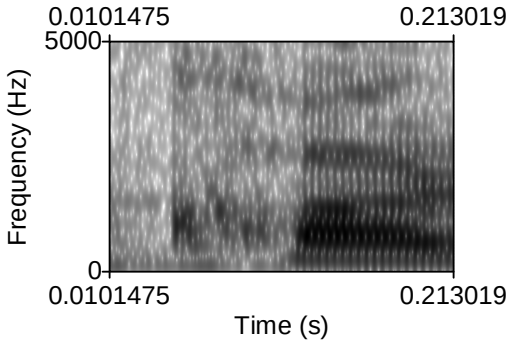
[ɑ] "عادی" [ɑ-diˈ] 'principal' "مال" [mɑl] 'hidden' "حتی" [ha-tɑˈ] 'until'

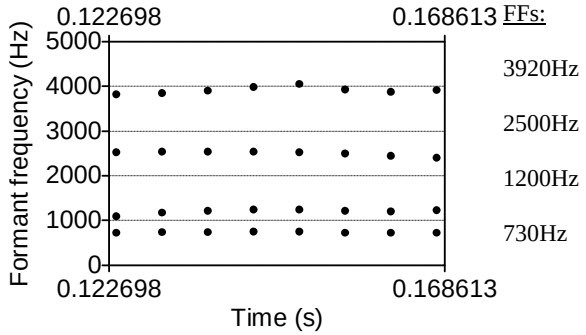
- example: "با" [bɑ] 'with':





- example: "کار" [kɑr] 'work':





Again, this will be covered at a later point in this chapter, but it bears noting that the duration of the aspiration burst following the velar plosive is approximately as long as the vowel itself indicating that Iranian, like Dari, allows phonetic aspiration of voiceless plosives.

A phenomenon related to the low back vowel and associated with Tehrani Persian, but widely attested in all Iranian Persian dialects and even western varieties of Dari involves the shift of /ɑ/ to [u] pre-nasally:

/ɑ/ → [u] / __ [+nas]

"نان" /nɑn/ → [nun] 'bread'

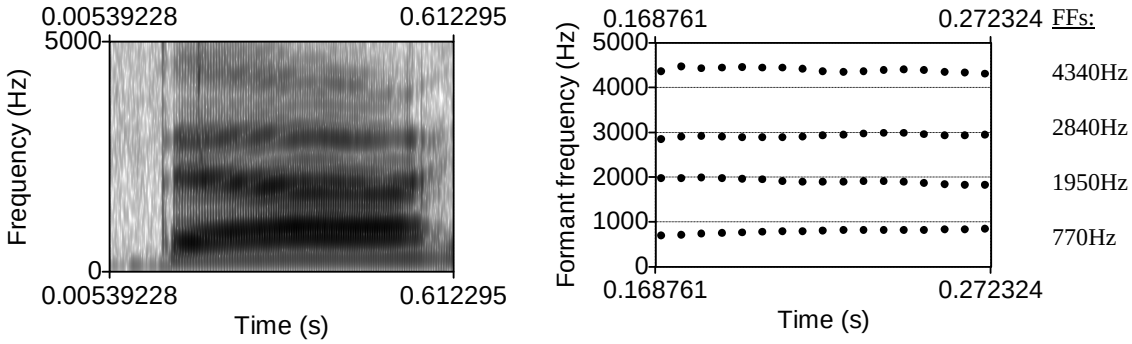
"نان" /xɑ-nɛ/ → [xu-nɛ] 'house'

This phenomenon is not possible in all instances of /ɑn/ or /ɑm/ indicating that there may be frequency or register effects, but this will be returned to briefly in chapter VII. Suffice it to say at this point that the resultant vowel [u] in no way differs from the canonic [u] discussed in section IV.C.2.

Finally, the Iranian Persian low short vowel is more fronted in location than its Dari counterpart, hence it is labeled [æ] rather than [a]. Its first formant frequency range is 670 to 950 Hz, and its second formant ranges from 1600 to 2100 Hz:

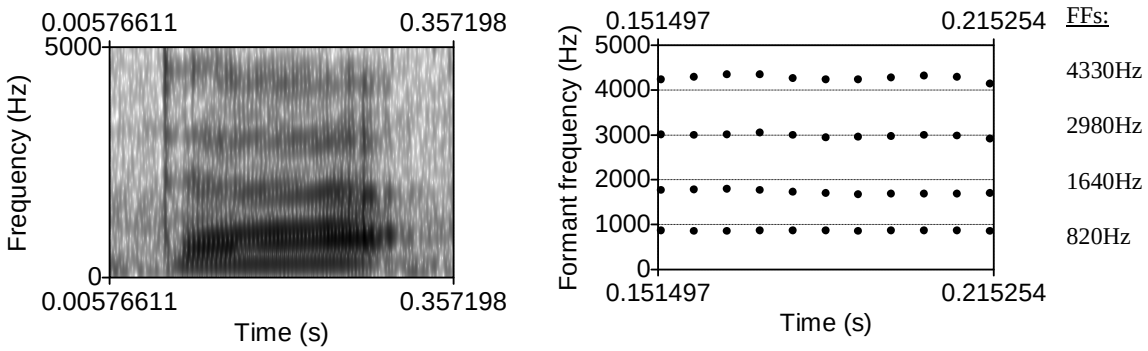
[æ] "از" [æz] 'from' "سر" [sær] 'head' "کرده" [kær-dæˈ] 'done'

- example: “ده” [dæ] ‘ten’:



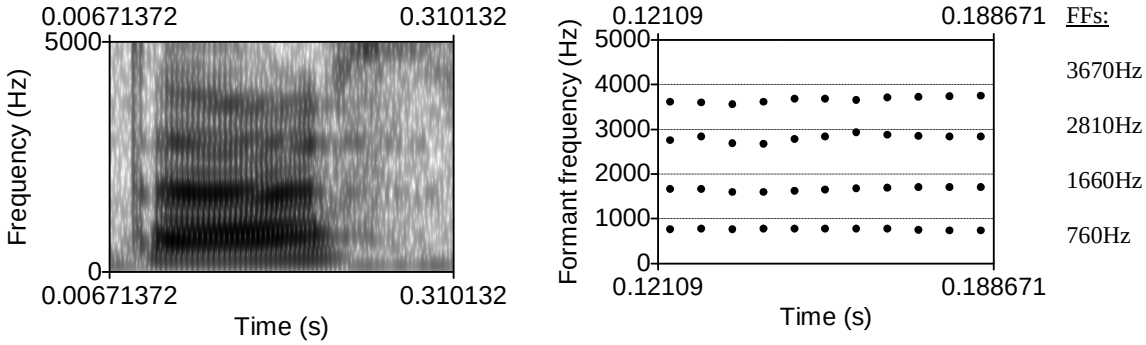
The frontness of this vowel contrasts with its Dari [a] counterpart, as made evident by the much lower second formant frequency in Dari:

- example: “ده” (Dari) [da] ‘ten’:



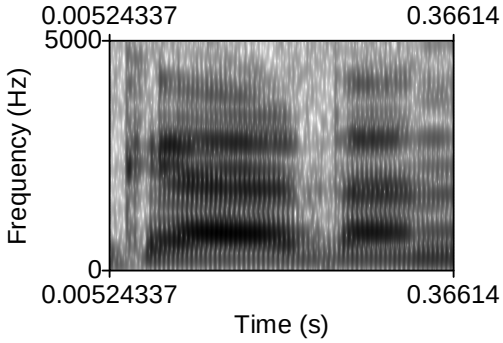
Because of the possible second formant frequency range however, Iranian will also allophonically realize [æ] as [a]:

- example: “درس” [dars] ‘study’:

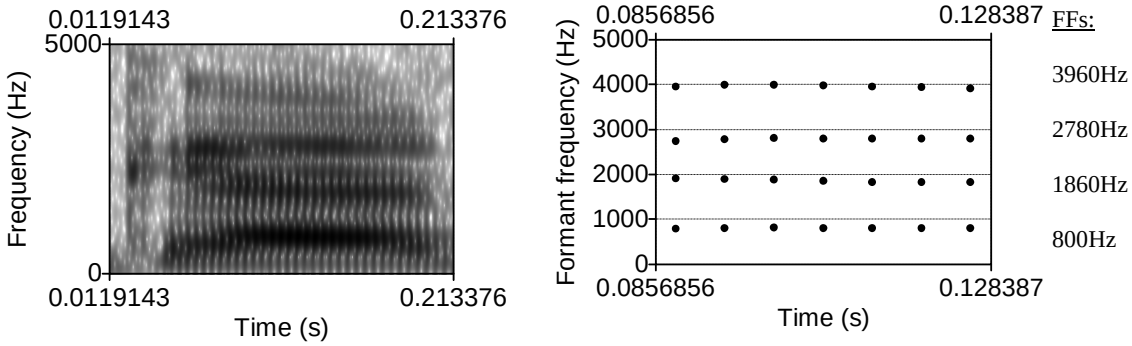


This allophonic variation appears to be tied to quantitative length rather than stress, as is evidenced by the lexical item “گردن” [gʲær-danˈ] ‘neck,’ in which the first vowel is over twice the length of the second vowel, despite the fact that the primary stress in the lexical item falls on the second syllable, not the quantitatively longer one:

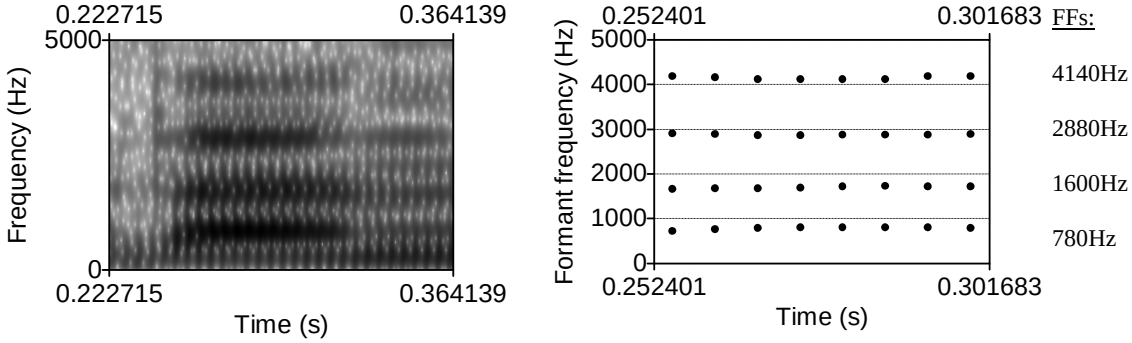
- example: “گردن” / gær-dænˈ/ → [gʲær-danˈ] ‘neck’:



[gʲær-]



[-dan]



The issue of palatalization will be covered in section IV.E.2 of this chapter, but note the substantial palatalization that takes place in examples like the above.

IV.C.6 Diphtongs: [ej], [ɑj]

Iranian lacks diphthongs that raise to the back vowel, as both [aw] and [ɑw] have converted to monophthongs:

/aw/ → [o] “نو” [no] ‘new’ (cf. Dari [naw])

/ɑw/ → [ɑv] “گاو” [gɑv] ‘cow’ (cf. Dari [gaw])

Incidentally, Tajik loses both of these diphthongs as well, because in Tajik both instances of /w/ become [v]:

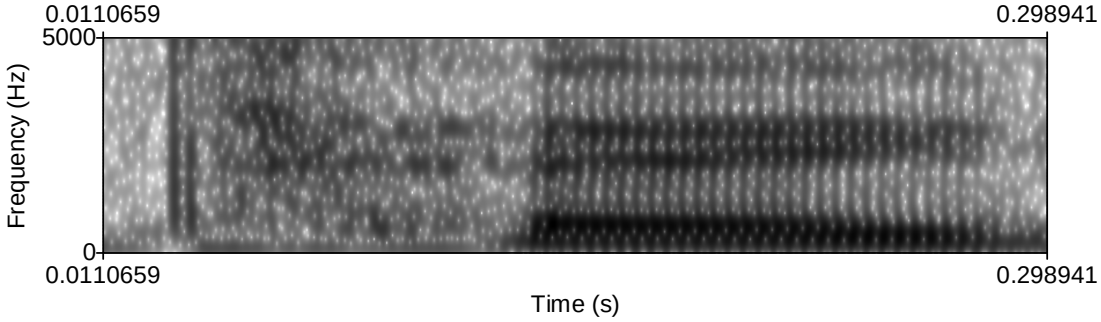
/aw/ → [av] “нав” [nav] ‘new’

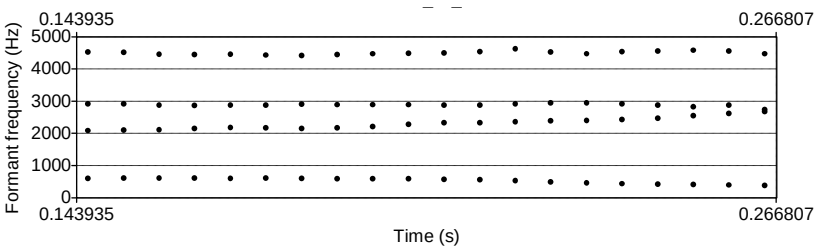
/ɑw/ → [ov] “гов” [gov] ‘cow’

Therefore, two remaining diphthongs are manifest in Iranian, [ej] and [ɑj]:

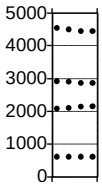
[ej] “ای” [ej] ‘oh…’ “غیر” [ɢejr] ‘un-’ “طی” [tej] ‘during’

- example: “طی” [t^hej] ‘during’:



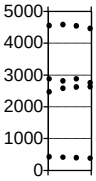


FF contour of [e]:



FFs:
4470Hz
2880Hz
2120Hz
610Hz

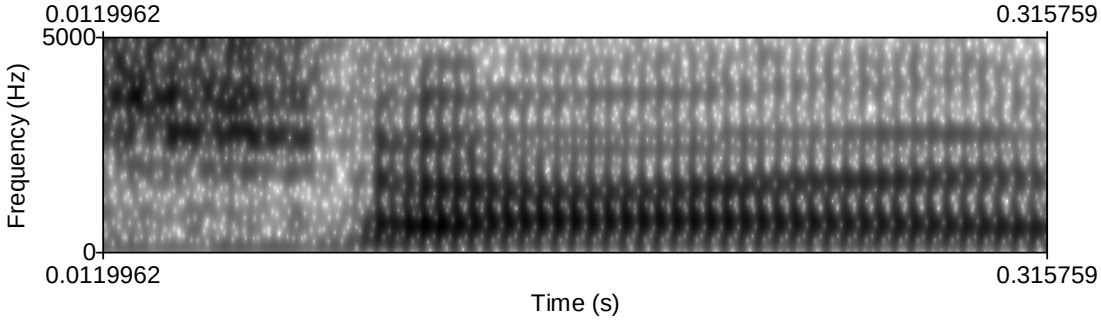
FF contour of [j]:

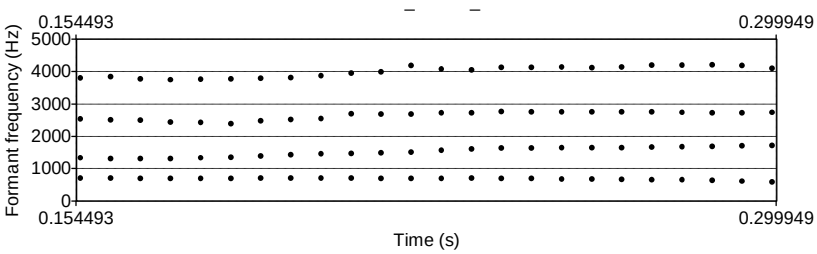


FFs:
4540Hz
2830Hz
2580Hz
390Hz

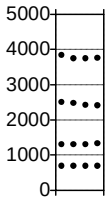
[ɑj] “آینه” [ɑj-nɛˈ] ‘mirror’ “مایل” [mɑjl] ‘willing’ “رآی” [rɑj] ‘vote’

- example: “ساعر” /ʃɑ-ʔɛrˈ/ [ʃɑjr] ‘poet’:



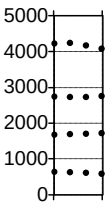


FF contour of [ɑ]:



FFs:
3770Hz
2460Hz
1310Hz
700Hz

FF contour of [j]:



FFs:
4180Hz
2730Hz
1700Hz
500Hz

The following is a complete overview of the development of Modern Persian diphthongs, noting origins. A full analysis will follow in chapter VI on OT and language change:

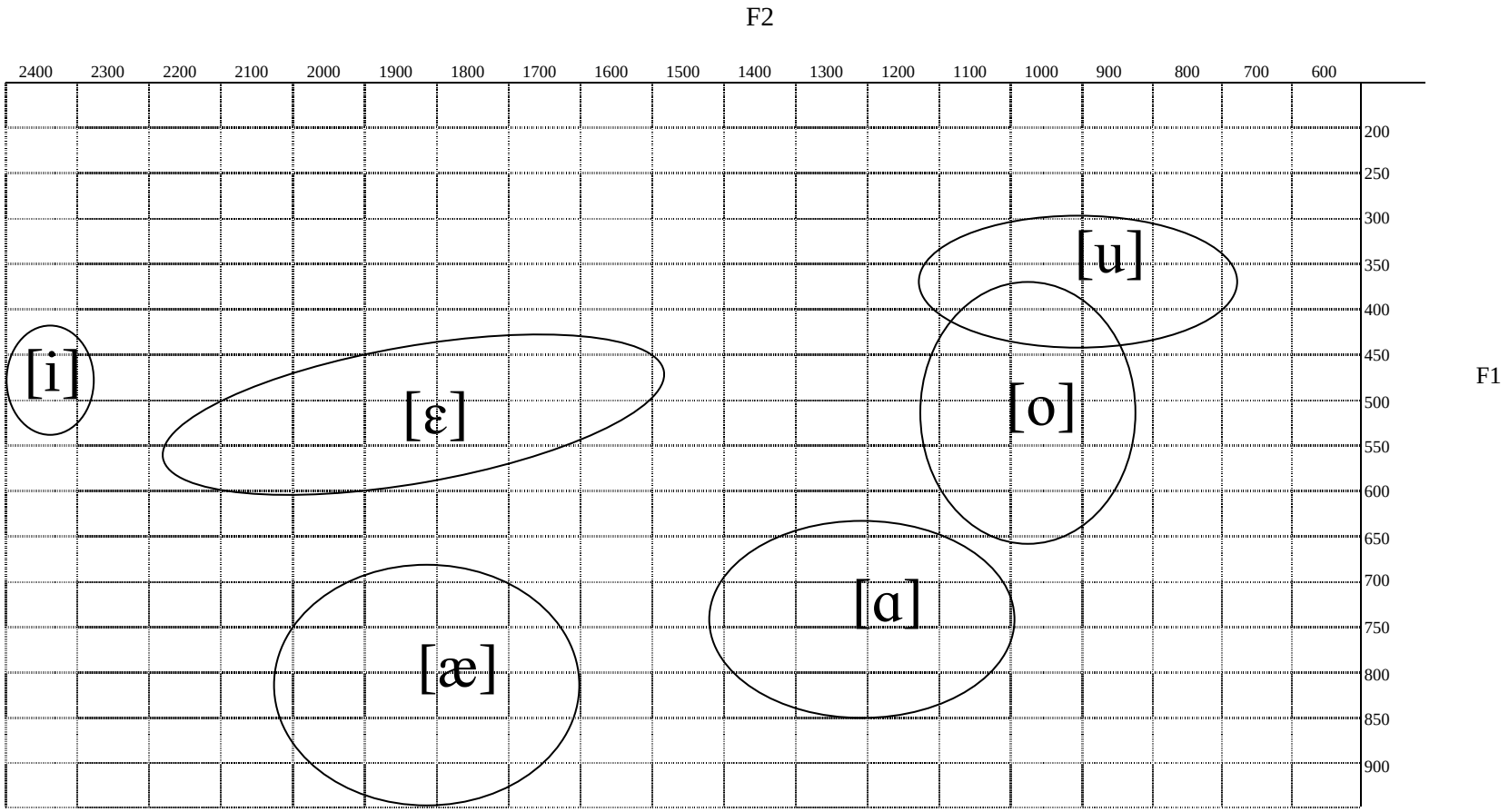
MP:		Iranian:	Dari:	Tajik:
/ē/; /aj/	→	[ej]	[aj]	[aj]
/aw/	→	[o]	[aw]	[av]
/āj/	→	[ɑj]	[ɑj]	[ɑj]
/āw/	→	[ɑv]	[ɑw]; [aw]	[ov]
/ōj/	→	[u]	[uj] ([oj][29])	[uj]

[29] Native speakers are aware that the formal register vowel is [o], but in the spoken variety, this lexical items with [oj] invariably render them [uj]

IV.D Iranian Vowels and Diphthongs: Overview

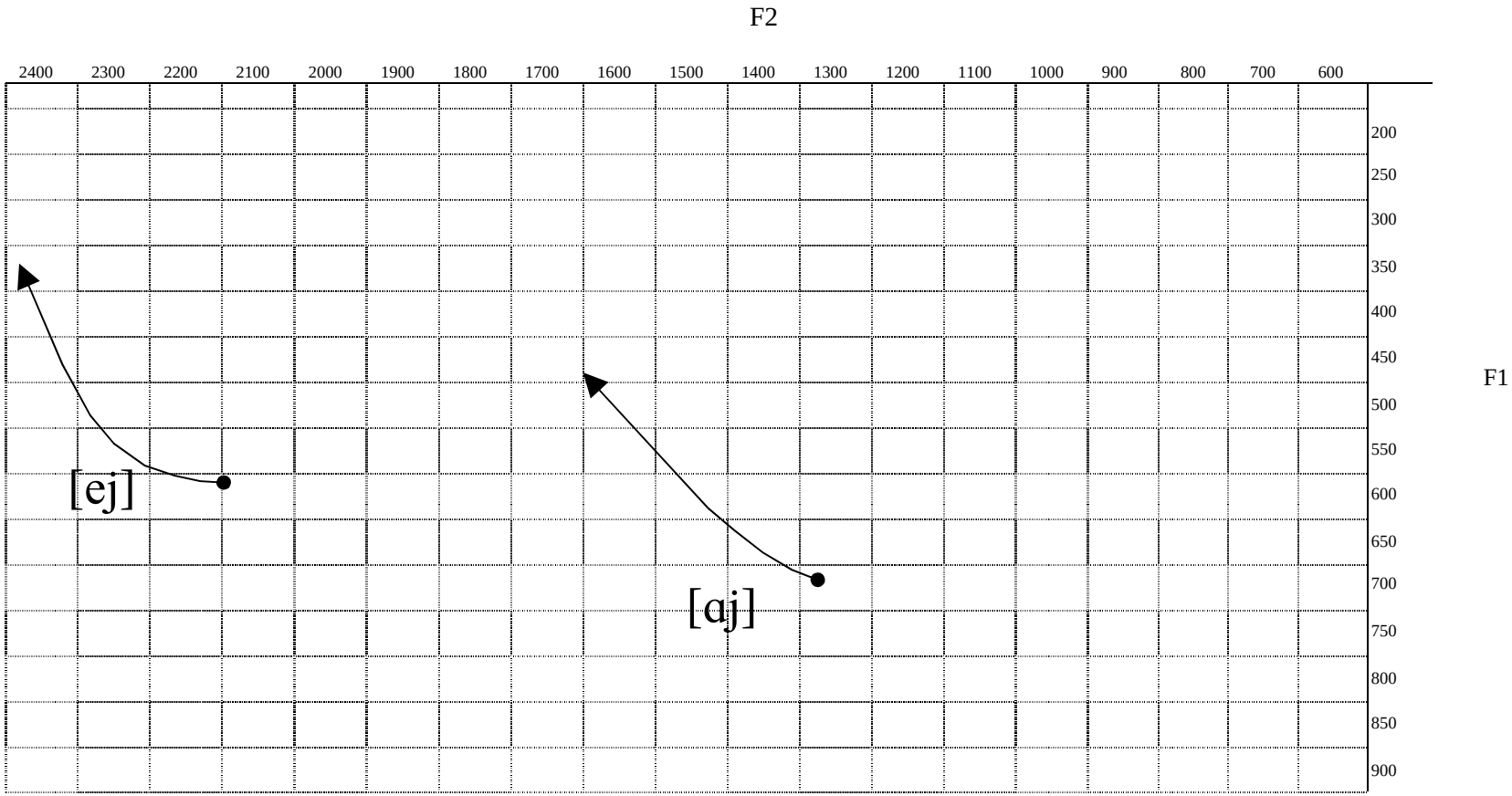
IV.D.1 Vowel Overview

V	F1:	F2:	V	F1:	F2:
[i]	390	2400	[u]	300	680
	530	2740		450	1200
[ɛ]	440	1420	[o]	380	870
	610	2270		660	1200
[æ]	670	1600	[ɑ]	640	1000
	950	2100		850	1500


F2
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
F1
[i]
[ɛ]
[u]
[o]
[ɑ]
[æ]

IV.D.2 Diphthong Overview

V_1	F1:	F2:		V_2	F1:	F2:
[ej]	[e]	610	2120	[j]	390	2580
[ɑj]	[ɑ]	700	1310	[j]	500	1700


F2
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
F1
[ej]
[ɑj]

IV.E Iranian Consonants: Introduction

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glottal
Nasal	m		n						
Plosive	p b	f v	t d				k g	ɢ	ʔ
Fricative				s z	ʃ ʒ		x	χ	h
Affricate					ʧ ʤ				
Approximate						j			
Trill			r						
Lateral			l						

As with the Dari section, the starred items on the list below are those that differ across the dialects:

- /v/ is the Iranian and Tajik realization of the PIE labial glide, while in Dari it is /w/. In other related dialects, this is realized as [ʋ].

- /h/ and /ʔ/ are less frequently elided in Iranian than they are in Dari and Tajik, and hence there are fewer suprasegmental phonological effects.

- [ɢ], [ɣ], and sometimes [χ] are allophones of one phoneme that is the result of a merger between original /q/ and /ɣ/. Possible phonological conditioning will be discussed in section IV.E.8.

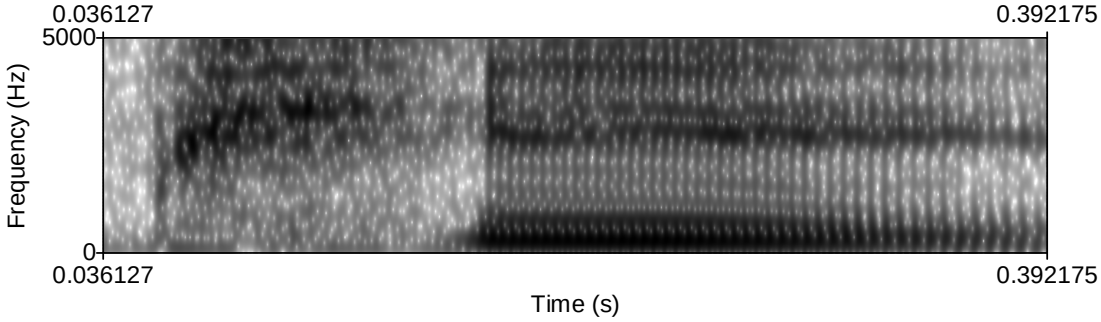
[b]	“بعد” [bæ-ʔædˈ] ‘after’	“طبیعی” [ta-bi-ʔiˈ] ‘natural’	“تب” [tæbˤ] ‘fever’
[p]	“پل” [pol] ‘bridge’	“تپه” [tɛ-pɛˈ] ‘hill’	“توپ” [tup] ‘cannon’
[f]	“فهرست” [fɛh-rɛstˈ] ‘list’	“صوفی” [su-fiˈ] ‘Sufi’	“معروف” [mæʔ-rufˤ] ‘famous’
[v]*	“وجود” [vo-judˈ] ‘existence’	“اول” [æ-vælˈ] ‘first’	“گاو” [gɑv] ‘cow’
[m]	“مست” [mæst] ‘drunken’	“همه” [hæ-mɛˈ] ‘all’	“هم” [hæm] ‘also’
[d]	“درس” [dærs] ‘study’	“کرده” [kʲær-dɛˈ] ‘done’	“رود” [rud] ‘river’

[t]	“تن” [tæn] ‘body’	“هفته” [hæf-tɛˈ] ‘week’	“هشت” [hæʃt] ‘eight’
[z]	“زود” [zudˤ] ‘quick’	“قاضی” [qɑ-ziˈ] ‘judge’	“از” [æz] ‘from’
[s]	“صد” [sæd] ‘hundred’	“دست” [dæst] ‘hand’	“بحث” [bæhs] ‘speech’
[n]	“نه” [næ] ‘no’	“اما” [æ-mɑˈ] ‘but’	“زن” [zæn] ‘woman’
[r]	“راست” [rɑst] ‘correct’	“قرن” [ɢærn] ‘century’	“شهر” [ʃæhr] ‘city’
[l]	“لیز” [liz] ‘slide’	“جلو” [ʤɛ-loʷˈ] ‘forth’	“مال” [mɑl] ‘item’
[ʤ]	“جو” [ʤu] ‘stream’	“حاجی” [hɑ-ʤiˈ] ‘Hajji’	“تاج” [tɑʤ] ‘crown
[ʧ]	“چه” [ʧe] ‘what	“ناچار” [nɑ-ʧɑrˈ] ‘inevitable’	“پیچ” [piʧ] ‘twist’
[ʒ]	“ژاله” [ʒɑ-lɛˈ] ‘frost’	“واژه” [vɑ-ʒɛˈ] ‘word’	“ولتاژ” [vol-tɑʒˈ] ‘voltage’
[ʃ]	“شصت” [ʃæst] ‘sixty’	“ایشان” [i-ʃunˈ] ‘they’	“باش” [bɑʃ] ‘be!’
[j]	“یک” [jek] ‘one’	“باید” [bɑ-jædˈ] ‘must’	“کی” [keʲ] ‘when’
[g]	“گل” [gol] ‘flower’	“زندگی” [zɛn-da-giˈ] ‘life’	“سگ” [sæg] ‘dog’
[k]	“کس” [kɛs] ‘person’	“خاکی” [xɑ-kiˈ] ‘beige’	“شک” [ʃæk] ‘doubt’
[ɢ]*	“قرن” [ɢærn] ‘century’	“اینقدر” [in- ɢadṛˈ] ‘thus’	“صندوق” [sæn-duɣˈ] ‘fund’
[x]	“خر” [xær] ‘donkey’	“آخر” [ɑ-xɛrˈ] ‘final’	“شاخ” [ʃɑx] ‘branch’
[h]*	“هندو” [hɛn-duˈ] ‘Hindu’	“بهار” [bæ-hɑrˈ] ‘spring’	“ماه” [mɑhˈ] ‘month’
[ʔ]*[30]	“اسب” [ʔæsbˤ] ‘horse’	“رئیس” [ræ-ʔisˈ] ‘chief’	n/a

IV.E.1 Iranian Aspiration

Iranian, like Dari, displays predictable aspiration on voiceless plosives in the syllable onset position:

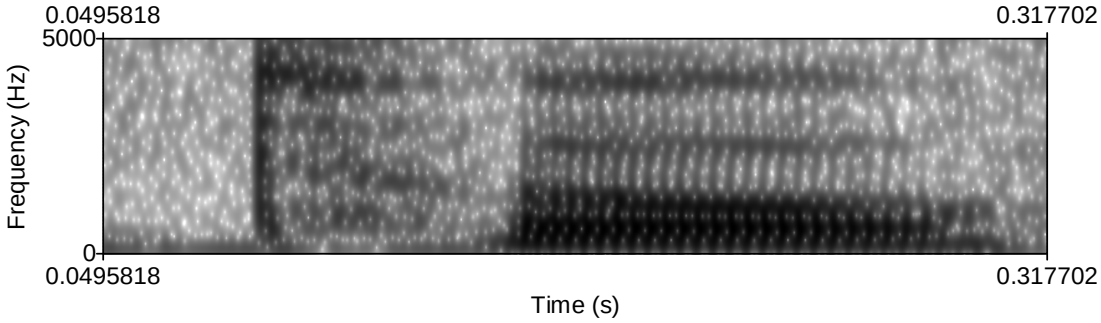
- example: “پی” [pʰi] ‘lard’:



[p | ʰ | i]

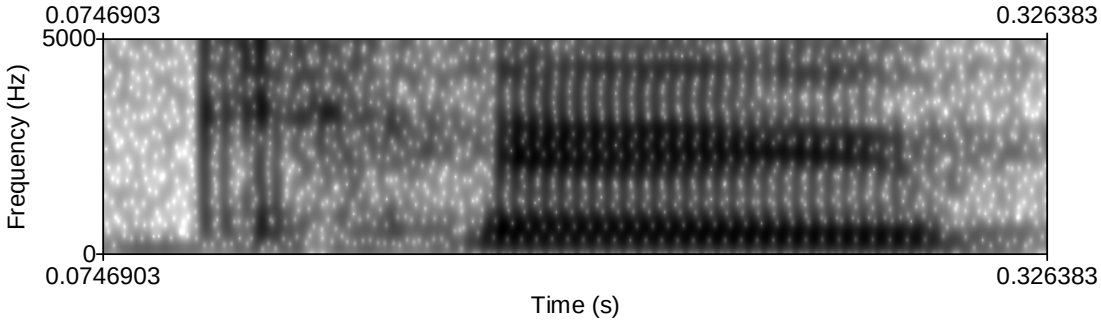
[30] The glottal stop is not contrastive in pre-vocalic word-initial position.

- example: “تو” [t^ho] ‘you’:



[t | ^{h} | o]

- example: “کی” [k^hej] ‘when’:



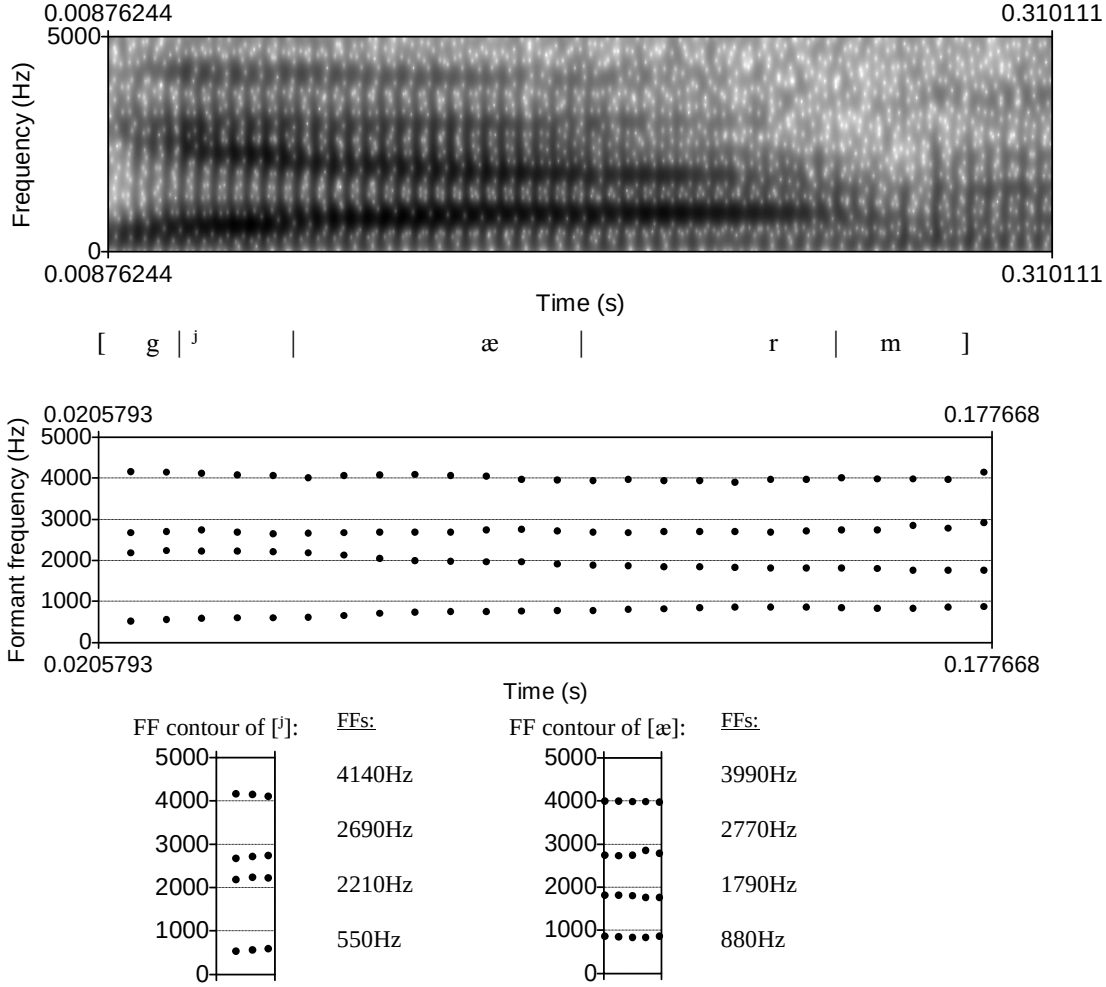
[k | ^{h} | e | j]

<u>IV.E.2 Iranian Phonology: Palatalization</u>

Palatalization is observed in both Iranian and Tajik, but crucially not in Dari. Tajik and Iranian palatalization differ in that the former is borrowed from Russian (and will be analyzed in the chapter that follows the present one), and the latter is phonologically conditioned. In Iranian, the following defines the phonological parameters for palatalization:

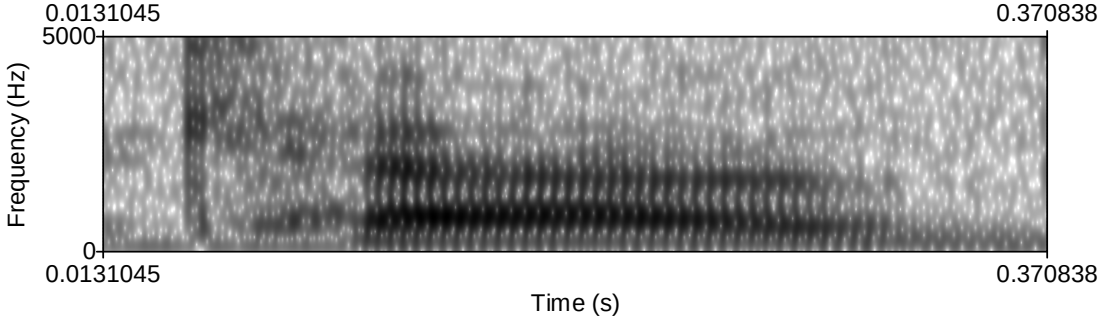
C[+Velar] → C[+Velar]^j / __V[+front –high]

- example: “گرم” [gʲærm] ‘warm’:



This palatalization also occurs with voiceless velar stops:

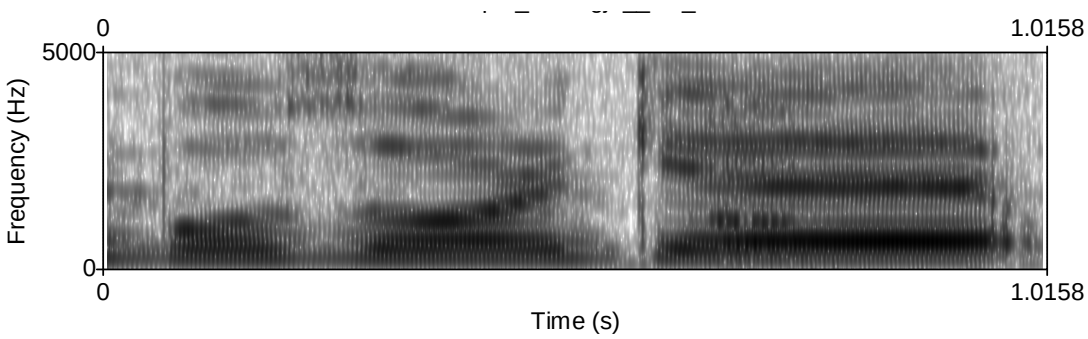
- example: “کرد” [kʲærd˥] ‘did’:



[k | ʰ | ʲ | æ | r | d]

It also occurs in both voiced and voiceless velar stops before the front mid vowel as well as the low front vowel, and even across word boundaries:

- example: “بزرگ است” /bo-zorgˈ#æst/[31] [bo-zorˈ-gʲɛ] ‘is large.’:



[b | o | z | o | r g | ʲ | ɛ]

IV.E.3 Iranian Phonology: h-drop and ʔ-drop

The /h/ and /ʔ/ are optionally deleted in certain highly frequent lexical items, but not universally as in Dari. One of the principal distinctions between Iranian and Dari is that Dari not only deletes nearly every case of /h/ and /ʔ/, but also that there is either

[31] The Iranian 3rd person singular conjugation of the verb ‘to be,’ /æst/ is often rendered [ɛ]. This will be further explored in section IV.5.G.

reduction of surrounding vowels or compensatory lengthening that is associated with some sort of vowel change. In Iranian, the ellipsis results only in some compensatory lengthening if any change at all:

	PP	Iranian:	Dari:	Gloss:
“قهر”	/qahr/	[ɢæhr̩] or [ɢæːr]	[qɑr]	‘angry’
“معلم”	/mu-ʔa-limˈ/	[mo-ʔæ-lɛmˈ]; [mwæ-lɛmˈ]	[ma-lɛmˈ]	‘teacher’

Note that in both of the above cases, Iranian remains more faithful to the initial form preserving the original vowel quality in the first example and preserving two distinct vowels in the second.

The phonological conditioning on h-drop and ʔ-drop in Iranian is fairly straightforward:

/ʔ/ →	∅ / V__#	“ممنوع” /mam-nuʔ/ →	[mam-nu] ‘prohibited’
→	:∅ / __C; C__	“معلوم” /mæʔ-lum/ →	[mæː-lum] ‘evident’
		“دفعه” /dæf-ʔɛ/ →	[dæ-fːɛ] ‘he understood’
/h/ →	:∅ / V__	“شهر” /qæhr/ →	[ɢæːr] ‘city’
→	:∅ / __C; C__	“فهمید” /fæh-mid/ →	[fæː-mid] ‘he understood’
		“صفحه” /sæf-hɛ/ →	[sæ-fːɛ] ‘page’
→	(:)∅ / V__V	“چهل” /ʧɛ-hɛl/ →	[ʧɛl] or [ʧel] ‘forty’

There are two final instances in which h-drop occurs outside of the parameters laid out above, and those are highly frequent tokens “هم” /hæm/ ‘also’ and the plural

suffix “ها” /hɑ/. Aside from frequency, this may also be accounted for by the fact that neither of these lexical items ever occurs in phrase-initial position.

There is left to mention one case in which the ellipsis of /h/ results in more reduction than its Dari counterpart, and that is in the morphology of the verb “خواستن” [xɑs-tæn] ‘to want.’ Recall the Dari conjugation, and compare with Iranian:

	Formal:	Dari:	Iranian:	Gloss:
“می خواهم”	/meˈ-xɑ-ham/	[meˈ-xɑ-jam]	[miˈ-xɑm]	‘I want’

IV.E.4 Iranian Phonology: n-drop

N-drop does not occur in Iranian Persian as in the other two dialects.

IV.E.5 Iranian Phonology: r-drop

Iranian will allow this ellipsis, but less so than Dari. Only three of the Dari r-ellipses are possible in Iranian, and perhaps relevant is the fact that those three are the only disyllabic morphemes allowing the ellipsis in Dari, and in point of fact, they all rhyme:

	Formal Iranian:	Spoken Iranian:	Gloss:
“دیگر”	/di-garˈ/	[di-gɛˈ]	‘other’
“اگر”	/a-garˈ/	[a-gɛˈ]	‘if’
“مگر”	/ma-garˈ/	[ma-gɛˈ]	‘can it be?’

Note in this case that Iranian changes the vowel /a/ to [ɛ] word-finally, further bolstering the earlier claim that Dari and Tajik retain the more conservative /-a/ ending (see section IV.C.3).

IV.E.6 Iranian Phonology: Definite Article Particle "را" (/rɑ/)

The /ɑ/ → [u] alternation addressed in IV.C.5 also occurs with the definite article enclitic /rɑ/ introduced in section III.E.6:

"من خانه را می بینم." /mæn xɑ-nɛ rɑ miˈ-bi-næm/ →[mæn xu-nɛ ru miˈ-bi-næm]
'I house rɑ see-1p-sing' 'I see the house.'

This effect is not universal, and like multiple other observed phenomena, probably subject to frequency or register effects.

Additionally, the /r/ may be elided when the enclitic is postpositional to a word-final consonant:

"من را دید." /mæn rɑ did/ →[mæ-nu did] 'He/she saw me.'
'I rɑ saw-3p-sing'

"آن را دیدم." /ɑn rɑ di-dæm/ →[uˈ-nu di-dæm] 'I saw that.'
'That-rɑ saw-1p-sing'

In these examples, Iranian contrasts with Dari in allowing the /r/-ellipsis and not the /n/-ellipsis, while Dari will disallow the /r/-ellipsis and favor the /n/-ellipsis:

"من را دید." /man rɑ did/ →[ma-ra did] 'He/she saw me.'
'I rɑ saw-3p-sing'

“آن را دیدم.” /ɑ̲n rɑ̲ di-dæm/ →[uˈ-ra diˈ-dam] ‘I saw that.’
‘That-rɑ saw-1p-sing’

In the fascinating case of Tajik, the result is mixed, allowing the /r/-ellipsis and not the /n/-ellipsis in the first case, while disallow the /r/-ellipsis and allowing the /n/-ellipsis in the second case. This is indeed enigmatic, and will be delved into later:

“من را دید.” /man ro̲ did/ →[ma-na did] ‘He/she saw me.’
‘I rɑ saw-3p-sing’

“آن را دیدم.” /o̲n ro̲ di-dam/ →[uˈ-ra diˈ-dam] ‘I saw that.’
‘That-rɑ saw-1p-sing’

IV.E.7 Iranian Phonology: Phonotactics

As mentioned in section III.E.12, Dari optionally permits onset consonant clusters (CCVC), but prohibits consonant + syllabic consonant constructions in the syllable coda (*CVCC̩), while in Iranian, the inverse is true:

*CCVC:

Mod. Pers:	Iranian:	Tajik:	Dari:	Gloss:
“شکم”	[ʃɛ-kæmˈ]	[ʃˈ-kamˈ]	[ʃkam]	‘belly’

CVCC:

Mod. Pers:	Iranian:	Tajik:	Dari:	Gloss:
“فکر”	[fɛkr̩]	[fiˈ-kar]	[fɛˈ-kar]	‘thought’

IV.E.8 Iranian Phonology: /q/ > [ɢ]

The final note on Iranian consonantal phonology concerns the uvular stop [ɢ]. First, it has been noted that this phoneme is derived from two original phonemes that have converged in Iranian:

	Iranian:	Dari:	Gloss:
"غیر"	[ɢejr]	[ɣajr]	'un-'
"قید"	[ɢejd]	[qajd]	'limit'

The resultant phoneme has several allophones that vary according to phonological conditioning:

[ɢ] /	#__	example: [ɢejr] 'un-'
	__ C[+Voice]	example: [bɑɢ-bunˈ] 'gardener'
[ɣ] / V__V		example: [fæ-ɣætˈ] 'only'
[χ] / __ C[-Voice]		example: [væχt] 'time'

Perhaps owing to this allophonic variation, even the most educated of Iranian Persian speakers contend that the graphemes "ق" and "غ" represent divergent sounds in Iranian Farsi, when there is in fact no phonological data to support this claim. Lazard (1992) notes that "in the contemporary Persian of Tehran, it is one and the same phoneme, whatever the origin." (p. 10)

IV.E.9 Iranian Phonology: Verbal Morphology and Possessive Pronouns

Iranian displays a similar phonological reduction to Dari in verbal morphology, but retains the requirement that word-final /-a/ be changed to [-ɛ]. Recall that the normative third person present tense ending in Persian verbs contains /-d/ which is elided in colloquial speech:

/miˈ	+	ko	+	næd/	→	*[miˈ-ko-næ]	→	[miˈ-ko-nɛ]
Prog.		‘do’		3p. sing.		‘he/she does’		
/miˈ	+	ko	+	nænd/	→	[miˈ-ko-næn]		
Prog.		‘do’		3p. pl.		‘they do’		

But, while Iranian does require that word-final /-a/ become [-ɛ], it does not raise the short low vowels in the present perfect verbal morphology and the possessive pronoun enclitics that Dari does:

						Dari:	Iranian:
/kar	-	da	+	am/	→	[ka-demˈ]	[kʲær-dæːmˈ]
‘do-past stem’			+	‘I am’		‘I have done’	
/ba	-	ʧaˈ	#	am/	→	[ba-ʧemˈ]	[bæ-ʧæːmˈ]
‘child’			+	‘my’		‘my child’	

IV.E.10 Iranian Phonology: Third Person Copula “است” (/æst/)

The third person singular conjugation of the verb to be, “است” /æst/ undergoes reduction to a greater and more predictable degree than Dari:

	Iranian:	Spoken Iranian:	Gloss:
“است”	/æst/	→ [st] → *[æ] → [ɛ]	‘is’

It appears that the following describes the colloquial Iranian realization of /æst/:

/æst/ →	[æst] / C#__	“مرد است” /mard#ast/ → [mærˈ-dæst] ‘(It) is a man.’
→	[st] / V#__	“چلو است” /ʧɛ-lo#ast/ → [ʧɛ-lostˈ] ‘(It) is rice.’
/æst/ →	[ɛ] / C#__	“ژاله است” /ʒā-lɛ#ast/ → [ʒɑ-lɛːˈ] ‘(It) is Zhaleh.’
→	[æst] / V#__ (and V → ∅)	“ژاله است” /ʒā-lɛ#ast/ → [ʒɑ-læstˈ] ‘(It) is Zhaleh.’

IV.E.11 Iranian Phonology: “یک” (/jɛk/)

One isolated case of word-final ellipsis occurs in the lexical item “یک” /jɛk/, meaning ‘one.’ This is isolated because there are no other word-final instantiations of /k/ which elide:

“یک” /jɛk/ → /jɛ/, but “کوچک” /ku-ʧɛkˈ/ → /ku-ʧɛkˈ/ (*/ku-ʧɛˈ/)

IV.E.12 Iranian Phonology: “گذاشتن” (/go-zɑʃ-tænˈ/)

Iranian exhibits the reduction /g/ → ∅ in the verb “گذاشتن” /go-zāʃ-tan/ ‘to leave’ when an inflectional prefix is added:

“می گذارم”	/mēˈ-gu-zɑ-ram/	→	/mēgˈ-zɑ-ram/	→	[mi-zɑ-ræm]	‘I leave’
“بگذار”	/biˈ-gu-zār/	→	/bigˈ-zār/	→	[bɛ-zɑrˈ]	‘Leave!’

V. From Proto-Persian to Tajik

V.A Tajik Phonology: Introduction

The Tajik phonemic inventory is also less conservative than Dari, but retains more distinctions than Iranian Persian. The vowel inventories of the three dialects succinctly sum up their degree of conservation: Iranian with six vocalic phonemes, Tajik with seven, and Dari with eight. The most notable distinction that is lost in Tajik is the following:

/ī/ ⟶ /i/
/i/ ↗

Beyond this, Tajik differs from the other dialects in its propensity for phonological borrowing from adjacent languages. Dari and Iranian borrow heavily from the lexical stock of European and Turkic languages, but not directly from their phonologies in the way that Tajik does from Russian (palatalization) and arguably Uzbek (central rounded high vowel).

This chapter will mirror the previous two chapters, introducing the Tajik vowel system in Section V.B, noting the various allophonic and free variations, and treating the individual vowels and diphthongs in Sections V.C.1 – V.C.6 with spectrogram analyses and formant frequency contours. Section V.D will chart the first and second formant frequencies for vowels and diphthongs, and Section V.E will deal with the consonant phonology of Tajik.

V.B Tajik Vocalic Phonemic Inventory

	Front	Central	Back
High:	/i/	/ʉ/	/u/
			/ʊ/
Mid:	/e/		/o/
Low:		/a/	

Diphthongs: /aj/; /oj/; /ʉj/

Tajik is unlike Iranian and Dari in its historical environment of polyglossia. Due to diglossia with Russian for the majority of the twentieth century, and relative proximity to Uzbek, Tajik ideolects vary widely based on the speaker's contact with the aforementioned languages. Take the following alternation:

MP:	Dari:	Iranian:	Tajik:
/ō/	[o]	[u]	[ʉ]

It is often assumed that [ʉ] is a borrowing from Uzbek that allows Tajik to retain the high back distinction that is lost in Iranian. A Northern Tajik who also speaks Uzbek, and has little contact with Southern Tajiks or Afghans will invariably pronounce the above vowel [ʉ], while a southern Tajik will display random variation between [ʉ] and [o], or even [u]. This does not take for granted that Tajik has borrowed the non-back rounded vowel from Uzbek, but rather asserts the possibility, and forwards the suggestion that Uzbek phonology provides and environment in diglossia that fosters retention of non-back rounded vowels. That being said, this paper will focus on the northern dialect, and

assume that exceptions are unpredictable. Furthermore, in section VI, an OT analysis will be provided for a phonologically based approach to the shift [o] → [ʉ] so that either case, borrowing from Uzbek or phonological shift, that wholesale shift will be accounted for.

It is assumed with respect to Middle Persian data and the present system that Tajik retains at least one vocalic phoneme in each of the original five Middle Persian vowel positions:

MP:		Dari:
/ī/ - /i/	→	/i/
/ū/ - /u/	→	/u/ - /ʊ/
/ā/ - /a/	→	/o/ - /a/
/ē/	→	/e/
/ō/	→	/ ʉ /

The following minimal pairs will demonstrate the constrastive status of the above vowels:

/i/	"شیر"	[ʃir]	'milk'
/e/	"شیر"	[ʃʲer]	'lion'
/u/	"پول"	[pul]	'money'
/ʊ/	"پل"	[pʊl]	'bridge'
/o/	"پال"	[pol]	'search'
/u/	"طول"	[tul]	'length'
/ʉ/	"تول"	[tʉl]	'to weigh'

V.C Tajik Vowels Illustrated

As in the previous two chapters, examples will be provided of the vowel's occurrence in initial, medial and final position for each vowel where available in the lexicon. Then a spectrogram analysis of each vowel will be given followed by a chart of the vowel's first four formant frequencies.

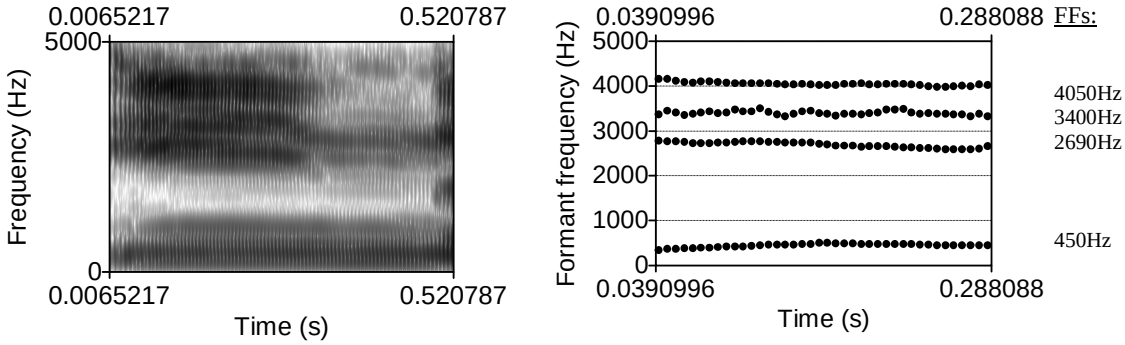
V.C.1 High Front Vowel: [i]

The Tajik high back vowel [i] has a first formant frequency range of 240 to 470 Hz and a second formant frequency range of 2250 to 2650 Hz. As mentioned above, Tajik does not retain the original distinction between long and short vowels /ī/ and /i/ as Dari and Iranian do, and so examples will be included below from both original long and short vowels in Modern Persian. It was also mentioned above that a Southern Tajik might optionally retain Dari-like pronunciation, and so it bears mentioning that the Tajik [i] that is a realization of the short high front vowel may be rendered [ɛ] as in Dari:

	Iranian:	Dari:	Tajik:	Gloss:
"دل"	[dɛl]	[dɛl]	[dil] or [dɛl]	'heart'

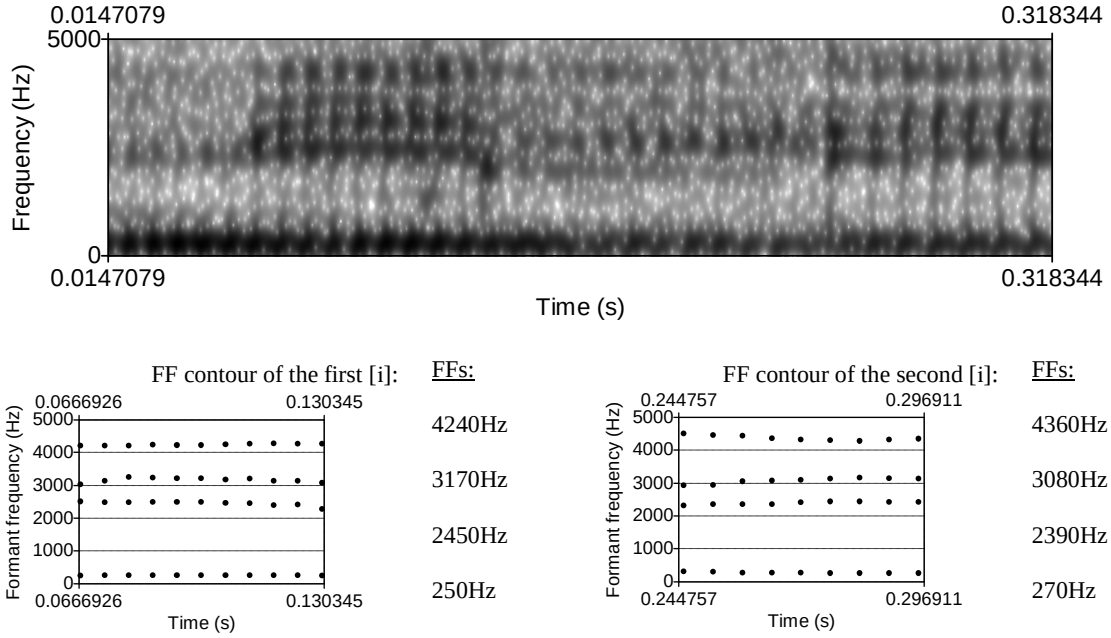
[i] "این" [in] 'this' "سن" [sin] 'age' "فنی" [fa-niˈ] 'scientific'

- example: “ин” / “این” [in] ‘this’:



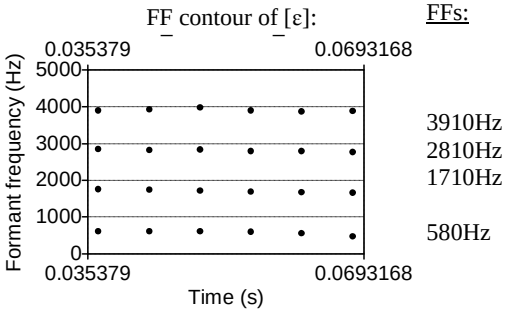
Below is another example illustrating both the original short vowel /i/ and the normative adjectival derivative ending (or “ezafe”) /-i/:

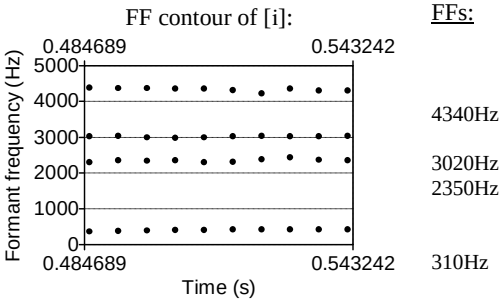
- example: “милӣ” / “ملی” [mi-liˈ] ‘national’:



Note that they are nearly identical, while in Dari and Iranian, they are distinct:

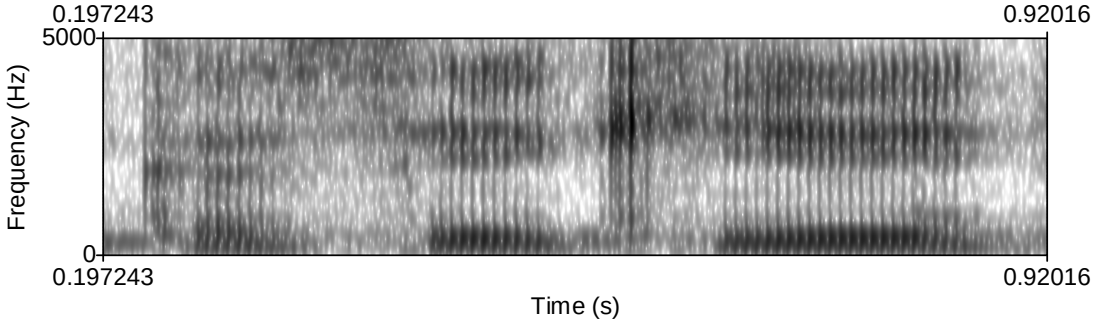
- example (Dari): “ملی” [mɛ-liˈ] ‘national’:

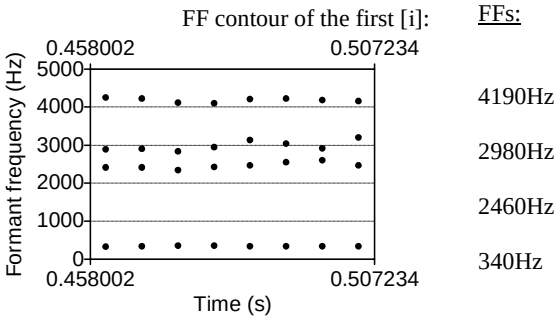


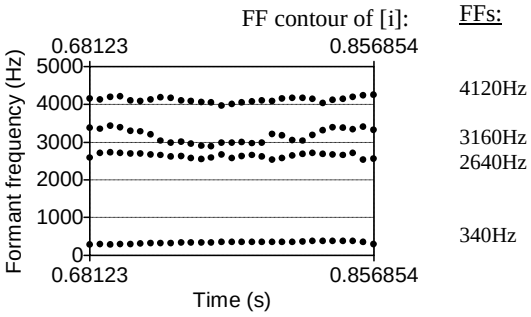


Two other instances of short /i/ that should be analyzed to determine if the convergence was indeed wholesale are the ezafe (explained at length in sections III.C.3 and IV.C.3), and the vowel in the relative pronoun /ki/:

- example: “каси ки” / “کسی که” [ka-si#ki] ‘a person who…’:

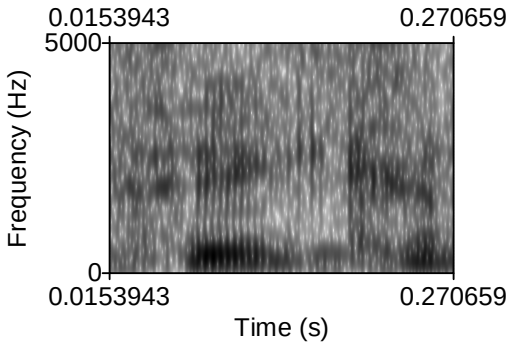


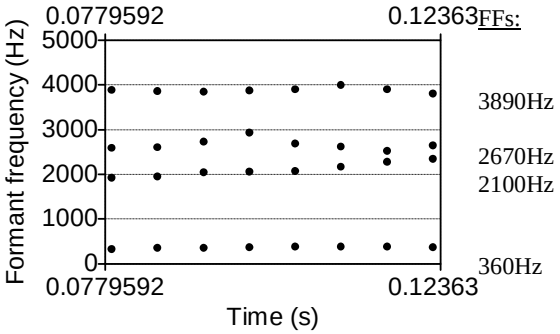




Because in Tajik the high front short vowel remains high (unlike Iranian and Dari), Tajik does not permit this vowel to reduce and centralize:

- example: “фикр” / “فکر” [fiˈ-kər] ‘thought’:



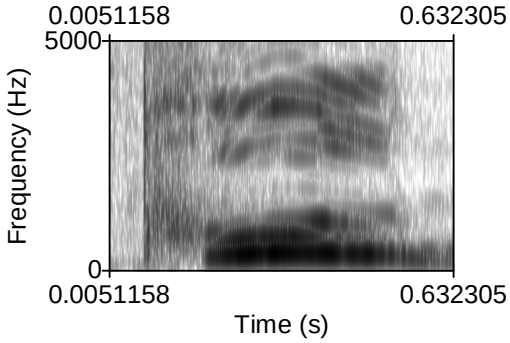


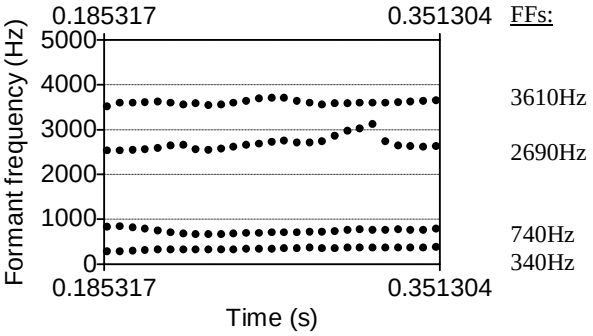
V.C.2 High Back Vowels: [u], [ʊ]

The Tajik high back vowel [u] varies very little in its articulation with a first formant frequency range of 300 to 400 Hz and a second formant frequency range of 690 to 800 Hz.

[u] “умда” “عمده” [um-daˈ] ‘main’ “пул” “پول” [pul] ‘money’ “ту” “تو” [tu] ‘you’

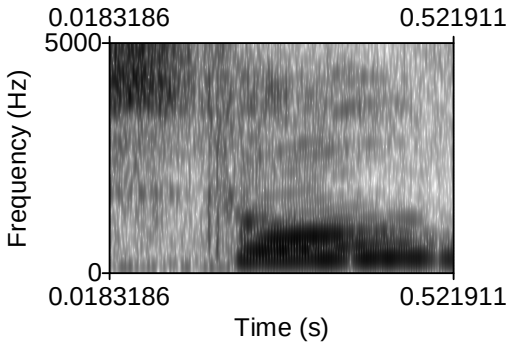
- example: “пул” / “پول” [pʰul] ‘money’:

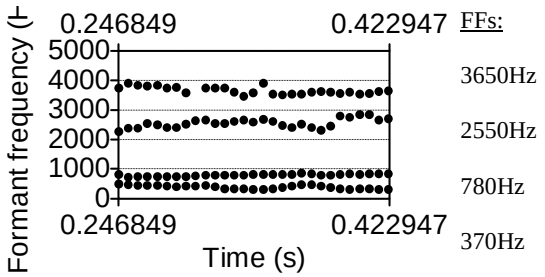




One question that arises is whether or not Tajik vowels are the same is those of Russian loanwords, and the answer is resoundingly that they are the same:

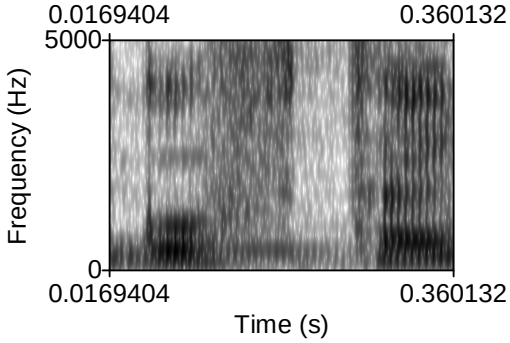
- example: “стул” [stul] ‘chair’:

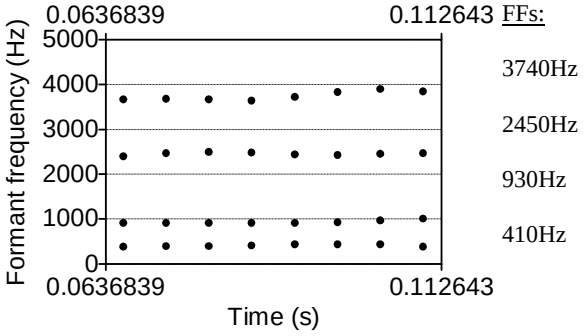




The Tajik high back short vowel [ʊ] has a first formant frequency range of 320 to 420 Hz and a second formant frequency range of 800 to 980 Hz:

- example: “гуфтан” / “گفتن” [gʊf-tanˈ] ‘to say’:



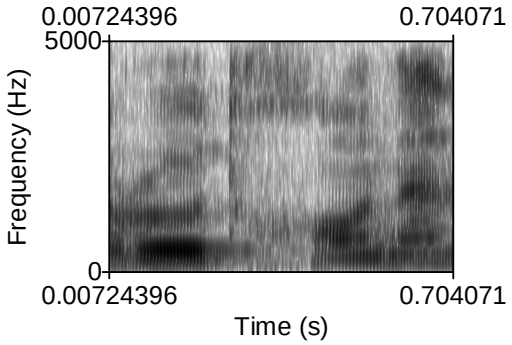


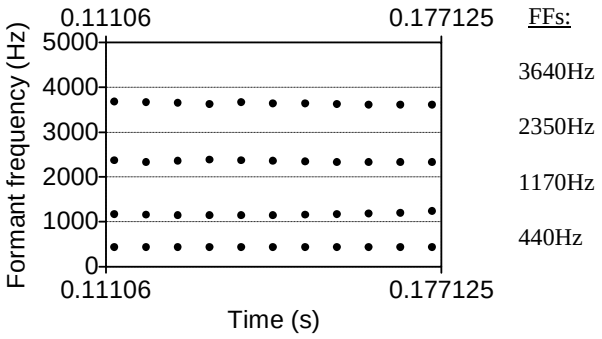
V.C.3 Mid High Vowel: [ʉ]

The mid high vowel [ʉ] is the Tajik realization of the original vowel /ō/, which in Dari remains a mid back vowel and in Iranian has converged with [u]. It is articulated at approximately the same height at its back counterpart (first formant frequency range 350

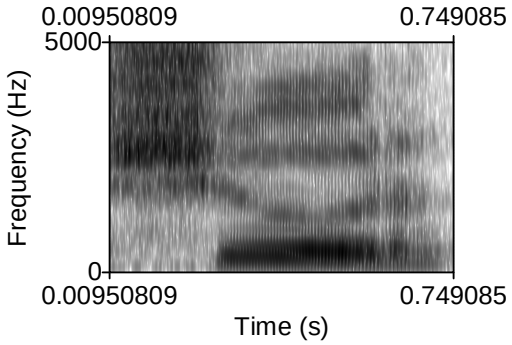
to 470 Hz), but further front, with a second formant frequency range from 1150 to 1530 Hz:

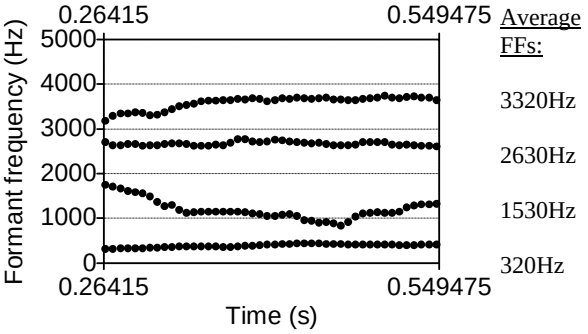
- example: “рӯдхона” / “رودخانه” [rʉd-xo-naˈ] ‘river’:





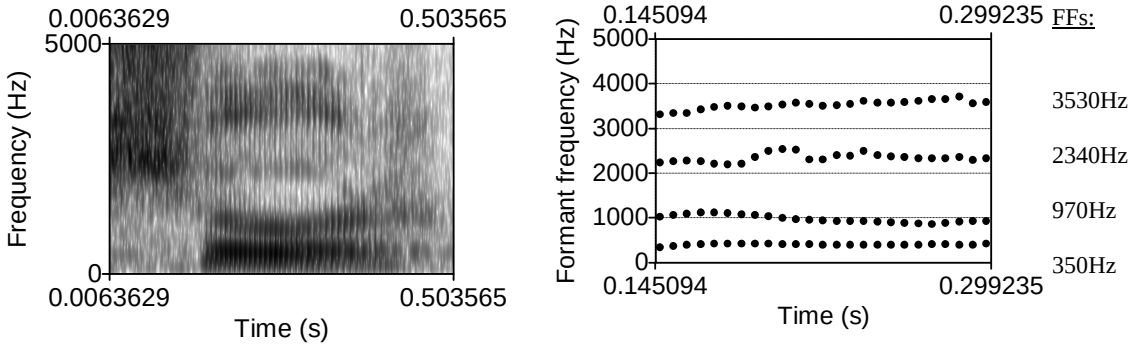
- example: “шӯр” / “شور” [ʃʉr] ‘salty’:





As mentioned above, these vowels may optionally be rendered as [u] or [o], as evidenced by the below, which was sampled from the same subject as the above. It bears noting that samples from Tajik speakers who consistently produced southern speech were deemed too close to Dari in their dialect and were subsequently not used in this study. In the end, it is understood that these three languages exist on a gradient phonologically, and hence an intermediate step should have predictable convergence on consistent forms, and in fact it does:

- example: “шур” / “شور” [ʃur] ‘salty’:

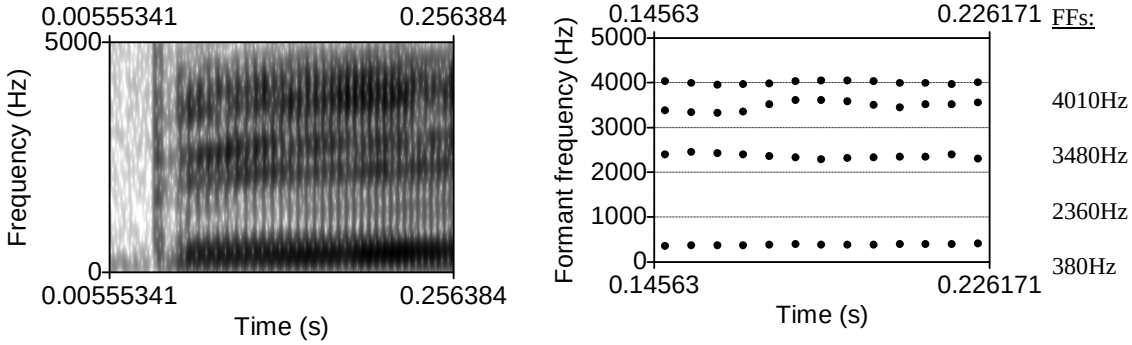


V.C.4 Mid Front Vowel: [e]

The Tajik mid front vowel [e] has a first formant frequency range from 370 to 560 Hz and a second formant frequency range from 2000 to 2400 Hz. Because the Russian front mid vowel is almost always palatalized in its onset, Tajik optionally palatalizes in the same fashion.

[e] n/a “шер” “شیر” [ʃer] ‘lion’ “монеъ” “مانع” [mo-neˈ] ‘prohibited’

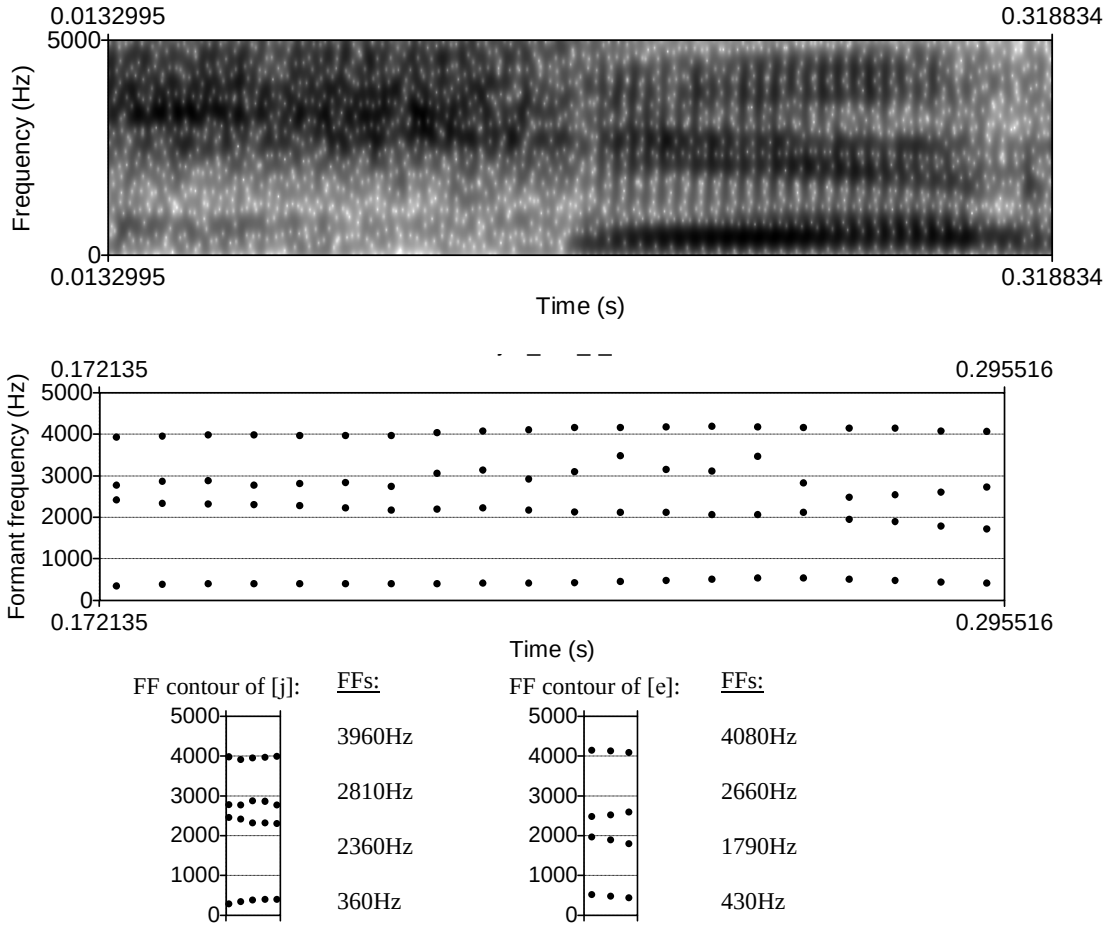
- example: “بی” [be] ‘without’:



Onset palatalization is most prevalent in the tense front mid vowel when preceded by alveolar and post-alveolar segments. This palatalization seems optional though, as [se]

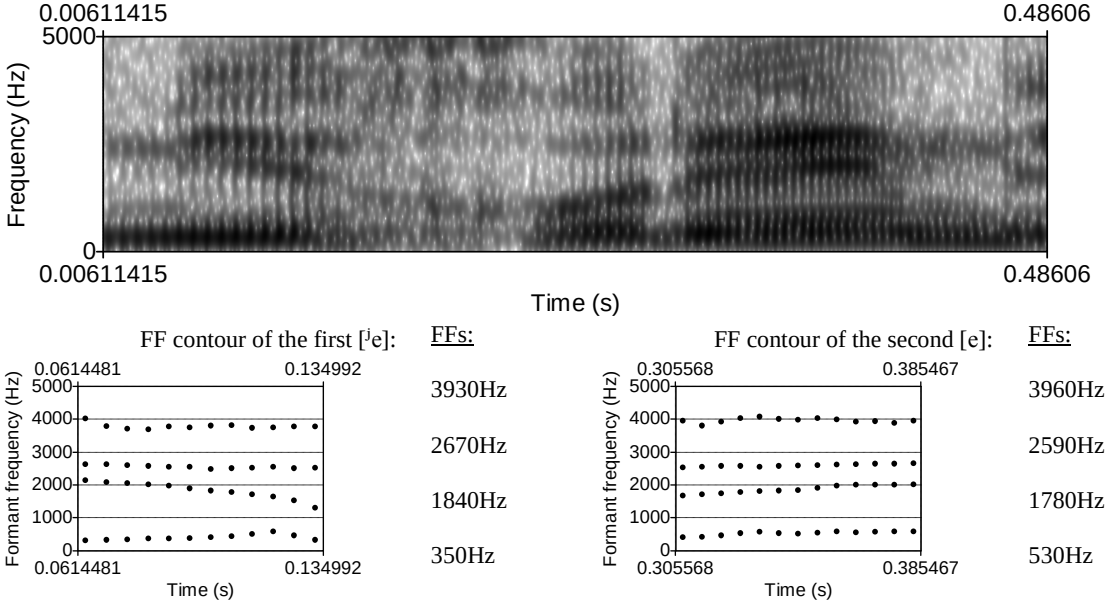
(‘three’) above contains a steady vowel and [ʃʲer] (‘lion’) contains a palatalized consonant:

- example: “شير” [ʃʲer] ‘lion’:



A systematic example of the front mid vowel [e] is in the verbal progressive prefix /me-/, and even this can be, and typically is, palatalized:

- example: “мехурем” / “می خوریم” [mʲe-xʉ-remˈ] ‘we eat’:

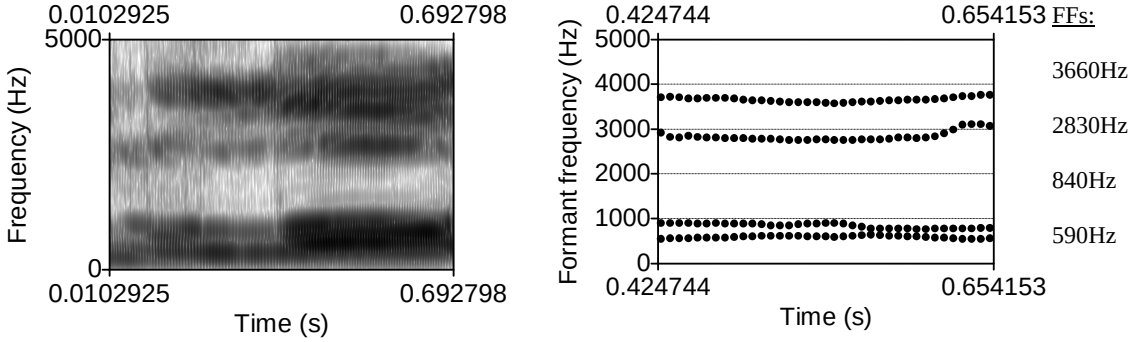


V.C.5 Mid Back Vowel: [o]

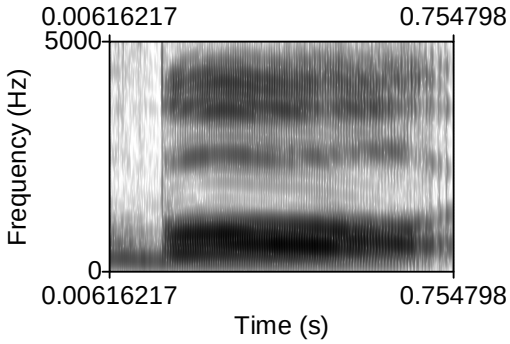
The mid back vowel [o] varies in height from a first formant frequency range of 420 to 550 Hz and in backness from 720 to 940 Hz:

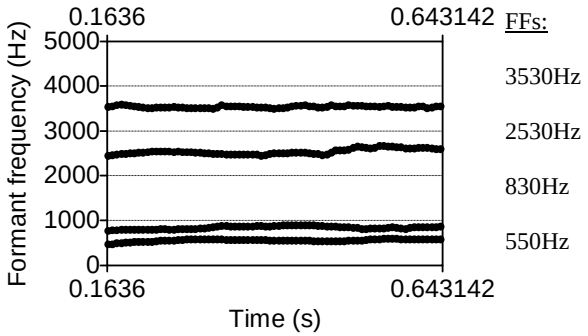
[o] “об” “آب” [ob] ‘water’ “рост” “راست” [rost] ‘right’ “то” “تا” [to] ‘until’

- example: “мулло” / “ملا” [mʊ-loˈ] ‘mullah’:



- example: “бо” / “با” [bo] ‘with’:



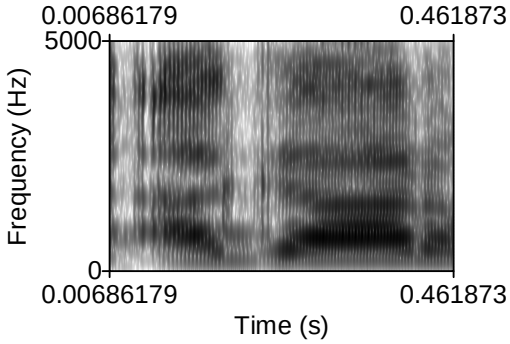


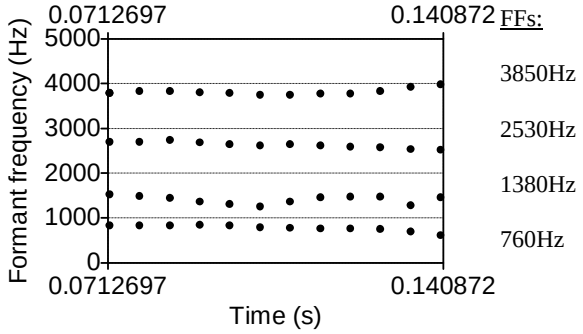
V.C.6 Low Vowel: [a]

The Tajik low vowel [a] has a first formant frequency between 630 and 850 Hz and a second formant frequency between 1200 and 1550 Hz:

[a] “аз” “از” [az] ‘from’ “сар” “سر” [sar] ‘head’ “карда” “کرده” [kar-daˈ] ‘done’

- example: “арар” “اگر” [a-garˈ] ‘if’:



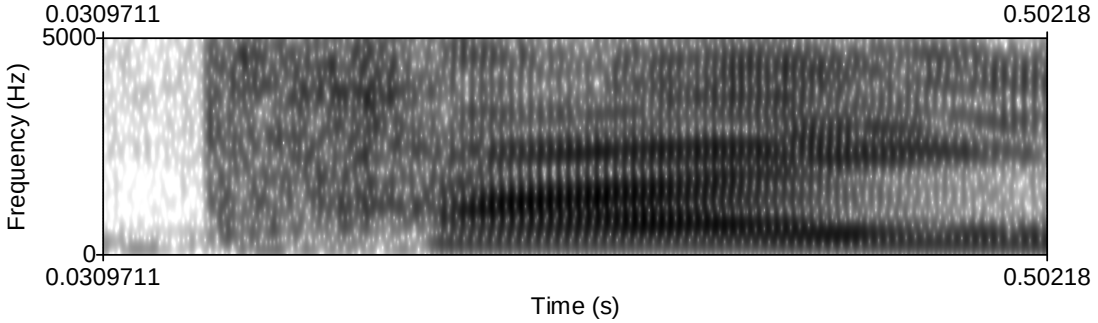


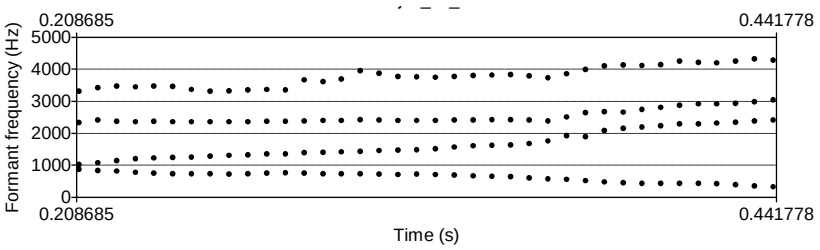
V.C.7 Diphthongs: [aj] [oj] [ʉj]

The back diphthongs /aw/ and /ɑw/ became [av] and [ov] respectively, and so have ceased to be diphthongs and become vowel-consonant combinations. Thus only front rising diphthongs remain:

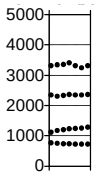
[aj] “айнак” [aj-nakˈ] ‘glasses’ “байн” “بین” [bajn] ‘between’ “тай” “طی” [taj] ‘during’

- example: diphthong [aj] from “тай” “طی” [taj] ‘during’:



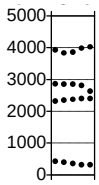


FF contour of [a]:



FFs:
3330Hz
2340Hz
1210Hz
730Hz

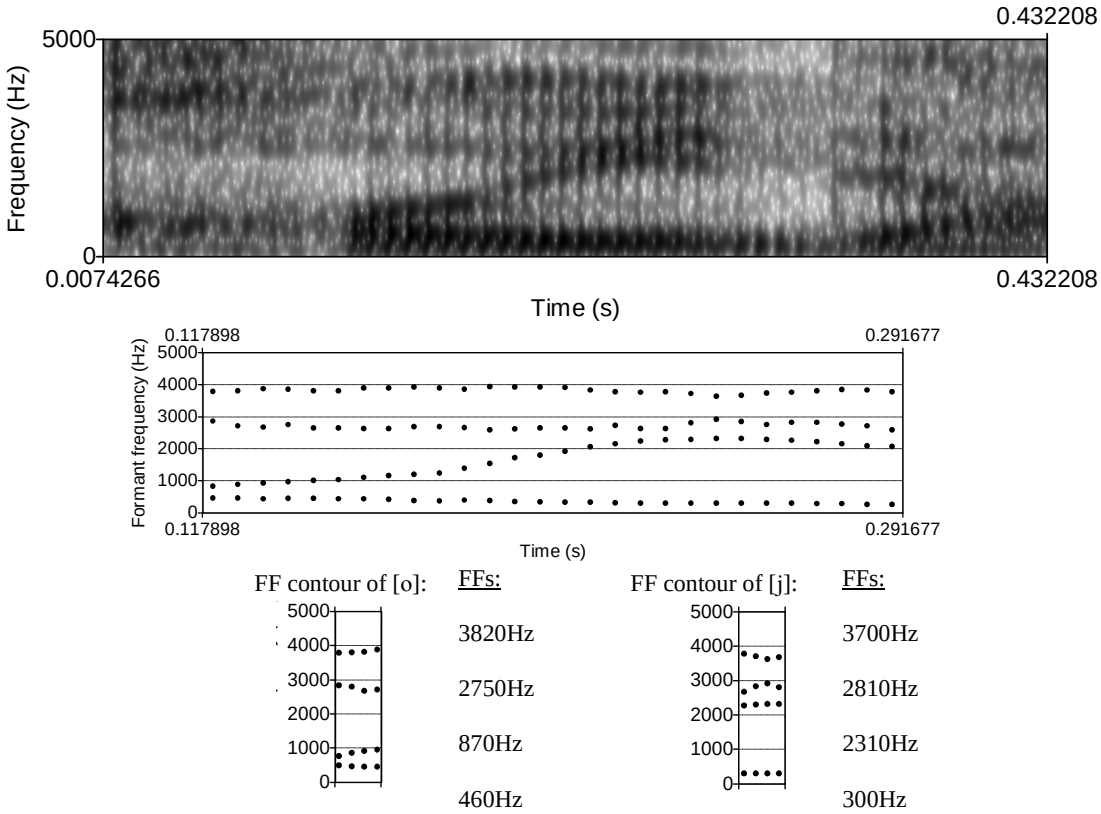
FF contour of [j]:



FFs:
3920Hz
2800Hz
2370Hz
360Hz

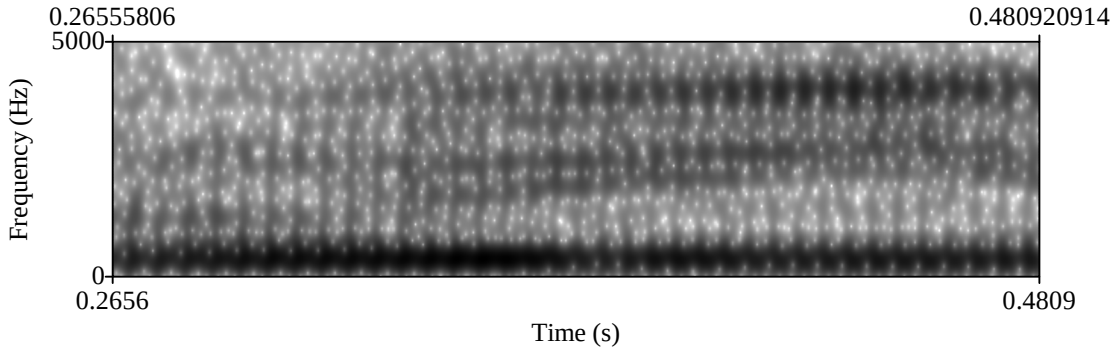
[oj] “ойна” “آینه” [oj-naˈ] ‘mirror’ “мойл” “مایل” [mojl] ‘willing’ “рой” “رأی” [roj] ‘vote’

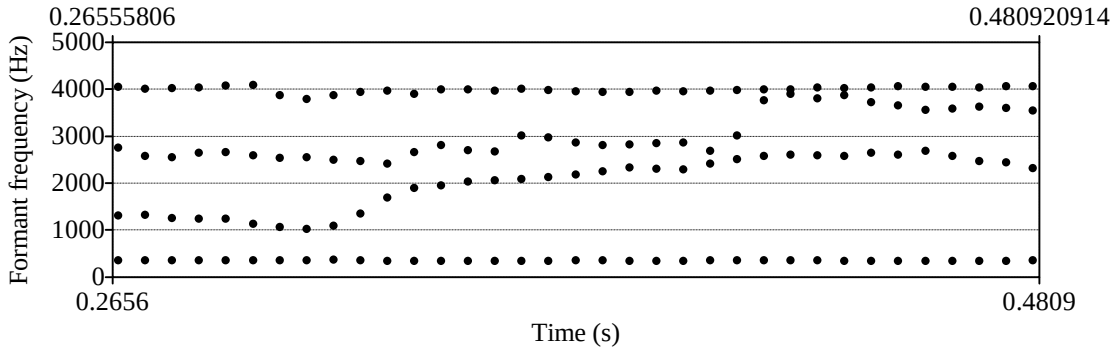
- example: diphthong [oj] from “қойда” “قاعده” [qojˈ-da] ‘base’:



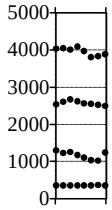
[ɯj] n/a “бӯйнок” “بویناک” [bɯj-nok] ‘smelly’ “мӯи” “موی” [mɯj] ‘hair’

- example: [uj] from “بوی” [buj] ‘smell’:





FF contour of [ʉ]:



FFs:

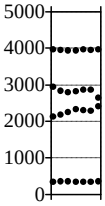
3890Hz

2500Hz

1230Hz

350Hz

FF contour of [j]:



FFs:

3960Hz

2650Hz

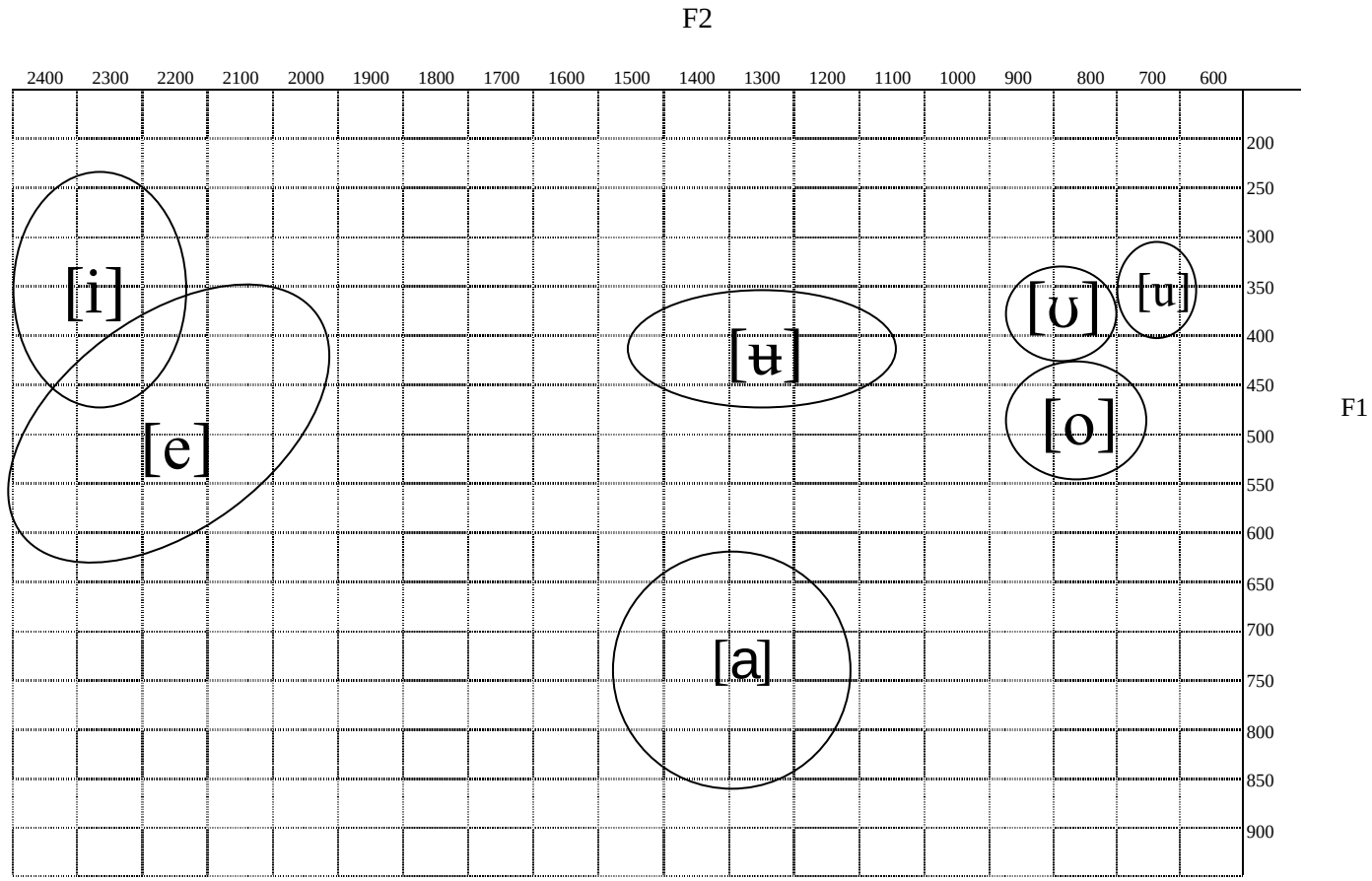
2410Hz

350Hz

V.D Tajik Vowels and Diphthongs: Overview

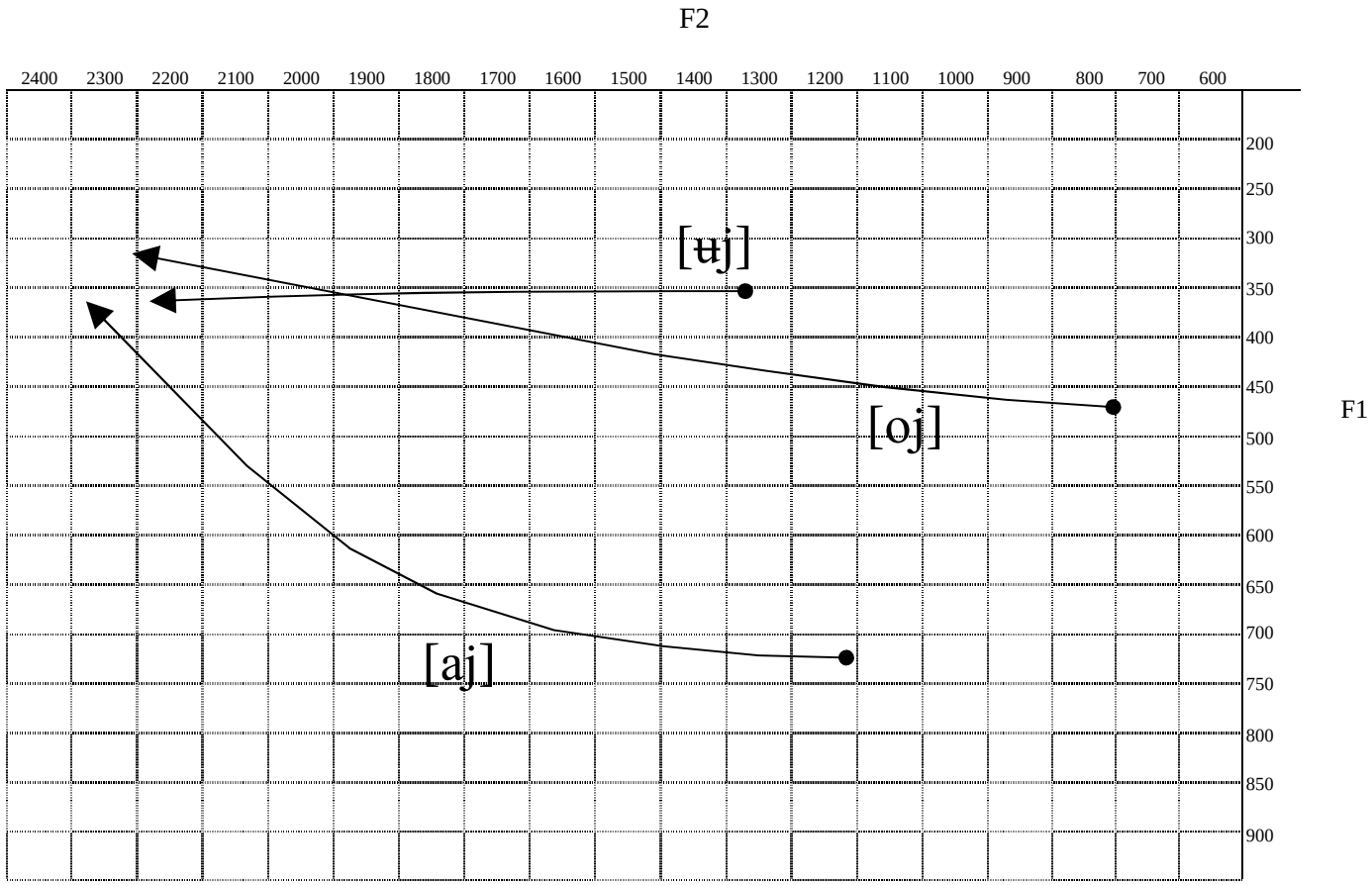
V.D.1 Vowel Overview

V	F1:	F2:	V	F1:	F2::
[i]	240	2250	[u]	300	690
	470	2650		400	800
			[ʊ]	320	800
				420	980
[e]	370	2000	[ʉ]	350	1150
	560	2400		470	1530
[a]	630	1200	[o]	420	720
	850	1550		550	940


F2
2400 2300 2200 2100 2000 1900 1800 1700 1600 1500 1400 1300 1200 1100 1000 900 800 700 600
200 250 300 350 400 450 500 550 600 650 700 750 800 850 900
F1
[i]
[e]
[ɨ]
[a]
[ʊ]
[u]
[o]

V.D.2 Diphthong Overview

	V_1	F1:	F2:	V_2	F1:	F2:
[aj]	[a]	730	1210	[j]	360	2370
[oj]	[o]	460	870	[j]	300	2310
[ʉj]	[ʉ]	350	1350	[j]	350	2410


F2
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
F1
[ʉj]
[oj]
[aj]

V.E Tajik Consonants: Introduction

With two minor exceptions, the Tajik consonantal phonemic inventory is the same as that of Dari. Tajik, like Dari, retains the original distinction between velar fricative /ɣ/ and uvular stop /q/. Tajik, like Iranian, changes /ʋ/ to [v], and it additionally borrows [ɫ] from Russian as an optional realization of /l/:

	Bilabial	Labio Dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glottal
Nasal	m		n						
Plosive	p b	f v	t d				k g	q	ʔ
Fricative				s z	ʃ ʒ		x ɣ		h
Affricate					ʧ ʤ				
Approximate						j			
Trill			r						
Lateral			l (ɫ)						

Below is a list of exemplary lexical items containing each consonant of Tajik in initial, medial, and final position, where possible:

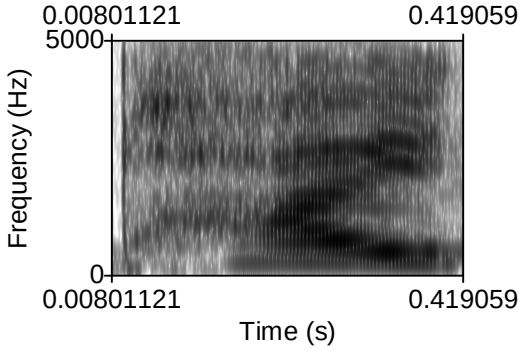
[b]	“болo” “بالا” [bo-ɫoˈ] ‘above’	“обӣ” “آبی” [o-biˈ] ‘blue’	“об” “آب” [ob] ‘water’
[p]	“панҷ” “پنج” [panʤ] ‘five’	“кападан” [ka-pa-danˈ] ‘to catch’	“тӯп” “توپ” [tʉp] ‘ball’
[f]	“фикр” “فکر” [fiˈ-kər] ‘thought’	“офтоб” “آفتاب” [oftob] ‘sun’	“соф” “صاف” [sof] ‘clear’
[v]	“ва” “و” [va] ‘and’	“ҳаво” “هوا” [ha-voˈ] ‘air’	“гов” “گاو” [gov] ‘cow’
[m]	“мо” “ما” [mo] ‘we’	“ама” “عمه” [a-maˈ] ‘uncle-mat.’	“кам” “کم” [kam] ‘little’
[d]	“дар” “در” [dar] ‘in’	“садо” “صدا” [sa-doˈ] ‘voice’	“бод” “باد” [bod] ‘wind’
[t]	“тавр” “طور” [tavr] ‘way’	“ҳато” “حتی” [ha-toˈ] ‘even’	“соъат” “ساعت” [sot] ‘hour’
[z]	“зад” “زد” [zad] ‘hit-past’	“озод” “آزاد” [o-zodˈ] ‘free’	“миз” “میز” [miz] ‘table’
[s]	“син” “سن” [sin] ‘age’	“аст” “است” [ast] ‘is’	“кас” “کس” [kas] ‘person’
[n]	“ном” “نام” [nom] ‘name’	“ин” “این” [in] ‘this’	“ман” “من” [man] ‘I’
[r]	“рӯз” “روز” [rʉz] ‘day’	“дард” “درد” [dard] ‘pain’	“дур” “دور” [dur] ‘far’
[l]	“лаб” “لب” [lab] ‘lip’	“ҳолo” “حالا” [ho-loˈ] ‘now’	“пул” “پول” [pul] ‘money’
[ʤ]	“ҷон” “جان” [ʤon] ‘body’	“кучo” “کجا” [kʊ-ʤoˈ] ‘where?’	“тоҷ” “تاج” [toʤ] ‘crown’
[ʧ]	“чой” “چای” [ʧoj] ‘tea’	“бача” “بچه” [ba-ʧaˈ] ‘child’	“мач” “مچ” [maʧ] ‘match’

[ʒ]	“жола” “ژاله” [ʒo-laˈ] ‘frost’	“вижа” “ویژه” [vi-ʒaˈ] ‘special’	“этаж” [e-taʒˈ] ‘storey’
[ʃ]	“шӯр” “شور” [ʃʉr] ‘salty’	“пашм” “پشم” [pa-ʃəmˈ] ‘wool’	“ош” “آش” [oʃ] ‘soup’
[j]	“як” “یک” [jak] ‘one’	“поён” “پایان” [po-jonˈ] ‘below’	“пай” “پی” [paj] ‘behind’
[g]	“гуфт” “گفت” [gʊft] ‘said’	“агар” “اگر” [a-garˈ] ‘if’	“саг” “سگ” [sagˠ] ‘dog’
[k]	“кор” “کار” [kor] ‘work’	“фикр” “فکر” [fiˈ-kər] ‘thought’	“пок” “پاک” [pok] ‘pure’
[ɣ]	“ғам” “غم” [ɣam] ‘sorrow’	“оғо” “آقا” [o-ɣoˈ] ‘sir’	“друғ” “دروغ” [dʊ-ruɣˈ] ‘lie’
[x]	“хурд” “خرد” [xʊrd] ‘small’	“пухт” “پخت” [pʊxt] ‘cooked’	“ях” “یخ” [jax] ‘ice’
[q]	“қайд” “قید” [qajdˠ] ‘cancel’	“ақл” “اقل” [aˈ-qəl] ‘wisdom’	“ҳақ” “حق” [haq] ‘right’
[h]	“ҳафт” “هفت” [haft] ‘eight’	“қаҳр” “قهر” [qaˈ-hər] ‘angry’	“роҳ” “راه” [roh] ‘way’
[ʔ][32]	“акс” “عکس” [ʔaks] ‘picture’	“идъо” “ادعا” [id-ʔoˈ] ‘claim’	n/a

V.E.1 Tajik Phonology: Aspiration

Tajik, like Dari and Iranian, exhibits a great deal of non-contrastive aspiration in voiceless plosives:

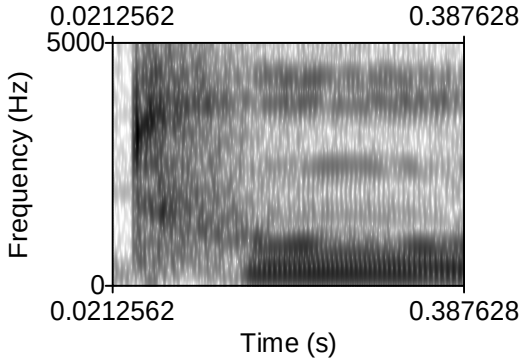
- example: “пай” “پی” [pʰaj] ‘behind’



[p| ʰ | a j]

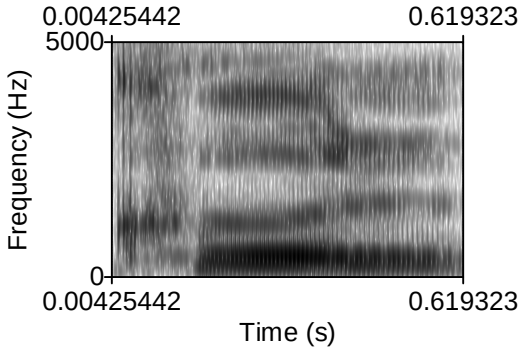
[32] The glottal stop is not contrastive in pre-vocalic word-initial position, and only ever pronounced in formal register. It is not a native Persian occurrence, but rather a product of the convergence of [ʔ] and [ʕ] in original Arabic tokens. Even in that environment, the glottal stop is typically elided.

- example: “ту” “تو” [t^hu] ‘you’



[t | ^{h} | u]

- example: “кӯл” “کل” [k^hʉl] ‘all’



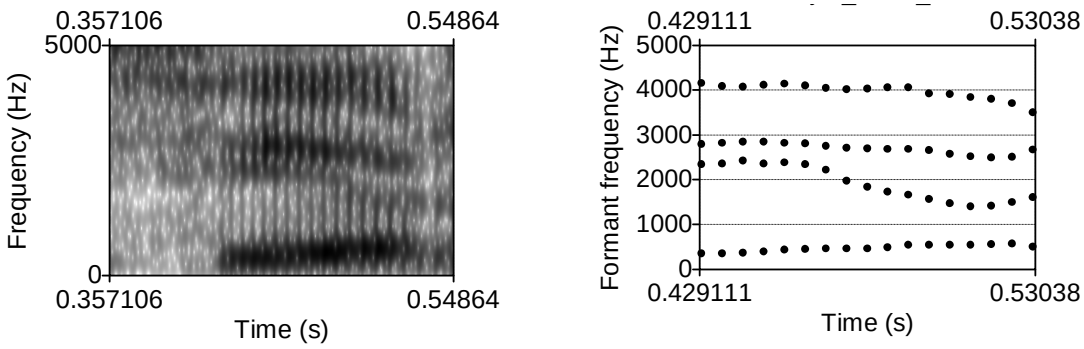
[k| ^{h} | ʉ | l]

V.E.2 Tajik Phonology: Palatalization

Tajik palatalization, unlike that of Iranian, is borrowed from Russian, as is apparent from the fact that it is phonetically identical to Russian palatalization, and occurs fairly systematically at the onset of the mid front vowel [e] and those instances of the high front vowel [i] that were derived from short /i/ ([ɛ] in Dari). Russian also realized a wholesale palatalization at the onset of /e/, which English only annunciates

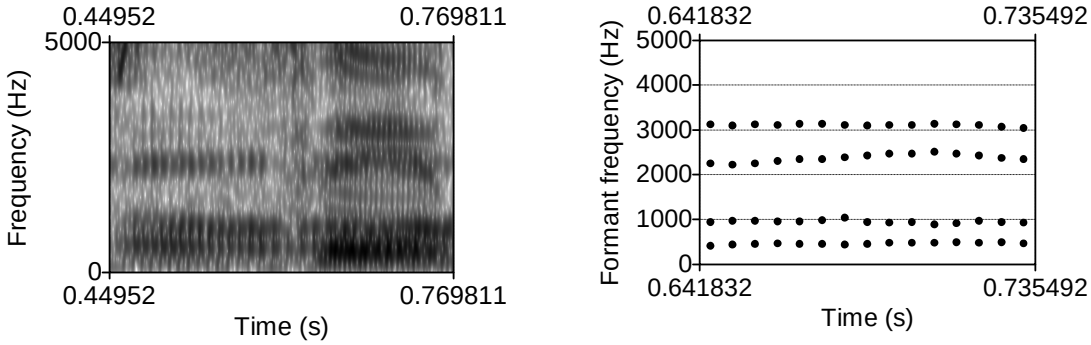
word-initially when articulating Russian names such as “Boris Yeltsin” and “Yevgeni Onegin.”[33] In fact, as a side note, Tajiks speaking English sound Slavic rather than Persian, due in part to the “Russian-ness” of their palatalization:

- example: “сер” / “سیر” [sʲer] ‘satiated; full’:



The palatalization is made clear by noting the drop in the second formant frequency, and also by comparing how stable that same lexical item is in Dari:

- example (Dari): “سیر” [ser] ‘satiated; full’:

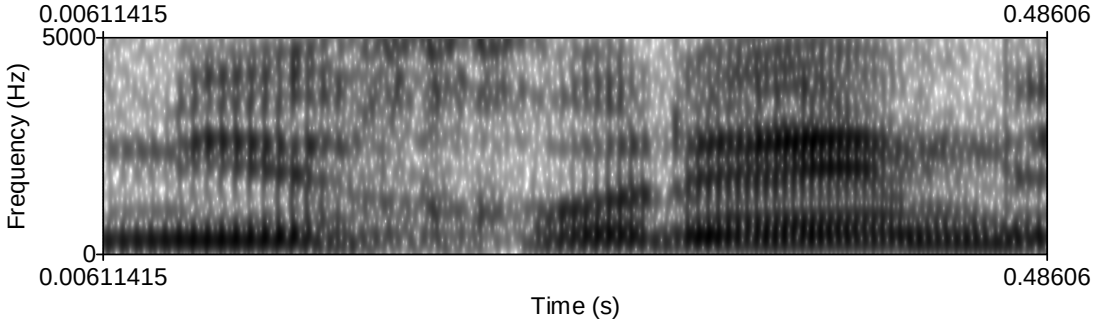


The following example demonstrates not only that Tajik palatalizes in non-Russian lexical items (like the above), but that even original Persian morphemes with no semantic

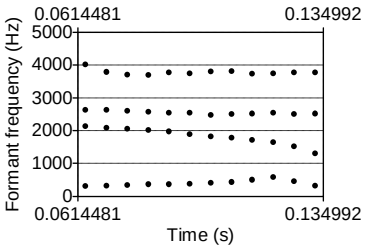
[33] It bears noting that in this particular example, ‘Евгений Онегин’ may be optionally rendered “Yevgyeni Onyegin,” demonstrating precisely the wholesale palatalization under consideration.

content may be palatalized. The example involves palatalization in the imperfect verbal prefix /me-/ as exemplified above:

- example: “мех□рем” / “می خوریم” [mʲe-xʉ-remˡ] ‘we eat’:



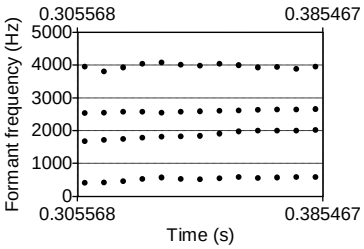
FF contour of the first [ʲe]:



FFs:

3930Hz
2670Hz
1840Hz
350Hz

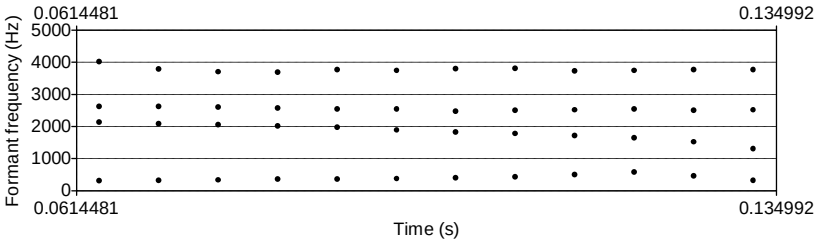
FF contour of the second [e]:



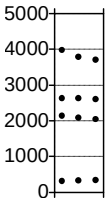
FFs:

3960Hz
2590Hz
1780Hz
530Hz

First Formant contour looks like a palatalized [ʲe] like Russian



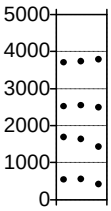
FF contour of [ʲ]:



FFs:

3780Hz
2610Hz
2070Hz
330Hz

FF contour of [e]:

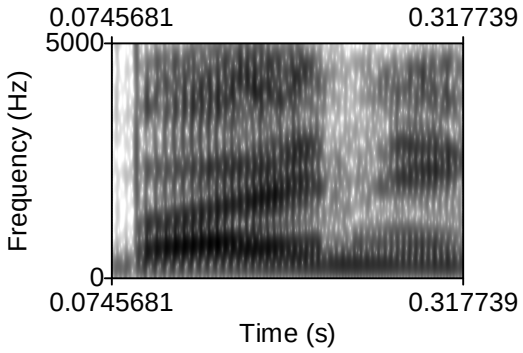


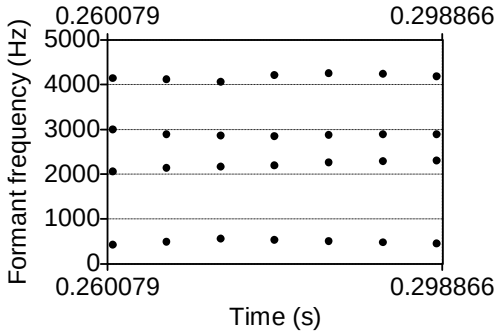
FFs:

3790Hz
2500Hz
1430Hz
420Hz

Palatalization also occurs in the ezafe (which has risen to [i] in Tajik):

- example: “байни” / “بین” [bajˈ-nʲi] ‘between’:





There is room for further research here, as not all instances of short /i/ ([i] in Tajik, [ɛ] in Dari and Iranian) allow palatalization, while all instances of /ē/ ([e] in Dari and Tajik, [i] in Iranian) do allow it, but this was hinted at above and will be returned to in detail in chapter VI. At risk of giving away the conclusion, it will be shown that historical length will play a role in the environmental conditioning of this palatalization.

V.E.3 Tajik Phonology: h-drop and ʔ-drop

The /h/ and /ʔ/ are deleted more frequently in Tajik than in Iranian, but not as frequently as in Dari, in which these sounds are almost never pronounced at all in colloquial speech. In Tajik, /ʔ/ → ∅ is without exception in colloquial speech:

			Iranian:	Dari:	Tajik:	Gloss:
“моьалим”	“معلم”	/moʔalim/	[mwæ-lɛmˈ]	[ma-lɛmˈ]	[ma-łimˈ]	‘teacher’
“аньанав□”	“عنعنوی”	/ʔan-ʔa-na-υiˈ/	[æn-ʔæ-næ-viˈ]	[a-na-na-wiˈ]	[a-na-na-viˈ]	‘traditional’

The allowance for the h-drop is phonologically conditioned, but in a much simpler way than its Iranian counterpart:

/h/ → ∅ / __ $C_n]_\sigma$

Thus in Tajik, word- and syllable-initial /h/ will remain while any instance of /h/ in coda position will be elided, and unlike Dari and Iranian, the ellipsis does not result in any type of compensatory lengthening or sound change:

	Proto-Persian:		Iranian:	Dari:	Tajik:	Gloss:
“кахр”	“قهر”	/qahr/	[ɢæ(h)ɾ̩]	[qɑr]	[qar]	‘angry’
“хама”	“همه”	/haˈ-ma/	[(h)æ-mɛˈ]	[aˈ-ma]	[haˈ-ma]	‘all’
“баҳор”	“بهار”	/ba-hɑrˈ/	[bæ-hɑrˈ]	[ba-ɑrˈ]	[ba-horˈ]	‘spring’
“субҳ”	“صبح”	/sʊbh̩/	[sob(h̩)]	[sʊb]	[sʊb]	‘morning’

V.E.4 Tajik Phonology: n-drop

As in Dari, Tajik allows for ellipsis of /n/ in certain frequently used lexical items, but only in the southern dialects. That is, in Southern Tajik, the following are acceptable ellipses, where in the northern dialect, the underlying consonant remains:

	Formal Tajik:	Spoken South Tajik:	Gloss:
“ман” “من”	/man/	[ma]	‘I’
“ин” “این”	/in/	[i]	‘this’
“он” “آن”	/on/	[u]	‘that’

In point of fact, the northern dialect has replaced “он” /on/ entirely with a separate lexical item “вай” /vaj/, a normative formal third person singular pronoun in Dari and Iranian.

Other n-drop related reductions in Southern Tajik will also mirror an earlier step in the Dari reduction, such as the plural demonstrative pronouns:

“инҳо”	“اینها”	/in-hoˈ/ → [in-oˈ] → *[i-joˈ]	‘they- here’
“онҳо”	“آنها”	/on-hɑˈ/ → *[o-noˈ] → [u-noˈ] → *[u-woˈ]	‘they- there’

And those same pronouns as part of locative adverbs ‘here’ and ‘there’ also reduce, as does the lexeme “ҷо” “جا” [ʤo] ‘place:’

“اینجا”	/inˈ-ʤo/ → [inˈ-ʤa] → [iˈ-ʤa]	‘here’
“آنجا”	/onˈ-ʤo/ → [unˈ-ʤo] → [unˈ-ʤa] → [uˈ-ʤa]	‘there’

For all of the above, the same exception applies to the Northern Tajik variety, but recall that the intent of this research is not to provide a complete view of modern Persian dialectology in its geographical distribution, but rather to express the availability of common Persian sound shifts.

V.E.5 Tajik Phonology: r-drop

Tajik allows for the same optional ellipsis of word-final /r/ as Dari does:

		Formal Tajik:	Spoken Tajik:	Gloss:
“дигар”	“دیگر”	/di-garˈ/	[di-gaˈ]	‘other’
“дар”	“در”	/dar/	[da]	‘in’
“агар”	“اگر”	/a-garˈ/	[aˈ-ga]	‘if’
“интавр”	“اینطور”	/inˈ-tavr/	[iˈ-ta][34]	‘like this’
“онтавр”	“آنطور”	/onˈ-tavr/	[uˈ-ta]	‘like that’
“кардан”	“کردن”	/kar-danˈ/	[ka-danˈ]	‘to do’

An interesting side note regarding the lexical item “дар” /dar/ meaning ‘in’ above, is that the reduced form [da] is not only unavailable in the northern dialect, but also occurs postpositionally. This follows the Turkic grammatical structure as laid out below:

Turkish:	“Ben ev-e gidiyorum.” I house-to go.	“O Ankara-da yaşıyor.” He Ankara-in lives.
North Tajik:	“Ман хона ба меравам.” [man#xo-na + ba#me-rʊm] I house+to go.	“Вай Душанбэ дар зиндаги мекуна.” [Vaj#du-ʃan-be + dar#zin-da-gi#me-kʊ-na] He Dushanbe+in life makes.
South Tajik:	“Ман ба хона меравам.” [man#ba#xo-na#me-rʊm] I to house go.	“У дар Душанбэ зиндаги мекуна.” [ʉ#da#du-ʃan-be#zin-da-gi#mʲe-kʊ-na] He in Dushanbe life makes.

[34] These forms, [iˈ-ta] and [uˈ-ta] are only available in the southern dialect, not the northern.

V.E.6 Tajik Phonology: Definite Article Particle “po” ([ro])

Recall that in Dari and Iranian, the initial consonant of the direct object enclitic /rɑ/ will elide when preceded by a consonant, but crucially not when preceded by a vowel. Tajik will mirror Dari in the post-consonantal consonant ellipsis and vowel reduction, but will also allow the consonant to reduce to a palatal semivowel post-vocalically:

Modern Persian:	Formal Dari:	Spoken Dari:	Gloss:
“تو را دید.”	/tu#rɑ#did/	[tuˈ-ra#did]	‘he/she saw you.’
“سگ را دید.”	/sag#rɑ#did/	[saˈ-ga#did]	‘he/she saw the dog.’

Tajik:	Formal Tajik:	Spoken Tajik:	Gloss:
“Тупо дид.”	[tu#ro#did]	[tuˈ-ja#did]	‘he/she saw you.’
“Сагро дид.”	[sag#ro#did]	[saˈ-ga#did]	‘he/she saw the dog.’

V.E.7 Tajik Phonology: Verbal Morphology and Possessive Pronouns

Tajik verbal morphology is like Dari in the normative third person present tense endings:

/me	+	kʊ	+	nad/	→	[m(^{j})e-kʊ-na]
Prog.		‘do’		3p. sing.		‘he/she does’
/me	+	kʊ	+	nand/	→	[me(^{j})-kʊ-nan]
Prog.		‘do’		3p. pl.		‘they do’

V.E.8 Tajik Phonology: Third Person Copula “аст” ([ast])

The third person singular conjugation of the verb to be, “аст” [ast] undergoes reduction, but to a lesser degree than that of Dari and Iranian:

Mod Pers:	Formal Tajik:		Reduced:	Gloss:
“است”	[ast]	→	[as] /C#__ [s] /V#__	‘is’

V.E.9 Tajik Phonology: Phonotactics

Tajik is like Iranian in prohibiting onset consonant clusters (*CCVC) and like Dari in prohibiting coda consonant clusters (*CVCC̣):

*CCVC:

	Iranian:	Tajik:	Dari:	Gloss:
“بلند”	[bo-lænd^{l}]	[ba-łand^{l}]	[bland] or [bə-land^{l}]	‘tall’
“برنج”	[bɛ-rɛnʤ^{l}]	[b^{l}-rinʤ^{l}]	[brɛnʤ] or [bə-rɛnʤ^{l}]	‘rice’
“شکم”	[ʃɛ-kæm^{l}]	[ʃ^{l}-kam^{l}]	[ʃkam] or [ʃə-kam^{l}]	‘belly’

*CVCC̣:

	Iranian:	Tajik:	Dari:	Gloss:
“چشم”	[ʧaʃm]	[ʧa^{l}-ʃəm]	[ʧʃəm] or [ʧɛ^{l}ʃəm]	‘eye’
“فکر”	[fɛkr̥]	[fi^{l}-kər]	[fɛ^{l}-kər] or [fə^{l}-kər]	‘think’
“شکر”	[ʃokr̥]	[ʃu^{l}-kər]	[ʃʊ^{l}-kər] or [ʃu^{l}-kər]	‘thank’

“طفل”	[tɛfl̩]	[tiˈ-fəl]	[tɛˈ-fəl]	‘child’
“قطل”	[ɢætl̩]	[qaˈ-təl]	[qaˈ-təl]	‘murder’
“قدر”	[ɢædr̩] or [ɢæd]	[qa-darˈ]	[qaˈ-dar]	‘amount’
“شغل”	[ʃoɣl̩]	[ʃʊˈ-ɣəl]	[ʃʊˈ-ɣəl]	‘occupation’
“شکل”	[ʃækl̩]	[ʃaˈ-kəl]	[ʃaˈ-kəl]	‘form’
“قهر”	[ɢæhr̩]	[qaˈ-hər] or [qaːr]	[qaˈ-hər] or [qɑr]	‘angry’
“شهر”	[ʃæhr̩]	[ʃaˈ-hər] or [ʃaːr]	[ʃaˈ-hər] or [ʃɑr]	‘city’

Thus to recap:

Dari: CCVC; *CVCC̩

Iranian:*CCVC; CVCC̩

Tajik: *CCVC; *CVCC̩

V.E.10 Tajik Phonology: Dark /l/ (/ɫ/)

Another seemingly unpredictable variation that appears optional and random is the occurrence of the velarized or “dark” /l/: [ɫ]. This occurs almost without exception in Russian lexical items:

“лощадь”	[ɫoˈ-ʃːadʲ]	‘horse’
“молодой”	[ma-ɫa-doj]	‘young’

“пол”	[poł]	‘gender’

It occurs in Tajik as an optional alternation:

“мулло”	“ملا”	[mʊ-łoˈ] or [mʊ-loˈ] ‘mullah’
“баланд”	“بلند”	[ba-łandˈ] or [ba-landˈ] ‘tall’

VI.A OT Analysis: Introduction

VI.A.1 OT Analysis: Introduction to Optimality Theoretic Sound Change

Language change will be thought of as a reflex of constraint reranking within the Optimality Theoretic framework, and OT will be utilized to provide a uniform account of the sound change in progress in Modern Persian. I propose that diachronic phonological change is in fact systematic and subject to the same set of universal, immutable, and violable constraints as is its synchronic counterpart, and hence language change would be more aptly referred to as constraint precedence change.

This paper assumes the reader to be versed in OT and its implications, and insofar as s/he is, a general introduction to the diachronic application of the theory is in order. It is commonly accepted that the mechanism of OT in terms of synchronic change involves a Lexicon, from which lexical inputs are drawn; a Generator, which provides the phonological system with potential outputs; and an Evaluator, which is a set of universal, violable constraints on faithfulness to and markedness of input forms ordered on a language particular basis to determine the proper phonetic output of a particular lexical item or phoneme. The three languages under consideration, for instance, have been established to be offspring of one ancient parent tongue, and so if inter-linguistic distinction is understood as variation in constraint ranking, it must stand to reason that language change must in fact be understood as change in constraint precedence. That being said, this dissertation is a reconstruction and as such a contribution to Persian philology, not a work that seeks to provide phonological theory with yet another model of language change. Indeed this dissertation attempts to present a justifiable picture of the

quantitative restructuring that gave birth to the modern dialects of Persian, and it is OT which does so.

Chapter VI is organized such that vowels are covered first, and then consonants. VI.A.2 will introduce the OT language change model beginning with Proto-Indo-European, and VI.A.3 will introduce the role that length and ATR will play in the OT model of Proto-Persian reconstruction.

VI.A.2 OT Sound Change and the Initial State

By nature, faithfulness constraints are conservative and hence will not be the driving force of language change. The initial state before a language change will then display faithfulness constraints outweighing markedness constraints. Thus, the constraint ranking for Proto-Persian vowels must be assumed to display faithfulness constraints outranking markedness constraints as below:

PIE *a	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	ID-IO(Low)	ID-IO(Rnd)	*[-High]	*[ATR]
☞ a						*	
o				*!	*		*
u			*!		*		*
ā	*!						

PIE *u	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	ID-IO(Low)	ID-IO(Rnd)	*[-High]	*[ATR]
a			*!	*	*	*	
☞ u							
o			*!			*	*
ū	*!						*

PIE *ā	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	ID-IO(Low)	ID-IO(Rnd)	*[-High]	*[ATR]
☞ ā						*	
ō				*!	*	*	*
ū			*!		*		*
a		*!				*	

PIE *ō	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	ID-IO(Low)	ID-IO(Rnd)	*[-High]	*[ATR]
☞ ō						*	*
ā				*!	*	*	
ū			*!				*
o		*!				*	*

Language change will be seen to occur as various markedness constraints are promoted to positions outranking the above faithfulness constraints as the respective ordering of constraints weaken due to the pressures of the constraint grounding: ease of articulation and/or ease of parsing. There follows a period of optionality and a gradual preference for the output generated by the raised constraint. Once the markedness constraints have created a divergent output, and lexicon optimization reinterprets the form in the lexicon, the markedness constraints are no longer relevant and hence are demoted again. Finally there occurs the promotion of a faithfulness constraint or constraints, restoring faithfulness to a high rank and hence preserving the new sound. Whether this occurs inter-generationally or otherwise is again beyond the scope of this research, but the abovementioned period of optionality is clearly seen in cases like the Tajik [o] / [ʉ] alternation and the Dari [ɛ] / [ɪ]. The former is a case of language change that appears to be spreading regionally, and as such, the current status of the change is [ʉ] in northern Tajik dialects and [o] in southern dialects. The latter is a case of random phonetic alternation in the high front lax vowel, probably indicating that Dari, like Iranian will

reinterpret [ɪ] as [ɛ], but the change is still in flux. For the sake of precision, it should be noted that the vocalic phoneme will be [ɛ] with variant [ɪ].

Rather than arbitrarily choosing an example to illustrate, I will allow the language to do this for itself, and when a suitable example arises (see section VI.F.1 of this chapter), the above will be overtly expressed with an example. The case to be covered there is an allophonic variation in Dari, and the theory essentially requires a period of allophonic variation (called optionality above.) Crucially, a universal constraint on language weakens the grammar enough to allow for the allophonic variation that will eventually give way to the new form and the period of variation will end again, as the new form becomes canonical.

VI.A.3 Length Matters

OT aside, this analysis must begin with an introduction to the significance of contrastive length in Indo-European vowels, noting that that to which we traditionally refer as length must certainly constitute either a qualitative or a quantitative distinction or even both, and this detail will prove non-trivial. Contrastive length is evident in the accepted PIE vocalic phonemic inventory:

Short Vowels:	a	e	i	o	u	ə
Long Vowels:	ā	ē	ī	ō	ū	
Diphthongs:	aj	ej	oj	aw	ew	ow
Long Diphthongs:	āj	ēj	ōj	āw	ēw	ōw
Syllabic Consonants:	l̥	m̥	n̥	r̥		

Crucial to this analysis will be the assumption that the length distinction is manifest in both a qualitative and quantitative difference rather than simply a quantitative

one. The basis for this assumption, as typical of the comparative method, is derived from the fact that the attested IE languages demonstrate a notable cross-linguistic variation of both qualitative and quantitative distinctions in that which has historically been considered length. Below are a few examples of modern IE vowel inventories noting pairs traditionally distinguished by length:

American English:	[ɪ, i]	[ɛ, e]	[ʊ, u]	[ɔ, o]	[æ, a], etc.
Dutch:	[ɪ, i]	[ʏ, y]	[ɛ, eː]	[ɔ, oː]	[ɑ, aː], etc.
Hindi:	[ɪ, i]	[ɛ, e]	[ʊ, u]	[ɔ, o], etc.	
Portuguese (Euro):	[ɛ, e]	[ɯ, u]		[ɔ, o], etc.	

Contrast the above with Hausa length distinctions:

[i, iː] [e, eː] [a, aː] [o, oː] [u, uː][35]

One notes immediately that there was certainly some degree of qualitative distinction and likely both a qualitative and quantitative distinction between the contrastive pairs in IE. For the sake of simplicity, the segments in question in the proto-form (Proto-Persian) will still be distinguished by the absence or presence of the long diacritic, but I will assume the distinction to be both qualitative and quantitative as indicated by attested IE. That is, I take proto-forms */ī, i/ to represent phonetic forms resembling [iː, ɪ], */ē, e/ to be [eː, ɛ], */ō, o/ to be [oː, ɔ], and */ū, u/ to be [uː, ʊ]. The case of /ā, a/ will be slightly more complicated, as attested IE indicates that there is no canonical tense/lax pair in the low central position, but that will be dealt with presently.

The phonology of Proto-Persian (PP) will be assumed to be the following:

[35] All five example sets were taken from *Handbook of IPA* (1999).

/ī/ /ū/
/i/ /u/
/ē/ /ō/

/a/
/ā/

As per the above specifications regarding the length distinction, the above may be alternately rendered, as laid out in Chapter I:

/i(ː)/ /u(ː)/
/ɪ/ /ʊ/
/e(ː)/ /o(ː)/

/a/
/ɑ(ː)/

Or

/i/ /u/
/ɪ/ /ʊ/
/e/ /o/

/a/
/ɑ/

In American usage of IPA, /ɑ/ is “for many linguists not distinct from /a/,” (Ladusaw & Pullam, p. 7), but the present course of analysis will assume that /ɑ/ is the tense counterpart of /a/, the latter fronted by its laxness, and the former backed by its ATR quality (the constraints governing these two shifts will be covered in section that follows this one, VI.A.4). That is to say that in this analysis, PP /a/ and /aː/ will be considered

low central vowels distinct for length. The length distinction will bear with it a distinction in ATR, and the ATR distinction will cause a change in frontness. This initial central ±ATR pair would be cumbersome and in fact confusing to represent in IPA (/ɐ/ and /ɐ̞ː/), so I will overtly state the following contrastive features:

	a	ā	æ	e /ɛ	ē	i/ɪ	ī	o/ɔ	ō	u /ʊ	ū	ʉ
high	-	-	-	-	-	+	+	-	-	+	+	+
low	+	+	+	-	-	-	-	-	-	-	-	-
back	-	-	-	-	-	-	-	+	+	+	+	-
front	-	-	+	+	+	+	+	-	-	-	-	-
round	-	-	-	-	-	-	-	+	+	+	+	+
long	-	+	-	-	+	-	+	-	+	-	+	-
ATR	-	+	-	-	+	-	+	-	+	-	+	+

When providing Proto-Persian lexical items they will henceforth be represented with the absence or presence of long diacritic (/ē/ versus /e/), and when referring to any Persian dialect, be it Dari, Farsi, Tajik, or any attested modern language, the appropriate IPA symbol will be used. With these contrastive features in mind, the following are the comparative vowels of Dari, Iranian and Tajik. In the left column are the three phonetic inventories organized by oral cavity position and in the right column the inventories are ordered by their historical length pairs:

<u>Dari:</u>

| | | |

Iranian:

									Long:		Short:
/i/					/u/				/i/		/ɛ/
									/u/		/o/
		/e/[36]			/o/						
			/ɛ/								
		/æ/			/ɑ/				/ɑ/		/æ/

Tajik:

									Long:		Short:
/i/			/ʉ/		/u/				/i/		
					/ʊ/				/u/		/ʊ/
		/e/			/o/				/ʉ/		
									/e/		
			/a/						/o/		/a/

Examples:

Dari/Iranian:	“دل”	“شیر”	“دیر”	“گرگ”	“کون”	“روز”	“تن”	“تا”
Tajik:	“дил”	“шир”	“дэр”	“гург”	“кун”	“руз”	“тан”	“то”
PP:	dil	šīr	dēr	gurg	kūn	rōz	tan	tā
Gloss:	‘heart’	‘milk’	‘long’	‘wolf’	‘arse’	‘day’	‘body’	‘until’
Dari:	[dɛl]	[ʃir]	[der]	[gʊrg]	[kun]	[roz]	[tan]	[tɑ]
Iranian:	[dɛl]	[ʃir]	[dir]	[gorg]	[kun]	[ruz]	[tæn]	[tɑ]
Tajik:	[dil]	[ʃir]	[der]	[gʊrg]	[kun]	[rʉz]	[tan]	[to]

The phonemic inventories of the three dialects indicate that Afghan the dialect, Dari is the most conservative, as it retains maximal phonemic distinction; in fact, it

[36] /e/ is included as a vowel of Iranian, but is not in the long/short columns on the right, because /e/ is in fact the Iranian reinterpretation of the historical diphthong */aj/. This was covered in section IV.C.4. /o/ is included in the short column, as it is both the modern manifestation of the short high back vowel */u/ and the diphthong */aw/. The current status of /ō/ and /ū/ in Modern Persian will warrant its own section, VI.C.2.

appears that Dari retains the entire Middle Persian vocalic inventory[37]. As mentioned in the previous chapter, Dari, like Middle Persian, has eight vocalic phonemes, Tajik has seven, and Iranian has only six. This is due to three convergent sound shifts, one in Tajik and two in Iranian. There are, of course, other sound shifts, but they did not affect the contrastive status of the original Middle Persian phonemic inventory. The below are examples of the three aforementioned alternations:

Tajik: /i/ → /ī/

سر / سیر	Dari: [sɛr] (‘secret’) / [sir] (‘garlic’)
сир / сир	Tajik: [sir] (‘secret’) / [sir] (‘garlic’)[38]
	Iranian: [sɛr] (‘secret’) / [sir] (‘garlic’)

Iranian: /ē/ → /ī/

شیر / شیر	Dari: [ʃer] (‘lion’) / [ʃir] (‘milk’)
шер / шир	Tajik: [ʃer] (‘lion’) / [ʃir] (‘milk’)
	Iranian: [ʃir] (‘lion’) / [ʃir] (‘milk’)

Iranian: /ō/ → /ū/

طول / تول	Dari: [tul] ‘length’ / [tol] ‘weight’
Ту*л* / то*л*	Tajik: [tul] ‘length’ / [tʉl] ‘weight’
	Iranian: [tul] ‘length’ / [tul] ‘weight’

[37] The status of MP short mid vowels /e/ and /o/ is unclear, but the balance of scholarship on MP indicates that /e/ and /o/ do not exist as independent phonemes.

[38] It bears noting that the Tajik lexical item “сир” [sir] (‘garlic’) has been widely replaced in common parlance by the Russian “чеснок,” [tʃʲes-nokˈ], but speakers, and in particular Southern Tajiks will still use “сир” [sir] as an optional alternative item.

Note in the third change that in Tajik, as in Iranian, the mid back vowel undergoes raising, but unlike Iranian it is also fronted to a central position, thus retaining a phonemic distinction. Many Tajik linguists speculate that this phonomenon is borrowed from Uzbek due to the proximity of the two languages, and this assumption is based on the fact that the high central rounded [ʉ] is retained in borrowed Uzbek lexical items (Perry, 2005, pp. 15-16). For three reasons, I will contend that this shift is not the product of borrowing, but rather a phonological shift similar to a much earlier Indo-European shift in Slavic of /ū/ → /ɨ/, which resulted in the modern Russian "ы."[39] First and most importantly, this must be at least considered, as the distribution of [ʉ] is predicatable in comparison with Dari and other historical data. This stands in contrast to the Tajik adoption of dark l ([ɫ]), which is clearly an exogenous development due to the Soviet influence and Russian lexical influx, but this will be covered in depth in section VI.H.9. Secondly, the segment employed by Uzbek is in fact not [ʉ], but either [y]/[ʏ] or [œ]/[ø], and thirdly, borrowing would require no theoretical justification, and so if the theory can account for a phonologically based shift from [o] → [ʉ], than an ad hoc explanation is altogether avoided.

[39] Although it most likely warrants no more than a footnote, it bears mention that this common shift is evidenced in Modern Russian and Persian cognates such as "быть" [bɨtʲ] and "بودن" [bu-danˈ], both meaning 'to be,' and so forth.

VI.A.4 From Middle Persian to Proto-Persian

Developments common to all three dialects will be considered to have occurred before the dialectal split, and in the case of the first, will be seen to effect subsequent changes. The first such change was covered at length above, but must be qualified with a constraint on length and ATR, mandating that long vowels will be +ATR and short vowels will be –ATR. This constraint is grounded in the mutually complementary physiological nature of length and ATR. That is, quantitative length provides the articulator with more time in which to advance the tongue root, and an advanced tongue root accommodates a longer articulation of sound. Furthermore, the distinction in ATR serves to maximize the perceptual difference between short and long vowels.

MAX-IO[αLong αATR] ("If input segment is αLong, output will be αATR.")

Thus, the tableaux for this initial development will simply be the following:

MP *i	αLg αATR	DEP-μ-IO	MAX-μ-IO	ID-IO(Lng)	ID-IO(Hg)	ID-IO(Fnt)	ID-IO(ATR)
i	*!						
☞ ɪ							*
u	*!					*	
ɨ						*!	
a					*!	*	

MP *ū	αLg αATR	DEP-μ-IO	MAX-μ-IO	ID-IO(Lng)	ID-IO(Hg)	ID-IO(ATR)	ID-IO (Low)
i			*!	*			
☞ ū							
ʊ	*!		*	*		*	
~~ʉ~~			*!	*			
ō					*!		

MP *ē	αLg αATR	DEP-μ-IO	MAX-μ-IO	ID-IO(Lng)	ID-IO(Hg)	ID-IO(ATR)	ID-IO (Low)
ɛ	*!		*	*		*	
☞ ē							
ī					*!		
ɪ	*!		*	*	*	*	
ɛ̄	*!					*	

MP *u	αLg αATR	DEP-μ-IO	MAX-μ-IO	ID-IO(Lng)	ID-IO(Hg)	ID-IO(Fnt)	ID-IO(ATR)
u	*!						
☞ ʊ							*
i	*!					*	
~~ʉ~~						*!	
a					*!	*	

MP *ō	αLg αATR	DEP-μ-IO	MAX-μ-IO	ID-IO(Lng)	ID-IO(Hg)	ID-IO(Fnt)	ID-IO(ATR)
ɔː	*!						*
☞ ō							
ū					*!		
ē						*!	
ɔ	*!		*	*		*	

The second sound change common to all dialects is /ā/ → /ɑ/. Note that Tajik subsequently changes /ɑ/ → /o/ as part of a "push-chain" effect, but my analysis assumes /ā/ → /ɑ/ to have occurred first, based on the following constraint on backness of low +ATR segments:

MAX-ATR +Low +Back ("Maximize Backness of Low +ATR segments.")

This constraint is based on ease of parsing and even articulation as ATR occurs naturally at the extreme positions of articulation in order to maximize the quality of the sound. This is similar to high back vowels favoring roundness, as the extension of the oral cavity lowers the pitch of the vowel even further, accentuating the height distinction.

So the tableau for the above change /ā/ → /ɑ/ is as follows:

MP *ā	αLg αATR	DEP-μ-IO	MAX-μ-IO	ID-IO(Fnt)	ID-IO(Lw)	ID-IO(ATR)	ID-IO (Bck)
a			*!				
☞ ɑː							*
ǣ				*!			
ē				*!	*		
ō					*!		*
əː	*!					*	

VI.B From Proto-Persian to Dari: Vowels

The Dari vocalic phonemic inventory is the product of but two sound changes from Middle Persian, and neither affects the contrastive status of the phonemes:

/i/ → /ɛ/
/ā/ → /ɑ/

These two changes are lowering and backing in nature, and the resultant system is as follows:

Dari:

/i/	/u/
/ɪ/	/ʊ/
/e/	/o/
(/ɛ/)	
/a/	
/a/ →	(/ɑ/)

/ā/ → /ɑ/ was covered above by the constraint MAX-ATR +Low +Back ("Maximize Backness of +Low +ATR segments"), and from an Optimality Theoretic

standpoint, the other shift, /ɪ/ → /ɛ/, is also intuitively tied to ease of articulation and parsing for the same reason. It can be succinctly expressed in the following constraint:

*-ATR +High (“High vowels must be +ATR.”)

Thus both of the above constraints are based on ease of parsing and even articulation as ATR occurs naturally at the extreme postions of articulation in order to maximize the quality of the sound.

The tableau for that above change would be as follows:

/ɪ/ → /ɛ/:

MP *ɪ	*-ATR+Hg	DEP-μ-IO	MAX-μ-IO	ID-IO(Fnt)	ID-IO(ATR)	ID-IO(Hg)
ɪ	*!					
☞ ɛ						*
ɪː	*!	*				
i					*!	
ī		*!			*	
ɨ				*!		

Unfortunately, *-ATR +High will also produce the undesirable result of /ʊ/ → /ɔ/:

MP *ʊ	*-ATR+Hg	DEP-μ-IO	MAX-μ-IO	ID-IO(Bck)	ID-IO(ATR)	ID-IO(Hg)
ʊ	*!					
☞ ɔ						*
ʊː	*!	*				
u					*!	
ū		*!			*	
ʉ				*!		

Thus, it must be explicitly stated that Back vowel outputs are faithful to their inputs in quality. Note that faithfulness constraints never need be justified, as they represent

conservation, which is implicit in a grammar at any state. In this case, the sweeping statement that back vowels retain their initial form is evidenced in the imperical data:

ID-$_{BACK}$-IO (“Back vowels in the output are faithful to their input segments.”)

/ɪ/ → /ɛ/:

MP *ɪ	ID-$_{BACK}$-IO	*-ATR+Hg	DEP-μ-IO	MAX-μ-IO	ID-IO(Fnt)	ID-IO(ATR)	ID-IO(Hg)
ɪ		*!					
☞ ɛ							*
ɪː		*!	*				
i						*!	
ī			*!			*	
ɨ̄					*!		

/ʊ/ → /ʊ/:

MP *ʊ	ID-$_{BACK}$-IO	*-ATR+Hg	DEP-μ-IO	MAX-μ-IO	ID-IO(Bck)	ID-IO(ATR)	ID-IO(Hg)
☞ ʊ		*					
ɔ	*!						*
ʊː	*!	*	*				
u	*!					*	
ū	*!		*			*	
ʉ̄	*!				*		

VI.C From Proto-Persian to Iranian: Vowels

The Iranian vocalic phonemic system, the least conservative, is the result of six shifts, which display all directional sound shifts, fronting, raising, backing and lowering:

/i/ → /ɛ/
/ē/ → /i/
/u/ → /o/
/ō/ → /u/
/a/ → /æ/
/ā/ → /ɑ/

The above set of changes can be expressed in simpler terms based on vowel position and length that actually indicate more succinctly the set of preferences that govern change from Proto-Persian into Iranian:

*+Long +Mid

- /ē/ → /i/
- /ō/ → /u/

*+Short +High

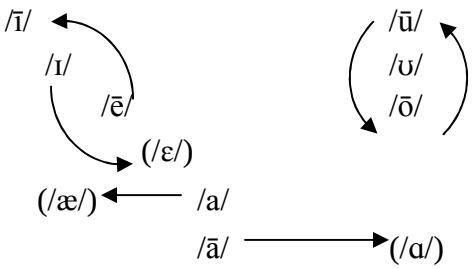
- /i/ → /ɛ/
- /u/ → /o/

*+Low +Central (or even *+Central)

- /a/ → /æ/
- /ā/ → /ɑ/

Thus the symmetry of the following will illustrate the validity of the above set of preferences in Iranian:

Iranian:



A further benefit of viewing these six shifts as three preferences is that if we are to understand constraints as a language's response to natural pressures, then "preference" is truly what constraints are expressing. For that reason, Iranian can be quantified using two sets of conspiring constraints rather than six baseless rules, those constraints will remain universal, not just to Persian, but to all Language.

The following two simple constraint pairs will give us all of the shifts from Proto-Persian into modern Iranian:

1. *-ATR +High: -ATR vowels are prohibited in high position.
   *+ATR -High: +ATR vowels are prohibited in –high position.

2. Max-Low (-ATR +Front): Maximize Frontness of –ATR inputs.
   Max-Low (+ATR +Back): Maximize Backness of +ATR inputs.

VI.C.1 *-ATR +High

Recall in the case of the first constraint above, that the Dari /ɪ/ → /ɛ/ shift was accounted for by *-ATR +High with faithfulness constraint ID-$_{BACK}$-IO outranking it so that /ɪ/ → /ɛ/ would occur without /ʊ/ → /ɔ/. In Iranian, we see that the *-ATR +High causes the back vowel to lower as well, even further justifying the legitimacy of the constraint. In this case, however, a further markedness constraint, *+Back –Round will allow /ʊ/ → /o/ and not /ʊ/ → /ɔ/:

*+Back –Round ("Unrounded back segments are prohibited." This constraint is grounded in acoustic phonetics and in empirical cross-linguistic data.)

/ɪ/ → /ɛ/:

MP *ɪ	*+Bk-Rnd	*-ATR+Hg	DEP-μ-IO	MAX-μ-IO	ID-IO(Fnt)	ID-IO(ATR)	ID-IO(Hg)
ɪ		*!					
☞ ɛ							*
ɪː		*!	*				
i						*!	
ī			*!			*	
ɨ					*!		

/ʊ/ → /o/:

MP *ʊ	*+Bk-Rnd	*-ATR+Hg	DEP-μ-IO	MAX-μ-IO	ID-IO(Bck)	ID-IO(open)	ID-IO(ATR)
ʊ	*!	*					
ʊː	*!	*	*				
☞ o							*
ū			*!				*
ʉ					*!		
ɔ						*!	

VI.C.2 *+ATR -High

It may be stated that *-ATR +High and *+ATR –High are conspiratorial in that they overtly state the direct correlation between height and ATR. That is to say that they formalize the notion that high vowels "prefer" to be +ATR and ATR vowels "prefer" to be raised. That being said, *+ATR -High will have to appear relatively low in the constraint ranking, as such a powerful constraint will militate against any +ATR vowels that are not /ū/ or /ī/. Thus, the following tableaux illustrate how low in the ranking *+ATR –High will fall:

/ē/ → /ī/:

MP *ē	*+Bk-Rnd	*-ATR+Hg	MAX-μ-IO	ID-IO(ATR)	ID-IO(Fnt)	*+ATR -Hg	ID-IO(Hg)
ē						*!	
☞ ī							*
ɛ			*!	*			
ɪ		*!	*	*			*
ō					*!		
ā					*!		*
ū					*!		*
ə			*!	*			

/ō/ → /ū/:

MP *ō	*+Bk-Rnd	*-ATR+Hg	MAX-μ-IO	ID-IO(ATR)	ID-IO(Bk)	*+ATR -Hg	ID-IO(Hg)
ō						*!	
☞ ū							*
ē					*!	*	
ɔ	*!		*	*			
ʊ	*!	*	*	*			
ā					*!	*	
a		*!			*		
ə			*!	*	*		

Meanwhile, other –High +ATR vowels - in this case there is only one vowel - are preserved by higher ranking faithfulness constraints on backness and ellipsis. Crucially, it should be noted that said faithfulness constraints appeared in the above tableaux as well, but had no effect on the final output evaluation:

/ā/ → /ā/:

MP *ā	*+Bk-Rnd	*-ATR+Hg	MAX-μ-IO	ID-IO(ATR)	ID-IO(Bk)	*+ATR -Hg	ID-IO(Hg)
a			*!	*			
☞ ā						*	
ē					*!		
ō					*!		
æ			*!	*			
ǣ					*!		
ū					*!		*
ə			*!	*			

In Iranian, /ā/ → /ā/ as in the other dialects, but this example is included so that it would be clear that *-High +ATR was outranked by enough faithfulness constraints that it did not affect /ā/.

So, to recap VI.C.1 and VI.C.2, markedness constraints *-ATR +High, and *+ATR -High conspired at different rankings in the evaluator with various faithfulness constraints to produce the following four seemingly contrary changes:

/ē/ → /i/
/ō/ → /u/

/i/ → /ɛ/
/u/ → /o/

They are seemingly contrary due to the fact that two mid vowels raise to high vowels and two high vowels lower to mid vowels. This demonstrates the power of OT as an explanatory theoretical framework, as all four changes are covered in one constraint ranking, and no other sound change occurs due to this hierarchy.

VI.C.3 Max-Low (-ATR +Front) and Max-Low (+ATR +Back)

Max-Low (-ATR +Front) or rather "Maximize Frontness of Low –ATR segments" is part of a conspiracy with Max-Low (+ATR +Back) mandating that in low vowels, ATR and frontness are inversely correlated. This is based on the physiology of speech in that in lower, more open vowels tongue root advancement occurs naturally with backing, and so the inverse is true as well. It further encapsulates the notion in Iranian that low central vowels are avoided in the grammar. In the below tableaux, there need not be –Low segments generated as input, as the constraints only apply to +Low segments:

/ā/ → /ɑ/:

MP *ā	Lw(+A +B)	*+Bk-Rnd	*-ATR+Hg	MAX-μ-IO	ID-IO(ATR)	ID-IO(Bk)	*+ATR -Hg
a	*!			*	*		
☞ ɑ		*				*	*
ā	*!			*			*
ǣ	*!						*
æ	*!			*	*		

/a/ → /æ/:

MP *a	Lw(-A +F)	*+Bk-Rnd	*-ATR+Hg	MAX-μ-IO	ID-IO(ATR)	ID-IO(Bk)	*+ATR -Hg
a	*!						
☞ æ							
ā				*!	*		*
ǣ				*!			
ɑ		*!				*	*

VI.C.1 through VI.C.3 have shown that a very simple set of interrelated conditions on height and ATR accounted for all six shifts in Iranian, and even unified an account of some seemingly contradictory data.

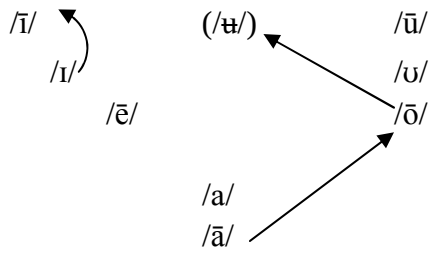
VI.D From Proto-Persian to Tajik: Vowels

The last of the three dialects to be considered is the Tajik phonemic system in which all of the changes exhibit raising:

/ā/ → /o/
/ō/ → /ʉ/
/i/ → /i/

In VI.C, it was shown that Iranian exhibits a resistance to long mid vowels, and in all three dialects there is some sort of shift in /ā/, but Tajik is indeed unique among the three

dialects in that the shifts create a system with no long low vowel and they add the central high rounded vowel /ʉ/. Fortunately for this analysis, all three sound shifts will be shown to occur due to the notion that ATR is a feature "preferred" by high vowels and "avoided" by low vowels. As before, a limited and complementary set of constraints will conspire to produce the shifts in Tajik. The following is a graphical representation of the sound shifts leading to modern Tajik:



/ā/ → /ō/, which occurred after /ō/ → /ʉ/ and /i/ → /ī/, will be seen to have occurred due to a constraint on ATR and lowness, indicating that ATR is a marked feature in low vowels. This is born out in numerous sound shifts including the great vowel shift of English, and related to the constraint generating /ō/ → /ʉ/, which like Iranian /ō/ → /ū/, will be seen to occur due to the *+ATR –High. Lastly /i/ → /ī/ will be shown to occur due to the complementary constraint *-ATR +High, which along with its counterpart was also seen to affect Dari and Iranian.

Before the OT analysis, a note should be made on the timing of these changes. This will be covered in greater detail in section VI.H.7.b, but for now, suffice it to say that this dissertation assumes /ā/ → /ō/ and /i/ → /ī/ to have both occurred later than /ō/ →

/ʉ/, and in fact, it will later be suggested that /ā/ → /ō/ and /i/ → /ī/ both occurred within the twentieth century.

VI.D.1 *+Low +ATR

Recall that in Iranian, /ā/ → /ɑ/ was generated by the constraint Max-Low (+ATR +Back) ("Maximize backness of low +ATR segments"). In the case of Tajik, which will also mandate a change in the segment /ā/, /ā/ → /o/ will occur due to the following markedness constraint:

*Low +ATR ("Low +ATR segments are prohibited.")

This constraint is based essentially in the physiological difficulty in advancing the tongue root while articulating low vowels, paired with the general tendency towards advanced tongue root in the articulation of non-low vowels:

PIE *ā	MAX-μ-IO	ID-IO(Hi)	ID-IO(Fnt)	*+Lw +ATR	ID-IO(Bk)	ID-IO(Low)
ā				*!		
☞ō					*	*
ū		*!				*
a	*!					
ǣ			*!		*	
ē			*!			*

This will furthermore not interfere with other vowels such as short /a/:

PIE *a	DEP-μ-IO	ID-IO(Hi)	ID-IO(Fnt)	*+Lw +ATR	ID-IO(Bk)	ID-IO(Low)
ā	*!					
☞a						
o					*!	*
e			*!			*
æ			*!			
ɵ						*!

VI.D.2 *+ATR –High and *-ATR +High

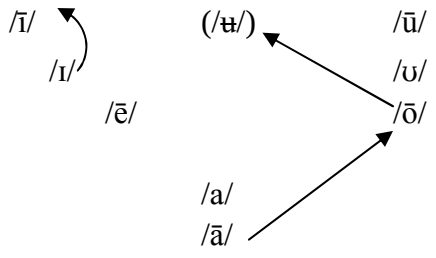
In Iranian, the constraint *+ATR -High caused both long mid vowels /ē/ and /ō/ to raise, but in Tajik it will only affect the back mid long vowel /ō/. Thus, it will be shown that in Tajik this constraint works alongside another markedness constraint on length as distinct from ATR: *+Long –Low, which relates to the fact that +Long vowel articulation is easiest in open vowels.

The convergence of /i/ and /ī/ as /ī/ will be shown to occur due to the constraint *-ATR +High, which is complementary to *+ATR –High, and these two constraints follow naturally from *+Low +ATR. That is to say that the following three constraints can be seen as part of one natural preference on the part of Tajik:

*+ATR –High
*-ATR +High
*+Low +ATR
} "ATR is a feature preferred in higher vowels."

The forth constraint which brings about Modern Tajik, *+Long –Low, will be seen to be part of another preference for length to be directly correlated with openness.

As illustrated above, this preference for height in +ATR segments will lead to three raising shifts in Tajik:



/ō/ → /ʉ/:

MP *ō	*+Bk-Rnd	*-ATR+Hg	MAX-μ-IO	ID-IO(Lw)	ID-IO(Fnt)	*+ATR -Hg	ID-IO(Hg)	*+Lng -Lw
ō						*!		
☞ ʉ							*	
ū							*	*!
ē					*!	*		
ɔ	*!		*					
ʊ	*!	*	*					
ā				*!		*		
a		*!		*				
ə			*!					

/i/ → /ī/:

MP *ī	*+Bk-Rnd	*-ATR+Hg	MAX-μ-IO	ID-IO(Low)	ID-IO(Fnt)	*+ATR -Hg	ID-IO(Hg)	*+Lng -Lw
i		*!						
☞ ī								*
ɛ							*!	
ē						!*	*	*
ū					*!			*
ʊ	*!	*						
ā				*!		*	*	
a				*!			*	
ɨ		*!						

VI.E Symmetry and Conspiracy in Proto-Persian Markedness Constraints:

The markedness constraints leading to the vowel systems characterizing the modern dialects of Persian all correspond to the following set of four preferences:

1. +ATR in Long vowels and –ATR in Short vowels:

   - Max-IO[αLong αATR]: ("If input segment is αLong, output will be αATR.")

2. +ATR in higher vowels and –ATR in lower vowels:

   - *-ATR +High: "High vowels must be +ATR."
   - *+ATR -High: "+ATR vowels must be High."

- *+Low +ATR: “Low vowels must be –ATR.”

3. In low vowels, ATR is directly correlated with backness.

   - Max-Low(-ATR +Fnt): “Maximize Frontness of –ATR Low vowels.”
   - Max-Low (+ATR +Back): “Maximize Backness of +ATR Low vowels.”

4. Long vowels will tend towards openness.

   - *+Long –Low: “Long vowels must be open.”

The strength of OT lies in the fact that the dozen shifts that occur after Proto-Persian can be accounted for by constraints that not only govern the shifts themselves, but also encapsulate the above tendencies. One seeming paradox that should be addressed is the following ostensible flaw in logic:

Long → ATR
ATR → High
Long → Low

Long → ATR and ATR → High should naturally indicate that Long → High, but according to #4 above Long → Low. This occurs due to the physiological foundations for the markedness causing the shift. That is, Long → ATR is based in the notion that increased length allows more time to advance one’s tongue root, and Long → Low is based in the similar idea that a greater duration will allow more time to open the oral cavity, hence they are both based on economy of production. ATR → High is related to the natural tendency of the tongue root to advance in higher vowels. Thus, *Long → High does not naturally fall out from Long → ATR and ATR → High because Long → Low cancels out *Long → High due to the provisions of duration.

VI.F Other Vowel Issues: Diphthongs

Diphthongs will present yet another instance that displays the power of OT by allowing one theory to pit competing universal constraints on language against each other to demonstrate how seemingly opposite shifts can occur. It will be shown that ease of articulation and ease of parsing will lead to differing changes in the diphthongal development of modern Persian. The current state of Modern Persian diphthongs is as follows:

	MP:	Mod Pers:	Dari:	Iranian:	Tajik:	Gloss:
/aʊ/	/ʤaʊ/	"جو"	[ʤaw]	[ʤo]	[ʤav]/ [ʤaw]	'barley'
/aj/	/kaj/	"کی"	[kaj]	[ke(j)]	[kaj]	'when'
/āʊ/	/gāʊ/	"گاو"	[gaw]	[gɑv]	[gov]/ [gav]	'cow'
/āj/	/pāj/	"پای"	[pɑj]	[pɑ]	[po(j)]	'foot'
/ōj/	/bōj/	"بوی"	[buj]	[bu]	[b~~ʊ~~(j)]	'smell'

With these in mind, there are four shifts for which an account must be made, and said account can be summed up in three constraint sets. Note that /ōj/ is not going to be dealt with in this section, but rather in section VI.G.1.a, as it represents a deviation in Dari's retention of original /ō/. The four shifts then to be considered in the present section are as follows:

1. /āʊ/ → [aw] (Dari)

This shift will occur due to the following constraint:

MAXIMAL DISTINCTION ("A diphthong's component features must be maximally distinct." This is obviously based on ease of parsing.)

2. /aʊ/ → [o] (sometimes rendered [ow] in literature) (Iranian)
/aj/ → [e] (or [ej]) (Iranian)

For this shift, I will draw on two constraints proposed by other scholars related to diphthong vowel height and harmony. MINIMAL DISTANCE, a constraint given by Miglio (1998, p. 9), requires "a minimal auditory distance between contrasting forms," and DIPHTHONG HEIGHT HARMONY, a constraint proposed by McMahon (2003, p. 137), is expressed as follows:

$$\$ \begin{bmatrix} \alpha\text{high} \\ \beta\text{back} \end{bmatrix} \begin{bmatrix} \alpha\text{high} \\ \beta\text{back} \end{bmatrix} \$$$

The combination of these two constraints will essentially mandate that for ease of articulation, diphthongal vocalic components should, contrary to MAXIMAL DISTINCTION, converge. These contrasting constraints capture the notion that ease of articulation and ease of parsing each have a crucial role to play in language change, and can lead to conflicting outcomes (certainly a desirable result for anyone who knows the nature of language change):

MINIMAL DISTANCE	("Minimize auditory distance between contrasting forms.")
DIPHTHONG HEIGHT HARMONY	("Diphthong segments should converge for Height and Backness.")

3. /w/ → [v] (Iranian and Tajik)

This final shift will require a revision in the inventory previously assumed. Evidence in extant Indo-European languages points to the existence of labiodental approximate /ʋ/, which in fact still exists in modern Hindi among other related languages. That being said, rather than positing that Dari and Tajik retain the original [w] and Iranian has shifted this to [v], I posited (and taken for granted above) that the original diphthong is in fact /aʋ/, and in the former /ʋ/ → [w], while in the latter /ʋ/ → [v]. That will require a constraint militating against labiodental approximates (grounded in ease of articulation and justified by the cross-linguistic markedness of the segment [ʋ]). That constraint will, in Dari and Tajik, conspire with an additional markedness constraint against voiced fricatives (also grounded in ease of articulation and justified by cross-linguistic markedness) to produce [w], and in Iranian, will conspire with a constraint against [w] to produce [v]. *[w] is based on the same facts as *[ʋ]:

*[ʋ]	("The segment [ʋ] is prohibited.")
*[w]	("The segment [w] is prohibited.")
*Fricative [+voice]	("Voiced Fricatives are prohibited.")

VI.F.1 Diphthongs: From PP to Dari

In Dari, /aw/, /aj/ and /āj/ remain in their original form and the onset vowel of /āw/ (or rather [ɑw]) shortens. /āw/ → [aw] will occur due to the constraint MAXIMAL DISTINCTION as the following illustrates:

MP*āw	DEP-μ-IO	ID-IO(Low)	ID-IO(Bk)	ID-IO(Fnt)	MAXDIST	ID-IO(Hg)
ō	*	*!	*		*	*
☞ aw	*				*	
ew	*	*!	*		*	
a	**!					*
ū	*	*!	*		*	*
ɑw	*		*!		*	
ā	*	*!	*		*	*
ow	*	*!	*		*	
æw	*			*!		

It is worth noting here that in the cases of both original MP /aw/ and /āw/, the maximal distance principle should juxtapose [æ] and [w], as those two segments are indeed maximally distinct. It is then extremely satisfying and fortunate for the proposed constraint that the Dari form [aw], both from original /aw/ and original /āw/, exists in register-tied variation with [æw][40], and this can be accounted for very simply by proposing a variation in the above constraint ranking of MAXDIST and ID-IO(Front), such that when the ranking is switched, the following tableau will be manifest, and [æw] will be the winning output:

[40] In point of fact, [æw] occurs markedly more frequently than [aw] in Dari.

MP *aw	DEP-μ-IO	ID-IO(Low)	ID-IO(Bk)	MAXDIST	ID-IO(Fnt)	ID-IO(Hg)
ō		*!	*	*		*
aw				*!		
ew		*!	*	*		
a	*!					*
ū		*!	*	*		*
αw			*!	*		
ā		*!	*	*		*
ow		*!	*	*		
☞æw					*	

It should be reiterated that this is not a paper that seeks to add to or subtract from OT, but rather a philological work seeking to quantify the changes responsible for modern Persian dialectology. That being said, the physiological justifications offered by OT are attractive theoretically, as sound change is truly tied to the mechanics of communication rather than stated as independent fact. At this point then, and as mentioned in the introduction, a note should be added on how the mechanism of OT will account for free variation in language. Any student of language change takes for granted that a language change is typically preceded by a period of optionality. That is to say that speakers of early English did not decide by consensus one day to have a vowel shift, great or small, and the next day implement it. Rather some pressure exerted over time gradually pushed the change until it was complete. It is my presumption that this may be represented according to the following set of steps:

MP *aw	DEP-μ-IO	ID-IO(Low)	ID-IO(Bk)	MAXDIST	ID-IO(Fnt)	ID-IO(Hg)
ō		*!	*	*		*
☞aw				*		
ew		*!	*	*		
a	*!					*
ū		*!	*	*		*
αw			*!	*		
ā		*!	*	*		*
ow		*!	*	*		
☞æw					*	

In the above tableau, MAXDIST and ID-IO (Front) are not strictly ordered, and the output may vary either by speaker, region, age, or what have you. While this appears to fly in the face of "strict ordering of constraints" (one of OT's main tenets), it really does not. That is, at the time of articulation (in OT: output generation), MAXDIST and ID-IO (Front) are indeed ordered such that either [aw] or [æw] is the output, but which it is may depend on a number of factors, and in fact may be random. To state this more precisely, if empirical data indicates that this segment is rendered [aw] 60% of the time and [æw] 40%, the following tableau may be utilized:

MP *aw	DEP-μ-IO	ID-IO(Low)	ID-IO(Bk)	MAXDIST	ID-IO(Fnt)	ID-IO(Hg)
ō		*!	*	*		*
☞$aw_{60\%}$				*		
ew		*!	*	*		
a	*!					*
ū		*!	*	*		*
αw			*!	*		
ā		*!	*	*		*
ow		*!	*	*		
☞$æw_{40\%}$					*	

This may also be the methodology employed in cases that appear to be frequency driven, such as many of the consonant changes. Whatever the case may be, this may be more concretely exemplified by the following oversimplification:

sneak+past (lexicon > 45 years old)	Past = +ed	Past = umlaut
☞sneaked		*
snuck	*!	
sneak+past (lexicon < 45 years old)	Past = umlaut	Past = +ed
sneaked	*!	
☞ snuck		*

This captures the notion that a sixty-year old American may never say 'snuck,' and a thirty year old American may never say 'sneaked,' even if both speak precisely the same dialect. It further encapsulates the other fact about language change that generational facts are relevant. That is, 'sneaked' will likely die within a few decades as its speakers die. That was admittedly a digression, but a necessary note on OT and language change.

VI.F.2 Diphthongs: From PP to Iranian

/aw/ → [o] and /āw/ → [ɑv] in Iranian appear to indicate a distinction between the way in which the grammar interprets these sounds. Immediately the question arises over why in Iranian, the short vowel + glide monophthongalizes while the long vowel + glide becomes long vowel + voiced fricative. Furthermore, in Tajik they both become vowel + voiced fricative, and in Dari they both remain V+[w].[41] For this it is necessary to understand the nature of MINIMAL DISTANCE and DIPHTHONG HEIGHT HARMONY. I propose that DIPHTHONG HEIGHT HARMONY (DHH) as defined:

[41] This is a worthwhile note, as it may be tempting to assume that because in Iranian the second diphthongal constituents shift differently, they are in fact different constituents, but the data from Tajik and Dari refute this claim.

$ $\begin{bmatrix} \alpha\text{high} \\ \beta\text{back} \end{bmatrix}$ $\begin{bmatrix} \alpha\text{high} \\ \beta\text{back} \end{bmatrix}$ $

can be further divided into DHH (High) into DHH (Back):

DHH (High): $ [αhigh] [αhigh] $ ("Diphthong components converge for height.")
DBH (Back): $ [βback] [βback] $ ("Diphthong components converge for backness.")

This distinction will allow the Iranian diphthongal shift to be viewed as twofold. First, the short vowel diphthongs will be seen to monophthongalize due to the frontness of the onset vowel, and the long vowel diphthongs will be seen to remain diphthongs due to the backness of the onset vowel.

MP *aj	DEP-μ-IO	MAX-μ-IO	DHH(H/B)	ID-IO(Hg)
☞ ej				
aj			*!	
e		*!		
ɑ		*!		
ij				*!
āj	*!		*	*
ɛ		*!		

The variation of /ej/ as [e] is simply the product of a constraint against diphthongs. The same holds true for /aw/:

MP *aw	DEP-μ-IO	MAX-μ-IO	DHH(H/B)	ID-IO(Hg)
☞ ow				
o		*!	*	
aw			*!	
ew			*!	
a		*!		
uw				*!
āw			*!	*
ā		*!		

In the latter case, the effect of the constraint *diphthong will not be an optional portion of the constraint ranking, but rather a second step mandating that the new phoneme /ow/ → [o], and providing as an optional shift /ej/ → [e]:

MP *ow	*Diphth	MAX-μ-IO
☞ o		*
ow	*!	

MP *ej	*Diphth	MAX-μ-IO
☞ e		*
ej	*!	

Or the variant:

MP *ej	MAX-μ-IO	*Diphth
e	*!	
☞ ej		*

As for the final Iranian diphthong, the principal distinction between underlying /āw/ and the three underlying diphthongs /āj/, /aw/ (or /æw/) and /aj/ (or /æj/) is that both components of the diphthong /āw/ (realized in all three dialects as /ɑw/) are back segments. That is to say that Back Harmony will play a crucial role in this analysis, and will simplify the process while accounting for the above mentioned paradox: "The short vowel + glide monophthongalizes while the long vowel + glide becomes long vowel + voiced fricative." The following are the crucial constraints:

DBH (Back): $ [βback] [βback] $ ("Diphthong components converge for backness." This is grounded in ease of articulation and minimization of effort.)

*C[+labial -obstruent] ("Labial glides and approximates are prohibited." This is also grounded in ease of articulation and backed by cross-linguistic markedness.)

*ID-IO (plosive) ("Input and output segments must match for the feature ±Plosive." Fatithfulness constraints need not be justified.)

Interestingly, recall that in section VI.C.3, the constraint Max-Low (+ATR +Back) was proposed mandating that in Low vowels, the backness of +ATR segments is maximized. This constraint will not be necessary in causing /ā/ → [ɑ] to occur in the diphthong /āw/, as it follows naturally from DBH (BACK), as will be seen below. This can only be seen to strengthen the implications of an OT based approach, because differing constraints may lead to similar shifts, providing for us the desirable theoretical ground that language change takes place for a variety of reasons and based on a variety of articulatory and auditory pressures.

MP*āw	DHH(B)	MAX-μ-IO	*C[+Lb-Ob]	ID-IO(Plo)	ID-IO (Low)	*+Lng-Lw
ō		*			*!	*
āw	*!		*			
ɑb		*		*!		
☞ ɑv		*				
ow		*	*!		*	
o		**!			*	
ov		*			*!	
ū		*			*!	*
ɑw		*	*!			
æw	*!	*	*			

A final note should be made on the constraint *diphthong in Iranian. Returning to the chart at the beginning of section VI.F of the five potential diphthongs, Iranian has not a single example in this list that is canonically a diphthong:

	MP:	Mod Pers:	Iranian:	Gloss:
/aw/	/ʤaw/	“جو”	[ʤo]	‘barley’
/aj/	/kaj/	“کی”	[ke(j)]	‘when’
/āw/	/gāw/	“گاو”	[gɑv]	‘cow’
/āj/	/pāj/	“پای”	[pɑ(j)]	‘foot’
/ōj/	/bōj/	“بوی”	[bu]	‘smell’

In multiple other instances, Iranian elides word final diphthongs, such as “جا” [ʤɑ] ‘place,’ which is [ʤɑj] in Dari and Tajik and “مو” [mu] ‘hair,’ which is [muɟ] in Dari and [mʉj] in Tajik, to name a few. Note that history backs up the originality of the word final semivowel as evidenced by the Middle Persian column above. This is not a consistant occurance, but warrants a comment, as it is a notable distinction between Iranian and the other dialects.

VI.F.3 Diphthongs: From PP to Tajik

In Tajik, the diphthongal shifts will be very simple, as the palatal diphthongs remain, and the labial diphthongs both convert from vowel + approximate to vowel + obstruent: /aw/ → [av] and /āw/→[ov]. Thus:

MP*āw	*C[+Lb-Ob]	DHH(B)	*+Lw+ATR	MAX-μ-IO	ID-IO(Plo)	*+Lng-Lw	ID-IO (Fnt)
ō				*		*!	
āw	*!	*	*				
ob				*	*!		
ɑv			*!	*			
ow	*!			*			
o				**!			
☞ ov				*			
ū				*		*!	
ɑj		*!	*	*			
æj				*			*!

MP*aw	*C[+Lb-Ob]	DHH(B)	*+Lw+ATR	MAX-μ-IO	ID-IO(Plo)	*+Lng-Lw	ID-IO (Low)
ō						*!	*
āw	*!	*	*				
aw	*!						
ab					*!		
☞ av							
ow	*!						*
o				*!			*
ov							*!
ū							*!
ɑj		*!	*				
æj		*!					

VI.F.4 Diphthong Overview

It has been shown that OT accounts for opposing shifts that occur within one single system. Ease of articulation can bring about one change and ease of parsing may bring about an opposing change, but more importantly than that, the theory advances a reason for conflicting changes. In Dari, /āw/ → [aw] took place because of the constraint MAXIMAL DISTINCTION based on ease of parsing, while in Iranian, /aw/ and /aj/ → [o] and [e] due to the constraints MINIMAL DISTANCE, and DIPHTHONG HEIGHT HARMONY, both based on ease of articulation.

VI.G Other Vowel Issues: Introduction

VI.G is dedicated to a few loose ends not yet covered in the preceding sections on vowels and diphthongs. The first three sections will be particular to Iranian and the fourth to Tajik. Section VI.G.1 deals with the fact that although we know that in Iranian /ō/ → /ū/ (and this remains /ō/ in Dari) and /aw/ → /o/, there is a rare, but nonetheless real

original /ō/ that remains in a very few cases in Iranian and in even fewer, that will become [u] in Dari. Section VI.G.2 will treat another notable quality that distiguishes Iranian from the other two dialects, and that is the abundance of the segment [ɛ] where the others would have [a]. Iranian also distinguishes itself with another phenomenon, [ɑ] → [u] / ___C[+nasal], and this will be covered in section VI.G.3. Finally, section VI.G.4 will briefly mention the actual geographic proliferation of the Tajik vocalic segment [ʉ].

VI.G.1 Iranian /ō/

To state categorically that Dari retains original /ē/ and /ō/ and Iranian loses them is not only an oversimplification, but rather a misleading statement, and hence, as with any misinformation, this section aims to add precision by confessing the oversimplification and correcting it. The fact is that the Iranian dialect does contain /ē/ and /ō/ in its phonemic inventory (see section VI.C), and in two instances, original /ō/ remains in Iranian and becomes /ū/ in Dari. This being said, the oversimplification was stated because the vast majority of MP instantiations of /ē/ and /ō/ shifted wholesalely to /ī/ and /ū/ respectively in Iranian and remained /ē/ and /ō/ in Dari. Tajik has been excluded from this portion, as the distintion from existing high vowels remains, but Tajik /ō/ becomes [ʉ] (This was shown in sections V.C.3 and V.C.4).

The Iranian phonemic inventory will be redefined then to include /ē/ (the historical reflex of the diphthong /aj/ with free variant [ej], as shown in section VI.F.2) and /ō/ (the historical reflex of [aw] with variant [ow], also shown in section VI.F.2). Additionally, Iranian retains a select few instances of original /ō/, but only in the word-final position of four monosyllabic lexical items. Cases of borrowing, such as "مانتو" [mɑn-to] 'shawl' will be ignored as it is clearly borrowed from the French "manteau." Iranian retains no original /ē/ outside of the diphthong /ej/. Likewise, Dari, which has been described as categorically retaining the original /ō/, will be shown in two of those four word-final instances and one historical diphthongal case to change this to /ū/. This being established, this section begins by stating the following as shifts in Dari and Iranian from the original MP vowel /ō/:

- MP [ō] in closed syllables → [ū] in Iranian and remains [ō] in Dari:

MP:	Mod P:	Dari:	Dari:	Iranian:	Iranian:	Gloss:
rōz	"روز" 'day'	[roz]	*[ruz]	[ruz]	*[roz][42]	
ʧōb	"چوب"	[ʧob]	*[ʧub]	[ʧub]	*[ʧob]	'wood'
kōr	"کور"	[kor]	*[kur]	[kur]	*[kor]	'blind'

[42] Recall that [roz] is a possible output in Iranian, but would be the Modern Iranian reflex of original */rawz/, which would remain *[rawz] in Dari and become *[ravz] in Tajik.

The above lexical items exemplify the vast majority of Middle Persian /ō/, and for that reason it was stated as canonical in preceding sections that Middle Persian /ō/ → /ū/ in Iranian and → /ō/ in Dari. For a more or less complete list Dari lexical items containing the segment /ō/, refer to Farhadi (1955, p. 12).[43] As mentioned in the introduction above, there are only four cases that violate the above generalization, and these shall be covered in sections VI.G.1.a through VI.G.1.d

VI.G.1.a /ō/: MP /ōj/ → Dari /ō/ or /ūj/; Iranian /ū/; Tajik /ū(j)/ or /ʉ(j)/

The diphthong /ōj/ was mentioned in section VI.F, and must be returned to in the context, not of diphthongs, but of cases that deviate from the above-stated "rule:" 'MP /ō/ in closed syllables → /ū/ in Iranian and remains /ō/ in Dari.' The present section deals with Middle Persian forms that were monosyllabic and ending in the diphthong /-ōj/. Their Dari, Iranian and Tajik forms vary in quality. Dari does not tolerate /ōj/ so it had two recourses, raise the vowel because of DBH (High) or drop the palatal glide.

- /ōj/ → /ū/ (Iranian) (numerous cases)
  → /ō/ or /ūj/ (Dari)
  → /ū(j)/ or /ʉ(j)/ (Tajik)

[43] Farhadi (1995, pp. 11-13) has in fact four lexicon lists, containing the Dari tokens retaining /ō/, /ē/, /aj/ and /aw/ all lost in Iranian, and partially altered in Tajik.

MP:	Mod P:	Formal Dari:	Spoken Dari:	Iranian:	Tajik:	Gloss:
rōj	“(ی)رو”	[ro]	[ruj]	[ru]	[ru(j)]/[r~~ʉ~~(j)]	‘face’
mōj	“(ی)مو”	[mo]	[muj]	[mu]	[mu(j)]/[m~~ʉ~~(j)]	‘hair’
bōj	“(ی)رو”	[bo]	[buj]	[bu]	[bu(j)]/[b~~ʉ~~(j)]	‘smell’
ʤōj	“(ی)جو”	[jo]	[ʤuj]	[ʤu]	[ʤu(j)]/[ʤ~~ʉ~~(j)]	‘ditch’

There is only one exception to this:

MP:	Mod P:	Formal Dari:	Spoken Dari:	Iranian:	Tajik:	Gloss:
ōj	“او”	/o/	[u]	[u]	[u]	‘s/he’

This case will be viewed as a shift that occurred for two reasons. Iranian resistance to /ō/ and Dari and Iranian resistance to the diphthong /ōj/. No new constraints will be introduced to deal with this change (and the exception will conspicuously be regarded as just that):

For Iranian /ōj/ → /ū/, *+ATR –High (see sections VI.C.2 and VI.D.2) and DIPHTHONG HEIGHT HARMONY for Height and Backness, or for short DHH(H/B), (see sections VI.F) will conspire in the following manner:

MP*ōj	*+ATR -Hg	DHH(H/B)	ID-IO (Fnt)	MAX-μ-IO	ID-IO(Hg)
☞ ū				*	*
āj	*!	*	*		
ūj		*!			*
ī			*!	*	*
ōw	*!	*			
īj			*!		*
ōe	*!				
ō	*!			*	
αj	*!	*		*	
æj		*!		*	

For formal Dari /ōj/ → /ō/, DIPHTHONG HEIGHT HARMONY for Height (DHH(H), see section VI.F.2) and a faithfulness constraint on Height will conspire to bring about the alternation in Dari formal register:

MP*ōj	DHH(H)	ID-IO(Hg)	MAX-μ-IO	ID-IO (Bck)
ū		*!	*	
āj	*!			*
ūj		*!		
ī		*!	*	*
ōw	*!			
ïj		*!		*
ōe	*!			
☞ō			*	
αj	*!	*	*	
æj	*!		*	*

For the colloquial Dari /ōj/ → /ūj/, the shift to colloquial speech will bring MAX-μ-IO (or "Ellipsis is prohibited.") above DHH(H), but note that crucially, there is one two-step shift between the same constraints to derive colloquial register from the existing tableau for fomal. This is desirable, as a more formal register should be the product of a highly ranked faithfulness constraint, is it should be the more conservative variant:

MP*ōj	MAX-μ-IO	DHH(H)	ID-IO (Hg)	ID-IO(Back)
☞ ūj			*	
āj		*!		*
ū	*!		*	
ī	*!		*	*
ōw		*!		
ïj			*	*!
ōj		*!		
ō	*!			
αj	*!	*		
æj	*!	*	*	

The above has left Tajik out of the analysis because Tajik allows both vowels (/u/ from original /ū/ and /ʉ/ from original /ō/) as potential outputs, and also allows for the absence or presence of the palatal glide. Thus, DIPHTHONG HEIGHT HARMONY for Height (DHH(H)) will cause /ōj/ → /ū/ or /ōj/ → /ʉ/ depending on the absence or presence of the constraint *diphthong, and a higher ranked faithfulness constraint could preserve the glide leaving /ūj/ or /ʉj/. Either of these effects will be seen as optional.

<u>VI.G.1.b /ō/: MP /ūh/ → Dari /ō/ and Iranian /ōh/</u>

- /ū/ → /ōh/ (Iranian) (one case)
  → /ō/ (Dari)

Example: MP gūh → "گوه" Iranian [goh] Dari [go] Tajik [gu(h)] 'dung'

The original Middle Persian form is not trully a closed syllable, but in point of fact ends in /h/. This /h/ is subsequently deleted in colloquial speech, and as the lexical item /goh/ is vulgar, it is less frequently written, and lost entirely at some point after the articulatory preference for /ōh/ over /ūh/ has taken effect. In modern colloquial Iranian the [h] is retained for emphatic usage. In the absence of other evidence, it appears that if this exception is phonologically conditioned, the relevant segment is the coda-positioned spirant. There are no other attested lexical items in Middle Persian with word-final /ūh/, and for that reason, any attempt at analysis would result in moot assumptions. However

this would prove to be fodder for interesting future study in the phonological and possibly even scatological implications of taboo vocabulary.

It is unclear why this vowel raises in Tajik while we would expect to see [gʉ].

VI.G.1.c /ō/: MP /ō/ → Dari /ō/ and Iranian /ōh/

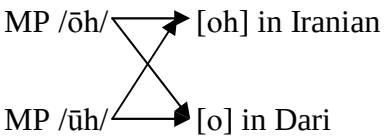
- /ō/ → /ōh/ (Iranian) (one case)
  → /ō/ (Dari)
  → /ū/ or /ō/ (Tajik)

Example: MP nō → "نه" Iranian [noh] Dari [no] Tajik [nu] or [nʉ] 'nine'

This is a bit less straight forward, as its spelling in Arabo-Persian script ("نه") would indicate that it is historically */noh/, with a short mid vowel[44], as contrary to its attested Middle Persian form /nō/. The reason for this is that the Arabic grapheme "ه" or /h/, much like the English usage of "h," is either used as the glottal approximate or as a word-final vocalic placeholder. Because of this ambiguity, I will consider this example to be another exceptional one because the status of the final /h/ (present in modern Iranian and absent in its Middle Persian predecessor) is unclear. If we may assume that the coda spirant is in fact the operative element, and that it was precisely that which preserved the mid vowel, this still does not account for the two lexical items (in section VI.G.1.d) that are reversed in Dari and Iranian.

[44] The Perso-Arabic script does not graphemically represent short vowels and so the Perso-Arabic "nh" would indicate a inter-consonantal short vowel.

However, the following can account for at least VI.G.1.a and VI.G.1.b if we assume that MP /nō/ was in fact originally /nōh/ as implied by modern Persian spelling then we can assume that /ōh/ and /ūh/ came together as [oh] in Iranian and [o] in Dari:



The following constraint is based on the marked nature of glottal-final syllables:

*V[+High] [h]]$_\sigma$ ("Vowels preceding glottal spirants must lower." This is due to ease of articulation as glottal spirants are conducive to more open articulation.):

MP*ūh	*V[+H][h]]$_\sigma$	ID-IO (Glot)	MAX-μ-IO	ID-IO (Low)	ID-IO(Hg)
☞ ōh					*
āh				*!	*
ūh	*!				
īh	*!				
ōw		*!			*
ō			*!		

And this does not affect the status of original MP /ōh/:

MP*ōh	*V[+H][h]]$_\sigma$	ID-IO (Glot)	MAX-μ-IO	ID-IO (Low)	ID-IO(Hg)
☞ ōh					
āh				*!	
ūh	*!				*
īh	*!				*
ōw		*!			
ō			*!		

For Dari, the systematic loss of /h/ is accounted for with a constraint against glottal segments (to be covered in section VI.H.2):

MP*ūh	*V[+H][h]]$_\sigma$	*Glottal	ID-IO (Glot)	MAX-μ-IO	ID-IO (Low)	ID-IO(Hg)
ōh		*!				*
āh		*!			*	*
ūh	*!	*				
īh	*!	*				
ōw			*!			*
☞ ō				*		

MP*ōh	*V[+H][h]]$_\sigma$	*Glottal	ID-IO (Glot)	MAX-μ-IO	ID-IO (Low)	ID-IO(Hg)
ōh		*!				
āh		*!			*	
ūh	*!	*				*
īh	*!	*				*
ōw			*!			
☞ ō				*		

Again, it is unclear why Tajik allows [nu].

<u>VI.G.1.d /ō/: MP /ō/ → Dari /ū/ and Iranian /ō/</u>

- /ō/ → /ō/ (Iranian) (two cases)
  → /ū/ (Dari and Tajik)

Examples: MP dō → "دو" Iranian [do] Dari/Tajik [du] 'two'

MP tō → "تو" Iranian [to] Dari/Tajik [tu] 'you'

In these final two instances, it will be shown that in Dari, word final /ō/ → /ū/, where in Iranian /ō/ remains. In all other instances of word-final /ō/, Dari retains /ō/, and Iranian shifts it to /ū/:

MP:	Mod P:	Dari:	Iranian:	Tajik:	Gloss:
abrō	“ابرو”	[a-bro^{l}]	[æ-bru^{l}]	[a-br~~u~~^{l}]	‘eyebrow’
nērōg	“نیرو”	[ne-ro^{l}]	[ni-ru^{l}]	[ne-r~~u~~^{l}]	‘power’

Thus, either Middle Persian /dō/ and /tō/ are actually graphemically misrepresented, or they are merely exceptions. In either case, it is impossible to ascertain from the available data. That leads us to the unfortunate and non-descriptive conclusion that they are anomalous.

It is certainly worth mentioning that these two lexical items are both highly frequent and highly basic. It is beyond the scope of this paper, but will be mentioned often and briefly covered in section VII that frequency and register of lexical items appear to affect the availability of changes under consideration. If this indeed proves true, it is satisfying that all four of the abovementioned exceptions are highly frequent and occur frequently in the basest of register.

VI.G.2 Iranian /ɛ/

It was mentioned in section IV.C.3 that Iranian systematically converts word-final /-a/ into [ɛ], and these remain /-a/ in Dari and Tajik:

	Mod Pers:	Iranian:	Dari:	Tajik:	Gloss:
Noun:	“خانه”	[xu-nɛ^{l}]	[xɑ-na^{l}]	[xo-na^{l}]	‘house’
Adjective:	“کهنه”	[koh-nɛ^{l}]	[kʊ-na^{l}]	[k~~u~~-na^{l}]	‘old’
Participle:	“کرده”	[k^{j}ær-dɛ^{l}]	[ka-da^{l}]	[ka(r)-da^{l}]	‘done’

3pers Sing:	"می کند"	[miˈ-ko-nɛ]	[meˈ-kʊ-na]	[mʲeˈ-kʊ-na][45]	'does'

In addition to these, there are a number of other instances in which Iranian realizes [ɛ] and even [i] where Dari and Tajik realize [a], as for instance the following:

MP:	Mod Pers:	Iranian:	Dari:	Tajik:	Gloss:
ʃaʃ	"شش"	[ʃɛʃ] or [ʃiʃ]	[ʃaʃ]	[ʃaʃ]	'six'
kas+i	"کسی"	[kɛˈ-si]	[kaˈ-sɛ]	[kaˈ-si]	'someone'
kūtʃak	"کوچک"	[ku-tʃikˈ]	[ku-tʃakˈ]	n/a	'small'
ras-	"رسیدن"	[rɛ-si-dænˈ]	[ra-si-danˈ]	[ra-si-danˈ]	'to arrive'

One sidenote that seems to add a great deal to this analysis is that the Middle Persian form "kas" can optionally be "keč," indicating that a palatal segment adjacent to the vowel is likely the operative element here in raising the vowel. This being true, and also noting that in Iranian, and only Iranian, the Middle Persian short low vowel /a/ is a front vowel, it may be claimed that this shift began as the raising of a front vowel adjacent to a palatal (see VI.G.2.b for the analysis).

The continuation of section VI.G.2 will be as follows: Section VI.G.2.a will deal with the systematic raising of word-final /-a/ to [ɛ] in Iranian, VI.G.2.b will deal with systematic non-word-final /-a/ to [ɛ] in Iranian, and VI.G.2.c will deal with exceptions to the abovementioned norm.

[45] It bears at least passing mention that Southern Tajik will allow for ellipsis of the syllable /kʊ/, leaving only /mje-na/ as the output.

VI.G.2.a Iranian Word-Final /ɛ/

The following constraint against open short (or +Low –ATR) vowels in word-final position is proposed based on the notion that in word-final position, the mouth will close in anticipation of a rested state (closure), unless the vocalic segment has the length to sustain its articulation. The specification of -ATR presupposes that ATR and durational length are indeed connected at the time of this transformation (between Middle Persian and Proto-Persian) and so in actuality it need not be specified because *+Long –Low (see section VI.D.2) will be relevant to this analysis. A similar constraint will cause /ʃaʃ/ → [ʃɛʃ], as the state of the articulators during articulation of palatal sounds is naturally higher:

MAX Height]# ("Maximize the height of word-final vowels.")

MP*a]	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	*+Lng -Low	MAX-Hg]#	ID-IO (Low)
a]					**!	*
ō]	*!			*		
i]			*!			
ā]	*!				**	
☞ ɛ]					*	*
∅]		*!				
ad]	*!				**	*

Meanwhile, word-final /ā/ is preserved in lexical items such as "جا" /□ɑ/ 'place' and "آمریکا" [ɑm-ri-kɑˈ] 'America':

MP*ā]	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	*+Lng -Low	MAX-Hg]	ID-IO (Low)
a]		*!		*	*	
ō]				*!		*
i]		*!				*
☞ ā]						
ɛ]		*!				*
ē]				*!		*
∅]		*!				

VI.G.2.b Iranian [ɛ] and Adjacent Palatals

Mentioned in the introduction to this section is another class of phonological conditioning causing /a/ to [ɛ] in Iranian, and that is in positions adjacent to palatals:

MP:	Mod Pers:	Iranian:	Dari:	Tajik:	Gloss:
ʃaʃ	“شش”	[ʃɛʃ] or [ʃiʃ]	[ʃaʃ]	[ʃaʃ]	‘six’
kūʧak	“کوچک”	[ku-ʧikˈ]	[ku-ʧakˈ]	[ku-ʧakˈ]	‘small’
kas+i	“کسی”	[kʲɛˈ-si]	[kaˈ-sɛ]	[kaˈ-si]	‘someone’

This is bolstered by the observation that there are two forms of the third lexical item in Middle Persian, /kas/ and /keʧ/. The following constraint is similar to MAX Height]# above and will bring about the raising attested without affecting other vowels:

C[+pal]$_n$ [MAX Height] C[+pal]$_n$ (“Maximize the height of vowels adjacent to palatal segments.”)

This will crucially occur after Iranian palatalization of velars, which will be covered in section VI.H.7.a:

MP*ʃaʃ	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	ID-IO (Bck)	*+Lng -Low	ID-IO(pal)	Pl[MX-Hg]Pl
ʃaʃ							**!
ʃɑʃ	*!			*	*		**
ʃiʃ			*!				
ʃuʃ			*!	*	*		
☞ ʃɛʃ							*
ʃʃ		*!					
sas						*!	**

Furthermore the alternation between [ʃɛʃ] and [ʃiʃ] in Iranian will be seen to occur as an optional demontion of the constraint ID-IO (High):

MP*ʃaʃ	DEP-μ-IO	MAX-μ-IO	ID-IO (Bck)	*+Lng -Low	ID-IO(pal)	Pl[MX-Hg]Pl	ID-IO(High)
ʃaʃ						**!	
ʃɑʃ	*!		*	*		**	
☞ ʃiʃ							*
ʃuʃ			*!	*			*
ʃɛʃ						*!	*
ʃʃ		*!					
sas					*!	**	

Meanwhile, other low vowels adjacent to palatals, such as the /ɑ/ of “شاش” /ʃɑʃ/ ‘urine,’ remain:

MP*ʃɑʃ	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	ID-IO (Bck)	*+Lng -Low	ID-IO(pal)	Pl[MX-Hg]Pl
ʃaʃ		*!		*			*
☞ ʃɑʃ							*
ʃiʃ			*!	*	*		
ʃuʃ			*!		*		
ʃɛʃ		*!		*			*
ʃʃ		*!					
sɑs						*!	

Another element to this raising is suggested by Lazard (1992, p. 20), in which he posits that certain of these /a/ → /ɛ/ shifts occur due to anticipatory harmony with a following high front vowel:

“رسیدن”	/ra-sī-dan/	→	[rɛ-si-dæn]			‘to arrive’
“می رسد”	/mē-ra-sad/	→	/mi-ra-sɛ/	→	[mi-rɛ-sɛ]	‘he arrives’
“رفیق”	/ra-fīq/	→	[rɛ-fiɣ]			‘friend’

The data here are inconclusive and would do nothing to elucidate the following examples:

MP:	Mod Pers:	Iranian:	Dari:	Tajik:	Gloss:
ēk	“یک”	[jɛk]	[jak]	[jak]	‘one’
unknown[46]	“سفید”	[sɛ-fidˈ]	[sa-fedˈ]	[sa-fedˈ]	‘white’
pidar	“پدر”	[pɛ-dærˈ]	[pa-darˈ]	[pa-darˈ]	‘father’

Furthermore a handful of Dari lexical items also exhibit raising:

MP:	Mod Pers:	Iranian:	Dari:	Tajik:	Gloss:
unknown	“شنبه”	[ʃæn-bɛˈ]	[ʃan-bɛˈ]	[ʃan-bɛˈ][47]	‘Saturday’
unknown	“به”	[bɛ]	[bɛ]	[bɛ]	‘b’

This should be compared with the normative example:

[46] While this is not attested, the token in OP is çpaita, indicating that there was no historical interstitial vowel and in fact the Modern Pashto token has not provided an epenthetic vowel: “سپین” /spin/ ‘white’

[47] Note that the nasal consonant in all three of these becomes bilabial (*/ʃan-be/ → [ʃam-bɛ]) in anticipation of the labial onset of the syllable that follows, but this is not relevant to the current discussion.

MP:	Mod Pers:	Iranian:	Dari:	Tajik:	Gloss:
?pad	“به”	[bɛ]	[ba]	[ba]	‘to, toward’

Finally there is even on case in which the Iranian vowel /a/ remains (the only word-final /a/ in Iranian) and is raised to [ɛ] in Dari and Tajik:

MP:	Mod Pers:	Iranian:	Dari:	Tajik:	Gloss:
nē	“نه”	[næ]	[nɛ]	[nɛ]	‘no’

Clearly more is at play here than can be systematically qualified.

VI.G.3 Iranian Pre-Nasal /ɑ/ → [u]

Dari speakers have been known to say “زبان به تهران رفت، زبون شد.” /za-bɑnˈ ba teh-rɑnˈ raft, za-bunˈ ʃʊd/, which means ‘Language (/za-bɑnˈ/) went to Tehran and became degraded (/za-bunˈ/).’ The implications of the above are that Iranians perverted the language by changing the /ɑ/ in the word /za-bɑnˈ/, meaning ‘language,’ to /za-bunˈ/, the Iranian pronunciation of /za-bɑnˈ/ and also a homophone with the word for ‘low; degraded.’ This tongue and cheek comment addresses /ɑ/ → [u] / __ [+nas], a phenomenon of Iranian mentioned in section IV.C.5, that essentially mandates that /ɑ/ → [u] before a nasal segment. This rule illustrates the weakness of most linguistic theoretical models in that this phenomenon is not possible in all instances of /ɑn/ or /ɑm/. In fact, the factors at play may be frequency or register effects or even both:

“نان” /nɑn/ → [nun] and *[nɑn]

“خانه” /xɑ-nɛˈ/ → [xu-nɛˈ] and *[xɑ-nɛˈ]

“زبان” /zæ-bɑnˈ/ → [zæ-bunˈ] and [zæ-bɑnˈ]

“قوران” /ɢur-ʔɑnˈ/ → [ɢur-ʔɑnˈ] and *[ɢur-ʔunˈ]

Native intuitions were elicited to begin assessment of the following as it deals with social and register effects. When a native speaker of Iranian is asked why /xɑ-nɛˈ/ must convert to [xu-nɛˈ] and not remain *[xɑ-nɛˈ], the response is that the word is simply [xu-nɛˈ] and not *[xɑ-nɛˈ], but when asked why [ɢur-ʔɑnˈ] cannot be changed to *[ɢur-ʔunˈ], the response is that the Qur’an is the Qur’an. It is also noted that [ɢur-ʔunˈ] is too like [ɢɛr-ʔunˈ], which means ‘penny.’ Clearly, there is a cultural or social effect or at least an ambiguity avoidance effect preventing a normative shift in certain instances.

Beyond this, register seems to affect the change as well. In the case of the plurals, it is observed that in more literary forms (or at least forms perceived to be more literary), the alternation is not allowed. Before an example, a note on Persian pluralization is warranted. In Persian, there are three normative plurals:

1. /-hɑˈ/ for inanimate nouns: /barg/ (‘leaf’) → /barg-hɑˈ/ (‘leaves’)
2. /-ɑnˈ/ for animate nouns: /mardˈ/ (‘man’) → /mar-dɑnˈ/ (‘men’)
3. Original Arabic pluralization: /waqt/ (‘time’) → /aw-qɑtˈ/ (‘times’)

Dari is the most conservative in terms of maintaining the above pluralizations, but it should be noted that even in Dari, the ending /-hɑˈ/ can be used for both animate and Arabic lexemes:

/mardˈ/ (‘man’) → /mar-dɑnˈ/ (‘men’) or /mard-hɑˈ/ (‘men’)

/waqt/ (‘time’) → /aw-qɑtˈ/ (‘times’) or /waqt-hɑˈ/ (‘times’).

That being established, the Iranian dialect tends in common parlance towards leveling the plural morphology as /-hɑˈ/ in all cases, and hence other forms are considered somewhat literary or dated. This morphology relates to the /ɑ/ → [u] / __ [+nas] shift in the following way. In Iranian, the following is noted:

“ایرانی”	/i-rɑ-niˈ/	→	[i-ru-niˈ]		‘Iranian’
“ایرانیها”	/i-rɑ-ni + hɑˈ/	→	[i-ru-ni-hɑˈ]		‘Iranians’
“ایرانیان”	/i-rɑ-ni + ɑnˈ/	→	[i-rɑ-ni-ɑnˈ] but	*[i-ru-ni-ɑnˈ] *[i-ru-ni-unˈ]	‘Iranians’

When surveyed, native speakers of Iranian explained the above data by insinuating that the /-ɑn/ ending was more proper, and hence the more colloquial /ɑ/ → [u] / __ [+nas] alternation could not take place. This appears to denote that while social factors are indeed at play, register plays a role as well.

Finally, there is the consideration of proper names. In Iranian, proper names or titles of respect containing the abovementioned combination will under no circumstances allow this shift:

“پدرام”	/pɛ-drɑm/	→	*[pɛ-drum]	‘Pedram- proper name’
“خانم”	/xɑ-næm/	→	[xu-næm]	‘Mrs.’

But in the case of names that bear substantive meaning, they may only shift in the case that the lexical item is a common noun, and not a proper name:

“غلام”	/ɣo-lɑm/	→	*[ɣo-lum]	‘Gholam- proper name’
“غلام”	/ɣo-lɑm/	→	[ɣo-lum]	‘slave’

This is fodder for future study, but for now, OT will account for the change in the following manner:

PITCH HARMONY (NASAL) (“Minimize frequency of vowel before a nasal.”) This is based in the acoustic fact that nasal sounds are much lower frequency, and hence the preceding vowel harmonizes by dropping frequency as well:

MP*ɑn	DEP-μ-IO	MAX-μ-IO	ID-IO(Back)	PH (Nasal)	ID-IO (Low)
an		*!	*	*	
ōn				*!	*
in		*!	*		*
ā		*!	*	*	
☞ ūn					*
∅		*!			
ɑn				*!	

One final note on this is that this shift has been cited as an Iranian shift, when in the case of the distal demonstrative pronoun “آن” /ɑn/, Dari allows the change too (See section VI.H.4). This is also a frequency effect on the change, as Iranian does mandate the change in all highly frequent tokens and disallows it in infrequent tokens where Dari

only allows it in the most frequent VC[+Nas] token in the language. Again, this is fodder for further study.

VI.G.4 The Status of Tajik /ʉ/

It has now been suggested several times that social and/or frequency effects likely play a role in shifts that are either optional or conditional. The case of Tajik /ʉ/ represents another question about how OT will handle language change in that /ō/ → /ʉ/ is neither optional nor conditional, but rather regional. That question is also beyond the scope of this paper, but worth adding to the note on the potential adequacy of any theory. How does a phonological theory account for the regional spread of language change? This question adds to the question raised in section VI.F.1, of how language change really takes place. Not only did speakers of early English not decide one day to have a great vowel shift, it surely did not happen to all speakers at once, but rather over time and space. Tajik appears to be in such a flux of time and space, and thus it is refered to in two dialects, northern and southern.

VI.H Proto-Persian Consonants: Introduction

Four wholesale consonant shifts were mentioned in Chapters III through V between Proto-Persian and the three principal dialects of today. One of these abovementioned shifts was assessed at length in sections VI.F.2 and VI.F.3, and so will not be reexamined here:

/υ/ → [w] (in Dari)

/υ/ → [v] (in Iranian and Tajik)

The second such shift is the universal and phonologically conditioned shift towards aspiration of voiceless simple onset plosives that will be dealt with in section VI.H.1:

1. V[+plos] → [+Asp]

The third shift, which affects all three dialects is the loss of glottal consonant segments /ʔ/ and /h/ and its subsequent phonological effects. This will be covered in section VI.H.2 and its subsections:

2. /ʔ/ → ∅ (to varying degrees in all three)
   /h/ ↗

The fourth and fifth consonant shifts that affect the dialects to varying degrees are r-drop and n-drop (sections VI.H.3 and VI.H.4):

3. /r/ → ∅ (to varying degrees in all three)

4. /n/ → ∅ (to varying degrees in Dari and Tajik)

Sections VI.H.5 and VI.H.6 will deal with two reductions peculiar to Dari, and they are indeed considered peculiar to Iranian speakers who have not encountered the forms before. In both of these instances, Dari consonants reduce to semivowels:

5. /ʤ/ → [j] (In a few Dari tokens)

6. /f/ → [w] (In a few Dari tokens)
   /b/ ↗

Section VI.H.7 will deal with Iranian and Tajik palatization, which not only does not occur in Dari, but its occurrence in Tajik is entirely distinct from Iranian in its phonological conditioning:

7. C [+Vel] → C^j / __ [æ] (in Iranian)

C → C^j / __ V[+Front] (in Tajik)

For the sake of interest, it should be noted that borrowed Arabic and Persian lexical items in Turkish will exhibit a palatalization effect identical to that of Iranian:

Mod Pers:	Turkish:	Iranian:	Turkish:	Dari:	Gloss:
“کباب”	kebab	[k^jæ-bɑbˈ]	[k^jɛ-babˈ]	[ka-bɑbˈ]	‘kabab’
“هر کس”	herkes	[hær#k^jæs]	[hɛr-k^jɛs]	[har#kas]	‘everyone’

Section VI.H.8 will cover the Iranian treatment of original /q/, which remains in Dari and Tajik, but exhibits wide allophonic variation in Iranian:

8. /q/ → */G/ → [G] / #__
__ C[+Voice]
→ [ɣ] / V__V (In Iranian)
→ [χ] / __ C[-Voice]

Finally, section VI.H.9 will cover Tajik’s optional loan segment [ɫ], as borrowed from Russian during the Soviet occupation:

9. /l/ → [ɫ] (In Tajik)

VI.H.1 Proto-Persian Consonants: Aspiration (Dari, Iranian, Tajik)

Aspiration is a logical starting point, as it exists in all three dialects and its distribution matches that of English, in word-initial simple onset voiceless plosives:

C[+plosive –voice]V → C[+asp]V

Thus, a constraint on allophonic aspiration based on ease of parsing (as this maximizes the distinction between the voiceless plosive and its voiced counterpart) can be wrought as follows:

C^hV ("word-initial simple onset voiceless plosives aspirate.")[48]

MP*p	C^hV	ID-IO(voice)	ID-IO (Lab)	MAX-μ-IO	DEP-μ-IO
☞ p^h					*
p	*!				
b		*!			
b^h		*!			*
h			*!		
∅				*!	

VI.H.2 Proto-Persian Consonants: h/ʔ-drop (Dari, Iranian, Tajik)

All three dialects exhibit some degree of glottal ellipsis, from colloquial Dari, which never phonologically realizes /h/ or /ʔ/, to Iranian which retains the option of ellipsis, but more frequently with /ʔ/ than with /h/. None of the dialects retain the original /ʕ/ in Arabic loanwords save occasionally in the affected speech of a Mullah or Imam, but never in natural or common parlance. In the below, the underlying form also represents

[48] The specification of word initial and simple onset are for convenience sake, but could just as easily be accounted for with additional phonotactic constraints against aspiration within a complex onset.

the formal register pronunciation, and the surface form hence represents the colloquial register pronunciation:

	Dari:	Iranian:	Tajik:	Gloss:
"همه"	/haˈ-ma/ → [aˈ-ma]	/hæ-mɛˈ/ → [hæ-mɛˈ]	/haˈ-ma/ → [haˈ-ma]	'all'
"بهار"	/ba-hɑrˈ/ → [ba-ɑrˈ]	/bæ-hɑrˈ/ → [bæ-hɑrˈ]	/ba-horˈ/ → [ba-horˈ]	'spring'
"صبح"	/sʊbḥ/ → [sʊb]	/sobḥ/ → [sob(ḥ)]	/sʊbḥ/ → [sʊb]	'morning'
"بعد"	/ba-ʔadˈ/ → [bɑd]/[bad][49]	/bæ-ʔædˈ/ → [bæːd]	/ba-ʔadˈ/ → [bad]	'after'

Immediately, the following distinctions between Iranian and Dari/Tajik become clear:

1. All three dialects will call for ellipsis of /ʔ/, but in Iranian, there is a phonological remnant.
2. Dari will delete every case of /h/ in colloquial speech, and Iranian retains it almost without exception. Tajik and Iranian ellipsis of /h/ will be proven to be predictable. In the former, h-drop is disallowed in syllable onset, and in the latter, the environment, while predictable, is more complex.
3. Dari (and only Dari) exhibits a qualitative vowel change in many lexical items associated with the h-drop that involves either the reduction of surrounding vowels or compensatory lengthening mentioned above. In colloquial Iranian,

[49] See section VI.H.2.a for an analysis of /aʔ/ → [ɑ].

the ellipsis results only in the quantitative compensatory lengthening, and in Tajik, there will not be quantitative compensatory lengthening:

Mod Pers:	Iranian:	Tajik:	Dari:	Gloss:
"قهر"	[ɢæhr̩] or [ɢæːr]	[qar]	[qɑr]	'angry'
"معلم"	[mo-ʔæ-lɛmˈ] or [mwæ-lɛmˈ]	[ma-ɬimˈ]	[ma-lɛmˈ]	'teacher'

Note that in both of the above cases, Iranian remains most faithful to the initial form preserving the original vowel quality in the first example and preserving two distinct vowels in the second.

VI.H.2.a h/ʔ-drop and Compensatory Lengthening (Dari)

It was mentioned in section III.E 4 that Dari and only Dari exhibits a sound change associated with the ellipsis of /h/ and /ʔ/ in colloquial speech. Below was the sample token illustrated in section III.B and III.E:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
"بعد"	/baʔd/ →[bad]	[bɑd]	'after'
"قهر"	/qahr/ →[qar]	[qɑr]	'angry'
"فهمیدن"	/fah-mi-danˈ/ →[fa(h)-mi-danˈ]	[fɑ-mi-danˈ]	'understand'

OT easily accounts for this by considering glottal sounds marked (*Glottal), and barring ellipsis. This constraint combined with Max-Low (+ATR +Back) from section VI.C.3 will lead to the change */aʔ/ → [ɑ] and */ah/ → [ɑ] in Dari:

MP*aʔ	*Glottal	ID-IO(low)	Lw(+A+Bk)	MAX-μ-IO	DEP-μ-IO
☞ ɑ				*	
aʔ	*!		*		
aː			*!		
a			*!	*	
i		*!		*	

Iranian will simply maintain MAX-μ-IO over Max-Low (+ATR +Back):

MP*aʔ	*Glottal	ID-IO(low)	MAX-μ-IO	Lw(+A+Bk)	DEP-μ-IO
ɑ			*!		
aʔ	*!			*	
☞ aː				*	
a			*!	*	
i		*!			

Tajik will mirror Iranian, save that a constraint *VV (to be covered in section VI.H.4.b.iv) will prohibit the quantitatively long vowel /aː/. Further recall from section VI.D.1 that in Tajik *Low +ATR prohibits the segment /ɑ/:

MP*aʔ	*Glottal	ID-IO(low)	*VV	*Low+ATR	MAX-μ-IO	Lw(+A+Bk)	DEP-μ-IO
ɑ				*!	*		
aʔ	*!					*	
aː			*!			*	
☞ a					*	*	
i		*!					

VI.H.2.b Exceptions to Compensatory Lengthening (Dari)

Two homophonous exceptions to this alternation are:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
"بهر"	/bahr/ → *[baːr], *[bɑr]	[bar]	'for; to'
"بحر"	/bahr/ → *[baːr], *[bɑr]	[bar]	'sea'

These exceptions seem to have no defining features that would set them apart as phonologically conditioned and hence will be considered to have retained the original vowel quality to distinguish them both from a highly frequent lexical item "بار" [bɑr] meaning 'time; instance.' If there is any truth at all to frequency effects in phonology (to be considered briefly in Chapter VII), less frequent lexical items will not interfere with this process, and that appears to be the case:

"بعد" /baʔd/ [bɑd] 'after'

"باد" /bɑd/ [bɑd] 'wind; be!-2nd pers singular imperative'

These above two cases stand in contrast to the lexical item "بهار" /ba-hɑrˈ/ meaning 'spring.' In this item and other items in which the glottal segment elides between a long vowel and a short vowel, the qualities of each vowel survive the ellipsis and in some cases will epend a glide:

Mod Pers:	Formal Dari:		Spoken Dari:	Gloss:
"بهار"	/ba-hɑrˈ/	→	/ba-ɑrˈ/	'spring'
"بهانه"	/bʊ-hɑ-naˈ/	→	/bʊ-(w)ɑ-naˈ/[50]	'excuse'

There is one exception in which /ɑʔa/ optionally reduces to [ɑ] in the highly frequent lexical item below:

[50] Glide-insertion in Dari will be dealt with in sections VI.H.2.e (for front glide insertion) and VI.H.2.f (for back glide insertion).

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“ساعت”	/sɑ-ʔatˈ/ →	[sɑ-atˈ] → [sɑt]	‘hour; clock’

Thus, section VI.H.2.c will revise the above tableaux to account for the fact that vowel quality is retained in cases of ellipsis between vowel segments mismatched for length.

VI.H.2.c Compensatory Lengthening in Non-/ah/ and /aʔ/ Cases (Dari)

In other instances, the final reduction does not occur both with /ɑʔa/ and /aʔɑ/:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“آهن”	/ɑ-hanˈ/	[ɑ-anˈ]	‘iron’
“بهار”	/ba-hɑrˈ/	[ba-ɑrˈ]	‘spring’

MP*ɑʔa	*Glottal	ID-IO(low)	Lw(+A+Bk)	MAX-μ-IO	ID-IO (Bk)	DEP-μ-IO
☞ ɑa			*	*		
aʔ	*!		*			
aː			*	*	*!	
a			*	**!	*	
ɑ				**!		
i		*!		*		

And the same holds for /aʔɑ/:

MP*aʔɑ	*Glottal	ID-IO(low)	Lw(+A+Bk)	MAX-μ-IO	ID-IO (Bk)	DEP-μ-IO
ɑ				**!		
aʔ	*!		*			
☞ aɑ			*	*		
a			*	**!		
i		*!				

VI.H.2.d /h/ and /ʔ/ in Verbal Morphology

In two examples of verb stem morphology, the ellipsis of /h/ results in different phonological effects, semivowel epenthesis in one and devoicing of an adjacent consonant in another:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
"می خواهم"	/meˈ-xɑ-ham/ → *[meˈ-xɑ-am]	[meˈ-xɑ-jam]	'I want'
"می دهم"	/meˈ-da-ham/ → [meˈ-dʰam] →	*[meˈ-tam], [meˈ-təm]	'I give'

The above two examples will have to be treated individually, as there is no phonological conditioning apparent. We will return to the above two verbal conjugation examples in section VI.J.1, but a note should be made about the availability in Dari of C[+Glottal] → Glide, as the following two sections, VI.H.2.e and VI.H.2.f, will illucidate:

VI.H.2.e Dari /h/ → [j] and /ʔ/ → [j]

Aside from the compensatory lengthening already addressed, Dari has available to it the ability to append a glide in place of the elided segment. This is a fairly natural process and as expected, and because Tajik does not elide syllable-initial /h/, this is considered a Dari phenomenon, as far as /h/ ellipsis is concerned:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“مجاهد”	/mʊ-ʤɑ-hɛdˈ/	[mʊ-ʤɑ-jɛdˈ]	‘jihadist’
“قاعده”	/qɑ-ʔɛ-daˈ/	[qɑ-jɛ-daˈ] → [qɑj-daˈ]	‘base; Qaeda’
“جهان”	/ʤɛ-hɑnˈ/	[ʤɛ-ɑnˈ] → [ʤɛ-jɑnˈ]	‘world; universe’

Note that in Tajik, these will be [mʊ-ʤo-hidˈ], [qoj-daˈ] and [ʤi-honˈ] respectively. This phenomenon of Dari is very simply accounted for by constraints already provided:

MP*ɑʔɛ	*Glottal	MAX-μ-IO	ID-IO(Lab)	DEP-μ-IO	ID-IO (closed)	ID-IO(pal)
ɑɛ		*!				
ɑʔɛ	*!					
☞ ɑjɛ					*	*
ɑpɛ			*!			
ɑwɛ			*!		*	

MP*ɛʔɑ	*Glottal	MAX-μ-IO	ID-IO(Lab)	DEP-μ-IO	ID-IO (closed)	ID-IO (pal)
ɛɑ		*!				
ɛʔɑ	*!					
☞ ɛjɑ					*	*
ɛpɑ			*!			
ɛwɑ			*!		*	

The same will naturally hold for /ɑhɛ/ and /ɛhɑ/:

MP*ɑhɛ	*Glottal	MAX-μ-IO	ID-IO(Lab)	DEP-μ-IO	ID-IO (closed)	ID-IO (pal)
ɑɛ		*!				
ɑhɛ	*!					
☞ ɑjɛ					*	*
ɑpɛ			*!			
ɑwɛ			*!		*	

MP*εhɑ	*Glottal	MAX-μ-IO	ID-IO(Lab)	DEP-μ-IO	ID-IO (closed)	ID-IO (pal)
εɑ		*!				
εhɑ	*!					
☞ εjɑ					*	*
εpɑ			*!			
εwɑ			*!		*	

There is however one case in Iranian in which the ellipsis of /h/ results in more reduction than its Dari counterpart, and that is in the morphology of the verb “خواستن” /xɑs-tæn/ ‘to want.’ Recall the Dari conjugation, and compare with Iranian:

Mod Pers:	PP:	Dari:	Iranian:	Gloss:
“می خواهم”	/meˈ-xɑ-ham/	[meˈ-xɑ-jam]	[miˈ-xɑm]	‘I want’

This shall be returned to in section VI.J.1.

VI.H.2.f Dari /h/ → [w] and /ʔ/ → [w]

Much like the data in section VI.H.2.e, Dari allows for C[+Glottal] → Glide when adjacent to back vowels as well, and naturally, the glide will be the back glide:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“بهانه”	/bʊ-hɑ-naˈ/	[bʊ-wɑ-naˈ]	‘excuse’
“یهود”	/ja-hudˈ/	[ja-udˈ] → [ja-wudˈ]	‘Jewish’
“آهو”	/ɑ-huˈ/	[ɑ-uˈ] → [ɑ-wuˈ]	‘deer’

This, as in section VI.H.2.e, is fairly straightforward:

MP*ʊʔɑ	*Glottal	MAX-μ-IO	ID-IO(plos)	ID-IO(pal)	ID-IO (closed)	ID-IO (lab)
ʊɑ		*!			*	
ʊʔɑ	*!					
☞ ʊwɑ					*	*
ʊpɑ			*!			*
ʊjɑ				*!	*	

MP*ʊhɑ	*Glottal	MAX-μ-IO	ID-IO(plos)	ID-IO(pal)	ID-IO (closed)	ID-IO (lab)
ʊɑ		*!			*	
ʊhɑ	*!					
☞ ʊwɑ					*	*
ʊpɑ			*!			*
ʊjɑ				*!	*	

If MAX-μ-IO is dropped below the ranked constraints in the above tableaux, the segment without the glide will be generated in the output:

MP*ʊʔɑ	*Glottal	ID-IO(plos)	ID-IO(pal)	ID-IO (closed)	ID-IO (lab)	MAX-μ-IO
☞ ʊɑ				*		*
ʊʔɑ	*!					
ʊwɑ				*	*!	
ʊpɑ		*!			*	
ʊjɑ			*!	*		

MP*ʊhɑ	*Glottal	ID-IO(plos)	ID-IO(pal)	ID-IO (closed)	ID-IO (lab)	MAX-μ-IO
☞ ʊɑ				*		*
ʊhɑ	*!					
ʊwɑ				*	*!	
ʊpɑ		*!			*	
ʊjɑ			*!	*		

The same will naturally hold for /ɑʔu/ and /ɑhu/:

MP*aʔu	*Glottal	MAX-μ-IO	ID-IO(plos)	ID-IO(pal)	ID-IO (closed)	ID-IO (lab)
au		*!			*	
aʔu	*!					
☞ awu					*	*
apu			*!			*
aju				*!	*	

MP*ahu	*Glottal	MAX-μ-IO	ID-IO(plos)	ID-IO(pal)	ID-IO (closed)	ID-IO (lab)
au		*!			*	
ahu	*!					
☞ awu					*	*
apu			*!			*
aju				*!	*	

And of course in the instances in which the epenthetic glide is not inserted, the constraint MAX-μ-IO is lowered:

MP*aʔu	*Glottal	ID-IO(plos)	ID-IO(pal)	ID-IO (closed)	ID-IO (lab)	MAX-μ-IO
☞ au				*		*
aʔu	*!					
awu				*	*!	
apu		*!			*	
aju			*!	*		

MP*ahu	*Glottal	ID-IO(plos)	ID-IO(pal)	ID-IO (closed)	ID-IO (lab)	MAX-μ-IO
☞ au				*		*
ahu	*!					
awu				*	*!	
apu		*!			*	
aju			*!	*		

VI.H.2 .g h/ʔ-drop and Compensatory Lengthening (Iranian)

As mentioned in section IV.E.3, Iranian also exhibits ellipsis of /h/ and /ʔ/, but typically has associated with it some degree of qualitative compensatory lengthening.

In Iranian, the ellipsis is environmentally conditioned except in the case of the optional deletion of word-initial /h/ in two highly frequent lexical items:

Mod Pers:	PP:	Iranian:	Dari:	Gloss:
“قهر”	/qahṛ/	[ɢæhṛ] or [ɢæːr]	[qɑr]	‘angry’
“معلم”	/maʔ-lim/	[mo-ʔæ-lɛmˈ] or [mwæ-lɛmˈ]	[ma-lɛmˈ]	‘teacher’
“هم”	/ham/	[hæm] or [æm]	[am]	‘also’
“ها”	/hā/	[hɑ] or [ɑ]	[ɑ]	‘suffix-plural’

The phonological conditioning on h-drop and ʔ-drop in Iranian as mentioned in chapter IV is as follows:

/ʔ/ →	∅ / V__#	“ممنوع” /mam-nuʔ/ →	[mam-nu] ‘prohibited’
	:∅ / __C; C__	“معلوم” /mæʔ-lum/ →	[mæː-lum] ‘evident’
		“دفعه” /dæf-ʔɛ/ →	[dæ-fːɛ] ‘he understood’
/h/ →	:∅ / V__	“شهر” /ʃæhr/ →	[ʃæːr] ‘city’
	:∅ / __C; C__	“فهمید” /fæh-mid/ →	[fæː-mid] ‘he understood’
		“صفحه” /sæf-hɛ/ →	[sæ-fːɛ] ‘page’
	(:)∅ / V__V	“چهل” /tʃɛ-hɛl/ →	[tʃɛl] or [tʃel] ‘forty’

Because the Iranian quantitative compensatory lengthening appears to be the norm, I will treat all cases above as /h/ (or /ʔ/) → :∅, and then consider those cases of /h/ or /ʔ/ → ∅ to have undergone reanalysis as Vː, and the constraint *VV (to be covered in

section VI.H.4.b.iv) to have caused V: → V to occur. The following constraints will apply:

*Glottal

ID-IO (#[C) ("Word-initial consonants are preserved." Ease of parsing obviates this faithfulness constraint.)

MP*/ mam-nuʔ/	ID-IO (#[C)	*Glottal	ID-IO(low)	MAX-μ-IO	DEP-μ-IO
☞ mam-nu				*	
mam-nuʔ		*!			
mam				***!	

MP*/mæʔ-lum/	ID-IO (#[C)	*Glottal	ID-IO(low)	MAX-μ-IO	DEP-μ-IO
☞ mæː-lum					
mæ-lum				*!	
mæʔ-lum		*!			

MP*/dæf-ʔɛ /	ID-IO (#[C)	*Glottal	ID-IO(low)	MAX-μ-IO	DEP-μ-IO
☞ dæ-fːɛ					
dæf-ʔɛ		*!			
dæ-fɛ				*!	

MP*/ʃæhr/	ID-IO (#[C)	*Glottal	ID-IO(low)	MAX-μ-IO	DEP-μ-IO
☞ ʃæːr					
ʃæhr		*!			
ʃær				*!	

MP*/ʧɛ-hɛl/	ID-IO (#[C)	*Glottal	ID-IO(low)	MAX-μ-IO	DEP-μ-IO
☞ ʧɛːl					
ʧɛ-hɛl		*!		*	
ʧɛl				*!	

Meanwhile, word initial /h/ and /ʔ/ are preserved (the latter occurs naturally):

MP*/hɑl/	ID-IO (#[C)	*Glottal	ID-IO(low)	MAX-μ-IO	DEP-μ-IO
☞ hɑl		*			
ɑl	*!				
al	*!			*!	

The case of the verb “خواستن” [xɑs-tæn] ‘to want’ (“می خواهم” /mēˈ-xā-ham/ → [miˈ-xɑm] ‘I want’) will covered in section VI.J.1.d.i.

VI.H.3 Proto-Persian Consonants: r-drop (Dari, Iranian, Tajik)

Dari allows for the optional ellipsis of word-final /r/ in certain frequently used lexical items. Iranian and Tajik allow for this alternation as well but to a lesser degree. In Iranian only the lexical items “دیگر” /di-garˈ/ ‘other’“مگر” /ma-garˈ/ ‘can it be?’ and “اگر” /æ-gærˈ/ ‘if’ will allow this ellipsis, and in Tajik, the variation will depend on dialect, northern or southern. For Dari, the following will apply:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“دیگر”	/di-garˈ/	[di-gaˈ] / [di-gɛˈ][51]	‘other’[52]
“در”	/dar/	[da]	‘in’
“اگر”	/a-garˈ/	[aˈ-ga]	‘if’

[51] The [di-gaˈ] / [di-gɛˈ] alternation reflects the difference between Dari or Tajik and Iranian in word final /a/, which becomes [ɛ] in Iranian and remains in the former two. This footnote refers to [aˈ-ga] / [a-gɛˈ] as well.

[52] As mentioned in section III.E.5, the lexical item “دیگر” /di-garˈ/ meaning other ‘other’ exhibits this ellipsis ([di-gaˈ] / [di-gɛˈ]), while “دیگر” /di-garˈ/ meaning ‘afternoon’ does not (*[di-gaˈ] / *[di-gɛˈ]). Likewise, “در” /dar/ meaning ‘in’ will allow r-drop ([da]), but “در” /dar/ meaning ‘door; opening’ will not (*[da]). This may indicate frequency or register effects, but that will be dealt with in section VII.

“اینطور”	/inˈ-tawr/	[iˈ-to]	‘like this’
“آنطور”	/ɑnˈ-tawr/	[uˈ-to]	‘like that’
“کردن”	/kar-danˈ/	[ka-danˈ]	‘to do’
“مگر”	/ma-garˈ/	[ma-garˈ]	‘but not’
“خیر”	/xajr/	[xaj]	‘okay; well’
“سیر”	/ser/	[sej]	‘satiated’
“بگیر”	/bɛˈ-gir/	[biˈ-gi]	‘get!’

In Tajik, this ellipsis will occur consisently in /di-garˈ/, /a-garˈ/ and /kar-danˈ/, but with /dar/, /inˈ-tawr/ and /ɑnˈ-tawr/, the shift will only occur in the southern dialects, which is logical, as the south borders Afghanistan.

This ellipsis will be seen to occur due to a markedness constraint on closed syllables:

$*C]_{\sigma}$ (“Closed syllables are prohibited.” This constraint is based on easy of articulation and backed up by cross-linguistic markedness.)

MP*/di-garˈ/	$*C]_{\sigma}$	ID-IO(low)	MAX-μ-IO	Lw(+A+Bk)	DEP-μ-IO
[di-garˈ]	*!				
☞ [di-gaˈ]			*	*	
[di-galˈ]	*!			*	
[di-gu]		*!	*	*	

This presents an immediate problem for the validity of the entire theory for two non-trivial reasons. First, */di-gar^{l}/ only becomes [di-ga^{l}] when it bears the meaning 'other,' and crucially not when it bears the meaning 'afternoon.' Second, even when the lexical item denotes 'other,' the ellipsis is an optional and register-related phenomenon. Chapter VII will provide suggested alterations to the theory that will account for stochastic and frequency related data, but for now, the above tableau will suffice as an explanation of the phonology of the Modern Persian r-drop, *when it occurs*. The particulars for when it does and does not occur will be dealt with later.

VI.H.3.a Proto-Persian Consonants: Onset Cluster r-drop (Dari)

In one case in Dari, /r/ that is part of an onset consonant cluster shifts to [j], rather than being elided:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
"برادر"	/brɑ-dar^{l}/	[bjɑ-dar^{l}]	'brother'

But this alternation appears anomalous, and is the only case of /brɑ/ → /bjɑ/ in the language:

"ابراظ"	/ɛ-brɑz^{l}/	*[ɛ-bjɑz^{l}]	'expression'

This is further support for a theory that incorporates frequency and register information, as the token 'brother' is both frequent and colloquial. In this lexical item, the following constraint on complex syllable onsets will produce [bjɑ-dar^{l}] as an output for input /brɑ-dar^{l}/:

*σ[CCV (“Complex syllable onsets are prohibited.”)

MP*/brɑ-darˈ/	*σ[CCV	ID-IO(vocalic)	MAX-μ-IO	DEP-μ-IO
[brɑ-darˈ]	*!			
☞ [bjɑ-darˈ]				
[blɑ-darˈ]	*!			
[baɑ-darˈ]		*!		

Again, in chapter VII, the subject will be broached of why this does not occur universally in Dari, but only in this one lexical item.

VI.H.4 Proto-Persian Consonants: n-drop (Dari, Tajik)

As with /r/ in VI.H.3, Dari and Tajik allow for ellipsis of /n/ in certain frequently used lexical items. Iranian does not allow this:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“من”	/man/	[ma]	‘I’
“این”	/in/	[i]	‘this’
“آن”	/ɑn/	[u][53]	‘that’
“کن”	/kʊn/	[ko]	‘do!’
“بچین”	/bɛ-ʧin/	[bɛ-ʧi]	‘cut!’

Tajik does not allow this ellipsis to the degree that Dari does, as section V.E.4 pointed out in that the ellipsis is only allowed in the first three of the above examples, and at that,

[53] In this example, a vowel change has taken place along with the ellipsis. The pre-nasal raising of the low back vowel [ɑ] was dealt with in chapter V and then extensively in section VI.G.3, as this lexical item seems to be the only token in which this occurs in Dari, while in Iranian /ɑ/ → [u]/__[+nas] is almost a wholesale shift. The lexical item “آن” /ɑn/, meaning ‘that,’ will be covered in sectionVI.H.4.a

only in the Southern Tajik varieties. That being established, this r-drop, like the word-final r-drop, can be accounted for with the markedness constraint $*C]_{\sigma}$ prohibiting closed syllables:

MP*/man/	$*C]_{\sigma}$	MAX-μ-IO	DEP-μ-IO
[man]	*!		
[mɑ]		*	*!
☞ [ma]		*	

Dari /kʊn/ → [ko] will be less straightforward, save that Dari has no lexical items with word-final *[-ʊ], and hence it appears that there is a general constaint against that operating in the language. Recall from sectionVI.C.1 that /ʊ/ → /o/ occurred in Iranian due to constraints on height, backness and rounding. In this case the constraint *+Back –Round ("Unrounded back segments are prohibited.") and *-ATR +High were the main operators in the shift. In Dari this shift will occur on similar grounds, but only in word-final position. The constraint is revised as follows:

*+Back –Round]# ("Back segments must be rounded word-finally.")

MP *ʊ	*+Bk-Rnd	αATRαlong	DEP-μ-IO	MAX-μ-IO	ID-IO(Bck)	ID-IO(ATR)	ID-IO(Hg)
ʊ	*!	*					
ɔ	*!						*
ʊː	*!	*!	*				
☞ o						*	*
u		*!					
ū			*!			*	
ʉ					*!		

VI.H.4.a The Problem of “آن” (/ɑn/) (Dari, Iranian, Tajik)

The flaw in the above tableau is that it does not account for the case of the distal demonstrative pronoun “آن” /ɑn/, or ‘that.’ The reason this is an issue is that post ellipsis /ɑn/ is never *[ɑ], and only [u] in all three dialects. Additionally, in its unreduced form, it may be [ɑn] or [un] as freely alternate forms again in all three dialects:

Mod Pers:	Formal Dr/Tj/Ir:	Spoken Dr/Tj:	Gloss:
“این”	/in/	[i]	‘this’
“آن”	/ɑn/ → [un]	[u]	‘that’

For this lexeme, three tableaux will be provided, one for the most formal case in which /ɑn/ → [ɑn], a second for the less formal /ɑn/ → [un], and a third for the colloquial Dari and Tajik /ɑn/ → [u]. For /ɑn/ → [ɑn], faithfulness retains the original form:

MP*/ɑn/	MAX-μ-IO	DEP-μ-IO	ID-IO(Bk)	ID-IO(Lw)	ID-IO(nas)	*C]$_\sigma$
☞ [ɑn]						*
[ɑ]	*!					
[un]				*!		*
[an]			*!			*
[ɑd]					*!	*

The second tableau deals with /ɑn/ → [un], and this tableau will mirror that which we saw in section VI.G.3, drawing on the constraint PITCH HARMONY (NASAL) (“Minimize frequency of vowel before a nasal.”) Note that the tableau is the same as the above, but PH (Nasal) has been promoted to the top of the constraint ranking:

MP*ɑn	PH (Nasal)	MAX-μ-IO	DEP-μ-IO	ID-IO(Back)	ID-IO (Low)	ID-IO(nas)	*C]$_\sigma$
an	*!	*		*			*
ōn	*!				*		*
in	*!			*	*		*
ā		*!		*		*	
☞ un					*		*
∅		*!		*	*	*	
ɑn	*!						*

The third and final tableau, dealing with /ɑn/ → [u], will be the same tableau as above with the constraint *C]$_\sigma$ (“Closed syllables are prohibited.”)

MP*ɑn	PH (Nasal)	*C]$_\sigma$	MAX-μ-IO	DEP-μ-IO	ID-IO(Back)	ID-IO (Low)	ID-IO(nas)
an	*!	*	*		*		
ōn	*!	*				*	
in	*!	*			*	*	
un		*!				*	
ā			*		*!		*
☞ u			*			*	*
∅			*		*!	*	*
ɑn	*!	*					

Immediately, there is a problem with this analysis, as the above two constraint rankings will cause all instances of /ɑ/ + [+nas] to convert to [u] + [+nas] or even simply [u] depending on the register. That is to say, without consideration of borrowing (from Iranian) and frequency effects (“آن” /ɑn/ ‘that’ is one of the most frequent lexemes in Persian, and ceraintly the most frequent of those containing /ɑ/ + [+nas]), the above tableau must be stipulatively offered as particular to the given lexical item. However, OT is employed to ground the facts in articulation and parsing, and must exist in harmony with use-based data. Thus, this dicussion will be returned to in Chapter VII.

VI.H.4.b n-drop Within Lexical Items and Phrases (Dari and Tajik)

Those same pronouns as part of locative adverbs ‘here’ and ‘there’ also reduce in Dari and Tajik, as does the lexeme “جا” /ʤɑ/ ‘place:’

Mod Pers:	Formal Dari/Southern Tajik:	Spoken Dari/Southern Tajik:	Gloss:
“اینجا”/“ин□о”	/inˈ-ʤɑ/ → [inˈ-ʤa] →	[iˈ-ʤa]	‘here’
“آنجا”/ “он□о”	/ɑnˈ-ʤɑ/ → [unˈ-ʤɑ] → [unˈ-ʤa] →	[uˈ-ʤa]	‘there’

Southern Tajik will reduce in identical fashion to the above, whereas in the northern variety, the n-drop will not occur:

Tajik:	Formal Northern Tajik:	Spoken Northern Tajik:	Gloss:
“ин□о”	/inˈ-ʤo/ → [inˈ-ʤo] →	[inˈ-ʤa]	‘here’
“он□о”	/onˈ-ʤo/ → [unˈ-ʤo] →	[unˈ-ʤa]	‘there’

In one case (and again, only in Dari), it is observed that the h-drop and n-drop conspire exhibiting compensatory lengthening in a similar fashion to that described in section III.E.4:

“من هم” /man#ham/ → [man#am] →*[ma#am] → *[maːm] → [mɑm] ‘I also’

MP*/man#ham/	*C]$_\sigma$	*Glottal	MAX-Lw(+A+B)[54]	MAX-μ-IO	DEP-μ-IO
[man#ham]	*!	*			
[man#am]	*!				
☞ [mɑm]					
[maːm]			*!		
[ma-ham]		*!			

[54] Recall the constraint Max-Low (+ATR +Back) from sections VI.A.4, VI.B and VI.C (“Maximize backness of a +ATR Low vowel.”).

Both Tajik and Farsi will exhibit the following shift in the above case:

“من هم” / “ман ҳам” /man#ham/ → [man#am] ‘I also’

The plural demonstrative pronouns in Dari also exhibit effects of both h-drop and n-drop to varying degrees depending on the register level used:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“اینها”	/in-hɑˈ/ → [in-ɑˈ] → [i-jɑˈ] → [jɑ]		‘they- here’
“آنها”	/ɑn-hɑˈ/ → *[ɑ-nɑˈ] → [u-nɑˈ][55] → [u-wɑˈ] → [wɑ]		‘they- there’

The following tableaux encapsulate all of the above permutations:

VI.H.4.b.i Dari Plural Demonstrative Pronouns- Formal Register

“اینها” [in-hɑˈ] ‘they- here (formal register)’

MP*/in + hɑ/	MAX-μ-IO	DEP-μ-IO	*C]$_\sigma$	*Glottal
☞ [in + hɑˈ]			*	*
[in + ɑˈ]	*!		*	
[i + hɑˈ]	*!			*
[i + ɑˈ]	*!			
[hɑ]	*!			*
[in]	*!		*	
[jɑ]	*!			

[55] See footnote 16.

“آنها” [ɑn-hɑˈ] ‘they- there (formal register)’

MP*/ɑn + hɑ/	MAX-μ-IO	DEP-μ-IO	*C]$_\sigma$	*Glottal
☞ [ɑn + hɑˈ]			*	*
[ɑn + ɑˈ]	*!		*	
[ɑ + hɑˈ]	*!			*
[ɑ + ɑˈ]	*!			
[hɑ]	*!			*
[ɑn]	*!		*	
[wɑ]	*!			

VI.H.4.b.ii Dari Plural Demonstrative Pronouns- Slightly Less Formal Register

In this case it is worth noting that the Dari h-drop is a wholesale ellipsis, where the n-drop is phonologically conditioned and apparently frequency driven, therefore, the language will more readily allow h-drop than n-drop. This being the case, we will expect to see that [in-ɑˈ] is a possible output of /in-hɑˈ/, but *[i-hɑˈ] is not, and that is exactly the case. In these tableaux, we see that the constraint against glottal segments is raised to the highest position:

“اینها” [in-ɑˈ] ‘they- here (formal register, slightly less)’

MP*/in + hɑ/	*Glottal	MAX-μ-IO	DEP-μ-IO	*C]$_\sigma$
[in + hɑˈ]	*!			*
☞ [i-nɑˈ]		*		
[i + hɑˈ]	*!	*		
[i + ɑˈ]		**!		
[hɑ]	*!	*		
[in]		**!		*
[jɑ]		**!		

Recall from VI.H.4.a (and previously VI.G.3) that PITCH HARMONY (NASAL) will cause /ɑ/ to raise to [u] before nasals in Iranian, and this constraint will play a role here in Dari and Tajik as well.

"آنها" [u-nɑˈ] 'they- there (formal register, slightly less)'

MP*/ɑn + hɑ/	*Glottal	PH (Nasal)	MAX-μ-IO	DEP-μ-IO	*C]$_{\sigma}$
[ɑn + hɑˈ]	*!	*			*
[ɑ-nɑˈ]		*!	*		*
[un-hɑˈ]	*!				
☞ [u-nɑˈ]			*		
[ɑ + hɑˈ]	*!		*		
[ɑ + ɑˈ]			**!		
[hɑ]	*!		*		
[ɑn]		*!	**		*
[wɑ]			**!		

VI.H.4.b.iii Dari Plural Demonstrative Pronouns- Colloquial Register

At this point it becomes necessary to suggest that an ellipsis or epenthesis when it occurs in a particular register, creates a new lexical item, in this case /i/ vice original /in/. The first bears the meaning 'this-formal,' and the second 'this-casual.' The notion is not insanely far fetched, as the majority of the world's languages have separate lexical items for pronouns distinct for register. The Persian personal pronouns "تو" /tō/ or /tū/ and "شما" /ʃʊmɑ/ are distinguished by familiarity much the same as "tu" and "vous" in French and the historical English "thou" and "you." The implication of this may be misleading, as the above mentioned personal pronoun pairs are historically distinct lexemes. A better example of colloquial usage reinterpreted with altered semantic content and morpho-

sytactic function is the English future formed with two differing lexical reinterpretations. The below illucidates:

‘I will go.’	→	‘I’ll go.’
‘I will you to go.’	→	*‘I’ll you to go.’
‘I am going to go.’	→	‘I’m gonna go.’
‘I am going to the store.’	→	*‘I’m gonna the store.’

Not only does this indicate that ‘‘ll’ and ‘gonna’ have undergone reinterpretation and no longer bear the same meaning of ‘will’ and ‘going to’, but it also bolsters two claims. First, the reduction is only allowed in the more frequent context, lending credence to frequency effects in language, and second, the even when a reduced form is reinterpreted with altered content, it retains as an option its original pronunciation at least for a time. Note that even the phrase ‘Good bye,’ has not replaced ‘God be with ye,’ in the English language, it has merely become the normative farewell derived from colloquial speech. That being said, it is fair to assume that a language may retain a separate colloquial and formal underlying form of one lexeme.

Returing to the data from above, in the colloquial examples, the input segment will be the colloquial /i + hɑˈ/, and the following constraint will apply:

PALATAL HARMONY (“Palatalization should occur between segments when the tongue position between them is palatal.” This is a clear case of ease of articulation, and it is the beginning of multiple processes that took place in historical English giving us ‘ship’ from OE “scip” and ‘child’ from OE “cild.”)

“اینها” [i-jɑˈ] ‘they- here (colloquial)’

MP*/i + hɑ/	*Glottal	PAL Harm	MAX-μ-IO	DEP-μ-IO	*C]$_{\sigma}$
[in + hɑˈ]	*!				*
☞ [i-jɑˈ]					
[i + hɑˈ]	*!		*		
[i + ɑˈ]		*!			
[hɑ]	*!		*		
[in]			*!		*
[jɑ]			*!		

“آنها” [u-wɑˈ] ‘they- there (colloquial)’

MP*/u + hɑ/	*Glottal	PAL Harm	MAX-μ-IO	DEP-μ-IO	*C]$_{\sigma}$
☞ [u-wɑˈ]					
[u + hɑˈ]	*!		*		
[u + ɑˈ]		*!	*		
[hɑ]	*!		*		
[u]			*!		*
[wɑ]			*!		

VI.H.4.b.iv Dari Plural Demonstrative Pronouns- Most Colloquial Register

For this, a simple and intuitive constraint is proposed, and one that was seen in earlier sections:

*VV (“Adjacent vowels are prohibited.” This follows from the notion of the optimal syllable structure being CV. It is bolstered by the notion that there is typically a diphthongalization or epenthetic glide insertion associated with VV forms.)

“اینها” [jɑ] ‘they- here (very colloquial)’

MP*/i + hɑ/	*Glottal	*VV	MAX-μ-IO	DEP-μ-IO	*C]$_{\sigma}$
[i + hɑˈ]	*!				
☞ [jɑˈ]			*		
[i + ɑˈ]		*!			
[i]			**!		
[hɑ]	*!				
[ɑ]			**!		

“آنها” [wɑˈ] ‘they- there (colloquial)’

MP*/u + hɑ/	*Glottal	*VV	MAX-μ-IO	DEP-μ-IO	*C]$_{\sigma}$
☞ [wɑˈ]			*		
[u + hɑˈ]	*!				
[u + ɑˈ]		*!			
[hɑ]	*!				
[u]			**!		
[ɑ]			**!		

VI.H.4.b.v Tajik/Iranian Plural Demonstrative Pronouns

As an aside, Iranian and Tajik Plural Demonstratives display some reduction, but not as much as Dari, and so the Dari reductions were analyzed first, as the Iranian and Tajik reductions may be viewed as a degree of shift within the Dari chain. The following is the available reduction in Farsi:

Mod Pers:	Formal Iranian:	Spoken Iranian:	Gloss:
“اینها”	/in-hɑˈ/ →	[i-nɑˈ]	‘they- here’
“آنها”	/ɑn-hɑˈ/ →	[u-nɑˈ]	‘they- there’

Tajik will again display dialectal differentiation between the northern and southern dialects, in the southern dialect will mirror Iranian, and the northern dialect will employ an alternate lexical item:

Tajik:	Formal Southern Tajik:	Spoken Southern Tajik:	Gloss:
“инҳо”	/in-hoˈ/ →	[i-nɑˈ]	‘they- here’
“онҳо”	/on-hoˈ/ →	[u-nɑˈ]	‘they- there’

Northern Tajik will be similar, but as mentioned in section V.E.4, has replaced “он - онҳо” /on/ - /on-hoˈ/ with “вай - вайҳо” /vaj/ - /vaj-hoˈ/:

Tajik:	Formal Northern Tajik:	Spoken Northern Tajik:	Gloss:
“инҳо”	/in-hoˈ/ →	[i-nɑˈ]	‘they- here’
“вайҳо”	/vaj-hoˈ/ →	[va-joˈ]	‘they- there’

VI.H.5 Proto-Persian Consonants: /ʤ/ > [j] (Dari)

Recall from section VI.H.2.e that Dari has the process C[+Glottal] → Glide available to it. There are two lexical items in Dari and one in Tajik (the second listed below) in which the voiced palatal affricate /ʤ/ optionally reduces to [j] in casual speech:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“پنج”	/panʤ/	[pajn]	‘five’
“مجلس”	/maʤ-lɛsˈ/	[maj-lɛsˈ]	‘meeting’

This second example (Dari and Tajik) can be accounted for much the same as C[+Glottal] → Glide in VI.H.2.e, by constraints already provided and an additional constraint against voiced affricates based on ease of articulation and justified by cross-linguistic markedness:

*C[+Fricative +Voice] ("Voiced fricatives are prohibited.")

MP*/maʤ-lɛsˈ/	*C[+Aff +Vc]	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO	ID-IO (closed)
[maʤ-lɛsˈ]	*!				
[ma-lɛsˈ]		*!			
☞ [maj-lɛsˈ]					*
[maz-lɛsˈ]			*!		

The first instance shows some syllable structure-driven metathesis that can be accounted for with the same tableau as above and the addition of the following phonotactic constraint:

V $C_1C_2C_3$ ("Coda clusters must increase in sonority."[56])

MP*/panʤ/	*C[+Aff +Vc]	V $C_1C_2C_3$	MAX-μ-IO	ID-IO(pal)	ID-IO (closed)
[panʤ]	*!				
[panj]		*!			
☞ [pajn]					*
[pan]			*!	*	

VI.H.6 Proto-Persian Consonants: /f/ and /b/ > [w] (Dari)

Dari phonology allows an optional alternation that converts certain instances of the phoneme /b/ and some pre-velar instances of /f/ to [w]:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
"دیشب"	/diˈ-ʃab/	[diˈ-ʃaw]	'last night'

[56] This constraint was known well before OT existed as a theory. Sievers (1881) and Jesperson (1904) are the earliest reference to the Sonority Sequencing Principle.

“سیب”	/seb/	[sew]	‘apple’
“خوب”	/xob/	[xow][57]	‘good’
“آب”	/ɑb/	[ɑw] → [aw]	‘water’
“خواب	/xɑb/	[xɑw] → [xaw]	‘sleep; dream’
“سبز”	/sabz/	[sawz]	‘green’
“شربا”	/ʃʊr-bɑˈ/	[ʃʊr-wɑˈ]	‘soup’
“باغبان”	/bɑɣ-bɑnˈ/	[bɑɣ-wɑnˈ]	‘gardener’
“افغانستان”	/af-ɣɑ-nɛ-stɑnˈ/	[aw-ɣɑ-nɛ-stɑnˈ]	‘Afghanistan’
“افگار”	/af-gɑr/	[aw-gɑr]	‘hurt; wounded’

The account for the above lexical items will be divided, as this process in Dari is not normative, and as far as this study is concerned, the above are the only lexical items exhibiting said reduction to [w].

VI.H.6.a Proto-Persian Consonants: /b/ > [w] / __]$_{\sigma}$# (Dari)

[57] The reduction of “خوب” /xob/ to [xow] will not occur in the normative adjectival usage of the lexical item, but rather in the interjective use, such as ‘okay,’ or ‘alright then.’ Thus, “خوب، خوب، استی؟” /xob#xob#asti/, (literally ‘well, then alright, are you?’) means, ‘Are you, then, well?,’ and is pronounced /xob#xow#asti/, and crutially all of the following are prohibited: */xob#xob#asti/, */xow#xow#asti/, and */xow#xob#asti/. As mentioned with “این” /in/ in VI.H.4.b.iii, “خوب” /xob/ has taken on two pronunciations with divergent semantic and morpho-syntactic qualities.

The first process, /b/ > [w] / __#, is simple in nature if we assume that CV[w] is in essence consonant + diphthong, and hence an open syllable, or rather that a glide is not in fact a consonant:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“دیشب”	/diˈ-ʃab/	[diˈ-ʃaw]	‘last night’
“سیب”	/seb/	[sew]	‘apple’
“خوب”	/xob/	[xow]	‘good’

But:

“فریب”	/freb/	*[frew]	‘deception’

OT will handle this brilliantly and V+diphthong will be shown to cause no ellipsis and still satisfy the constraint *C]$_\sigma$:

MP*/seb/	*C]$_\sigma$	MAX-μ-IO	ID-IO (Lab)	DEP-μ-IO
[seb]	*!			
☞ [sew]				
[se]		*!		
[sej]			*!	
[sea]			*!	

There is of course no account for why this occurs in some lexical items and not others, but that, as mentioned, will be dealt with in chapter VII. In this example, as with all that follow, I will include lexical items in which the reduction does not occur to highlight the fact that the shift is not wholesale.

VI.H.6.b Proto-Persian Consonants: /-ɑb/ > [-aw] / __# (Dari)

The process /-ɑb/ > [-aw] / __# is much the same as that covered in section VI.H.6.a, but the vowel shortens. It is convenient that this was a process we witnessed in

Dari diphthongal phonology covered in section VI.F.1 in which we noted that Dari converts /ɑw/ to [aw]:

	Formal Dari:	Spoken Dari:	Gloss:
“آب”	/ɑb/	[ɑw] → [aw]	‘water’
“خواب	/xɑb/	[xɑw] → [xaw]	‘sleep; dream’

But:

“باب”	/bɑb/	*[bɑw]; *[baw]	‘issue’

/ɑb/ → [aw] will occur due to a conspiracy between *C]$_\sigma$ and the constraint MAXIMAL DISTINCTION (“If a diphthong exists, its component features must be maximally distinct.”), as the following illustrates:

MP*/ɑb/	*C]$_\sigma$	MAX-μ-IO	MAXDIST	ID-IO (fnt)	ID-IO (Lab)	DEP-μ-IO
[ɑb]	*!					
☞ [aw]						
[ɑw]			*!			
[æw]				*!		
[ɑj]			*!		*	
[ō]			*!		*	

VI.H.6.c Proto-Persian Consonants: /b/ > [w] / __C]$_\sigma$ (Dari)

In VI.H.6.a, it was shown that /b/ > [w] occurred due to constraints against closed syllables, and in section VI.H.6.b, /ɑw/ > [aw] occurred due to constraints on Dari

diphthong structure. In the present section, /b/ > [w] / __C]$_\sigma$ will occur due to phonotactic constraints:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
"سبز"	/sabz/	[sawz]	'green'
"قبظ"	/qabz/	[qawz]	'constipated'

Recall the constraint V $C_1C_2C_3$ ("Coda clusters must increase in sonority.") from section VI.H.5:

MP*/sabz/	V $C_1C_2C_3$	MAX-μ-IO	ID-IO(lab)	ID-IO (voice)	ID-IO (closed)
[sabz]	*!				
[sab]		*!			
☞ [sawz]					*
[saz]		*!			
[sarz]			*!		*
[sabt]				*!	*

VI.H.6.d Proto-Persian Consonants: /b/ > [w] / C_σ -[__ (Dari)

/b/ > [w] / C_σ -[__, in the case of /ʃʊr-baˈ/, is distinct from that of /baɣ-banˈ/, in that /baɣ-banˈ/ will be subject to the same constraints affecting /f/ > [w] / __]$_\sigma$ C[+Velar] in section VI.H.6.e below, and hence it will be covered there with the examples provided there. That is, /baɣ-banˈ/ will become [baɣ-wanˈ] due to the proximity of /b/ to the velar segment [ɣ].

	Formal Dari:	Spoken Dari:	Gloss:
“شربا”	/ʃʊr-bɑˈ/	[ʃʊr-wɑˈ]	‘soup’
“باغبان”	/bɑɣ-bɑnˈ/	[bɑɣ-wɑnˈ]	‘gardener’

But:

“قربان”	/qʊr-bɑnˈ/	*[qʊr-wɑnˈ]	‘martyr’
“تربیه”	/tar-bjaˈ/	*[tar-wjaˈ]	‘training’

Rather than propose an ad hoc solution to this (such as *rb), let it be noted that ‘soup’ and ‘farmer’ are not prestigious lexical items such as ‘vichiçoise’ or ‘Attorney General,’ and hence there is likely a register effect here. Otherwise, this particular sound change is left as an anomaly and the notion of low register items will be returned to in section VII.

VI.H.6.e Proto-Persian Consonants: /f/ > [w] / __ $]_{\sigma}$ C[+Velar] (Dari)

The cases of /f/ > [w] / __ $]_{\sigma}$ C[+Velar] are more straightforward, as [w], a labio-velar allows /f/ to retain labiality and pick up the velarity of the following segment through anticipatory coarticulation:

HARMONY CC (velar) (“Adjacent consonants should match for the feature velar”)

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“افغانستان”	/af-ɣɑ-nɛ-stɑnˈ/	[aw-ɣɑ-nɛ-stɑnˈ]	‘Afghanistan’
“افگار”	/af-gɑr/	[aw-gɑr]	‘hurt; wounded’

But:

“افکار” /af-kɑr/ *[aw-kɑr] ‘thoughts’

MP*/af-gɑrˈ/	H-CC(Velar)	MAX-μ-IO	ID-IO(lab)	ID-IO (voice)	ID-IO (fric)
[af-gɑrˈ]	*!				
[av-gɑrˈ]	*!			*	
☞ [aw-gɑrˈ]				*	*
[af-kɑrˈ]	*!				
[a-gɑrˈ]		*!			
[a-fɑrˈ]		*!			

Note that the same tableau will produce the following shift:

“باغبان” /bɑɣ-bɑnˈ/ [bɑɣ-wɑnˈ] ‘gardener’

MP*/bɑɣ-bɑnˈ/	H-CC(Velar)	MAX-μ-IO	DEP-μ-IO	ID-IO(lab)	ID-IO (voice)	ID-IO (fric)
[bɑɣ-bɑnˈ]	*!					
[bɑ-bɑnˈ]		*!				
☞ [bɑɣ-wɑnˈ]						
[bɑɣ-gɑnˈ]				*!		
[bɑ-ɣɑnˈ]		*!				
[bɑ-ɣa-bɑnˈ]			*!			

VI.H.7 Proto-Persian Consonants: Palatalization (Iranian, Tajik)

Palatalization is observed in both Iranian and Tajik, but crucially not in Dari. Tajik and Iranian differ entirely though, in the phonological conditioning on the process in that the former is borrowed from Russian, and the latter arises from within its own phonology. Additionally, it will be shown that Tajik palatalization arises due to conditions on the vowel that follows the segment that undergoes palatalization, and Iranian palatalization arises due to a condition on both the segment that undergoes palatalization and the vowel that follows it.

VI.H.7.a Proto-Persian Consonants: Iranian Palatalization

In Iranian, the following defines the phonological parameters for palatalization:

C[+Velar] → C[+Velar]^{j} / __V[+front –high]

This seems conceptually straightforward because in terms of point of articulation, the palate exists directly between the velum and the point of articulation of a front vowel. In constraint form, we will express this using the constraint PALATAL HARMONY ("Palatalization should occur between segments when the tongue position between them is palatal.") from section VI.H.4.b.iii:

MP*/gærm/	PAL HARM	MAX-μ-IO	ID-IO(back)	ID-IO (voice)	ID-IO (high)
[gærm]	*!				
☞ [g^{j}ærm]					

Unfortunately, this is inadequate and false as it stands, because the above tableau would cause the following to take place as well:

MP*/ki/	PAL HARM	MAX-μ-IO	ID-IO(back)	ID-IO (voice)	ID-IO (high)
[ki]	*!				
☞ *[k^{j}i]					

For this reason, the following simple constraint will be added to the tableau based on ease of parsing:

*^{j}i ("Palatals are prohibited before high front vowels.")

MP*/ki/	*^{j}i	PAL HARM	MAX-μ-IO	ID-IO(back)	ID-IO (voice)	ID-IO (high)
☞ [ki]		*				
[k^{j}i]	*!					

That final constraint will be significant in Tajik palatalization as well.

VI.H.7.b Proto-Persian Consonants: Tajik Palatalization

Tajik palatalization is a feature clearly borrowed from Russian, and hence occurs fairly systematically at the onset of the mid front vowel /e/ and those instances of the high front vowel /i/ that were derived from short /i/ ([ɛ] in Dari and Iranian). This will require a note on the timing of the change and potentially the effect of orthography on language development. It bears noting from the outset that Tajik borrowing from Russian must be viewed as a strictly twentieth century phenomenon, as the Soviet occupation took place during that time. That being said, it appears non-trivial that the imposition of the Cyrillic script on the Tajik language may have been a driving force in the phonological changes heretofore addressed. The following illuminates:

MP:	Perso-Arabic:	Cyrillic:	Example:			
[i]	ی	‘и’	“سین”	“син”	[sin]	‘s’
[ɪ]/[ɛ]	∅	‘и’	“سن”	“син”	[sɛn]	‘age’

In the above, note that Cyrillic has no graphemic representation of the long and short distinction in the high front vowel. Further note that the Cyrillic grapheme “е” is phonetically realized, as evidenced in [je] as in “Елцин” ‘Yeltsin’ and “перестройка” [pje-rje-stroj^{l}-ka]. Considering that the above two graphological elements match two phonological processes in Tajik, it may be assumed that both changes followed the imposition of the Cyrillic alphabet in 1940. Thus, I assume that front vowel palatalization preceded the /ɪ/ → [i] shift and that that shift took place as indicated in

section VI.D.2, but was initiated by orthographic convention. Because the constraint *^{j}i blocked the palatalization in /i/, the process then occurred in the vowels /e/ and /ɪ/. Subsequently, the shift /ɪ/ → [i] occurred later and Tajik was left with [^{j}i] as an allophonic variation of /i/ that only occurs in [i] from the original short vowel and crucially not from /ī/. This note on high vowel palatalization must bear the all too familiar caveat of "sometimes." There is no apparent phonological conditioning on the optional palatalization, save that it only occurs in historical short high front vowels. As mentioned, the palatalization before the mid front vowel is essentially without exception, but the palatalization phenomenon tends to occur more widely among older, more educated Tajiks (who are invariably proficient and literate in Russian). Thus the phenomenon is more in use in larger cities than the countryside. The following constraint will account for the first phase of Tajik palatalization:

FRONT VOWEL PALATALIZATION ("Front vowels take onset palatalization." This is based on ease of articulation and backed up by cross-linguistic attestation.)

MP*/ser/	ID-IO(high)	*^{j}i	FV PAL	MAX-μ-IO	ID-IO (ATR)
☞ [s^{j}er]					
[ser]			*!		
[s^{j}ir]	*!	*			
[sir]	*!				

In the above tableau, the free variation between [e] and [ʲe] is accounted for, and the tableau that follows will demonstrate how the same constraint ranking will not cause /i/ → [ʲi] to occur.

MP*/sir/	ID-IO(high)	*ʲi	FV PAL	MAX-μ-IO	ID-IO (ATR)
[sʲir]		*!			
☞ [sir]			*		
[sʲer]	*!				
[ser]	*!				

Finally, and following the assumption that the Tajik /i/ → /ī/ occurred after palatalization, the allophonic existence of palatalization in some high front vowels, can be accounted for with the same tableau:

MP*/sɪr/	ID-IO(high)	*ʲi	FV PAL	MAX-μ-IO	ID-IO (ATR)
☞ [sʲɪr]					
[sɪr]			*!		
[sʲir]		*!			
[sir]					*!

After this point in history, there were three Tajik front vowels, and two of them had onset palatalized free variants:

/e/ ([ʲe])
/ɪ/ ([ʲɪ])
/i/

According to the tableaux in section VI.D, the following changes occurred:

/e/ ([ʲe])	→	[e] ([ʲe])	→	/e/ ([ʲe])
/ɪ/ ([ʲɪ])	→	[i] ([ʲi])	→	/i/ ([ʲi])
/i/	→	[i]	→	/i/

VI.H.8 Proto-Persian Consonants: /q/ > [ɢ] (Iranian)

The voiceless uvular stop /q/ and voiced velar fricative /ɣ/ are tolerated and their distinct status is maintained by the phonologies of Dari and Tajik, but not Iranian, which has merged the segments as [ɢ] with allophonic variants [ɣ] and [χ]. It bears mention here that the orthographies of the three dialects all maintain a distinction between /q/ and /ɣ/, although they have converged in Iranian:

Mod Pers/Tajik:		Iranian:	Dari:	Tajik:	Gloss:
“غیر”	“ғайр”	[ɢejr]	[ɣajr]	[ɣajr]	‘un-’
“قید”	“ҚАЙД”	[ɢejd]	[qajd]	[qajd]	‘limit’
“باغ”	“боғ”	[bɑɣ]	[bɑɣ]	[boɣ]	‘garden’
“حق”	“хаҚ”	[hæɣ]	[(h)aq]	[haq]	‘right; law’

One other note here is that /q/ (“ق”) only exists in lexical items borrowed from Arabic, while /ɣ/ (“غ”) exists in both Arabic and original Persian items.

The resultant phoneme has several allophones that vary but exhibit predictable phonological conditioning:

[ɢ] /	#__	example: [ɢejr] (“غیر” ‘un-’)
	__ C[+Voice]	example: [bɑɢ-bunˈ] (“باغبان” ‘gardener’)
[ɣ] /	V__V	example: [fæ-ɣætˈ] (“فقط” ‘only’)
	__#	example: [hu-ɣuɣ] (“حقوق” ‘rights; laws’)

[χ] / __ C[-Voice] example: [væχt] ("وقت" 'time')

In the two sections that follow, two options will be advanced as possible theoretical frameworks within which the shift took place. VI.H.8.a, the reanalysis approach, will advance the notion that Iranian /q/ took on the allophonic variations listed above, and because the majority of resultant /q/ was realized as [ɣ], /q/ and /ɣ/ were reanalyzed as one segment. VI.H.8.a will advance the theory that both Iranian /q/ and /ɣ/ shifted to [ɢ] due to differing phonological pressures, and then the resulting phoneme /ɢ/ is the segment that undergoes the allophonic variation above. Crucially, they will both be derivable within OT.

VI.H.8.a /q/ > [ɢ]: The Reanalysis Approach (Iranian)

This analysis begins under the crucial assumption that there were two ordered shifts. First, the various allophonic variations of the segment /q/ were realized due to the markedness of the voiceless uvular stop, and then, because most segments underlyingly /q/ were realized as [ɣ], /ɣ/ and /q/ were reanalyzed as one segment bearing the allophonic properties of /q/. The reason this assumption is crucial to this analysis is that it is clear from the data and from native elicitation that the graphemes "ق" and "غ" represent the same sound to an Iranian speaker, so much so that even the most educated speaker may have to pause to consider which of the two is orthographically represented in a lexical item. Thus, regardless of the theoretical background, there is no justification

for the claim that a voiced velar fricative and a voiceless uvular stop will coincidentially produce the same allophonic variation.

First are the allophonic variations that arose within the Iranian dialect in the phoneme /q/:

/q/ → [ɢ] / #__; __ C[+Voice]
/q/ → [ɣ] / V__V; __ #
/q/ → [χ] / __ C[-Voice]

Ironically, because the above environments encompass the gambit of potential environments, [q] is never phonetically realized in modern Iranian. Its historical spelling betrays the original sound which is preserved in Dari, Tajik and many dialects of Arabic. Its three allophones are derived from the following four markedness constraints that bring about the above phonological processes:

*[q] ("Voiceless uvular stops are prohibited." This is based on ease of articulation and parsing and justified by cross-linguistic markedness of the segment. As this is concerned, note that many dialects of Modern Arabic elide this segment as well.)

*V-[+Uvular]-V ("Intervocalic Uvular segments are prohibited." This is based on ease of articulation.)

*[+Uvular]# ("Word-final Uvular segments are prohibited." This is based on ease of articulation.)

*V-[-Voice]-V ("Intervocalic voiceless segments are prohibited." This is based on ease of articulation, and very widely attested.)

The first alternation, /q/ → [ɢ] / #__, is simply a case of voicing, but for the sake of consistency, a tableau will have to be created that will derive all of the above alternations. Thus, the highest ranked constraints will have no bearing on this evaluation process. The lexical item below, /qɑf/, is the name of the letter "ق" or /q/ in Dari and Tajik:

MP*/qɑf/	*V[Uvul]V	*[Uvul]#	*V[-voice]V	*q	ID-IO(Dors)	ID-IO(Uvul)	ID-IO (voice)
☞ [ɢɑf]							*
[qɑf]				*!			
[ɣɑf]						*!	*
[gɑf]						*!	*
[ʔɑf]					*!	*	

The above tableau will also account nicely for the same alternation when it occurs before a voiced consonant (/q/ → [ɢ] / __ C[+Voice]), such as in the lexical item "مقدار" /mɛq-dɑrˈ/, meaning 'amount:'

MP*/mɛq-dɑrˈ/	*V[Uvul]V	*[Uvul]#	*V[-voice]V	*q	ID-IO(Dors)	ID-IO(Uvul)	ID-IO (voice)
☞ [mɛɢ-dɑrˈ]							*
[mɛɣ-dɑrˈ]				*!			
[mɛk-dɑrˈ]						*!	*
[mɛ-dɑrˈ]						*!	*
[mɛʔ-dɑrˈ]					*!	*	

For the shift /q/ → [ɣ] / V__V; __ #, there is an exemplary lexical item containing the phoneme in both environments, "حقوق," /ho-quqˈ/ or 'rights; rules:'

MP*/ho-quqˈ/	*V[Uvul]V	*[Uvul]#	*V[-voice]V	*q	ID-IO(Dors)	ID-IO(uvul)	ID-IO(voice)
☞ [ho-ɣuɣˈ]						**	**
[ho-quɣˈ]	*!		*	*		*	*
[ho-ɣuqˈ]		*!		*		*	*
[ho-ɢuɢˈ]	*!	*					**
[ho-ʔuʔˈ]					*!	**	

The above is a prime example of the violability of lower ranked constraints. In the tableau, /ho-ɣuɣˈ/ clearly violates faithfulness constraints on the features of voicing and uvularity, but it still wins as the output, because higher ranking constraints rule out competing outputs.

The final allophonic variation, /q/ → [χ] / __ C[-Voice] is also accounted for by the same tableau:

MP*/væqt/[58]	*V[Uvul]V	*[Uvul]#	*V[-voice]V	*q	ID-IO(Dors)	ID-IO(uvul)	ID-IO(voice)
☞ [vækt]							
[væqt]				*!			
[vækt]						*!	
[væɢt]							*!
[vægt]						*!	*
[væʔt]					*!	*	

At this point in time, another constraint, unique to reanalysis came into play and caused the phonemes to converge:

REANALYSIS α →β ("When most allophones of phoneme α are identical to phoneme β, reanalyze α as β")

[58] Note that this form, /væqt/ is from the Proto-Persian /ʋaqt/, but for the purposes of this section, we assume the changes between /ʋaqt/ and /væqt/ to have taken place through other constraints particular to Iranian. Those constraints are illustrated in section VI.F.

MP*/bɑ-ɣi/	REANAL-αβ	ID-IO (Voice)	ID-IO (uvular)
☞ /bɑ-ɣi/			
/bɑ-qi/	*!		
/bɑ-xi/	*!	*	
/bɑ-ɢi/	*!		*

MP*/bɑ-qi/	REANAL-αβ	ID-IO (Voice)	ID-IO (uvular)
☞ /bɑ-ɣi/		*	*
/bɑ-qi/	*!		
/bɑ-xi/	*!	*	
/bɑ-ɢi/	*!		*

The above example was chosen because that is an example of ambiguity in Iranian, as "باقی" and "باغی" are both [bɑ-ɣiˈ], but the former, meaning 'remainder' is originally /bɑ-qiˈ/, and the latter meaning 'pertaining to gardens' is underlyingly /bɑ-ɣiˈ/. When the reanalysis occurred, the sounds converged on the allophone set represented by /q/.

VI.H.8.b /q/ > [ɢ]: The Convergence Approach (Iranian)

The second possibility is that both /ɣ/ and /q/ converged as [ɢ], and that the allophonic variation followed as constraints on the newly formed phoneme [ɢ]. Intuitively, /ɣ/ → [ɢ] occurred as some feature of /ɣ/ became intolerable to the grammar of Iranian. A constraint against voiced fricatives or even against fricatives themselves or

voice obstruents themselves is common[59] and hence it can be assumed that *Fricative and *Velar [+voice] initiated the change. But to provide a bit more precision, the following constraint will encapsulate the above intuition without militating against all fricatives:

*Dorsal Fricative (“Dorsal fricatives are prohibited.” This is based on ease of articulation and backed by relative cross-linguistic markedness.)

*Velar [+voice] (“Voiced Velars are prohibited.” This too is based on ease of articulation and backed by relative cross-linguistic markedness.)

MP*/ɣ/	*Dors Fric	ID-IO (voice)	*Vel[+vce]	ID-IO(Dors)	ID-IO(Uvul)	MAX-μ-IO
[q]		*!			*	
☞ [ɢ]					*	
[ɣ]	*!		*			
[χ]	*!					
[g]			*!			
[ʔ]		*!		*	*	

/ɣ/ → [ɢ] occurred due to similar constraints that were called upon in the previous section:

MP*/qɑf/	*q	ID-IO(Dors)	ID-IO(Uvul)	ID-IO (voice)
☞ [ɢɑf]				*
[qɑf]	*!			
[ɣɑf]			*!	*
[gɑf]			*!	*
[ʔɑf]		*!	*	

Once /ɣ/ and /q/ converged as [ɢ], the following changes occurred:

/ɢ/ → [ɣ] / V__V; __ #

[59] Recall in the nature of constraints that *fricative or *Obs [+voice] does not connote that fricatives and voiced obstruents are an abomination and a rarity, but rather that they are *more* marked then say voiceless stops or vowels.

/ɢ/ → [χ] / __ C[-Voice]
/ɢ/ → [ɢ] / #__; __ C[+Voice]

Using the same examples as the previous section, /ɢ/ → [ɣ] / V__V; __ #, will be seen to occur due to the same constraints that brought about /q/ → [ɣ] / V__V; __ #, only there is no voicing mismatch:

MP*/ho-ɢuɢˈ/	*V[Uvul]V	*[Uvul]#	*V[-voice]V	*q	ID-IO(Dors)	ID-IO(uvul)	ID-IO(voice)
☞ [ho-ɣuɣˈ]						**	**
[ho-quɣˈ]	*!		*	*		*	*
[ho-ɣuqˈ]		*!		*		*	*
[ho-ɢuɢˈ]	*!	*					**
[ho-ʔuʔˈ]					*!	**	

/ɢ/ → [χ] / __ C[-Voice] is accounted for by the same tableau as the previous section as well:

MP*/væɢt/	*V[Uvul]V	*[Uvul]#	*V[-voice]V	*q	ID-IO(Dors)	ID-IO(uvul)	ID-IO(voice)
☞ [væχt]							
[væqt]				*!			
[væɣt]						*!	
[væɢt]							*!
[vægt]						*!	*
[væʔt]					*!	*	

Also, because the word-initial and pre-voiced consonant positions are not addressed by the above constraints, it is fair to say that in these environments, /ɢ/ will not be affected by the above constraint ranking.

VI.H.9 Proto-Persian Consonants: Dark /l/ (/ɫ/) (Tajik)

Another optional allophonic variation that occurs only in Tajik is the wholesale velarization of the lateral /l/. This is clearly borrowed from Russian, whose instantiations of /l/ are invariably "dark" or in IPA, [ɫ]. In Tajik, this velarization occurs almost without exception in borrowed Russian lexical items:

"полковник"	[pał-kovˈ-nʲik]	'colonel'
"стол"	[stoł]	'table'

It also occurs often, but not predictably in original Arabic and Persian tokens as an optional alternation:

"мулло"	"ملا"	[mʊ-łoˈ] or [mʊ-loˈ]
"баланд"	"بلند"	[ba-łandˈ] or [ba-landˈ]

Because this is not phonologically conditioned, it will not be addressed as a phonological process, but rather as an available borrowing. The question of how borrowing takes place within an Optimality Theoretic framewaork would be an interesting consideration for future study. In this particular case, a borrowing constraing might be formed as such:

BORROW C_α ("A substrate varietal when juxtaposed with a superstrate varietal with quality α of consonant C will borrow quality α.")

This will be left for potential future study, but it would only apply to wholesale shifts, and so this analysis will still differ with the proposal of borrowed [ʉ], as [ʉ] is predictably distributed in Tajik and [ɨ] is not.

VI.J Proto-Persian Phono-Morphological Interplay (Dari, Iranian, Tajik)

As in most any language, phonological change due to morphological inflection is greater than lexeme-internal phonology, but subject to the same set of constraints that drive the change. In the previous section, VI.H, it was shown that various phonological reductions take place at the syllable level that further distinguish the three dialects. Section VI.J is devoted to the word and phrase-level reductions and shifts that even further differentiate the dialects.

Section VI.J.1 will examine reduction in the verbal and suffixal morphologies of the three dialects, and it will be concluded that unlike strict phonological reduction, none of the three dialects exhibits more reduction than the others; they are simply different. Section VI.J.2 will treat the phonology of the enclitic /rɑ/, or definite direct object marker, and note not only its reduction in the three dialects, but also vowel change and phonological interaction with other reductions. Finally, section VI.J.3 will deal with the various reductions associated with the third person singular conjugation of the copula verb, /ast/, and again the conclusion will be drawn that all dialects exhibit a great deal of reduction in morphological change. In point of fact, the data will show that while Dari has exhibited by far the most phonological reduction at the syllabic level, Iranian will exhibit the most reduction at the lexico-morphological level. Finally, section VI.J.4 will deal with the phonology of the ‘ezafe’ enclitic, already mentioned in this dissertation, and notes will be added on the possessive enclitic and negative copula.

VI.J.1 Verbal Morphology and Possessive Pronouns (Dari, Iranian, Tajik)

This section by no means intends to provide a complete account of Persian verbal inflection, but rather of the anomalous phonological processes associated thereto. Before the analysis however, a rudimentary explanation of common Persian verb conjugation should be provided for reference purposes, and it should also be referred back to in future discussions of verbal morphology and its associated processes. As a sidenote, the following represents a simplification of common Persian conjugation, and neglects the grammatical peculiarities of spoken Tajik. The grammar of spoken Tajik is so unique among the dialects that it merits separate treatment. The verbal morphology laid out below will be considered common Persian and formal Tajik, as distinct from spoken Tajik, which will be discussed separately from the present course of discussion, because these morphological processes are presented as a framework for phonological discussion, and Tajik grammar is not relevant to this strictly phonological analysis.[60] The same applies to Iranian and Dari peculiarities that will be appended to each conjugation in the interest of being thorough, but they will not be part of the phonological discussion, as the forms differ lexically, not phonologically.

Common Persian verbal morphology is simple and essentially Indo-European. Modern Persian has three persons (first, second and third), two numbers (singular and plural), three tenses (past, present and future), two aspects (perfect and imperfect), and two moods (subjunctive and indicative). It has no dual number and no gender (even in nouns and adjectives). The following is a fully conjugated regular verb for reference, and

[60] The forms laid out in this section will not exhibit nearly as much reduction in Tajik as the other two dialects, which falls out naturally from the fact that they represent formal register variations for Tajik, and hence will not allow widescale reduction.

the standard transcription system used shall be that which has been established as Proto-Persian (essentially formal Dari), with footnotes to indicate dialectal variations. In each, the negative proclitic "نه" /naˈ/ will be added in italics following the table of examples to demonstrate its normative position in the negative conjugation and potential effect on stress:

a. Infinitive

Infinitive stem	+	Inifinive Ending: "ان" /anˈ/			
"خرید" /xa-ridˈ/	+	"ان" /anˈ/	= "خریدن"	/xa-ri-danˈ/[61]	'to buy'

"نخریدن" /na□-xa-ri-dan/ 'to not buy'

b. Present tense (indicative):

imperf prefix	+	present stem	+	personal ending			
"می" /meˈ/	+	"خر" /xar/	+	"ام"	/-am/	= "می خرم"	/meˈ-xa-ram/
				"ای"	/-i/	= "می خری"	/meˈ-xa-ri/
				"د"	/-ad/	= "می خرد"	/meˈ-xa-rad/
				"یم"	/-em/[62]	= "می خریم"	/meˈ-xa-rem/
				"اید"	/-ed/	= "می خرید"	/meˈ-xa-red/
				"ند"	/-and/	= "می خرند"	/meˈ-xa-rand/

"نمی خرم" /naˈ-me-xa-ram/ 'I do not buy'

[61] This is /xæ-ri-dænˈ/ in Iranian, but the /a/ ←→ /æ/ distinction between Dari/Tajik and Iranian is henceforth considered to be trivial and will not be mentioned again. Every instance of PP /a/ in this section is thus to be considered /æ/ in Iranian.

[62] This is /-im/ in Iranian, and henceforth all instances of PP /e/ in this section shall be considered /i/ in Iranian.

As mentioned, there are other verb forms employing auxiliaries, and hence they are lexical in nature and not pertinent to this phonological discussion. In the case of the present tense, Dari and Tajik will not differentiate the present progressive from the non-progressive, where Iranian will. Spoken Iranian will employ the verb “داشتن” /dɑʃ-tæn/ ‘to have’ in the colloquial formation of the present progressive: “می نویسد” /mi-næ-vi-sæd/ ‘he writes’ vs. “دارد می نویسد” /dɑ-ræd#mi-næ-vi-sæd/ ‘he is writing’ (literally ‘he has he is writing.’). Again, for the sake of morpho-phonology, the two forms “دارد” /dɑ-ræd/ ‘he has’ and “می نویسد” /mi-næ-vi-sæd/ ‘he writes’ are the only items relevant to the current discussion.

c. Present tense (subjunctive – identical to indicative save prefix):

subjnct prefix	+	present stem	+	personal ending			
“به” /bɛˈ/	+	“خر” /xar/	+	“ام”	/-am/	= “بخرم”	/bɛˈ-xa-ram/
				“ای”	/-i/	= “بخری”	/bɛˈ-xa-ri/
				“د”	/-ad/	= “بخرد”	/bɛˈ-xa-rad/
				“یم”	/-em/	= “بخریم”	/bɛˈ-xa-rem/
				“اید”	/-ed/	= “بخرید”	/bɛˈ-xa-red/
				“ند”	/-and/	= “بخرند”	/bɛˈ-xa-rand/

“نخرم” /naˈ-xa-ram/ ‘I do not buy’ (subjunctive)

Tajik will not employ the prefix “به” /bɛˈ/ in the subjunctive, and so the distinction between present indicative and present subjunctive forms will be indicated by the present or absence of the indicative imperfective prefix “ме” /mɛˈ/: “мехарам” /mʲeˈ-xa-ram/ ‘I

buy’ vs. “харам” /xaˈ-ram/ ‘that I buy.’ Interestingly, the prefix is retained on verbs that begin in /o/ from PP /ā/ such as “омадан” /o-ma-danˈ/ ‘to come,’ such that the subjunctive conjugation is “биёям” /bjoˈ-jam/ ‘that I come,’ but the prefix is hypercorrectively appended to the root even in indicative conjugations: “мебиёям” /meˈ-bjo-jam/ ‘I come.’ This stands in contrast to the Dari and Iranian present indicative “ می آیم” /meˈ-jɑ-jam/ (Dari) or /miˈ-jɑm/ (Farsi) ‘I come.’

d. Past tense (indicative - preterite):

no prefix	+	infinitive stem +	personal ending			
∅	+	“خرید” /xa-ridˈ/ +	“ام”	/-am/	= “خریدم”	/xa-riˈ-dam/
			“ای”	/-i/	= “خریدی”	/xa-riˈ-di/
			∅		= “خرید”	/xa-ridˈ/
			“یم”	/-em/	= “خریدیم”	/xa-riˈ-dem/
			“اید”	/-ed/	= “خریدید”	/xa-riˈ-ded/
			“ند”	/-and/	= “خریدند”	/xa-riˈ-dand/

“نخریدم” /naˈ-xa-ri-dam/ ‘I did not buy’

e. Past tense (imperfective):

imperf prefix	+	infinitive stem +	personal ending			
“می” /meˈ/	+	“خرید” /xa-rid/ +	“ام”	/-am/	= “می خریدم”	/meˈ-xa-ri-dam/
			“ای”	/-i/	= “می خریدی”	/meˈ-xa-ri-di/
			∅		= “می خرید”	/meˈ-xa-rid/
			“یم”	/-em/	= “می خریدیم”	/meˈ-xa-ri-dem/
			“اید”	/-ed/	= “می خریدید”	/meˈ-xa-ri-ded/
			“ند”	/-and/	= “می خریدند”	/meˈ-xa-ri-dand/

“نمی خریدم” */na ˈ-me-xa-ri-dam/* *‘I was not buying’*

f. Present prefect tense (Past Participle = infinitive stem + /-aˈ/):

No prefix	+	Past Participle	+	Copula Verb ‘to be’ conjugated (present)			
∅	+	“خریده” /xa-ri-daˈ/	+	“ام”	/am/	= “خریده ام”	/xa-ri-da + amˈ/
				“ای”	/i/	= “خریده ای”	/xa-ri-da + iˈ/
				“است”	/ast/	= “خریده است”	/xa-ri-da + astˈ/
				“ایم”	/em/	= “خریده ایم”	/xa-ri-da + emˈ/
				“اید”	/ed/	= “خریده اید”	/xa-ri-da + edˈ/
				“اند”	/and/	= “خریده اند”	/xa-ri-da + andˈ/

“نخریده ام” */na ˈ-xa-ri-da#amˈ/* *‘I have not bought’*

Tajik, as mentioned above, has developed its own grammar with regards to the colloquial formation of tenses, and hence it will not be necessary to analyze the colloquial forms of the above verbal conjugations, as they will only be attested in formal settings, and no reduction will take place. For the sake of thoroughness, the following will constitute the Tajik present perfect in all persons and numbers:

No prefix	+	Past Participle	+	Nominal Suffix /-iˈ/ (with intervocalic /g/)
∅	+	“харида” /xa-ri-daˈ/ + /g/ + /-iˈ/		= “харидаги” /xa-ri-da-giˈ/

Optionally, Tajik will allow the normative past tense in combination with the lexical item “кайҳо” /kaj-hoˈ/ ‘ago’ to denote present perfect tense: “кайҳо харидам” /kaj-ho#xa-riˈ-dam/ ‘I have bought.’ In any case, the common Persian morphology is considered formal in Tajik and the above two forms are used colloquially.

g. Past prefect tense (Past Participle = infinitive stem + /-aˈ/):

Past Participle	#	Copula Verb 'to be' conjugated (past)			
"خریده" /xa-ri-daˈ/	#	"بودم"	/bu-damˈ/	= "خریده بودم"	/xa-ri-daˈ#bu-damˈ/
		"بودی"	/bu-diˈ/	= "خریده بودی"	/xa-ri-daˈ#bu-diˈ/
		"بود"	/bud/	= "خریده بود"	/xa-ri-daˈ#budˈ/
		"بودیم"	/bu-demˈ/	= "خریده بودیم"	/xa-ri-daˈ#bu-demˈ/
		"بودید"	/bu-dedˈ/	= "خریده بودید"	/xa-ri-daˈ#bu-dedˈ/
		"بودند"	/bu-dandˈ/	= "خریده بودند"	/xa-ri-daˈ#bu-dandˈ/

"خریده نبودم" /xa-ri-daˈ# naˈ-bu-dam/ 'I was not buying'

h. Future tense

Future is a compound formed with the inifitive stem and a conjugation of the verb "خواستن" /xɑ-stanˈ/ 'to want:'

"خواه" /xɑh-/[63] +	pers ending	# Infinitive stem (of future verb)

"خواه" /xɑh-/ +	"ام"	/-am/	# "خرید" /xa-ridˈ/ = "خواهم خرید" /xɑ-hamˈ# xa-ridˈ/
	"ای"	/-i/	# "خرید" /xa-ridˈ/ = "خواهی خرید" /xɑ-hiˈ# xa-ridˈ/
	"د"	/-ad/	# "خرید" /xa-ridˈ/ = "خواهد خرید" /xɑ-hadˈ# xa-ridˈ/
	"یم"	/-em/	# "خرید" /xa-ridˈ/ = "خواهیم خرید" /xɑ-hemˈ# xa-ridˈ/
	"اید"	/-ed/	# "خرید" /xa-ridˈ/ = "خواهید خرید" /xɑ-hedˈ# xa-ridˈ/
	"ند"	/-and/	# "خرید" /xa-ridˈ/ = "خواهند خرید" /xɑ-handˈ# xa-ridˈ/

"نخواهم خرید" /naˈ-xɑ-ham#xa-ridˈ/ 'I will not buy'

Colloquial Dari reverses the conjugated portion of the above in the following fashion:

[63] Note that this is one of only three present tense conjugations that omits the imperfective prefix "می" /meˈ/. The other two are the verb "هستن" /ha-stanˈ/ 'to be,' and the verb "داشتن" /dɑʃ-tanˈ/ 'to have.' Interestingly, English also disallows the imperfective present tense of both verbs in all but a few very specific contexts.

“خواهد” /xɑ-had^{l}/[64] # Infinitive stem + pers ending

“خواهد” /xɑ-had^{l}/	# “خرید” /xa-rid^{l}/	+ “ام” /-am/	= “خواهد خریدم”	/xɑ-had^{l}# xa-ri^{l}-dam/
	# “خرید” /xa-rid^{l}/	+“ای” /-i/	= “خواهد خریدی”	/xɑ-had^{l}# xa-ri^{l}-di/
	# “خرید” /xa-rid^{l}/	+∅	= “خواهد خرید”	/xɑ-had^{l}# xa-rid^{l}/
	# “خرید” /xa-rid^{l}/	+ “یم” /-em/	= “خواهد خریدیم”	/xɑ-had^{l}# xa-ri^{l}-dem/
	# “خرید” /xa-rid^{l}/	+ “اید” /-ed/	= “خواهد خریدید”	/xɑ-had^{l}# xa-ri^{l}ded/
	# “خرید” /xa-rid^{l}/	+ “ند” /-and/	= “خواهد خریدند”	/xɑ-had^{l}# xa-ri^{l}-dand/

“نخواهد خریدم” /na^{l}-xɑ-had#xa-rid^{l}-am/ ‘I will not buy’

Tajik again will only use the auxiliary form in formal language and will prefer one of two nominalized forms (as with the present perfect) for the future tense. The first is simply “харидан□” /xa-ri-da-ni^{l}/ for all forms, and the second is as follows:

imperf +	Past Participle	+ /-gi^{l}/ +	‘to be’	
“ме” +	“харида” /xa-ri-da^{l}/ +	/gi^{l}/ +	“ам” /am/	“мехаридагиам” /me-xa-ri-da-gi-am^{l}/
			“и” /i/	“мехаридагии” /me-xa-ri-da-gi^{l}/
			“аст” /ast/	“мехаридагиаст” /me-xa-ri-da-gi-ast^{l}/
			“ем” /em/	“мехаридагием” /me-xa-ri-da-gi-em^{l}/
			“ед” /ed/	“мехаридагиед” /me-xa-ri-da-gi-ed^{l}/
			“анд” /and/	“мехаридагианд” /me-xa-ri-da-gi-and^{l}/

So to recap, Persian has the following verbal morphological inflections:

[64] This also omits the imperfective prefix “می” /me^{l}/, and again only in the auxiliary usage of this verb.

Infinitival verb stem: “خرید” /xa-rid^{l}/ (Specific to the verb ‘to buy’)

Present verb stem: “خر” /xar/ (Specific to the verb ‘to buy’)

Present Participle: “خریده” /xa-ri-da^{l}/ (Specific to the verb ‘to buy,’ infinitival+/-a^{l}/)

Infinitival suffix: “ان” /-an^{l}/

Negative proclitic: “نه” /na^{l}-/

Imperfect prefix: “می” /me^{l}-/

Subjunctive prefix: “به” /bɛ^{l}-/

Personal endings	“ام”	/-am/	= 1st person singular
	“ای”	/-i/	= 2nd person singular
	“د”	/-ad/	= 3rd person singular
	“ایم”	/-em/	= 1st person plural
	“اید”	/-ed/	= 2nd person plural
	“اند”	/-and/	= 3rd person plural
Copula verb:	“ام”	/am/	= 1st person singular
	“ای”	/i/	= 2nd person singular
	“است”	/ast/	= 3rd person singular
	“ایم”	/em/	= 1st person plural
	“اید”	/ed/	= 2nd person plural
	“اند”	/and/	= 3rd person plural

Future auxiliary: “خواه” /xɑh-/ + personal endings

It is important to note that the relationship between present stem “خر” /xar/ and past stem “خرید” /xa-rid^{l}/ in the above example is exemplary of a regular verb

conjugation. The regular verb stems only account for about half of the 240 or so verbs in the language. For the other forms please refer to Appendix I (chapter IX) of this dissertation.

In the present section, VI.J.1.a will deal with a phenomenon common to all dialects of verb ending-final stop deletion, and VI.J.1.b will deal with the one exception to that, which is not deletion but a change from oral to nasal stop. Crucially, both will demonstrate that *C[+stop]# is clearly a constraint pertinent to verbal morphology but not to the grammar as a whole. Section VI.J.1.c will cover the phonology of perfect tense enclitics including a note on possessive enclitics that act similarly, and VI.J.1.d will deal with a number of verb stem-internal reductions. Section VI.J.1.e will then treat reductions associated with verbal inflectional affixes, and finally, VI.J.1.f will offer a brief commentary on an apparent stress hierarchy within the enclitics and affixes of Persian verbal morphology. Another issue that pertains to the above verbal permutations, but is actually pertinent on a higher level is the reduction of the copula verb /ast/ (to be covered in section VI.J.3).

VI.J.1.a Verbal Morphology: /d#/ → ∅ (Dari, Iranian, Tajik)

The normative third person present tense endings as covered above are /-ad/ for the singular and /-and/ for the plural, but the morpheme-final oral stops are invariably elided in both in the colloquial speech of all three dialects. The third person singular ending /-and/ is also found in past conjugations, and because the future tense utilizes a present-conjugated auxiliary, both the singular and plural third person conjugations are

found in future tense conjugations as well. In the present tense, Tajik and Dari will simply elide the segment leaving a word-final open syllable:

/meˈ	+	kʊ	+	nad/	→	[meˈ-kʊ-na]
Imp.		‘do’		3p. sing.		‘he/she does’
/meˈ	+	kʊ	+	nand/	→	[meˈ-kʊ-nan]
Imp.		‘do’		3p. pl.		‘they do’

As noted in section VI.G.2, Iranian will not tolerate the low vowel /-a/ in word-final posisiton in its grammar, and so in the case of the abovementioned shift, Iranian will not only exhibit this ellipsis, but also a shift in /-a/ (or in the specific case of the Iranian dialect, /-æ/) to /-ɛ/:

/miˈ	+	ko	+	næd/	→	*[miˈ-ko-næ]	→	[miˈ-ko-nɛ]
Prog.		‘do’		3p. sing.		‘he/she does’		
/miˈ	+	ko	+	nænd/	→	[miˈ-ko-næn]		
Prog.		‘do’		3p. pl.		‘they do’		

Out of the six personal endings in Persian, all three that end in an oral stop exhibit ellipsis of the same, and those that end in a nasal stop remain intact, thus it is clear that Persian verbal morphology does not tolerate a morpheme final oral stop, but does tolerate nasal stops in final position. Thus, the following constraint will account for the above deletions:

*+V_nC_nC[oral stop]# (“Morpheme final oral stops are prohibited.” This falls within the constraint *$C]_\sigma$ already established.):

MP*/-ad/	*+C[oral]#	ID-IO(nas)	MAX-μ-IO
☞ [-a]			
[-ad]	*!		*
[-an]		*!	

MP*/-and/	*+C[oral]#	ID-IO(nas)	MAX-μ-IO
☞ [-an]			
[-and]	*!		*
[-ad]	*!	*	
[-a]		*!	

The problem with this tableau is that it will create the undesirable effect of causing the 2nd person plural ending /-ed/ to change to *[-e]:

MP*/-ed/	*+C[oral]#	ID-IO(nas)	MAX-μ-IO
[-en]		*!	
[-ed]	*!		*
☞ *[-e]			*

This will be examined in the next section.

VI.J.1.b Verbal Morphology: /d#/ → /n#/ (Dari, Iranian, Tajik)

The normative reduction in the second person plural ending in colloquial speech is /-ed/ → /-en/, or less commonly /-id/ → /-in/ in Iranian. This shift will occur in Southern Tajik, while in the northern variety, the morpheme will remain /-ed/. It seems logical that in the absence of further evidence, that nasal is an analogous shift based on the 3rd person plural reduced ending /-an/. There are even dialects of Dari in which both

the 2nd and 3rd person plural endings are both rendered /-en/, which can be confusing, but nonetheless aptly demonstrates the availability of analogy in verbal morphemes.

Other examples of analogous change are the Romance shift to the word “quinque” for ‘five’ and the Russian shift to “девять” [dje^{l}-vjat^{j}] for ‘nine.’ In those two cases, the lexical items do not resemble their Indo-European counterparts, as ‘five’ should begin in /p/ as the Greek “πέντε” [pen^{l}-te] or the Persian “پنج” [panʤ], as should the Russian for ‘nine’ begin in /n/ as in the Latin “novem” or the Persian “نه” [noh]. In Latin, the /p/-word had been traded out for one more resembling “quattuor” (four), and in the case of Russian, its original /n/-word had been traded out for one more closely resembling “десять” [dje^{l}-cjat^{j}] or ‘ten.’ These examples were chosen because both analogous changes chose the lexical item “next to” them in the abstract linguistic sense. In the case of the Persian second person plural ending, a new analogy constraint must be proposed that encapsulates that notion:

ANALOGY (Plural verb stem) (“When a segment is prohibited, morph to a grammatically adjacent segment that does not cause ambiguity.”)

Albeit somewhat ad hoc, it stands to reason that any explanation of analogous sound change must be by nature ad hoc, as in these cases the language simply opts to copy an adjacent segment. The underlying meaning of this constraint is that whatever the offending segment (in this case /-d/), it may be replaced with another plural verbal ending that does not cause ambiguity:

MP*/-ed/	*+C[oral]#	ANL(PVS)	ID-IO(nas)	MAX-μ-IO
☞ [-en]			*	
[-ed]	*!	*		*
[-e]		*!		*
[-em]		*!		

VI.J.1.c Present Perfect Enclitic Phonology (Dari, Iranian, and Tajik)

Recall that the present prefect tense in Persian is formed with a past participle (infinitive stem + /-aˈ/) and short form of the verb 'to be' and that the negative is a prefixal stressed negative clitic /na-ˈ/:

"خریده" /xa-ri-daˈ/ +	"ام"	/am/	= "خریده ام"	/xa-ri-da#amˈ/
	"ای"	/i/	= "خریده ای"	/xa-ri-da#iˈ/
	"است"	/ast/	= "خریده است"	/xa-ri-da#astˈ/
	"ایم"	/em/	= "خریده ایم"	/xa-ri-da#emˈ/
	"اید"	/ed/	= "خریده اید"	/xa-ri-da#edˈ/
	"اند"	/and/	= "خریده اند"	/xa-ri-da#andˈ/

"نخریده ام"	/naˈ-xa-ri-da#amˈ/
"نخریده ای"	/naˈ-xa-ri-da#iˈ/
"نخریده است"	/naˈ-xa-ri-da#astˈ/
"نخریده ایم"	/naˈ-xa-ri-da#emˈ/
"نخریده اید"	/naˈ-xa-ri-da#edˈ/
"نخریده اند"	/naˈ-xa-ri-da#andˈ/

VI.J.1.c.i Present Perfect Enclitic Phonology (Dari)

Dari will realize a number of shifts into colloquial speech that will render these forms distinct from the preterite past tense forms by stress alone in all but the 3rd person singular – in bold below. The negative will cause no effect to this, save that there will be

the addition of the proclitic bearing primary stress. The below stressed syllables will retain stress, but it will be secondary:

Past:	Present Perfect:	Past (neg):	Pres Perfect (neg):
[xa-ri^{l}-dam]	[xa-ri-dem^{l}]	[na^{l}-xa-ri$_{l}$-dam]	[na^{l}-xa-ri-dem$_{l}$]
[xa-ri^{l}-di]	[xa-ri-di^{l}]	[na^{l}-xa-ri$_{l}$-di]	[na^{l}-xa-ri-di$_{l}$]
[xa-rid^{l}]	**[xa-ri-da^{l}]**	**[na^{l}-xa-rid$_{l}$]**	**[na^{l}-xa-ri-da$_{l}$]**
[xa-ri^{l}-dem]	[xa-ri-dem^{l}]	[na^{l}-xa-ri$_{l}$-dem]	[na^{l}-xa-ri-dem$_{l}$]
[xa-ri^{l}-den]	[xa-ri-den^{l}]	[na^{l}-xa-ri$_{l}$-den]	[na^{l}-xa-ri-den$_{l}$]
[xa-ri^{l}-dan]	[xa-ri-dan^{l}]	[na^{l}-xa-ri$_{l}$-dan]	[na^{l}-xa-ri-dan$_{l}$]

Because these reductions are varied in nature, they will have to be treated with a number of constraints, but as ususal, one constraint ranking will be able to account for all of the shifts:

/xa-ri-da + am^{l}/	→	[xa-ri-dem^{l}]
/xa-ri-da + i^{l}/	→	[xa-ri-di^{l}]
/xa-ri-da + ast^{l}/	→	[xa-ri-da^{l}]
/xa-ri-da + em^{l}/	→	[xa-ri-dem^{l}]
/xa-ri-da + ed^{l}/	→	[xa-ri-den^{l}]
/xa-ri-da + and^{l}/	→	[xa-ri-dan^{l}]

Two of the above are unique and require special treatment, the 1st person singular, as it contains a vowel shift, and the 3rd person singular, as it involves a complete deletion. The remainder of this section will cover first the other four forms (2nd person singular and all plural forms), and then the 3rd and 1st person singular will be dealt with individually, the latter of which being treated fully in the section VI.J.4.

Having already derived the ending [-en^{l}] from /-ed^{l}/ in section VI.J.1.b and [-an^{l}] from /-and^{l}/ in section VI.J.1.a, those ellipses will be taken for granted due to the

constraints ANALOGY (Plural verb stem) and *+V_nC_nC[oral stop]#, and the only change left to deal with below is /-a/ → ∅:

/xa-ri-da + iˈ/ → [xa-ri-diˈ]
/xa-ri-da + emˈ/ → [xa-ri-demˈ]
/xa-ri-da + edˈ/ → [xa-ri-denˈ]
/xa-ri-da + andˈ/ → [xa-ri-danˈ]

A constraint on syllable structure is straightforward, but valuable here:

*VV ("Adjacent vowels are prohibited." This is from section VI.H.4.b.iv.)

Another constraint on reduction precedence is proposed:

REDUCE (-Stress) ("If a segment or combination of segments is prohibited, reduce unstressed segments first.")

Additionally, the following constraints, also used in other portions of this paper, will be pertinent:

*+C[oral]# ("Morpheme final oral stops are prohibited." from VI.J.1.a)

*Glottal ("Glottal segments are prohibited." from VI.H.2)

MP*/xa-ri-da + iˈ/	*+C[oral]#	ANAL(PVS)	*VV	Reduce(-Str)	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-diˈ]						*
[xa-ri-da-iˈ]			*!			
[xa-ri-dajˈ]					*!	
[xa-ri-daˈ]				*!		*
[xa-ri-da-jiˈ]					*!	
[xa-ridˈ]						**!

This will also hold for all of the plural endings:

MP*/xa-ri-da + em'/	*+C[oral]#	ANAL(PVS)	*VV	Reduce(-Str)	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-dem']						*
[xa-ri-da-em']			*!			
[xa-ri-dajm']					*!	
[xa-ri-dam']				*!		*
[xa-ri-da-jem']					*!	
[xa-rid']						***!

MP*/xa-ri-da#ed'/	*+C[oral]#	ANAL(PVS)	*Glottal	*VV	REDUCE(-str)	ID-IO(pal)
☞ [xa-ri-den']						
[xa-ri-ded']	*!					
[xa-ri-de']		*!				
[xa-ri-da-ed']	*!			*		
[xa-ri-da-e']		*!		*		
[xa-ri-da-jen']						*!

MP*/xa-ri-da#and'/	*+C[oral]#	ANAL(PVS)	*Glottal	*VV	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-dan']						**
[xa-ri-da-an']				*!		*
[xa-ri-dand']	*!					*
[xa-ri-da-jan']					*!	
[xa-ri-da-a]				*!		**
[xa-ri-da']						***!

Interestingly, given the facts associated with the 3rd person singular conjugation of the copula verb "است" or /ast/, the above tableau will also derive the 3rd person singular present perfect form despite the fact that it is simply a complete deletion of the copula verb. In section VI.J.3, the copula verb will be covered, and it will be established that /ast/ → [-s] / V#__. Leaving the bulk of the exposition on the copula verb for that section, it will merely be stated that the above tableau - with additions trivial to the above analysis - will render the 3rd person singular as well as the four forms already covered in this section:

MP*/xa-ri-da#astˈ/	*+C[oral]#	ANAL(PVS)	*Glottal	*VV	*C]$_{\sigma}$	DEP-μ-IO
☞ [xa-ri-daˈ]						
[xa-ri-da-hastˈ]	*!		*			
[xa-ri-dasˈ]					*!	
[xa-ri-da#hatˈ]	*!		*		*	
[xa-ri-da-haˈ]			*!			
[xa-ri-da-aˈ]				*!		
[xa-ri-da-staˈ]						*!

Conspicuously not covered in the present section is the 1st person singular present perfect form /xa-ri-da + amˈ/ → [xa-ri-demˈ], and the reason it was ignored is that it appears to be an analogous shift patterned after the vowel lowering exhibited in the possessive enclitic. Before delving into that specific case, the phonology of the ezafe will have to be explained in detail, and then the possessive pronouns, and these explanations along with an analysis of the distinctions between Iranian, Dari and Tajik will take place in section VI.J.4.

VI.J.1.c.ii Present Perfect Enclitic Phonology (Iranian)

Iranian will display some surprising shifts in this regard as well:

/xa-ri-dɛ + amˈ/ → [xæ-ri-dæːmˈ]
/xa-ri-dɛ + iˈ/ → [xæ-ri-diːˈ]
/xa-ri-dɛ + astˈ/ → [xæ-ri-dɛːˈ][65]
/xa-ri-dɛ + emˈ/ → [xæ-ri-diːmˈ]
/xa-ri-dɛ + edˈ/ → [xæ-ri-diːdˈ] or [xæ-ri-diːnˈ]
/xa-ri-dɛ + andˈ/ → [xæ-ri-dæːnˈ]

For the most part (being in all cases but the third person singular), Iranian seems to be following a pattern of quantitative compensatory lengthening. Recall from sections

[65] The process by which /ast/ → [ɛ] will be covered in section VI.J.3.

IV.E.3 and VI.H.2.g that Iranian, unlike Dari, diplays a quantitative compensatory lengthening in instances of /h/ and /ʔ/-ellipsis. The same will hold for the majority of the above present perfect conjugations. Thus, the same tableaux that derived the plurals above and the second person singular will derive all of the Iranian present perfect forms, save that the constraint *VV will not take effect:

MP*/xa-ri-dæ + æmˈ/	*+C[oral]#	ANAL(PVS)	Reduce(-Str)	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-dæːmˈ]					
[xa-ri-dæˈ]			*!		*
[xa-ri-dæ-jæmˈ]				*!	
[xa-rid-æmˈ]					*!

MP*/xa-ri-dæ + iˈ/	*+C[oral]#	ANAL(PVS)	Reduce(-Str)	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-diːˈ]					
[xa-ri-dæˈ]			*!		*
[xa-ri-dæ-jiˈ]				*!	
[xa-rid-iˈ]					*!

MP*/xa-ri-dæ + ɛˈ/	*+C[oral]#	ANAL(PVS)	Reduce(-Str)	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-dɛːˈ]					
[xa-ri-dæˈ]			*!		*
[xa-ri-dæ-jɛˈ]				*!	
[xa-rid-ɛˈ]					*!

MP*/xa-ri-dæ + imˈ/	*+C[oral]#	ANAL(PVS)	Reduce(-Str)	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-diːmˈ]					
[xa-ri-dæˈ]			*!		*
[xa-ri-dæ-jimˈ]				*!	
[xa-rid-imˈ]					*!

MP*/xa-ri-dæ + id^{l}/	*+C[oral]#	ANAL(PVS)	Reduce(-Str)	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-diːn^{l}]					
[xa-ri-dæ^{l}]			*!		*
[xa-ri-dæ-jin^{l}]				*!	
[xa-rid-in^{l}]					*!
[xa-ri-diːd^{l}]	*!				
[xa-ri-did^{l}]	*!				

MP*/xa-ri-dæ + ænd^{l}/	*+C[oral]#	ANAL(PVS)	Reduce(-Str)	ID-IO(pal)	MAX-μ-IO
☞ [xa-ri-dæːn^{l}]					
[xa-ri-dæ^{l}]			*!		*
[xa-ri-dæ-jæn^{l}]				*!	
[xa-rid-æn^{l}]					*!
[xa-ri-dæːnd^{l}]	*!				
[xa-ri-dænd^{l}]	*!				

The same will apply to the possessive pronouns, which will be covered in detail in section VI.J.4.a.

Tajik will not undergo reduction in these forms, and hence the underlying form will also be the surface representation, as the auxiliary form of the present perfect is only used in formal settings. This was covered in section VI.J.1.

VI.J.1.d Verb Stem Internal Reduction (Dari, Iranian, Tajik)

An ancient historical reduction as old as the days of Avestan and Old Persian remains in evidence as a diachronic and synchronic alternation in Persian, and it shall be referred to as partial verb stem ellipsis. Ali Jahanshiri (2007) cites the following as examples from Old Persian and Avestan of this shift in a diachronic context, and note that in the below examples, both incidents of CV and VC are elided:

Avestan or Old Persian:	Middle Persian:	Modern Persian	Gloss
ʃijāti	ʃāti → ʃādi	/ʃādi/	‘joy’
ʧavant	ʧant → ʧand	/ʧand/	‘how much?’
duʋar	dar	/dar/	‘door’
xʃājaθija	ʃāha → ʃāh	/ʃāh/	‘king’
nēʋak	nēk	/nēk/	‘good’

He further provides examples of synchronic variation in Modern Persian. Isolated dialectal data have been removed, and the remaining list is essentially differenciated by Formal and Spoken Iranian. Verb forms have also been removed as they will be covered extensively below:

Formal Iranian:	Spoken Iranian:	Gloss:
/xʋā-har/	/xʋār/	‘sister’
/pi-rā-han/	/pi-rān/	‘shirt’
/ba-rā-jam/	/ba-rām/	‘for me’

As a sidenote, none of the above three forms will be attested in Tajik, because Tajik has borrowed different forms for each of the above. That is, Tajik will use various Turkic lexemes for ‘sister:’ “апа” /a-paˈ/ in southern Tajik and “буча” /bu-ʧaˈ/ or “ава” /a-vaˈ/ in northern speech. The form “خواهر” or “хвоҳар” /xvo-harˈ/ is only used to denote ‘younger sister’ and does not reduce. Likewise, “برادر” or “биродар” /bi-ro-darˈ/

connotes ‘older brother,’ and the Turkic “ака” /a-kaˈ/ and “додар” /do-darˈ/ mean ‘younger brother.’ Thus, “бирода р” /bi-ro-darˈ/ does not reduce to [bjo-darˈ] as in Dari. The next example, /pi-rā-han/ or ‘shirt’ has been replaced with the Turkic “курта” /kur-taˈ/, and in the final example, spoken Tajik employs normative Persian prepositions postpositionally, a syntactic loan from Turkic, hence nullifying all three examples for Tajik.

VI.J.1.d.i Verb Stem Internal Reduction: /ha/ → ∅

In the case of “خواستن” /xɑ-stanˈ/ ‘to want’, Iranian and Tajik display greater reduction than Dari, as the latter will prefer semivowel epenthesis to ellipsis. In the case of “دادن” /dɑ-danˈ/ ‘to give,’ in Dari and Tajik the adjacent consonant will undergo devoicing that will not take place in Iranian:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“می خواهم”	/meˈ-xɑ-ham/ → *[meˈ-xɑ-am]	[meˈ-xɑ-jam]	‘I want’
“می دهم”	/meˈ-da-ham/ → [meˈ-dʰam] →	[meˈ-tam]	‘I give’

Mod Pers:	Formal Iranian:	Spoken Iranian:	Gloss:
“می خواهم”	/miˈ-xɑ-hæm/ → *[miˈ-xɑ-æm]	[miˈ-xɑm]	‘I want’
“می دهم”	/miˈ-de-hæm/ → [miˈ-dʰæm] →	[miˈ-dæm]	‘I give’

Mod Pers:	Formal Iranian:	Spoken Iranian:	Gloss:
“мехваҳам”	/meˈ-xo-ham/ → *[meˈ-xo-am]	[meˈ-xom]	‘I want’
“медеҳам”	/meˈ-de-ham/ → [meˈ-dʰam] →	[meˈ-tam]	‘I give’

Thus in Dari, Iranian and Tajik, the operative constraints will be the following three:

*Glottal	(“Glottal segments are prohibited.” This is from section VI.H.2)
*VV	(“Adjacent vowels are prohibited.” This is from section VI.J.1.c)
MAX-$_{LEX}$-IO	(“Ellipsis in lexical items is prohibited.” This differentiates the verb stem vowel from the inflectional suffix.)

In Dari, the following will be the tableaux for the verb “می خواهم” /meˈ-xɑ-ham/:

MP*/meˈ-xɑ-ham/	*Glottal	*VV	MAX-$_{LEX}$-IO	MAX-μ-IO	DEP-μ-IO
[meˈ-xɑ-ham]	*!				
[meˈ-xɑ-am]		*!		*	
[meˈ-xɑm]			*!	**	
☞ [meˈ-xɑ-jam]					*
[meˈ-xam]				**!	

In Iranian and Tajik, DEP-μ-IO and MAX-μ-IO are promoted above MAX-$_{LEX}$-IO, and the same tableaux will apply:

MP*/miˈ-xɑ-hæm/	*Glottal	*VV	DEP-μ-IO	MAX-μ-IO	MAX-$_{LEX}$-IO
[miˈ-xɑ-hæm]	*!				
[miˈ-xɑ-æm]		*!		*	
☞ [miˈ-xɑm]				**	*
[miˈ-xɑ-jæm]			*!		
[miˈ-xæm]				**!	

As a sidenote, the verb “آمدن” /ɑ-ma-danˈ/, although it has traditionally been considered to exhibit verb stem reduction /ma/ → ∅, will not be analyzed in this section:

“می آیم” */meˈ-ɑ-am/ → [meˈ-jɑ-jam] (Dari) → [miˈ-jɑm] (Iranian) ‘I come’

“آمدم” /ɑ-maˈ-dam/ → [ɑ-maˈ-dam] (Dari) → [ɑ-mæˈ-dæm] (Iranian) ‘I came’

These are in actuality suppletive forms, the former based on Old Persian verb “ay-” ‘to go,’ and the latter based on Old Persian “ā+gam-” ‘to come.’ That being said, */meˈ-ɑ-am/ → [meˈ-jɑ-jam] (Dari) → [miˈ-jɑm] (Iranian) mirrors exactly the process witnessed above: /meˈ-xɑ-ham/ → [meˈ-xɑ-jam] (Dari) → [miˈ-xɑm] (Iranian) ‘I want.’

The case of “می دهم” /me^{l}-da-ham/ is not as simple in Dari and Tajik, but in Iranian the above tableau will still account for the reduction:

MP*/me^{l}-da-ham/	*Glottal	*VV	DEP-μ-IO	MAX-μ-IO	MAX-$_{LEX}$-IO
[mi^{l}-dæ-hæm]	*!				
[mi^{l}-dæ-æm]		*!		*	*
☞ [mi^{l}-dæm]				**	**
[mi^{l}-d^{h}æm]	*!			*	**
[mi^{l}-t^{h}æm]	*!			*	**

In Dari and Tajik, however, a two step shift appears to have occurred:

/me^{l}-da-ham/ → /me^{l}-dham/ (or /me^{l}-d^{h}am/) → /me^{l}-tam/

This first shift will require the addition of a new constraint particular to verb stem reduction and witnessed in a number of languages in frequently used verbal forms such as the English ‘I’ve,’ ‘I’d’ and ‘she’s.’ The notion is that the short vowel of the stem reduces, and ironically the verb ‘to be able,’ will exhibit this in both English and Iranian (cf. /kæn/ → [kn̩] and /ta-ʊɑn/ → *[tʊɑn] → [tɑn]):

VERB STEM REDUCT (1) (“Short low vowels in frequent verbs are ellided.” This is the first formal OT mention of frequency, and it is based on the Sonority Sequencing Principle. This will be returned to in the section that follows.)

*Asp [+voice] (“Voiced Aspirates are prohibited.” This is based on ease of articulation and parsing, and attested in cross-linguistic markedness and attested language change.)

MP*/me^{l}-da-ham/	VSR	*VV	DEP-μ-IO	MAX-μ-IO	ID-IO(voice)	*Asp +Vce	*VC^{h}V
[me^{l}-da-ham]	*!						
[me^{l}-da-am]		*!		*			
[me^{l}-dam]				**!			
☞ [me^{l}-d^{h}am]				*		*	
[me^{l}-t^{h}am]				*	*!		*
[me^{l}-tam]				**!	*		

Following this initial change, a prohibition on voiced aspirates coupled with the general faithfulness constraint blocking ellipsis devoices the segment and the output is [me^{l}-t^{h}am], with variant [me^{l}-tam]:

MP*/me^{l}-d^{h}am/	*Asp +Vce	MAX-μ-IO	DEP-μ-IO
[me^{l}-d^{h}am]	*!		
[me^{l}-dam]		*!	
☞ [me^{l}-t^{h}am]			
[me^{l}-tam]		*!	

This raises a familiar question regarding why this does not cause the phenomenon to occur in other less frequent cases, and surely the answer then lies in frequency effects to be dealt with briefly in chapter VII. The frequency question is particularly pertinent in this present section on verb stems, as it is often noted in language classes that "irregular" verbs are really nothing more than the most frequent verbs, and typically it is the most frequent verbs – 'to be,' 'to have,' 'to go' and what have you - that are the most "irregular."

Finally, there is one exception regarding the Dari and Tajik conjugation of the verb 'to want' above. It was shown that Iranian showed full ellipsis of the /hæ/ portion of the conjugated verb: "می خواهم" /mi^{l}-xɑ-hæm/ → [mi^{l}-xɑm] 'I want,' while Iranian and

Tajik preferred semivowel epenthesis: “می خواهم” /meˈ-xɑ-ham/ → [meˈ-xɑ-jam] ‘I want.’

This is indeed not the case when the verb “خواستن” /xɑ-stanˈ/ is used as an auxiliary in Dari and Tajik, in which case it will reduce as in Iranian:

“خواهد کردم” /xɑ-had#kar-dam/ → [xɑt-ka-dam][66] ‘I will do.’

It is not out of the question to presume that these are two separate verbs, one an auxiliary and the other a lexical verb, much the same as the English ‘will,’ and this being the case, the constraint MAX-$_{LEX}$-IO, which caused the distinction between Iranian and Dari will be nullified in the Dari and Tajik auxiliary “خواستن” /xɑ-stanˈ/. This strengthens the justification for the constraint.

A further justification for the constraint and this entire portion of the analysis is the short list of lexical items provided in the introduction:

Formal:	Spoken:	Gloss:
/xʊā-<u>ha</u>r/	/xʊār/	‘sister’
/pi-rā-<u>ha</u>n/	/pi-rān/	‘shirt’
/ba-rā-<u>ja</u>m/	/ba-rām/	‘for me’

The first two items on the list are fully lexical, while the third is in fact /ba-rā/ ‘for’ + ezafe /ɛ/ + /am/ ‘me-direct object.’ This is pertinent because only the last case will

[66] There is more reduction here than covered in in the present section, but r-drop has been discussed, and word-final devoicing warrants no discussion, as this would be an obvious constraint.

exhibit a distinction in Dari and Iranian (as seen above), while the first two will be identical in Dari and Iranian:

*/meˈ-ɑ-am/	→	[meˈ-jɑ-jam] (Dari)	→	[miˈ-jɑm] (Iranian)
/meˈ-xɑ-ham/	→	[meˈ-xɑ-jam] (Dari)	→	[miˈ-xɑm] (Iranian) 'I want'
*/ba-rɑˈ-am/	→	[ba-rɑˈ-jam] (Dari)	→	[ba-rɑmˈ] (Iranian) 'for me'
/xʊɑ-har/	→	[xʊɑr] (Dari)	→	[xʊɑr] (Iranian) 'sister'
/xɑˈ-had/ (aux.)	→	[xɑt] (Dari)	→	[xɑt] (Iranian) 'will-fut.'

VI.J.1.d.ii Verb Stem Internal Reduction: /aʊ/ → ∅

The reduction covered in the present section is common to all three dialects and so dialect specific phenomena such as Iranian /miˈ-tɑ-nam/ → [miˈ-tu-nam] will be ignored and only the common Proto-Persian vowel stem reduction will be treated. Note the patterned change below:

Mod Pers:	PP:		Comn Pers:	Gloss:
"می شوم"	/meˈ-ʃa-ʊam/ →	*[meˈ-ʃa-am] →	[meˈ-ʃam]	'I become'
"می روم"	/meˈ-ra-ʊam/ →	*[meˈ-ra-am] →	[meˈ-ram]	'I go'
"می توانم"	/meˈ-ta-ʊɑ-nam/ →	*[meˈ-ta-ɑ-nam] →	[meˈ-tɑ-nam]	'I can'
"می آورم"	/meˈ-ɑ-ʊa-ram/ →	*[meˈ-ɑ-a-ram] →	[meˈ-jɑ-ram]	'I bring'

For the sake of comparison, a less frequent token identical in form to the first two items above will not exhibit the same reduction:

“می دوم” /meˈ-da-ʋam/ → *[meˈ-da-am] → *[meˈ-dam][67] ‘I run’

For the four examples for which the change is available, the reduction is easily derived from constraints already mentioned, save that the VERB STEM REDUCT (VSR) constraint must be revisited. Recall that VSR (1) specified the following:

VERB STEM REDUCT (1) (“Short vowels in frequent verbs are ellided.” This is from section VI.J.1.d.i.)

*[ʋ] (“The segment [ʋ] is prohibited.” This was mentioned in VI.F, and needs little justification as the segment /ʋ/ no longer exists in any dialect of Modern Persian:

MP*/meˈ-ra-ʋam/	VSR (1)	*[ʋ]	*VV	MAX-μ-IO	DEP-μ-IO
[meˈ-ra-ʋam]	*!	*			
[meˈ-r-ʋam]		*!		*	
☞ [meˈ-ra-m]				**	
[meˈ-ra-am]	*!			*	
[meˈ-am]			*!	***	

MP*/meˈ-ta-ʋɑ-nam/	VSR (1)	*[ʋ]	*VV	MAX-μ-IO	DEP-μ-IO
[meˈ-ta-ʋɑ-nam]	*!	*			
[meˈ-tʋɑ-nam]		*!		*	
☞ [meˈ-tɑ-nam]				**	
[meˈ-ta-nam]	*!			**	
[meˈ-ta-ɑ-nam]	*!		*	*	

MP*/meˈ-ɑ-ʋa-ram/	VSR (1)	*[ʋ]	*VV	MAX-μ-IO	DEP-μ-IO
[meˈ-ɑ-ʋa-ram]	*!	*			
[meˈ-ɑ-ʋram]		*!		*	
☞ [meˈ-ɑ-ram]				**	
[meˈ-a-ram]	*!			**	
[meˈ-ɑ-am]			*!	***	

[67] This token remains [meˈ-da-ʋam] in conjugated form.

Again, this will produce the undesirable result of causing /meˈ-da-ʋam/ → *[meˈ-dam] in the verb 'to run,' but for now, the above tableaux are specific to those more frequent verbs.

VI.J.1.d.iii Verb Stem Internal Reduction: /ū/ → ∅

Recall from above the constraint RVS (1) from section VI.J.1.d.i. The constraint is straightforward enough, but the number "1" was not, and at this point it will be explained. A weakness of OT must be pointed out, in that the following intuition cannot easily be expressed. This set of verb stem reductions under consideration are a clear case of reduction in order of sonority. That is, the most sonorous segments are elided first. A simple constraint "Elide in order of sonority." Will have no means of stopping and indeed, the output will invariably be ∅. Verbs may reduce, but disappearance is highly undesireable. Thus the following is proposed from a representation of the sonority scale from Burquest & Payne (1993, p. 101):

VSR (1) ("Low lax vowels in frequent verbs are ellided.")
VSR (2) ("Low tense vowels in frequent verbs are ellided.")
VSR (3) ("Mid lax vowels in frequent verbs are ellided.")
VSR (4) ("Mid tense vowels in frequent verbs are ellided.")
VSR (5) ("High lax vowels in frequent verbs are ellided.")
VSR (6) ("High tense vowels in frequent verbs are ellided.")
VSR (7) ("Semivowels in frequent verbs are ellided.")
VSR (8) ("Flaps in frequent verbs are ellided.")
VSR (9) ("Laterals in frequent verbs are ellided.")
VSR (10) ("Nasals in frequent verbs are ellided.")
VSR (11) ("Voiced fricatives in frequent verbs are ellided.")
VSR (12) ("Voiceless fricatives in frequent verbs are ellided.")
VSR (13) ("Voiced plosives in frequent verbs are ellided.")
VSR (14) ("Voiceless plosives in frequent verbs are ellided.")
VSR (15) ("Complex plosives in frequent verbs are ellided.")

Hence the “1” above represented the place on the sonority scale, and any higher number presupposes each lower number in the tableau. Hence, the constraint VSR (10) denotes “In frequent verbs, all segments ≥ nasals in sonority are ellided.” Thus the following reduction will be accounted for with VSR (6):

“می گویم” /meˈ-go-jam/ → *[meˈ-go-am] → [meˈ-gam] ‘I say’

MP*/meˈ-goj-am/	VSR (6)	*VV	MAX-μ-IO	DEP-μ-IO
[meˈ-gu-jam]	*!			
[meˈ-gom]	*!		*	
☞ [meˈ-gam]			*	
[meˈ-gm]			**!	
[meˈ-go-am]		*!		

It would appear that this constraint is unnecessary, as other simpler constraints may be operative, but in this case it may be demoted in formal speach, and the underlying form may be produced, and in fact it is:

MP*/meˈ-go-am/	*VV	MAX-μ-IO	DEP-μ-IO	VSR (6)
☞ [meˈ-go-jam]				*
[meˈ-gom]		*!		*
[meˈ-gam]		*!		
[meˈ-gm]		**!		
[meˈ-go-am]	*!			

VI.J.1.d.iv Verb Stem Internal Reduction: /nī/ → ∅

One interesting reduction that is in fact historical rather than phonological involves the Proto-Persian conjugation of the verb ‘to sit:’

“می نیشینم” /meˈ-nī-ʃī-nam/ → [meˈ-ʃī-nam] ‘I sit’

This is a relic of Old Persian in that the canonical Modern Persian verb /nī-ʃas-tanˈ/ is from Old Persian "*ni+šād" 'to establish.' Modern Persian recalls the historical prefix and preserves it as a relic in formal and canonical use only. In colloquial usage the prefix is elided. Dari stands alone as the only dialect which elides the prefix "*ni" in the past tense:

"نشست"	/ni-ʃast/	→	[ʃīʃt]	'he/she/it sat'

VI.J.1.d.v Verb Stem Internal Reduction: C# → ∅

The Persian command form is simply the subjunctive prefix added to the present tense stem. In Dari, this final consonant elides in many, but not all verbs:

"بچین"	/biˈ-ʧīn/ →	[bɛˈ-ʧi]	'cut!'
"بگیر"	/biˈ-gīr/ →	[bɛˈ-gi]	'take!'
"بخور"	/biˈ-xōr/ →	[bɛˈ-xo]	'eat!'
"بخیز"	/biˈ-xēz/ →	[bɛˈ-xe]	'arise!'
"کن"	/kun/[68] →	[ko]	'do!'

As in the other cases, there are other less frequent items in which this does not occur, such as the following:

"بخر"	/biˈ-xar/ →	*[bɛˈ-xa]	'buy!'

[68] Recall that the prefix /bi-/ is optional when forming the subjunctive of the verb "کردن" /kar-dan/ and is elided altogether in Tajik.

“بلیس” /biˈ-līs/ → *[bɛˈ-li] ‘lick!’

For the instances in which it does occur, it may be viewed as a simple case of the constraint *C]# at work, and hence needs no further explanantion:

MP*/bɛˈ-gīr/	*C]#	MAX-μ-IO	DEP-μ-IO
[bɛˈ-gīr]	*!		
[bɛˈ-gr]	*!	*	
☞ [bɛˈ-gī]		*	
[bɛg]	*!	**	

VI.J.1.d.vi Verb Stem Internal Reduction: Compound Verbs

Finally Dari and Tajik exhibit reduction in verbs that have directional enclitics “در,” “بر” and “ور”:

“می برایم” /mēˈ-bar-ā-jam/ → [meˈ-brɑ-jam] ‘I come up’

“می درایم” /mēˈ-dar-ā-jam/ → [meˈ-drɑ-jam] ‘I come out’

“می ورایم” /mēˈ-ʋar-ā-jam/ → [mewˈ-rɑ-jam] ‘I come off’

This is the same phenomenon mentioned in section III.E.13.a, and will be seen to take place due to the constraint VSR (1), noting that the enclitic has become part of the verb stem. In the cases mentioned in III.E.13.a, this needs no extra caveat, as those were already part of the verb stem:

MP*/meˈ-da-rɑ-am/	VSR (1)	*VV	MAX-μ-IO	DEP-μ-IO
[meˈ-da-rɑ-am]	*!			
[meˈ-da-rɑ-jam]	*!			
☞ [meˈ-drɑ-jam]			*	
[meˈ-drɑ-am]		*!	**	
[meˈ-gu-am]		*!		

VI.J.1.e Verbal Affix Phonology (Dari, Iranian, Tajik)

Persian vowel-initial verb stems react differently between dialects when an enclitic or prefix is added to them. Section VI.J.1.e.i will deal with the negative proclitic /na+/ on vowel-initial verbs, section VI.J.1.e.ii will cover the imperfect aspect prefix /mē+/, and VI.J.1.e.iii will deal with the subjunctive prefix /bɛ+/[69]. It should be noted from the start that Persian proclitics and prefixes have a different morpho-phonology and so will be treated with a differing set of tableaux. Furthermore, it will be established that the subjunctive prefix, unlike its imperfect counterpart (essentially the indicative) has no innate vowel, but rather its output is assigned one through harmony. This, as mentioned in the last footnote, will be detailed in section VI.J.1.e.iii.

First, recall the Persian verbal proclitics and prefixes as laid out in section VI.J.1:

/na-/	example: “نخرید”	/naˈ-xar-idɩ/	‘s/he did not buy’	Negative proclitic
/mē-/	example: “می خرید”	/mēˈ-xar-idɩ/	‘s/he was buying’	Imperfect Prefix
/be/	example: “بخر”[70]	/bɛˈ-xar/	‘buy!’	Subjunctive Prefix

As these prefixes are vowel final, they will react with vowel initial verbs, and in general, the following patterns will arise:

[69] For the sake of this analysis, this vowel will, and has been referred to as underlyingly /ɛ/, when in reality, it appears to be a blank vowel position with a harmonizing output in the grammar. This will be covered in section VI.J.1.e.iii.

[70] As mentioned above, Tajik omits this prefix, leaving only “xap” /xar/ ‘buy!’

		Iranian:	Dari:	Tajik:
Negative proclitic:				
/na/ + /a-/	→	[næː]	[na]	[naː]
/na/ + /u-/	→	[næ-jo]	[nʊ]	[naː][71]
/na/ + /ā-/	→	[næ-jɑ]	[nɑ]	[na + bjo] [72]
/na/ + /ī-/	→	[næ-ji]	[ni]	[na + bi]
Imperfect Prefix:				
/mē/ + /a-/	→	[mi]	[me]	[me] or [me + a]
/mē/ + /u-/	→	[mi-jo]	[me]	[me + a]
/mē/ + /ā-/	→	[mi-jɑ]	[mjɑ]	[me + bjo] or [mjo]
/mē/ + /ī-/	→	[miː]	[mej]	[mej]
Subjunctive Prefix:				
/be/ + /a-/	→	[bɛ]	[be]	∅
/be/ + /u-/	→	[bɛ-jo]	[be]	∅
/be/ + /ā-/	→	[bɛ-jɑ]	[bjɑ]	∅ or rarely [bjo]
/be/ + /ī-/	→	[bɛ-ji]	[bi]	∅

For the above three prefixes, four verbs have been picked that exemplify the four possible verb-initial vowels:

"انداختن"	/an-dāx-tan/	'to throw'
"افتادن"	/uf-tā-dan/	'to fall'
"آمدن"	/ā-ma-dan/	'to come'
"ایستادن"	/ī-stā-dan/	'to stand'

[71] All ʊ-initial verbs in Tajik have been reinterpreted as a-initial, and hence will exhibit the same phonology as other a-initial verbs.

[72] Recall from section VI.J.1 that Tajik does not employ the prefix "би," /бɛ/ at all for the subjunctive, but adds it almost invariably on conjugated forms of vowel initial verbs.

These verbs have been conjugated into various tenses and moods in order to demonstrate the entirely of their interaction with the prefixes:

/na/: Present subjunctive, Past indicative

/mē/: Present indicative, Past progressive

/bɛ/: Present subjunctive

VI.J.1.e.i Verbal Affix Phonology: The Negative Proclitic

The negative particle when affixed to a vowel-initial verb will be reduced (/na-/ → /n-/) in virtually all Dari forms, while in Iranian, the vowel of the negative proclitic as well as the initial vowel of the verb will be preserved either through quantitative lengthening or though an epenthetic semivowel. In Tajik, /na-/ + short vowel initial verbs will mirror Iranian and /na-/ + long vowel initial verbs will involve the epenthetic insertion of the subjunctive prefix /bɛ-/:

- /na/ + /a-/ → [næː]; [na] in "انداختن" /an-dāx-tan/ 'to throw'

"ناندازم" → /næ + æn-dɑ-zæm/ → [næːn-dɑ-zæm] Iranian 'that I not throw-subjunctive'

/na + an-dɑ-zam/ → [nan-dɑ-zam] Dari

/na + an-do-zam/ → [naːn-do-zam] Tajik

“نانداخت” → /næ + æn-dɑxt/ → [næːn-dɑxt] Iranian ‘he/she/it did not throw.’

→ /na + an-dɑxt/ → [nan-dɑxt] Dari

→ /na + an-doxt/ → [naːn-doxt] Tajik

In the Dari, the above two examples demonstrate that *VV holds as expected from other tableaux:

MP*/naˈ + an-dɑ-zam/	*VV	MAX-$_{LEX}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [nanˈ-dɑ-zam]				*	
[naˈ + jan-dɑ-zam]			*!		
[naːn-dɑ-zam]	*!				
[an-dɑ-zam]				**!	

In Iranian and Tajik, *VV would not allow for /næ/ + /æ-/ → [næː] and /na/ + /a-/ → [naː], so a more specific constraint is proposed:

*$V_\alpha V_\beta$ (“Adjacent non-matching vowels are prohibited.” This is based on ease of articulation and justified by cross-linguistic proliferation of semivowel epenthesis in cases of $V_\alpha V_\beta$.)

This speculation is desirable as Iranian and Tajik tend toward quantitative compensatory lengthening in like vowels and either semivowel or consonant epenthesis in differing vowels, and Dari tends toward deletion. The below tableau will account for both Tajik and Iranian (Iranian vowels are used):

MP*/næˈ + æn-dɑ-zæm/	*$V_\alpha V_\beta$	MAX-LEX-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [næːnˈ-dɑ-zæm]					
[næˈ + jæn-dɑ-zæm]				*!	
[næn-dɑ-zæm]			*!		
[æn-dɑ-zæm]			**!		

Note that in Tajik, the above tableau can also be applied to /na/ + /u-/ procliticization as well as /na/ + /a-/, because in Tajik when /ʊ/ occurs in verb initial position, the vowel shifts to /a/. Thus, “افتادن” /uf-tā-dan/ ‘to fall,’ will be rendered “афтидан” /af-tī-danˈ/ ‘to fall’ in Tajik. The conversion of /ā/ → /ī/ within the verb stem is not relevant to this portion of the analysis, but for the sake of thoroughness, it should be noted that “یدن” /-ī-danˈ/ is considered the regular verb ending, and hence the shift is hypercorrective to the regular form. This alternation is available to Dari as well.

This same tableau will allow for the semivowel epenthesis in the other three examples as well:

- /na/ + /u-/ → [næ-jo]; [nʊ] in “افتادن” /uf-tā-dan/ ‘to fall’

“نافتم” → /næ + of-tæm/ → [næ-jof-tam] Iranian ‘that I not fall- subjunctive”

/na + ʊf-tam/ → [naf-tam] Dari

/na + ʊf-tam/ → [naːf-tam] Tajik

“نافتادم” → /næ + of-tɑ-dæm/ → [næ-jof-tɑ-dæm] Iranian ‘I did not fall”

/na + ʊf-tɑ-dam/ → [nʊf-tɑ-dam] Dari

/na + ʊf-tɑ-dam/ → [naːf-ti-dam] Tajik

In Dari, the above tableaux will derive the correct output:

MP*/naˈ + ʊf-tam/	*VV	MAX-$_{LEX}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [nʊf-tam]				*	
[naˈ + jʊf-tam]			*!		
[naːfˈ-tam]	*!				
[nafˈ-tam]		*!			
[ʊf-tam]				**!	

And the same will hold for Iranian:

MP*/næˈ + of-tæm/	*$V_\alpha V_\beta$	MAX-$_{LEX}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [næˈ + jof-tæm]				*	
[nof-tæm]		*!			
[næf-tæm]			*!		
[næˈ + of-tæm]	*!		**		

- /na/ + /ā-/ → [næ-jɑ]; [nɑ] in “آمدن” /ā-ma-dan/ ‘to come’

“نیآیم” → /næ + ɑ-jæm/ → [næ-jɑm] Iranian ‘that I not come-subjunctive’

/na + ɑ-jam/ → [njɑ-jam] Dari

/na + o-jam/ → [na-bjo-jɑm] Tajik

“نآمد” → /næ + ɑ-mæd/ → [næ-jɑ-mæd] Iranian ‘he/she/it did not come’

→ /na + ɑ-mad/ → [nɑ-mad] Dari

→ /na + o-mad/ → [na-o-mad] Tajik

In Dari, the above tableaux will hold:

MP*/naˈ + ɑ-jam/	*VV	MAX-$_{LEX}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [njɑˈ-jam]				*	
[naˈ + jɑ-jam]			*!		
[naːˈ-jam]	*!				
[naˈ-jam]		*!			
[ɑˈ-jam]				**!	

Iranian is also accounted for:

MP*/næˈ + ɑ-jæm/	*$V_\alpha V_\beta$	MAX-$_{LEX}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [næˈ + jɑ-jæm]				*	
[nɑˈ-jæm]		*!			
[næˈ-jæm]			*!		
[næˈ + ɑ-jæm]	*!				
[ɑ-jæm]			**!		

In Tajik, no further analysis is needed with /ā/-initial verbs, as the appenthetic /bi/ mentioned earlier renders the verb no longer vowel-initial, and in the second example, the output matches the underlying form.

- /na/ + /ī-/ → [næ-ji]; [ni] in “ایستادن” /ī-stā-dan/ ‘to stand’

“نایستم”	/næ + i-stam/	→	[næ-ji-stam]	Iranian ‘that I not stand- subjunctive’
	/na + i-stam/	→	[ni-stam]	Dari
	/na + i-stam/	→	[ni-stam]	Tajik
“نایستاد”	/næ + i-stɑd/	→	[næ-ji-stɑd]	Iranian ‘he/she/it did not stand’
	/na + i-stɑd/	→	[ni-stɑd]	Dari
	/na + i-stɑd/	→	[ni-stod]	Tajik

Dari and Tajik:

MP*/na' + i-stam/	*VV	MAX-$_{LEX}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [ni-stam]				*	
[na' + ji-stam]			*!		
[na:'-stam]	*!				
[na'-stam]		*!			
[i-stam]				**!	

Iranian:

MP*/næ' + i-stæm/	*$V_\alpha V_\beta$	MAX-$_{LEX}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [næ' + ji-stæm]				*	
[ni-stæm]		*!			
[næ-stæm]			*!		
[næ' + i-stæm]	*!		**		

Thus, one constraint ranking accounted for all variations of preclitic + vowel initial verbal morpho-phonology.

VI.J.1.e.ii Verbal Affix Phonology: The Imperfective Prefix

In the case of verbal inflectional prefixes, a different set of constraints will be active, as is to be expected due to the differing nature of clitics and affixes. Recall the following:

/mē/ + /a-/	→	[mi]	[me]	[me] or [me-a]
/mē/ + /u-/	→	[mi-jo]	[me]	[me-a]
/mē/ + /ā-/	→	[mi-jɑ]	[mjɑ]	[me + bjo] or [mjo]
/mē/ + /ī-/	→	[miː]	[mej]	[mej]

The above demonstrates that in the case of /a-/ verbs, all three dialects will elide the verb-initial /a-/, save an dialectal retention in northern Tajik. In the case of /u-/, /ā-/ and /ī-/ verbs, Iranian will remain true to its phonology as seen in the negative proclitic, and with /u-/ and /ī-/ verbs, Dari will remain true to the proclitic forms. Finally, Dari will retain an epenthetic semivowel with /ā-/ verbs, not witnessed in the case of the proclitic. Southern Tajik will essentially mirror Dari, save instances particular to dialectal Tajik and consonant epenthesis.

There is one common thread that differentiates proclitic phonology from that of prefixes, and that is the inability of prefixal vowels to elide. Referring back the introductory chart in VI.J.1.e, it is clear that Dari allows reduction of the proclitic vowel, but never the prefixal vowel. Furthermore, Tajik has two dialectal options available to it in the case of the imperfective prefix, and neither of them involves ellipsis of the prefixal vowel. Thus, in the case of prefixes, it will be proposed that prefixal vowels cannot be deleted, and verb-initial long vowels will not elide, but short vowels will:

MAX-$_{+ATR}$-IO (“Ellipsis of +ATR vowels is prohibited.” This is intuitive.)

- /mē/ + /a-/ → [mi]; [me] in “انداختن” /an-dāx-tan/ ‘to throw’

“می اندازم” → /mi + æn-dɑ-zæm/ → [min-dɑ-zæm] Iranian ‘I throw’

/me + an-dɑ-zam/ → [men-dɑ-zam] Dari

/me + an-do-zam/ → [men-do-zam] Southern Tajik

/me + an-do-zam/ → [me-an-do-zam] Northern Tajik

“می انداخت” → /mi + æn-dɑxt/ → [min-dɑxt] Iranian ‘he/she/it was throwing.’

/me + an-dɑxt/ → [men-dɑxt] Dari

/me + an-doxt/ → [men-doxt] Southern Tajik

/me + an-doxt/ → [me-an-doxt] Northern Tajik

With the addition of the above constraint ID-$_{+ATR}$-IO in place of MAX-$_{LEX}$-IO, the following tableau will apply in Dari and Southern Tajik:

MP*/ me + an-dɑ-zam /	*VV	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [men-dɑ-zam]				*	
[me + jan-dɑ-zam]			*!		
[maːn-dɑ-zam]	*!	*			
[meːn-dɑ-zam]	*!	*			
[an-dɑ-zam]				**!	

Recalling that in Iranian, *VV is revised as *$V_{\alpha}V_{\beta}$, the above tableau will also render the desired result in Iranian:

MP*/ mi + æn-dɑ-zæm /	$*V_{\alpha}V_{\beta}$	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [min-dɑ-zæm]				*	
[mi + jæn-dɑ-zæm]			*!		
[mi + æn-dɑ-zæm]	*!				
[æn-dɑ-zæm]		*!		**	

In northern Tajik, the underlying form matches its surface representation, and so no analysis is necessary.

- /mē/ + /u-/ → [mi-jo]; [mʊ]; [me-a] in "افتادن" /uf-tā-dan/ 'to fall'

"می افتم" → /mi + of-tæm/ → [mi-jof-tam] Iranian 'I fall.'

/me + ʊf-tam/ → [mef-tam] Dari

/me + af-tam/ → [me-af-tam] Tajik

"می افتادم" → /mi + of-tɑ-dæm/ → [mi-jof-tɑ-dæm] Iranian 'I was falling."

/me + ʊf-tɑ-dæm/ → [mef-tɑ-dam] Dari

/me + af-ti-dæm/ → [me-af-ti-dam]Tajik

Dari:

MP*/ me + ʊf-tam /	*VV	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [mef-tam]				*	
[me + jʊf-tam]			*!		
[mʊːf-tam]	*!	*			
[meːf-tam]	*!	*			
[ʊf-tam]		*!		**	

Iranian:

MP*/ mi + of-tæm /	$*V_{\alpha}V_{\beta}$	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
[mof-tæm]		*!		*	
☞ [mi + jof-tæm]			*		
[mi + of-tæm]	*!				
[mif-tæm]		*!		**	

Again in Tajik, the output form will match the input form making analysis unnecessary.

This accounts nicely for the fact that Dari will elide the historical short high back vowel and Iranian will not. The Iranian realization of that historical segment is +ATR and in Dari and Tajik it is –ATR. The remainder of the verb-initial vowels will also be accounted for with the same tableau.

- /mē/ + /ā-/ → [mjɑ]; [mjɑ] in "آمدن" /ā-ma-dan/ 'to come'

"می آیم" → /mi + ɑ-jæm/ → [miˈ-jɑm] Iranian 'I am coming'

/me + ɑ-jam/ → [me-jɑ-jam] Dari

/me + o-jam/ → [me-bjo-jam] → [me-bjom] Tajik

"می آمد" → /mi + ɑ-mæd/ → [mi-jɑ-mæd] Iranian 'he/she/it was coming'

/me + ɑ-mad/ → [mjɑ-mad] Dari

/me + o-mad/ → [mjo-mad] Tajik

In Dari (recall that the remaining phonological alternations in this verb were covered in sectionVI.J.1.d.i. This section will deal solely with the prefixal morpho-phonology.), the following will occur:

MP*/ me + ɑ-jam /	*VV	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [me-jɑ-jam]			*		*
[mɑ-jam]		*!		*	
[me-jam]		*!		*	
[mɑː-jam]	*!	*			
[ɑ-jam]		*!		**	

Iranian (Again, refer to sectionVI.J.1.d.i for further reductions):

MP*/ mi + ɑ-jæm /	*$V_\alpha V_\beta$	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
[mɑ-jæm]		*!		*	
☞ [mi + jɑm]			*		
[mi + ɑ-jæm]	*!				
[mi-jæm]		*!		**	

And in Tajik, the addition of the epenthetic consonant will render this section irrelevant.

- /mē/ + /ī-/ → [miː]; [mej] in “ایستادن” /ī-stā-dan/ ‘to stand’

“می ایستم” → /mi + i-stam/ → [miː-stam] Iranian ‘I stand’

/me + i-stam/ → [mej-stam] Dari

/me + i-stam/ → [mej-stam] Tajik

“می ایستاد” → /mi + i-stɑd/ → [miː-stɑd] Iranian ‘he/she/it was standing’

/me + i-stɑd/ → [mej-stɑd] Dari

/me + i-stɑd/ → [mej-stod] Tajik

Dari and Tajik:

MP*/ me + i-stam /	*VV	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [mej-stam]			*		*
[mi-stam]		*!		*	
[me-stam]		*!		*	
[mi:-stam]	*!	*			
[i-stam]		*!		**	

Iranian:

MP*/ mi + i-stæm /	*$V_\alpha V_\beta$	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
[mi-stæm]		*!		*	
☞ [mi: + stæm]					
[i-stæm]		*!		**	

VI.J.1.e.iii Verbal Affix Phonology: The Subjunctive Prefix

The final prefixal verbal morph is "به" or historical /bi-/ for Dari and Iranian, and ∅ for Tajik. The vowel in the phonemic representation is deceiving as this prefix undergoes harmonic shifts in conjugation both in consonant-initial verbs and in vowel-initial verbs:

"بردن" /bur-dan/ 'to carry' → /bi/ + /ba-ram/ → /bu + ba-ram/ 'that I carry-subj.'

"گرفتن" /grif-tan/ 'to obtain' → /bi/ + /gi-ram/ → /bī + gī-ram/ 'that I obtain-subj.'

"خیستن" /xē-stan/ 'to arise' → /bi/ + /xē-zam/ → /bɛ + xē-zam/ 'that I obtain-subj.'

Thus, through harmony, this segment may be viewed as follows:

/bV/ → /bu/ / __(C_n) V[-Front] (NB /bu/ → [bʊ] Dr/Tj; [bo] Ir)

→ /bi/ / __(C_n) V[+High +Front]

→ /bɛ/ / __(C_n) V[+Mid +Front]

Iranian differs from Dari in its harmony with the low vowels /ā/ and /a/ in that in Iranian the prefix remains /bɛ/ and in Dari it becomes /bʊ/. At any rate, this will be left as a case of vowel harmony, but the true nature of the underlying form must be established, and evidence from Middle Persian points to /be/. This is odd, considering that /o/ and /e/ are not firmly established as attested Middle Persian phonemes. Whatever the case, it shall be assumed to be /be/. Fortunately for this particular section, the harmony does not affect the morpho-phonology:

- /be/ + /a-/ → [bɛ]; [be] in "انداختن" /an-dāx-tan/ 'to throw'

"باندازم" → /be + æn-dɑ-zæm/ → [bɛn-dɑ-zæm] Iranian 'that I throw-subjunctive'

→ /be + an-dɑ-zam/ → [ben-dɑ-zam] Dari

→ /∅ + an-do-zam/ → [an-do-zam] Tajik

Dari:

MP*/ be + an-dɑ-zam /	*VV	MAX-$_{+ATR}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [ben-dɑ-zam]				*	*
[ban-dɑ-zam]		*!	*		
[be-an-dɑ-zam]	*!				
[beːn-dɑ-zam]	*!	*			
[an-dɑ-zam]		*!	**		

Iranian:

MP*/ be + æn-dɑ-zæm /	$*V_{\alpha}V_{\beta}$	MAX-$_{+ATR}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [bɛn-dɑ-zæm]			*		
[bæn-dɑ-zæm]		*!	*		
[be-æn-dɑ-zæm]	*!				
[æn-dɑ-zæm]		*!	**		

- /be/ + /u-/ → [bɛ-ju]; [be] "افتادن" /uf-tā-dan/ 'to fall'

"بافتم" → /be + of-tæm/ → [bɛ + jof-tæm] Iranian 'that I fall- subjunctive"

/be + ʊf-tam/ → [bef-tam] Dari

/∅ + af-tam/ → [af-tam] Tajik

Dari:

MP*/ be + ʊf-tam /	*VV	MAX-$_{+ATR}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [bef-tam]			*		
[bʊf-tam]		*!	*		
[be + ʊf-tam]	*!				
[be:f-tam]	*!	*			
[ʊf-tam]		*!	**		

Iranian:

MP*/ be + of-tæm /	$*V_{\alpha}V_{\beta}$	MAX-$_{+ATR}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [bɛ + jof-tæm]				*	*
[bof-tæm]		*!	*		
[bɛ + of-tæm]	*!				
[bef-tæm]		*!			
[of-tæm]		*!	**		

- /bɛ/ + /ā-/ → [bjɑ]; [bjɑ] "آمدن" /ā-ma-dan/ 'to come'

“بیآیم”	/be + ɑ-jæm/	→	[bɛ-jɑm]	Iranian ‘that I come-subjunctive’
	/be + ɑ-jam/	→	[bjɑ-jam]	Dari
	/be + o-jam/	→	[bi-jom]	Tajik

Dari:

MP*/be + ɑ-jam/	*VV	MAX-$_{+ATR}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [bjɑ-jam]		*		*	*
[be-ɑ-jam]	*!				
[bɑ-jam]		*	*!		
[bɛ-jam]		*	*!		

Recall that this verb is a suppletive form, and so more reduction appears to take place here than simply the prefixal morpho-phonology, but that is outside the focus of this section. It initially appears odd that Tajik, which forms the subjunctive with ∅-prefix vice original /bɛ/, should retain the prefix in this instance. In point of fact, as mentioned, Tajik has hypercorrectively epended the prefix /bɛ/ to all conjugated forms of /ā/-initial verbs, and so */bi-bjo-jam/ would in fact be the Tajik form if the prefix were added. The following tableau accounts for Iranian, and Tajik will be said to mirror Iranian as with the formation of the imperfective present tense of the verb “خواستن” / “хостан” /xā-stanˈ/ ‘to want’ in section VI.J.1.d.i:

MP*/ be + ɑ-jæm /	*$V_\alpha V_\beta$	MAX-$_{+ATR}$-IO	ID-IO(pal)	MAX-μ-IO	DEP-μ-IO
☞ [bɛ-jɑm]				*	
[bɑ-jæm]		*!		*	
[be-ɑ-jam]	*!				
[ɑ-jam]		*!		**	

- /be/ + /ī-/ → [bɛ-ji]; [bej] / “ایستادن” /ī-stā-dan/ ‘to stand’

“بایستم” → /be + i-stæm/ → [bɛ-ji-stæm] Iranian ‘that I stand- subjunctive’

/be + i-stam/ → [bej-stam] Dari

/be + i-stam/ → [bej-stam] Tajik

Dari and Tajik:

MP*/be + i-stam/	*VV	MAX-$_{+ATR}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [bej-stam]				*	*
[bi-stam]		*!	*		
[be-i-stam]	*!				
[beː-stam]	*!	*			
[i-stam]		*!	**		

Iranian:

MP*/be + i-stæm/	*$V_\alpha V_\beta$	MAX-$_{+ATR}$-IO	MAX-μ-IO	ID-IO(pal)	DEP-μ-IO
☞ [bɛ-ji-stæm]			*		
[bi-stæm]		*!	*		
[be-i-stæm]	*!				
[i-stæm]		*!	**		

One final case of reduction particular to Dari is that of the subjunctive form of the verb “ماندن” /mɑn-dan/ ‘to remain; to put,’ which exhibits the following reduction:

“بماند” /bʊˈ-mɑ-nad/ → /bɑ-nad/

“بمان” /bʊˈ-mɑn/ → /bɑn/

This is a case similar to those mentioned in VI.J.1.d.1:

VI.J.1.d.1.i /ha/ → ∅
VI.J.1.d.1.ii /aʊ/ → ∅
VI.J.1.d.1.iii /ū/ → ∅
VI.J.1.d.1.iv /nī/ → ∅

In this case, /ɛm/ → ∅, it appears similar to "آمدن" /ɑ-ma-danˈ/ 'to come' (/ma/ → ∅), but having established that this is a suppletive form, the similarity will be considered coincidental. In this case, it appears that the following occurred:

"بماند" /bʊˈ-mɑ-nad/ → */bmɑ-nad/ → /bɑ-nad/

The grammar reduced the vowel /ʊ/ and did not permit the resultant two consecutive labial segments, and so it elided the first in order of sonority. Without further evidence, this will be left as a speculation.

VI.J.1.f Stress Heirarchy of Verbal Morphology (Dari, Iranian, Tajik)

Recall from section VI.J.1, the following are the tenses of Persian conjugated in the 1st person singular affirmative and negative, noting stress:

a. Infinitive

"خریدن"	/xa-ri-danˈ/	'to buy'
"نخریدن"	/naˈ-xa-ri-dan/	'to not buy'

b. Present tense (indicative):

"می خرم"	/meˈ-xa-ram/	'I buy'
"نمی خرم"	/naˈ-me-xa-ram/	'I do not buy'

c. Present tense (subjunctive – identical to indicative save prefix):

“بخرم”	/bɛˈ-xa-ram/	‘I buy’
“نخرم”	/naˈ-xa-ram/	‘I do not buy’ (subjunctive)

d. Past tense (indicative - preterite):

“خریدم”	/xa-riˈ-dam/	‘I bought’
“نخریدم”	/naˈ-xa-ri-dam/‘I did not buy’	

e. Past tense (imperfective):

“می خریدم”	/meˈ-xa-ri-dam/	‘I was buying’
“نمی خریدم”	/naˈ-me-xa-ri-dam/	‘I was not buying’

f. Present prefect tense (Past Participle = infinitive stem + /-aˈ/):

“خریده ام”	/xa-ri-da + amˈ/	‘I have bought’
“نخریده ام”	/naˈ-xa-ri-da#amˈ/	‘I have not bought’

g. Past prefect tense (Past Participle = infinitive stem + /-aˈ/):

“خریده بودم”	/xa-ri-daˈ#bu-damˈ/	‘I was buying’
“خریده نبودم”	/xa-ri-daˈ# naˈ-bu-dam/	‘I was not buying’

h. Future tense

“خواهم خرید”	/xɑ-hamˈ# xa-ridˈ/	‘I will buy’
“نخواهم خرید”	/naˈ-xɑ-ham#xa-ridˈ/	‘I will not buy’

Colloquial Dari:

“خواهم خرید”	/xɑ-hadˈ# xa-riˈ-dam/	‘I will buy’
“نخواهد خریدم”	/naˈ-xɑ-had#xa-ridˈ-am/	‘I will not buy’

Immediately, a stress hierarchy emerges:

1. /na^{ˈ}/ (negative prefix) always bears primary stress.

2. In the absence of /na^{ˈ}/ the other prefix, either /me^{ˈ}/ or /bɛ^{ˈ}/ bears primary stress.

3. In the absence of a prefix, the verb stem or present participle will bear ultimate stress.

4. Auxiliary verbs will bear their own ultimate stress, secondary to the main verb.

VI.J.2 Proto-Persian Definite Article "را" (/rɑ/) (Dari, Iranian, Tajik)

Recall from previous sections that the Persian lexicon contains an enclitic /rɑ/ that appends postpositionally to direct objects and direct object phrases and indicates definiteness:

"من آن مرد را دیدم." /man ɑn mard-rɑ di^{ˈ}-dam/ 'I saw that man.'
'I [that man]-rɑ saw-1p-sing'

"من آن مرد بلند را دیدم." /man ɑn mard-e b^{ɛ}-land-rɑ di^{ˈ}-dam/ 'I saw that tall man.'
'I [[that man]-e tall]-rɑ saw-1p-sing'

"من یک مرد دیدم." /man jak mard di^{ˈ}-dam/ 'I saw a man.'
'I one man saw-1p-sing'

This enclitic is treated differently by the dialects in its phonology, and hence will be covered in the three sections that follow.

VI.J.2.a Definite Article "را" (/rɑ/) (Dari)

The vowel of the enclitic /rɑ/ is reduced to [a] in colloquial Dari speech, and the consonant elides when suffixed to a word-final consonant:

Mod Pers:	Formal Dari:	Spoken Dari:	Gloss:
“تو را دید.”	/tu + rɑ#did/	[tuˈ + ra#did]	‘he/she saw you.’
“سگ را دید.”	/sag + rɑ#did/	[saˈ-ga#did]	‘he/she saw the dog.’

These phenomena will be dealt with as the result of constraints on complex onsets across morphological boundaries and word-final unstressed long vowels:

*C+C… (“Consonant clusters across morpheme boundaries are prohibited.” This is due to ease of articulation and the ability of language to avoid deriving marked forms in the grammar.)

*V[+Long -Stress]# (“Word-final unstressed long vowels are prohibited.” This is a typical reduction tendency.)

MAX-$_{LEX}$-IO (“Ellipsis in lexical items is prohibited.” This presumes that a clitic is not a lexical item in the sense that a noun would be.)

MP*/sagˈ + rɑ/	*C+C	*V[L-S]#	MAX-$_{LEX}$-IO	MAX-μ-IO
[sagˈ + rɑ]	*!	*		
[sagˈ + ra]	*!			
[saˈ-gɑ]		*!		*
[saˈ-ra]			*!	*
☞ [saˈ-ga]				*

MP*/tuˈ + rɑ/	*C+C	*V[L-S]#	MAX-$_{LEX}$-IO	MAX-μ-IO
[tuˈ + rɑ]		*!		
[tuˈ + a]				*!
[trɑ]	*!		*	
☞ [tuˈ-ra]				

VI.J.2.b Definite Article “را” (/rɑ/) (Iranian)

In Iranian, the same r-ellipsis phenomenon will take place, but in the case of the vowel reduction, rather than the Dari /ɑ/ → [a], we will see /ɑ/ → [u], similar to the alternation addressed in IV.C.5, VI.G.3 and VI.H.4.a:

Mod Pers:	Formal Iranian:	Spoken Iranian:	Gloss:
“تو را دید.”	/to + rɑ#did/	[toˈ + ru#did]	‘he/she saw you.’
“سگ را دید.”	/sæg + rɑ#did/	[sæˈ-gu#did]	‘he/she saw the dog.’

Similarly to Dari, the grammar of Iranian will contain a constraint on word-final unstressed long vowels, but rather than prohibiting them, Iranian will force their height to be maximized for ease of articulation:

MAX(High)V[+Long -Stress]# (“Maximize the height of word-final unstressed long vowels.”)

MP*/sægˈ + rɑ/	*C+C	*MAX (hg)[L-S]#	ID-IO(back)	MAX-$_{LEX}$-IO	MAX-μ-IO
[sægˈ + rɑ]	*!	*			
[sægˈ + ru]	*!				
[sæˈ-gɑ]		*!			*
[sæˈ-ru]				*!	*
☞ [sæˈ-gu]					*
[sæˈ-gi]			*!		

MP*/toˈ + rɑ/	*C+C	*MAX (hg)[L-S]#	ID-IO(Back)	MAX-$_{LEX}$-IO	MAX-μ-IO
[toˈ + rɑ]		*!			
[toˈ + u]					*!
[trɑ]	*!	*		*	
☞ [toˈ-ru]					
[toˈ + ri]			*!		

VI.J.2.c Proto-Persian Consonants: Definite Article "را" (/ro/) (Tajik)

Like Dari, the vowel of the Tajik enclitic /ro/ is reduced to [a] in colloquial speech, and the consonant /r/ elides following another consonant:

Mod. Pers.:	Formal Tajik:	Spoken Tajik:	Gloss:
"تو را دید."	/tu#ro#did/	[tuˈ-ra#did]	'he/she saw you.'
"سگ را دید."	/sag#ro#did/	[saˈ-ga#did]	'he/she saw the dog.'

Having established my hypothesis that /ā/ → /o/ in Tajik is a late change, it stands to reason that the above tableaux for Dari took their effect on colloquial Tajik before the input vowel of the enclitic had become /o/.

VI.J.2.d Proto-Persian Consonants: n-drop and "را" (/rɑ/) Interaction

Thus far, it has been shown that Iranian, Dari and Tajik share the same conditions for r-drop in the enclitic /rɑ/. Immediately, the question arises of how the three dialects may differ in their treatment of said r-drop when it occurs in conjunction with a preceding word-final deletion. That is, if two deletions coincide, will one or the other happen more readily? In the case of word-final n-drop, Dari will retain the consonant of the enclitic, and still elide the final /n/ of the lexical item:

"آن را دیدم." /ɑn rɑ di-dam/ →[uˈ-ra diˈ-dam] 'I saw that.'
'That-rɑ saw-1p-sing'

“من را دید.” /man rɑ did/ →[ma-rɑ did] ‘He/she saw me.’
‘I rɑ saw-3p-sing’

This stands in strict opposition to the notion that lexical items retain their integrity with more strength than non-lexical items like enclitics. The only way of dealing with this will be to ascert that each register has its own constraint ranking with its own set of outputs at the lexical level. Once the outputs are generated they must enter a set of phrase level constraint rankings. Thus, /ɑn/ and /man/ above have two register dependent inputs: /ɑn/ and /man/ or /u/ and /ma/, so that /ɑn-rɑ/ and /u-rɑ/ or /man-rɑ/ and /ma-rɑ/ are separate lexical inputs.

Iranian does not allow the n-drop and so will exhibit the following reduction: Iranian:

“آن را دیدم.” /ɑn rɑ di-dæm/ →[uˈ-nu di-dæm] ‘I saw that.’
‘That-rɑ saw-1p-sing’

“من را دید.” /mæn rɑ did/ →[mæ-nu did] ‘He/she saw me.’
‘I rɑ saw-3p-sing’

Interesting, if not frustrating, is the fact that Tajik, which allows n-drop in both cases, will elide the enclitic /r/ in the first example and the pronominal /n/ in the second:

“манро дид.” /man ro did/ →[ma-na did] ‘He/she saw me.’
‘I rɑ saw-3p-sing’

“Онро дидам.” /on ro di-dam/ →[uˈ-ra di-dam] ‘I saw that.’
‘That-rɑ saw-1p-sing’

Again, the reason for noting this is thoroughness, but the tableaux that accounted for these shifts in sections VI.J.2.a and VI.J.2.b will apply to each example as it stands. Tajik remains anomalous in this as in many cases. In point of fact several alternations have been analyzed thus far in which portions of the Tajik grammar followed a pattern akin to Dari and other portions mirrored Iranian in nature, and unfortunately, the only theoretical escape for the researcher is to claim that this is due to borrowing.

VI.J.3 Proto-Persian Consonants: 3rd Per Sing Copula "است" (/ast/) (Dari, Iranian, Tajik)

It is worth noting at this point that the three dialects vary in their treatment of the entire copula verb "بودن" /bu-danˈ/, not just the 3rd person singular. It was stated above that Dari has available to it an epenthetic form /aˈ-sta/ of the verb conjugation /ast/. This may prove slightly more complicated than to claim that this conjugation is an analogous form derived from the long form of the verb 'to be.' Before a phonological analysis, a synopsis of the Persian verb to be is warranted. The copula verb conjugation provided in section VI.J.1 was in point of fact only the short form of the Persian copula. There are three distinct verbs 'to be' in Modern Persian:

	"بودن" Short form: /bu-danˈ/		"بودن" Long form: /bu-danˈ/		"استن" Long Form: /a-stanˈ/	
1st person singular	"ام"	/am/	"هستم"	/haˈ-stam/	"استم"	/aˈ-stam/
2nd person singular	"ای"	/i/	"هستی"	/haˈ-sti/	"استی"	/aˈ-sti/
3rd person singular	"است"	/ast/	"هست"	/hast/	"است"	/ast/
1st person plural	"ایم"	/em/	"هستیم"	/haˈ-stem/	"استیم"	/aˈ-stem/
2nd person plural	"اید"	/ed/	"هستید"	/haˈ-sted/	"استید"	/aˈ-sted/
3rd person plural	"اند"	/and/	"هستند"	/haˈ-stand/	"استند"	/aˈ-stand/

One notable distinction between Dari and Iranian is that Dari uses "استن" /a-stanˈ/ (the third column above) almost exclusively as its canonical copula, and Iranian uses the short form of "بودن" /bu-danˈ/ (the first column above)[73]. Tajik, which reserves the option of null copula, will favor the short form of "بودن" /bu-danˈ/ when a copula is phonetically realized. The question remains then of the function of the long form of "بودن" /bu-danˈ/, and this will be seen to vary by dialect. In Iranian, long form "بودن" /bu-danˈ/ connotes the existential 'to be,' meaning 'to exist.' Thus, "خدا هست" /xu-dā#hast/ means "God exists," and "یک کتاب روی میز هست" /jak#ki-tāb#ru-ji#mēz#hast/ means "There is (exists) a book on the table." For all other uses, the short form of the verb 'to be,' or "است" /ast/, is used and means 'is.'

In Dari, because *Glottal will cause all instances of /h/ to elide, "استن" /a-stanˈ/ and "بودن" /bu-danˈ/ have been reinterpreted as the same verb with infinitive "بودن" /bu-danˈ/, and the sole differentiation between /hast/ and /ast/ much like any other word containing /h/, is register. That is, Dari speakers will frequently use the long form of "بودن" /bu-danˈ/ hypercorrectly as a formal pronunciation of "استن," but with no difference in semantic content.

[73] This nomenclature is purely Iranian, as Dari and Tajik speakers consider "بودن" /bu-danˈ/ to be the infinitive of all three forms, and reckon the difference as suppletive register/dialectal variation.

Tajik, when a copula is used, will have all three forms available to it, but as stated will favor by far the short from of "بودن" /bu-dan^{l}/. The Tajik null copula is in wide use in both the northern and southern dialects and is likely a borrowed feature of either Russian or Turkic, the latter of which has no copula in the third person singular and the former of which has none at all in the present tense. When Tajik does in fact employ the long form of "بودن" /bu-dan^{l}/ or "استن" /a-stan^{l}/, it appears that the former, like Iranian, at least connotes the existential usage of the copula.

At this point it becomes apparent that Dari's adoption of [a-sta] (mentioned in Chapter III) as a colloquial form of /ast/ may in fact be the historically regular 3rd person conjugation of "استن" /a-stan^{l}/. This stands in contrast to the accepted view that [a-sta] is an erroneous reinterpretation of /ast/ analogous with regular verb conjugation, whose 3rd person singular form ends in /-ad/ which reduces to /-a/ in Dari. That is, [a-sta] is in fact */a-stad^{l}/ reduced.

The 3rd person singular conjugation of the copular verb "بودن" /bu-dan^{l}/ 'to be,' aside from the semantic facts mentioned, has its own phonology that varies according to dialect, and can in fact serve as somewhat of a Shibboleth among Afghan and Iranian dialects. Because the lexical item under question will remain the 3rd person singular 'is,' it will not be necessary to provide a gloss for the item each time it arises.

In Dari, “است” /ast/ undergoes reduction to varying degrees according to register, and in some cases undergoes epenthesis:

	Formal Dari:		Reduced:	Spoken Dari:
“است”	/ast/ → [ast]	→	[as] or rarely [a]	[as^{l}-ta]

Iranian will be similar, but will omit the epenthetic form altogether, and in normative colloquial speech will elide the entire coda:

	Formal Iranian:	Spoken Iranian:
“است”	/æst/ → [æst]	→ [st] → *[æ] → [ɛ]

Tajik is distinct from Dari and Iranian in its treatment of the copula as optional, and in its third person singular conjugation:

	Formal Tajik:		Spoken Tajik:
“است”	/ast/ → [ast]	→	[ai]

The following rule sets describe the colloquial realization of /hast/ in Dari:

/ast/ → [as] / C#__ “مرد است” /mard#ast/ → [mar^{l}-das] ‘(It) is a man.’

/ast/ → [s] / V#__ “خانه است” /xā-na#ast/ → [xā-nas^{l}] ‘(It) is a house.’

Iranian is somewhat more complicated in that for each potential ending before the copula /æst/, the grammar retains the option of either eliding the copula vowel or eliding the copula consonant cluster:

/æst/ → [æst] / C#__ “مرد است” /mard#ast/ → [mærˈ-dæst] ‘(It) is a man.’

/æst/ → [st] / V#__ “چلو است” /ʧɛ-lo#ast/ → [ʧɛ-lostˈ] ‘(It) is rice.’

Or:

/æst/ → [ɛ] / C#__ “مرد است” /mard#ast/ → [mærˈ-dɛ] ‘(It) is a man.’

/æst/ → [ɛ] / V#__ “چلو است” /ʧɛ-lo#ast/ → [ʧɛ-lo-ɛˈ] ‘(It) is rice.’

In the case of preceeding lexical items ending in [-ɛ], the following somewhat surprising alternation takes place:

/æst/ → [ɛ] / C#__ “ژاله است” /ʒā-lɛ#ast/ → [ʒɑ-lɛːˈ] ‘(It) is Zhaleh.’

/æst/ → [ɛ] / V#__ “ژاله است” /ʒā-lɛ#ast/ → [ʒɑ-lastˈ] ‘(It) is Zhaleh.’

<u>VI.J.3.a 3rd Person Singular Copula “است” (/ast/) in Dari</u>

The constraints that bring about the Dari and Tajik reduction in the copula are simple, and in fact a great deal like those seen in section VI.J.1.a on regular verb conjugation. Recall the constraints *+V_nC_nC[oral stop]# (“Morpheme final oral stops are prohibited” from section VI.J.1.a), *Glottal from section VI.H.2, and REDUCE (-Stress) and *VV from section VI.J.1.c.

For the shift /hast/ or /ast/ → [s] / V#__, the following tableau is valid for all three dialects as the phrase “چه است؟” or “چیست؟,” that is [ʧi#ast] or [ʧi#hast] (the difference is arbitrary) or reduced form [ʧist] meaning ‘what is…?’ is common to all three:

MP*/ʧi#hast/	*+C[oral]#	*Glottal	*VV	REDUCE(-str)	MAX-μ-IO
☞ [ʧis]					***
[ʧas]				*!	***
[ʧi#has]		*!			*
[ʧi#hast]	*!	*			
[ʧi-jast]	*!				
[ʧi-as]			*!		**

For the shift /(h)ast/ → [as] / C#__, the same tableau will apply, and the phrase used will be "سگ هست" (/sag#(h)ast/ – lit. 'dog' 'is') or 'It is a dog:'

MP*/sag#hast/	*+C[oral]#	*Glottal	*VV	REDUCE(-str)	MAX-μ-IO
☞ [sa-gas]					**
[sag#hast]	*!	*			
[sag#has]		*!			
[sag#jast]	*!				
[sagst]	*!				
[sags]					***!

VI.J.3.b 3rd Person Singular Copula "است" (/ast/) in Iranian

For the Iranian shift, the analysis will be more complicated, and as noted above, will require a two-fold analysis, /æst/ → [æst] or [st] and /æst/ → [ɛ] or [æst] with ellipsis of preceeding /-ɛ/. Recall that in Iranian, /hæst/ is distinct from /æst/, and the former will retain its full form regardless, so /ʧī#hæst/ will remain [ʧī#hæst] ('what is there; what exists?'), but /ʧī#æst/ will become /ʧīst/ ('What is (it)?'). The first analysis will be straightforward. Recall from section VI.J.1.d.i the constraint MAX-$_{LEX}$-IO ("Ellipsis in lexical items is prohibited."), and note the following constraint:

$MAX_{MORPH}IO$ (“Ellipsis of whole morphemes is prohibited.”).

/æst/ → [æst] / C#__ “مرد است” /mard#ast/ → [mærˈ-dæst] ‘(It) is a man.’

/æst/ → [st] / V#__ “چلو است” /ʧɛ-lo#ast/ → [ʧɛ-lostˈ] ‘(It) is rice.’

MP*/ʧi#æst/	*VV	ID-IO (Pal)	REDUCE(-str)	MAX-μ-IO	DEP-μ-IO
☞ [ʧˈist]				*	
[ʧæs]			*!	**	
[ʧi#æs]	*!			*	
[ʧi#æst]	*!				
[ʧi-jæst]		*!			*
[ʧi-æs]	*!			**	

For the shift /æst/ → [æst] / C#__, the same tableau will apply, and the same phrase from above will will be used: “سگ هست” (/sæg#æst/ – lit. ‘dog’ ‘is’) or ‘It is a dog:’

MP*/sæg#æst/	*VV	ID-IO (Pal)	REDUCE(-str)	MAX-μ-IO	DEP-μ-IO
☞ [sæ-gæst]					
[sæg#æs]				*!	
[sæg#jæst]		*!			*
[sægst]				*!	
[sægs]				**!	

In the case of /æst/ → [ɛ], it is a logical assumption that /æst/ has been reinterpreted as the formal input form of /ɛ/, and they are separate underlying forms. They occurred as follows: /ast/ → /a/ (still attested in some dialects) and then /a/ → /ɛ/ due to the constraint MAX Height]# (“Maximize the height of word-final vowels.”) from section VI.G.2.a:

MP*/ast/	*C]$_{\sigma}$	DEP-μ-IO	MAX-μ-IO
☞ [a]			**
[as]	*!		*
[ast]	*!		
[a-sta]		*!	
∅			***!

It is interesting to note that without DEP-μ-IO, [a-sta] is a possible output, and indeed a common one in Dari. This segment was then reinterpreted as /a/, a variant of /ast/, through lexicon optimization, and the following tableau from VI.G.2.a caused /a/ → /ɛ/:

MP*a]	DEP-μ-IO	MAX-μ-IO	ID-IO(High)	*+Lng –Low	MAX-Hg]#	ID-IO (Low)
a]					**!	*
ō]	*!			*		
i]			*!			
ā]	*!				**	
☞ ɛ]					*	*
∅]		*!				
ad]	*!				**	*

Then the two sentences from above will be derived as follows:

MP*/ʧi#ɛ/	MAX$_{MORPH}$IO	*VV	ID-IO (Pal)	REDUCE(-str)	MAX-μ-IO
☞ [ʧi-ɛ]					*
[ʧɛ]				*!	**
[ʧi]	*!	*			*
[ʧ]	*!	*			

For the shift /æst/ → [ɛ] / C#__, the same tableau will apply, and the phrase used will be “سگ است” (/sæg#æst/ – lit. ‘dog’ ‘is’) or ‘It is a dog:’

MP*/sæg#ɛ/	*VV	ID-IO (Pal)	REDUCE(-str)	MAX-μ-IO	DEP-μ-IO
☞ [sæ-gɛ]					
[sæg]				*!	
[sæg#jɛ]		*!			*

Finally there remains the case of the copula /ɛ/ or /ast/ when the final segment in the preceeding lexical item is /-ɛ/. This is problematic, as the above tableau will yield undesirable results. Recall from above:

/æst/ → [ɛ] / C#__ “ژاله است” /ʒā-lɛ#ast/ → [ʒɑ-lɛːˈ] ‘(It) is Zhaleh.’

/æst/ → [æst] / V#__ (and V → ∅) “ژاله است” /ʒā-lɛ#ast/ → [ʒɑ-læstˈ] ‘(It) is Zhaleh.’

The above tableau will cause “ژاله است” /ʒā-lɛ#ast/ → *[ʒɑ-lɛstˈ] unless we assume that the verbal morphology took effect before the Iranian /a/ → [ɛ] / __# took place. If this is the assumption, and it is a fair assumption that /a#ast/ → [ast] and remained canonical, then the above system is justified and it may be assumed that word-final /ɛ/ in Iranian still enters the evaluator as /a/ when followed by the copula /ast/.

VI.J.3.c 3rd Person Singular Copula “аст” (/ast/) in Tajik

Tajik is unique in its reduction of this form, and deserves separate treatment, as the forms are entirely distinct. To begin with, the affirmative copula is optional in all forms in Southern Tajik, and omitted entirely in the Northern variety. Furthermore, in

the north, the Turkic particle “ми” /mi/ is employed at the end of the sentence in copular questions:

Common Tajik:	Southern Spoken:	Northern Spoken:	Gloss:
“Карим касал аст.” /ka-rim#ka-sal#ast/	[ka-rim#ka-sal#(aj)]	[ka-rim#ka-sal]	“Karim is sick.”
“Карим ҳаст?” /ka-rim#hast/	[ka-rim#haj]	[ka-rim#haj#mi]	“Is Karim there?”
“Ин китоб аст?” /in#ki-tob#ast/	[in#ki-tob#(aj)]	[in#ki-tob#mi]	“Is this a book?”

As a sidenote, the remainder of the short copula forms remain consistant with Dari and Iranian, save their optionality in the South and their omission in the North:

Common Tajik:	Southern Spoken:	Northern Spoken:	Gloss:
“Ман касал ам.” /man#ka-sal#am/	[ma#ka-sal#(am)]	[man#ka-sal]	“I am sick.”
“Ту касал и.” /tu#ka-sal#i/	[tu#ka-sal#(i)]	[tu#ka-sal]	“You are sick.”
“Онҳо касал анд?” /on-ho#ka-sal#and/	[U-no#ka-sal#(an)]	[u-no#ka-sal#mi]	“Are they sick?”

The realization of “аст” /ast/ as [aj] and “ҳаст” /hast/ as [haj] appears anomalous and may be realated to a similar form in Turkic, but this unclear at this point.

VI.J.4 The Ezafe and Enclitic Vowel Lowering (Dari, Iranian and Tajik)

As mentioned at various points in this dissertation, there exists a phenomenon in Persian referred to as the 'ezafe,' first mentioned in section III.C.3. The enclitic /ɛ/ is epended postpositionally to nouns or noun phrases when followed by a modifier (either nominal or adjectival):

"دوست من"	/dost#ɛ#man/ → [dos-tɛ-man^{l}] 'friend-e + I'	'my friend'
"دوست من"	/dost#ɛ#xob/ → [dos-tɛ-xob^{l}] 'friend-e + good'	'good friend'

Note in the above that Iranian and Tajik will be different in pronunciation, but the same in usage, so Iranian would render the above [dus-tɛ-mænˡ] and [dus-tɛ-xubˡ] respectively, and Tajik will realize them as [d~~u~~s-t^{j}i-manˡ] and [d~~u~~s-t^{j}i-xubˡ]. The phonology of the epenthesis is predictable:

/ɛ/ → [ɛ] / C__; V [+Back] __

/ɛ/ → [jɛ] / V [+Front] __

/-a + ɛ/ → [ɛ] (in Dari and Tajik)

That is, the ezafe will epend to a back vowel or consonant without effect, to a front vowel with an epenthetic palatal glide inserted, and to the short low mid vowel causing said vowel to elide. These general rules will hold for Tajik although the ezafe is realized as /i/, and they will hold for Iranian as well, but the final provision will not be necessary, as

in Iranian /-a/ → [-ɛ], and so /ɛ/ → [jɛ] / V [+Front] __ will apply to all cases that are /-a/ in Dari and Tajik. Examples are as follows:

	Modern Persian:		Gloss:
Back vowel: "پتوی سبز"	/pa-tuˈ-ɛ#sabzˈ/	[pa-tuˈ-ɛ#sabzˈ]	'green blanket'
Front vowel: "فارسی تهران"	/fɑr-siˈ-ɛ#tɛh-rɑnˈ/	[fɑr-siˈ-jɛ#tɛh-rɑnˈ]	'Tehran Iranian'
/-a/ Dari/Tajik: "بچه خوب"	/ba-ʧa-ɛ#xobˈ/	[ba-ʧɛ#xobˈ]	'good child'
/-a/ Iranian: "بچه خوب"	/ba-ʧɛ-ɛ#xubˈ/	[ba-ʧɛ-jɛ#xubˈ]	'good child'

The following constraints will bring about these shifts:

*V[αfront]V[αfront] ("Adjacent vowels matched for ±Front are prohibited." This, much like Maximal Distance in section VI.F is based on ease of parsing.)

$MAX_{MORPH}IO$ ("Ellipsis of whole morphemes is prohibited." This is from the previous section.)

Of the three above shifts, the first two are identical in all three dialects, and so they shall be covered first. /ɛ/ → [ɛ] / C__; V [+Back] __ and /ɛ/ → [jɛ] / V [+Front] __ are simply the constraint *V[αfront]V[αfront] mandating a distinction between adjacent segments for differentiation. It must be stated overtly that although this does not affect the output, the tableaux for Dari and Tajik will be distinct from Iranian, which will be as follows:

MP*/pa-tu + ɛ + sabzˈ/	MAX-μ-IO	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV
☞ [pa-tu + ɛ + sabzˈ]					*
[pa-tu + jɛ + sabzˈ]				*!	
[pa-tu + sabzˈ]	*!	*			
[pa-tɛ + sabzˈ]	*!	*			

Dari and Tajik will be identical save that MAX-μ-IO is demoted to the lowest position:

MP*/pa-tu + ɛ + sabzˈ/	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV	MAX-μ-IO
☞ [pa-tu + ɛ + sabzˈ]				*	
[pa-tu + jɛ + sabzˈ]			*!		
[pa-tu + sabzˈ]	*!				*
[pa-tɛ + sabzˈ]	*!				*

This same tableau will account for the fact that in front vowel-final lexical items, an epenthetic glide is inserted between the lexical item and the modifier or modifier phrase:

MP*/far-si + ɛ + tɛh-ranˈ/	MAX-μ-IO	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV
☞ [far-si + jɛ + tɛh-ranˈ]				*	
[far-si + ɛ + tɛh-ranˈ]			*!		
[far-si + tɛh-ranˈ]	*!	*			
[far-sɛ + tɛh-ranˈ]	*!	*			

As above, Dari and Tajik will exhibit a divergent constraint ranking, but it will not distinguish the output from the above:

MP*/far-si + ɛ + tɛh-ranˈ/	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV	MAX-μ-IO
☞ [far-si + jɛ + tɛh-ranˈ]			*		
[far-si + ɛ + tɛh-ranˈ]		*!			
[far-si + tɛh-ranˈ]	*!				*
[far-sɛ + tɛh-ranˈ]	*!				*

With the final environment, /-a + ɛ/ → [ɛ] in Dari and Tajik and /-a + ɛ/ → [-a + ɛ] in Iranian, the same tableau will account for both, only as above, a faithfulness constraint will be elevated above all of the other constraints in the Iranian tableaux:

MP*/ba-tʃɛ + □ + xubˈ/	MAX-μ-IO	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV
☞ [ba-tʃɛ-jɛ-xubˈ]				*	
[ba-tʃɛ + ɛ + xubˈ]			*!		*
[ba-tʃɛ + xubˈ]	*!	*			*
[ba-tʃɛj + xubˈ]	*!	*		*	

MP*/bæ-tʃɛ + æʃˈ/	MAX-μ-IO	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV
☞ [bæ-tʃɛ-jæʃˈ]				*	
[bæ-tʃæʃˈ]	*!				
[bæ-tʃɛʃˈ]	*!				
[bæ-tʃɛ + æʃˈ]			*!		

Crucially, Dari and Tajik will lower the faithfulness constraint MAX-μ-IO ("Ellipsis is prohibited."):

MP*/ba-tʃa + ɛ + xubˈ/	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV	MAX-μ-IO
☞ [ba-tʃɛ-xubˈ]					*
[ba-tʃa + ɛ + xubˈ]				*!	
[ba-tʃa + xubˈ]	*!				*
[ba-tʃa + jɛ + xubˈ]			*!		
[ba-tʃaj + xubˈ]	*!				*
[ba-tʃɛj + xubˈ]			*!		*

VI.J.4.a The Ezafe and Enclitic Vowel Lowering (Dari)

The previous section detailed the phonology of the ezafe, and the current section will deal with the possessive enclitic. At the end, a note will be made regarding the 1st person singular conjugation of the present perfect tense, as it was indicated in VI.J.1.c that this would be returned to.

The possessive enclitic and short form copula verbs are identical for 1st person singular, and this will be pertinent to the analysis:

Copula Verb:			Possessive Pronoun:		
“ام”	/am/	‘I am’	“ام”	/am/	‘my’
“ای”	/i/	‘you-s. are’	“ات”	/at/	‘your-s.’
“است”	/ast/	‘s/he is’	“اش”	/aʃ/	‘his/her’
“ایم”	/em/	‘we are’	“ما”	/mɑ/	‘our’
“اید”	/ed/	‘you- pl. are’	“تان”	/tɑn/	‘your-pl.’
“اند”	/and/	‘they are’	“شان”	/ʃɑn/	‘their’

The Dari vowel lowering phenomenon shared by the 1st person present perfect and the possessive pronoun enclitic morphology is as follows:

/kar	-	da	+	am/	→	[ka-daːmˈ] → [ka-demˈ]
‘do-past part.’			+	‘I am’		‘I have done’
/xa - ri -		da	+	am/	→	[xa-ri-daːmˈ] → [xa-ri-demˈ]
‘buy-past part.’			+	‘I am’		‘I have bought’
/ba	-	ʧaˈ	#	am/	→	[ba-ʧaːmˈ] → [ba-ʧemˈ]
‘child’			+	‘my’		‘my child’
/ba	-	ʧaˈ	#	at/	→	[ba-ʧaːtˈ] → [ba-ʧetˈ]
‘child’			+	‘your-fam.’		‘your child’
/ba	-	ʧaˈ	#	aʃ/	→	[ba-ʧaːʃˈ] → [ba-ʧeʃˈ]
‘child’			+	‘his/her’		‘his/her child’

The grammar of Dari will be seen to necessitate an ezafe enclitic when the postpositional possessive is used, and in this case, unlike the case of noun+adjective, Iranian and Tajik do not:

MP*/ba-ʧa + ɛ + aʃʲ/	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV	MAX-μ-IO
☞ [ba-ʧɛʃʲ]					**
[ba-ʧa + ɛ + aʃʲ]				*!	
[ba-ʧa + aʃʲ]	*!			*	*
[ba-ʧa + jɛʃʲ]			*!		*
[ba-ʧa + ɛʃʲ]				*!	*
[ba-ʧɛ + aʃʲ]				*!	*
[ba-ʧaʃʲ]	*!				**

Iranian:

MP*/ba-ʧɛ + aʃʲ/	MAX-μ-IO	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV
☞ [ba-ʧɛ-æʃʲ]					*
[ba-ʧɛ + jɛʃʲ]				*!	
[ba-ʧɛ + jæʃʲ]				*!	
[ba-ʧɛ + ɛʃʲ]			*!		*
[ba-ʧɛʃʲ]	*!	*			
[ba-ʧæʃʲ]	*!	*			

Tajik:

MP*/ba-ʧa + aʃʲ/	MAX-μ-IO	$MAX_{MORPH}IO$	$*V_{\alpha\text{-ft}}V_{\alpha\text{-ft}}$	ID-IO(pal)	*VV
☞ [ba-ʧaːʃʲ]					*
[ba-ʧa + jaʃʲ]				*!	
[ba-ʧaʃʲ]	*!	*			

So, due to the fact that the 1st person singular conjugation of the copula verb and possessive enclitic are the same, and Dari requires the use of the ezafe with possessive pronouns, the forms are reanalyzed as analogous, and the Dari 1st person singular present perfect conjugation morphology matches the normative 1st person singular possessive.

Unfortunately for this analysis is the availability of the following forms in colloquial Iranian:

/kær	-	dɛ	+	æm/	→	[kær-dæːmʲ]
‘do-past part.’			+	‘I am’		‘I have done’

/xæ - ri	- dɛ	+	æm/	→	[xæ-ri-dæːmˈ]
‘buy-past part.’		+	‘I am’		‘I have bought’
/bæ -	ʧɛˈ	#	æm/	→	[bæ-ʧæːmˈ]
‘child’		+	‘my’		‘my child’
/bæ -	ʧɛˈ	#	æt/	→	[bæ-ʧæːtˈ]
‘child’		+	‘your-fam.’		‘your child’
/bæ -	ʧɛˈ	#	æʃ/	→	[bæ-ʧæːʃˈ]
‘child’		+	‘his/her’		‘his/her child’

This was called unfortunate because, like the copula verb following lexical items with /-ɛ/ in Persian (/bæ-ʧɛ#æst/ → [bæ-ʧæst] ‘(it) is a child’- cf. section VI.J.3.b), the word-final /-ɛ/ becomes [-æ] or is elided altogether. This, like the aforementioned case in VI.J.3.b, will be considered a relic that predates the shift /a/ → [ɛ] / __#.

VI.J.4.b Borrowed Enclitic Vowel Lowering (Dari, Iranian and Tajik)

One final case of lowering that for now will be viewed as anomalous is that of the negative present tense copula:

Neg. enclitic:	‘to be’:		Dari/Tajik:	Iranian:
“نه” /naˈ/ +	“استم” /aˈ-stam/	=	“نیستم” /neˈ-stam/	/niˈ-stæm/
	“استی” /aˈ-sti/	=	“نیستی” /neˈ-sti/	/niˈ-sti/
	“است” /ast/	=	“نیست” /nest/	/nist/
	“استیم” /aˈ-stem/	=	“نیستیم” /neˈ-stem/	/niˈ-stim/
	“استید” /aˈ-sted/	=	“نیستید” /neˈ-sten/	/niˈ-stid/
	“استند” /aˈ-stand/	=	“نیستند” /neˈ-stand/	/niˈ-stænd/

Tajik, in keeping with the affirmative forms, will maintain the option in colloquial speech of further copular reduction:

“ман не”	/man#nje/ → /ma#nje/
“ту не”	/tu#nje/ → /tu#nje/
“вай нестай”	/vaj#nje/ → /vaj#nje-staj/[74]
“мо не”	/mo#nje/ → /mo#nje/
“шумо не”	/ʃʊ-mo#nje/ → /ʃʊ-mo#nje/
“онҳо не”	/on-ho#nje/ → /u-no#nje/

In the absence of additional historical data, this shall remain without explanation, but it does bear note that although Iranian does not retain usage of the verb “استن” /a-stanˈ/, the above negative copula is clearly derived from “استن” /a-stanˈ/ and remains in use in Iranian.

VI.K Proto-Persian Phonotactics (Dari, Iranian, Tajik)

In section III.E.12, it was noted that Dari is unique among the three dialects under discussion in that it optionally permits onset consonant clusters. However, Iranian is unique among the dialects in allowing consonant + syllabic consonant in the coda. Dari and Iranian are opposite in this regard. Tajik will allow neither onset consonant clusters nor coda consonant clusters. Simply put, the following illustrates:

Dari:	CCVC; *CVCC̩
Iranian:	*CCVC; CVCC̩
Tajik:	*CCVC; *CVCC̩

[74] Note that the addition of [aj] is analogous with the affirmative third person singular form.

CCVC:

	Iranian:	Tajik:	Dari:	Gloss:
“بلند”	[bo-lænd^{l}]	[ba-łand^{l}]	[bland] or [b^{a}land^{l}]	‘tall’
“برنج”	[bɛ-rɛnʤ^{l}]	[b^{l}-rinʤ^{l}]	[brɛnʤ] or [b^{a}rɛnʤ^{l}]	‘rice; iron’
“شکم”	[ʃɛ-kæm^{l}]	[ʃ^{l}-kam^{l}]	[ʃkam] or [ʃ^{a}kam^{l}]	‘belly’

*CVCC̣:

	Iranian:	Tajik:	Dari:	Gloss:
“چشم”	[ʧʃɛʃm̩] or [ʧʃɛʃ]	[ʧa^{l}-ʃam]	[ʧ^{ɛ}ʃam] or [ʧɛ^{l}ʃam]	‘eye’
“فکر”	[fɛkr̩] or [fɛk]	[fi^{l}-kar]	[fɛ^{l}-kar] or [fə^{l}-kar]	‘thought’
“شکر”	[ʃokr̩]	[ʃu^{l}-kar]	[ʃʊ^{l}-kar] or [ʃu^{l}-kar]	‘thanks’
“طفل”	[tɛfl̩]	[ti^{l}-fal]	[tɛ^{l}-fal]	‘child’
“قطل”	[ɢætl̩]	[qa^{l}-tal]	[qa^{l}-tal]	‘murder’
“قدر”	[ɢædr̩] or [ɢæd]	[qa-dar^{l}]	[qa^{l}-dar]	‘amount’
“شغل”	[ʃoɣl̩]	[ʃʊ^{l}-ɣal]	[ʃʊ^{l}-ɣal]	‘occupation’
“شکل”	[ʃækl̩]	[ʃa^{l}-kal]	[ʃa^{l}-kal]	‘shape; form’
“قهر”	[ɢæhr̩]	[qa^{l}-har] or [qaːr]	[qa^{l}-har] or [qɑr]	‘angry’
“شهر”	[ʃæhr̩]	[ʃa^{l}-har] or [ʃaːr]	[ʃa^{l}-har] or [ʃɑr]	‘city’

Mercifully, this may be the simplest analysis yet, as the constraints have already been overtly stated above, and both of the constraints are based on the notion of the ideal syllable structure as CV:

*CCVC (“Onset consonant clusters are prohibited.”)

*CVCC̩ (“Syllablic consonants are prohibited in syllable codas.”) The parentheses are added because the constraint could be more broadly stated as *C̩, but as with any theory, it is superfluous.)

The lexical items used for the below tableaux will be “فکر” /fɛkr̩/ ‘thought’ and “شکم” /ʃkam/ ‘belly.’

For Dari, *CVCC̩ will cause the epenthetic vowel to be inserted between the consonant and syllabic consonant in the coda:

MP*/fɛkr̩/	*CVCC̩	ID-IO(Stress)	DEP-μ-IO	MAX-μ-IO
[fɛkr̩]	*!			
[fɛk]				*!
☞ [fɛˈ-kar]				
[fɛ-karˈ]		*!		
[fɛr]				*!

And CCVC will still be allowed:

MP*/ʃkam/	*CVCC̩	ID-IO(Stop)	DEP-μ-IO	MAX-μ-IO
[ʃɛ-kamˈ]			*!	
☞ [ʃkamˈ]				
[ʃɛkm̩]	*!			
[ʃjam]		*!		

For Iranian, the inverse will be true as *CCVC will cause the complex onset to take an interconsonantal epenthetic vowel (CVCVC):

MP*/ʃkam/	*CCVC	ID-IO(Stress)	ID-IO(Stop)	MAX-μ-IO	DEP-μ-IO
☞ [ʃɛ-kamˈ]					*
[ʃkam]	*!				
[ʃɛkm̩]		*!			
[ʃjam]			*!		

And CVCC̩ will still be permitted:

MP*/fɛkr̩/	*CCVC	ID-IO(Stress)	ID-IO(Stop)	DEP-μ-IO	MAX-μ-IO
☞ [fɛkr̩]					
[fɛk]					*!
[fɛ-karˈ]		*!		*	
[fɛˈ-kar]				*!	
[fɛr]			*!		

One note on this second tableau for Iranian is that in casual Iranian speech the constraint *CVCC̩ can optionally be raised above MAX-μ-IO causing the output [fɛk], which is available in the speech and only not desirable among Irish Persians:

MP*/fɛkr̩/	*CCVC	ID-IO(Stress)	ID-IO(Stop)	DEP-μ-IO	*CVCC̩	MAX-μ-IO
[fɛkr̩]					*!	
☞ [fɛk]						*
[fɛ-karˈ]		*!		*		
[fɛˈ-kar]				*!		
[fɛr]			*!			

For Tajik, both constraints, *CCVC and *CVCC̩, will be part of the evaluator, and in contrast to the above counterparts, in Tajik, MAX-μ-IO will outrank DEP-μ-IO in the derivation:

MP*/fɛkr̩/	*CCVC	*CVCC̩	ID-IO(Stress)	ID-IO(Stop)	MAX-μ-IO	DEP-μ-IO
[fɛkr̩]		*!				
[fɛk]					*!	
[fɛ-karˈ]			*!			
☞ [fiˈ-kar][75]						*
[fɛr]				*!		

MP*/ʃkam/	*CCVC	*CVCC̩	ID-IO(Stress)	ID-IO(Stop)	MAX-μ-IO	DEP-μ-IO
☞ [ʃɛ-kamˈ]						*
[ʃkamˈ]	*!					
[ʃɛkm̩]		*!	*			
[ʃjam]				*!		

VI.L Conclusion

Doubtlessly the bulk of the analysis, and by far the most difficult to justify, Chapter VI has set out to accomplish three goals. First, Chapter VI has sought to bring together the data presented in Chapters III – V and unify them as one Persian with three sets of peculiarities. Second, this chapter has sought to tie all of the changes observed in Modern Persian to some fact of either physiological or historical relevance, and in so doing to avoid the road of ad hoc explanation. Third, this chapter has sought to demonstrate that OT is a theoretical tool by which the first two goals may be accomplished.

[75] The normative high front vowel in Tajik shifts to [i], but this is considered trivial to the phonotactic portion of this analysis.

VII. Implications for Use-Based Approach to Language Change

VII.A Introduction

In recent years there has been an increasing amount of attention paid to the effects of frequency on phonological processes in actual language. Due to the fact that sound change is dynamic and cannot be easily captured, the bulk of research on the issue derives from inquiry into the nature of synchronic variation (see Labov, p. 1994). The overall significance of this has been that some gradient diachronic sound changes, typically manifest in synchronic variation, have been accounted for by the implementation of frequency effects, and are thus tied to probability. That is to say, a more frequent lexical item will be given more instantiation in actual language, and hence the sound change process will be accelerated in those more frequent items. The ultimate intent of a frequency based approach to exceptional sound change is to demonstrate that usage based facts (cf. Bybee, 2000, 2001) can be implemented to avoid the inclusion of stipulated exception rules. The purpose of this section is to suggest a potential explanation for the phonological processes discussed thus far in this dissertation that have been shown to occur only a percentage of the time. Furthermore, I have chosen data that indicate that while frequency facts can indeed shed light on the copious instances of gradience in phonological processes, they create the need for a new set of exceptions to account for incongruous frequency data, and in some cases they simply fail. I will conclude that frequency alone cannot account for the availability of language shifts and in fact the process is likely a great deal more dynamic than it had been assumed.

Bybee (2000) suggests this use-based model using the notion of exemplars, by which linguistic data is stored at the word or even phrase level (pp. 257-258), and each use of a word or phrase potentially alters the stored item. Thus the grammar emerges entirely from use, such that the morphological affixes and inflectional information are inferred from their regularity among stored items (due to the fact that inflected items are stored separately from non-inflected ones.) The lure of this approach is that it allows for a much sought after connection between phonetics and phonology in a system that utilizes phonetic facts such as ease of articulation to account for phonological processes such as place assimilation. This idea is of course not new, as articulatory phonology and Gesture Theory (cf. Browman and Goldstein, 1992) have been considered in the past. It is not satisfactorily explanatory to proclaim that a change is not complete or to express exceptions to wholesale shifts, when frequency or (in a related vein) register are clearly at play. This will test the notion of frequency effects using only Dari, and using only a small set of informants as a control set. For this reason, the results will not be included as proof for or against a use-based approach, but rather as the beginnings of what would prove an interesting line of inquiry. I have chosen to use only Dari in this Chapter as it displays the most reduction of the three dialects, and hence a more robust specimen may be obtained.

To remain with the notion of ease of articulation, this chapter will examine three of the phonological processes of reduction in Dari. The first will be the deletion of word final /n/ in frequent lexical items (see section VI.H.4), namely those in the word-final position of pronouns and a few other highly frequent items. These alternations have been

seen to even take place morpheme-internally and within phrasal constructions. The second process examined will be word final and complex coda initial deletion of /r/ in some frequent items and not in others (see section VI.H.3). In this section, it will be demonstrated that preservation of lexical distinction can serve as a retardant of such frequency conditioned phonological change. The third and final process examined will be that of the alternation of the borrowed Arabic sequences /aʔ/ and /aʕ/ with [ɑ] (see section VI.H.2.a). It will be shown that while frequency facts do perhaps play a significant role, register facts will account for the data with more efficacy.

Phillips (1994, 1998, 2000) provides compelling evidence for a distinction between changes which originate from the least frequent lexical items and proceed to the most frequent, and those displaying the opposite order. Her distinction is drawn along the lines of phonetically motivated sound change, such as assimilation and reduction, and non-phonetically motivated change, such as analogical morphological spreading. From the outset, I do not intend to address this distinction or even consider the latter type of change, and so if her predictions are correct, the following changes are expected to occur from most to least frequent lexical items.

VII.B /n/-drop

The first process under consideration is one of the deletion of word final nasal in pronominals and a couple of other frequent tokens. The reason for starting with this change is that it is a naturally frequent environment for language change, and hence no further frequency analysis is necessary. As a side note, Phillips' analysis will predict that phonetically motivated change should affect these frequent lexical items first, and

morphological changes will affect them last. A case in point is that of analogical leveling of case marking in English except for the case of pronominals, which retain such case marking. This of course stands in sharp contrast to the currently analyzed Afghan Persian data in which a phonological change affects only these more frequent lexical items. Thus, Phillip's predictions are upheld.

As previously discussed, these changes occur in the colloquial variant of Dari, and the formal counterpart is simply the token before the ellipsis. Thus, the colloquial variant will represent the change, and the formal variant will represent the base form in terms of pronunciation:

Formal Dari:		Colloquial Dari:	
/man/	→	[ma]	'I'
/ɑn/	→	[u]	'he, she, it, that'
/in/	→	[i]	'he, she, it, this'

This change will be seen to occur systematically in some other nasal-final lexical items, but not in all. Notice the following:

Formal Dari:		Colloquial Dari:	
/hɑn/	→	[o]	'yes; yeah'
/kʊn/	→	[ko]	'do!'
/tan/	→	*[ta]	'body; *person classifier*'
/ʤɑn/	→	*[ʤɑ]	'dear'
/alʔɑn/	→	*[alʔɑ]	'now'
/nɑn/	→	*[nɑ]	'bread; food'

In the above, it is obvious that frequency has played a role, as the affirmative item 'yes' displays the deletion, and the last four do not. As for the second person singular imperative, it bears recalling that the majority of Persian verbs have come to be used in a

compound construction using “کردن” /kar-dan/ ‘to do,’ such that “نوشتن” /na-ʊiʃ-tan/ ‘to write’ is colloquially “نوشته کردن” /na-ʊiʃ-ta#kar-dan/, “پرسیدن” /pur-si-dan/ ‘to ask’ is “پرسان کردن” /pur-sān#kar-dan/, and what have you. Thus, the above imperative is highly frequent, and in the corpus to be utilized later, “کردن” /kar-dan/ is by far the most frequent verb after “بودن” /bū-dan/ ‘to be.’

It further bears recalling that this n-deletion takes place word-internally and at the post-lexical level:

Formal Dari:		Colloquial Dari:	
/ɑnhɑ/	→	[unɑ] → [wɑ]	‘they, those’
/inhɑ/	→	[inɑ] → [jɑ]	‘they, these’
/man#ham/	→	[maam] → [mɑm]	‘me too’

Note that in all of the above cases both /n/ and /h/ delete in the spoken variant, and in the third example, /aa/ has changed to [ɑ], the normative interpretation of long /a/. This being the case it would appear that Dari displays a highly frequency sensitive sound change.

VII.C /r/-deletion

The second process under consideration is that of word final /r/ deletion and one case of morpheme internal /r/ deletion that also seems to point towards a frequency based analysis of incomplete sound changes. By incomplete, I mean not in the fashion of the Neo-Grammarian sound change, which is an unconditional wholesale shift in the phonemic inventory of a language. Although it is a distinct possibility that these are in

fact such changes, only they are in the early stages of metamorphosis. This phenomenon involves the deletion of word final and coda cluster initial /r/ in frequent lexical items in the dialect. There are however examples of lexical items with equal or greater frequency that do not incorporate the potential for this sound change:

Formal Dari:	Colloquial Dari:	Gloss:
/dar/ →	[da]	'in'
/agar/ →	[aga]	'if'
/digar/ →	[diga]	'other'
/ʧī-tawr/→	[ʧˈto]	'which'
/īn-tawr/→	[ito]	'like this'
/ān-tawr/→	[uto]	'like that'
/xajr/ →	[xaj]	'fine; then'
/kar-dan/→	[ka-dan]	'to do'

but

/digar/ →	[digar]	'afternoon'
/tawr/ →	[tawr]	'manner; way'
/bar/ →	[bar]	'on'
/har/ →	[har]	'every'

It was mentioned earlier that "دیگر" /dī-gar/ had two separate meanings with divergent frequencies in the grammar: 'other' and 'afternoon,' but another lexical item appears twice above too. "طور" /tawr/ 'manner.'[76] This would indeed point to a frequency effect as the token meaning 'other,' is clearly more frequent than the token meaning

[76] Note that /īn-tawr/ and /ān-tawr/ are simply the demonstrative pronouns prefixed to the lexical item /tawr/, and so in essence, they are based on an identical lexeme.

‘afternoon,’[77] as is the token meaning ‘like so’ more frequent than that which means ‘manner.’

In the current line of research, I have compiled a small corpus from online news sources of some 26,888 Persian tokens, and run them through a frequency indexer. This is in fact a very small corpus, but due to the fact that this is a preliminary investigation, a small corpus will suffice. The results show that “در” /dar/ ‘in’ is the most frequent token with 1081 instantiations. “اگر” /a-gar/ ‘if’ and “دیگر” /dī-gar/ ‘other’ are also rather high with 14 and 68 respectively. An issue arises, however with the item ‘بر’ /bar/ ‘on’ in that it occurs 93 times out of the possible 26,888 in the corpus. This is unexpected as it outranks both “اگر” /a-gar/ and “دیگر” /dī-gar/ in frequency. I contend that this warrants the consideration of distinction preservation in sound change, as the preposition “به” /ba/ ‘at; to’ occurs some 752 times in the corpus. If “بر” /bar/ were to participate in the ellipsis, the two prepositions would fall together as /ba/, and their semantic distinction would be confused. Furthermore, “هر” /har/ ‘all’ appears 33 times in the corpus, and deletion is not permitted. Why “هر” /har/ should not be seen to participate in the sound change is enigmatic, as there is seemingly no other lexical item with which confusion

[77] In fact, /dī-gar/ is specifically used for the period of time between around four o’clock and the beginning of twilight.

might be caused. One possibility is that “هر” /har/ is always followed by some other lexical item which it modifies:

/har#rōz/	[arroz]	‘every day’
/har#kas/	[arkas]	‘everyone’
/har#ʤʊmʔa/	[arʤʊma]	‘every Friday’

Unfortunately, the data above clearly demonstrate that even word-internally, these sound changes take place, and furthermore, “در,” being a preposition, is also always followed by a lexical item, and yet the sound change is universal.

Beyond this, these findings are problematic in that not only are we required to account for “بر” /bar/ by positing an exception for ambiguity avoidance, but there is also the issue of “دیگر” /dī-gar/ and its distinct phonological treatment due to semantic facts. That is to say, that no matter the meaning of the lexical item at the time of utterance, the frequency facts will hold and the word should exhibit ellipsis throughout the lexicon. I will leave the issue at this: frequency effects will be retarded, or in this case prevented, when the result of the phonological process is ambiguity, and some other effect seems to be at play in the item “دیگر” /dī-gar/, allowing the change in one semantic environment and disallowing it others. “هر” /har/ is simply an exception, a stipulation that is never desirable when working towards the goal of proving a theoretical framework.

VII.D /aʕ/ → [ɑ]

VII.D.1 /aʕ/ → [ɑ]: Frequency

This final sound change occurs as the Dari analysis of Arabic loanwords containing the sequence /aʕ/. Recall that in Dari both /aʕ/ and /ah/ can become /ɑ/ due to ellipsis and compensatory lengthening, but for the sake of simplicity, /ah/ will be ignored to factor out that distinction from the analysis.

It can be assumed that once the intervocalic pharyngeal approximant is deleted, the resultant sequence is treated as a long /ā/. That is to say that the borrowed Arabic item "بعد" /baʕd/, meaning 'after,' as discussed becomes [baad], and then can be realized either as [bad] or as [bɑd], the former being formal register Dari and the latter being colloquial. The crux of the issue at hand is that the lexical item "دفعتا" /da-faʕ-tan/ is realized as [da-fa-tan] in Dari, but never as *[da-fɑ-tan]. There are additionally quite a few lexical items between those two in terms of acceptability of sound change.

An experiment was conducted in which six native Dari speakers ranging in age from twenties to sixties were surveyed for acceptability judgments in twenty-five Persian lexical items in random order. All of the informants used for this survey were from Kabul, and hence speak standard Kabuli dialect. They were briefly informed of the phenomenon, and then asked to rate each item on a scale of one to ten using the following as their guide:

۱- درست است، من اینطور می گویم. (It is correct, I would say it like this)

۲- درست است، اما شاید به هر دو طریق بگویم. (It is correct, but I might say it both ways)

۳- درست است، اما من اینطور نمی گویم. (It is correct, but I do not say it like that)

۴- شنیده ام، مگر به نظرم این تلفظ درست نیست. (I have heard it, but I think it is bad)

۵- به نظرم، هیچ کس اینطور نمی گوید. (In my opinion no one would say it like that)

I combined the results of their judgments with the frequency facts, and discovered that after the first two items, they were blatantly not a match:

underlying		alternation	gloss	judgement	freq.
/baʕd/	→	[bɑd]	'after'	2.34	26
/maʕnɑ/	→	[mɑ-nɑ]	'meaning'	2.34	9
/naʕnɑʕ/	→	[nɑ-nɑ]	'mint'	2.34	0
/waʕda/	→	[wɑ-da]	'promise'	2.67	0
/baʕarusxɑna/	→	[bɑrusxɑna]	'to the wedding room'	2.67	0
/maʕtal/	→	[mɑ-tal]	'waiting'	3	0
/naʕl/	→	[nɑl]	'horseshoe'	3	0
/baʕzi waqt/	→	[bɑzi waqt]	'sometimes'	3.3	4
/maʕlum/	→	[mɑ-lum]	'effect'	4	0
/maʕmulan/	→	[mɑ-mu-lan]	'usually'	5	0
/daʕwat/	→	[dɑ-wat]	'invitation'	5.34	0
/yaʕnɑ/	→	[yɑ-nɑ]	'meaning'	5.34	6
/taʕmir/	→	[tɑ-mir]	'building'	5.67	0
/maʕrfi/	→	[mɑr-fi]	'introduced'	6	14
/maʕlɛm/	→	[mɑ-lɛm]	'teacher'	6.34	0
/ɑbɛ maʕdani/	→	[ɑ-bɛ mɑdani]	'sparkling water'	6.5	0
/taʕlim/	→	[tɑ-lim]	'instruction'	7	0
/maʕmul/	→	[mɑ-mul]	'usual'	7	10
/baʕzan/	→	[bɑ-zan]	'sometimes'	7.5	0

/maʕdan/	→	[mɑ-dan]	‘mine’	8.17	0
/maʕni/	→	[mɑ-ni]	‘meaning’	8.34	9
/taʕqib/	→	[tɑ-qib]	‘chase; pursuit’	8.5	11
/saʕd/	→	[sɑd]	‘auspicious’	9.5	0
/waʕz/	→	[wɑz]	‘admonition’	9.5	0
/dafaʕtan/	→	[da-fɑ-tan]	‘definitely’	10	0

Judging by the results in say /baʕd/, /maʕnā/, and /dafaʕtan/, there may be said to be some connection between frequency and rate of change (by grammaticality judgments). In fact, I will maintain that frequency is indeed an important factor, but only insofar as it relates to other facets of language use. Returning to the notion of the colloquial to formal distinction, we turn to register in an attempt to account for the apparently frequency independent gradience in the above shift. Note the following:

/maʕtal kardan/	‘to wait’	*spoken*
/sabr kardan/	‘to wait’	*both spoken and formal*
/entɛzɑr kardan/	‘to wait’	*formal*

We find in this example that in Afghan Persian as well as the aforementioned duality of the dialect, that register plays an important role in production, as the first example above would be found in conversation, but would be atypical in any form of news media, whereas the last would be found in formal settings, but not in conversation. The middle example might be found in either, with a certain degree of semantic distinction (‘to wait’ vs. ‘to have patience’). A great deal of the Persian lexicon is found to exhibit such variation, and if some analogy between register and the grammaticality judgments in the above twenty-five items can be drawn, a solution to this problem is found.

VII.D.2 /aʕ/ - [ɑ]: Register

If it could be shown that in the intermediate lexical item above, /sabr kardan/, the more colloquial parse permits sound changes, but its formal parse does not, then a register based account would be in order. This is, however, not intended to refute a frequency based account, as a case such as that of "دیگر" /dī-gar/ is undoubtedly based on frequency of parse, but rather to prove that other facets within a use based approach must be considered. I begin by examining one case of the /aʕ/ phenomenon in which, much like "دیگر" /dī-gar/, there is phonological variation between two semantic interpretations of one lexical item. The word under consideration is "لعل" /laʔl/ meaning either 'ruby' or 'diamonds (the suit in cards).' Using the same one to ten rating system, we find the following:

underlying	alt	gloss	judgement
a. /laʕl/→	[lɑl]	'ruby'	7.5 (cf. the above judgment scale)
b. /laʕl/→	[lɑl]	'diamonds'	2.34 (cf. the above judgment scale)

This demonstrates that although the 'ruby' parse is by far more frequent, and more canonical as the meaning of the lexical item, the shift occurs in the less frequent, but more colloquial usage. Or rather 'diamonds,' being said in a specific setting is less frequent, but that setting being more colloquial, allows the alternation to take place. Bybee (2000) notes that, "casual speech among familiars typically shows more reduction. Thus words that are used in casual situations will undergo change at a faster rate." She further notes that these tokens tend to be of a higher overall frequency. (p. 254)

However, where does this leave a theory based on frequency of use, when they are not the most frequent? I postulate that a usage based theory is not off track, but rather one based solely on the notion of frequency is. Frequency is one of several factors affecting sound change, and I contend that a language with two discrete registers will provide a proving ground for the notion of register as another factor independent of frequency.

In order to solidify this intuition, the participants were given a second survey in which they were asked to comment on the register of the same twenty-five lexical items. They were again asked to rate the items from one to five according to the following scale provided:

۱- بسیار اصطلاحی است.	(It is very colloquial)
۲- اصطلاحی است؛ در زبان رسمی کمیاب است.	(It is colloquial and rare in formal)
۳- در زبان اصطلاحی و رسمی استفاده می شود.	(It is found in both)
۴- رسمی است؛ در زبان اصطلاحی کمیاب است.	(It is formal and rare in spoken)
۵- بسیار رسمی است.	(It is very formal)

This time the results were less than satisfying, as the speakers considered all of the items to be more or less common in both settings, colloquial and formal, and no one chose either 1 or 5 for any of them. Clearly colloquial use is not conscious to these speakers. It is telling however, that the two scores of 2 (colloquial) are high on the list and all of the scores of 4 (formal) are low in the list:

underlying		alternation	gloss	judgement	col/for
/baʕd/	→	[bɑd]	‘after’	2.34	3
/maʕnɑ/	→	[mɑnɑ]	‘meaning’	2.34	2
/naʕnɑʕ/	→	[nɑnɑʕ]	‘mint’	2.34	3
/waʕda/	→	[wɑda]	‘promise’	2.67	3
/baʕarusxɑna/	→	[bɑrusxɑna]	‘to the wedding room’	2.67	2
/maʕtal/	→	[mɑtal]	‘waiting’	3	3

/naʕl/	→	[nɑl]	'horseshoe'	3	3
/baʕzi waqt/	→	[bɑzi waqt]	'sometimes'	3.3	3
/maʕlum/	→	[mɑlum]	'effect'	4	3
/maʕmulan/	→	[mɑmulan]	'usually'	5	3
/daʕwat/	→	[dɑwat]	'invitation'	5.34	3
/yaʕnɑ/	→	[yɑnɑ]	'meaning'	5.34	3
/taʕmir/	→	[tɑmir]	'building'	5.67	3
/maʕrfi/	→	[mɑrfi]	'introduced'	6	3
/maʕlɛm/	→	[mɑlɛm]	'teacher'	6.34	3
/ɑbɛ maʕdani/	→	[ɑb mɑdani]	'sparkling water'	6.5	4
/taʕlim/	→	[tɑlim]	'instruction'	7	3
/maʕmul/	→	[mɑmul]	'usual'	7	3
/baʕzan/	→	[bɑzan]	'sometimes'	7.5	3
/maʕdan/	→	[mɑdan]	'mine'	8.17	3
/maʕni/	→	[mɑni]	'meaning'	8.34	4
/taʕqib/	→	[tɑqib]	'chase; pursuit'	8.5	3
/saʕd/	→	[sɑd]	'auspicious'	9.5	4
/waʕz/	→	[wɑz]	'admonition'	9.5	4
/dafaʕtan/	→	[dafɑtan]	'definitely'	10	3

VII.D.3 /aʕ/-[ɑ]: The Lexicon

It is at this point that this chapter must be labeled a work in progress, as all of the informants were educated in their language and hence knew not only correct spellings, but proper usage in formal settings. For this reason, the question is left open for later study with illiterates, which should prove more productive. For the remainder of this paper, I will suggest that each word in which the sound change is significantly acceptable fits into one of two register types: inherently casual environment (such as /laʕl/ 'diamonds'), or possessed of a formal register counterpart (cf. /maʕtal kardan/ vs.

/entɛzɑr kardan/). Furthermore, each word in which the sound change is significantly *un*acceptable fits into one of two register types: inherently formal environment (such as /laʔl/ ‘ruby’), or possessed of a casual register counterpart.

VII.D.3.a /aʔ/-[ɑ]: The Register Counterpart

Upon first glance at the list provided, one pair of register pairs immediately emerges, that of /maʔnɑ/ and /maʔni/ (both signifying ‘meaning’):

UR	judge	freq.	Coll./Form
/maʔnɑ/	2.34	9	2
/maʔni/	8.34	9	4

It becomes apparent that these two lexical items whose frequencies are in fact equal, display enough of an analog between grammaticality judgment of the altered token and the register judgment to be satisfying. There are also cases within the data set of contradictory grammaticality in pairs that seem to be contrastive for frequency.:

UR	judge	freq.	Coll./Form
/baʔzi waqt/	3.3	4	3
/baʔzan/	7.5	0	3

In this case, the speakers’ judgments on the formality of the forms are equal, while the frequency accounts for the acceptability of the change. The above two pairs provide at least a preliminary demonstration of the notion that frequency effects and register effects are not of necessity tied to one another, and thus, the exemplar theory may need some addition.

Obviously, if every lexical item in the language contained this particular conditioning environment, this would be an easier task, however they do not and this register counterpart portion of the proposal will require the addition of more lexical items:

UR	judge	freq.	Formal Counterpart	
/waʕda/	2.67	0	/mɑlɑqat/	'promise; meeting'
/maʕtal/	3	0	/entɛzɑr/	'wait'

UR	judge	freq.	Colloquial Counterpart	
/taʕqib/	8.5	11	/pʊʃt kardan/[78]	'follow, chase'
/dafaʕtan/	10	0	/jak dafaji/	'definitely'

It becomes obvious, by the addition and consideration of register pairs, that the facility with which /waʕda/ and /maʕtal/ become [wɑda] and [mɑtal] may in fact be related to the fact they are each the more colloquial of synonym pairs. Likewise, /taʕqib/ and /dafaʕtan/ are the more formal of synonym pairs, and hence will not be prone to the change. Furthermore, if frequency is to be retained, it may even be noted that /taʕqib/ is more prone to the alternation due to its relative frequency within the formal register portion of the lexicon.

VII.D.3.b /aʕ/-[ɑ]: Register Environment

The remaining lexical items are not seen to possess such counterpart tokens. They are however found in environments conditioning the level of formality with which they are used. For the sake of simplicity, I omit from this analysis the items whose

[78] An interesting case for register as a determinant of change availability, /taʕqib kardan/ is used for chases in an official capacity such as police or fugitive, while /pʊʃt kardan/ is used for the general notion of chasing.

grammaticality judgments range from a score of 4 to that of 7 (from /maʕlum/ to /maʕmul/ above), as there is nothing to be proved using the middle ground in the present analysis. Beginning with the most acceptable items, it is noted that much the same as the case of /laʕl/, which freely permits the change in a card playing environment, these items are most common in casual environments:

UR	judge	freq.	Casual Environment
/naʕnɑʕ/	2.34	0	kitchen environment (‘mint’)
/ba ʕarusxɑna/	2.67	0	family environment (‘to the wedding room’)
/naʕl/	3	0	farm environment (‘horseshoe’)

Similarly, the following items are all found in more formal settings, warranting more formal usage:

UR	judge	freq.	Casual Environment
/saʕd/	9.5	0	a proper name, ‘auspicious’
/waʕz/	9.5	0	religious environment ‘religious advice’

This leaves only two lexical items, /maʕdan/ and /baʕd/. /maʕdan/, I will account for in a similar fashion to “بر” /bar/ above, in claiming that the result of the alternation creates an ambiguity:

UR			judge	freq.
/maʕadan/	→	[mɑdan]	8.17	0

(/mɑdan/ = plural of /maʕdan/)

Thus, the alternant [mɑdan] is deemed incorrect by the speakers due to its interference with the plural /mɑdan/. This is at best stipulative, but indeed a necessary sort of account of retardation of change.

The final change under consideration is in fact that most frequent lexical item in the list, and may perhaps shed the most light on the issue at hand. /baʕd/ is the only item to show a distinct connection between grammaticality judgment and frequency with a score of 2.34 for the pronunciation [bɑd] and a frequency of 26/26,888. Interestingly, unlike the above items, it has no counterpart, colloquial or formal, and the item 'after' will never be seen to be found in any particular environment. Furthermore, the instantiation of [bɑd] creates an item that is identical with both 'wind' and 'be-3rd person singular subjunctive.' I find the data elucidating.

If we assume that there are two registers, colloquial and formal, and two domains frequent and infrequent, it stands to reason that there must be some sort of interplay. Returning to the data from above:

UR	judge	freq.	Formal Counterpart
/waʕda/	2.67	0	/mɑlɑqat/
/maʕtal/	3	0	/entɛzɑr/

UR	judge	freq.	Colloquial Counterpart
/taʕqib/	8.5	11	/pʊʃt kardan/
/dafaʕtan/	10	0	/jak dafaji/

We note that across the four items, register accounts for the sound change data, but within the second pair, frequency will account for the fact that [tɑqib] is somewhat more

acceptable than [dafɑtan]. This will explain why an item such as /maʕmul/ ([mɑmul]) although it has a frequency of 10/26,888 receives a grammaticality judgment of only 7 in the above chart. It is frequent *within* the formal register, and hence ranks better than most formal register lexical items. I contend that /baʕd/ is quite simply frequent in both registers, and hence ranks high in both frequency and grammaticality of change, but is given a medium score in terms of formality. This may account for the fact that it seems to ignore the ambiguity it creates.

VII.E Conclusions

The above data has laid a framework for analysis of gradient language change in terms of register, and crucially related, but not tied to frequency effects. This is of course, not to say that frequency does not play a role in language change, but often more frequent lexical items are either more frequent because they are more colloquial forms, or more colloquial because they are used more frequently, muddling the distinction. This chapter supports the notions of a usage based model, but suggests a multi-dimensionality, which must be considered in order to account for the above data. In order for Bybee's claims to hold, lexical items such as "دیگر" /dīgar/ and "لعل" /laʕl/ that are phonologically distinct in different semantic interpretations, must contain two distinct, but closely related exemplars in order for the change to take place in one and not the other. Fortunately, this is not as far-fetched as it seems, given data such as 'going to' and 'gonna' or the like. Specifically, both register and frequency as independent from one another, must thus be considered independently, and most likely interactively, in light of the above data. The

concept of lexical items bearing their own register independent of usage is a difficult concept to formalize, but the data cannot be overlooked.

VIII. Appendix I: Proto-Persian Verb Stems

As mentioned in section VI.J, Persian has two stems, a present and a past, and these have been traditionally viewed as widely suppletive. This is inadequate and untrue for all but three verbal tokens. An initial caveat regarding verbs is warranted. As with much of the evidence, Dari leads the pack in terms of synchronic change, and Iranian does in terms of diachronic change. That is to say, Dari shows the greatest allophonic variation and synchronic change, but retains the greatest number of original Persian phonemes. This gradient of change holds true for the preservation of intact verb forms. All three dialects, but Dari especially, exhibit an optional compound verb substitute for the original verb form. This typically entails an eventive form of the verbal followed by the conjugated form of one of the following six verbs:

“کردن”	/kar-dan^{l}/	‘to do’
“زدن”	/za-dan^{l}/	‘to strike’
“خوردن”	/xōr-dan^{l}/	‘to take’
“شدن”	/ʃu-dan^{l}/	‘to become’
“دادن”	/dā-dan^{l}/	‘to give’
“گرفتن”	/grif-tan^{l}/	‘to get’

The following illustrates not only the compounding ability, but Dari’s preference for it. Tajik will generally line up with Iranian in this case:

Gloss:	Dari Norm:		Iranian/Tajik Norm:	
‘to speak’	“گپ زدن”	/gapˈ#za-dan/ ‘speaking # to hit’	“صحبت کردن”	/soh-bætˈ#kær-dæn/ ‘to speak’
‘to rotate’	“دور خوردن”	/dawrˈ#xor-dan/ ‘round# to take’	“پیچیدن”	/pi-tʃi-dænˈ/ ‘to turn’
‘to prefer’	“ترجیح دادن”	/tar-ʤiˈ#dɑ-dan/ ‘preference#to give’	“گزیدن”	/go-zi-dænˈ/ ‘to prefer’
‘to decide’	“تصمیم گرفتن”	/tas-mimˈ#grɛf-tan/ ‘decision#to take’	“تسمیم گرفتن”	/tæs-mimˈ#gɛ-rɛf-tæn/ ‘decision#to take’

The majority of compounds are formed with the verb “کردن” /kar-danˈ/ ‘to do,’ and this is decidedly one of the Dari/Iranian Shibboleths:

Gloss:	Dari Norm:		Iranian/Tajik Norm:	
‘to ask’	“پرسان کردن”	/pʊr-sɑnˈ#ka-dan/ ‘asking # to do’	“پرسیدن”	/por-si-dænˈ/ ‘to ask’
‘to write’	“نوشته کردن”	/na-wɛ-ʃtaˈ#ka-dan/ ‘written # to do’	“نوشتن”	/næ-vɛʃ-tænˈ/ ‘to write’
‘to cook’	“پخته کردن”	/pʊx-taˈ#ka-dan/ ‘cooked#to do’	“پختن”	/pox-tænˈ/ ‘to cook’
‘to sleep’	“خواب کردن”	/xawˈ#ka-dan/ ‘sleep#to do’	“خوابیدن”	/xɑ-bi-dænˈ/ ‘to sleep’

This caveat aside, this appendix will simply list non-compound verbs, as the above have recurrent verbs and hence redundant stems. Persian has somewhere between 200 and 300 verbs (although it could be argued that compounding has extracted all but twenty or thirty from common parlance in Dari), and these are widely variant in their stem alternations. For the sake of simplicity, the following is the derivation of the past

stem from the present, and it is satisfying to note that the change will be almost identical to its distant cousin language English:

Present stem + */d/

Example: /kan/ → /kand/ ("کندن" /kan-dan/ 'to pluck; pull out')

"می کنم" /mē + kan + am/ 'I pluck' "کند" /kan + d/ 'he/she/it plucked'

The morpheme */d/ has allomorph [t] when the preceding segment is voiceless:

Example: /kuʃ/ → /kuʃt/ ("کشتن" /kuʃ-tan/ 'to kill')

"می کشم" /mē + kuʃ + am/ 'I kill' "کشت" /kuʃ + t/ 'he/she/it killed'

So already there is a simple division by voicing of the past stem ending, voiced and unvoiced dental. Beyond this, there will be a great number of alternations, both vocalic and consonantal, but for simplicity, it may be divided into six subsets by past stem ending: /-d/, /-ft/, /-st/, /-ʃt/, /-xt/ and suppletive (sections A.I.1-6). Below are the alternations including number of instances of each alternation.[79] In each, the alternation is represented as /present stem ending/ → /past stem ending/ with a parenthetical reference to the number of verbs out of the list exhibiting this class of alternation, and the examples are given as "Persian token in infinitive form" then /present/ → /past/ then 'gloss:'

[79] No two speakers of Persian, regardless of dialect, will agree on how many verbs are truly in use. The selections in this dissertation have been made considering the widest degree of attestation. That is, if only one speaker will produce a certain verb conjugation, but the form is common, it is included. If one only exists in literature and the form is anomalous, it has been omitted.

VIII.A -/d/ Past Stems:

a. ∅ → /īd/ (96)

“آشامیدن”	/ā-ʃām/ →/ ā-ʃā-mīd/	‘to drink’
“آمرزیدن”	/ā-murz/ → /ā-mur-zīd/	‘to forgive’
“ارزیدن”	/arz/ → /ar-zīd/	‘to be worth’
“انجامیدن”	/an-ʤām/ → /an-ʤā-mīd/	‘to finish’
“اندیشیدن”	/an-dīʃ/ → /an-dī-ʃīd/	‘to reflect’
“بالیدن”	/bāl/ → /bālīd/	‘to boast’
“بخشیدن”	/baxʃ-/ → /bax-ʃīd/	‘to forgive; give’
“بریدن”	/bur/ → /bu-rīd/	‘to cut’
“بسیجیدن”	/ba-sīʤ/ → /ba-sī-ʤīd/	‘to intend’
“بلیدن”	/bal/ → / ba-līd/	‘to swallow’
“بوسیدن”	/bus/ → /bu-sīd/	‘to kiss’
“بوییدن”	/bōj/ → /bō-jīd/	‘to smell’
“پاسیدن”	/pās/ → /pā-sīd/	‘to sprinkle’
“پالیدن”	/pāl/ → /pā-līd/	‘to search’
“پرسیدن”	/purs/ → /pur-sīd/	‘to ask’
“پریدن”	/par/ → /pa-rīd/	‘to fly’
“پژوهیدن”	/pa-ʒuh/ → /pa-ʒu-hīd/	‘to search’
“پسندیدن”	/pa-sand/ → /pa-san-dīd/	‘to admire’
“پوسیدن”	/pōs/ → /pō-sīd/	‘to rot’
“پوشیدن”	/pūʃ/ → /pū-ʃīd/	‘to cover; wear’
“پوییدن”	/pūj/ → /pū-jīd/	‘to search’
“پیچیدن”	/pētʃ/ → /pē-ʧīd/	‘to twist’
“تراشیدن”	/ta-rāʃ/ → /ta-rā-ʃīd/	‘to shave’
“ترسیدن”	/tars/ → /tar-sīd/	‘to fear’
“ترشیدن”	/turʃ/ → /tur-ʃīd/	‘to sour’
“ترکیدن”	/tark/ → /tar-kīd/	‘to burst’
“تنیدن”	/tan/ → /ta-nīd/	‘to weave’
“جنگیدن”	/ʤang/ → /ʤan-gīd/	‘to fight’
“جنابیدن”	/ʤa-nāb/ → /ʤa-nā-bīd/	‘to move’

“جوشیدن”	/ʤōʃ/ → /ʤō-ʃīd/	‘to boil’
“جویدن”	/ʤaʊ-/ → /ʤa-ʊīd/	‘to jump’
“چاپیدن”	/ʧāp/ → /ʧā-pīd/	‘to overturn’
“چائیدن”	/ʧāj/ → /ʧā-jīd/	‘to catch a cold’
“چربیدن”	/ʧarb/ → /ʧar-bīd/	‘to increase’
“چریدن”	/ʧar/ → /ʧa-rīd/	‘to graze’
“چسپیدن”	/ʧasp/ → /ʧa-spīd/	‘to stick’
“چشیدن”	/ʧaʃ/ → /ʧa-ʃīd/	‘to taste’
“چکیدن”	/ʧak/ → /ʧa-kīd/	‘to drip’
“چمیدن”	/ʧam/ → /ʧa-mīd/	‘to strut’
“چوشیدن”	/ʧōʃ/ → /ʧō-ʃīd/	‘to suck’
“خاریدن”	/xār/ → /xā-rīd/	‘to itch’
“خراشیدن”	/xa-rāʃ/ → /xa-rā-ʃīd/	‘to scratch’
“خروشیدن”	/xa-rūʃ/ → /xa-rū-ʃīd/	‘to shout’
“خریدن”	/xar/ → /xa-rīd/	‘to buy’
“خزیدن”	/xaz/ → /xa-zīd/	‘to creep’
“خشکیدن”	/xuʃk/ → /xuʃ-kīd/	‘to dry’
“خمیدن”	/xam/ → /xa-mīd/	‘to stoop’
“خندیدن”	/xand-/ → /xan-dīd/	‘to laugh’
“خوابیدن”	/xāb/ → /xā-bīd/	‘to sleep’
“خیسیدن”	/xēs/ → /xē-sīd/	‘to soak’
“درویدن”	/di-raʊ-/ → /di-ra-ʊīd/	‘to reap’
“دریدن”	/dar/ → /da-rīd/	‘to bite; sting’
“دزدیدن”	/duzd/ → /duz-dīd/	‘to rob; steal’
“دویدن”	/daʊ-/ → /da-ʊīd/	‘to run’
“رسیدن”	/ras/ → /ra-sīd/	‘to arrive’
“رقصیدن”	/raqs/ → /raq-sīd/	‘to dance’
“رمبیدن”	/rumb/ → /rum-bīd/	‘to collapse’
“رمیدن”	/ram/ → /ra-mīd/	‘to be startled’
“رنجیدن”	/ranʤ/ → /ran-ʤīd/	‘to be bothered’
“زاریدن”	/zār/ → /zā-rīd/	‘to regret’
“زیبیدن”	/zēb/ → /zē-bīd/	‘to suit’
“ستیزیدن”	/stēz/ → /stē-zīd/	‘to argue’

“سزیدن”	/saz/ → /sa-zīd/	‘to merit’
“سگالیدن”	/sa-gāl/ → /sa-gā-līd/	‘to think’
“سنجیدن”	/sanʤ / → /san-ʤīd/	‘to weigh’
“ششیدن”	/ʃaʃ/ → /ʃa-ʃīd/	‘to urinate’
“شوریدن”	/ʃōr/ → /ʃō-rīd/	‘to riot’
“فهمیدن”	/fahm/ → /fah-mīd/	‘to understand’
“کبیدن”	/kūb/ → /kū-bīd/[80]	‘to beat’
“کشیدن”	/kaʃ/ → /ka-ʃīd/	‘to pull’
“کوچیدن”	/kōʧ/ → /kō-ʧīd/	‘to move’
“کوشیدن”	/kōʃ/ → /kō-ʃīd/	‘to try’
“گائیدن”	/gāj/ → /gā-jīd/	‘to copulate’
“گردیدن”	/gard/ → /gar-dīd/	‘to walk’
“گرویدن”	/gi-raʊ-/ → /gi-ra-ʊīd/	‘to be inclined to’
“گزیدن”	/gaz/ → /ga-zīd/	‘to prefer’
“گنجیدن”	/ganʤ/ → /gan-ʤīd/	‘to be contained’
“گندیدن”	/gand/ → /gan-dīd/	‘to rot’
“گواریدن”	/ga-ʊār/ → /ga-ʊā-rīd/	‘to chew’
“لاییدن”	/lāj/ → /lā-jīd/	‘to bark’
“لرزیدن”	/larz/ → /lar-zīd/	‘to tremble’
“لغزیدن”	/laɣz/ → /laɣ-zīd/	‘to trip’
“لمیدن”	/lam/ → /la-mīd/	‘to roll’
“لنگیدن”	/lang/ → /lan-gīd/	‘to limp’
“لولیدن”	/lōl/ → /lō-līd/	‘to gyrate’
“لیسیدن”	/līs/ → /lī-sīd/	‘to lick’
“ماستیدن”	/māst/ → /mās-tīd/	‘to congeal’
“مالیدن”	/māl/ → /mā-līd/	‘to rub’
“نازیدن”	/nāz/ → /nā-zīd/	‘to boast’
“نالیدن”	/nāl/ → /nā-līd/	‘to moan’
“نامیدن”	/nām/ → /nā-mīd/	‘to name’
“نشمیدن”	/naʃm/ → /naʃ-mīd/	‘to eat’
“نوشیدن”	/nūʃ/ → /nū-ʃīd/	‘to drink’

[80] This verb has an alternate (presumably original) form as /kūf-tan/, but has adopted /kū-bī-dan/ as analogous with the copious /-īd/ form verbs.

“نیوشیدن”	/na-jūʃ/ → /na-jū-ʃīd/	‘to hear’
“وزیدن”	/ʋaz/ → /ʋa-zīd/	‘to blow’
“هراسیدن”	/ha-rās/ → /ha-rā-sīd/	‘to fear’

For three reasons, the /īd/ derivation of the past stem, A.I.1.a, is assumed to be the canonical past tense stem derivative. First, it accounts for over a third of the verbs in Persian. Second, it appears that verbs formed from normative advectivals and nominals are typically formed this way:

“نام” /nɑm/ ‘name’ → “نامیدن” /nɑ-mī-danˈ/ ‘to name’

“ترش” /turʃ/ ‘sour’ → “ترشیدن” /tur-ʃī-danˈ/ ‘to sour’

Third, a number of other verbs have alternate forms that are analogous reanalyses as /īd/ stem verbs:

“کفتن” /kūb/ → /kūft/ ‘to beat’ → “کبیدن” /kūb/ → /kū-bīd/ ‘to beat’

“سودن” /sā/ → /sūd/ ‘to rest’ → “سائیدن” /sā/ → /sā-jīd/ ‘to rest’

Thus, such analogous tokens have been eliminated from the /īd/ section, but no explanation will be warranted as the following will be regarded as verb class 1: /stem/ + /ī/ (derivational infix) + /d/

b. /ā/ → /ūd/ (13)

“آلودن”	/ā-lā-/ → /ā-lūd/	‘to pollute’
“افزودن”	/af-zā/ → /af-zūd/	‘to add; increase’
“اندودن”	/an-dā/ → /an-dūd/	‘to smear’

“آزمودن”	/āz-mā/ → /āz-mūd/	‘to test’
“فرمودن”	/far-mā/ → /far-mūd/	‘to command’
“نمودن”	/na-mā/ → /na-mūd/	‘to seem’
“پیمودن”	/paj-mā/ → /paj-mūd/	‘to measure’
“پالودن”	/pā-lā/ → /pā-lūd/	‘to filter’
“ربودن”	/ru-bā/ → /ru-būd/	‘to rob’
“ستودن”	/stā/ → /stūd/ (orig. /sdā/)	‘to praise’
“سرودن”	/srā/ → /srūd/	‘to sing’
“زدودن”	/zu-dā/ → /zu-dūd/	‘to polish’
“سودن”	/sā/ → /sūd/	‘to rest’
“اندودن”	/an-dā/ → /an-dūd/	‘to smear’
“زدودن”	/zu-dā/ → /zu-dūd/	‘to polish’
“پالودن”	/pā-lā/ → /pā-lūd/	‘to filter’
“آزمودن”	/āz-mā/ → /āz-mūd/	‘to test’
“فرمودن”	/far-mā/ → /far-mūd/	‘to command’
“نمودن”	/na-mā/ → /na-mūd/	‘to seem’
“پیمودن”	/paj-mā/ → /paj-mūd/	‘to measure’

c. → /ād/ (8)

i. ∅ → /ād/ (4)

ii. ∅ → /d/ / /ā/ + __ (3)

iii. /ahd/ → /ād/ / Compensatory Lengthening (1)

c.i	“فرستادن”	/frist-/ → /fri-stād/	‘to send’
	“افتادن”	/uft/ → /uf-tād/	‘to fall’
	“ایستادن”	/īst/ → /ī-stād/	‘to stand’
	“نهادن”	/nah/ → /na-hād/	‘to place; store’
c.ii	“زادن”	/zā-/ → /zād/	‘to be born’
	“گشادن”	/gu-ʃā-/ → /gu-ʃād/	‘to release’
	“ستادن”	/stā/ → /stād/	‘to take’

c.iii	“دادن”	/dah-/ → /dād/	‘to give’

d. /n/ → /d/ (5)

“آفریدن”	/ā-fa-rīn-/ → /ā-fa-rīd/	‘to create’
“چیدن”	/ʧīn-/ → /ʧīd/	‘to gather; to cut’
“گزیدن”	/gu-zīn-/ → /gu-zīd/	‘to prefer’
“ریدن”	/rīn-/ → /rīd/	‘to defecate’
“زدن”	/zan-/ → /zad/	‘to strike; hit’

e. ∅ → /d/ / /n/ + __ (7)

“کندن”	/kan-/ → /kand/	‘to pluck’
“افکندن”	/af-kan/ → /af-kand/	‘to throw’
“ماندن”	/mān/ → /mānd/	‘to remain’
“راندن”	/rān/ → /rānd/	‘to drive’
“خواندن”	/xān/ → /xānd/	‘to read’
“افشاندن”	/af-ʃɑn/ → /af-ʃɑnd/	‘to scatter; to sprinkle’
“آگندن”	/ā-gan/ → /ā-gand/	‘to fill; to stuff’

f. ∅ → /d/ / /r/ + __ (14)

i. ∅ → /d/ / /r/ + __ (7)

ii. ∅ → /d/ / /r/ + __ (with stem vowel shift /ā/ → /u/) (5)

iii. ∅ → /d/ / /r/ + __ (with stem vowel shift /ī/ → /u/) (1)

iv. ∅ → /d/ / /r/ + __ (with stem vowel shift /a/ → /u/) (1)

f.i	“آوردن”	/ā-ʊar-/ → /ā-ʊard/	‘to bring’
	“خوردن”	/xōr/ → /xōrd/	‘to eat; take’
	“پروردن”	/par-ʊar/ → /par-ʊard/	‘to foster; to rear’
	“افسردن”	/af-sar/ → /af-sard/	‘to freeze’
	“گساردن”	/gu-sār/ → /gu-sārd/	‘to consume; to drink’
	“گستردن”	/ga-stur/ → /ga-sturd/	‘to spread’

f.ii	“آزردن”	/ā-zār-/ → /ā-zurd/	‘to annoy’
	“سپردن”	/spār-/ → /spurd/	‘to grant’
	“افشردن”	/af-ʃār/ →/af-ʃurd/	‘to press’
	“فشاردن”	/fi-ʃār[1]/ → /fi-ʃurd/	‘to squeeze’
	“شمردن”	/ʃmār/ → /ʃmurd/	‘to count’
f.iii	“مردن”	/mīr-/ → /murd/	‘to die’
f.iv	“بردن”	/bar-/ → /burd/	‘to carry’

<u>VIII.B -/ft/ Past Stems:</u>

a. ∅ → /ft/ (2)

“بافتن”	/bāf-/ → /bāft/	‘to weave’
“شکافتن”	/ʃkāf-/ → /ʃkāft/	‘to split’

b. /b/ → /ft/ (7)

i. /b/ → /ft/ (5)

ii. /b/ → /ft/ (with stem vowel shift /V:/ → /V/) (2)

b.i	“فریفتن”	/frēb-/ → /frēft/	‘to decieve’
	“یافتن”	/jāb/ → /jāft/	‘to find’
	“کوفتن”	/kūb/ → /kūft/	‘to beat; to pound’
	“شتافتن”	/ʃtāb/ → /ʃtāft/	‘to hurry’
	“تافتن”	/tāb/ → /tāft/	‘to twist’
b.ii	“آشفتن”	/ā-ʃūb-/ → /ā-ʃuft/	‘to be agitated; to rebel’
	“رفتن”	/rūb-/ → /ruft/	‘to sweep’

c. /υ/ → /ft/ (10)

i. /υ/ → /ft/ (3)

ii. /υ/ → /ft/ (1)

iii. /υ/ → /ft/ (1)
iv. /υ/ → /ft/ (1)
v. /υ/ → /ft/ (2)
vi. ∅ → /ft/ (1)
vii. ∅ → /ft/ (with stem vowel ellipsis and epenthesis of /i/) (1)

c.i	"رفتن"	/raυ-/ → /raft/	'to go'
	"کافتن"	/kāυ-/ → /kāft/	'to dig'
	"شیفتن"	/ʃēυ/ → /ʃēft/	'to enamor'
c.ii	"شنیدن"	/ʃnaυ-/ → /ʃnīd/	'to listen'
c.iii	"شدن"	/ʃaυ-/ → /ʃud/	'to become'
c.iv	"پذیرفتن"	/pa-zīr-/ → /pa-zī-ruft/	'to accept'
c.v	"نهفتن"	/na-hunb-/ → /na-huft/	'to hide'
	"سفتن"	/sunb-/ → /suft/	'to pierce'
c.vi	"گفتن"	/gū-/ → /guft/	'to say'
c.vii	"گرفتن"	/gīr-/ → /grift/	'to take'

<u>VIII.C -/st/ Past Stems:</u>

a. ∅ → /st/ (12)
i. ∅ → /st/ (3)
ii. ∅ → /ist/ (4)
ii_1 ∅ → /ist/ (with vowel shift /V:/ → /V/ or /V/ → /V:/) (2)
iii. /ō/ → /ust/ (3)

a.i	"گرستن"	/gi-ri-/ → /gi-rist/	'to cry'
	"آراستن"	/ā-rā/ → /ā-rāst/	'to adorn'

	“زیستن”	/zī/ → /zīst/	‘to live’
a.ii	“دانستن”	/dān-/ → /dā-nist/	‘to know; consider’
	“مانستن”	/mān/ → /mā-nist/	‘to appear’
	“توانستن”	/ta-ʋān/ → /ta-ʋā-nist/	‘to be able’
	“بائیستن”	/bāj/ → /bā-jist/	‘to be necessary’
a.ii$_1$	“نگریستن”	/ni-gar-/ → /ni-ga-rīst/	‘to look’
	“یارستن”	/jār-/ → /jā-rast/	‘to have the courage’
a.iii	“شستن”	/ʃō-/ → /ʃust/	‘to wash’
	“جستن”	/ʤō/ → /ʤust/	‘to search’
	“رستن”	/rō/ → /rust/	‘to grow’

b. /h/ → /st/ (4)

“جستن”	/ʤah-/ → /ʤast/	‘to jump’
“کاستن”	/kāh-/ → /kāst/	‘to decrease’
“رستن”	/rah-/ → /rast/	‘to escape’
“خواستن”	/xʋāh-/ → /xʋāst/	‘to want’

c. /z/ → /st/ (1)

“خاستن”	/xēz-/ → /xāst/	‘to arise’

d. /n/ → /st/ (4)

i. /n/ → /st/ (2)

ii. /nd/ → /st/ (2)

d.i	“نشستن”	/ni-ʃīn-/ → /ni-ʃast/	‘to sit’
	“شکستن”	/ʃkan-/ → /ʃkast/	‘to break’
d.ii	“بستن”	/band-/ → /bast/	‘to tie’

“پیوستن”	/paj-ʊand-/ → /paj-ʊast/	‘to connect’

VIII.D -/ʃt/ Past Stems:

a. /ʃ/ → /ʃt/ (2)

“کشتن”	/kuʃ-/ → /kuʃt/	‘to kill’
“سرشتن”	/sriʃ-/ → /sriʃt/	‘to knead; to mold’

b. /r/ → /ʃt/ (11) (/ā/ __)

i. /r/ → /ʃt/ (10)

ii. /rd/ → /ʃt/ (1)

iii. /r/ → /ʃt/ (with vowel shift /V/ → /V:/) (1)

a.i	“داشتن”	/dār-/ → /dāʃt/	‘to have’
	“انباشتن”	/an-bār/ → /an-bāʃt/	‘to store up’
	“گماشتن”	/gu-mār/ → /gu-māʃt/	‘to appoint’
	“گذاشتن”	/gu-zār/ → /gu-zāʃt/	‘to leave; to let’
	“گذشتن”	/gu-zar/ → /gu-zaʃt/	‘to pass’
	“کاشتن”	/kār/ → /kāʃt/	‘sow; to plant’
	“نگاشتن”	/ni-gār/ → /ni-gāʃt/	‘to draw’
	“پنداشتن”	/pin-dār/ → /pin-dāʃt/	‘to imagine’
	“انگاشتن”	/an-gār/ → /an-gāʃt/	‘to suppose’
a.ii	“گشتن”	/gard-/ → /gaʃt/	‘to turn’
a.iii	“آغشتن”	/ā-ɣār/ → /ā-ɣaʃt/	‘to chew’

c. /s/ → /ʃt/ (with stem vowel shift /ī/ → /i/) (2)

“نوشتن”	/na-ʊīs-/ → /na-ʊiʃt/	‘to write’
“رشتن”	/rīs-/ → /riʃt/	‘to spin’

d. /l/ → /ʃt/ (1)

“هشتن”	/hil/ → /haʃt/	‘to let’

VIII.E -/xt/ Past Stems:

a. /z/ → /xt/ (23)

i. /z/ → /xt/ (21)

ii. /z/ → /xt/ (with stem vowel shift /a/ → /u/) (2)

a.i	“ریختن”	/rēz-/ → /rēxt/	‘to pour’
	“افروختن”	/af-rūz/ → /af-rūxt/	‘to kindle; to burn’
	“افراختن”	/a-frāz-/ → /a-frāxt/	‘to elevate’
	“آمیختن”	/ā-mēz/ → /ā-mēxt/	‘to mix’
	“آموختن”	/ā-mūz/ → /ā-mūxt/	‘to learn’
	“انداختن”	/an-dāz/ → /an-dāxt/	‘to throw’
	“اندوختن”	/an-dūz/ → /an-dūxt/	‘to accumulate’
	“انگیختن”	/an-gēz/ → /an-gēxt/	‘to incite’
	“آویختن”	/ā-ʋēz/ → /ā-ʋēxt/	‘to hang’
	“باختن”	/bāz/ → /bāxt/	‘to lose’
	“بیختن”	/bēz/ → /bēxt/	‘to sift’
	“دوختن”	/dōz/ → /dōxt/	‘to sew’
	“گریختن”	/grēz/ → /grēxt/	‘to flee’
	“گداختن”	/gu-dāz/ → /gu-dāxt/	‘to melt’
	“نواختن”	/na-ʋāz/ → /na-ʋāxt/	‘to caress’
	“پرداختن”	/par-dāz/ → /par-dāxt/	‘to pay’
	“ساختن”	/sāz/ → /sāxt/	‘to create’
	“سوختن”	/sūz/ → /sūxt/	‘to burn’
	“تاختن”	/tāz/ → /tāxt/	‘to assault’
	“آختن”	/āz/ → /āxt/	‘to draw’
	“پرهیختن”	/par-hēz/ → /par-hēxt/	‘to obtain’
a.ii	“پختن”	/paz-/ → /puxt/	‘to cook’
	“گسیختن”	/gu-sal-/ → /gu-sēxt/	‘to break’

b. /ʃ/ → /xt/ (2)

"فروختن"	/frūʃ-/ → /frūxt/	'to sell'
"دوختن"	/dōʃ/ → /dōxt/	'to milk'

c. /s/ → /xt/ (with stem vowel shift /V:/ → /V/) (1)

"شناختن"	/ʃnās-/ → /ʃnaxt/	'to know'

<u>VIII.F Suppletive Forms:</u> (4)

"دیدن"	/bīn-/ → /dīd/	'to see'
"کردن"	/kun-/ → /kard/	'to do'
"بودن"	/bāʃ-/ → /būd/	'to be'
"آمدن"	/ā-/ → /ā-mad/	'to come'

IX. Appendix II: Field Methodology

IX.A Questions for Minimal Pair Ellicitation and Segment Clarification

Below are the short answer questions I provided to the participants (English translations follow each question in italics):

۱- فرق بین ملا، امام، مولوی، مولانا، و آیت لله این است که _____...

1. The difference between mullah, Imam, Mawlawi, Mawlana, and Ayatollah is...

۲- در هندوستان معمولاً سن نامزاد شدن _____...

2. In India the marrying age is...

۳- "پی" چند تا معنا دارد، مثلاً _____... (در پی کسی یا چربی)

3. The word spelled “PI” has several meanings, such as...

۴- لطفاً با کلمهٔ "طی" یک جمله بسازید.

4. Please make a sentence with the word “during.”

۵- "شیر" هم دو معنا دارد، مثلاً _____ ...

5. The word spelled “شیر” has two meanings, and they are...

۶- لطفاً این جمله را بخوانید (گپ زدنی): "من از ترشی سیر سیر شدم."

6. Please read the following sentence: “I have had my fill of pickled garlic.”

۷- لطفاً با هر دو لغات "کی" سوأل بسازید. (چه وقت و چه کسی)

7. Ask a question using both of the meanings of the word spelled “KI.”

۸- کسی که _____ می نویسد شاعر است.

8. A person who writes poetry is a...

۹- معلمان درس _____، و شاگردان درس _____.

9. A teacher _____, and a student _____ .

۱۰- چند لغت به یاد دارید که با "سر" هم قافیه اند؟

10. List as many words as you can think of that rhyme with [sar].

۱۱- من هر روز ساعت 7 صبح از خانه _____، و ساعت ۵ شام به خانه _____.

11. I ____ home every day at 7, and ____ home at five.

۱۲- رئیس جمهور گواتیمالا کیست؟

12. Who is the president of Guatemala?

۱۳- کسی که در باغ کار میکند _____ است.

13. One who works in the garden is a...

۱۴- ۳ + ۵ = _____.

14. 5 + 3 =

۱۵- ۲ + ۴ = _____.

15. 4 + 2 =

۱۶- ۵ + ۷ = _____.

16. 7 + 5 =

۱۷- ۹ + ۱ = _____.

17. 1 + 9 =

۱۸- ۱۷ + ۱ = _____.

18. 1 + 17 =

۱۹- برای هر ماه، لطفاً یک تاریخ بگویید که برای شما مهم است (مثلاً سالگرد برادر تان) یا تاریخی که رخصتیست (مثلاً نو روز).

19. For each month give a date that is significant for you or is a holiday.

۲۰- لطفاً به چیزهای که در داخل این تعمیر وجود دارند با استفاده از رنگها جملات بسازید. (مثلاً: "آن کتاب سفید است.")

20. Please describe the objects in this room/area using colors.

۲۱- حالا با استفادهٔ صفت های دیگر به چیزهای که در این تعمیر وجود دارند جملات بسازید. ("این میز کوچک است.")

21. Please describe the same objects using adjectives.

۲۲- حالا با شماره ها ("مثلاً: در این اطاق 9 چراغ استند."

22. Now use numbers.

۲۳- از هر حیوان که به یاد تان بیآید لطفاً نام ببرید.

23. List every animal you can think of.

۲۴- در مورد هر اعضا خانوادهٔ تان که می خواهید بگویید که چند تا دارید و نام آن عضو چیست (مثلاً "من چهار برادر دارم و نامهای شان ... استند." می توانید جعل کنید.).

24. Please tell how many of each family member you have and their names (you can invent them if you like).

۲۵- کلمهٔ "تو" دو معنا دارد. لطفاً هر دو را در یک جمله استفاده کنید.

25. The word spelled “TU” has two meanings, please make a sentence with both.

۲۶- لطفاً لغات "آسمان" و "آفتاب" را در یک جمله استفاده کنید.

26. Please make a sentence with “sun” and “sky.”

۲۷- کلمهٔ دیگری برای "راهٔ آهن" _____ است.

27. Another word for “railroad” is ____.

۲۸- لطفاً لغت "داوطلبانه" را معنی کنید.

28. Please define the word “voluntary.”

۲۹- من از ولایت (استان) _____ استم.

29. I am from _____ province.

۳۰- نام های والدین من _____ استند. (نباید راستی باشد.)

30. My parents' names are ____ (You can invent if you want.)

۳۱- "شور" دو معنا دارد (مثلاً شورش و بانمک). لطفاً هر دو را در یک جمله استفاده کنید.

31. The word spelled "SHUR" has two meanings, please make a sentence with both.

۳۲- "تور" هم دو معنا دارد (مثلاً سفر و افزار ماهی گیری). لطفاً هر دو را در یک جمله استفاده کنید.

32. The word spelled "TUR" has two meanings, please make a sentence with both.

۳۳- "حکومتی" چه معنی دارد؟ لغت دیگری برای اش وجود دارد؟

33. Describe the meaning of "governmental." Are there other words for this.

۳۴- "دور" دو معنا دارد (مثلاً نزدیک نه و دوران). لطفاً هر دو را در یک جمله استفاده کنید.

34. The word spelled "DUR" has two meanings, please make a sentence with both.

۳۵- لطفاً "پول" و "پل" را در یک جمله استفاده کنید.

35. Please make a sentence with the words "bridge" and "money."

۳۶- لطفاً با کلمات "وقت" و "اوقات" یک جمله بسازید.

36. Please make a sentence with the words "time" and "times."

۳۷- یک لغت دیگری برای "قهر" وجود دارد؟ چه چیزهای شما را قهر میسازند و چرا؟

37. Is there another word for "angry?" What makes you angry?

۳۸- من فکر میکنم که "قندهار / کندهار" تلفظ درست است، چون که _____...

38. Do you think that the city name should be "Qandahar" or "Kandahar," and why?

۳۹- خانهٔ تان چند تا اطاق دارد؟

39. How many rooms are in your house?

۴۰- "عقرب" چیست؟

40. What is the meaning of “AQRAB” (Scorpio)?

۴۱- "به چشم" چه معنا دارد، و چرا میگویند؟

41. What is the meaning of the phrase “BA CHASHM?”

۴۲- چه وقت "شکر" میگویید، و چرا؟

42. When do you use the phrase “SHUKR?”

۴۳- فرق بین "سعی کردن" و "کوشش کردن" چیست؟ لطفاً هر دو را در جملات استفاده کنید.

43. What is the difference between “try” / “attempt?” Please use them in sentences.

۴۴- فرق بین "تور" و "طور" چیست؟

44. What is the difference between “TUR” and “TUR?” (different Arabic spelling)

۴۵- کلمهٔ دیگری برای "دریا" _____ است.

45. Is there another word for “DARYAA?”

۴۶- دیروز چه کار کردید، امروز چه کار می کنید، و فردا چه کار خواهید کرد؟ بعد، لطفاً برای روز هفته، بگویید که چه کار می کنید (نباید راستی باشد.)

46. What did you do yesterday, today, and tomorrow, and for every day of the week, what do you do?

۴۷- فرق بین لغات "بینا" و "بنا" چیست؟ لطفاً در یک جمله استفاده شان کنید.

47. What is the difference between “BINA” and “BANA?” Use them in sentences.

۴۸- ضد "نو" چیست؟

48. What is the opposite of “new?”

۴۹- "دل" یک چند استفاده اصطلاحی دارد. لطفاً هر استففاده که به یاد تان می یاید بگویید.

49. What are some uses of the word “DIL?” Use it in some sentences.

۵۰- فرق بین "کمک" و "امداد" چیست؟

50. What is the difference between “help” and “aid?”

۵۱- لطفاً "سر" (مثل سر تا پا)،"سر" (مثل راز)، و"سر" (مثل لیز خوردن) را در جملات استفاده کنید.

51. Please use “SAR,” “SER,” and “SUR” in a sentence.

۵۲- "عهده" چی معنا دارد؟

52. What is the meaning of “promise.”

۵۳- لغت "محترم" در کدام حالت استفاده می کنید؟

53. When do you use the word “respectfully?”

۵۴- در وطن شما، دوختن چه اهمیت دارد؟ لطفاً تشریح کنید.

54. What is the significance of sewing in your homeland?

IX.B Questions for Narrative Ellicitation

Below are the discussion questions I provided to the participants (English translations follow each question in italics):

1- گاهی در زنده گی تان در حالتی بودید که احساس می کردید که به مرگ نزیک استید و یا اینکه یکی از عزیزان تان در بند بود و یا به مرگ نزیک بود و شما خود را بیچاره احساس کردید؟ اگر با چنین حالتی مواجه شده اید لطفاً در باره اش حرف بزنید.

1. Were you ever in a situation in which you felt that you were going to die? Has anyone close to you ever been arrested or in danger of death?

2- چرا از وطن تان به جای دیگر رفتید؟ آیا این یک تصمیم خوب بود یا نه؟ پیش از آمدن به امریکا به کدام جاها رفتید؟

2. Why did you move from your home country? Was it a good decision? Where did you move before coming to the US?

3- (افغانها و تاجیکها مخصوصاً) بسیار رژیم ها در سال های 1300 در منطقه های پارسی زبان مداخله کرده اند (مثلاً اتحاد شوروی، طالبان، پاکستان، ایالات متحده و کشورهای غربی دیگر، و غیره). به نظر شما، چرا؟ واین مداخلات چه تأثیر داشته است (و خواهد داشت)؟ لطفاً در مورد هر یکی از این گروها حرف بزنید:

- اتحاد شوروی
- طالبان
- پاکستان
- ایالات متحده و کشورهای غربی
- دیگران

3. (For Afghans and Tajiks) in the last century, many entities have come to power and/or interfered in the Persian-speaking area of the world (i.e. USSR, Taliban, Pakistan, USA, other western countries, etc.). In your opinion, why? What effect has their interference had (and will continue to have)? Talk about the following:
- USSR
- Taliban
- Pakistan
- USA and other western countries
- others

۴- (ایرانیها مخصوصاً) در تاریخ قرن بیستم میلادی در ایران بسیار تغییرات سیاسی، جمعیتی، فرهنگی (و غیره) به وجود آمد. به نظر شما، شخصیتهای ذیل در منطقه چه اثر داشته اند:
- رظا شاه پهلوی
- امام خمینی
- آیت الله خامنه ئی
- رئیس جمهور سابق خاتمی
- محمود احمدی نژاد

4. (For Iranians) In the 20th century, many political, societal, cultural, and other changes have come into Iran. In your opinion, what effect have the following players had?:
- Reza Shah Pahlavi
- Imam Khomeini
- Ayatollah Khameini
- Ex-President Khatami
- President Mahmoud Ahmadi-Nejad

۵- (تاجیکها مخصوصاً) در تاریخ قرن بیستم میلادی در تاجیکستان، به خاطر تغییرات در رهبری سیاسی، مردم بسیار رنج دید. به نظر شما، شخصیتهای ذیل در منطقه چه اثر داشته اند، و اگر با آنها می توانستید حرف بزنید، چه می گویید:
- رهبران شوروی و اعضاء مجلس عالی
- کهر محکموف
- رحمان نبییف
- صعید عبدلله نوری و مقاومت متحده تاجیکستان
- نهضت اسلامی تاجیکستان
- امام علی رحمنوف

5. (For Tajiks) Tajikistan has been through much suffering in the 20th century due to changes in the political leadership. In your opinion, what effects have the following people had on the region, and what would you say to them if you had the chance?:
- The Soviet leaders and members of the General Council
- Kakhar Makhkamov
- Rahmon Nabiyev
- Said Abdullah Nuri, leader of the Tajikistan United Resistance
- Islamic Front of Tajikistan

- President Emomali Rahmanov

۶- اگر بین جورج بوش و محمود احمدی نژاد یک ملاقات صورت می گرفت، بالای حالت فعلی چه تاثیر می داشت و چی تغیرات را به بار می آورد؟

6. If George Bush and Mahmoud Ahmadi-Nejad held a meeting, what effect would it have on the current situation, and what changes might it bring about?

7- اگر با رئیس جمهور (احمدی نژاد، کرزی، یا رحمانوف) حرف می زدید، به ایشان چه می گفتید و چرا؟

7. If you could sit down and talk with the president of your homeland (Hamed Karzai, Mahmoud Ahmadi-Nejad, or Emomali Rahmanov) what would you say?

8- لغت "ناموس" چه معنا دارد؟ لطفاً اهمیت این لغت را تشریح کنید.

8. What is the meaning of "naamus" in your culture? Please explain.

9- یکی از شرکت کنندگان این پژوهس در میان مصاحبهٔ اش گفته است که "یکی از عادت های ایالات متحده است که همیشه یک کشور را به ضد دیگر استفاده می کند." با این عقیده موافق استید یا نه، و چرا؟

9. During an interview, one of the participants in this project said: "One of the things that the US does is constantly pit one country against another." Do you agree with this statement? Why or why not?

10- در نظر شما لهجه های ایرانی، افغانی، و تاجیکی چه فرق دارند؟ لهجه های دیگر چطور صدا می دهند؟

10. In your opinion, what is the difference between the dialects of Iran, Afghanistan and Tajikistan? What do the other dialects sound like to you?

11- کسانی که از وطن خود به یک کشور مثل آمریکا می روند، معمولاً پیش از رسیدن و بعد از رسیدن، گمان شان تغییر می دهد. پیش از آمدن به آمریکا چه فکر کردین،و از کدام نظر حقیقت از آین فکر مختلف است؟

11. When a person moves from their native land to another country like America, often their attitude about the country changes. What did you expect before coming, and in what ways did your experience differ from those expectations?

X. Bibliography

X.A Historical Persian and Dialectology

Beekes, R. S. P. (1997). Historical phonology of Iranian. *The Journal of Indo-European Studies, 25*(1,2), 1-26.

Chambers, J. K., Trudgill, P., & Schilling-Estes, N. (Eds.). (2002). *The handbook of language variation and change.* Oxford: Blackwell.

Cheung, J. (2007). *Etymological dictionary of the Iranian verb.* Leiden: Brill.

Christensen, A. (1930-1935). *Contributions à la dialectologie Iranienne, I, II*. Copenhagen: Historisk Filosofiske/filologiske Meddelelser.

Darling-Buck, C. (1949). *A dictionary of selected synonyms in the principle Indo-European languages*. Chicago: University of Chicago Press.

Edelman, D. I. (1995). Ещё раз о взаимодействии языковых уровней в истории иранских языков[81]. *Voprosy Jazykoznannja, 4(*July/April), 25-42.

Insaf'pur, G.R. (1994). *کمال فرهنگ فارسی (یا فرهنگ دری برای ملل ایرانی، تاجیک، افغان، خوارزم، فارسی زبانان از بنگال و هند و پاکستان قفقاز و آسیای غربی کرد.)* Tehran: Zawar. [82]

Gray, L.H. (1965). *Indo-Iranian phonology with special reference to the middle and new Indo-Iranian languages*. New York: AMS Press.

Ivanov, V. B. (1995). Границы слова и инкапсуляция в персидском, таджикском и дари.[83] *Voprosy Jazykoznannja*, *4*(July/April), 42-60.

[81] Russian: 'Jeshchyo Raz o Vzaimodjejstvii Yazykovykh Urovnjej v Istorii Iranskikh Jazykov ' "Once More on the Interaction of Linguistic Levels in the History of Iranian Languages."

[82] Farsi: ' Kamol Farhang-e Farsi (ya Farhang-e Dari baraye Milal-e Irani, Tajik, Afghan, Khwarzm, Farsi-zabanan az Bangal va Hend va Pakistan ta Qavqaz va Asia-ye Gharbi-ye Kurd)'"Complete Persian Dictionary (Or Dictionary of Dari for Iranians, Tajiks, Afghans, Khorazams, Farsi Speakers of Bengal, Pakistan, and up to the Caucuses and Western Kurdish Asia)."

Kent, R.G. (1953). *Old Persian: Grammar, texts, lexicon*. New Haven: American Oriental Society.

Lazard, G. (1957). La formation de la langue Persane. *Travaux et Mémoires de l'Institut d'Études Iraniennes*. Paris: Peeters.

MacKenzie, D. N. (1971). *A concise Pahlavi dictionary*. London: Oxford University Press.

Matini, J. (1994). یک زبان و سه نام![84] In M. Marashi (Ed.), *Persian studies in North America: Studies in honor of Mohammad Ali Jazayery* (pp. 9-21). Bethesda: Iranbooks.

Oranskij, I.M. (1998). *زبانهای ایران*[85] Tehran: Sukhan.

Pisowicz, A. (1985). *Origins of the new and middle Persian phonological systems*. Krako'w: Nakladem Uniwersytetu Jagiellon'skiego.

Sayah, H. (1980). *Etymology of the Indo-European languages including the New Iranian: with some interrelationship between this and other language families due to kinship or borrowing*. Tehran: S. N.

Shakibi Gilani, J. (1988). *A comparative probe in the Iranian semi-languages and dialects*. Nashville, TN: Babylonia Language and Translation Center.

Shakibi Gilani, J. (1989). *Shenosnameye vazhegane Parsi: A concise etymological dictionary of the Persian language*. Nashville: Shakibi Gilani.

---

[83] Russian: 'Granicy Slova i Inkapsuljacija v Pjersidskom, Tadzhikskom, i Dari'"Word Boundaries and Encapsulation in Persian, Tajik, and Dari."
[84] Persian: 'Yek Zabaan veh Seh Naam!' "One Language and Three Names!"
[85] Persian: '*Zabanhaye Irani*'"Iranian Languages"

Shezha'i, M. (2000). Articles on research in the languages and literatures of Iran, Afghanistan, and Tajikistan. *Mujallah-i Zabanshinasi*, *15*(1), 79-94.

Shukla, S. Extensive notes and collected data. Georgetown University, Washington, DC.

Sokolova, V. S. (1953). *Очерки по Фонетике Иранских Языков.*[86] Moscow-Leningrad: Izdatjel'stvo Akadjemiji Nauk USSR.

Spooner, B. (1994). Are we teaching Persian? or Farsi? or Dari? or Tojiki? In M. Marashi (Ed.), *Persian studies in North America: Studies in honor of Mohammad Ali Jazayery* (pp. 175-190). Bethesda: Iranbooks.

Stilo, D.L. (1994). Phonological systems in contact in Iran and Transcaucasia. In M. Marashi (Ed.), *Persian studies in North America: Studies in honor of Mohammad Ali Jazayery* (pp. 75-94). Bethesda: Iranbooks.

Testen, D. (1997). Old Persian and Avestan phonology. In A. Kaye (Ed.), *Phonologies of Asia and Africa* (Vol. 2) (pp. 569-600). Winona Lake, IN: Eisenbrauns.

Weber, D. (1997). Pahlavi phonology. In A. Kaye (Ed.), *Phonologies of Asia and Africa* (Vol. 2) (pp. 601-636). Winona Lake, IN: Eisenbrauns.

Wei, J. (1962). *Dialectal differences between three standard varieties of Persian: Tehran, Kabul, and Tajik.* Washington, DC: Center for Applied Linguistics.

X.B Kabuli Persian

Bogdanov, L. (1930). Stray notes on Kabuli-Persian. *Journal and Proceedings of the Asiatic Society of Bengal*, *26*, 1-23.

[86] Russian: '*Ochjerki po Fonjekikje Iranskikh Jazykov*.' "Studies on the Phonetics of Iranian Languages."

Dorofeeva, L. N. (1959). *О Языке 'Фарсии' Афганистана*.[87] Moscow-Leningrad: Kratkje Soobshchjenija Instituta Vostokovjedjeija A. N. USSR.

Dorofeeva, L.N. (1960). *Язык Фарсии-Кабули*.[88] Moscow-Leningrad: Izdatjel'stvo Vostochnoj Litjeratury.

Ilham, M.R. (1970). *روش جدید در تحقیق زبان دری*.[89] Kabul: Da Puhani Mataba'ah.

Farhadi, A. R. (1955). *Le Persan parlé en Afghanistan. Grammaire du Kabôli*. Paris: Libraire C. Klincksieck.

Henderson, M. M. T. (1972). Phonology of Dari. (Doctoral dissertation, University of Wisconsin, Madison, 1972).

Morgenstierne, G. (1964). *Indo-Iranica; Mélanges présantés à Georg Morgenstierne à l'occasion de son soixante-dixième anniversaire*. Wiesbaden: O. Harrassowitz.

Morgenstierne, G. (1981). *Monumentum Georg Morgenstierne*. Leiden: E. J. Brill.

Morgenstierne, G. (1970). *Report on a linguistic mission to Afghanistan*. Cambridge, MA: Harvard University Press.

Nawid, S. (1994). بررسی مختصر گویشهای دری کابل و فارسی تهران[90] In M. Marashi (Ed.), *Persian studies in North America: Studies in honor of Mohammad Ali Jazayery* (pp. 22-42). Bethesda: Iranbooks.

Nekhat-Saidi, M. N. (1969). *دستور زبان مؤثر دری*[91]. Kabul: Da Kabul Pohantun.

[87] Russian: 'O Yazykje 'Farsii' Afganistana' "On the 'Farsi' of Afghanistan."
[88] Russian: '*Yazyk Forsii-Kabuli*' " The Kabuli Farsi Language."
[89] Dari: '*Ravesh-e jadid dar Tahqiq-e Zaban-e Dari*' "The New Trend in the Study of the Dari Language."
[90] Farsi: "A Brief Study of the Kabuli Dari and Tehrani Farsi Varieties."
[91] Dari: '*Dastur-e Zaban-e mo'asur-e Dari*' "Effective Grammar of the Dari Language."

Phillott, D. C. (1919). *Higher Persian grammar for the use of the Calcutta University, showing differences between Afghan and modern Persian, with notes on rhetoric*. Calcutta: Calcutta University.

Shamal, A.R. (1965). A contrastive study of the segmental phonemes of English and Dari (Afghan Persian). (Master's thesis, Georgetown University, 1965).

Wahab, S. (1995). *Arthur Paul Afghanistan collection bibliography*. Lincoln, NE: Dageforde Publishers.

Yamin, H. (1986). *دستور زبان دری*.[92] Kabul: Pohantun-e Kabul.

X.C Tehrani Persian

Ahsan, 'A. (1976). *Modern trends in the Persian language*. Islamabad: Iran Pakistan Institute of Persian Studies.

Alamolhoda, S.M. (2000). *Phonostatistics and phonotactics of the syllable in modern Persian*. Helsinki: Finnish Oriental Society.

Fouchécour, C.H. (1985). *Eléments de Persan*. Paris: Publications Orientalists de France.

Hadley, G.D. (1972). *Introductory grammatical remarks on the Persian language, 1776*. Menston, Yorkshire: Scolar Press.

Ja'far Mu'-infar, M. (1973). *Phonologie quantitative du Persan, d'après le livre des rois de Firdaus-i*. Paris: Dunod.

Jahanshiri, A. (December 14, 2007). *How did 'miravam' become 'miram'?* Retrieved April 2, 2007, from http://jahanshiri.ir/pa/miram.html.

[92] Dari: '*Dastur-e Zaban-e Dari*' "A Grammar of the Dari Language."

Jones, W. (1969). *A Grammar of the Persian language, 1771*. Menston, Yorkshire: Scolar Press.

Lazard, G. (1992). *A grammar of contemporary Persian*. California and New York: Mazda Publishers.

Lazard, G. (1957). *Grammaire du Persan contemporan*. Paris: Libraire C. Klincksieck.

Mace, J. (2003). *Persian grammar: For reference and revision*. London, New York: Routledge Curzon.

Mahootian, S. (1997). *Persian*. London: Routledge.

Mask-ub, S. (1992). *Iranian nationality and the Persian language*. Washington DC: Mage Publishers.

McLendon, S.V. Some phonemic aspects of formal Iranian. (Master's thesis, Georgetown University, 1960).

Nawabi, Y. M. (1969). *A bibliography of Iran; a list of books and articles on Iranian subjects, mainly in European languages*. Tehran: Khajeh Press.

Obolenshy, S., Panah, K.Y., & Nouri, F.K. (1973). *Spoken Persian*. Ithaca, NY: Spoken Language Services.

Rubinchik, I.U.A. (1971). *Современный Персидский Язык*[93]. Moscow: Nauka Publishing House.

Tabataba'i, M. (1982). A linguistic survey of French borrowings in modern Persian: The adaptation, innivation, and identification of French loan words. (Doctoral dissertation, Georgetown University, 1982).

[93] Russian: '*Sovremennyi Persidskii Iazyk*' "The Modern Persian Language."

Windfuhr, G.L. (1979). *Persian grammar: History and state of its study*. The Hague, New York: Mouton Publishers.

Windfuhr, G.L. (1990). Persian. In B. Comrie (Ed.), *The world's major languages*. Oxford, New York: Oxford University Press.

Windfuhr, G.L. (1997). Persian phonology. In A. Kaye (Ed.), *Phonologies of Asia and Africa* (Vol. 2) (pp. 675-697). Winona Lake, IN: Eisenbrauns.

X.D Dushanbe Persian

Bashiri, I. (1994). Russian loanwords in Persian and Tajiki languages. In M. Marashi (Ed.), *Persian studies in North America: Studies in honor of Mohammad Ali Jazayery* (pp. 109-140). Bethesda: Iranbooks.

Dodykhodoeva, L. (Unpublished). The sociolinguistic situation and language policy of the autonomous region of mountainous Badakhshan: The case of the Tajik language.

Ido, S. (2005). *Tajik*. Munich: Lincom.

Lazard, G. (1956). Caractères distinctifs de la langue Tadjik. *Bulletin de la Société de Linguistique de Paris, 52*(1), 117-186.

Perry, J.R. (2005). *A Tajik Persian reference grammar*. Leiden: Brill.

Rastorgueva, V. S. (1963). *A short sketch of Tajik grammar*. The Hague: Mouton & Co.

Rzehak, L. (2001). *Vom Persischen zum Tadschikischen*. Weisbaden: Reichert.

Sokolova, V. S. (1949). *Фонетика Таджикского Языка*.[94] Moscow-Leningrad: Izdatjel'stvo Akadjemiji Nauk USSR.

X.E Fieldwork Methodology

Baugh, J. & Scherzer, J. (Eds.). (1986). *Language in use*. Englewood Cliffs, NJ: Prentice-Hall.

Briggs, C.L. (1986). *Learning how to ask*. Cambridge/New York: Cambridge University Press.

Cameron, D., et al. (1998). *Researching language: Issues of power and method*. London/New York: Routledge.

Duranti, A. (1997). *Linguistic anthropology*. New York: Cambridge University Press.

Fasold, R. (1984). *The sociolinguistics of society*. Oxford and New York: Blackwell.

Johnstone, B. (2000). *Qualitative methods in sociolinguistics*. New York: Oxford University Press.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society* (97-120).

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Milroy, L. & Gordon, M. (2003). *Sociolinguistics: Methods and interpretation*. Cambridge: Blackwell.

Milroy, L. (1987). *Language and social networks*. Cambridge: Blackwell.

[94] Russian: '*Fonjetika Tadzhikskovo Jazyka*' "Phonetics of the Tajik Language."

Mishler, E. (1991). *Research interviewing*. Cambridge, MA: Harvard University Press.

Wolfram, W. (1998). Scrutinizing linguistic gratuity: Issues from the field journal of sociolinguistics. *Journal of sociolinguistics, 2*(2), 271-279.

Wolfram, W. & Fasold, R. (1974). *The study of social dialects in American English*. Englewood Cliffs, NJ: Prentice-Hall.

X.F Optimality Theory and Use-Based Phonology

Alderete, J. (1997). Dissimilation as local conjunction. *Proceedings of North East Linguistics Society, 27*, 17-31.

Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Browman, C. & Goldstein, L. Articulatory phonology: An overview. *Phonetica, 49,* 155-180.

Burquest, D.A., & Payne, D.L. (1993). *Phonological analysis: A functional approach.* Dallas, TX: Summer Institute of Linguistics.

Bybee, J. (2000). Lexicalization of sound change and alternating environments. In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 250-268). Cambridge: Cambridge University Press.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Calabrese, A. (1988). *Towards a Theory of Phonological Alphabets*. PhD Dissertation, MIT.

Holt, D.E. (Ed.). (2003). *Optimality theory and language change*. Dordrecht: Kluwer.

International Phonetic Association. (1999). *Handbook of the international phonetic association*. New York: Cambridge University Press.

Kager, R. (1999). *Optimality theory*. New York: Cambridge University Press.

Kenstowicz, M. (1994). *Phonology in Generative grammar*. Cambridge, MA: Blackwell

Kirchner, R. (1996). Synchronic chain shifts in optimality theory. *Linguistic Inquiry, 27*, 341-350.

Labov, W. (1994). *Principles of linguistic change*. Cambridge, MA: Blackwell.

McMahon, A. (2000). *Change, chance, and optimality*. New York: Oxford University Press.

Miglio, V. (1998). The great vowel shift: An OT model for unconditioned language change. *The 10th International Conference on English Historical Linguistics.* University of Manchester.

Moreton, E, & Smolensky, P. (2002). Typological consequences of local constraint conjunction. In L. Mikkelsen & C. Potts (Eds.), *WCCFL Proceedings, 21*. Cambridge, MA: Cascidilla Press.

Phillips, B.S. (1994). Word frequency and the actuation of sound change. *Language, 60*, 320-342.

Phillips, B.S. (1998). Word frequency and lexical diffusion in English stress shifts. In R. Hogg & L. van Bergen (Eds.), *Historical linguistics 1995: Vol. 2: Germanic linguistics* (pp. 223-232). Amsterdam: John Benjamins.

Phillips, B.S. (2000). Fast words, slow words. *American Speech 75*(4), 414-416.

Pierrehumbert, J. (2002). Word specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII*. Berlin: Mouton de Gruyer.

Prince, A. & Smolensky, P. (2004). *Optimality Theory*. Oxford: Blackwell.

Smolensky, P. (1993). Harmony, markedness, and phonological activity. Handout to talk at Rutgers Optimality Workshop 1, October 23, 1993. (ROA-87, http://roa.rutgers.edu).

X.G Miscellaneous

Jesperson, O. (1904). *How to teach a foreign language*. London: G. Allen & Unwin.

Sievers, E. (1881). *Grundzöge der Phonetik: zur Einförung in das Studium der Lautlehre der Indogermanischen Schprachen*. Leipzig: Breitkopf & Hörtel.