Iranian Dialectology and Dialectometry

By

Chundra Aroor Cathcart

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Linguistics in the Graduate Division of the University of California, Berkeley

Committee in charge:

Professor Andrew J. Garrett, Chair
Professor Gary B. Holland
Professor Martin Schwartz

Spring 2015
Abstract

Studies in Iranian Dialectology and Dialectometry

by

Chundra Aroor Cathcart

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Andrew Garrett, Chair

This dissertation investigates the forces at work in the formation of a tightly knit but ultimately non-genetic dialect group. The Iranian languages, a genetic sub-branch of the larger Indo-European language family, are a group whose development has been profoundly affected by millennia of internal contact. This work is concerned with aspects of the diversification and disparification (i.e., the development of different versus near-identical features across languages) of this group of languages, namely issues pertaining to the development of the so-called West Iranian group, whose status as a legitimate genetic subgroup has long remained unclear. To address the phenomena under study, I combine a traditional comparative-historical approach with existing quantitative methods as well as newly developed quantitative methods designed to deal with the sort of linguistic situation that Iranian typifies. The studies I undertake support the idea that West Iranian is not a genetic subgroup, as sometimes assumed; instead, similarities between West Iranian languages that give the impression of close genetic relatedness have come about due to interactions between contact and parallel driftlike tendencies. I develop new methodologies which make it possible to demonstrate which similarities are due to contact and which are due to drift.

I present a qualitative comparative-historical treatment of Iranian subgrouping. Several genetic subgroups have been proposed within Iranian; these include a Southwest Iranian subgroup comprising Persian and related dialects, as well as a Sakan subgroup. In addition to this, Iranian languages are typically divided into higher order East and West subgroups, a division that some scholars hold to be genetic. I evaluate these claims in detail. I show that, as generally agreed upon in the literature, there is evidence for a Southwest Iranian and a Sakan subgroup, though this support is limited to a few scant isoglosses. I provide previously unadduced evidence against an East-West subgrouping, and conclude that there is little support for such a genetic divide, and point to a number of phonological and morphological innovations common to West Iranian languages, creating the illusion of a genetic linguistic group; I demonstrate that these features must have come about via contact or shared parallelism among West Iranian languages.

Methodologies from the field of computational phylogenetics are used to investigate the
genetic and areal classification of Iranian. The first part of the chapter uses a Bayesian phylogenetic algorithm to analyze lexical trait data. Lexical items in the Swadesh-200 word list are coded according to cognacy for a set of Old, Middle and New Iranian languages. In addition to lexical characters, I code phyletic (i.e., unlikely to come about in parallel) phonological and morphological characters. In one dataset, etymon-meaning characters are coded as loans when they unambiguously represent borrowings; in another dataset, loans go uncoded. When languages of all chronological levels are included, known subgroups are not replicated unless phyletic characters are heavily weighted or clade constraints are imposed. When only New Iranian languages were included, known subgroups are replicated without clade constraints. The dataset in which loans are coded produces a Southwest Iranian subgroup. The dataset in which loans are uncoded produces a Southwest Iranian subgroup within a West Iranian subgroup. In the second part of the chapter, distance-based metrics are used to observe patterns shown by phenetic (i.e., likely to come about in parallel) and other typological characters. Networks involving typological characters of different types (i.e., phonological and morphosyntactic) show an East-West split between languages of all chronological levels. These results provide additional evidence for the idea that West Iranian is an areal grouping rather than a genetic subgroup.

I introduce a novel quantitative means of modeling irregular sound change designed to investigate whether West Iranian languages have developed formally similar functional items (i) due to shared tendencies expected among a group of closely related languages, or (ii) via language contact. I analyze plausibly cognate functional items in two Middle and eight New Iranian languages. I compare their observed forms to the “expected” outcomes of these forms that would result from purely regular sound change. I integrate this information into a quantitative model that measures the Levenshtein distance between the observed and expected forms. I use mixed-effects linear regression to model these distance measures as a function of LANGUAGE, with random intercepts for each FUNCTIONAL ITEM; since I wish to observe how non-Persian languages differ from Persian, the mean New Persian value was dummy coded as the intercept. I find that while most New West Iranian languages undergo roughly the same amount of reduction, Balochi, a language with particularly conservative historical phonology, shows significantly more irregular phonological reduction than New Persian in its functional vocabulary. I interpret this result as a detection of contact between New Persian and Balochi. Given this result, it is plausible that the similarity in functional items seen across West Iranian is due to New Persian influence, though this contact may not be quantitatively detectable due to similar trends in regular historical phonology.

Finally, I adapt methodologies from the literature on dialectometry to measure the extent to which historical phonological variation shown by New West Iranian languages reflects a geographic signal. New West Iranian languages show a great deal of variation in their reflexes of certain Proto-Iranian sounds, often manifested as doublets. Variation of this sort is generally ascribed to lexical borrowing between dialects in contact, and certain reflexes are thought to be attributable to particular dialects or dialect regions. I aggregate this variation according to the etymological reflex in which each variant occurs, and observe correlations between distances based on aggregate variation and different geographic distance
measures. I find that distances based on certain varying reflexes of certain Proto-Iranian sounds show a highly significant correlation with geographic distance, while others do not (e.g., *u-*). Crucially, evidence from Middle Persian and Parthian shows us that *u-* is preserved until a relatively late date, meaning that the distribution of reflexes of *u-* cannot reflect an earlier stage of linguistic geography (patterns of which have become non-linear over time). These models show for the most part that varying reflexes are due to one-off diffusion events; however, a change *u > b* appears to have come about in parallel at least twice, clarifying a longstanding question in the Iranological literature. This highlights the joint role of contact-induced and parallel, drift-like changes in the formation of West Iranian as a linguistic group.
## Contents

<table>
<thead>
<tr>
<th>List of Figures</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
</tbody>
</table>

### 1 Introduction
1.1 Reconstruction and subgrouping ........................................... 1
1.2 Criteria for Subgrouping .................................................... 2
1.3 Contributions of this dissertation to the study of linguistic diversification .. 4

### 2 Iranian development and diversification
2.1 Introduction ................................................................. 5
2.2 Iranian languages: the traditional taxonomy ................................ 5
2.3 PIE to PIIr to PIr ................................................................. 8
   2.3.1 From Proto-Indo-Iranian to Proto-Iranian ................................ 9
2.4 Subgrouping within Iranian ................................................... 10
   2.4.1 Sakan ................................................................. 10
   2.4.2 Southwest Iranian ...................................................... 11
      2.4.2.1 Southwest Iranian-defining changes ................................ 11
         2.4.2.1.1 PIr \( *\acute{s} > \theta \) ............................................. 11
         2.4.2.1.2 PIr \( *\acute{z} > [\ddot{o}] > d. \) ................................ 12
         2.4.2.1.3 PIr \( *\acute{y} > s. \) .................................................. 12
         2.4.2.1.4 PIr \( *\theta r > \varsigma [s]. \) ........................................ 13
         2.4.2.1.5 PIr \( *\theta t > s(i)y. \) .............................................. 13
         2.4.2.1.6 PIr \( *\acute{z} > \theta (i)y. \) ........................................ 14
         2.4.2.1.7 PIr \( *\acute{n} > (x)\acute{n}. \) ........................................ 14
         2.4.2.1.8 PIr \( *ja- \) rel. pron. \( \rightarrow \) OP haya m., taya n., hayā f. 15
         2.4.2.1.9 PIr \( *pa\acute{s}c\ddot{a}- \) OP pasā- \( \rightarrow \) MP ps /pas/ \( \rightarrow \) NP pas 'behind'. .................................................. 15
      2.4.2.2 Less straightforward Southwest Iranian-defining innovations ........ 15
         2.4.2.2.1 PIr \( *\acute{s}t > st? \) ................................................ 15
         2.4.2.2.2 PIr \( *\acute{s}r > \varsigma? \) .......................................... 15


2.4.3 The rest of Iranian .................................................. 16
2.4.4 Other postulated dialectal affinities ............................ 16

2.5 East vs. West Iranian ................................................. 17
2.5.1 No Eastern subgroup .............................................. 17
2.5.2 Southwest, Central, and Northeast, and Southeast subgroups .... 17
2.5.3 Western vs. Eastern subgroups (+ Avestan) ..................... 17
2.5.4 Western vs. Eastern subgroups (− Avestan) ................. 18
   2.5.4.1 Lenition of PIr *b(r)−, *d−, *g− .............................. 18
   2.5.4.2 PIr *-ft−, *-xt− ............................................. 19

2.5.5 East vs. West Iranian lexis ..................................... 21
   2.5.5.1 PIr *kaufa− ‘mountain’ ..................................... 21
   2.5.5.2 Real East-West lexical isoglosses? ......................... 22
   2.5.5.3 East Iranian .................................................. 22
      2.5.5.3.1 *maiθa− ‘day’ .......................................... 22
      2.5.5.3.2 *kapā− ‘fish’ ........................................... 23
   2.5.5.4 West Iranian .................................................. 23
      2.5.5.4.1 *hižu(ā)− (against other Iranian *hiža−) ‘tongue’ 23
   2.5.5.5 Summary of East Iranian vs. West Iranian shared innovations 25

2.6 East Iranian vs. West Iranian historical trends ................... 25
   2.6.1 West Iranian treatment of *θ .................................. 25
   2.6.2 West Iranian treatment of *θn .................................. 25
   2.6.3 Other West Iranian phonological patterns .................... 27
      2.6.3.1 Reflexes of PIr *š, etc. .................................. 27
      2.6.3.2 PIr *r + coronal clusters ................................ 29
      2.6.3.3 PIr *y− .................................................... 30
      2.6.3.4 West Iranian functional items ............................ 32

2.7 Conclusion ............................................................ 33

3 Computational classification of the Iranian Languages 34
3.1 Computational phylogenetic linguistics ............................ 35
   3.1.1 Lexical Characters .............................................. 36
   3.1.2 Phyletic Characters ............................................ 36
   3.1.3 Phenetic Characters .......................................... 37
   3.1.4 Previous Literature .......................................... 37
3.2 Character Coding ................................................... 38
   3.2.1 Phenetic Characters .......................................... 38
      3.2.1.1 Phonological Characters ................................ 39
         3.2.1.1.1 Rhotacism of dentals ................................. 39
         3.2.1.1.2 Loss of PIr *θ ....................................... 39
         3.2.1.1.3 PIr *śr > sr, ḵ/s, ś, rs ............................ 39
         3.2.1.1.4 PIr *śy > sp ........................................ 40
         3.2.1.1.5 Lateralization of dentals .......................... 40
         3.2.1.1.6 PIr *-ft−, *-xt− > -vd−, -yd− .................... 40
## 5 Geography and contact-induced variation in West Iranian

5.1 Introduction .................................................. 73
5.2 Etymological variation in West Iranian ........................... 74
5.3 Contact vs. Parallelism ........................................ 77
5.4 Methods ...................................................... 77
  5.4.1 Linguistic distance measures .................................... 78
  5.4.2 Geographic distance measures ................................... 79
  5.4.3 Models ................................................... 79
5.5 Results ....................................................... 79
  5.5.1 Results for all West Iranian languages ......................... 79
  5.5.2 Results for Northwest Iranian languages ....................... 82
5.6 Discussion .................................................. 84
5.7 Conclusion .................................................. 86

## 6 Conclusion .................................................. 87

Appendix A: Chapter 3 Supplementary Materials ....................... 89
Appendix B: Chapter 4 Supplementary Materials ....................... 121
Appendix C: Chapter 5 Supplementary Materials ....................... 132
Bibliography .................................................. 135
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Map of Iranian Languages</td>
<td>6</td>
</tr>
<tr>
<td>3.1</td>
<td>NeighborNet with Binary Typological/Recurrent Characters</td>
<td>47</td>
</tr>
<tr>
<td>3.2</td>
<td>Typological data; consensus tree</td>
<td>49</td>
</tr>
<tr>
<td>3.3</td>
<td>Lexical data; all languages; loans coded</td>
<td>50</td>
</tr>
<tr>
<td>3.4</td>
<td>Lexical data; all languages; loans uncoded</td>
<td>51</td>
</tr>
<tr>
<td>3.5</td>
<td>Lexical data; contemporary languages; loans coded</td>
<td>52</td>
</tr>
<tr>
<td>3.6</td>
<td>Lexical data; contemporary languages; loans uncoded</td>
<td>53</td>
</tr>
<tr>
<td>3.7</td>
<td>NeighborNet with Lexical Characters, loans coded</td>
<td>54</td>
</tr>
<tr>
<td>4.1</td>
<td>Approximate geographic locations of West Iranian languages treated in this chapter</td>
<td>59</td>
</tr>
<tr>
<td>4.2</td>
<td>Multidimensional scaling of inter-language distances</td>
<td>70</td>
</tr>
<tr>
<td>5.1</td>
<td>Schema of West Iranian dialects in MacKenzie 1961:71, according to non-varying phonological and morphological features</td>
<td>76</td>
</tr>
<tr>
<td>5.2</td>
<td>Correlation coefficient ( r ) for the relationship between great-circle distance and palatal distance, lateral distance, and labiovelar distance</td>
<td>81</td>
</tr>
<tr>
<td>5.3</td>
<td>Density curves for each distance type</td>
<td>82</td>
</tr>
<tr>
<td>5.4</td>
<td>Plots of palatal, lateral, and labiovelar distance as functions of great-circle distance (distance measures involving SW Iranian languages are in dark print; distance measures involving only NW Iranian languages are light print)</td>
<td>83</td>
</tr>
<tr>
<td>5.5</td>
<td>Correlation coefficient ( r ) for the relationship between geographic distance and overall linguistic distance, palatal distance, lateral distance, and velar distance (excluding reflexes of (*u)-)</td>
<td>85</td>
</tr>
</tbody>
</table>
List of Tables

4.1 Distance between EXPECTED and OBSERVED outcomes by language, all forms 67
4.2 Distance between EXPECTED and OBSERVED outcomes by language, forms equivalent to plausible cognates in Persian 68
4.3 Distance between Proto-Iranian etyma and lexical items in different languages 70

5.1 $R^2$ and $p$ values for correlations between linguistic and geographic distances (asterisks denote significance codes) for all West Iranian languages 80
5.2 $R^2$ and $p$ values for correlations between linguistic and geographic distances (asterisks denote significance codes) for Northwest Iranian languages 82
5.3 $R^2$ and $p$ values for correlations between velar and labial distance and geographic distance for all Western Iranian and only Northwest Iranian languages 84

6.1 Distance measures between expected and observed functional items 125
6.2 Distances between Proto-Iranian etyma and reflexes, in lossless ASCII form 127
‘Because the philological race is the most stupid under heaven,’ said the man in black; ‘they are possessed, it is true, of a certain faculty for picking up words, and a memory for retaining them; but that any one of the sect should be able to give a rational answer, to say nothing of an acute one, on any subject — even though the subject were philology — is a thing of which I have no idea.’

George Borrow, *Lavengro*
Acknowledgements

This dissertation would not have come to fruition but for the support and guidance of a number of people. I am grateful to my committee for the lessons that each member has imparted to me. I appreciate Andrew’s encouragement to be bold and push the envelope; I aspire to Gary’s level of care and precision; and I have a fondness for Martin’s whimsy, and for the creative and aesthetic sensibilities that reside in his work. I am very fortunate to have had the experience of learning from and exchanging ideas with all of the aforementioned. I hope to some day fully do justice to the scholarly example that they set.

I owe a tremendous debt to the administrators of the Berkeley Linguistics Department, particularly Belén Flores and Paula Floro. Their tireless work and advocacy on my behalf has helped me clear a number of hurdles, administrative and otherwise. I cannot thank them enough. I have been fortunate enough to study alongside people who have helped to foster an environment of diverse and eclectic interests. Special thanks go to Will Chang, who offered copious advice regarding the quantitative methods I wished to develop for this dissertation — though I accept full responsibility for aspects of these methodologies that are in need of further refinement. I am grateful to the folks at the Sino-Tibetan Etymological Dictionary and Thesaurus, particularly Jim Matisoff and J.B. Lowe, for broadening my views on historical linguistics and giving me the opportunity to develop some technical skills that I otherwise would have put off.

Thanks also to Michael Weiss for comments on a draft of this dissertation, and to Alexis Manaster Ramer for feedback on an early version of Chapter 2. I must also acknowledge a debt to Stephanie Jamison, Brent Vine and Craig Melchert for guidance they have provided at a number of junctures during my graduate career. Additionally, I have never forgotten those who encouraged me to pursue linguistics during my salad days; Elena Bashir, Victor Friedman, Lenore Grenoble, C.M. Naim, and Blake Wentworth — many thanks to all of you.

Friends and family have helped me break up the monotony of coursework and writing. I am grateful to my parents for their encouragement and support. For picnics, A’s and Warriors games, walks, runs, and miscellaneous shenanigans, I thank Eoin, Christine, Phaedra, Vivian, Alex, Dan, Vin, Luke, Harold, and others. I cannot thank Tammy enough for her love, patience and support. My grandmother, a polyglot who made it look easy, was one of the first people to stimulate my interest in languages and comparative linguistics. This dissertation is dedicated to her memory.
Chapter 1

Introduction

1.1 Reconstruction and subgrouping

Linguistic reconstruction at its most ambitious seeks to serve as a record of human prehistory. By demonstrating genetic relatedness between a large number of languages, we can attempt to trace the spread and diversification of those languages’ stock back to an ancestral homeland. By reconstructing binarily branching linguistic subgroups within a larger stock, we can hope to pinpoint the divergence of two groups of previously unified people.

The reconstruction of subgroups was formalized by Hoenigswald (1960, 1966) via pairwise comparison of form-meaning correspondences across languages. But a number of problematic issues in subgrouping persist. First, subgrouping models which reconstruct shared innovations against archaisms are often simplistic: members of a diversifying speech community may remain in contact as innovations diffuse across it; or, even if a speech community undergoes a clean split, speakers from the resulting diversified speech communities may come back into contact. Additionally, across historical linguistics, scholars have generally failed to agree on standard criteria for subgroup-defining innovations (see §1.2).

These and other shortcomings have long been recognized in historical linguistics (Schrader 1907; Schuchardt 1885; Gilliéron and Roques 1912), and have been dealt with in different ways and with different models found in Schmidt 1872 (the Wave Model); Southworth 1964 (the “diachronic isogloss map”); van Driem 2001 (“fallen leaves”); Toulmin 2009; François 2015, Kalyan and François forthcoming (“glottometry”).

Much of our current understanding of linguistic diversification comes from the Austronesian literature. A crucial concept is that of the “linkage,” or a set of genetically related languages, or a dialect network, across which key defining innovations have diffused (Ross 1988; among others Geraghty 1983, Babel et al. 2013). The empirical evidence brought to bear on the subject of linguistic diversification by Austronesian languages often leads scholars (e.g., Donohue et al. 2008) to impute putative shared genetic innovations to deep areality. (It goes without saying that genetically related languages were once areally very proximate, but here, the areality is epiphenomenal.)

In sum, similarities seen across languages can be due to a number of factors, including
the following:

- Shared genetic innovations
- Areally diffused or contact-induced features
- Parallelism, hypothetically comprising
  - Parallel developments due to drift (cf. Sapir 1921:ch. 7) or advergence (Renfrew 2000) between genetically related languages
  - Chance parallelism, with no genetically inherited precursor

A number of approaches to the study of linguistic diversification, such as glottometry, seek to tease apart innovations that look more genetic and ones that look less genetic. A set of tools that will make it possible to distinguish between contact-induced change and parallelism among closely related languages (between related languages, it is assumed that any parallel developments will be due to drift and/or advergence) are much needed, a desideratum that many other scholars are currently working to address. This dissertation seeks to at least partly advance this goal.

1.2 Criteria for Subgrouping

When a clean split between languages is observable, we face the challenge of deciding whether the innovations that define it, both on their own and as a whole, are of the sort that are diagnostic of subgrouping. How likely are they to have come about via chance? How likely are they to be shared? At the moment, there is no principled way of answering questions of this sort. Specialists are often left wondering whether they should posit a subgroup just because they can.

What, then, are the sort of innovations that should be taken to define a subgroup? This is also the cause of disagreement in the literature. On treating sound changes as diagnostic of subgrouping, see Ringe et al. (2002:66–8): “Sound changes are usually so ‘natural’ that they can easily be repeated in different lines of descent.” Following from this is the idea that an “unnatural”-looking sound change could possibly serve as a shared innovation capable of defining a subgroup. This has generally been the consensus regarding sound changes like the RUKI change, where Proto-Indo-European *s becomes Proto-Indo-Iranian *š when directly preceded by *r, *r̥, u, k, or i. This change’s unusual quality has led scholars like Ringe et al to posit it as an innovation defining not only Indo-Iranian, but Balto-Slavic as well, despite the fact that there is some evidence that RUKI was diffused to, not inherited by, the latter clade (e.g., it is not fully operative, failing to affect forms such as Lithuanian visas ‘all’).

Further complicating matters are attempts (chiefly by phonologists) to present RUKI as a phonetically natural change or alternation, at least synchronically (Vennemann 1974). In any event, scholars have posited subgroups based on (often) trivial, recurrent sound changes, simply because they can. Iranian, for one, has long been defined by (inter alia) the change

2
*s > *h, a common change that occurs often cross-linguistically (though increasing evidence has been adduced in favor of the idea that this change postdates Iranian unity). But the distinction between “natural” and “unnatural” sound changes remains a fairly subjective choice. Compendia of sound changes (e.g., Kümmel 2007, Hamed and Flavier 2007) provide absolute frequencies of occurrence of sound changes in a large sample of languages and stocks.

Given the importance of morphology (particularly functional morphology, cf. Goddard 1975) in linguistic reconstruction, morphological innovations have been put forth as strong diagnostics of joint development, but this too needs qualification: these innovations must be sufficiently unusual enough to be deemed unlikely to take place in parallel. This can be problematic when arguing for an odd morphological innovation: the more unusual such an innovation is, the harder it is to convince an audience that it in fact took place (as opposed to a simpler account). On the other hand, morphological changes can be recurrent, and likely have precursors in usage that are common to a number of language varieties, related or otherwise (Joseph 2012 discusses the merger of genitive and dative in Young Avestan and Classical/Epic Sanskrit; both developments, while absent in the older language, may have their seeds in an earlier merger between genitive and dative pronominal clitics). This distinction between trivial and non-trivial morphological changes can be seen in the ongoing debate regarding Italo-Celtic unity. Some morphological innovations shared by the two branches (such as the i-genitive, which has been linked to derivation of the vrkī-type — ignoring also the fact that *-osja-genitives can be found in Italic and arguably in Celtic) could easily have come about in parallel, while others (e.g., superlatives in *ismmo-) are less likely to have done so. Watkins (1966) and Cowgill (1970) cite a number of Italo-Celtic morphological innovations that are (respectively) trivial and non-trivial, non-probative and potentially subgroup defining.

Lexis is another domain of linguistic innovation, and vocabulary items can often easily be reconstructed for certain subgroups as against others. However, in small numbers, lexical innovations are not seen as particularly convincing diagnostics of subgrouping, given their proneness to being borrowed (e.g., Aikhenvald 2006:168; Kalyan and François:18). Lexical items are prone to areal diffusion, and furthermore, semantic innovations underlying the specialization of a lexical item in a new context are often likely to recur. This, however, has not stopped scholars from positing subgroups based on a scant number of lexical innovations (often as few as one: Parpola 2012 posits a subgroup containing Avestan, Scythian, Saka, and Ossetic on the basis of a single word, *jažata- ‘god’).

While lexical innovations in isolation may not tell us much about subgrouping, advances in computational phylogenetics have made it possible to observe overall patterns of lexical replacement among related languages, and return an optimal tree with an internal structure that reflects these patterns. While the results of this methodology can show compatibility between a particular view of subgrouping and lexical replacement, they should not necessarily supersede subgroups based on traditional comparative-historical methods, but, in the absence of traditional subgroup-defining innovations, are often quite informative. At the same time, shared patterns of lexical replacement may be a byproduct of linguistic fission, but may also be a byproduct of other factors as well.

In addition to the features mentioned above, subgroups may share features that are harder
to reconstruct, like prosody or poetics; parallel changes in related languages that are not reconstructible to a common ancestor may be effects of usage-based factors at play in a common ancestor.

1.3 Contributions of this dissertation to the study of linguistic diversification

This dissertation comprises a handful of studies designed to investigate the sociolinguistic forces at work in the West Iranian languages, a dialect group which displays few, if any, subgroup-defining innovations. The Iranian languages, a genetic sub-branch of the larger Indo-European language family, are a group whose development has been profoundly affected by millennia of internal contact. This work is concerned with aspects of the diversification and disparification (i.e., the development of different versus near-identical features across languages) of this group of languages, namely issues pertaining to the development of the so-called West Iranian group, whose status as a legitimate genetic subgroup has long remained unclear. To address the phenomena under study, I combine a traditional comparative-historical approach with existing quantitative methods as well as newly developed quantitative methods designed to deal with the sort of linguistic situation that Iranian typifies. The studies I undertake support the idea that West Iranian is not a genetic subgroup, as sometimes assumed; instead, similarities between West Iranian languages that give the impression of close genetic relatedness have come about due to interactions between contact and parallel driftlike tendencies. New methods let me demonstrate which similarities are due to contact and which are due to drift.
Chapter 2

Iranian development and diversification

2.1 Introduction

This chapter deals with the internal subgrouping of the Iranian languages and presents a survey of features probative of various Iranian-internal subgroups. While some subgroups, e.g., Southwest Iranian and Sakan, are relatively uncontroversial, there is still ongoing debate regarding an East-West genetic grouping within Iranian.

Here I assess the evidence in support for East and West Iranian genetic subgroups, and find that it is scant. At the same time, I draw attention to a number of remarkable innovations common to West Iranian languages that cannot be genetically shared, since in many cases they postdate the Old Iranian period, and instead must have come about in parallel or via contact.

I also sketch out some desiderata for future study of West Iranian similarities, which serve as the basis for this dissertation’s remaining chapters.

2.2 Iranian languages: the traditional taxonomy

The Iranian languages are a diverse group; they have developed in contact with each other for millennia. The literature on Iranian languages typically makes a two-way distinction between East and West. These labels largely correspond to the geographical location of the languages to which they apply, but occasionally do not. Typologically, the East-West division is highly informative. Each group is tightly knit in terms of the grammatical patterns displayed by its members, which have also undergone similar trends in their historical phonology. However, there is disagreement over whether East and West Iranian are in fact genetic subgroups (these conflicting views are summarized below). (Smaller subgroups within Iranian, namely Southwest Iranian and Sakan, are fairly uncontroversial.) A simplified list of Iranian languages follows, divided according to the standard categories of East and West (I give spellings with diacritics here, but omit them in later parts of this dissertation).

1. West Iranian
Figure 2.1: Map of Iranian Languages
(a) Southwest Iranian
   i. Old: Old Persian (OP)
   ii. Middle: Middle Persian (MP; Inscriptional, [Book] and Psalter Pahlavi, Manichean, Pāzand1)
   iii. New: New (Modern) Persian (NP), Lurī (Lur), Kumzārī (Knz, Thomas 1930), Baškardī (Bshk, Skjærvø 1988; Voskanian and Boyajian-Surenianis 2007), Lārestānī (Lar, Kamioka and Yamada 1979), Bandarī (Band, Pelevin 2010), dialects of Fars (Frs, incl. [Judeo-]Širāzī)

(b) Northwest Iranian
   i. Old: Median (fragmentary)
   ii. Middle: Parthian (Pth; Inscriptional, Manichean)
   iii. New: Kurdish (Kd, McCarus 1958; MacKenzie 1961), Balōčī (Bal, Barker 1969; Korn 2005), Zāzakī (Zaz), Gorānī (Gor), Awromanī (Awr, Benedictsen and Christensen 1921) Tāleṣī (Tal, alternatively Talysh), Sīvandī (Siv, Lecoq 1979), Semnānī (Sem), Sangesarī (Sang, Azami and Windfuhr 1972), the Dialect of Gāz (Gaz), the Central Dialects (CD) and the Caspian Dialects (Casp), comprising Gīlakī (Gil) and Māzandarānī (Maz, Nawata 1984)

2. East Iranian (these have at times been divided according to North and South, e.g., Oranskij 1977:197, but there’s not widespread agreement on this)

(a) Old: Avestan (Old and Young; often called “Central”), Scythian (fragments)

(b) Middle: Sogdian (Sog; in three scripts: Christian Sogdian [CSog], Manichean Sogdian [MSog], and Sogdian Sogdian [SSog], of the Muγ documents and Buddhist texts, though some of the latter are written in the Sūtra script), Khwarezmian (Khw), Bactrian (Bct), Alanic (fragmentary), Sarmatian (fragmentary), Jassic (fragmentary, Nemeth 1959), “Parnian” (substrate found in Parthian, Sims-Williams 1989b: 171, and possibly Armenian, Olsen 2005)

(c) New: Pašto (Psht), Ormurī (Orm, Morgenstierne 1929), Parācī (Par, ibid.), Ossetic (Oss, Iron dialect, unless noted as [D]igor, Thordarson 2009), Yaŋnobī (Ygh, Andreev and Peščereva 1957); so-called “Pamir” languages include Šuyń (Shg, Zarubin 1960), Sangličī-Iškāšmī (SI), Yidyā-Munj(ān)ī (YM), Yazyulāmī (Yzg)

(a) Sakan subgroup
   i. Middle: Tumšuqese (Tshq) and Khotanese (Khot) Saka
   ii. New: Waxī (Wkh)

---

1 Generally considered to be later and more corrupt than other Middle Persian varieties.
This chapter assesses the validity of an East-West subgrouping in Iranian. I find that the argument for an East-West divide hinges on a few lexical and lexeme-specific morphological isoglosses, which are not a particularly good diagnostic of subgrouping (as opposed to a change in the grammatical system). I then outline possible scenarios of how the Iranian configuration came to be — particularly how West Iranian, a group with no good shared subgroup-diagnosing innovations came to be so low in linguistic disparity — a cohesive dialect group capable of displaying a sort of genetic-looking signal.

2.3 PIE to PIIr to PIr

Proto-Indo-Iranian is an uncontroversial subgroup defined by the operation of key phonological and morphological developments.

1. Generalization of the PIE augment *(h₁)e- to all aorist verbs (also affects Greek and Armenian)
2. Labio- and Plain Velar Merger: PIE *k₁, *k₂w > PIIr *k; *g₁, *g₂w(h₁) > *g(h₁)
3. The Law of the Palatals: Early PIIr *k₁, *g(h₁) > PIIr *č, *ɟ(h₁) before (PIE) *e, *i
4. Indo-Iranian Vowel Merger: PIE *e > PIIr *a; PIE *o > PIIr *a, PIE *e̞, PIE *e̞, *o̞,
   *a > PIIr *a
5. Brugmann’s Law: PIE *o > PIIr *a in an open syllable, *a in a closed syllable
6. Bartholomae’s Law: \{ s \} → \{ z \} / D (D represents a segment with breathy voice)
   (the process D → T / _ \{ s \} already took place in PIE, where D represents a
   segment with modal voicing and T represents a voiceless segment)
7. Pedersen’s Law or RUKI: PIE *s, *z (allophone of /s/ before D(h)) > PIIr *š, *ž
9. Laryngeal Merger: PIE *h₁, *h₂, *h₃ > PIIr *H; PIIr *CH > *C (exceptions due
to Inner-IE laryngeal developments)
10. Laryngeal Vocalization: *H > /H/ (-)C_Cè, -C_(C), -C_{sR} (Lipp 2009:II
   485-7)
11. PIE *n > *an / _ V, *a / _ \{ C \} ; PIE *m > *am / _ R, *a / _ \{ O \}

²Kobayashi (2004:24, 25, 137–8) holds that the regular PIIr outcome of the PIE syllabic nasals is *a, with
*am and *an restored via analogy.
12. **Liquid merger**: PIE *t₁, t_r > PIr r (the status of this merger is questionable; we see what appears to be PIE *l in peripheral IIr dialects); PIE *t₁, t_r > PIr *r

13. PIE *oh₂om → PIr *Vnām in the genitive plural ending for thematic nouns

14. PIE *h₁e₂go₄-’I’ → *h₁e₂gh₂-om, and other pronominal changes (see Meillet 1922:26–7)

15. PIr *c > *š / t, t, possibly p

16. PIr */s/ → [ś] / t, t, possibly p

### 2.3.1 From Proto-Indo-Iranian to Proto-Iranian

1. PIr *pᵉ* *tᵉ* *kᵉ > PIr *f *θ *x

2. PIr *p *t *k > PIr *f *θ *x before consonants

3. PIr *s > PIr *h, except before *p, *t, *k, *n, after *t (see Kümmel 2012 on the preservation of *s when *n precedes it and a resonant follows it in Avestan; see also Hintze 1998; Tremblay 2005a on this change postdating Proto-Iranian)

4. PIr *b₁h *d₁h *g₁h > PIr *b *d *g

5. PIr *t₁tᵉ > PIr *st, *d₁dᵉ > PIr *zd

6. PIr *H > θ in non-initial syllables (Lipp 2009:11485)

7. PIr *ps > PIr *fš, *bz > PIr *βž

8. PIr *ts, *č > PIr *š, PIr *j > *ž (see below)

9. PIr *čš > PIr *š

10. PIr *(č)sč > *š

11. PIr *-šn- > *-šn- (e.g., YAv ašnō : OInd aśnas gen.sg. ‘stone’), *-šn- > *sn- (e.g., YAv snath- : OInd snath- ‘strike’)

Authors represent the Proto-Iranian palatals (< PIIR *č, *j) according to different conventions. Some use affricates (e.g., *ts, *dz; *č, *j; an advantage of this convention is that PIIR *ts behaves identically to PIIR *č in Iranian, cf. the separate developments of the Avestan and Persian word for ‘fish’: PIIR *matsia- > PIr *matsia- > YAv masiaa-, OP *māthiya(ka)- > MP m’hī(g) / m’hīg/: Ved. matsya-), believing Old Persian ąd and d to be more likely to develop from the occlusive element. Some (e.g., Cheung 2007) simply write s and z, believing that these sounds were alveolar. Some scholars (e.g., Sims-Williams 1998:136) believe that the Sakan change of PIIR *ču, *j > *š, *ž shows that these sounds were still alveo-labial in PIr; however, the low F1–3 of [w] could coarticulatorily lower the spectral energy of preceding
[s], yielding a secondary [c] or [ʃ]. Some authors even believe that a PIr affricate *ć [tʃ] was preserved in some Old Iranian dialects in some contexts. Tocharian A ančwasi ‘in steel’, B eńcuwo, ińcuwo ‘steel’ are most definitely Iranian loans, perhaps from “Old Sakan” *ančyän- (Tremblay 2005b:424); cf. Khwarezmian hnčw ‘iron-tipped’, perhaps a back-loan (Schwartz 1974:409, fn. 33). But parallel examples show that nasals could bring about the occlusion of a following fricative, or preserve affrication (Martin Schwartz draws my attention to CSog 'nc’y /antₐaj/ ‘rest’ < *han-ćjāja- and related forms, which have not undergone the change *ćj > š(y), vs. CSog ptš’dy < *pati-ćjātaka-).3 And again, the palatal quality could ultimately come from the glide [w] (cf. above). So, it is most likely that the PIr reflexes of the PIE palatals were simply pronounced [s] and [z]; I write *ś and *ź to minimize confusion with other sibilants.

Voiceless consonants spirantize before other consonants; however, Avestan shows pt as a reflex of PIr *pt whereas the rest of Iranian reflects *ft. This has been taken as an archaism by some scholars (e.g., Hoffmann, Beekes). Others (myself included, cf. also Skjærvø 2009b) believe that this is one of many redactional phonological idiosyncrasies of Avestan. Martin Schwartz reports W.B. Henning’s teaching that pt is an innovation in the archetype reflecting the regional pronunciation of Kermān and Yazd (centers of Zoroastrian learning with some consonant fortition developments).

### 2.4 Subgrouping within Iranian

Some inter-relationships between Iranian languages are fairly uncontroversial. This section serves to describe agreed upon subgroups (see Skjærvø 2009b; Windfuhr 2009, inter alia), and also to detail relevant changes affecting the Iranian languages outside of these subgroups.

Southwest Iranian and Sakan are generally accepted subgroups, defined by innovations described in the following sections.

#### 2.4.1 Sakan

Sakan, a group consisting of Tumshuqese and Khotanese Saka and Wakhi, shows at least two phonological isoglosses, first noted by Morgenstierne (1938:469). This is, of course, a fairly small amount of evidence in favor of a genetic subgroup, but the nature of the data (in that these changes are confined to few forms) makes it difficult to argue with respect to whether or not these innovations are shared.

- The Ossetic word sist, D sistæ ‘flea’ (probably < *šiš-) is ascribed by Thordarson

---

3Sims-Williams (1989d:261) attributes Sogdian c in ‘nc’y to a dissimilation *ćj...j > č...j. This account may also explain Brahmi Sogdian ica- ‘comfortable’ < *ūžč̱(i)āta- (Sims-Williams 1996:308), if formed on the basis of a dissimilated stem *ūžč̱ai- < *ūžč̱iā-. 
(2009:13) to contact with a “cognate dialect,” probably of a Sakan affiliation (cf. Wkh šiš ‘id.’ < *šuš-).

- PIr *źu > ž (?): *hiźu- ‘tongue’ > Khot bišaa- /bižaː/, Wkh zik (with secondary de-palatalization?); *źyar- > Late Khot sīr- /ʒiːːɾ/ ‘to go ill, turn bad’ (Cheung 2007:475).

I don’t know that the palatal quality (where it exists) of the above reflexes necessarily speaks to the preservation of the quality of PIr *č, *j, given the ability of [w] to palatalize an adjacent [s] or [z] (discussed above). There is no Tumshuqese evidence for the change *źu > ž, and the idea that Wakhi depalatalizes an earlier *ź is merely an assumption, making this change’s status as a Sakan-defining one somewhat unclear.

2.4.2 Southwest Iranian

The Southwest Iranian subgroup comprises Old, Middle and New Persian (including a number of dialects, e.g., Farsi, Tajik, Hazaragi, Aimaq, Tat-Persian) as well as Kumzari, Baskardi, Larestani, Bandari, dialects of Fars, and Luri (including the dialects Buyer-Ahmad, Mamasani, Kohgiluya, Baxtiyari, and perhaps Lak). Old, Middle and New Persian are generally thought to represent a direct conduit of descent; however, this is complicated by the fact that the Achaemenid corpus is relatively small, and probably does not represent the full range of dialectal variation that evolved into Middle Persian (see Hoffmann 1976a:61-62, Gershevitch 1962).4

Dialect admixture has created a great deal of variation in Southwest Iranian (particularly Persian) reflexes of Proto-Iranian sounds. A distinction is generally drawn between “proper Old Persian” (“echt-Altpersisch,” e.g., Hoffmann 1976a:62 et passim) and “Median” elements in Old Persian (the label “Median” is an inference regarding the source of non-“proper”-looking lexical items; we have no direct Median attestations, and the only word we know of attributed directly to the Median language is σπάκα ‘dog’, recorded by Herodotus [1.110.1]). This logic can be extended to Middle and New Persian; reflexes that display the innovations listed below are generally taken to be authentic Persian forms, and others due to admixture from Northwest Iranian languages (Tedesco 1921; Lentz 1926; Paul 2005). (Middle and New Persian contain likely loanwords from East Iranian languages as well.)

2.4.2.1 Southwest Iranian-defining changes

The following are generally accepted Southwest Iranian innovations. Some developments require more comment than others.

4For instance, in many situations, we can see the breakdown of the Old Persian case system as it leads into Middle Persian, e.g., būnim ‘earth’ ACC (Darius, Naqš-i Rustam a) → būmām ‘id.’ (Artaxerxes, Persepolis a) > MP būm /bʊm/. However, we see morphological changes that are not reflected in later languages, e.g., Proto-Iranian *xratu- ‘wisdom’ → OP xraθu-, which cannot yield Phl hlt, MMP xrd /xrad/ > NP xirad; (Mayrhofer 1996:402) suggests that OP xratum acc.sg. (Xerxes, Persepolis) is a Median loan.
2.4.2.1.1 PIr \(^*\dot{s} > \theta\). At the same time, a number of words found in Persian show \(s\) for PIr \(^*\dot{s}\), usually ascribed to “Median” or other NW Iranian admixture. Word-medially, OP \(\theta > MP \ h > NP \ h\). (Non-Persian Southwest Iranian dialects show the change \(^*\theta- > -h\) as well.) Word-initially, the picture is unclear. OP \(\theta-\) corresponds to MP, NP \(\dot{s}-\) across the board, with at least one exception, \(^*\dot{s}ata-p\dot{a}d(a)- \ ‘100 feet’ > NP \ hadba ‘centipede’\) (Morgenstierne 1932:55). Forms like this and others outside of NP proper have been thought to continue a hyper-Achaemenid dialect of Old SW Iranian (Lentz 1926:301 claims that Širāzī tanzīdan ‘weigh’, a cognate of NP sanjīdan ‘id.’, is such a dialectal form, connecting it to OP aθangainā/ī- ‘stony’). Either (i) the MP and NP forms are NW Iranian borrowings, or (ii) there was a change of the type PIr \(^*\dot{s}-\) > OP \(\theta-\) > MP \(s-\); the latter scenario seems to be the case, given forms like PIr \(^*\thetaaxta-\) > NP saxt ‘hard’ (cf. Khw \(\thetayd\)), and if PIr \(^*taigra-(či-)\) > Scythian \(^*θaigra-(či-)\) → OP month name \(θaigra(či-)\) (as per Lubotsky 2002:199) > NP \(s\dot{ir} ‘garlic’\). This correspondence was noted by Salemann (1901:263-264). If PIr \(^*a\dot{sanga}-\) > OP \(aθanga-\) regularly yields MMP \(sng /sang/ > NP \ sang\) (and it is not a NWIr loan), then a change of OP \(\theta-\) > MP \(s-\) may postdate the aphaeresis of unstressed vowels.

2.4.2.1.2 PIr \(^*ź > [\ddot{d}] > d\). The same caveats as above apply. The regular reflex appears in the following words:

- \(^*aźam ‘I’ > OP adam\)
- \(^*jaź- ‘worship, sacrifice’ > OP yad-\)
- \(^*źrd- ‘heart’ > MP dyl /dil/, NP dil\)
- \(^*źrāduni- (cf. Ved hrāduni-) ‘hail’ > SBshk dɔrāyen\)
- \(^*źrajah- ‘sea’ > OP drayah-, MP dry’(b) /draja:b/, NP daryā ‘river’\)

However, PIr \(^*ź > z\) in the following:

- \(^*jaź- ‘worship, sacrifice’ > MMP yz- /yaz/\)
- \(^*uαźra- > NP gurz ‘mace’\)

Forms of the latter type are usually attributed to a Northwestern source.

2.4.2.1.3 PIr \(^*šy > s\). Again, the same caveats as above apply. Some examples of this change follow:

- \(^*a\dot{š}ya-bāra- ‘rider’ > OP asabāra- > MP ‘sw’r > NP savār\)
- \(^*šu\dot{a}ka- ‘dog’ > MP sg /sag/, NP sag, Kmz sōy, Frs (Pāpuni) sag, Lur (Feyli) sag\)
- \(^*šuśi- ‘louse’ > (Pre-Kmz sōš) > Kmz śōš\)
- \(^*gau-śu\dot{a}nta- ‘holy cow’ > Kmz yosēn ‘sheep’\)
*šuunta aramati- ‘holy earth’ → Armenian (loanword from MP?) Sandaramet ‘hell, abyss’

New Persian shows not only sp (the reflex shown by all other Iranian languages besides SWIr, Sakan, Kurdish and Balochi), but sf:

*ašya- ‘horse’ > OP aspa-, NP asb

*šuiš- ‘louse’ > NP sepeš

*šuanta- ‘holy’ > NP espand ∼ sipand ∼ esfand ‘wild rue’

*šuunta aramati- ‘holy earth’ > MP spndrmt /spandarmad/ (cf. Arm Sandaramet)

*šuaita- ‘white’ > NP sipēd ∼ sīfēd

The change *šu > s also affects Kurdish and Balochi. PIr *šu also > h in Baškārdī, e.g., *šuiša ‘louse’ > heš. PIr *šu > t in Judeo-Širāzī, e.g., *šuiša ‘louse’ > teš. The h- may be secondary from *s-, e.g., Bshk yahmōn ‘sky’ (: NP asmān < PIr *ašman-), but it is unlikely that Judeo-Širāzī t- reflects *s- rather than *θ.⁵

2.4.2.1.4 PIr *θr > ç [s]. The same caveats as above apply.

*θrai- ‘three’ > MMP sh /se(:)/, NP sih

*a-puθrā- > Phl ‘pws /a:bus/, NP ābist(an) (+ *tāna- ‘offspring’), SBshk yōpes ‘pregnant’

Old Persian shows θr only in the name of the deity Mithra (Mθra ∼ Mtra), and this is likely a Median loan, but the regular reflex is ç, e.g., OP puça- < PIr *puθra-. Middle and New Persian frequently show hr, clearly from a Northwest source:

*šuθra- (cf. Ved śvitrā- ‘white’) > Phl spyr /spihr/ > NP sipihr ‘sky, firmament’ (if not from Greek σφαίρα ‘sphere, globe’)

*ciθra- > Phl cyhl /tʃihr/ > NP čihr ‘form, face’ (cf. Elam Tī-iš-šā-an-tam-ma = *Tiçantama- < OP PN Ciça’taxma- < *čθrǝntaxma-)

This change also affects Kurdish and Balochi, and similar-looking changes to š take place in Sogdian, Khwarezmian and Semnani.

⁵There may be evidence that speakers of Old Southwest Iranian dialects took the change *s > θ even further, as evidenced from Achaemenid-era Elamite Ti-ia-ma = OP (or a closely related dialect) *θavva- ‘black’, an “archaic equivalent of *Syāva-” (Tavernier 2007:316, 319, 330), depending on how we interpret this form and what we believe to be the primary Old Persian outcome of *š-.
2.4.2.1.5 PIr *θ̣i > š(i)y. This is a fairly solid Persian innovation, as seen in the following examples; the rest of Iranian shows no palatalization of *θ before a glide (though similar-looking palatalizations can be seen in far-flung East Iranian languages, e.g., *hraja-pəθ̣ja-kə > bəsk ‘own’, Humbach 1989:195, *əθ̣ja-(ka)- ‘duck’ > Yidyā yeχko, Rastorgueva and Ėdel’man 2003:1).

*həθ̣jam > OP hašiyam ‘truth’ acc. sg.

*həi-pəθ̣ja- > Phl NPŠE, hwyš, MMP xwy(b)š /xweʃ/ ‘self’

*pəθ̣a- (Skjærvø 1989:364) ‘before’ > MP pyš /peʃ/ > NP peš

It is not clear if non-Persian Southwest Iranian languages undergo palatalization; these languages tend to show ∅ (< *h?), e.g., *paθ̣a- > NBshk p̣i ‘before’, Kmz p̣i ‘from’ (?) (Skjærvø 1989:364); this cannot be secondary < *š, given forms like Kmz ṣ̌iš ‘louse’ < *šiš-, which preserve *š as such.

2.4.2.1.6 PIr *ši > θ(i)y. The fate of PIr *ši is somewhat up in the air in Southwest Iranian. Old Persian inscriptions show -θi- (e.g., vithiya ‘house’ loc. sg., frathiya ‘punish’ passive stem; Kent 1951:34 calls paštīyā ‘written text’ a “dubious example”); however, we see (what is generally interpreted as) šiy in some onomastic items attributed to Old Persian, e.g., Elamite Ši-ia-a-na ∼ Ši-ia-a-na ∼ Ši-ia-e-na = OP (?) *Šyaina- ‘eagle’, Elamite Ši-ia-ma = OP *Šỵava- ‘black’, alongside Elamite Ti-ia-ma = OP *θỵava- ‘black’, an “archaic equivalent of *Sỵava-” (Tavernier 2007:316, 319, 330).

We see a variety of reflexes in Middle and New Southwest Iranian languages:

- PIr *ši > s: *š(i)a-mrgə- (cf. YAv sačna- ‘eagle’, with dissimilation of the first glide?) > Phl synmxlw /sɔm məruw/ ‘a fabulous bird’, NP šinγ (MacKenzie 1971a:74); *kašapa- > Bandari kəsəpošt ‘turtle’ (+ pošt ‘back’, with perhaps haplology) (Rastorgueva and Ėdel’man 2003:IV 338);

- PIr *ši > s(i)y: *šịa-ga- > MP sy’h > NP sỵa ‘black’

- PIr *š < θ(i)y > h: *māš(i)xaka- ‘fish’ (Hoffmann (1976b:637, fn. 25) ascribes the long vowel to Vṛddhi) > Phl m’hyg /məhiɡ/ > NP māh, Band mūyi, Kmz mī, Lar ma’i; PIr *tus-sč(i)xka- > MMP twhyg/twihg/ ‘barren’

This knotty issue is discussed in further detail in §2.6.3.1.

2.4.2.1.7 PIr *źn > (x)šn. From what evidence we have, it looks as though PIr *źn became -šn- word-medially (e.g., *jāna- > Phl ʒən /ʒən/ > NP jašn ‘festival’, *gažna- > NP gašn ‘abundance’) and xšn- word-initially (e.g., *зван- > OP xšṇasa- ‘know’ inch. > MMP šṇṣ- /ʃṇaṣ/, Phl šṇṣ- /ʃṇaṣ/ > NP šinas- ‘recognize’). Some forms, likely loans from Northwest or East Iranian, contain zn, e.g., NP gavazn ‘deer’ (cf. MSog γwzn-, Khot ggünzna- < PIr *gau-až-na-).
The word-medial outcome of *-źn- distinguishes SWIr from NWIr, but from what evidence we have, initial *źn- > (x)śn- in NWIr, e.g., *źnā-sa- > *(x)śnā-sa- > Pth ‘śn’s- /i>jna:s/ ‘recognize, get to know’ inch., Awr ažnās- (with voicing), Gil šenās-, Xunsārī išnās-, Qohruđī ešnās- (Cheung 2007:466).

Avestan shows word and morpheme-initial xšn-, e.g., *fra-źnin- > YAv fraxšnin- ‘prescient (?)’, as well as word-initial žn- (e.g., *źnu- > YAv žnu- ‘knee’, *źnā-tar- > YAv žnātar- ‘knower’. Word-internally, it shows both -šn- and -sn-. According to Hoffmann and Forssman (2004:102), -šn- is the regular outcome, with -sn- (as seen in OAv yasna- < *iaź-na-) due to analogy. Schwartz (2010) proposes that -sn- is the elsewhere condition of the regular outcome, with -šn- found in rhotic environments.

Initial *źn- yields Khw ū-, e.g., *źnā-sa- > m/ū’s /u>as/ (?) ‘become acquainted, accustomed’ inch.; cf. *xšnaθra- ‘that which reciprocates’ (?) > ‘xnwry ‘thank’ (MacKenzie 1971b:532; Cheung 2007:467). These forms serve as evidence that *źn- and *xšn- do not merge in Khwarezmian (cf. Pth śnuhr → Armenian šnorh ‘thanks’).

Sogdian shows initial (or at least stem-initial) *źn- > (x)śn- (though *źnaka- > BSog ẓn’kh ‘knowledge, jñāna’ goes against this generalization), e.g., *fra-źnā-sa- > MSog fšn’s- ‘recognize’ (Cheung 2007:467); Sogdian has -zn- for medial *-źn-, e.g., *gaźna- > MSog γzn- /γzn/ ‘treasure’ — essentially the same distribution of NWIr.

Khotanese Saka has -zn- for medial *-źn-, e.g., vaysña/vazña/ ‘now’ < *a-ņa-źnā- (Bailey 1979:376). The behavior of initial *źn is unclear.

For New East Iranian, Ormuri shows n < *-źn-, e.g., injān ‘day before yesterday’ < *anja-āţna-. Some Pamir languages show -zd-, e.g., Sgl āluzd ‘id.’ < *a-unsafe-azna- (Rastorgueva and Ėdel’man 2003:1 295).

2.4.2.1.8 PIr *ja- rel. pron. → OP haya m., taya n., hayā f. This pronoun becomes the ezafe clitic, widespread across West Iranian. We have no information about the “Median” relative pronoun (and cannot reconstruct anything, since ezafe are highly reduced, usually to =i or =e).

2.4.2.1.9 PIr *paśč˘˘a- > OP pasā- > MP ps /pas/ > NP pas ‘behind’. Tedesco (1945:128, fn. 4) explains OP kaščiy, čiščiy ‘something, anything’ as NW Iranian loans (analogue restoration is unlikely to have given -š-, rather than -s-).

2.4.2.2 Less straightforward Southwest Iranian-defining innovations

2.4.2.2.1 PIr *št > st? SWIr shows št and st in variation (as well as *zd and *zd, e.g., mužd ~ muszd ‘reward’ < *mižda-). Skjærvø (1989:364) considers SWIr to be defined by a phonological development *št > st, and that all instances of št (e.g., OP h-u-frašta- ‘well punished’ against h-u-frasta-) are due to NWIr admixture. Kent (1951:34) argues that the change is analogical with past participles of dental stems, operating within Old Persian (and not serving as a diagnostic of a SWIr subgroup). I find Kent’s view (or at least an analogical account) more tenable in some cases, particularly in accounting for forms like *źaušt˘˘a- >
OP *daušt-ā- ‘friend’ (> MP *dust /dost/ > NP *dost), which shows the Southwest Iranian reflex of PIr *ź and would have to be viewed, if not as a form unaffected by Kent’s analogical change, as a mixed form showing both Persian and Median phonology (this phenomenon is seen in Achaemenid compounds, e.g., Elamite mi-iš-ba-tan-na = Med *vispa + OP *dana- ‘containing all tribes, all types of men’ (Tavernier 2007:34, 78), and perhaps in simplex forms like OP *ganda- as well, on which see 2.6.3.1). However, st forms like Phl/MMP *ngwst /anqust/ ‘finger’ (< *angušta-, alongside NP *mušt, as well as Phl *mуст ∼ mušt ‘fist’, NP mušt < *mušti-), Phl b’lyst /ba:лист/, MMP b’ryst /ba:rist/ ‘highest, summit’ are unlikely to be due to analogical change. The extent and mechanism(s) of this change remain unclear.

2.4.2.2.2 PIr *śr > ç? There is scant evidence for this development (which is phonologically parallel to *θr > ç), e.g., ç in *ni-śrai ‘restore’ > OP niyaçray-caus. (× daraya- ‘hold’ caus., Kent 1951:188), MP *sry’ /nisa:j/ ‘conveying, dispatch’. Middle and New Persian often show sr:

- *śrau̯aia- ‘hear’ (caus.) > *śrau̯aia- > MP sr’y /srâj/ ‘sing’, NP surû(dan)/sarây
- *śrau̯ni- > NP surûn ‘buttocks’
- *hūaśrū- > NP xusrû ‘mother-in-law’
- *aśra/u- ‘tear’ > MMP ’rs /ars/ > NP ars (also MMP ’sr)

In some words, NP shows š (possibly loans from a language like Sogdian):

- *hūaśrū- > NP xušû ‘mother-in-law’
- *ašruka- > NP ašk ‘tear(drop)’

2.4.3 The rest of Iranian

The remaining Iranian languages (i.e., all non-Southwest and non-Sakan languages) show some interesting innovations:

- PIr *šu > sp. Can change secondarily to šp (e.g., in NP šipiš ‘louse’, via assimilation) ∼ šb etc. Ossetic has fs. Yazdī (a NW Iranian language) has sv, probably secondary (PIr *sp > sv as well), rather than an archaism. NP shows sf in some strata. Balochi and Kurdish are unaffected by this rule, agreeing with Persian in their outcome (s).

- PIr *źu > zb (generally). We see some secondary reflexes like Ossetic vz and Kurdish zm. We often see zw as well, but it’s hard to gauge whether this is an archaism or a secondary lenition.
2.4.4 Other postulated dialectal affinities

A number of other small subgroups and dialect groups have been posited within Iranian.

Beekes (1997) remarks that Semnani agrees with Median, due perhaps to the sound change \( *hu > f \) (this change exists in many New Iranian dialects found in Media Major, but the status of this change in Median, a fragmentary language, is controversial; for discussion, see Skjærvø 1983).

Azami and Windfuhr (1972:36-37) note a small number of isoglosses between Khwarezmian and Sangesari, most notable of which is the change \( *fr > š \) (also seen in Baluchi).

Yaghnobi has been paired with Sogdian (Bielmeier 1989), based primarily on shared lexis and shared morphological features, such as plurals in -t, the vocalization of \( *r- > u-\) in the present stem of ‘do’ (Sogdian kwn-, Yaghnobi kwn— however, similar developments take place in Old Persian kunautiy 3sg pres, Bal kanag— and others). At the same time, Yaghnobi shares a key morphological isogloss, the 3rd person plural in r (Ygh -or), with Khwarezmian and Khotanese Saka (Schwartz 1969). Furthermore, the Sogdian collective marker -t(’)/t(a:)/ agrees with with Ossetic -tæ; a cognate also appears to be present in some Sarmatian items, e.g., the tribal name Σαρματαί ‘Sarmatians’.

The fragmentary Middle Iranian languages Sarmatian, Alanic, and Jassic show a strong affinity with Ossetic. The Old Iranian language Scythian appears also to be dialectally related, but there is less evidence and hence less agreement. Notable traits include the change \( *rj > l \) (as well as the change \( *r > l \) in the environment of a high vowel) and spirantization of initial \( *p-\). Scythian attests the latter, but there is no evidence of the former. There is remarkable congruence between some of these languages in the greeting ‘good day’: Jassic daban horz (Németh 1959:14), Alanic ταπαγχὰς, cited in Tzetzes’ Theogony, Oss I dæ bon xorz, D dæ bon xposéarz (Kim 2003:54-55).

Old Avestan is generally taken to be the direct predecessor of Young Avestan, since there are virtually no Old Avestan innovations that are not shown by Young Avestan as well (de Vaan 2003:8-10).

2.5 East vs. West Iranian

While the inter-relationships discussed in the previous section are generally accepted, there is no complete and final consensus on whether East and West Iranian are genetic groups. While most Iranists are far from agnostic on the issue of East vs. West Iranian subgrouping, it is difficult to find a terse pronouncement on the first-order branching of Iranian in the handbooks, etc. Some opinions found in the literature follow.

2.5.1 No Eastern subgroup

Sims-Williams (1996:651) states that East Iranian is not a genetic grouping, but a Sprachbund; most of its shared characteristics are retentions, rather than innovations, and the innovations that it shares are relatively trivial. Wendtland (2009) finds that there are no
shared phonological or morphological characteristics between the East Iranian languages, and argues against Northeast and Southeast subgroups (a division provisionally suggested in Morgenstierne 1926 and followed in Oranskij 1977, Kieffer 1989 and elsewhere).

2.5.2 Southwest, Central, and Northeast, and Southeast subgroups

Skjærvø (2009b:50-1) says that “Proto-Iranian split into at least four distinct proto-Iranian dialect groups.” These groups are characterized by the treatment of the palatals and the clusters *śu, *ẓu. He describes an Old Central Iranian “represented by most of the remaining dialects” that has undergone the changes *śu > sp, *ẓu > zb. This is somewhat problematic if Kurdish and Balochi are taken to be descendants of Old Central Iranian, since Kurdish and Balochi have not undergone the change *śu > sp (showing s instead; they could not have undergone a change *sp > s, since Old Iranian *sp is preserved as sp). Attempts to put Kurdish and Balochi in the Southwest Iranian group run into the same problem, as *s > s, *ẓ > z against regular Southwest h (< *θ), d.

2.5.3 Western vs. Eastern subgroups (+ Avestan)

Kuiper (1976:251) claims that “[t]he term ‘East Iranian’ presupposes that at Zarathustra’s time ... the eastern dialects were already so much differentiated and geographically apart from West Iranian as to justify the distinction.” Beekes (1988:10) also says, “On one side we can reconstruct ... Proto-East-Iranian as distinct from West Iranian.” However, not much is done to substantiate these views.

2.5.4 Western vs. Eastern subgroups (− Avestan)

Oranskij (1977:197-202) gives a number of isoglosses differentiating East Iranian languages (excluding Avestan) from West Iranian ones (this presupposes that Avestan was the first to branch off from the larger body of Iranian; cf. also Schmitt 2000:63). These require Ormuri and Parachi to be (re-re)classified as West Iranian. These include

- Lenition of PIr *b(r)-, *d-, *g- in East Iranian
- Depalatalization of PIr *(-)č- in East Iranian
- Loss of PIr *h- in East Iranian
- Voicing of PIr *-ft-, *-xt- in East Iranian

To this catalog of innovations affecting East Iranian (but not Avestan), I would add an additional development:

- PIr *-θn- > n (this includes Ormuri and Parachi, but perhaps not Wakhi, e.g., aršt ‘elbow’ Rastorgueva and Édel’man 2003:I 214, if < *araθni-)
Many of these changes are trivial, and likely to recur multiple times; some do not cleanly and completely define all of East Iranian. Édel’man (1992) rejects most of these isoglosses, but retains two: the lenition of initial voiced stops in East Iranian, and the voicing of PIr *-ft-, *-xt-.

In the following subsection, I seek to address Édel’man’s claims, with an eye to the following questions:

1. How well are these claims borne out by the data? Are proposed innovations for a Proto-East or Proto-West Iranian supported by all languages in each putative genetic group?

2. How falsifiable are these claims?

3. Are these features strong diagnostics of subgrouping?

I find that there are some exceptions to one of the proposed innovations, though they are few in number and somewhat obscure. Not only this, but Édel’man relies on multiple minority-view assumptions in order to argue for an otherwise implausible mechanism of change. The remaining innovations are somewhat weak as diagnostics of subgrouping, consisting of sound changes that are likely to recur, as well as a few of lexical isoglosses, which could be due to areal diffusion rather than indicative of joint genetic development.

2.5.4.1 Lenition of PIr *b(r)-, *d-, *g-

This development is not unambiguously complete across East Iranian. Ossetic shows γ- in the conservative Digor dialect (Iron q-), but no lenition in b- or d-. Ossetic b- could conceivably be due to a secondary fortition from *β-, but there is no evidence for a fortition that would affect other fricatives or glides such as *γ- (e.g., Oss uad ‘wind’ < PIr *γατα-). Khotanese Saka has b- [β] and d- [δ], but gg- [g], unambiguously a plosive; Tumshuqese Saka has b- and d- for *b- and *d- “despite possessing special symbols for the corresponding fricatives” (Sims-Williams 1989c:168). Wakhi shows lenited initial consonants. Again, the Khotanese (and perhaps Tumshuqese, depending on what the orthography represents) forms could show secondary fortition from a hypothetical Proto-EI *γ- (as well as *β- and *δ- in the case of Tumshuqese). Yaghnobi shows d-, plausibly from *δ- (cf. Sogdian). Ultimately, it is not entirely clear whether this development operated across East Iranian, given various secondary developments that have subsequently taken place. Suffice it to say, fortition and particularly lenition are rather trivial sound changes, highly likely to recur multiple times in the history of a single language much less across closely related ones. The fact that lenition happens word initially and not necessarily medially, while striking, does not mean that the change cannot have been recurrent: supporting this idea is the fact that Yazdi, a West Iranian language,

---

6Fragments of Scythian, thought to be closely related to Ossetic, show an orthographic merger between reflexes of PIr *γ- and *b-, e.g., *baγμαρ-ασγα > Бαγμαρ-ασγα ‘(having) myriad horses’, *γανατ-ασγα > Бαγανατ-ασγα ‘(having) victorious horses’ (Harmatta 1970:86); whether or not there is a one-to-one letter-to-phoneme correspondence is unclear.
appears to undergo lenition of initial *g- to y- in words like yūz ‘walnut’ < *gauža- (Vahman and Asatrian 2002:24), while intervocalic *-g- is preserved as g (according to Ivanow 1940:46, though he does not provide particularly good evidence of this). For this reason, lenition of initial voiced stops does not serve as a thoroughly convincing subgroup-defining innovation.

2.5.4.2 PIr *-ft-, *-xt-

East Iranian languages tend to voice PIr *-ft-, *-xt-; West Iranian languages tend to devoice PIr *bd, *gd. As a phonological development, this is not a particularly strong diagnostic of subgrouping; voicing is generally thought of as a low-level sound change, likely to recur multiple times (this is in essence the view of Sims-Williams 1996 on this sound change). Edelman (1992) however interprets this as an analogical phenomenon stemming from changes to clusters that were affected by Bartholomae’s Law in Proto-Indo-Iranian. If this scenario is correct, it would mean that East and West Iranian split, and then underwent separate, subgroup-defining innovations.

The three-way phonation of Proto-Indo-Iranian occlusives yielded past participles of the following types:

\[
\begin{align*}
T + t & \rightarrow T(s)t \\
D + t & \rightarrow T(s)t \\
D^h + t & \rightarrow D(z)d^h
\end{align*}
\]

Old Avestan preserves BL across the board,7 e.g., aogodā ‘utter’ 3sg inj. (< PIIr *Haugdha ← **Haug-tha); this is generally undone in Young Avestan, e.g., aoxta ‘id.’ (with some exceptions, e.g., vərəzdā- ‘grown’: OInd vrddhā-, ubdaēna/ī- ‘wool’). This analogical undoing makes sense: Iranian lost the distinction between modal and breathy voice, and speakers extended the pattern shown by roots ending in D to roots ending in etymological Dª (in the case of PIIR -b-ª, they would have to metanalytically rely on the pattern shown by -p-, since no roots ending in PIE *-b- survived in PIIR). Edelman argues the following:

1. PIIR *bh, *dh, *gh had fricative allophones [β], [ð], [ɣ] (p. 53) which were preserved in East Iranian and Avestan (keeping a distinction between PIIR D and Dª).

2. After Avestan branched off, East and West Iranian split; West Iranian undid BL (on the basis of the voiceless and modally voiced series), while East Iranian extended BL to the voiceless and modally voiced series

Some complications stemming from this view follow. First, voicing affects underived East Iranian forms like ‘seven’, e.g., Oss avd < *hafta; Edelman doesn’t explain the mechanism by which voicing would be extended from past participles to a numeral.

Additionally, BL didn’t just occur in past participles; suffixes beginning with -s- were also affected by its operation (and subsequently “restored”). In East Iranian (as well as in West Iranian), patterns from voiceless and modally voiced stops are extended to breathy voiced stops.

7Though Rix et al. 2001:658, fnm. 4–5 argues that the t of OAv vaft- ‘sing’ goes back to a participle *y(a)fta- ← **y(a)bº-tha in which BL was undone; additionally, if OAv duuafša- ‘torment’ is to be connected to Greek tuptlós ‘blind’; Old Irish dubh ‘black’ (Duchesne-Guillemin apud Kellens and Pirart 1991:II 263) with some sort of Schwebeablaut, then BL has been undone in this form as well.
stops (Dʰ-s → T-s on the basis of T-s, D-s), e.g., PIr **gṛbʰ-sća- → *gṛf-śa- > Khwarezmian γfs- ‘be quiet, keep silent’ (Cheung 2007:120). It’s surprising that we don’t see a single instance of this putative morphological development here, though it could also be the case that BL-affected inchoatives such as YAv uboj- ‘press down’ (not recognizably an inchoative) were too morphologically opaque for this pattern to be generalized to newer forms reflecting the same verbal suffix.

There is additionally the question as to whether this change operated completely within East Iranian. Certain Sogdian clusters, written βṭ, γṭ in the Manichean script, are interpreted either as fully voiced (Gershevitch 1954) or partially voiced (Sims-Williams 1989e:179), (perhaps a sort of reverse of Avestan clusters written fsδ, xδ (cf. Monna 1978, passim)). Yaghnobi devoices these clusters entirely. If these clusters are partially voiced, then partial voicing must have happened at a relatively late date, since these clusters have not merged with clusters that did historically have partial voicing but were ultimately devoiced, e.g., *baga-tama- > *baytam > MSog βṭtm /vaxtam/'devātidēva' (Gershevitch 1954:194).

Furthermore, Khotanese Saka may not show complete voicing in all forms where we would expect it, e.g., autta-/otta/ ‘reached’ < *āfta-: OInd āp- (?)(Emmerick 1989:215); patāvutta ‘shaven’, connected to OInd vap- ‘shave’ (Bailey 1967:38). This is unlikely to be an Indic loan, since Gandhari Prakrit fully assimilates heterorganic clusters, e.g. sataṁma ‘7th’ < OInd septamā-, as do most Prakrit varieties (cf. Pk. atta ‘obtained’), and would have preserved no trace of the labial element. The form autta occurs alongside in a doublet with byauda ‘id.’. Bailey (1967:38) argues that autta is devoiced because of the long vowel in *āfta-, whereas byauda < short *api-āfta-. This historical phonology seems odd and ad hoc. It is well known that voiced consonants can phonetically lengthen the duration of the preceding vowel, but I know of no claims that long vowels can serve to preserve voicelessness in a following consonant. Furthermore, this claim does not explain the lack of voicing in patāvutta, which probably reflects *-ufta- < *-afta-, not *-uāfta-. It is possible that these forms are borrowed from a peripheral East Iranian dialect in which voicing did not take place. The possibility that this form is West Iranian seems unlikely.

Additionally, BL doesn’t seem to be fully undone in West Iranian either, e.g., Bal šabt- ‘tup (a ewe’ (< *fra-abbixa; cf. *hafta- > hapt ‘seven’).

In short, the most parsimonious scenario is one in which BL broke down (a widespread but perhaps not complete trend across Iranian), and most East Iranian dialects voiced *ft and *xt — a phonological (not morphological) change, and one not particularly diagnostic of subgrouping.

### 2.5.5 East vs. West Iranian lexis

Some lexical (and lexeme-specific morphological) isoglosses appear to separate East and West Iranian (for further detail see Sims-Williams 1996:651). Many previously proposed East Iranian and West Iranian vocabulary items have been discarded, since they have been found to bleed across East-West lines, such as the following archaisms:

- East Iranian *gari- ‘mountain’, found also in MP gar ‘id.’, Bal gar ‘abyss’
- East Iranian *anda-* ‘blind’, but also Pth hnd ‘blind’ /hand/ (perhaps attributable to the Parnian East Iranian element; see Sims-Williams 1989c:171), Zaz agil-hend ‘truth blind’ (Paul 1998a:175), the first element ← Arabic

An innovative form, *kaufa-*, is discussed below.

### 2.5.5.1 PIr *kaufa- ‘mountain’

Handbooks tend to consider PIr *kaufa- the typical West Iranian term for ‘mountain’, against the Eastern archaism *gari- (Sims-Williams 1989b:169; Sims-Williams 1996:651). The traditional etymology given is PIIR *kaupʰa- (Bartholomae 1895:8). Rastorgueva and Édel’man (2003:IV 371) connect *kaufa- to PIE *keu-p ‘bend’ (Pokorny 1969:no. 950, with a number of enlargements, e.g., *keu-k, *keu-p, *keu-bʰ-, etc.). The *-f- is then irregular. This irregular historical phonology could further support the idea that *kaufa- was a West Iranian innovation, were it not for the fact that it is attested in Young Avestan as ‘dome-shaped mountain’ (Humbach and Ichaporia 1998:73), e.g., kaôf¯ o nom. sg. (Yt 19.3), kaafaŋm gen. pl. (Yt 14.21).

A bit of clarification is needed on the derivational history of this form. Rastorgueva and Édel’man (loc. cit.) ascribe the variation between the reflexes *kaupa- and *kaufa- to Pre-Iranian dialectology and areal tendencies in various later Iranian languages. Alternations between *-p- and *-f- are virtually unattested, not nearly as common as alternations between *-k- and *-x-, e.g., *madakã- vs. *madaxã- ‘locust’, *sûrãkã- vs. *sûrãxã- ‘hole’, *ôiaka- vs. *ôiixa- ‘egg’, *ahmûka- vs. *ahmûxa-, etc. (Klingenschmitt 2000:203, fn. 40), so parallels would be welcome.

It seems likely that PIIR *kaupʰa- adheres to the same sort of derivational pattern as PIE *roth₂o- ‘thing with wheels, chariot’ ← *rot-eh₂- ‘wheel’, e.g., *koup-eh₂- ‘hump’ → *kouph₂o- ‘thing with a hump/peak/dome’. Regardless of which of the above scenarios is correct, the presence of this lexical item, as well as the irregular *f- it contains, has no real bearing as a West Iranian innovation, as Avestan is affected (and, for what it’s worth, its derivation likely predates Iranian).

### 2.5.5.2 Real East-West lexical isoglosses?

Below, I discuss three lexical items that fall on either side of the East-West line, taking into account their semantic and derivational histories. The goal here is to get a feel for how unique or unusual these items are, and whether we can confidently take them to be shared genetic innovations.

### 2.5.5.3 East Iranian

#### 2.5.5.3.1 *maiðã- ‘day’

PIr *maiðã- ‘day’ > Shg meθ, Yzg miθ, Zebâkî mî, Ygh met, Yidyã-Munî miç, Sanglîčî miç, Orm miθ (prob. a Shughni loan), Buddhist and Manichean

---

*Thanks to Michael Weiss for this suggestion.*
Sogdian *myδ, Christian Sogdian *myθ and Khwarezmian *myθ /meθ/. Morgenstierne (1938:229) argues against the diphthong *-ai-, since Sariqol¯ı has the monophthong mǎθ (cf. məyz ~ məyz ~ məyz ‘urine’ < *maiz- (Paxalina 1971:267)), and instead sets up a protoform *məθja-. He gives references connecting the form to Lithuanian mėtas, Albanian mot ‘time’, a cognate of Old Indic māti- ‘measure’ (Demiraj 1997:278). PIr *məθ(i)ja- would be a thematized derivative of an abstract noun *məti-. It is not clear if this form works for other languages, e.g., Shg xᵲbaθ ‘him/herself’ < *hyai-pəθia- (though *-a- is short here, Zarubin 1960:363). On its own, Sogdian *myθ looks as though it could have been subject to the palatalization targeting stressed syllables seen in words like zyn ‘gold’ < *zárani-, pyrδnn ‘saddle’ < *párí-dana- (Sims-Williams 1989:e:181), but it’s not clear that this account explains the shape of the other forms. If we assume an original *məθ(i)ja-, metathesis to *maiθa- must have occurred in Khwarezmian, for which PIr *məθ(i)ja- < xbsk ‘own’ (Humbach 1989:195).

Morgenstierne (1974:45–6) rejects the Lithuanian connection, and posits a protoform *məθa- which he connects to Avestan məθa- “interchanging (with the night??)” (loc. cit.). This semantic explanation seems reasonable enough.9

Many of these languages seem to preserve the timespan sense, e.g., Buddhist Sogdian *str’ myθ ‘later in the day, the next day’ (Gershevitch 1954:13), Yidya do yū mixel ‘in one day’ Morgenstierne (1938:loc. cit.) (but Srq məθ wand(εw) ‘live’ [lit. ‘see day’] (Paxalina 1971:252)).

In short, it seems most likely that the Sogdian, etc. word is formally equivalent to OAv məθa-, and has developed the semantically natural sense of ‘day’, possibly independently, in each of these languages. Sariqol¯ı reflects *ma(i)θia-, and is thus alone in undergoing a morphological innovation of the type diagnostic of subgrouping. The rest of East Iranian shows a semantically natural development without any sort of non-trivial derivational change.

2.5.5.3.2 *kapā- ‘fish’ *kapā- ‘fish’ > SSog kp (nom.sg. kpy < *kapah), Khw k(y)b, Khot kava, Wkh kyp, Psht kb ‘fish’; Modern Ossetic has kæsag, replacing an older form kæf (cognate to the other Iranian forms), found in early Ossetic versions of the Gospels of Mark and John (Bailey 1945:22), and the compound iæv-gæf ‘caviar, roe’ (lit. “fish millet”). Cf. also the Scythian hydronym Παντικ´απης < *panti-kapa-, lit. ‘fish path’; a connection between these forms and Elamite Ka₄-ab-βa has been tentatively proposed (Tavernier 2007:225), but there is no independent evidence to support this interpretation.

Abaev (1989:I 576) links the word to Vedic kapan´¯a-, Latvian kāpe ‘caterpillar’, noting the putative cognacy between Slavic *ryba (< *rūba-) ‘fish’ and German Raupe ‘caterpillar’ (however, the connection with kapan´¯a- does not appear in Mayrhofer 2001:I 299). Even if the Slavic and German forms are not cognates (as per Orel 1995:164), Old High German ruppe, rûpe ‘caterpillar; eelpout’ exhibits polysemy indicating a natural semantic connection

---

9The following step is unnecessary, but worth mentioning: if we accept Insler’s (1971:174) reading of OAv məθa- ‘enduring’ at Y33.9, we can capitalize upon the fact that the meanings ‘day, 24-hour timespan’ and ‘live, dwell, pass the night’ share a natural metonymic relationship (SPEND DAY/NIGHT ↔ DAY/NIGHT), cf. Written Tibetan žag ‘day (24 hours)’ and its Old Chinese cognate *sjök ‘lodge the night, pass a night’ (Matisoff 2003:323, 328).
It is not clear whether the -f in Ossetic *kæf is via regular historical phonology, given the intervocalic changes to *-p- seen in *upa- > *ba-, *tāp- > *tāv(y)n ‘to heat’ (Thordarson 2009:154). Cheung (2002:18–9) claims that PIr *p “occasionally” yields postvocalic final -f, picturing a development *p V > *b V > *v > -f.11

It is not entirely clear what the suffix of this etymon is. MSog kp-, CSog qp- (light stem) and Khw kb (spelled kyb in pausa) point to a short final vowel, but Bailey (1979:56) derives Khot kava- (n.sg kavā, kava, pl. kave; adjectival kavaṁie, kaviṁe, kaviṁām) from *kapā-. from *kapā-; the plural in -c indicates that this form is a member of the ā-declension (Emmerick 1989:218). And it is not clear if the Vedic form (if cognate) contains a secondary enlargement, or exhibits morphology that was lost by Iranian cognates.

As with *maiθa-, the semantic change of ‘caterpillar’ ⇒ ‘fish’ may have parallels elsewhere, making it not a particularly strong diagnostic of subgrouping. Even if we reject this etymology, we are left with a lone lexical item that may or may not be a shared innovation.

2.5.5.4 West Iranian

2.5.5.4.1 *hiźu- (against other Iranian *hiźu(ā)-) ‘tongue’  West Iranian languages agree in terms of suffixation in the word for ‘tongue’. This feature was observed by Morgenstierne (1938:425), who suggested the influence of *dantan-.

All Iranian languages continue PIr *hiźuH(ā)- ‘tongue’.12 Some Iranian forms are unsuffixed, e.g., *hiźū- > YAv hizuuā-, Psht ḏōba (*-aka- should become Psht -ay), Shg zīv. We also see the *-aka- suffix across most of East Iranian, a widespread Indo-Iranian feature: PIr *hiźūka- > Sogdian ḏβ’k, C žb’q, Khwarezmian žb’k, Khotanese Saka ḏīśāa- /βīṣā/, Yzg zveg, Ygh zivok, Wkh zik, Zābāki zevuk, Ossetic ævzæg; suffixation with *-aka- is a widespread Indo-Iranian tendency and doesn’t really qualify as an innovation that is diagnostic of subgrouping. Parachī (zu)bān and Ormuri zubān are marked as loans by Morgenstierne (1929:302, 413). All West Iranian words for ‘tongue’ reflect PIr *hiźūāna-, e.g., OP ḏ-z-a-n-m (acc.sg.), MP ‘uzu’ /uzwam/, Pth ‘zb’n /azba:n/, NP zabān, zuwān, Kumzārī zuwān, Lar zabu (with loss of final *-n? cf. dudu ‘tooth’ < *dantān-), Kd azmān, zimān, Zaz zuwān, Maz zevon. Given this distribution, it is possible to reconstruct a Proto-West Iranian form *hiźūāna-, an innovation against the rest of Iranian. But given the fact that this is a lone lexical item (and that lexis is notoriously areal), it would be unwise to place much faith in this isogloss as a diagnostic of subgrouping.

Various analogical accounts of *hiźu- have been proposed. As mentioned above, Morgenstierne cited the possible influence of *dantan- ‘tooth’. Lommel (1922:261) suggests that the acc. sg. suffix -ānam was extended from OP *dāfānam ‘mouth’ acc. sg. (cf. YAv zafr-

---

10 Ved kapanā- may derive from a verbal root originally meaning ‘to advance upon, change levels of movement’ (Schwartz 1969:446).

11 Both kef (which he derives from *kapā-) and keref ‘greedy’ (dubiously connected with OAv kar(a)pan-) are cited, but the latter etymology should probably be thrown out, semantics aside; since *xšapā-, with its long final vowel, yields I. æxesae, D. æxesæe ‘night’, with -v-, it remains unclear how an -an- stem could yield the proper conditioning environment for final devoicing of *-v.

12 On the irregular historical phonology of this form, see Fay 1895:ccxxvii; Lipp 2009:I 188–90.
‘mouth’, acc. sg. thrizafanom ‘having three mouths’), *dantănam ‘tooth’ acc. sg. (> NP dahān, dandān), influenced by gen. pl. *-ānām. Vis-à-vis this suggestion, Kent (1943:227) says, “...of necessity this remains a mere speculation. It is a curious coincidence that Old Latin díngua (classical lingua) ‘tongue’ is an ā-stem, but the Germanic cognate exemplified by Gothic tugó has been transformed into an -n-stem; yet this can be hardly taken as evidence that the OP word for ‘tongue’ was a stem in -n rather than a stem in -na-, for the Germanic names of the ‘ear’ and the ‘eye’ also had become -n-stems, and the corresponding Iranian words suffered no such alteration.” It is worth noting that in its sole OP attestation, hizāna- unambiguously means ‘tongue (organ)’: u-t-a n-a-h-m u-t-a g-u-š-a u-t-a h-z-a-n-m f-r-a-j-n-m ‘I cut off his nose and ears and tongue...’ (DB II 74). Given this concreteness of meaning, it seems unlikely that the genitive plural exerted any influence, as it might in a semantic development like ‘of the tongues’ ⇒ ‘language, speech’, for example.

The generalization of an-stem morphology to buccal words appears to have been common in Old Iranian, and has at least one parallel in Old Indic as well. While Old Avestan has only the neuter root noun āŋh- ‘mouth’ (Y 28.11 ōāŋhā inst. sg., Y 31.3 āŋhō gen. sg.), Young Avestan has āŋhan- ‘mouth’ nt. (V 3.29 āŋhanō gen.sg.). In RV, we see ās- alongside āsān-; it is not clear, however, whether the an-stem is reconstructible to PIIR. In Young Avestan, we also see vimitō. dantan- ‘having deformed teeth’ (V 2.29 vimitō. dantanō nom. pl.).

It seems like the most likely locus of the extension of an-stems is PIIR *ās-. The locative *ās-ān could have served as the basis of an an-stem paradigm, possibly in Proto-Indo-Iranian, or separately in Iranian and Indic (given discrepancies in vowel length between YAv āŋhanō gen.sg. and Ved āsānī loc.sg.). This inflectional pattern could then have been extended to YAv dantān-, and also to *hiźuān- in some Old Iranian dialects. Forms like heteroclitic YAv zafar/n- ‘mouth’ along with derivatives like thrizafana- could also have provided influence. Ultimately, while *hiźuāna- is reconstructible as a lexical-morphological isogloss for West Iranian, it is unlikely to represent a subgroup-defining innovation, as extension of an-stems seems to have been a natural tendency in Iranian with several parallels from Avestan.

2.5.5.5 Summary of East Iranian vs. West Iranian shared innovations

Ultimately, there are not too many isoglosses that can be reconstructed in favor of a Proto-East or West Iranian. We have a couple of phonological ones (the completeness of which are in question), and some lexical items, which could have been diffused areally. A small number of lexical isoglosses do not a subgroup make. We do not posit a special genetic relationship between Persian, Sogdian, Khwarezmian and Mazandaran (cf. Borjian 2008:74) on the basis of lexical innovations shared by the three languages (such as the use of the root *gaub- ‘speak, praise’), so it is not clear why we would assign the same weight to the above items.

2.6 East Iranian vs. West Iranian historical trends

There’s not much of an empirical basis for a reconstructible Proto-West or Proto-East Iranian. But these labels still make sense. Why is that the case? For one thing, we see a number of
trends in historical phonology and morphology that are not reconstructible for the purpose of subgrouping. It is not always clear if these developments are due to parallelism or contact. In the upcoming sections, I list a few innovations of this type common to (or widespread within) West Iranian.

2.6.1 West Iranian treatment of *θ

Old Persian and Median preserve Proto-Iranian *θ (though in some contexts on Old Persian, it is palatalized). However, intervocally, *θ > h in most West Iranian languages (Balochi undergoes fortition to t). This is in sharp contrast to East Iranian, where *θ is overwhelmingly preserved as an obstruent. This asymmetry stands out. Could there have been something about the phonetic quality of a hypothetical Proto-West Iranian *θ that was different from its counterpart in Proto-East Iranian? Or perhaps this is the wrong question to ask, given the diachronic instability of [θ]. Was a similar factor at play in East Iranian, or was there some sort of sociolinguistic pressure responsible for the longstanding preservation of *θ in these languages?

An additional confound is the fact that *θn becomes n across East Iranian (Wakhi being an exception), whereas West Iranian languages preserve the fricative element (there may be exceptions to this claim; see below).

2.6.2 West Iranian treatment of *θn

Southwest Iranian has šn for *θn:

- *araθnī- > OP arašni- ‘cubit’ > Phl ḷšn MMP *ryšn /arɛfn/ > NP ăreš(n)
- *dmāna-paθnī- > MMP b’nbyšn /bambiʃn/
- *ham-paθnī- > NP ămvasnī (?) ‘rival wife’

We also see Phl ʼlnc > NP āranj ‘elbow’, perhaps from a source closely related to Sogdian (cf. ʼryan).

The clearest evidence of the outcome of *θn in NW Iranian is seen in the Babylonian Akkadian transmission of a Median PN: Akk Pa-at-ni-e-ša = Med *Paθnīeša- < *Paθnī-aša- 'looking for [desiring?] a wife' (Tavernier 2007:273). The t points unambiguously to *θ, indicating that the cluster was unchanged in Old NW Iranian. Gk Πισσούθνης = *Pišiθna-

---

13 Proper names like Mithradates and demonyms like Part’eew are generally taken as evidence that “an older stage of Parthian” (Korn 2005:81) preserved *θ (Sundermann 1989:123).

14 Schwartz (1974:401) argues that MMP d’š(y)n /daʃʃin/, Phl d’šn (as well as NP dāšan?) cannot reflect *dəθna- due to spelling, and because MPth d’š(y)n cannot be borrowed from MP. He argues for an etymology dāšinen- ‘that which is presented or indicated with the right hand’, cf. OInd daksinā- ‘right (handed)’ → daksinē- ‘gift due to priest (for performance of a rite)’. Cf. also Bct λαψνο ‘gift’ (Sims-Williams 2007:226).
(with haplology, cf. YAv Piši.šíiaodna-, a bahuvrhihi with the second element meaning ‘actions’ and the first unknown,\textsuperscript{15} Benveniste 1966:123-5).

The outcome in later languages, however, is more of a mystery, due to secondary changes, the paucity of evidence, and the possibility of borrowing from Persian. For Middle NW Iranian, Manichean Parthian has $d's(y)n = \text{MMP } d's(y)n$ (though perhaps not a reflex of *$d\tilde{\theta}na-$), $b'ny\tilde{s}n = \text{MMP } b'ny\tilde{s}n$. However, given that these words are found in the bilingual Manichean Turfan texts, it’s entirely possible that these forms are actually loans from Middle Persian.

For New NW Iranian, Rastorgueva and Édel’mann (2003:I 213) give Bal hariš, harš, harša and Awr aražni ‘cubit’ ($*\tilde{s}n- > \text{Awr } \tilde{z}n$). Korn (2005:154, fn. 406) says that the Balochi forms must be NP borrowings because fricative + nasal clusters are assimilated to the nasal, e.g., $*\tilde{c}a\tilde{m}s\tilde{m}- > \text{cam(m) } ‘\text{eye}.\text{ Benveniste (1935:105) derives the Zazaki } -i\tilde{s} \text{ infinitive suffix } -i\tilde{s} \text{ from } *-\tilde{\theta}na-$ (with *$n$ lost, and a preceding thematic vowel).

A number of Iranian loans into Armenian show the change $*\theta n- > n$. Gippert (1993:348) identifies Armenian danak ‘knife’ as a borrowing of NW Middle Iranian $*\text{dahnak } < \text{P}Ir \ *\text{d\tilde{\theta}na-}$ (alongside Arm dašnak ‘id.’ $< \text{SW Middle Iranian } \ *\text{da\tilde{s}nak } < \text{P}Ir \ *\text{d\tilde{\theta}na-ka-}$, cf. NP dašna ‘dagger’); however, with no actual Iranian attestation of this word (in this particular phonological shape), it is difficult to be sure of its linguistic provenance (could it perhaps be a “Parnian” item? See Olsen 2005 for a discussion of possible East Iranian loans into Armenian). Hübschmann (1897:I 20) links the Armenian female name Ašxēn either to $*\text{axšaina- } ‘\text{brown’ or } *\text{xšai\tilde{\theta}n-}$, taking $*\text{axšēn}$ to be a possible MP outcome (perhaps via dissimilation?). He also cites Oss āxsin ‘Herrin’. Cf. also Périkhanian 1993:16. However, we have no direct attestations of these supposed NW Iranian forms.

In short, the Proto-NWIr reflex of PIr $*\theta n$ was unchanged, but most Middle and New NWIr languages show $\tilde{s}n$ (or a secondary change from it, e.g., Awr $\tilde{z}n$, Bal $n$ $\sim \tilde{s}?$). Some of this may be due to Persian lexical borrowing. But in the case of the Zazaki infinitive suffix (if $< *\theta na-$), borrowing seems unlikely; why would Zazaki borrow the (albeit deverbal) MP abstract noun suffix $-i\tilde{s}n$ and reinterpret it as an infinitive verbal morpheme? Furthermore, Zazaki tends to show the least amount of influence from Persian of all NWIr languages; hence this change is likely to represent an independent parallel development.

\subsection{Other West Iranian phonological patterns}

Reflexes of certain Proto-Iranian sounds show a great deal of variation across West Iranian, to the extent that it is not always clear what the linguistic source of a particular reflex is (if not more than one). Chapter 5 investigates the areal signal displayed by these patterns.

\subsubsection{Reflexes of PIr $*\tilde{s}$, etc.}

Ilya Gershevitch (1962) argued that $*\tilde{s}$, $*\tilde{z} > \text{OP } \theta$, $d$ was not a legitimate sound law and instead represented some sort of register variation.

\textsuperscript{15}Perhaps related to OInd piś- ‘crush, destroy’? This might give the name an original pejorative sense.
His reasoning was this: we only know of one actual word explicitly attributed to the Medians, σπάκα (‘dog’), recorded by Herodotus (1.110.1). Given this item, Gershevitch suggests, we can be reasonably confident that words with sp for PIr *šy are of Median origin (e.g., aspa- (‘horse’ alongside asa-bārā- ‘horseman’)), but we cannot be equally confident that variation of the type *s > s ∼ θ, *ši > šy ∼ θy (on which see §2.4.2.1.6) represents a Median stratum rather than intra-Persian dialectal differentiation.

Part of Gershevitch’s argument (p. 11) concerns reflexes of PIr *gažna- (‘treasure’ (> Pth,Sogd γzn, NP gašn ‘galore’), found indirectly in Achaemenid-era documents as *ganza- and *ganda (e.g., Elamite kan-da-ra, kan-za-ba-ra ‘treasurer’). The latter form is generally ascribed to Old Persian, the former Median. The former shows a distinctive Median metathesis of *zn (cf. Henning 1963a). Gershevitch argues that an OP equivalent *ganda- of Med *ganza- is definitive proof that d was an optional pronunciation of z in OP: the Median loan was borrowed, and then the optional change *z > *d was implemented. However, it may be the case that *ž > d was a subgroup-defining Proto-SWIr change, and that scribes aware of a z : d correspondence (e.g., via OP dāna- vs. Med zāna- ‘race, people’) between Old Persian and Median were simply attempting to “Persianize” the form, and it need not be the case that *ž > d was only an optional, register-specific rule.16

Gershevitch (1962:19–22) takes θy ∼ šy alternation as evidence for variation between θ and s as reflexes of PIr *ž in Old Persian. Others (cf. Hoffmann 1976b:637, fn. 25) disagree; Klingenschmitt (2000:203) ascribes alternation between pre-OP *-ia- and *-ia- to analogical suffix alternation (“Suffixwechsel”), not phonological conditioning. Pre-OP *θi would yield ši, and pre-OP *θi would yield θi. This accounts for some variation within Persian: *tūđiakā- > MP tuhāq, *māđiā-kāhja > MP māhāq, vs. *kāšijāpa-kā- > MP kašavāq (expected ı̊ kašābāq), NP kašav, kašaf, *yašijāh- > MP wyš /weːʃ/ > NP beš ‘more’. This leaves variation in onomastic items like Elamite Ši-ia-ma = OP Šyava.17 versus Elamite Ti-ia-ma = OP *θyāva- ‘black’, if we are willing to identify them with *šiāya- (most scholars seem amenable to this view, though it is impossible to be entirely sure). According to the view given above, word-medially, palatalization of the type *θi > *ši precedes consonant + glide epenthesis. Word-initially, consonant + glide epenthesis appears to pre-date *θi > *ši change, since MP sy’h and NP syah ‘black’ can reflect either OP *θāya- or *šāya-, but not *šāya- (this is my main reason for taking the OP onomastic item *θāya- seriously). For this reason, it looks as though pre-Old Persian contained a doublet *šāya- ~ *šāya-, the disyllabic member of which underwent palatalization (which was blocked in the trisyllabic member). Given that this *-ia- ~ *-ia- allomorphy doesn’t occur in a suffix, we cannot appeal to suffix alternation here, but perhaps are dealing with a phenomenon similar to Sievers’ or Lindeman’s Law.

The fate of the PIr clusters *ši and *θi across West Iranian remains unclear, blurred by large-scale contact. With some exceptions, we can see that certain reflexes of these clusters have taken hold as lexical “prototypes” across West Iranian. For instance, most West

---

16Cf. other examples of loanword adaptation, e.g., English hammer → Tongan hamala → Samoan sāmala, where Samoan speakers were aware of an s : h correspondence between Samoan and Tongan, and took this into account in the loanword adaptation process (Geraghty 1983:102).

17Elamite š also = OP s; however, see Asatrian 2012:106–7 for a variety of Persian forms with š-, e.g., šāh-tūt ‘mulberry’ (the first member is originally ‘black’, not ‘king’).
Iranian languages show \( h < \ast \theta \) in their word for ‘fish’, except for Parthian \( m^\prime sy^\prime g /ma\hat{\imath}\text{-}sja\hat{\imath}/ \), Zazaki \( m\hat{\imath}\text{-}se \) and Kurdish \( m\hat{\imath}\text{-}\text{as} \). PIr \( \ast\text{-}\text{i}\text{-}\text{i} \) shows \( \text{siy-} \), e.g., MP syh \( > \) NP syâh ‘black’, Zaz siyâ, Qohrûdi sigâh, with the characteristic loanword phonology seen in other Persian borrowings (e.g., NP piyâz \( \rightarrow \) Qoh piqâz ‘onion’). Furthermore, PIr \( \ast\text{-}\text{ka}\text{-}\text{s} \) shows a plain \( s \) in more than one language, e.g., Bandarî kâsâpošt ‘turtle’, 18 Balochi kâsâp \( \sim \) kâsib ‘turtle, tortoise’. (Korn 2005:284 notes that a “genuine” Balochi word should contain \( š \). However, long \( \check{i} \) reflects disyllabic \( \ast\text{i}\hat{\imath}\); cf. Korn 2005:105. It could then be the case that kâsâp reflects \( \ast\text{k}\hat{\imath}\text{s}\hat{\imath}\text{i}\text{-}\text{i} \) \( \text{a-} \) gen.sg.).

19 These patterns of historical phonological variation across West Iranian are likely an interaction of regular sound change, analogy (i.e., the continuation of different case forms), and contact between closely related languages.

We see some additional idiosyncrasies concerning the cluster \( \ast\text{-}\text{su} \). The “proper” Southwest Iranian reflex of this form is \( s \), but Persian shows \( sp \) as well at each chronological level as well as \( sf \) (restricted to New Persian). Given \( \sigma\tau\text{-}\text{oxa} \), Gershevitch was content to take forms with \( sp \) as Median loans in Persian. I believe that Henning (1963b:71, fn. 13) suggests that incidences of NP \( sp \sim sf \) for PIr \( \ast\text{su} \) could be due to (possibly Persian-internal?) dialectal differentiation rather than admixture with Median or another NW Iranian language when he says,

> Why then should we assume that Ir. \( zu \) (and \( su \)) should necessarily become either \( zb \) (\( sp \)) or \( z \) Khot. \( \check{z} \) (\( s \) Khot \( \check{s} \) Wkh \( \check{s} \)) and exclude the possibility of other developments?...The modern Persian \( sf = sp \) seems to be confined to words with original \( su \)...(barring a few arabicized forms) and could be attributed to the influence of a dialect in which \( su \) resulted directly in \( sf \).”

New Persian shows at least three relatively consistent outcomes (i.e., limited to particular etymological reflexes) for PIr \( \ast\text{-}\text{su} \):

1. \( \ast\text{su} > s \); e.g., \( \ast\text{a}\text{-\text{s}u}-\text{b}\text{a}\text{r} \) > OP asabara ‘horseman’ > Phl \( \text{sw}^\prime r /\text{asw}^\prime r/ \) > NP suvâr ‘rider’; \( \ast\text{-s}\text{u} \text{-ka} \) > NP sag ‘dog’

2. \( \ast\text{su} > sp \sim \check{sp} \); e.g., \( \ast\text{-a}\text{s}\text{u} \) > NP asb [\( \check{a}sp \)] ‘horse’, \( \ast\text{-s}y\text{-i} \) > NP sepeš \( \sim \) šepeš ‘louse’

3. \( \ast\text{su} > sf \sim sp \); e.g., \( \ast\text{-}\text{s}\text{u} \text{-\text{a}nt} \) > NP esfand \( \sim \) espand ‘wild rue’; \( \ast\text{-s}\text{u} \text{-\text{ait}a} \) > NP sefid \( \sim \) sepid ‘white’

(1) gives the expected Southwest Iranian outcome. NP asb ‘horse’ has a probable antecedent in OP aspa-, and could represent the phonological treatment of the old Median stratum; while sepeš \( \sim \) šepeš ‘louse’ (< Phl spys) does not have an attested antecedent in Old Persian,

---

18It is possible that this form reflects an anticipatory dissimilation \( s...\check{s} > s...\check{s} \), rather than \( s < \ast\check{s} \). However, assimilation of the type \( [s \ldots \check{f}] > [\check{f} \ldots \check{f}] \) is well documented, given the nature of motor planning for complex segments (Garrett and Johnson 2013); while a dissimilatory change to the opposite effect is in theory possible, it would be articulatorily suboptimal.

19In a sense, pace Korn, it may be the case that \( sy \) and \( \check{sy} \) are not, respectively, Persian and Balochi outcomes of these clusters. Multiple languages show these outcomes depending on whether \( \ast\check{s} \) is in contact with a glide or an epenthetic vowel. It is not clear if this is a pan-West Iranian rule or not.
it may represent the same stratum, as would Phl \textit{wsp}, MMP \textit{wysp} < OP \textit{vispa-}. But forms in which \textit{sp} and \textit{sf} vary may come from another source. Martin Schwartz (p.c.) remarks that it is unlikely that \textit{sf} came about via a secondary change from \textit{sp}, given that instances of \textit{sp} from etymological sources other than PIr \textit{*šu} do not undergo such a development. Schwartz (2006:223) says that for \textit{esfand} \textasciitilde \textit{esfand} ‘wild rue’, the variant with \textit{sf} “cannot be due to the mediation of Arabic, which has only \textit{harmal}.”

What dialect could have given rise to forms with this sort of variation? Some clues may be apparent in the Yazdi language. Yazdi is one of the Central Dialects (i.e., Northwest Iranian) spoken in the city of Yazd, a long-standing center of Zoroastrian worship in Islamic Iran which also boasts the oldest Jewish community in the country; the Zoroastrian lect, Gabri (a derogatory term; NP \textit{gabr} ‘non-Muslim’ = Turkish \textit{gavur} = Arabic \textit{k\textacutes}ir ‘infidel’), stands alongside a Jewish lect. Some Zoroastrian Yazdi forms seem to have \textit{sv} for PIr \textit{*šu}, e.g., \textit{s(e)v\textbar}ıd ‘white’ (Bailey 1936:349), \textit{svaka} ‘dog’ (Krahnke 1976:231, citing the field notes of Michael M.J. Fischer; Bailey gives \textit{sag}). From the terms given by Bailey, it is hard to tell whether \textit{sv} continues a preform \textit{*šu} or \textit{*sf}—his Yazdi’s \textit{v} reflects (along with \textit{*m} and \textit{*u}) \textit{f}, e.g., \textit{*draf\textbar{s}a-} \textasciitilde \textit{dru\textbar{š}} ‘banner’—or if it could possibly come from “Median” \textit{sp}. According to Vahman and Asatrian (2002:20), PIr \textit{*šu} “remains unchanged in a very important lexeme: \textit{sva}, \textit{seba}, \textit{sava}, \textit{sewa} ‘dog’.” However, the presence of the word \textit{svarz} ‘spleen’ < PIr \textit{*spr\textbar{̆}źa(n)} seems to indicate that PIr \textit{*sp} also became \textit{sv} in Yazdi. So while it is likely that NP forms with \textit{sf} are loanwords, Yazdi is unable to explain their distribution (i.e., why \textit{sf} for \textit{*šu}, but not \textit{*sp}?) — the issue raised above by Henning.

2.6.3.2 PIr \textit{*r} + coronal clusters

Middle and Modern Persian show \textit{l} for a number of \textit{*r} + coronal clusters:

- PIr \textit{*rd\textbar{e}}- > MMP \textit{dyl} /\textit{dil}/ \textasciitilde NP \textit{dil} ‘heart’
- PIr \textit{*yarda-} \textasciitilde NP \textit{gul} ‘flower’
- PIr \textit{*sarda-} > Phl \textit{ŠNT} (Aram.) /\textit{sa\textbar{l}/} (MMP \textit{s}’r) > NP \textit{s\textbar{l}
- PIr \textit{*p(a)rdanku-} > NP \textit{palang} ‘panther’ (cf. Vedic \textit{p\textbar{̆}d\textbar{k}u-} ‘snake’, meaning ‘tiger, panther’ in lexicata, SSog \textit{pwr\textbar{δ}nk} ‘panther, leopard’)
- PIr \textit{*br\textbar{z}ant-} > MMP \textit{bw\textbar{n}d} /\textit{buland}/ \textasciitilde NP \textit{buland} ‘high’
- PIr \textit{bar\textbar{z}ā\textbar{d}(a)}. \textsuperscript{21} > MMP \textit{b\textbar{l}y} /\textit{ba\textbar{laj}/} \textasciitilde NP \textit{b\textbar{l\textbar{a}}} ‘height’
- PIr \textit{mar\textbar{z}-} > Phl \textit{m\textbar{l}-} /\textit{ma\textbar{l}/-} ‘rub, sweep’ \textasciitilde \textit{m\textbar{l\textbar{d}an} ‘rub, polish’
- PIr \textit{r\textbar{z\textbar{i}f\textbar{a}-} > MMP \textit{lwp}/f /\textit{a\textbar{luf}/} > NP \textit{\textbar{a\textbar{l\textbar{u}}h} ‘eagle’

\textsuperscript{20}This form is likely a Wanderword; see Lubotsky 2001.
\textsuperscript{21}Perhaps a de-instrumental \textit{d}-stem built to PIE \textit{*wh\textbar{erg\textbar{h}}-ch\textsubscript{1}-}, cf. Latin \textit{merc\textbar{ē}s}, \textit{merc\textbar{ē}dis} ‘wages’, \textit{her\textbar{ē}s}, \textit{her\textbar{ē}dis} ‘heir’ (Weiss 2009:304-5).
• PIr *źarnu-man-ı ‘gold neck’ > NP dâl-man ‘black eagle’ (Schwartz 1971:292, fn. 14)\(^{22}\)

• PIr g(a)rama-ka- > NP gala ‘flock’ (Schwartz 1971:292, fn. 14)

• PIr *pr̥tw- → *pr̥θw- > MMP pwhl /puhl/ > NP pûl ‘bridge’

This change, from what we can see, post-dates Old Persian.\(^{23}\) However, there are a large number of exceptions to this rule; for example, NP buland is in a doublet with burz, thought to represent a Northwest Iranian form (Beekes 1997:3). For some etyma, Persian lacks l, while a non-Persian reflex displays it, e.g., NP supurz ‘spleen’ versus Kd sipîl ‘id.’ < PIr *spr̥zān. On the outcome of PIr *rd/rź in Kurdish, MacKenzie (1961:78) says the following: “I do not think it is possible to be certain which is the true Kurdish development, but whether we consider the many words with l/l as native or loan-words their preponderance is significant.”

We encounter other vexing aspects of the distribution of l. For instance, Gorani zîl ‘heart’ shows a decidedly non-SW Iranian outcome of PIr *ź-, showing that not all forms with l can be accounted for by considering them to be Persian loans. Was *rd > l then a lexically diffused sound change, or did Gorani somehow partially “undo” the Southwest Iranian historical phonology of that word based on knowledge of the correspondence NW z : SW ð? On a case-to-case basis, it can be difficult to securely establish a dialectal or areal source for some of these forms, but it may be possible to quantitatively measure the geolinguistic signal displayed by these varying outcomes.

2.6.3.3 PIr *u-

The fate of initial *u- is highly irregular across West Iranian (see Schwartz 1982 for some possible conditioning environments).

• PIr *u-i-nāśa- > MP wināh > NP gunāh ‘sin’

• PIr *u-i-čāra- > Phl we’l-, MMP wyc’r- /wiza:r- / > NP guzār(dan)

• PIr *urtka- > MP gurdag > NP gurda ‘kidney’

• PIr *urka- > MP gurg > NP gurg ‘wolf’

• PIr *urpa-ka- > Phl gwîlkb’ /gurbag/ > NP gurba ‘cat’ (cf. YAv urupi- ‘dog, fox (?)’ < *urpi-)

• PIr *u(i)jāna- > MP gyān > NP jān ‘life, soul’

• PIr *yahāna-ka- > MP wihân(aq) > NP bahâna ‘reason, pretext’ (cf. Dîgî Ossetic ræuonæ < PIr *fra-yahâka-, from *yah- ‘dress’; see Gershevitch 1952:483-4)

---

\(^{22}\) Assuming that this form is a word equation with YAv zarnunmainiš; dâl on its own could also be connected with PIr *sarta- (assuming that YAv zairita- ‘golden’ ← *zarta- due to the influence of zairi- ‘gold’).

\(^{23}\) There are no good direct precursors of the forms listed above, but we see forms like OP ardata- ‘silver’ < PIr *(a)rźata-, cf. Yazdî älî (Kent 1951:171).
• PIr *u-di-dana- (?) > Phl wyð’n, MMP wy’n /wijan/ > NP giyân ‘tent’ (cf. OInd vi-dhā-‘furnish, spread, diffuse’?)
• PIr *uata- > Phl wt, MMP wd /wad/ > NP bad ‘bad’
• PIr *uāta- > Phl w’t, MMP w’d > NP bād ‘wind’
• PIr *u(n)šati- > Phl/MMP wəst/wi:st/ > NP biwarnings‘tent’ (cf. OInd vi-dhā-‘furnish, spread, diffuse’?)
• PIr *uata- > Phl wt, MMP wd /wad/ > NP bad ‘bad’
• PIr *uata- > Phl w’t, MMP w’d > NP bād ‘wind’
• PIr *anśati- > Phl/MMP wəst/wi:st/ > NP biwarnings‘tent’ (cf. OInd vi-dhā-‘furnish, spread, diffuse’?)
• PIr *aṃda- > Phl ghl /gul/ > NP gul ‘flower, rose’
• PIr *aṃda- > Phl ghl /gul/ > NP gul ‘flower, rose’
• PIr *aṃda- > Phl ghl /gul/ > NP gul ‘flower, rose’
• PIr *aṃda- > Phl ghl /gul/ > NP gul ‘flower, rose’
• PIr *aṃda- > Phl ghl /gul/ > NP gul ‘flower, rose’
• PIr *uata- > Phl wt, MMP wd /wad/ > NP bad ‘bad’
• PIr *uata- > Phl wt, MMP wd /wad/ > NP bad ‘bad’
• PIr *uata- > Phl wt, MMP wd /wad/ > NP bad ‘bad’
• PIr *uata- > Phl wt, MMP wd /wad/ > NP bad ‘bad’

The historical phonology of the forms given above shows some dominant trends: we see that PIr *ur- generally > MP gu-, PIr *u- > MP w- > Early New (Judeo-)Persian [β]- > NP b-, but there are some environments where it is difficult to make a generalization, particularly the fate of *ur-, which shows change to both g- and b- in conditioning environments that are not particularly clear, as well

24 YAv varoṣa-‘rose’ points to (and Senn val(a) ‘flower’ seems to point to — perhaps also Pth w’r wari-/)

25 Schmitt (1989a:69) and others give this reading — departing from vazrakā- (found in Kent 1951) — on the basis of the NP form, but the morphology (-ka- suffixed to an athematic stem in -r-) seems unusual for Old Persian.

26 See MacKenzie 2003 for this interpretation based on variable spellings found in a manuscript of a commentary on the Book of Ezekiel.
as variation in the words for ‘sparrow’ and ‘rice’. Additionally, a handful of New Persian words show \(v\)- for PIr \(*u\)-, e.g., vazay ‘frog’, vazidan ‘blow’, varzidan ‘work’, vîr ‘memory’ (MacKenzie 2003:110); the verbs are marked as loans by Cheung (2007:426, 431).

The forms descended from \(*uazr(a)-(ka-)\) were likely NWIr loans into Old Persian since they show \(zr\) as opposed to expected \(dr\) (though our best evidence of the change \(*źr > OP dr\) is word-initial, e.g., drayah- ‘sea’). In any case, the relevant historical phonology (i.e., changes to \(*u-\)) postdates the loan, and the fact that the word was borrowed at an early date does not tell us anything about the provenance of the changes that took place at a later date.

Lentz (1926:280–1) seems to consider \(*u- > b-\) the regular SWIr outcome. MacKenzie (1971a:76) takes the change \(*u- > b-\) as a feature shared by Persian and Northern and Central Kurdish dialects, whereas “[i]n most other W.Ir dialects \(w-\) is little modified in this position, while in Bal. it has developed into \(g(w)\)-.”

But whatever the provenance of these changes, a cursory glance at the West Iranian data shows that the picture is far from clear. While \(*u-\) is often modified, forms with \(g-\) and \(b-\) preponderate, likely due to contact. For instance, Kumzârî ward ‘flower’ sits alongside bistâ ‘twenty’ (Thomas 1930:814, 825).\(^{27}\) Additionally, while Zazaki usually shows \(v-\) (e.g., vâ ‘wind’), the word for ‘blood’ is guâni < \(*yahuni\)\(^{28}\) (Paul 1998b).

This variation is no doubt due to contact, but given the shallow chronology of changes of this sort, it can be difficult to establish a change as Persian as opposed to non-Persian.

### 2.6.3.4 West Iranian functional items

West Iranian languages share a large number of similar-looking or identical nominal and verbal morphemes that cannot be genetically shared innovations, as they postdate the Old Iranian period. These are discussed at length in Windfuhr 2009 and elsewhere. We can be certain that these morphemes post-date any possible period of West Iranian unity, but it is not clear if they are seen across West Iranian languages due to borrowing, or have been grammaticalized (or otherwise developed) in parallel. Chapter 4 presents an attempt to tease apart areality and parallelism in this domain of shared West Iranian features.

### 2.7 Conclusion

In this chapter, I have shown that there is not much support for East and West Iranian as genetic groups. I have demonstrated that the lenition of PIr \(*b(r)-\), \(*d-\), \(*g-\) seen across East Iranian is also seen some West Iranian languages, and that there are exceptions to the East Iranian voicing of \(*-ft-\), \(*-xt-\) as well as the West Iranian devoicing of \(*bd\), \(*gd\). I have shown that this genetic division rests on a scant number of lexical isoglosses, making it quite weak.

\(^{27}\)Though Martin Schwartz (p.c.) notes that this form, rather than being tantalizingly conservative, is likely a loan from Arabic.

\(^{28}\)The word for blood shows irregular historical phonology across west Iranian. NP xûn (MP xôn) has either undergone a metathesis between \(*u\) and \(*h\), or was subject to the same irregular \(x\)-prothesis as MP xâyag, NP xaya ‘egg’. Parthian has guûn, with unexpected \(y\)-. Sivandi has fîn.
At the same time, I have made note of some uncanny trends and similarities across West Iranian languages that are non-genetic, but that, were it not for evidence from the oldest West Iranian specimens, we might reconstruct as shared innovations. These features are investigated in detail in the following chapters.
Chapter 3

Computational classification of the Iranian Languages

In this chapter, I make use of methods imported from computational phylogenetics in order to investigate the genetic subgrouping and other internal dialectal affinities of the Iranian languages.

In historical linguistics, computational phylogenetics can be used to test the validity of proposed higher-order subgroups for which there is not good evidence in the form of one or more shared phonological or morphological innovations. Generally, efforts of this sort use lexical character data in order to see if such subgroups are compatible with patterns of lexical replacement, if not traditional innovations used to define subgroups in traditional comparative-historical linguistics. Additionally, there is a sizable literature which attempts to replicate uncontroversial phylogenies using typological character data (Dunn et al. 2008 et seq.). In particular, I seek to replicate uncontroversial subgroups within Iranian, such as Southwest Iranian, and to see whether patterns of lexical replacement support the idea of a West Iranian subgroup. In addition to lexical character data, I use characters based on typological and recurrent features in order to gauge the extent to which East and West Iranian serve as good areal groups.

I find that typological and recurrent characters produce a division between East and West Iranian languages, though there is not a particularly strong split between these groups. I find that lexical character data consistently produces a West Iranian clade, but no East Iranian clade, a result found by other scholars. The lexicostatistical result reached for West Iranian is in theory compatible with a view in which West Iranian is a genetic subgroup; however, in the absence of other strong subgroup-defining innovations for West Iranian, this result may also indicate that West Iranian is an areal group. Lexical borrowing between West Iranian languages has clearly taken place at the time of their earliest attestation, and it may be the case that patterns of lexical replacement in West Iranian owe to prolonged (or punctuated) contact dating back millennia.
3.1 Computational phylogenetic linguistics

Evolutionary biology makes use of a number of computational methods for inferring phylogenies of different taxa. Different software programs have the capability to process NEXUS files (Maddison et al. 1997), which contain matrices of taxa and the biological characters that they display. These programs use different algorithms in order to infer phylogenies. These methods have been co-opted for use in various subfields of linguistics; their use is comprehensively described in McMahon and McMahon 2005; Nichols and Warnow 2008; Dunn 2015.

Some methodologies are distance based. The traditional algorithm used in lexicostatistics is the unweighted pair group method with arithmetic mean (UPGMA). It observes a pairwise distance matrix to construct a distance-based dendrogram. An assumption of this method, often viewed as a disadvantage in historical linguistics, is its assumption of a constant rate of evolution.

Another distance based method is Neighbor Joining (NJ, Saitou N. and Nei M. 1987). This algorithm works according to a distance matrix of all taxa involved. It starts from a star network (in which all taxa are represented as points branching out from a central node) and proceeds to join together taxon pairs that are close to each other in two-dimensional space via new nodes, and then joining the newly-formed nodes together with additional nodes. It can be implemented in little time. However, it can be unreliable when dealing with large amounts of data and large-diameter trees.

The NeighborNet algorithm is based on NJ. However, while NJ creates a new node for every pair of proximate nodes, NeighborNet waits until a node is paired with two proximate nodes, and then creates two new nodes. When there is uncertainty or no clear optimal or hierarchical branching structure, NeighborNet leaves this unresolved as webbing between branches. The webbing is often quite informative, though not necessarily with respect to strict phylogeny.

Other methodologies are character based. Many of these involve Bayesian algorithms which start with an initial model tree and follow a random walk through model tree space, generating a number of hot chains alongside a tree composed of immobile cold chains. These different tree probabilities form a consensus tree. These methods are generally NP-hard (non-deterministic polynomial-time hard), meaning that they cannot be solved efficiently in polynomial time, and require heuristics (such as the random walk) to solve them, which can be time-intensive.

Several programs, such as Bayesian Evolutionary Analysis Sampling Trees (BEAST, Lemey et al. 2009), use Bayesian Monte Carlo Markov Chains to infer optimal trees. These programs are designed to investigate the origin of virus outbreaks from molecular sequence data, but can process different data types as well, such as simple binary data.

Prior to introducing different character types, it is helpful to mention the term homoplasy, which refers to recurrence among characters. A good biological example of a homoplasy is the independent development of wings among birds, bats and flying insects; this trait (wings) has recurred multiple times, an example of convergent evolutionary behavior. (There is an infelicitous lack of isomorphy between biological and linguistic notions of “convergence”;
linguistic convergence presupposes some sort of areal unity among participants, whereas parallelisms can occur between areally discontiguous languages.)

3.1.1 Lexical Characters

The use of lexical characters to infer subgrouping is the focus of lexicostatistics, a much-criticized methodology pioneered by Morris Swadesh in the 1950s. An appraisal of lexicostatistics as a whole (and not simply as it relates to the data analyzed here) is outside of the scope of this paper. A danger of lexicostatistics is the possibility of inferring subgrouping according to shared patterns in multiple languages’ vocabulary replacement, when in reality there has simply been drift on a large scale; such patterns can also arise due to contact rather than subgrouping. Lexical characters can additionally exhibit homoplasy; i.e., over time, a lexical item could potentially fill a semantic role, be replaced, and later return to the same semantic role. Part of lexicostatistics’ bad reputation stems from its association with glottochronology, a methodology that infers absolute chronologies from lexicostatistics, with the underlying assumption of a constant rate of vocabulary replacement. For heavy criticism of this methodology, see *inter alia* Matisoff 2000. However, historical linguists who are not lexicostatisticians are often willing to define a subgroup according to a shared vocabulary, or in an extreme case (Parpola 2012), on the basis of a single word (*ṣaṇḍaṇa*—‘god’ in Avestan, Scythian, Saka, and Ossetic). Recall also that some stable East-West lexical isoglosses were established in the previous chapter, but that I do not consider these strong diagnostics of subgrouping on their own.

Lexical characters are generally assigned according to a Swadesh-200 word (really meaning) list, or an areally or culturally appropriate meaning list such as the CALMSEA list (proposed in Matisoff 2000 for languages of Southeast Asia). The objective is to fill each meaning slot with the basic or default vocabulary item in a given language that corresponds to a particular meaning. At times, there may be more than one vocabulary item in a given slot. This is referred to as a lexical POLYMORPHISM and can be either embraced, or suppressed by eliminating the item judged to be less basic. When Swadesh lists for a group of related languages are compared and cognate forms in different languages are matched per meaning slot, a matrix is produced that can be processed by computational phylogenetic software in order to generate optimal phylogenetic networks according to apparent shared rates of vocabulary replacement between languages.

3.1.2 Phyletic Characters

The biological term PHYLETIC refers to a character with the capacity to define a phylum, and importantly, unlikely to recur, or exhibit homoplasy. In linguistics, phyletic characters are ones that can potentially define linguistic subgroups. These include phonological and morphological innovations that are unlikely to have arisen via parallel innovation. Lexical characters, when taken as a whole (i.e., 200 or so per language) provide information from which genealogies can be inferred, but a single phonological or morphological change that is unusual or unlikely to recur multiple times is taken as an even stronger diagnostic of linguistic
subgrouping. Characters of this sort should be weighted over characters that are likely to recur.

### 3.1.3 Phenetic Characters

The biological term *phenetic* refers to traits related to phenotypes rather than genetic traits. In historical linguistics, phenetic characters are ones that are recurrent or homoplastic, meaning that they are likely to have come about independent of any shared genetic development. They can include common, phonetically-natural sound changes (e.g., $s > h$) or morphsyntactic changes that represent common diachronic tendencies (e.g., the parallel development of ergativity in Indic and Iranian languages).

Other phenetic characters include typological and structural characters, such as the presence of various phonemes in a given language’s inventory, the presence of vowel harmony, locative infinitives, etc. Characters such as these which pertain to a language’s “appearance” can be used to measure typological distance. Distance-based algorithms are generally used with such data.

### 3.1.4 Previous Literature

Tischler 1973 is an early lexicostatistic and glottochronological study of Indo-European. It confirms the Indo-Hittite hypothesis, and reports glottochronological dates of 3500 BCE (with a 200-word Swadesh list) and 3800 BCE (with a 100-word Swadesh list) for PIE.

Lexicostatistics has been somewhat successful in subgrouping Austronesian, though results have been brought into question. Thurston (1994:575) says that “lexicostatistical methods in the Pacific have been important in providing an initial organisation of the languages into working categories, especially at the lower taxonomic levels of Austronesian, but they have probably been less satisfying at higher levels of the taxonomy than traditional comparative-historical methods.”

Dyen et al.’s (1992) seminal study is a purely lexicostatistic analysis of modern Indo-European languages. The authors use UPGMA to plot languages in multidimensional space, and successfully classify languages according to subgroup. The dataset has been reused in later studies, but it is highly flawed, as many Indo-Europeanists are well aware. The character states are often incorrect, and the lexical data itself is often of poor quality, devoid of diacritics and with little metadata apart from the sources used.

Additional studies have made use of computational methodologies for the purposes of historical linguistic subgrouping and classification. Warnow et al. 1995, Ringe et al. 2002, and Nakhleh et al. 2005 make up a series of studies presenting ongoing research on the computational classification of Indo-European languages, with a focus on the oldest attested languages of each branch. The authors rely on lexicostatistic data, phonological innovations, and morphological innovations.

In another character, an Indo-European language will be coded 1 or 2 depending on whether it underwent (respectively) full centum or satem treatment of the PIE palatovelars (this view precludes alternatives to the idea that satem behavior is a phyletic, subgroup-defining
innovation; for discussion, see Clackson 2007:51). However, there are cases in which satem treatment is not complete (acknowledged by the authors), such as the famous Lithuanian doublet *akmuõ* ~ *ašmuõ* ‘stone’. All the same, Lithuanian is coded as 2. There is no way for a language to be both centum and satem. In theory, this is a good way to avoid coding homoplastic or recurrent characters, but it completely ignores the role of language contact in the role of the development of Balto-Slavic.

Dunn et al. 2008 *et seq* explore the possibility of inferring phylogenetic relationships based on structural traits. Finding agreement between a structural phylogenetic analysis of Oceanic languages and an older tree constructed using the comparative method, they conclude that such a methodology is “a valid way of extracting linguistic prehistory.” Donohue et al. 2008 *et seq* have argued that the signal picked up by the use of structural features in linguistic comparison as performed by Dunn et al. is areal rather than phylogenetic.

Gray and Atkinson 2003 and Bouckaert et al. 2012 are purely lexicostatistic studies of Indo-European. Unlike Ringe et al., they use modern Indo-European data (taken from Dyen et al.’s (1992) lexicostatistic experiment) in addition to old Indo-European data. They claim that their results confirm the “First Agriculturalists” hypothesis of Renfrew 1997, which claims that Proto-Indo-European originated in Anatolia ca. 7500 BCE and spread to Europe via farming dispersals. Chang et al. 2015 makes use of virtually the same data, but employs ancestry constraints between ancestral and descendant languages (e.g., Old and Modern Irish, Ancient and Modern Greek, etc.) and restriction site characters; the results of this work support the Steppe Hypothesis, i.e., that PIE was originated in the Pontic-Caspian Steppe ca. 4500 BCE.

Blažek 2013 employs Sergei Starostin’s “recalibrated glottochronology” to investigate the subgrouping of Middle Iranian languages, building off of previous results by Sergei Starostin’s team carried out in 2004 (*non vidi*). Both experiments produce a West Iranian subgroup, but not an East Iranian one. This work uses a 100-word Swadesh list, and includes polymorphisms. Branching structure is based on ratios of shared vocabulary between sister languages.

A number of studies have used NeighborNet in order to cluster languages according to their typological properties. Bakker et al. (2011) demonstrate via NeighborNet that creoles form a natural class distinct from non-creoles according to typological characters. Szmrecsanyi (2011) makes use of NeighborNet, linear regression, and other models to compute the linguistic distance between regional varieties of English.

### 3.2 Character Coding

This section includes a discussion of the lexical and typological characters used in this chapter’s experiments.

#### 3.2.1 Phenetic Characters

This section presents a number of parallelisms (i.e., “homoplastic characters”) found in the Iranian languages. In a computationally tractable dataset of Iranian languages, parallel/recurrent
phonological and morphological characters fall under the umbrella of homoplasy. They do not have much value as phyletic characters, which are better diagnostics for subgrouping. For the typological dataset, I omitted Wāzirī; I added features from Mazandaranī from Lecoq 1989 and was able to gather additional information regarding vocabulary from a native speaker.

3.2.1.1 Phonological Characters

3.2.1.1.1 Rhotacism of dentals A change from dentals to *r recurs within Iranian among geographically separate languages; it is seen in the “Tatic” development (Schwartz 2012) of Judeo-Yazdi šer- ‘go’ (< *čjuta-), Judeo-Isfahani čer- ‘know’ (< *čait-), Kumzārī spīr, N Baškārdī espīr ‘white’ (< *suaita-). Unfortunately, due to the availability of information, only the latter two languages were included (this N Baškārdī trait, unseen in S Baškārdī, was coded as general Baškārdī).

3.2.1.1.2 Loss of PIr *T Old Iranian *T has been lost over time. In the Middle Iranian period, it is already absent from Middle Persian and Parthian, though it is generally still appears in East Iranian. East Iranian languages lose traces of *θ in various positions. For instance, PIr *θ > h. Bactrian preserves intervocalic *θ, but undergoes the change *θr > hr in parallel with Parthian, e.g., *puθra- > πουρο ‘son’. A second parallelism with Parthian vis-à-vis the treatment of *T is seen in Khwarezmian: PIr *caθyāra- > Pth, Khw cf’r /cafa:r/ ‘four’. No New East Iranian language retains θ as such, except Šuynī (Édel’man 1980:298). However, dialects of Yaghnobi, which today shows the dialectal variation t ∼ s, probably preserved *θ into the 19th century, if the transcription of de Ujfalvy (1896:6), who gives a form therāi alongside sēriāi for ‘3’ (< PIr *θraja-). Some, like Ossetic, have re-hardened *θ to t.

Henning (1958:108) pointed to phonological parallels between Old Persian and Sogdian, namely the change of PIr *θr > Sogd š, OP ç (thought phonologically to be /s/; for an alternative view, see Kümmel 2007). This phonological behavior is characteristic also of Khwarezmian, and appears in some new NW Iranian languages (e.g., Semn še ‘3’).

It is necessary to note that Sogdian and Khwarezmian do not always show this behavior; it is dependent on whether θ and r are adjacent, perhaps due to prosodic factors. Hence, there are doublets, e.g., Buddhist Sog ḍry, Manichean Sog šy ‘three’.

3.2.1.1.3 PIr *śr > sr, ç/s, š, rs This character is somewhat of a gray area. Some languages, such as Old Persian, reflect solely ç (the number of OP reflexes is small in any event). In some cases, Middle Persian appears to continue this behavior, e.g., *ni-šrai- > OP niyaçaɾāray- ‘lean’ caus. (× dāraya- ‘hold’ caus., Kent 1951:188), MP ns’y /nisaj/ ‘conveying, dispatch’. Elsewhere, Middle Persian and New Persian preserve the cluster *śr, e.g., *śrāyaʃa- ‘hear’ (caus.) > MP sr’y /saraj/ ‘sing’, NP sarąy.

New Persian shows a number of doublets:

- PIr *hyaʃru- > NP xuṣrū, xuʃu ‘mother-in-law’
• PIr *hūašura- > NP xasar, xusū (perhaps with elision of final r rather than a continuation of OP *ç) ‘father-in-law’

• PIr *ašra(-ka)- > NP ašk, ars ‘tear’ (cf. Khw šwkHzū)

It is possible to view xusū and ars as “authentic” NP forms (ars in particular, given that it undergoes metathesis typical of NP, e.g., *čaxra- > čarx ‘wheel’, *namra- > narm ‘soft’), whereas xušū and ašk could be loans from some East Iranian language given their phonological behavior (cf. Balochi šarr < *śrīra-, called an “Eastern substrate word” in Elfenbein 1989:398), but this line of reasoning relies on somewhat unclear evidence—that is, it seems less straightforward to me than identifying Šuṅni dil as a Persian loan.

Similarly, Balochi has a triple reflex for this cluster: the aforementioned šarr; *aśra- > Balochi ars (Raxšānī, Barker 1969, s.v.), ārs (Collett 1909, s.v.) ‘tear’; *hūašru- > Balochi vassū ‘mother-in-law’.

In short, the outcome of PIr *šr is difficult to encode in a single, simple phonological character. I do not include it as a phenetic character.

3.2.1.1.4 PIr *šu > sp The development of the Proto-Iranian cluster *šu is a strong diagnostic of subgrouping. The change PIr *šu > s defines SW Iranian, and the change PIr *šu > š/sš unites Khotanese, Tumshuqese, and Wakhi. However, the rest of Iranian has sp or something incredibly similar. The change *šu > sp cannot be a subgroup-defining change given its distribution. It affects all of Iranian (except for SW Iranian, Kurdish, Balochi, and Saka). It is quite old, attested in the Median word σπάκα ‘dog’ (< *šu aka-) given by Herodotus (1.110), as well as the Scythian proper name Βαιορασπος (< *baiu ar-aśu a- ‘myriad horses’), found at Tanaïs (6th cent. BCE). I believe that diffusion is the only possible explanation for this feature’s distribution in Iranian. If we believe it to be a phyletic character, it forces us to envision a clade which excludes SW Iranian, Saka, and also Kurdish and Balochi.1

3.2.1.1.5 Lateralization of dentals A change of dentals to l (unconditioned by a preceding *r, as opposed to Middle/New Persian *rd > l) is seen in Bactrian, Pashto, and Yidgha. It appears also in the Nuristani languages (e.g., Prasun lūšt ‘daughter’ < *dužit-, Mayrhofer 1984:384) and has affected Romani as well. It may be an areal feature.

3.2.1.1.6 PIr *-ft-, *-xt- > -vd-, -yd- Voicing of medial fricative + plosive clusters is common across east Iranian.

1Interestingly, a commonly noted feature of diffused sound changes is incompleteness of operation, but the change to sp has been quite thorough. The exceptions are Old, Middle and New Persian *rd > l) is seen in Bactrian, Pashto, and Yidgha. It appears also in the Nuristani languages (e.g., Prasun lūšt ‘daughter’ < *dužit-, Mayrhofer 1984:384) and has affected Romani as well. It may be an areal feature.
3.2.1.1.7 Presence of retroflexion A number of East Iranian languages have retroflexion, but under different circumstances. Balochi has retroflexion only in loans. Khotanese Saka, Pashto, Or Muhammad, Parāce, Iskašmi and Wakhi have retroflexion in inherited Iranian words.

The development of retroflexion in the aforementioned languages is not uniform between them.

Khotanese Saka has 

\[ *rt > d; *rt > Vd; *xš > kš; š before *ā; *fš > ks; *šr > š; *š > š; *θi > ťh; *-r...n- > -r...ŋ-, but *rn > rr. \]


It would conceivably be possible to code each individual sound change according to its operation in a given language, but this creates an exceptionally large number of characters for a phonological phenomenon that affects only a subset of Iranian. I chose instead to distinguish retroflexion in loanwords from inherited retroflexion.

3.2.1.1.8 Fortition of PIr fricatives In Khotanese Saka, Balochi, and Parâce, the Proto-Iranian fricatives have generally been hardened to stops, in the case of Balochi (in some dialects of which stops are spirantized post-vocally), or aspirated stops (in Khotanese Saka and Parâce). Avestan curiously shows pt for PIr *ft.

3.2.1.2 Morphosyntactic Characters

3.2.1.2.1 Collective *-tā A shared collective marker is seen in Sogdian -t(‘) and Ossetic -tw, and also appears to be present in some Sarmatian items, e.g., the tribal name Σαρμαται ‘Sarmatians’.

3.2.1.2.2 3rd plural in r This feature is seen in Yaghnobi -or, Khwarezmian -ri, and Khotanese Saka -are (Schwartz 1969).

3.2.1.2.3 Causative in *-āuajā- Causatives and denominatives in *-āuajā- are identified for Khwarezmian and Khotanese Saka in Schwartz 1969. Present also in the Wakhi causative suffix -0v-.

3.2.1.2.4 The ezafe clitic The majority of West Iranian languages share a number of morphosyntactic properties. A salient one is the use of the ezafe clitic =e/=i. This feature is generally absent in Balochi, though it appears in collocations borrowed from New Persian. Parthian of the Sasanian period has an ezafe clitic with the shape cy /ce:/ (Boyce 1964:31). In Middle Persian, Parthian, New Persian, Kumzâri, Bûskârdi, Lûrestâni, Kurdish, Zâzâki, Semmani, Gorani, and Central Dialects, the ezafe is head marking. In Zâzâki, the clitic inflects for gender. The following example shows a New Persian noun phrase with an ezafe.

\[(3.1) \text{gol} =e \text{ sorx} \]

\[=E\text{EZ} \text{ red} \]

‘red flower’
In Taleshi, Tat, Siv, Az, Casp, it marks the dependent, as seen in the following Taleši noun phrase (taken from Lecoq 1989:299):

(3.2) \( \text{isbi } =_a \text{ asb} \)
white =EZ horse
‘white horse’

This innovation postdates the Old Iranian period. The development of the \( eزafe \) can be traced from Old Persian to Middle Persian via the grammaticalization of the relative pronoun \( haya- \) into an enclitic. Constructions like \( \text{Gaumata haya maguš ‘Gaumata the Magian’ (Darius, Behistûn, passim) developed into constructions like MMP š’h-y wzrg ‘noble king’.} \)

Windfuhr (2009:28) attributes the right-branching syntactic structure of Persian to Elamite influence. Hence, it is clearly not a shared genetic innovation.

3.2.1.2.5 The \( m\overline{e} \)-imperfective While the \( eزafe \) postdates Old Iranian, the \( m\overline{e} \)-imperfective postdates Middle Iranian, being grammaticalized into New Persian from MP \( hmy \) ‘forever’ (Windfuhr 2009:24). All SW Iranian languages have this prefix, as does Mazandaraní; Kurdish, Balochi, and Zazaki have other means of marking imperfective verbal forms.

3.2.1.2.6 Preverbal \( b\overline{i}/b\overline{e} \)-subjunctive/perfective The preverb \( b\overline{i} \) marked the perfective in Early New Persian (Windfuhr 2009:24). In later New Persian, however, it took on a subjunctive role, seen also in Kurdish, Zazakî, and Balochi. In Mazandaraní, it retains its perfective-aspect-marking role.

3.2.1.2.7 Object marking with \( r\overline{a} \) Middle Persian \( r\overline{a}y \), grammaticalized from OP \( r\overline{a}dy \) ‘on account of’ (Windfuhr 2009:33) marked both indirect objects and direct objects (to a slightly lesser extent) and had a few other grammatical functions as well. The indirect-object-marking function is absent in NP, but retained in Tajik Persian. This particle exists in other New West Iranian languages, but not Kurdish. In Zazaki, it only marks the indirect object.

3.2.1.2.8 The 3rd person singular pronominal enclitic (Windfuhr 1989:259) gives the following distribution for the 3rd sg. pronominal enclitic: “*\( hai \) in Avestan (besides *\( šai \)), Kurdish, Harzani in the North, Khuri, Sedehi, Abyăne‘i...Bandar Abbasi, Minabi, Baškardi, Kumzārî, Balōcī in the Southeast vs. \( šai \) in Old Persian and the remainder of West Iranian.” I do not include this character because it is not entirely clear what areal information it bears.

3.2.1.2.9 Constituent Order This is not a very informative character, as all languages observed have an unmarked SOV order.
3.2.2 Lexical Characters

Iranists often point to an East/West division in vocabulary, but attempts to define such a lexical split give way to many exceptions (these were discussed in the previous chapter; I briefly reintroduce them here). For instance, *kaufa-* is often treated as a uniquely West Iranian word for ‘mountain’, but kaofa- ‘(dome-shaped) mountain’ (Humbach and Ichaporia 1998:73) appears in Young Avestan. Sims-Williams (1989b:169) gives *gari- ‘mountain’ as one of several Eastern “retention[s] of ancient lexical items lost in WMIran.” But there is an exception to this in MP gr ‘mountain’ (cf. with different semantics, Balochi gar ‘abyss’, Korn 2005:150). Sims-Williams (1996) gives *anda- ‘blind’ as another East Iranian word, but we have also Pth ḥnd ‘blind’, Zaz agil-hend ‘truth-blind’ (Paul 1998a:175). (The Parthian word could be taken as part of the so-called “Parnian” East Iranian element in the language, likely an imprint of the Scythian Dahae.)

Traditional lexicostatistic research tends to use a 200-word Swadesh list, which contains basic vocabulary items thought to be present in any cultural context. Many have remarked upon the inabilities of the Swadesh list to capture all relevant cultural information in certain regions of the world, and have developed appropriate word lists for different areas. Ringe et al. use an extended Swadesh list containing additional items common to the Indo-European world.

In this study, I have employed a Swadesh-200 list. In reality, a large number of the characters are lexically uninformative (although, of course, the preservation of stable cognate sets is highly informative for historical phonological purposes), especially in West Iranian, where the regional vocabulary may have been shaped by Persian dominance. However, regional patterns do exist, such as the salience of *karka- ‘fowl’ in Northwest Iranian (however, languages showing this tendency, e.g., Mazandarani, were excluded from my lexical dataset due to the absence of comprehensive wordlists for many NW Iranian languages). Furthermore, some East-West differences emerge as well, on the basis of *kapa- as the default word for ‘fish’ in all East Iranian languages—except for modern Ossetic, which has kesag, which Thordarson (2009:58) connects to Wanetsi kosa ‘id.’, replacing an older form kaf (< *kapa-), found in early Ossetic versions of the Gospels of Mark and John (Bailey 1945:22).

Following Ringe et al., I employ root equations rather than whole word equations. This means that two words with the same root (and same meaning) in different languages are coded as cognate even if one has undergone some sort of morphological change, whether suffixation, prefixation, or analogical change; e.g., Avestan xraatu- and Old Persian xratu- ‘wisdom’ would be coded as cognates, despite the fact that the latter underwent leveling of a weak stem. Similarly, forms suffixed with -aka- were not coded differently from those without (e.g., Persian mard, Tajik mard, Larestâni mard, but Kumzâri murtk ‘man’), given that use of *-aka- in Indo-Iranian can be seen as “shading into pleonastic meaninglessness” (Jamison 2009) and is widespread throughout the branch. Interesting instances of analogy or morphological derivation were observed in the morphological datasets. The methodology of accepting only root equations may seem unsatisfying to some: at the moment, however, there is no principled cutoff point at which two derivationally related forms containing the same root (e.g., PIE *dʰǵʰ omen- ‘of the earth, earthling, human’) can be
considered separate lexical items (Trask 2000:234 introduces the term “oblique cognates” to refer to relationships like the aforementioned).

The lexical dataset was assembled as follows: The data for Avestan, Old Persian, Wakhi, Balochi, Persian, Waziri, Pashto, and Ossetic was taken from IELEX\(^2\) (Dunn and Ludewig 2012). This data, much of which comes from Dyen et al. 1992, was recoded, since much of the coding for Iranian in that study is incorrect (e.g., NP dil ‘heart’ is given a different character state than Pashto zoř ‘id.’, but both reflect Pir *źrd*-). To this data I added Zazaki (Paul 1998b), Kumzari (Thomas 1930), Baškardı (Skjærvø 1988), Larestani (Kamioka and Yamada 1979), Bandari (Pelevin 2010), Ormuri, Parachi (Morgenstierne 1929), Ishkashmi, Yazghulami, Zebaki (Grierson 1920), Shughni (Zarubin 1960), Yaghnobi (Andreev and Peščereva 1957), Sogdian (Gershevitch 1954; Sims-Williams 1989b,e; Gharib 1995), Middle Persian, Parthian (MacKenzie 1971a; Boyce 1977), Khwarezmian (Humbach 1989), Bactrian (Sims-Williams 1989a), and Khotanese Saka (Bailey 1979). Of these additions, several factors led to potential inconsistencies between datasets: some languages, such as Bactrian, have considerably smaller corpus sizes; some, like Khwarezmian, lack complete dictionaries; and some (Khotanese Saka, Middle Persian, Parthian, Ormuri, Parachi) lack bidirectional word lists. Additionally, the bulk of Middle Persian and Parthian forms were taken from the vocabulary of the bilingual Manichean Turfan texts, and are thus quite similar to each other. Some languages were omitted due to a lack of comprehensive resources—I regret having to omit Yidgha-Munjı, and would have liked to include additional NW Iranian languages, but for sparse data.

In coding the lexical dataset, it was crucial to maintain awareness of loanwords. In one school of thought, the adoption of a loanword is treated strictly as a change in state, and cannot be modeled as a shared genetic state with the loanword’s donor, or with other languages that have borrowed the same item. For example, Persian, Yaghnobi and Wakhi dil ‘heart’ must be assigned separate character states, since the latter two languages have borrowed the word from Tajik Persian. The same is true of NP daryā, Psht daryob, Shg daryō, Ygh dariyo ‘river’; we expect d for Pir *ź* (here, the etymon is *źraːj*ah-) only in SW Iranian (and only part of the time!)\(^3\).

Persian loanwords are found throughout modern languages spoken in Iran and Tajikistan. Additionally, Ormuri and Parachi have borrowed from Pashto at different chronological stages, and the older borrowings are quite difficult to distinguish. Difficulty arises in Pamir languages as well, as loans are often altered phonologically, e.g., Shughni nařq ‘ornament’ < Tajik Persian naqš (Edel’man and Dodykhudoeva 2009:791).

Persian loans are difficult to identify in the absence of key Persian sound changes. For this reason gul, the word for ‘flower’ in over half the Iranian languages, is clearly Persian, as it comes from Pir *u(a)rda- and shows the distinctive Persian sound change *rd > l. Similarly,

\(^2\)http://ielex.mpi.nl/

\(^3\)The word *źasta- ‘hand’, which appears with a z only in Avestan has undergone dissimilation (perhaps from an old affricate [dz]) of the first consonant to d in the overwhelming majority of Iranian (Morgenstierne 1927:39). Martin Schwartz brought my attention to a 2006 conference presentation by Hamasik Kirakosian which mentioned an instance of zast in the Old Azari Fahdevayat, a genre of “Neo-Median” verse, but Dr. Kirakosian (p.c.) confirmed our suspicion that this was a mispointing of the Arabic letter dāl.
NP aftāb ‘sun’ (< *abi-tapa-) appears in Pamir languages, e.g., Shg őftōb.4

In SW Iranian languages like Kumzari, Bashkardi, and Larestani, it is often difficult to tell whether something is a Persian loan or an inherited word (given that the historical phonology of these languages is very close to that of Persian). In this case, asking whether Bandari müği ‘fish’ is a Persian loan (< māhī) is reminiscent of the philosophical question regarding the audibility of a tree falling in a lonely forest. In many cases, the lexical field is one into which other languages have borrowed a word from Persian, but the phonology is not informative; I treat cases like this as inheritances.

There are additionally paradoxical situations like that of NP supurz ‘spleen’ versus Kd sipil ‘id.’ < PIr *sprzan-. We expect PIr *rź > OP rd > MP, NP l against NW Iranian rz (e.g., the latter member of the NP doublet bala ∼ burz ‘height’ is generally ascribed to Median influence, cf. Beekes 1997:3), but here the roles are reversed. (On the outcome of PIr *rd/rź in Kurdish, recall MacKenzie 1961:78: “I do not think it is possible to be certain which is the true Kurdish development, but whether we consider the many words with l/t as native or loan-words their preponderance is significant.”) The Swadesh-200 word list does not contain ‘spleen’ (though it contains ‘liver’, Yaghnobi šupurda5), and there is no comparable situation in my dataset, but it is not clear how such a problem should be dealt with. We could assume that the Kurdish word is a Persian loan, while New Persian has borrowed from a NW Iranian language (perhaps Kurdish). Thus we would code two separate character states, despite the fact that Kurdish sipil affirms the presence of a reflex of *sprzan- in Persian, and vice versa!

Some difficult cases of loan identification are listed at the end of this section.

It is not always appropriate to code loanwords (see Chang et al. 2015:205 for discussion). And while it is crucial to code an Albanian borrowing from Latin (e.g. L canis → Alb qenē ‘dog’) as a separate rather than a shared state, lest the software infer that the two languages are more closely related to each other than to other Indo-European languages, in the case of Iranian, it may not be as crucial, since in a given non-Persian language, a form borrowed from Persian often stands alongside a cognate “native” form (see for instance Shugni yak ∼ yīw ‘one’, discussed below) and in many cases, may have had a prehistoric form of the same cognate class. For instance, Ormuri zubān, Parachi (zu)bān ‘tongue’ are identified as Persian loans (< NP zubān) by Morgenstierne (1929:302, 413) and poetic forms (I believe this to be the case as well, since they are virtually identical to NP, and all other East Iranian languages reflect either PIr *hiźu- or *hižuaka-, with *hižuāna- found across West Iranian), but even if the Persian loan replaced an earlier word for ‘tongue’, it is highly unlikely that the earlier, “native” form would have reflected an etymon other than *hižu(ka)-. So the change of state due to the loan may not be highly relevant.

4 Though the evidence is somewhat murky, we generally expect the *-b- of *abi- to be voiced in Pamir languages, e.g., Yazghulamī aṣeṇ `bridle’ < *abi-dān(d)ja-, avšūst ‘mitten’ < *abi-dasta- (Edel’man 1980:303), Šuṃni biṣiṣ- ‘swell, inflate’ < *abi-aič-(-sa-), bidāj- `irrigate’ < *abi-tač- (Cheung 2007:127).

5 This form shows the nature of complications posed by loanwords—this form has the distinctly Yaghnobi trait of palatalizing earlier *sp < PIr *sū (seen also in Ishkashmi and Pashto), but has the distinctly Persian trait of showing d for PIr *z. Fortunately here, Yaghnobi is alone in our sample of languages in undergoing the change spleen > liver, and receives a unique character state regardless.
By comparing the results achieved by coding loanwords versus leaving loanwords uncoded, we can observe the influence of language contact in shaping West Iranian’s lexical profile. If a West Iranian clade is produced when loans are coded, this provides a degree evidence (albeit non-traditional evidence) for a West Iranian genetic subgroup within Iranian. If such a clade is produced only when loans are coded, we can infer that in terms of lexical replacement, West Iranian makes more sense as an areal group. (A third possibility is that neither configuration will produce a West Iranian subgroup, indicating that there is neither a genetic nor an areal basis for a West Iranian group, as far as vocabulary is concerned.)

Polymorphisms were suppressed. If a language showed multiple forms for one meaning, one of which was a loan, the loan was not considered. Overly artistic or specialized words were excluded as well, if necessary. My objective was to select the “default” word in each language. At times this was done arbitrarily. At times, two languages would show the same polymorphism, but none of the polymorphic forms were present in the other languages; this made it unnecessary to suppress the polymorphism. For verbs, if the cognate lexical item surfaced in the verbal predicate, it was coded as a cognate, even if not strictly a verb—e.g., Larestani ‘eskāl ‘hunt’ (n.—a verbal form was not given in the word list that I used) was coded as cognate to the NP light verb šekār kardan ‘hunt’ (v.); the Larestani light verb xana kerda ‘laugh’ (v.) was coded as cognate to NP xandīdan ‘laugh’ (v.).

3.2.2.1 Loans or Inherited Forms?

Below, I give a few examples of vocabulary items that demonstrate the difficulty of teasing apart inherited vocabulary from loans.

3.2.2.1.1 PIr *aiyaka- ‘one’ > yak Middle Persian has Pahlavi ’ywk /eːk/ (< *ēwk < PIr *aiya-ka-) and a reduced Manichean form yk /jak/. The latter form, which is reflected in New Persian as yak, must have undergone some sort of irregular phonological reduction (see also Cantera 2009:19 and the next chapter). In a number of New Iranian languages, yak is the word for ‘one’ and often appears alongside an unsuffixed form, e.g., Šuynī yīw. Are all languages which show yak defined by a multistep change from *aiyaka-, or is the form simply borrowed from New Persian or Tajik? While virtually any lexical item can be borrowed, languages in the region do not generally borrow numerals lower than five (cf. Brahui, which has borrowed all but the numbers 1-4 from Balochi ). I have chosen to code all instances of New Iranian yak for ‘one’ as inheritances, since they are unlikely to be borrowed, and may have come about due to parallel phonological innovation. In cases like that of Šuynī, where a form identical to the Persian or Tajik word stands alongside a form displaying different historical phonological and morphological developments, the polymorphism was suppressed in favor of the “native”-looking word.

3.2.2.1.2 PIr *rā̄sta- ‘right’ > rāst Kent (1951:34) considers a change of PIr *rā̄sta- → OP rāsta- ‘straight, right’ to be due to analogy with past participles of dental stems (PIE *-t̪-t̪- > PIr *-st̪-). The OP form stands against Avestan rāsta- and is continued by MP r’st and NP rāst. However, a number of other (i.e., non-SW) Iranian languages show an
almost identical form, e.g., Balochi rāst, Ossetic rast, Shughni rōst ‘right side’. In Ossetic, PIr *s and *š merge to s. But the Balochi and Shughni forms look either as if their ancestor underwent an analogical change similar to Old Persian, or as if they have been borrowed from New Persian or Tajik (the vowel in the Shughni form does not preclude this possibility, given the vowel change in āftōb ‘sun’ ← NP aftāb). While these forms look suspiciously as though they were borrowed from Persian (Persian rāst has been borrowed into Indic as well), I cannot rule out the possibility that the ancestors of Balochi and Shughni underwent an analogical change identical to that proposed by Kent for Old Persian. For that reason, such forms have been coded as inherited forms, and not borrowings.

3.2.2.1.3 PIr *hakaram > agar ‘if’  Horn (1893:25) derives NP agar from OP *hakaram ‘once’. This word appears throughout New Iranian as the word for ‘if’, and has been borrowed into Indic. Pamir languages that have aga(r) generally voice PIr *-k- intervocalically, making it virtually impossible to pinpoint phonological evidence of a loan. Though the form is certainly a loan into Indic, and discourse markers of this sort are easily borrowable within a linguistic area (cf. Turkish ama ‘but’, borrowed by a number of languages in the Balkan Sprachbund), simply looking over the forms, there is no solid evidence that Iranian languages have borrowed agar from Persian.

3.3 Results

This section presents the results of simulations based on character types discussed in the previous section.

3.3.1 Non-lexical characters

A NEXUS file containing 27 Iranian taxa and 31 binary typological/recurrent characters was read as a NeighborNet file using SplitsTree (Huson and Bryant 2006). These included all of the characters in §§3.2.1.1-3.2.1.2. The network produced by these features can be seen in Figure 3.1. A relatively clear East-West patterning is visible, though this division lacks split strength, given the large amount of reticulation or webbing along the interior of the network.

Much of the branching structure is unresolved in the network diagram. It is then useful to obtain a summary tree for such data, using a method where a tree with binary branching is forced. To do this, I processed the same dataset in BEAST version 1.7.3 (Lemey et al. 2009), a software program for Bayesian phylogenetic analysis with Markov chain Monte Carlo (MCMC) sampling of the posterior distribution. I excluded non-contemporary languages from this simulation, since Old and Middle Iranian languages tend to form clades according to their chronological level. The sampling process ran for 10000000 iterations, with a thinning interval of 1000. A posterior tree sample was produced (with the first 1000 samples discarded as burn-in), from which a summary tree was returned using the program TreeAnnotator version 1.7.3 (Heled and Bouckaert 2013). I used binary Stochastic Dollo Characters (SDC), which follow a probability distribution which penalizes a character state change of 0 to 1.

48
This character type is not always suited to linguistic studies of this sort (see discussion in Chang et al. 2015:217), but serve the purpose of an exploratory study like this chapter’s. When higher order subgroups in a set of languages under investigation are known a priori, it is ideal to use a character type (such as Restriction Site Characters) which allow for parallel character changes across languages. But in cases where higher order subgroups remain controversial, as is the case with West/East Iranian, it can be highly informative to make use of a character type that places restrictions on the number of times a trait can be born, like SDC. A coalescent tree prior was used. (The foregoing specifications were used for all BEAST simulations described in this chapter.)

The summary tree for the non-lexical data is shown in Figure 3.2. Clear East and West Iranian clades are visible and annotated.

It becomes clear that East Iranian and West Iranian clades can easily be produced from recurrent characters whose status as subgroup-defining remains controversial.

3.3.2 Lexical characters

Two datasets based on lexical characters were processed in BEAST version 1.7.3 (Lemey et al. 2009), a software program for Bayesian phylogenetic analysis with Markov chain Monte Carlo (MCMC) sampling of the posterior distribution. A posterior tree sample was produced, from which a summary tree was returned using the program TreeAnnotator (version 1.7.3). I used binary Stochastic Dollo Characters (SDC), which follow a probability distribution
which penalizes a character state change of 0 to 1. This character type is not always suited to linguistic studies of this sort (see discussion in Chang et al. 2015:217), but serve the purpose of an exploratory study like this chapter’s. When higher order subgroups in a set of languages under investigation are known a priori, it is ideal to use a character type (such as Restriction Site Characters) which allow for parallel character changes across languages. But in cases where higher order subgroups remain controversial, as is the case with West/East Iranian, it can be highly informative to make use of a character type that places restrictions on the number of times a trait can be born, like SDC.

In the first dataset, loans were coded when recognizable. In the second, loans went
uncoded. Since languages of like chronological periods will often cluster together, often defying genetic affiliation, I thought it prudent to observe (i) languages of all chronological periods and (ii) only contemporary languages in the simulations of interest.

3.3.2.1 Languages of all chronological periods

3.3.2.1.1 Loans coded The summary tree for the dataset of languages of all chronological levels in which loans were coded is shown in Figure 3.3.

A SW Iranian clade within a West Iranian clade is produced, but no NW Iranian clade. No East Iranian clade is produced.

3.3.2.1.2 Loans uncoded The summary tree for the dataset of languages of all chronological levels in which loans were uncoded is shown in Figure 3.4.

Both a SW Iranian and a NW Iranian clade are produced within an West Iranian clade. No East Iranian clade is produced.

3.3.2.2 Contemporary languages

3.3.2.2.1 Loans coded The summary tree for the dataset of only contemporary (i.e., modern) languages in which loans were coded is shown in Figure 3.5.

A SW Iranian clade within a West Iranian clade is produced, but no NW Iranian clade. No East Iranian clade is produced.

3.3.2.2.2 Loans uncoded The summary tree for the dataset of only contemporary languages in which loans went uncoded is shown in Figure 3.6.

Again, a SW Iranian clade within a West Iranian clade is produced, but no NW Iranian clade. No East Iranian clade is produced.

3.3.2.2.3 Distance-based modeling of lexical character data So far, Bayesian phylogenetic inference using lexical character data has consistently produced a West Iranian clade. This clade is at least compatible with a West Iranian genetic subgroup, but there is no particularly strong supporting evidence for a West Iranian subgroup, outside of lexicostatistic experiments. However, some historical linguists tend to doubt the idea that clades based on lexical character data necessarily correspond to genetic subgroups; they could just as well be areal (Donohue et al. 2012).

I was unable to apply methodologies from Donohue et al. 2012 to this chapter’s data, due primarily to a well-established, strongly articulated subgrouping of Iranian languages against which to compare these results. However, as a final means of gauging the “subgroupiness” of the lexical character data, I created a NeighborNet network in SplitsTree using the same data (for contemporary languages, with loans coded). It is evident from this graph (shown in Figure 3.7) that West Iranian does not split off cleanly from the rest of Iranian, and in fact shows a great deal of webbing, consistent with an areal network.
3.3.2.3 Summary

Like Blažek’s (2013) results, the foregoing simulations consistently produced a West Iranian subgroup, but no East Iranian subgroup with a common ancestor to the exclusion of West Iranian. A West Iranian clade was present both when loans were coded as such, and when they were uncoded.

The consistent presence of a West Iranian clade across these two datasets could indicate
that there is a reality to West Iranian as a genetic subgroup. However, there are alternative explanations as well. Similarities in patterns of lexical replacement across West Iranian could be due to precursor-driven drift, or even advergence, “process of mutual influence when two separate languages, which are in fact genetically related through descent from a common ancestor, occupy adjacent territories and continue to interact” (Renfrew 2000:14). In fact, loanword transfer between West Iranian languages may be of such an archaic date in some cases that subsequent sound changes have made it impossible to distinguish lexical items as
loans. Unfortunately, the lack of a documented historical record for languages other than Persian make it difficult to tell whether a West Iranian subgroup is supported by lexical data at all points throughout history, and is not in fact a product of advergence or prolonged contact.

3.4 Discussion and Conclusion

This chapter’s results can be summarized as follows:

1. Bayesian phylogenetic inference using lexical character data consistently produces a West Iranian clade

2. Bayesian phylogenetic inference using non-lexical character data consistently produces both an East and a West Iranian clade
3. Distance-based analyses of lexical character data produces an East-West division with no clear split strength.

4. Distance-based analyses of typological and other non-lexical recurrent characters produces an East-West division, albeit with a great deal of webbing around the center of the network.

The network representing the non-lexical data shows a pattern usually associated with areal clustering, for characters usually thought to display an areal signal (Donohue 2012). The network representing lexical data does not show a particularly tree-like structure either. At the same time, both datasets produce East and West Iranian clades when subject to Bayesian phylogenetic analysis. The West Iranian clade that produced is, as mentioned above, compatible with a West Iranian genetic subgroup, but there does not seem to be much in the way of external genetic support.

The fact that East Iranian forms a cohesive group in one scenario (non-lexical characters), but not in the other (lexical characters) requires comment. It is sometimes argued that
lexicostatistical results pick up an areal, rather than genetic signal; the same is argued for typological characters. If both statements are true, why then should East Iranian show the effects of one type of areality, but not the other? It may be the case that archaisms are contributing to the East Iranian typological clade (e.g., the presence of θ), or that some similarities across East Iranian are due to parallelism or drift. It could then follow that the cohesiveness of West Iranian, both in terms of its lexical and non-lexical profile, is due to more recent contact, and possibly the propagation of more innovative typological features.

However, the results of this chapter, particularly the lexicostatistical studies, contradict the results of the previous chapter. In this chapter, a West Iranian clade was consistently produced, and lexical character data failed to produce an East Iranian clade. The findings of the previous chapter show, conversely, that as far as traditionally reconstructible innovations go, the West Iranian ones are easiest to attribute to natural tendencies of morphological change (e.g., *hiźyan- ‘tongue’, devoicing of clusters where Bartholomae’s Law operated). I had greater difficulty explaining away East Iranian innovations such as the lexical item *kapa- ‘fish’. However, if the development of *kapa- is a shared East Iranian innovation, it is at odds with the fact that this chapter’s results do not support the idea of a unified lexical profile
for East Iranian. It is worth noting that in some areal groups and Sprachbünde, languages converge in terms of structural features, but not necessarily lexis (François 2011). It could be the case that this is what happened in East Iranian (i.e., structural but no real lexical convergence), and that *kapa- is one of very few East Iranian lexical innovations. Either way, these contradictions make it difficult to say anything about the genetic status of East or West Iranian at the moment.
Chapter 4

Phonological reduction and grammatical convergence in West Iranian

4.1 Introduction

This chapter investigates the development of functional items in West Iranian languages. Across West Iranian, functional items in the same etymological set tend to be phonologically quite similar, and are often identical. A longstanding question in the literature asks whether this formal congruence is due to (1) shared tendencies among these closely related languages (i.e., parallelism) or (2) contact (i.e., borrowing between languages). In order to address this question, I use a novel methodology designed to detect contact between closely related languages, and find that sociolinguistic pressure has played a significant role in the development of similarities among West Iranian functional items.

West Iranian is a group of typologically similar and closely related languages that have been in contact for most of their development. The most widely spoken West Iranian language is New Persian (NP, a label subsuming both Classical and Contemporary varieties). New Persian has exerted a great deal of influence on other West Iranian languages over history, and continues to do so today. Loans from Persian are common in non-Persian West Iranian languages (as well as in East Iranian, Indic, and Turkic). It is easy to identify a Persian loan into a non-Persian West Iranian language if the form in question appears to show one or more distinctive Persian sound changes that would be “irregular” in the non-Persian language, yielding a “significant result” as described by Bloomfield (1933:361), in his assessment of the debate between neogrammarians and dialect geographers: neogrammarians, assuming the regularity of sound change, explain many instances of apparent irregular change via borrowing between dialects, whereas dialect geographers believe in irregular sound change; Bloomfield criticizes the latter group for occasionally appealing to dialect mixture despite

---

1This chapter reproduces, by kind permission of Brill, most of the contents of Cathcart forthcoming (copyright Koninklijke Brill NV, Leiden, 2015).
the fact that “[i]n many other cases, however, they prefer to say that irregular sound-change was at work, and, strangely enough, they do this in cases where only the neo-grammari- 

This sort of appraisal of irregular-looking change is not so easy to carry out for the 

To address this question, this chapter employs a quantitative methodology designed to 

Using what is known about the regular historical phonological behavior of these languages, I generate expected forms of these functional items that would have hypothetically resulted via regular sound change alone. I measure distances between these expected forms and the observed forms that actually occur in West Iranian languages. I integrate these distances into a quantitative model that compares degrees of irregular sound change across West Iranian, particularly with respect to Middle and New Persian. Results (§4.6) show that while most West Iranian languages undergo degrees of irregular sound change that are not significantly different from Persian (Middle or New, depending on their chronological period), Balochi, a West Iranian language with particularly conservative historical phonology, appears to undergo significantly more irregular phonological change than New Persian; this large difference in irregular sound change results in forms which are mostly equivalent to their New Persian cognates. I suggest that this points to the sociolinguistic pressure of New Persian on Balochi, and serves to show that Balochi has borrowed a large portion of its functional items from Persian during its history. This is similar to Bloomfield’s reasoning: Bloomfield argues that in the face of synchronic irregularity, hypotheses invoking dialect borrowing are more meaningful than those that appeal to the idea of sporadic sound change. Similarly, I suggest that borrowing is more likely than inflated degrees of irregular sound change in a section of the lexicon (i.e., functional vocabulary) where irregular sound change is the norm.

This chapter is structured as follows: §4.2 lists the languages under study in this chapter and describes the genetic and sociolinguistic relationships between them; §4.3 discusses irregular sound change as it applies to this chapter; §4.4 gives a detailed outline of the historical phonology of the functional items under study; §4.5 outlines methodological
preliminaries pertaining to the study; §4.6 discusses modeling and results, showing that Balochi appears to undergo a significantly greater degree of irregular sound change in the development of functional items identical to their New Persian counterparts (suggesting borrowing); §4.7 investigates whether this result is an effect or correlate of a wider tendency toward non-conservative behavior on the part of non-literary languages like Balochi, but finds that Balochi is in fact more conservative in its historical phonology than New Persian.

4.2 West Iranian

As noted in previous chapters, the West Iranian languages are a typologically close-knit group of related languages spoken in Central and Southwest Asia. Within the West Iranian group, Southwest Iranian forms a subgroup defined by key inherited phonological and morphological innovations shared by the languages in it. The remaining West Iranian languages are called Northwest Iranian, and as a group are not defined by any shared genetic innovations. Tedesco’s (1921) monograph is the locus classicus of West Iranian dialectology, establishing a number of phonological and morphological isoglosses cutting between Northwest and Southwest Iranian; West Iranian dialectology has been subsequently discussed in length in the works of Lentz (1926); MacKenzie (1971a); Krahnke (1976); Paul (1998a), and Korn (2003), inter alios. The following is an abbreviated list of West Iranian languages, organized by dialect group and chronological level. Languages treated in this chapter are in italics.

Southwest Iranian:
- Old: Old Persian
- Middle: Middle Persian (Durkin-Meisterernst 2004)
- New: New Persian (Steingass 1892; Lazard 1963), Kumzari, Bashkardi, Luri, Bandari

Northwest Iranian
- Old: Median (limited to onomastic remnants)
- Middle: Parthian (Durkin-Meisterernst 2004)
- New: Balochi (Raxšanî dialect of Western Balochi, Barker 1969), Kurdish (Sulemaniya subdialect of Sorani dialect, McCarus 1958), Zazaki (Paul 1998a,b), Mazandarani (Nawata 1984; Borjian and Borjian 2008), Sangesari (Azami and Windfuhr 1972), Yazdi (i.e., Zoroastrian Dari, Ivanow 1940; Vahman and Asatrian 2002), Auromani (Benedictsen and Christensen 1921), Taleshi, Tati, Gilaki, etc.

Old, Middle and New Persian are in a relationship of direct descent. However, no Northwest Iranian language has been demonstrated to be a direct descendant of any other attested Iranian language.

Figure 4.1 shows approximate locations where these languages are or were spoken. West Iranian languages have been in at least sporadic contact for most of their development; while
loanword transfer between Northwest and Southwest Iranian has been bidirectional (e.g., Lentz 1926), Persian has exerted intense sociolinguistic influence on other West Iranian languages (Borjian 2009).

As mentioned above, it is easy to identify a Persian loan into a non-Persian West Iranian language if the form in question appears to show one or more distinctive Southwest Iranian innovations. For instance, the regular New Persian outcomes of Proto-Iranian *ź and *rd are d and l respectively, as opposed to z and rd in Balochi, another West Iranian language. The New Persian reflex of PIr *źr- is dil ‘heart’. Balochi has a doublet consisting of dil ∼ zird for the same etymon. If we were naïve to New Persian historical phonology, we might say that Balochi dil appears to have undergone an irregular sound change of some sort, zird being the expected form. However, given what we know regarding Persian sound change, we can easily identify the former word as a Persian loan into Balochi, yielding a “significant result” as described by Bloomfield.

For the functional items under study in this chapter, the picture is not so clear as in the example given above. For instance, the New Persian definite object marker -rā (< PIr *rādī) has undergone no irregular sound change in its history, but its equivalent counterpart in Balochi, -rā, appears to have undergone an irregular elision of *-d; via regular sound change, we expect PIr *rādī to become Bal ŏ-rad. Since the elision of final *-d is irregular in Balochi but regular in Persian, we could conclude that -rā was simply borrowed from New Persian by Balochi. However, a grammatical morpheme like -rā is precisely the sort of item in which we might expect an irregular phonological change (like elision of *-d) to occur. Thus, on a case-by-case basis, it is not clear if the appearance of an irregular sound change in a non-Persian West Iranian language’s functional item points to borrowing (from Persian, most likely), or the type of sporadic change that is common in grammatical morphemes.
4.3 Irregular Sound Change

Historical linguists today generally allow for the same exceptions to the regularity of sound change that the neogrammarians did, acknowledging that apparent irregularity can be due to borrowing or analogical change, as well as sporadic phonological reduction.

Irregular phonological reduction (alternatively known as phonetic reduction or erosion, though these terms can refer to regular developments as well) is a frequent correlate of the grammaticalization process. It is generally viewed as a possible byproduct of the process, following (or occurring during the final stages of) the development from lexical item to functional item (Heine and Kuteva 2009:35). Reduction can also occur in contexts independent of the grammaticalization process, such as spontaneous speech. Irregular phonological reduction is manifested segmentally in different ways. It can involve vowel contraction, e.g., Old English nāwiht ⇒ Modern English nought; it can involve elision of segments, or both; it can involve gestural overlap, e.g., Modern English did you ~ [d3dɔ] ∼ [d3]̅; and it can involve combinations of these phenomena, e.g., Basque ez dakit ‘I don’t know’ ∼ [stait] (Trask 2000:15).2

For the purposes of this chapter, all types of irregular phonological change (i.e., irregular phonological reduction, analogical change, dialect borrowing, etc.) are bundled together and modeled as any sort of deviation from the expected regular sound change that a given form in a given language ought to undergo (cf. Joseph 2004, where reduction is viewed as a type of analogy). Given the nature of phonological change in West Iranian functional items (and the absence of distinctive, language-specific sound changes), different types of irregular phonological change are qualitatively difficult to tease apart. But this is not a problem for my analysis, since I seek not to measure amounts of irregular phonological reduction alone, but skews in the amount of irregular change undergone by a language in a domain of the lexicon where irregular change is common, thus satisfying my adaptation of Bloomfield’s diagnostic.

4.4 Historical Phonology of Persian Functional Morphemes

The data analyzed in this chapter consist of thirteen Persian functional morphemes and their plausible cognates3 in other West Iranian languages (if they exist; I deal with complications

---

2I use > to represent a regular diachronic development, → to represent an analogical change or borrowing, and ⇒ to represent an irregular phonological reduction (or excrescence).

3I use the term “plausible cognate” to refer to members of the same etymological set. It may be the case that some of the forms listed in this section are borrowings from New Persian, and hence not legitimately cognate to their New Persian counterparts, but for reasons discussed in this section, it is difficult to tell. In this study, I make no a priori assumptions as to whether a functional item has been borrowed by New Persian; I am concerned only with the difference between the forms that we see in a language and the forms we would expect via regular sound change.
In this section, I provide an outline of the historical development of these forms. The fact that Persian is attested at three chronological levels makes it possible to tell at what chronological stage phonological reduction took place in the grammaticalization process (if at all, and if grammaticalization took place at all); for this reason, I outline the Persian forms’ historical phonology in detail, noting whether irregular reduction appears to have taken place (underlined segments in the Persian forms denote segments that have been irregularly elided or inserted), but do not devote the same attention to non-Persian forms, whose history is less well documented. For each Persian item listed below, I give the Proto-Iranian etymon (in cases where it is reconstructible), the Persian historical phonology, and the Persian forms’ plausible cognates in other West Iranian languages. Unglossed items are synonymous with the preceding item.

Of these forms, three show no irregular phonological reduction in Persian. These are:

1. NP -rā (definite object marker)
   - PIr *rādi ‘sake, account’ loc. sg.
   - OP rādiy ‘sake’ > MP rāy > NP -rā definite object marker
   - Pth rāō ‘on account of, for the sake of, because of, for’, Bal -rā indirect/direct object marker, Maz -ra IO/DO marker, Zaz -rē, -rā IO marker, Yzd -ra oblique marker, Awr -ra IO marker

   - PIr *aiya-čī-
   - OP *aiya-čī ‘one thing’ > ēwč (with syncope, cf. Skjærvø 2009a:201) > MP ēc ‘any’ ⇒ NP hīč ‘anything’ (with Cockney-style prothesis)

3. NP har ‘each, every’
   - PIr *haryā-
   - OP harua ‘all’ > MP harw ‘all, every’ > NP har ‘each, every’

---

4I have been forced to exclude a couple of notable items due to etymological uncertainties. The ezafe particle is common to most West Iranian languages. The Persian development is as follows: PIr *ja- (relative pronoun) ⇒ OP haya- ⇒ MP -i(y) (ezafe) > NP -i. However, we do not know whether other West Iranian ezafees (if not simply borrowed from Persian) developed from a relative pronoun *ja-, or a form like OP haya- (an analogical innovation, Adiego Lajara 2000). Additionally, a verbal prefix bi- which marks the subjunctive or perfect is common to various West Iranian languages. It is unclear if these verbal prefixes are etymologically the same morpheme, and what its etymology is: Horn (1901:161) identifies the preverb *u-, roughly meaning ‘away’; Josephson (2013) connects it to MP bē- ‘out, away’ (additionally, it bears a curious resemblance to NP bāyad (ki) ‘it is necessary (that)’).

5All but the attested Old Persian transcriptions have been normalized (OP c = ē, OP iy = ĭ, OP uv = ū).
• Pth harw ‘all, every’, Kd har ‘each, every’, Maz har, Sang har, Yzd har

The remaining ten show some degree of irregular phonological reduction and/or analogical change in the history of Persian:

4. NP az ‘from’
   • PIr *hač¯ a ‘in association with’
   • OP ḥac¯ a ‘from’ ⇒ MP az > NP az
   • Pth až, Bal aš (Makrani ač), Kd ž, Maz -j¯ a, Sang az, Yzd az, Awr ja

5. NP ki (complementizer)
   • PIr *kahja
   • OP kahya ‘what’ gen. sg. ⇒ MP k¯ e ‘who, which’ ⇒ NP ki (complementizer)
   • Pth k¯ e ‘who, which’, Bal ki (complementizer), Kd ka, Maz ke, Zaz ki, Sang ku, Yzd g¯ i, Awr ka

6. NP agar ‘if’
   • PIr *ha-kr(-)t- ‘once’
   • Pre-OP *ha-kr(-)t→ OP ḥakaram (on the basis of kara- ‘maker’, Kent 1951:212) ⇒ MP agar ‘if’ > NP agar
   • Pth ag, Bal agar, Kd agar, Maz agar, Zaz eger, Sang aga(r), Yzd agar, Awr agar (ar)

7. NP ba- ‘with’
   • PIr *pati-
   • OP pati(y) ‘against, near, upon’ ⇒ MP pad ‘to, at, in, on’ → Early NP ba(d) (cf. Lazard 1963:248)6 ⇒ Modern NP ba
   • Pth pad ‘to, at, in, on’, Bal pa ‘for’, Kd ba ‘in, at; by; to’, Zaz bi ‘with, through’, Sang ba ‘with, in, at, through’, Yzd ba ‘to’, Awr ba, pa(δ/y) ‘to, in, on, for’

8. NP dīgar ‘other’
   • PIr *duitija-ka(ra)-
   • OP *dāutižakāram ‘2nd time’ ⇒ MP ḏudīgar ‘2nd’ ⇒ NP ḏīgar ⇒ ḏīgar ‘other’
   • Pth bidīg ‘2nd, other’, Bal digar ‘other’, Maz dogar, Sang digar

9. NP mē- (imperfective prefix)

---

6 This change is likely due to the analogical influence of other prepositions like NP (a)bā ‘with’ (< MP abāg), bē ‘without’, bar ‘upon’ (Martin Schwartz, p.c.).
• Plr *ham-aiu-
• OP *hamaiwa- > MP hamèw ‘always’ ⇒ NP me- imperfective prefix
• Pth hamèw ‘always’, Sang mî (imperfective prefix), Awr ma-/mø-

10. NP magar (adversative connective)
• MP ma__agar ‘hopefully not’ ⇒ NP magar ‘but’
• Bal magar, Maz mago, Yzd mager, Kd magar ‘unless, if not’

11. NP yak ‘one’
• Plr *aiu-ka-
• OP *aiuakahia- gen. sg. (aiva- attested) > *ěiakè (adapted from Cantera 2009:19)
⇒ MP yak8 (alongside regular ēk) > NP yak
• Pth ēw; Bal yak (cf. ēwak ‘alone’); Kd yak; -ēk; Maz atta, yak; Zaz žew/žū; Sang yakka; Yzd yak; Awr yak, yoå

12. NP injâ ‘here’
• MP ēn gygë ‘this place’ (⇐ virtual *(h)aina- [a demonstrative not yet present in Old Persian] + OP *niyaka-) ⇒ NP injâ ‘here’
• Maz inje, Zaz injâ, Sang anjû, Awr injâ, Kd injâ ‘then, in that case’

13. NP dar ‘into’
• Plr *antar- ‘inside, within’
• OP antar ‘within, among’ > MP andar ‘in’ ⇒ NP dar ‘in, into, within, among’
(alongside regular andar ‘in, into, within’)
• Pth andar ‘in’, Maz dar ‘inside’, Sang de ‘in, into’, Yzd dar

In some of the cognate sets given above, forms in West Iranian languages appear to have undergone degrees of irregular reduction that are equal to those shown by their Persian cognates. This is generally the case in Parthian when compared to Middle Persian. Equal degrees of irregular reduction sometimes result in formally different cognates. For instance, MP az and Pth až both show irregular elision of Plr *h- in *hača, but regular, language-specific changes to Plr *-č- result in non-identical forms. Equal degrees of irregular change can also result in formally identical cognates, as in the case of MP andar and Pth andar, which have undergone no irregular change.

7While not a functional item per se (and not the primary means of marking indefiniteness is most West Iranian languages), this item is generally of high frequency and prone to contraction (cf. also poetic Hindi/Urdu ek ∼ ik ‘one’).
8This development must postdate the development of OP y- to MP j-.
In other cases, West Iranian languages appear to have undergone degrees of irregular change that are greater than those shown by their Persian cognates. For instance, recall NP -rå (definite object marker), which has undergone no irregular sound change in its history, despite the fact that its plausible cognate, Bal -rā, appears to have undergone an irregular elision of *-d; via regular sound change, we expect PIr *rādi to become Bal ą-rād. As in the case of Bal -rā, non-Persian West Iranian forms that undergo more irregular change than their Persian cognates may be formally identical to said Persian cognates (given the regular sound changes listed in Appendix B).

On a case-by-case basis, it is difficult to interpret whether forms that appear to have undergone more irregular change than their identical Persian cognates are borrowings. For instance, it is easy enough to declare that Bal -rā is a borrowing from New Persian, but still somewhat difficult to justify; Bal -rā could just as easily be an inherited form that underwent phonological reduction during the grammaticalization process. For this reason, I aim to incorporate measures of irregular change into a quantitative model that can determine whether a given language undergoes roughly the same amount of irregular change in a subset of its functional vocabulary as its closely related, chronologically contemporary peer languages — or more. If languages undergo significantly different degrees of irregular change in the formation of similar-looking functional items, it should be cause for surprise. If we assume that under analogous circumstances, related languages tend to undergo equal degrees of irregular change, then a significant difference in degrees of irregular change could point to pronounced language contact.

4.5 Modeling Preliminaries

In order to conduct a quantitative analysis of the data described above, there exist at least two desiderata. First, an adequate means of measuring distances between observed and expected forms is required. Second, these distances need to be incorporated into a model that is capable of measuring language-specific distances, ideally with respect to Persian. Given the uneven nature of the attestation of these functional items, a model capable of accounting for missing data and group-specific idiosyncrasies is desirable. These issues are discussed in the following subsections.

4.5.1 Levenshtein Distance

A great deal of quantitative linguistic work has made use of Levenshtein edit distance as a measure of phonological difference. In its most basic form, Levenshtein distance is a count of the smallest number of deletions, additions, and substitutions needed to transform one string into another. For example, the Levenshtein distance between /dəɡ/ and /fəɡ/ is 2, as the two strings differ minimally according to two edits.

Most work to date has used Levenshtein distance to measure the synchronic distance between words within and across languages. Less common and more problematic are approaches that measure the Levenshtein distance between words in two languages in
an ancestry relationship (cf. Heggarty 2000, which discusses preliminaries to measuring phonological distance between Latin and various Romance language descendants). Such a practice seems in theory feasible to me, so long as the most recent common ancestor of multiple languages is used as the pivot of comparison. For instance, measuring the distance from Latin fundus ‘deep’ to (respectively) Spanish hondo and Italian fondo is more meaningful than measuring the distance between Proto-Indo-European *bʰudʰ-(m)n-o- and the two modern Romance cognates. The former comparison shows that Italian is more conservative than Spanish (in this particular lexeme); the latter quantifies both languages as equally innovative. Weighted distance measures can be developed on the basis of phonetic similarity, but weighted measures tend to be blind to complex diachronic trajectories.

Multiple techniques for normalizing Levenshtein distance have been proposed in the literature. Nerbonne et al. (1996) calculate relative Levenshtein distance by dividing the Levenshtein distance between two strings by the length of the longer string (also called normalized Levenshtein distance [LDN]). This can be done by dividing the distance by the sum of the length of both strings. Another method is Levenshtein distance normalized divided (Wichmann et al. 2010), defined as the average LDN for all comparisons of words referring to the same concept, divided by the average LDN for pairwise comparisons of words that do not refer to the same concept. Log-transformation is another appropriate means for normalizing positive numerical data such as Levenshtein distances, as it reduces skews introduced by outlying data points. (In the case of data containing non-negative values including zero, it is necessary to add 1 to each value being log-transformed.) Normalization of one of the types described above is crucial for the purposes of large-scale crosslinguistic work, as non-normalized distance measures will consistently be smaller for languages with shorter words on average.

I believe, however, that normalization of this type is less appropriate for the measures investigated in this chapter. In this study’s dataset, expected forms of cognates differ in length across languages. If Language A and Language B both have two edits between the expected and observed outcomes of a given vocabulary item, but the expected form is longer in Language A than in Language B, then normalization will result in a measure of less irregular change in Language A than in Language B. This issue can be circumvented by dividing distances by the length of their Proto-Iranian ancestors, or by accounting for group-level differences in length using mixed-effects linear regression models with random intercepts by functional item (see below).

There are additional complications that may arise from this measure. It is hypothetically possible that a single irregular change in a language’s distant past could lead to a large distance between expected and observed forms (in the case where a number of regular changes that otherwise would not apply operate on forms affected by this irregular change), whereas an irregular change in a language’s very recent past might not. This distance measure thus has the potential to exaggerate the amount of irregular change in the former circumstance, when both scenarios involve only one irregular change. As a means of checking the model against my own intuitions regarding sound change, I manually tallied up the number of irregular changes that are likely to have occurred in the history of each form (again, if we make no
a priori assumptions regarding borrowing) and incorporated these measures into a model alongside the distance-based ones.

4.5.2 Mixed-Effects Models

Linear regression measures the relationship between a vector of real-valued responses and \( n \) vectors of predictor values \( x^1, ..., x^n \). A response value \( y_i \) is modeled as a linear function of predictor values \( x^1_i, ..., x^n_i \), plus a value drawn from a normally distributed error term \( \epsilon \sim N(0, \sigma^2_\epsilon) \). The formula can be written as

\[
y_i = \beta^0 + \beta^1 x^1_i + \cdots + \beta^n x^n_i + \epsilon_i,
\]

where \( \beta^0 \) denotes the intercept and \( \beta^1, ..., \beta^n \) denote the slopes associated with the predictor vectors \( x^1, ..., x^n \).

Predictors can be either continuous or categorical. If a categorical predictor contains more than two levels, it is necessary to recode it as multiple binary predictors that take the value 0 or 1, each with a level-specific slope.

Thus, the formula

\[
y_i = \beta^0 + \beta^1 x^1_i + \cdots + \beta^n x^n_i + \epsilon_i
\]

also characterizes a regression with one \( n \)-level categorical predictor variable. For level \( k = 1, ..., n \), \( x^k_i = 1 \) and \( x^\neg k_i = 0 \). Models of this sort serve as the basis for analysis of variance (ANOVA).

Basic linear regression models are flat, and assume that all data are normally distributed and lack group-specific skews. Hierarchical linear models, on the other hand, are well-suited to collections of data with missing values and cluster-level idiosyncrasies. Cluster- or group-level idiosyncrasies can be accounted for by using random effects in a mixed-effects model. The simplest type of random effect is the random intercept, which serves as a group-specific adjustment to the fixed intercept \( \beta^0 \). (Random slopes are not applicable to this chapter’s data.) A random intercept takes the form of an adjustment to \( \beta^0 \) for group \( j \) that is normally distributed around 0, i.e., \( u^0_j \sim N(0, \sigma^2_u) \).

For the purposes of analyzing the data described above, a hierarchical (i.e., mixed-effects) model is a necessity. Not all thirteen functional items analyzed had cognates in all of the languages observed. Furthermore, certain functional items may undergo more irregular change than others. The exclusion of functional items of certain types in a given language might bias the language’s mean distance value with respect to the intercept.

Linear regression has a number of coding systems for categorical predictor variables. One of the most basic systems is DUMMY CODING, which compares each level of a variable to a reference level mean. This value can represent the overall mean of the response variable’s values, or the mean of these values for one specific level of the predictor variable. In this study, the mean of interest is that of New Persian. For that reason, the mean distance value for New Persian is dummy coded as the model’s intercept. The basic model used in the following sections thus takes the formula

\[
y_{ij} = \mu^{NP}_j + u_j + \beta^\text{Lang}_j \cdot X^\text{Lang}_j + \epsilon_{ij}
\]

for each functional item \( j = 1, ..., J \) in language \( i \in \{ \text{Balochi, Kurdish, Mazandarani, Zazaki, Sangesari, Yazdi, Awromani} \} \), where \( \mu^{NP}_j \) is the intercept; \( u_j \) is a deflection for each functional item \( j = 1, ..., J \); \( \beta^\text{Lang}_j \) is a \( 7 \)-length vector of language-specific slopes for Balochi, Kurdish, Mazandarani, Zazaki, Sangesari, Yazdi, and Awromani; \( X^\text{Lang}_j \) is the \( i \)th row vector in a \( 7 \times 7 \) identity matrix \( X^\text{Lang} \) (valued 1 in the \( i \)th column, and 0 in all others); and \( \epsilon_{ij} \) represents the error term. This model was implemented in the R (R Core Team 2014) package \texttt{lme4} (Bates et al. 2013) with the call
function lmer(distance \sim Balochi + Kurdish + Mazandarani + Zazaki + Sangesari + Yazdi + Awromani + (1|functional_item), data) \text{(distance here stands in for the multiple distance measures used, discussed above and below).}

4.6 Method and Results

In constructing linear regression models, I treat \textit{degree of irregular phonological change} (i.e., the Levenshtein distance between the expected and observed forms) as a function of (New West Iranian\textsuperscript{9}) \textit{language}, with random intercepts by \textit{functional item}. I seek to measure the dependent variable in four different forms, since there are multiple means of normalizing Levenshtein distances. These are:

- Raw, unnormalized Levenshtein Distance, termed LD
- Levenshtein Distance divided by the length of the functional item’s Proto-Iranian etymon, termed RLD
- \text{Log}(x + 1)-transformed Levenshtein Distance, termed LLD
- Qualitative, manually tallied counts of irregular changes (in order to determine how well automated distance measures accord with impressions regarding sound change), termed QC

Data used to generate \textit{expected} forms and perform statistical analyses can be found in Appendix B. Coefficients for each language as a predictor of each distance type are given in Table 4.1. For NP, \(\hat{\beta}\) represents the estimated intercept (i.e., the \(\mu^{NP}\) in the formula given in the previous section), but represents language-specific estimated slopes for the remainder of the languages.

The \texttt{dropterm()} function from \texttt{R}'s \texttt{MASS} package (Venables and Ripley 2002) was used to carry out the likelihood ratio test for each language. Balochi shows significantly more irregular change than New Persian for each distance measure except for RLD (LD: \(\chi^2(1) = 3.27, p = 0.07\); LLD: \(\chi^2(1) = 3.9, p = 0.04\) QC: \(\chi^2(1) = 16.7, p < 0.001\)).

Additional mixed-effects linear regression was carried out for forms that were identical to their plausible cognates in New Persian. Here, separate models were fitted for each non-Persian language with Persian dummy coded as the intercept, since different non-Persian languages have different functional items identical to their Persian plausible cognates. Coefficients for these can be found in Table 4.2. A visualization of inter-language Levenshtein distances using multidimensional scaling can be found in Figure 4.6; this figure shows how similar to each other West Iranian languages are in terms of their functional vocabularies (Zazaki is a clear New West Iranian outlier).

\textsuperscript{9}I give distance values for Middle Persian and Parthian (where the Parthian value represents the estimated intercept for MP) for purely general interest, but these numbers are not included in this chapter’s quantitative analysis.
Table 4.1: Distance between EXPECTED and OBSERVED outcomes by language, all forms

<table>
<thead>
<tr>
<th>Lang.</th>
<th>$\hat{\beta}_{LD}$</th>
<th>$\hat{\beta}_{RLD}$</th>
<th>$\hat{\beta}_{LLD}$</th>
<th>$\hat{\beta}_{QC}$</th>
<th>SE$_{LD}$</th>
<th>SE$_{RLD}$</th>
<th>SE$_{LLD}$</th>
<th>SE$_{QC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP</td>
<td>0.461</td>
<td>0.069</td>
<td>0.297</td>
<td>0.461</td>
<td>0.218</td>
<td>0.034</td>
<td>0.118</td>
<td>0.184</td>
</tr>
<tr>
<td>Pth</td>
<td>-0.061</td>
<td>-0.006</td>
<td>-0.07</td>
<td>-0.06</td>
<td>0.259</td>
<td>0.039</td>
<td>0.121</td>
<td>0.225</td>
</tr>
<tr>
<td>NP</td>
<td>1.538</td>
<td>0.226</td>
<td>0.849</td>
<td>1.153</td>
<td>0.315</td>
<td>0.049</td>
<td>0.132</td>
<td>0.240</td>
</tr>
<tr>
<td>Bal</td>
<td>0.631</td>
<td>0.087</td>
<td>0.219</td>
<td>1.106</td>
<td>0.300</td>
<td>0.050</td>
<td>0.115</td>
<td>0.269</td>
</tr>
<tr>
<td>Kd</td>
<td>0.183</td>
<td>0.024</td>
<td>0.025</td>
<td>-0.04</td>
<td>0.300</td>
<td>0.050</td>
<td>0.115</td>
<td>0.269</td>
</tr>
<tr>
<td>Maz</td>
<td>0.236</td>
<td>0.065</td>
<td>0.057</td>
<td>0.032</td>
<td>0.292</td>
<td>0.049</td>
<td>0.112</td>
<td>0.261</td>
</tr>
<tr>
<td>Zaz</td>
<td>-0.25</td>
<td>-0.032</td>
<td>-0.14</td>
<td>0.054</td>
<td>0.339</td>
<td>0.057</td>
<td>0.130</td>
<td>0.303</td>
</tr>
<tr>
<td>Sang</td>
<td>0.010</td>
<td>0.017</td>
<td>0.001</td>
<td>0.014</td>
<td>0.283</td>
<td>0.047</td>
<td>0.108</td>
<td>0.254</td>
</tr>
<tr>
<td>Yzd</td>
<td>0.066</td>
<td>0.014</td>
<td>0.038</td>
<td>0.185</td>
<td>0.291</td>
<td>0.049</td>
<td>0.111</td>
<td>0.261</td>
</tr>
<tr>
<td>Awr</td>
<td>0.497</td>
<td>0.082</td>
<td>0.157</td>
<td>0.471</td>
<td>0.311</td>
<td>0.052</td>
<td>0.119</td>
<td>0.278</td>
</tr>
</tbody>
</table>

Table 4.2: Distance between EXPECTED and OBSERVED outcomes by language, forms equivalent to plausible cognates in Persian

<table>
<thead>
<tr>
<th>Lang.</th>
<th>$\hat{\beta}_{LD}$</th>
<th>$\hat{\beta}_{RLD}$</th>
<th>$\hat{\beta}_{LLD}$</th>
<th>$\hat{\beta}_{QC}$</th>
<th>SE$_{LD}$</th>
<th>SE$_{RLD}$</th>
<th>SE$_{LLD}$</th>
<th>SE$_{QC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Pth</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NP</td>
<td>1.625</td>
<td>0.221</td>
<td>0.881</td>
<td>1.125</td>
<td>0.399</td>
<td>0.045</td>
<td>0.147</td>
<td>0.309</td>
</tr>
<tr>
<td>Bal</td>
<td>0.875</td>
<td>0.138</td>
<td>0.315</td>
<td>1.5</td>
<td>0.226</td>
<td>0.041</td>
<td>0.095</td>
<td>0.267</td>
</tr>
<tr>
<td>Kd</td>
<td>1.625</td>
<td>0.251</td>
<td>0.895</td>
<td>1.375</td>
<td>0.386</td>
<td>0.056</td>
<td>0.164</td>
<td>NA</td>
</tr>
<tr>
<td>Maz</td>
<td>0.25</td>
<td>0.02</td>
<td>0.063</td>
<td>0</td>
<td>0.25</td>
<td>0.02</td>
<td>0.063</td>
<td>NA</td>
</tr>
<tr>
<td>Zaz</td>
<td>1.5</td>
<td>0.308</td>
<td>0.824</td>
<td>1.25</td>
<td>0.677</td>
<td>0.126</td>
<td>0.308</td>
<td>0.568</td>
</tr>
<tr>
<td>Sang</td>
<td>0.5</td>
<td>0.083</td>
<td>0.1277</td>
<td>0.25</td>
<td>0.5</td>
<td>0.083</td>
<td>0.127</td>
<td>0.25</td>
</tr>
<tr>
<td>Yzd</td>
<td>0.666</td>
<td>0.361</td>
<td>0.963</td>
<td>1.666</td>
<td>0.235</td>
<td>0.065</td>
<td>0.095</td>
<td>0.471</td>
</tr>
<tr>
<td>Sng</td>
<td>0.333</td>
<td>0.083</td>
<td>0.135</td>
<td>0.333</td>
<td>0.333</td>
<td>0.083</td>
<td>0.135</td>
<td>0.333</td>
</tr>
<tr>
<td>Yzd</td>
<td>1.571</td>
<td>0.275</td>
<td>0.866</td>
<td>1.285</td>
<td>0.368</td>
<td>0.06</td>
<td>0.17</td>
<td>0.28</td>
</tr>
<tr>
<td>Awr</td>
<td>0.142</td>
<td>-0.028</td>
<td>-0.057</td>
<td>0</td>
<td>0.26</td>
<td>0.061</td>
<td>0.105</td>
<td>0.21</td>
</tr>
<tr>
<td>NP</td>
<td>1.5</td>
<td>0.285</td>
<td>0.895</td>
<td>1.5</td>
<td>0.27</td>
<td>0.076</td>
<td>0.109</td>
<td>0.27</td>
</tr>
<tr>
<td>Awr</td>
<td>0.25</td>
<td>0.035</td>
<td>0.101</td>
<td>0.25</td>
<td>0.25</td>
<td>0.035</td>
<td>0.101</td>
<td>0.25</td>
</tr>
</tbody>
</table>
The `dropterm()` function was again used to carry out the likelihood ratio test for each language. Balochi shows significantly more irregular change than New Persian for raw, unnormalized LD ($\chi^2(1) = 9.12, p < 0.01$), RLD ($\chi^2(1) = 7.5, p < 0.01$), LLD ($\chi^2(1) = 7.52, p < 0.01$), and QC ($\chi^2(1) = 13.7, p < 0.001$).

Thus far, results show that Balochi tends to undergo significantly more irregular sound change than New Persian in its functional vocabulary (except for in RLD measures of its functional vocabulary as a whole), particularly in items that are equivalent to their New Persian counterparts. This result could be tentatively interpreted as demonstrating that Balochi has borrowed its functional vocabulary that is identical to Persian from Persian, since it does not make much sense for Balochi to have undergone an overwhelming degree of irregular change to achieve the result of forms that are formally identical to their New Persian counterparts. However, it could also be the case that Balochi is less phonologically conservative than New Persian; this alternative explanation is investigated below. From a methodological perspective, Levenshtein distance-based metrics accord for the most part with qualitative judgments regarding amounts of irregular change that have taken place.

### 4.7 Phonological Change in West Iranian Lexical Items

The previous section showed that Balochi undergoes significantly more irregular phonological change in its functional vocabulary than New Persian. This section seeks to investigate whether this is a correlate of a wider tendency toward non-conservative historical phonology in Balochi. Why should one language undergo significantly more irregular change in its functional vocabulary than a closely related, neighboring language? New Persian has existed as a literary language for roughly a millennium, while Balochi was not written down before the 19th century (Elfenbein 1989:351), and is still not used as a literary language by many of its speakers. It has been held that literary languages are more stable over time in their lexicon and phonology than nonliterary languages (cf. Bergsland and Vogt 1962:128). If we can show New Persian to be more phonologically conservative (in non-functional items) than Balochi, the fact that Balochi undergoes more irregular phonological change (presumably irregular phonological reduction) in its functional vocabulary should be unsurprising.

To investigate this idea, I assembled a set of reliable (i.e., unlikely to be borrowed)¹⁰ cognate lexical items in the languages observed above. Levenshtein distances were measured between each lexical item and its Proto-Iranian etymon (here, only the LD, RLD and LLD measures were used). For example, the distance between PIr *raučah- and NP roz ‘day’ is 5, as is the distance between PIr *raučah- and Kurdish rož ‘id.’. Distances (found in the appendix) were integrated into a mixed-effects linear regression model for New Iranian languages (as before, Middle Iranian distances are given for show, but not part of any quantitative analysis), in which LANGUAGE was treated as a fixed effect, and ETYMA (corresponding to different lexical items) were assigned random intercepts. As before, New Persian was dummy coded as the intercept for the New Iranian model.

¹⁰This criterion was determined with the help of the Leipzig-Jakarta list, but also claims made in the various Iranological sources consulted in the course of this study.
Figure 4.2: Multidimensional scaling of inter-language distances
Table 4.3: Distance between Proto-Iranian etyma and lexical items in different languages

<table>
<thead>
<tr>
<th>Lang.</th>
<th>$\beta_{LD}$</th>
<th>$\beta_{RLD}$</th>
<th>$\beta_{LLD}$</th>
<th>$SE_{LD}$</th>
<th>$SE_{RLD}$</th>
<th>$SE_{LLD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>2.84</td>
<td>0.59</td>
<td>1.28</td>
<td>0.18</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Bal</td>
<td>-0.44</td>
<td>-0.09</td>
<td>-0.13</td>
<td>0.15</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Kd</td>
<td>0.28</td>
<td>0.03</td>
<td>0.09</td>
<td>0.3</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>Maz</td>
<td>0.24</td>
<td>0.03</td>
<td>0.07</td>
<td>0.29</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>Zaz</td>
<td>0.47</td>
<td>0.06</td>
<td>0.10</td>
<td>0.29</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>Sang</td>
<td>0.32</td>
<td>0.06</td>
<td>0.11</td>
<td>0.3</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>Yzd</td>
<td>0.62</td>
<td>0.12</td>
<td>0.16</td>
<td>0.3</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>Awr</td>
<td>-0.006</td>
<td>-0.006</td>
<td>0.01</td>
<td>0.31</td>
<td>0.06</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Coefficients and standard errors for each language as a fixed effect are given in Table 4.3, separated according to chronological period. Coefficients for non-Persian languages represent the difference between their estimated mean distances and the intercept.

The dropterm() function was used to carry out the likelihood ratio test for each language. For all three distance measures, there is significantly less distance between Proto-Iranian and Balochi than there was between Proto-Iranian and New Persian (LD: $\chi^2(1) = 8.41, p = .003$; RLD: $\chi^2(1) = 8.33, p = .003$; LLD: $\chi^2(1) = 9.43, p = .002$). Additionally, Yazdi is significantly further from Proto-Iranian than New Persian (LD: $\chi^2(1) = 4.46, p = .03$; RLD: $\chi^2(1) = 4.55, p = .03$; LLD: $\chi^2(1) = 4.04, p = .04$).

These results demonstrate that Balochi is significantly more conservative in the historical phonology of its non-functional lexical items than New Persian. Hence, the greater degree of irregular change shown by Balochi in its functional vocabulary is not a correlate of a larger tendency toward non-conservative historical phonology on the part of the language.

4.8 Discussion and Conclusion

The results of this study’s quantitative analysis show that Balochi appears to undergo significantly more irregular phonological change in its functional vocabulary than New Persian, while other West Iranian languages do not. Additionally, Balochi is significantly more conservative in its regular historical phonology than New Persian; it is thus unusual that it should undergo a greater degree of phonological reduction (if these two types of sound change are correlated). In the opposite direction, Zazaki and Yazdi are significantly less phonologically conservative than New Persian; these languages were shown to have undergone a smaller degree of irregular phonological change, though these differences were not significant.

In fact, it is precisely because Balochi is so phonologically conservative that a quantitative model detects a greater degree of irregular change in its functional vocabulary. For each functional item, the expected form for New Persian is closer to the observed form and further from Proto-Iranian, while the expected Balochi form is further from its observed form and
closer to Proto-Iranian.

If we interpret these results with respect to West Iranian sociolinguistics, they suggest that the Balochi numbers are skewed due to sociolinguistic pressure from Persian. It is likely that Balochi has borrowed a large number of its functional items from Persian during its history, and that the significantly higher degree of irregular change seen in Balochi is due to borrowing rather than a greater degree of irregular phonological reduction. This interpretation is particularly convincing given that Balochi functional items show a large degree of phonological convergence (rather than divergence) with New Persian ones. Though the quantitative data doesn’t make it explicit, it is still plausible — and highly likely — that other West Iranian languages have undergone large-scale borrowing of functional items from Persian, but the fact that their regular historical phonology is so similar to that of Persian makes it virtually impossible for the methodology employed in this chapter to detect significant differences in irregular change between them and New Persian that might point to borrowing.

Related languages can undergo different degrees of irregular reduction in cognate functional items, since amounts of reduction should be determined by language-specific parameters (e.g., usage, prosody, etc.). The above conclusion rests on the premise that if one language shows more irregular change than the other in a group of cognate functional items and these items are mostly identical across the two languages, such behavior is due to convergence between the two languages, if these two languages are in contact. However, if this type of behavior is attested between two related languages that have not been in contact since their genetic divergence (a sort of “pseudo-convergence”), it could potentially invalidate the above conclusion. This issue is worthy of future investigation.
Chapter 5

Geography and contact-induced variation in West Iranian

5.1 Introduction

The field of dialectometry, and offers a number of quantitative techniques for measuring linguistic distance between speech varieties. This literature has been characterized as “seek[ing] to quantify the degree of current similarity between dialects in a way that might, for example, be useful to language planners and educators,” rather than looking to establish genetic relatedness between them (Zuraw 2003:174). At the same time, dialectometry is highly informative from a diachronic standpoint, and can be used, among other things, to trace historical diffusions of linguistic innovations. In contrast to single-feature dialectology, dialectometry proceeds from a relatively large catalog of linguistic features, and computes aggregate distance between dialects on the basis of these. Linguistic measures found in dialectometry can involve the following:

- Measuring the aggregate Levenshtein edit distances between cognate vocabulary items across dialects (Kessler 1995; Nerbonne et al. 1996)
- Coding binary or n-ary feature matrices according to the presence or absence of a number of features
- Coding matrices according to the features of varying features in parallel corpora across dialects (Szmrecsanyi 2011)

Feature matrices of the type described in the last two bullet points can be used to compute inter-dialectal Euclidean or Hamming distance. The Hamming distance between two vectors $u$ and $v$ is the total number of all vector indices $i$ where $u_i \neq v_i$. These distance measures can be subjected to a number of additional analysis techniques: Multidimensional scaling can be used to observe speech varieties’ positions relative to each other in 1, 2, or 3-dimensional linguistic space. Principal components or exploratory factor analysis can be used to uncover latent elements that explain large portions of the variance.
Additionally, one can see how well linguistic distance measures correlate with non-linguistic distance measures. Non-linguistic distance measures used in the literature include:

- Geographic distance measures (Haynie 2012), which come in multiple types, including:
  - Great-circle or as-the-crow-flies distance: length of the shortest line between two points, taking into account the curvature of the earth
  - Least-cost distance: length of the shortest path between two points, taking into account obstacles and possibly surface friction of the landscape; this can be measured using cost rasters based on a various types of geospatial data, e.g., elevation, vegetation type, etc. (some authors have measured this using Google Maps’ walking directions, e.g., Szmrecsanyi 2012)
- Cophenetic distance (i.e., distance between nodes in a cluster) based on a priori dialect groupings

While linguistic distance measures of the type described above are in their most basic sense a metric encapsulating the synchronic distance between two speech varieties, they can additionally shed light on the historical diversification of speech varieties, as well as historical diffusion of linguistic features between them. For instance, clustering based on aggregate linguistic distance often accords with a priori dialect groupings established on the basis of single-feature studies and other traditional dialectological research. But more importantly, significant correlations between linguistic distance based on a set of features and geographic distance (with a substantial amount of the variance explained) can reflect patterns of geographic spread. Dialectometry that seeks to quantify overall similarity tends to aggregate as much data as possible. Alternatively, we can look at a number of varying items subsumed by one linguistic features (or a relatively small set) — so long as there is enough data to provide statistical power.

This chapter assesses the role of geography in the patterning of reflexes of Proto-Iranian sounds that vary irregularly across languages in terms of their historical phonology. For some Proto-Iranian sounds, linguistic distances correlate significantly with geographic distances, indicating either a one-off diffusion event or single trajectory of diffusion for words affected by a sound change. However, in the case of at least one sound change, the linguistic distances based on it are clearly bimodal, pointing to multiple developments and/or diffusion events, and hence do not correlate significantly with geographic distance.

5.2 Etymological variation in West Iranian

The previous chapters have established that, due to prolonged contact, Iranian languages are full of varying reflexes (i.e., doublets) of Proto-Iranian sounds. Different “prototypes” (a term borrowed from Rastorgueva and Édel’man 2003) appear in different Iranian languages. Take *ārāthni-ča- > *āranč > CSogd ’rync /ārinč/, Phl ‘lnc > NP āranj ‘elbow’: given the well-established outcome n of *θn in Sogdian and most East Iranian languages (vs. SWIr šn)
and the virtual formal identity between the Persian and Sogdian forms, we can be reasonably sure that the form was borrowed by Persian from a Sogdian-like dialect, if not Sogdian proper. The same goes for non-Southwest Iranian forms with \( d \) for PIr \( *\ddot{z} \) (other than in the word for ‘hand’).

“Lower-level” or shallower sound changes which exhibit variation can be hard to ascribe to a source, particularly if diffused changes (or changes found in borrowed lexical items) appear only during the New Iranian period. If a lexical item agrees across a number of languages in showing a (possibly sporadic or irregular) feature, e.g., an unexpected prothetic \( h- \) or \( x- \) in the word for ‘egg’ (PIr \( *\ddot{a}la-\)ka-), scholars generally accept that this agreement is due to lexical borrowing between languages, though the original source of the word is not always identifiable (for an interpretation of such forms, see Korn 2005:155–159).

Often, when it comes to such forms, we employ a “majority rules”-type of procedure whereby if a sound change is seemingly regular and well-attested in a given language, but shows up sporadically in a neighboring language, it is assumed to be a loan from the former to the latter. At the same time, we find musings in the literature as to whether these generalizations are always correct (recall the confusion of MacKenzie 1961:78 as to what the authentic Kurdish reflex of \( *r\ddot{z} \)).

This approach has yielded some valuable observations on the nature of contact between West Iranian languages, but can be somewhat atomistic. A desideratum in Iranian dialectology is a model that allows us to look at agreement of this sort across the lexicon, not just in individual lexical items.

In chapter 2, I listed a number of Proto-Iranian sounds whose reflexes exhibit a great deal of variation within and across West Iranian languages. Chronologically, the earliest variation concerns reflexes of PIr palatals (e.g., \( *s \) and \( *\ddot{z} \), as well as \( *\ddot{y}u \) and \( *\ddot{y} \), \( *\ddot{i} \), \( *\ddot{r} \), etc.). We are relatively certain that reflexes like \( h < \theta < *\ddot{y} \), \( s < *\ddot{y}u \) are Southwest Iranian in origin (though recall that Gershevitch 1962 believes that only \( s < *\ddot{y}u \) is a secure SWIr reflex). The provenance of chronologically shallower changes is less clear. For example, the clusters \( *r\ddot{z} \), \( *rt \) and \( *rd \) show sporadic change to \( l \) across West Iranian dialects. Lateralization is tentatively thought to be a Persian innovation, spreading to other dialects via loanwords. However, MacKenzie (1961:77) rightly states that “the outcome of the groups -rd- and -rz in the various non-Persian dialects is far from certain, words having been borrowed in every direction.” (In the case of Gorani \( zil \), there may have been diffusion of the change, or some sort of mixture or blend of SWIr and NWIr features due to contact.) Finally, there are low-level changes affecting Proto-Iranian segments such as \( *u- > b- \sim g- \sim v/w- \). In Persian, these changes have operated between MP and NP, and are thus relatively recent (i.e., taking place somewhere between ca. 500 and 1000 CE). Do these reflect an areal signal as well? Some authors seem to think so, or attribute variants to a linguistic source; recall Lentz’s (1926) apparent consideration of \( b- \) as the regular SWIr outcome of \( *u- \), as well as MacKenzie’s (1971a) idea that \( b- \) is a feature shared by Persian and Northern and Central Kurdish dialects.

West Iranian dialects are generally classified according to the features they have in common (e.g., Figure 5.1); in terms of historical phonology, they are usually classified according to their “expected” or “proper” outcomes (e.g., \( *s > h \) for NP, despite the fact that it often
In this chapter, I analyze aggregate data pertaining to the Proto-Iranian sounds listed above. Feature aggregation may show whether phonological patterns displayed by the items in question reflect historical contact.

I look first at the entire feature catalog, to investigate the role of geography in the overall signal displayed by West Iranian phonological variation. I then divide the catalog into three partitions, consisting of features of the following type, intended to capture variation in outcomes of Proto-Iranian sounds of interest:

1. PrI *laupāša- ‘fox’
   (a) h in reflex of *laupāša- (e.g., NP rubāḥ)
   (b) s in reflex of *laupāša- (e.g., S Tāti luās)

2. PrI *spržan- ‘spleen’
   (c) rz in reflex of *spržan- (e.g., NP supurz)
   (d) l in reflex of *spržan- (e.g., Kd sipīl)

3. PrI *yāta- ‘wind’
In a given language, etymological doublets are coded, even if a variant is clearly due to contact (e.g., Balocī has both zird and dil as words for ‘heart’, though the latter is clearly a Persian loan). Problems stem from coverage issues: some less-studied languages may contain some etymologically obscure (although see Rastorgueva and Ředel’mann 2003 for etyma beginning with a through k). For instance, we know enough about Persian historical phonology to identify the first member of NP dāl-man ‘eagle’ with Plr *źarnu- or *źar(i)ta- ‘gold, yellow’. A similarly opaque etymology might go unnoticed in a more obscure language.

Forms used and sources from which they were taken can be found in Appendix C, as well as the various distances calculated from them and linguistic metadata.

5.3 Contact vs. Parallelism

As mentioned in chapter 1, a large body of work historical linguistics, particularly in Austronesian historical linguistics, draws a dichotomy between cohesive linguistic groups that have been shaped by areal diffusion (“linkages”), and legitimate subgroups (Ross 1988; Garrett 2006). Concerns regarding deep genetic relationships versus deep areal relationships have played a role in a number of debates, including those about the validity of so-called “structural phylogeny” (Dunn et al. 2008; Donohue et al. 2008). The literature on glottometry attempts to quantify linguistic features’ “subgroupiness” across a continuum between linkages and subgroups proper (Kalyan and François forthcoming). The key distinction in the literature mentioned here tends to be between genetic inheritance and contact. The foregoing chapters have made it increasingly clear that it is uninteresting to debate whether West Iranian cohesiveness is due to genetic or non-genetic factors, given that there are only a small number of possible West Iranian genetic innovations (these not particularly diagnostic of subgrouping), and that several key features common to West Iranian postdate its earliest attested records.

The work presented thus far in this dissertation is in line with views that privilege the role of areality in linguistic diversification and disparification. However, we often lack a means of distinguishing between singular contact events and multiple parallel events — in this chapter, my objective is to work toward this latter goal.

5.4 Methods

The standard practice in dialectometry, following Séguy 1971, is to calculate pairwise distances between languages, and model these distances as a function of pairwise geographic distance. Most of Séguy’s linguistic distances were calculated on the presence or absence of a particular lexical item in a dialect. In a large scale dialectal survey, we care about a number of dependent variables (pertaining to the different features under study); unfortunately, linear regression
can take into account multiple independent factors, but only one dependent response variable. Incorporating these features into one distance measure allows us to measure the effect of predictors of interest on the entire feature catalog. A consequence is that these distance measures become reliant on other distance measures.

An additional consequence is this: once we have integrated linguistic features into a distance measure, it becomes difficult to take advantage of statistical tools that control for cluster-level idiosyncrasies (such as random effects). Certain subsets of the feature catalog may be skewed in a particular direction. Additionally, there may be cells missing data. Euclidean distances can be calculated between vectors with missing data in one of at least two ways:

\begin{align}
(5.1) & \quad 1. \text{ If a cell } n \text{ in a vector } I \text{ has no value, then } I_n \text{ and } J_n \text{ will be excluded when calculating the distance between vector } I \text{ and vector } J \\
& \quad 2. \text{ If a cell } n \text{ in a vector } I \text{ has no value, it will be replaced with } \sum_{i \neq n} I_i/\text{len}(I) - 1 \text{ before Euclidean distance is calculated}
\end{align}

Both of these techniques run the risk of underestimating the distance between two vectors.

The issue of non-normality within the data used to calculate distance measures can be dealt with in a number of ways. In theory, random effects can be assigned on the basis of various differential measures between languages (i.e., if we are comparing a literary language and a non-literary one, etc.).

Wieling et al. (2011) approach similar data from a different angle in a study of pronunciation distances between Dutch dialects. Rather than taking pairwise inter-dialectal distances as their dependent variable, they instead measure distances between a given dialect and standard Dutch. They represent geography with a tensor product of latitude and longitude, and model pronunciation distance as a function of the geographic smooth. This allows them to fit a number of random effects to the data. Wieling et al. (2014) use the same geographic smooth as a predictor of vocabulary choice between Standard Italian and Tuscan variants, a Bernoulli-distributed response. This methodology is advantageous in that the use of random effects is well suited to uneven data coverage. However, a nice feature of distance-based methodologies is that they demonstrate how pairs of languages vary together in particular features; this variation is essentially factored out of a mixed model via random intercepts and slopes.

Cronbach’s $\alpha$ is a measure of a dataset’s internal consistency. Many studies require a Cronbach’s $\alpha$-level of 0.7 or higher (Nerbonne 2009; Szmrecsanyi 2011). The dataset used in this chapter reached the 0.7 cutoff.

The distance measures used in this chapter are discussed below.

### 5.4.1 Linguistic distance measures

The linguistic feature catalog used in this study was based on lexeme-/etymon-specific phonological features. Phonological features are more (“diachronically”) abstract than low level (cf. Nerbonne 2009:179), in a manner according with the discussion of these features in the
Iranian literature. For instance, *šu has a number of reflexes in NW Iranian languages that go back to an earlier *sp, e.g., most NWIr sp, Maz sb, Yazdī sv; these are all coded as representing the reflex sp The catalog comprised 74 binary features pertaining to 31 lexeme-/etymon-based feature sets (exx. given above; Cronbach’s standard α = .91). Doublets are coded, even if a variant is clearly due to contact. Euclidean distances based on each language pair were calculated for each vector of features. Distance was calculated with R’s dist() function. There were some missing values, which were dealt with as per (5.1.2).

5.4.2 Geographic distance measures

Language locations consisted of coordinates of sites where languages were documented. In cases of languages with broader geographical coverage, centroids of the areas where they are or were spoken were used as coordinates. Geographical distances between locations were measured with the R package gdistance (van Etten 2012), using a raster of elevation data. Two types of distance were used:

- Great-circle or as-the-crow-flies distance: length of the shortest line between two points, taking into account the curvature of the earth
- Least-cost distance: length of the shortest path between two points, taking into account obstacles and possibly surface friction of the landscape

5.4.3 Models

Correlations between these distance types were measured via linear regression. For all models, geographic distance types (i.e., great-circle or least-cost) served as the independent variable. For each model, the dependent variable consisted of linguistic distance types based on

- Overall distance: the entire feature catalog
- Subsections of the feature catalog
  - Palatal distance: features pertaining to PIr etyma containing *š, *ši, *šu (N = 15)
  - Lateral distance: features pertaining to PIr etyma containing *r/rt, *r/rd, *r/rdz (N = 26)
  - Labiovelar distance: features pertaining to PIr etyma containing *u- (N = 34)

Correlations were measured for all West Iranian languages, and in some models, only Northwest Iranian languages.
5.5 Results

5.5.1 Results for all West Iranian languages

$R^2$ and $p$ values for each distance type are found in Table 5.1.

Table 5.1: $R^2$ and $p$ values for correlations between linguistic and geographic distances (asterisks denote significance codes) for all West Iranian languages

<table>
<thead>
<tr>
<th></th>
<th>Great-circle distance</th>
<th>Least-cost distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall linguistic distance</td>
<td>$R^2 = .16$</td>
<td>$p = .002^{**}$</td>
</tr>
<tr>
<td>Palatal distance</td>
<td>$R^2 = .10$</td>
<td>$p = .01^*$</td>
</tr>
<tr>
<td>Lateral distance</td>
<td>$R^2 = .29$</td>
<td>$p &lt; .001^{***}$</td>
</tr>
<tr>
<td>Labiovelar distance</td>
<td>$R^2 = .03$</td>
<td>$p = .19$</td>
</tr>
</tbody>
</table>

Stanford (2012:249) writes the following on correlations between geographic and linguistic distance in the literature: “[S]imple geographic distance typically accounts for 16% to 38% of variation...” That is to say, for the most part, $0.16 \leq R^2 \leq 0.38$. There is no particularly good consensus regarding the cutoff for meaningful values of $R^2$. How much variance needs to be explained by geography in order for us to see it as a predictor of dialect distance? As seen in Table 5.1, some linear models have $R^2$ values well below 0.16, but $p$-values below 0.05 (see *inter alia* Baayen 2008:114–6 on the relationship between statistical power and the amount of variance explained).

Some patterns of interest emerge from the data: Great-circle distance seems to correlate better with linguistic distance than least-cost distance based on elevation. This may be an issue tied up with the overall geographic scale covered by West Iranian languages (cf. Haynie 2012:55: “There is no evidence for overall patterns of environmental influences on contact throughout the entire Eastern Miwok territory, but within each of the traditional divisions of Eastern Miwok we find patterns within the linguistic network that reflect particular environmental influences”). Environmental features figure more strongly into the formation of smaller, shallower dialect groups than larger, deeper ones.

The chronological stratification of the partitions of the feature catalog makes it possible to assess some statements made in the dialectometry literature regarding time depth and the relationship between geography and dialect distance. We find claims that linguistic features actuated and diffused at a more recent date share a stronger relationship with geography than older features; cf. Goebl (2006:423): “We know that in the instance of the foundation of a linguistic island, as it relies basically on the factor of migration, older relations between language and space which are considered to be ‘natural’ are radically suspended and expire.”

Looking at partitions of our feature catalog, we see the following:

1. Lateral distance (i.e., distance based on reflexes of $*t/\ddot{t}$, $*r/\ddot{r}d$, $*r/\ddot{r}j$) correlates significantly ($p < .001$) with great-circle distance and least-cost distance
2. Palatal distance (i.e., distance based on reflexes of *š, *ši, *šu) correlates significantly with great-circle distance (\(p = .01\)) and insignificantly with least-cost distance.

3. Labiovelar distance (i.e., distance based on reflexes of *u-) correlates insignificantly with both great-circle distance and least-cost distance.

Do these results reflect the intuition that recently actuated and diffused features correlate more strongly with geographic distance than more older ones (keeping in mind the caveat that lexical items containing features that spread at an earlier date can be borrowed at any time)?

We can establish a rough chronology for change affecting the relevant groups according to our Persian attestations:

- The change of PIr *šu to \(s \sim sp\) predates Old Persian, e.g., PIr *ašua- > OP asa- \(\sim\) aspa- (\(<\) Median?) ‘horse’
- The change of PIr *r/r\!t, *r/r\!d, *r/r\!ź to \(l\) postdates Old Persian and predates Middle Persian, e.g., OP b(a)rd- ‘be high’ > MP bul(and) ‘high’
- The change of PIr *u- to \(b\)- postdates Middle Persian and predates New Persian, e.g., MP wad > NP bad ‘bad’
  - The dating of the change *u- > g- is less clear; PIr *u-i-nāša- > MP wināh > NP gunāh ‘sin’, but PIr *uṛka- > MP gurdag > NP gurda ‘kidney’, PIr *uṛka- > MP gurg > NP gurg ‘wolf’, PIr *u(i)i-āna- > MP gyro > NP jān ‘life, soul’

The graph in Figure 5.2 shows the correlation coefficient \(r\) plotted against the chronology of each change, from oldest to most recent. Figure 5.3 shows the density curves for distances based on each partition of the overall dataset. While noise is evident from the local minima seen in all plots, it is clear that labiovelar distances are lower in kurtosis (i.e., peak sharpness) than the other distance types; low kurtosis is a common indicator of bimodality (Darlington...
The plot of labiovelar distance as a function of greater-point distance (bottom plot, Figure 5.4) also shows bimodally distributed distances between Southwest Iranian and Northwest Iranian languages; this bimodal distribution is partially responsible for the insignificant correlation.

Figure 5.3: Density curves for each distance type

![Density curves for each distance type](image)

5.5.2 Results for Northwest Iranian languages

$R^2$ and $p$ values are found in Table 5.2. Great-circle distance now accounts for 24% of the variance for overall linguistic distance, and 28% of the variance for lateral distance. Crucially, when only Northwest Iranian languages are observed, there is a significant relationship between geographic distance and labiovelar distance. Least-cost distance has more explanatory power than great-circle distance, albeit only slightly more. This may indicate that differences in elevation may have played a greater role in the spread of chronologically shallow changes among closer-related languages, albeit only slightly.

The overall lack of a correlation between geography and labiovelar distance may also be due to more than one areally-diffused change affecting PIr *y*-.

<table>
<thead>
<tr>
<th></th>
<th>Great-circle distance</th>
<th>Least-cost distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall linguistic distance</td>
<td>$R^2 = .26$</td>
<td>$p = .001^{**}$</td>
</tr>
<tr>
<td>Palatal distance</td>
<td>$R^2 = .13$</td>
<td>$p = .02^{*}$</td>
</tr>
<tr>
<td>Lateral distance</td>
<td>$R^2 = .30$</td>
<td>$p &lt; .001^{***}$</td>
</tr>
<tr>
<td>Labiovelar distance</td>
<td>$R^2 = .13$</td>
<td>$p = .02^{*}$</td>
</tr>
</tbody>
</table>

The feature sets used to calculate palatal and lateral distance were bipartite, those used to calculate labiovelar distance were tripartite. Theoretically, this shouldn’t affect the Euclidean distance calculated,
Figure 5.4: Plots of palatal, lateral, and labiovelar distance as functions of great-circle distance (distance measures involving SW Iranian languages are in dark print; distance measures involving only NW Iranian languages are light print).

but if different binary features in the same feature set represent different linguistic phenomena, the distances calculated may contain some noise. To test this idea, the dataset from which labiovelar distance was initially calculated was split into two separate datasets, containing feature sets of the following type (PfR *uata* ‘wind’ serves as an example etymon):

4. Dataset 1 (“Velar distance”):
   (a) g- in reflex of *uata-
   (b) v- in reflex of *yata-
5. Dataset 2 ("Labial distance"):

(c) b- in reflex of *yāta-
(d) v- in reflex of *yāta-

Furthermore, since velar distance could potentially represent an interaction between two chronologically separate developments (cf. the early change of PIr *y- > g- / _r seen in MP gurdag, etc.), I excluded two feature sets pertaining to etyma beginning with the sequence *y_r-. New velar distance and labial distance measures were calculated, and modeled as functions of geography for all West Iranian and only Northwest Iranian languages; $R^2$ and $p$ values are found in Table 5.3. For all West Iranian languages, these new calculations show a significant relationship between velar distance and geographic distance (least-cost has higher $R^2$ value). For only Northwest languages, least-cost distance accounts for 31% of the variance in the relationship between least-cost and velar distance. No significant relationship between labial distance and geographic distance is seen.

### Table 5.3: $R^2$ and $p$ values for correlations between velar and labial distance and geographic distance for all Western Iranian and only Northwest Iranian languages

<table>
<thead>
<tr>
<th></th>
<th>All West Iranian</th>
<th>Great-circle distance</th>
<th>Least-cost distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar distance</td>
<td></td>
<td>$R^2 = .12$</td>
<td>$R^2 = .14$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p &lt; .01^{**}$</td>
<td>$p = .01^{**}$</td>
</tr>
<tr>
<td>Labial distance</td>
<td></td>
<td>$R^2 = -.01$</td>
<td>$R^2 = -.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = .69$</td>
<td>$p = .89$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Only Northwest Iranian</th>
<th>Great-circle distance</th>
<th>Least-cost distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar distance</td>
<td></td>
<td>$R^2 = .27$</td>
<td>$R^2 = .31$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = .001^{**}$</td>
<td>$p &lt; .001^{***}$</td>
</tr>
<tr>
<td>Labial distance</td>
<td></td>
<td>$R^2 = .02$</td>
<td>$R^2 = .01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = .16$</td>
<td>$p = .22$</td>
</tr>
</tbody>
</table>

5.6 Discussion

We see the following relationships between the previous section’s results and the chronologies of the actuation and diffusion of these sound changes:

- Changes affecting the Proto-Iranian palatals were actuated at an early date; hence there is a looser correlation between geography and distance (and an insignificant correlation with least-cost distance) based on varying reflexes of these sounds, given the amount of time that lexical items containing these reflexes have had to spread.

- Lateralization of *r + CORONAL clusters is more recent, yet still quite old (pre-MP, i.e., ca. 400 CE); whether languages agree in terms of lexical items affected by this change is strongly determined by geography.

- Change of the type *y- > g- is relatively recent (except before *-r-); topographical distances appear to play a highly significant role in the spread of words showing this change between Northwest Iranian languages.
• Change of the type \( ^*u- > b- \) cannot be dated with respect to \( ^*u- > g- \) change; it may reflect a recurrent tendency (cf. MacKenzie 1961:76), or a more “trivial” sound change.

The inclusion of distances between NP and other Iranian languages introduced a lot of noise which obscured the geolinguistic signal, given the fact that it is geographically distributed across Iran, and hence in close proximity to linguistically dissimilar languages (and its center of gravity has moved away from some linguistically similar languages). As shown in Figure 5.5, results did not show a linear relationship between the date of the actuation of a sound change and the correlation between geographic distance and linguistic distance based on said sound’s diffusion. However, as mentioned above, lexical items reflecting a given sound change can be borrowed at any point in time, and most of the irregular phonological change seen in Iranian is likely to be mediated by lexical borrowing. Results showed that the role played by topography is more visible when observing shallower changes in smaller linguistic groups (cf. Haynie 2012:55).
5.7 Conclusion

This chapter’s results show that for the most part, the shared phonological patterns seen across West Iranian reflect a significant geographic signal, and thus can be attributed to singular areal trajectories of diffusion, though the transfer of lexical items may have taken place over a prolonged period of time. However, in the case of at least one shallow, low-level change, that of \*u- > b-, words displaying this change appear to have originated in and diffused from two locations, or perhaps came about due to separate spheres of Persian influence on other West Iranian languages.

These results can be brought to bear on previous views of West Iranian dialectology, in particular the fate of \*u-. The change \*u- > b- was previously thought to represent a single, isolated areal development, but patterns of areal variation show at least two such developments. Unfortunately, these results bring us no closer to pinpointing the sources of each development, if they are in fact different (and not, as tentatively suggested above, the result of two contact/diffusion events between Persian and adjacent languages).

These results highlight the relevance of accounting for parallelism in the development of dialect groups and isoglosses. It will be valuable in future work to formulate possible accounts of how such parallelism has come about — to treat drift as explanandum, not explanans (cf. Lightfoot 2000:89). Unfortunately, the historical details of the change PIr \*u- > b across West Iranian are fairly limited.

Also of interest — and the object of valuable future inquiry — are non-Persian forms where a conservative reflex of PIr \*u- coexists with a sound change that has taken place during the history of the Persian language. For instance, Sivandi vare, Semani vará ‘lamb’ (presumably from < *yarna-ka-) show initial u-, but appear also to have undergone a sound change known to Persian, OP -rn- > MP -rr- (e.g., *hyarnah-/ *farnah- > MP xwarrah > NP farr ‘glory’; cf. also Shirazi vol ‘spider’, likely from *yarna-, Schwartz 1971:292, fn. 14). Similarly, varf ‘snow’ (PIr *yafra-), found in Sivandi, Qohrudi, Mazandarani, and Sangesari (where it exists in a doublet with vafr), shows the same conservatism in the initial vowel, but the same metathesis (-fr > -rf) that is found in NP barf ‘id.’ (< wafr). It is possible that these sound changes (-fr > -rf, final obstruent + sonorant metathesis) operated beyond the scope of Persian (the latter resolves a violation of the sonority hierarchy), via parallelism or areal diffusion. It is also possible that non-Persian languages borrowed words like NP barra and barf, and subsequently adapted them by “undoing” the \*u > b change (it is unlikely that they would have borrowed these words from Middle Persian, and furthermore, metathesis took place between Middle and New Persian).
Chapter 6

Conclusion

In writing this dissertation, I had two primary goals: first, I wished to conduct a comprehensive investigation of the status of West Iranian as a genetic subgroup; additionally, I sought to determine whether non-genetic innovations common to West Iranian languages were due to contact, parallel development, or both.

In Chapter 2, I demonstrated that as far as traditional comparative-historical reconstruction is concerned, there is scant evidence for an East-West divide. In Chapter 3, lexicostatistic evidence provided support for a West Iranian subgroup, and typological and other recurrent characters provided support for an East-West divide; at the moment, it seems safest to interpret these results as demonstrating an areal signal, a conclusion that other scholars have reached on the basis of different data (e.g., Sims-Williams 1996; Blažek 2013). In Chapter 4, I introduced a novel quantitative method of modeling irregular sound change designed to investigate whether West Iranian languages have developed formally similar functional items (i) due to shared tendencies expected among a group of closely related languages, or (ii) via language contact; results of this study indicate that contact has shaped the functional vocabulary of New West Iranian languages, at least to some degree. The dialectometric studies in Chapter 5 showed that while most irregular variation in reflexes of certain Proto-Iranian sounds is due to contact, a portion of it may be due to parallel tendencies.

My findings have broad implications for Iranian studies. They are compatible with the following scenario, partially fleshed out in the literature: innovations began to spread across Southwest Iranian, situated at a geographic extreme of the larger body of Common Iranian speakers, effectively forming a subgroup. Speakers of Southwest Iranian dialects came back into contact with the Medes and other speakers of “Northwest” Iranian. Lexical items were bidirectionally transferred between Southwest and Northwest Iranian languages, starting prior to the Achaemenid empire and continuing until the present day, leading to some of the variation described above, and leveling a large amount of the disparity in vocabulary across the incipient West Iranian dialect group. At the same time, a number of morphosyntactic differences persisted, visible across Northwest/Southwest lines. By the New Iranian period, most West Iranian languages had adopted formally near-identical functional morphemes, but still exhibited widely varying morphosyntactic properties (e.g., constituent order, etc.). As chapter 4’s results show, at least some of this convergence is due to borrowing that must
have taken place after the Middle Iranian period, and it seems likely that a great deal of it took place during the 16th–18th century Safavid dynasty, thought to be responsible for widespread linguistic Persianization in Iran (Borjian 2009).

These studies contribute to our understanding of language contact and dialect group formation. Elsewhere, it has been shown that in some multilingual, small-scale agricultural, socially non-stratified environments (e.g., Banks and Torres Strait Islands, Vanuatu), speakers seek to maximize the lexical distance between the languages that they speak, but often converge on nearly isomorphic grammatical structures; this situation is ascribed to ongoing contact and multilingualism between these languages. West Iranian shows different effects of contact: languages have leveled lexical distinctions on a large scale, but show pronounced differences (e.g., head-marking vs. dependent-marking, prepositions vs. postpositions, etc.) in morphosyntax. While New West Iranian languages have converged for the most part in terms of the functional morphemes they use, it is clear that only the forms, and not their usage, have been borrowed between languages. This, in contrast to the stable multilingual scenario, points to punctuated or chronologically delimited contact events, perhaps taking place alongside various changes in political power (Iran has been under rule by large empires for most of its historical record).

This dissertation offers new methodologies designed to tease apart linguistic similarities that are due to language contact as opposed to parallel tendencies. I have developed a means of modeling irregular sound change that is capable of detecting when a language has borrowed a functional item, particularly useful in cases where irregular phonological reduction has taken place. Additionally, I have used existing geolinguistic and dialectometric tools to demonstrate whether a feature shared by multiple languages has come about via a singular contact event versus a number of parallel developments, a distinction that has often been overlooked or blurred due to the fact it has generally proved difficult to draw. I hope that future use in linguistics will help to refine these tools, and that they will continue to increase our understanding of the mechanisms of language contact.
Appendix A: Chapter 3 Supplementary Materials

Lexical Characters

In the following Swadesh-200 word list, each word is given with two character states. The first was used in datasets where loans were coded; the second was used in datasets where loans went uncoded.

<table>
<thead>
<tr>
<th>Lexical Character</th>
<th>Loan State</th>
<th>Uncoded State</th>
</tr>
</thead>
<tbody>
<tr>
<td>I:</td>
<td>Pashto ZE (A), (A); Persian man (A), (A); Ormuri az (A), (A); Avestan azom, mâm (A), (A); Shughni wuz (A), (A); Bactrian 'z' (A), (A); Yaghnobi man (A), (A); Wakhi WUZ (A), (A); Old Persian mâm, adam (A), (A); Balochi MAN (A), (A); Kumzari meh (A), (A); Khotanese Saka aysu (A), (A); Middle Persian 'n' (A), (A); Waziri ZE (A), (A); Zebaki az (A), (A); Zazaki ez, min (obl.) (A), (A); Larestani ma (A), (A); Bashkardi mon (A), (A); Ishkashmi az (A), (A); Bandari me (A), (A); Khwarezmian n'z (A), (A); Parthian z,-m (A), (A); Kurdish ez, min (A), (A); Sogdian 'zw' (A), (A);</td>
<td></td>
</tr>
<tr>
<td>all:</td>
<td>Wakhi KUXT, CU, CUST (E), (E); Sogdian 'γ'c (C), (C); Balochi KULL, DRUH, SARO, THEGH, THEWAGH (F), (F); Khotanese Saka hama-,harbiśśa- (A), (A); Kurdish giş, hemû, tev (B), (B); Yaghnobi háma,yákay (H), (H); Old Persian visa (A), (A); Ossetic appat (J), (J); Ishkashmi dizgdak,gul (G), (G); Pashto TOL (I), (I); Parachi kull (D), (D); Larestani 'ama(š) (B), (B); Waziri HAMAGI, GHWUT (B), (B); Avestan vispe (A), (A); Zebaki juk,saf (H), (H); Shughni buqaθ (K), (K); Bactrian oispo (A), (A); Persian hame (B), (B); Zazaki heme,péro (B), (B);</td>
<td></td>
</tr>
<tr>
<td>and:</td>
<td>Zazaki ú,ki (B), (B); Persian VA (B), (B); Khotanese Saka va (B), (B); Larestani -o, va (B), (B); Sogdian tty (F), (F); Kumzari wa (B), (B); Waziri AU (B), (B); Wakhi ET, SE, WOZ (F), (F); Khwarezmian 'wd (F), (F); Avestan ča (A), (A); Pashto AU (B), (B); Old Persian -čä, utā (F), (F); Zebaki i,wo (B), (B); Shughni atā (F), (F); Ossetic aemae (C), (C); Ishkashmi za (D), (D); Parachi ū (B), (B); Balochi GUDA, DI, WA, O (B), (B); Yaghnobi -at,ham (E), (E);</td>
<td></td>
</tr>
<tr>
<td>animal:</td>
<td>Balochi JANWAR, ZANWAR, SHANWAR (H), (A); Zazaki heywān (C), (B); Sogdian stwpōyy (I), (I); Waziri DZANAWAR (E), (A); Ossetic caeraegoj, XAJUAN (G), (G); Persian janvar, HEYVAN (A), (A); Larestani junvar (D), (A); Pashto HAJVAN (F), (B); Kumzari haiwan (B), (B); Shughni aywūn (J), (B);</td>
<td></td>
</tr>
</tbody>
</table>
ashes: Wakhi PERG (B), (A); Pashto IRA (A), (A); Sogdian ḥš’kw (A), (A); Persian KHAKESTAR (D), (D); Khotanese Saka āḥāra- (A), (A); Avestan ātriiām (A), (A); Balochi PHUR (B), (B); Waziri TRA (A), (A); Kūrdish xuwi (H), (G); Yaghnobi xoxistār (G), (D); Bandari pūr (B), (B); Parachī bhāγ (C), (C); Zazaki wel (H), (G); Middle Persian dwryst (I), (F); Larestani ‘ašk (D), (E); Yaghnobi xoxistār (G), (D); Bandari pūr (B), (B); Parachī bhāγ (C), (C); Zazaki wel (H), (G); Middle Persian dwryst (I), (F); Larestani ‘ašk (D), (E);

at: Ossetic cur, (MAE - AEM) (G), (G); Waziri KSHE, PA, PERI (E), (E); Larestani ‘a tek-e (D), (D); Zazaki het(i),verā (C), (C); Wakhi DU, TE (A), (A); Balochi SAR-A, A (B), (B); Pashto DE...SERA, TA NEZDE (F), (F); Sogdian kw (H), (H);

back: Yaghnobi arka,Gurk,pušt (A), (A); Persian POSHT (A), (A); Wakhi DUM, URQA (B), (B); Balochi PHUSHT (A), (A); Bandari kūla (E), (E); Middle Persian pwSt (A), (A); Larestani kamār, gorda (D), (D); Sogdian prc(h) (A), (A); Parthian pwSt (A), (A); Zebaki dam,med (B), (B); Ishkashiši kāmuq (G), (G); Waziri SHAMZAI (F), (F); Ossetic parštā (A), (A); Kūrdish zek (D), (D); Middle Persian wd (D), (I); Parachī bad (I), (E); Ormuri γanj,xarāb (B), (B); Waziri KHEROP, BAD (F), (D);

bad: Avestan aḵō, aγō (A), (A); Pashto BAD (G), (E); Balochi HARAB, GANDAGH, GANDAGH (B), (D); Khotanese Saka dara- (K), (J); Persian BAD (D), (E); Larestani ne-xaš (E), (;); Wakhi SUK (C), (C); Zebaki šak (C), (G); Shughni ganda,šakki (B), (B); Yaghnobi ganda,bad,šum (B), (B); Old Persian gasta (B), (B); Ossetic aevzaer (H), (F); Kūnmari ban’jeh (D), (E); Ishkashiši kāmuq (G), (G); Sogdian byz- (J), (H); Middle Persian wd (D), (I); Parachī bad (I), (E); Ormuri γanj,xarāb (B), (B); Waziri KHEROP, BAD (F), (D);

bark: Waziri PATIKAI (D), (D); Larestani pušs-e derax (A), (A); Pashto POST (A), (A); Shughni čilik, pust, anqoq (F), (F); Kūrdish qal, qasaq (C), (C); Wakhi DERUXTE PIST (A), (A); Persian PUST (A), (A); Sogdian c’wnt (E), (E); Ossetic c’ar (E), (E); Yaghnobi pust,pustloq (A), (A); Balochi GAWAZ (B), (B);

because: Shughni čarö didi (B), (B); Pashto DZEKA CE, VALI CE (B), (B); Persian CHUN (B), (C); Balochi PHA HAW-AN SAUV, PHA HAW-AN KHAN, KI (C), (C); Zazaki weło-k (D), (D); Ossetic umaen aemae (E), (E); Wakhi CIZER (B), (B); Zebaki ke, tsiz-bā ke (B), (B); Avestan zī (A), (A);

belly: Balochi LAF (C), (C); Waziri GADOLYAI (G), (G); Avestan udaram (A), (A); Ishkashiši der (A), (A); Kūrdish zik (D), (D); Sogdian kōr’k (A), (A); Middle Persian pwrdg? (A), (A); Kūzmari shukum (F), (F); Zebaki der (A), (A); Yaghnobi dāra (A), (A); Wakhi WANJ, WERD, DOR (B), (B); Larestani ‘aškām (F), (F); Khotanese Saka aha,ūra,garba- (A), (A); Parachī nʃaf,ʃi (“I), (I); Shughni qič (J), (J); Ossetic guybyn (H), (H); Zazaki pize (E), (E); Persian SHEKAM (F), (F); Pashto GEDA, NAS, XETA (G), (G); Ormuri škamba (stomach) (F), (F);
| big       | Zebaki kata (H), (H); Shughni γylla, xidir (H), (H); Larestani gap (E), (E); Wakhi LUP (B), (B); Balochi MAZAAN, MAZ EN (A), (A); Middle Persian wzrg (D), (D); Yaghnobi kätta (H), (H); Zazaki gird, pil, xišn (C), (C); Kurdish mezin, gewre, gir (A), (A); Sogdian mzyx (A), (A); Bashkardi gozer (D), (D); Waziri STER (F), (F); Ossetic ystyr (F), (F); Ishkashmi katta (H), (H); Ormuri ustur (F), (F); Persian bozorg (D), (D); Bandari gap (E), (E); Kumzari gayp (E), (E); Parthian wzrg (D), (D); Avestan mas-, maza (A), (A); Pashto LOJ (G), (G); |
| bird      | Kurdish balinde, çivîk (D), (D); Waziri MARGHAI (C), (C); Zazaki mirlık, teyr (C), (C); Balochi MURGH (C), (C); Wakhi UNGUS (B), (B); Khwarezmian 'mγ (C), (C); Larestani paranda (F), (F); Zebaki (parinda) (G), (F); Persian MORGH, parande (C), (C); Kurdish tayr (E), (E); Middle Persian mwrw (C), (C); Ossetic marg' (C), (C); Yaghnobi paranda, jondor (H), (F); Avestan viš (A), (A); Khotanese Saka mura- (C), (C); Sogdian mγ' (C), (C); Pashto MURGE (C), (C); Parthian mwrw (C), (C); Shughni parandā (I), (I); Bandari morg (C), (C); |
| bite      | Yaghnobi kan, xišoy (H), (H); Persian DANDAN GEREFTAN (B), (B); Wakhi DENDUK DI-, GUP DI- (B); Sogdian zβ'-t (I), (I); Kumzari kha'adish (pst) (E), (E); Shughni pirënd (L), (L); Avestan dás- (A), (A); Parthian gšt (J), (J); Middle Persian gc (J), (J); Pashto CICEL (G), (G); Balochi WARAGH, WARTHAA (C), (C); Waziri CHICHEL (G), (G); Kurdish geztin (D), (D); Larestani kap kanda, gereta (F), (F); Ossetic xaecyn (H), (H); Khwarezmian bγ'heh' (K), (K); |
| black     | Yaghnobi šow, siyoh (A), (A); Wakhi SU, SIO (A), (A); Pashto TOR (D), (D); Waziri TOR (D), (D); Persian SIAH (A), (A); Ishkashmi šū (A), (A); Middle Persian šaw (Arm. lw) (A), (A); Sogdian šw (A), (A); Kurdish rēş (B), (B); Avestan siāuua-, sāmō (A), (A); Balochi SIYAH (A), (A); Parthian sy'w (A), (A); Larestani meški (C), (C); Shughni tēr, siyō (D), (D); Kumzari siyeh (A), (A); Zazaki siyā (A), (A); Parachi paddō (E), (E); |
| blood     | Persian khun (A), (A); Parachi hīn (A), (A); Kurdish xwīn (A), (A); Waziri WINA (A), (A); Ormuri šun (A), (A); Wakhhi XUN, WUSEN (A), (A); Kumzari khwāyym (A), (A); Parthian gwxn (A), (A); Avestan vohuni (A), (A); Zazaki günī (A), (A); Yaghnobi waxin, xun (A), (A); Ossetic tug (B), (B); Larestani xūn (A), (A); Sogdian (y)xwrn (A), (A); Ishkashmi wēn (A), (A); Pashto VINA (A), (A); Balochi HON (A), (A); Middle Persian xwn (A), (A); Khwarezmian hwny (A), (A); Shughni wixin, xun (A), (A); |
| blow      | Ossetic dymyn (G), (G); Yaghnobi puf kun (H), (H); Avestan vāiî (A), (A); Larestani bād 'onda (D), (D); Persian VAZIDAN (A), (A); Sogdian w's (A), (A); Wakhi KULUMUT, MEST (B), (B); Balochi KASHAGH, KHASHTA (C), (C); Waziri CHALEDEL (E), (E); Khotanese Saka dam- (G), (G); Pashto LEGEDEL (F), (F); Shughni puf čidow (H), (H); Parthian dm- (G), (G); |
bone: Khotanese Saka āstaa- (A), (A); Ishkashmi wastuk (A), (A); Middle Persian 'stg (A), (A); Kurdish hestî (A), (A); Parachi hadđ (B), (B); Ossetic staeg (A), (A); Kumzari khar (C), (C); Sogdian 'st-k- (A), (A); Ormuri stor'yan (A), (A); Pashto HED, HADUKAJ (B), (B); Avestan ast-, as (A), (A); Larestani 'ossoxun (A), (A); Zazaki este (A), (A); Wakhi YUSC, USTUXON (A), (A); Yaghnobi sitak (A), (A); Balochi HAD (B), (B); Waziri HADIKAI (B), (B); Persian ostokhan (A), (A);

breast: Yaghnobi vána,cic (F), (F); Zazaki sênê,virär (C), (C); Khwarezmian (y')ft'n (A), (A); Kurdish pêşîr, çiçik (B), (B); Persian pistān (B), (B); Zebaki bar (E), (E); Sogdian 'štnh (A), (A); Khotanese Saka tcıjsa (G), (G); Kumzari sınō (C), (C); Shughni bar, bat, sınä, biš (E), (E); Ishkashmi pešbar,cici (E), (E); Parachi sina (C), (C); Middle Persian pyst'n (B), (B); Avestan fštānō (A), (A); Larestani šir (D), (D);

breathe: Middle Persian dm- (H), (H); Balochi SAH ZIRAGH (B), (B); Avestan an- (A), (A); Pashto TANAFFUS KAVEL (F), (F); Waziri SAYA (BREATHE) (B), (B); Persian NAFAS KASHIDAN (D), (D); Avestan an- (A), (A); Pashto TANAFFUS KAVEL (F), (F); Waziri SAYA (BREATHE) (B), (B); Persian NAFAS KASHIDAN (D), (D);

burn: Avestan dažiiete (A), (A); Pashto SVADZEDEL, SVADZEL (D), (D); Zazaki veşäyiš (E), (E); Waziri BALEDEL, SWEL (C), (C); Larestani sota (D), (D); Khwarezmian 'br'z- (F), (F); Middle Persian h'w (E), (E); Yaghnobi suč- (D), (D); Khotanese Saka suv-,dajs- (A), (A); Persir SIKHTAN (D), (D); Ossetic sudzyn (D), (D); Parthian h'w,bruy (E), (E); Shughni suzaž čidow (D), (D); Kurdish sotn, şewtî (D), (D); Parachi thew- (B), (B); Waziri THAU- (B), (B); Balochi BALAGH (C), (C); Sogdian swc (D), (D); Persir SIKHTAN (D), (D); Ossetic suč- (D), (D); Khotanese Saka dam- (H), (H); Larestani nafas kerda (E), (E); Zazaki nefes (breath) (C), (C);

child: Persian BACHCHE (D), (D); Parthian fryznd (M), (J); Avestan aporēnlaiukō (A), (A); Zazaki qeček,gede,leyr,doman,qiž,tut (C), (C); Zebaki cuT (J), (C); Ossetic syvaellon, sabi (H), (H); Pashto TIFL (G), (G); Kumzari rōk (m.), ditk (f.) (E), (E); Larestani bec (D), (D); Shughni bača,kūdak,tīfīl (B), (B); Parachi bačī (I), (D); Yaghnobi pulla,bača,gudak,farzand (K), (D); Sogdian 'jwn (L), (J); Waziri KUDUK, ZA, ZUMAN (B), (B); Waziri WORKAI (F), (F); Middle Persian fryznd (M), (J); Balochi CHUKH (C), (C); Ishkashmi zāman (J), (I);

cold: Ishkashmi gulbāduk (H), (H); Larestani 'abr (A), (A); Parthian myg (G), (G); Shughni abri (A), (A); Pashto TORA URIADZ, XERA URIADZ (F), (F); Ossetic nūg (G), (G); Kurdsch erw (A), (A); Yaghnobi abr,tīra (A), (A); Waziri WERYEZ (E), (E); Avestan a'ró̂m, snaoðõ, maγem (A), (A); Kumzari nim (< Ar?) (D), (D); Yasgulyam varm (I), (I); Sogdian pr'yβ'k (A), (A); Ormuri yewar (A), (A); Parachi a'ır (A), (A); Waziri KUDUK, WETIS (B), (B); Persian ABR (A), (A); Khotanese Saka pryaura (A), (A); Balochi JH UR (C), (C);Persian ABR (A), (A);

cold: Pashto SOR (A), (A); Avestan saratō, aotō (A), (A); Balochi SARTH, GWAHAR (A), (A); Ossetic uazal (B), (B); Kurdish sar (A), (A); Parachi eštäwo (D), (D); Persian
SARD (A), (A); Kumzari sard (A), (A); Sogdian srt (A), (A); Wakhi SUR, SOZ (A), (A); Ishkashmi sard (A), (A); Larestani sard, xonok (A), (A); Ormuri câk (C), (C); Yaghnobi sort, xunuk (A), (A); Bashkardi sîrt (A), (A); Waziri SOR (A), (A); Shughni sard (A), (A); Zazaki sard (A), (A);

come: Balochi AGH, AKHTA, ATKA (A), (A); Khotanese Saka āta- (B), (B); Waziri ROTLEL (D), (D); Ossetic caeuyn (E), (E); Shughni yad (A), (A); Avestan ā jasaiti, ā aëiti (A), (A); Kumzari hâmëd (pst) (C), (C); Zazaki aëitiš (C), (C); Middle Persian 'md- (C), (C); Wakhi WEZI- (A), (A); Yaghnobi vvow- (G), (G); Bactrian aga- (A), (A); Sogdian ”γ- (A), (A); Zebaki is- (F), (F); Kurdish hatin (B), (B); Old Persian ā-jam-, aítij (A), (A); Parthian 'šm'- (C), (C); Persan AMADAN (C), (C); Ishkashmi āGad (A), (A); Pashto RATLEL (D), (D);

count: Khotanese Saka Sumăr (A), (A); Persian SHOMORDAN (A), (A); Shughni asöb (B), (B); Ossetic nymajyn (A), (A); Sogdian ptšmr (A), (A); Zazaki aëitiš (C), (C); Middle Persian 'šm'- (A), (A); Waziri GANREL (C), (C); Middle Persian ā-jam-', aítij (A), (A); Pashto ISOB TSER- (B), (B);

cut: Persian BORIDAN (E), (E); Avestan korountai (A), (A); Old Persian *karti- (A), (A); Balochi CHAKAGH, CHAKITHA (C), (C); Paruchi maëc (F), (F); Khotanese Saka źajš-. (H), (H); Ossetic kaerdyn (A), (A); Pashto PREKAVEL (A), (A); Sogdian 'nkr-nt (A), (A); Yaghnobi pakk- (G), (G); Middle Persian t's (H), (H); Shughni tèb (J), (J); Larestani boleda (E), (E); Wakhi RESED- (B), (B); Parthian q'f? (I), (I); Zazaki qesnayiš (D), (D); Waziri PREKREL (A), (A); Zebaki keT- (past ppl) (A), (A); Kumzari qaṣai'kin (imp sg) (), ()

day: Sogdian myθ (D), (D); Avestan aiiara, azan- (A), (A); Zazaki rož (B), (B); Bashkardi res (B), (B); Ossetic bon (C), (C); Balochi ROSH (B), (B); Yaghnobi met, ruz (D), (D); Ishkashmi roz (B), (B); Parthian rwc,rwž (B), (B); Paruchi dewâs, ruč (B), (B); Middle Persian rwc,rwz (B), (B); Ormuri rož (B), (B); Waziri VREZ, WREZ (B), (B); Pashto VRADZ (B), (B); Yazgulyam miθ (D), (D); Old Persian rauča (B), (B); Shughni meθ, rüz (D), (D); Wakhi ROR, REWOR (B), (B); Larestani rüz roz (B), (B); Zebaki mi (D), (D); Persian RUZ (B), (B); Bandari rüz (B), (B); Khwarezmian myθ (D), (D);

die: Middle Persian myr- (A), (A); Avestan miriête (A), (A); Wakhi MERI- (A), (A); Sogdian myr (A), (A); Zebaki murum (A), (A); Zazaki mériš (A), (A); Shughni Mar (A), (A); Yaghnobi mir-, marg vu- (A), (A); Old Persian mariyataiy (A), (A); Paruchi mer- (A), (A); Persian MORDAN (A), (A); Parthian myr- (A), (A); Kumzari mürd (pst) (A), (A); Pashto MREL (A), (A); Ishkashmi mul (A), (A); Khwarezmian 'my-' (A), (A); Khotanese Saka mar- (A), (A); Balochi MIRAGH, MURTHA (A), (A); Ossetic maëlyn (A), (A); Kurdish mirin (A), (A); Waziri MREL (A), (A);

dig: Zazaki kendiš (A), (A); Yaghnobi kan-, kow- (A), (A); Parthian kn- (A), (A); Middle Persian kn- (A), (A); Bashkardi kûç, kuhl (A), (A); Avestan kanaiti (A), (A); Kumzari
tikayna (pst) (A), (A); Khwarezmian kn- (A), (A); Sogdian qn- (A), (A); Ossetic k’axyn (D), (D); Ormuri waxay-ēk (D), (D); Old Persian kantiy (A), (A); Balochi JANAGH, PHATAGH, KATAGH (A), (A); Larestani kanda (A), (A); Wakhi PUS– (B), (B); Khotanese Saka ka’ggan- (A), (A); Shughni čán (A), (A); Pashto KINEL, KINDEL (A), (A); Waziri KANDEL (A), (A); Parachi kusèw (D), (D); Persian BILZADAN (C), (C);

dirty: Parthian rymn (B), (B); Bandari sehār (F), (F); Avestan āhitō (A), (A); Wakhi RIM (B), (B); Parachi kačal (J), (J); Balochi MELAR (C), (C); Shughni čal¯ın, čirkin, γažd, qarq (I), (I); Sogdian ”γwstk (L), (L); Larestani gana (E), (E); Ormuri čir (I), (I); Yaghnobi xira,loynók (K), (K); Waziri KHACHEN, KHIRAN (G), (G); Zazaki šimin (B), (B); Ossetic c’izi (H), (H); Pashto CATAL, XIREN (G), (G); Persian KASIF (D), (D);

dog: Zazaki kutik (B), (B); Kurdish kûçik, seg (A), (A); Ormuri kuçuk,spuk (A), (A); Khotanese Saka śve (< Ind?) (C), (C); Wakhi SUC (A), (A); Waziri SPAI (A), (A); Larestani sag (A), (A); Parachi espö,kuçuk (A), (A); Middle Persian sg (A), (A); Ossetic kuydz (B), (B); Avestan spā (A), (A); Pashto SPAJ (A), (A); Kumzari sōgh (A), (A); Balochi sag (A), (A); Yaghnobi kut (B), (B); Bashkardi sax (A), (A); Ishkashmi kud (B), (B); Persian sag (A), (A); Sogdian ’kwt (B), (B); Shughni kud (B), (B); Zebaki ked (B), (B);

drink: Pashto CSEL (E), (E); Kurdish vexwarin (A), (A); Bashkardi ’xwar (A), (A); Ossetic muazyn, CYMYN (F), (F); Wakhi PEV–, PU(W)- (B), (B); Kumzari khordish (A), (A); Balochi WARAGH, WARTHA (A), (A); Persian NUSHIDAN (D), (D); Zazaki šimitiš (C), (C); Larestani xarda (A), (A); Waziri TSHEL (E), (E); Avestan xvaraiti (A), (A); Ormuri xr-, xr-;

dry: Zazaki wišk (A), (A); Yaghnobi qoq, xušk (B), (B); Old Persian uška (A), (A); Middle Persian hwšk (A), (A); Pashto VUC (A), (A); Kurdish zuha (A), (A); Ossetic XUS (A), (A); Shughni qoq,xušk (B), (B); Wakhi WESK, XUK (A), (A); Ishkashmi kâk (B), (B); Waziri WUCH, SIR (A), (A); Parachi oškār, hušku (A), (A); Balochi HUSHK (A), (A); Parthian hwšk (A), (A); Avestan huškō (A), (A); Persian KHOSHK (A), (A); Sogdian škwyy (A), (A); Kumzari hishk (A), (A); Khotanese Saka huSka- (A), (A); Ormuri wôka (A), (A);

dull: Pashto PEC (B), (B); Yaghnobi kunt (A), (A); Persian KOND (A), (A); Ossetic k’uymyx (C), (C); Shughni gand, gund (A), (A);

dust: Shughni gard, γubòr,sit (M), (B); Wakhi GERD, XSUREM (B), (B); Parachi čärk (I), (H); Ossetic CYREN, ryg, TAERK, TAEVD (H), (G); Balochi DATO, DANZ (C), (C); Ishkashmi šit (< Wakhi?) (J), (I); Bashkardi dulā/āx (E), (E); Sogdian γwrm(h) (L), (J); Yaghnobi Girek, čang, xok (K), (D); Pashto GARZ (B), (B); Larestani gard-o- xāk (F), (D); Waziri KHAIRPAL (G), (F); Avestan paqsmuš (A), (A); Persian KHAK (D), (D);
ear: Sogdian γwš (A), (A); Kumzari gōsh (A), (A); Avestan gaošō (A), (A); Parthian ‘zgwlg (A), (A); Balochi GOSH (A), (A); Shughnī γu-γ (A), (A); Persian gush (A), (A); Yaghnobi Guš (A), (A); Old Persian gauša (A), (A); Bandari giš (A), (A); Khotanese Saka gi (A), (A); Khwarezmian γwx (A), (A); Zebaki Gāl (A), (A); Waziri GHOZH (A), (A); Zazaki goš (A), (A); Ishkashmi Gōl (A), (A); Ossetic x'us (A), (A); Middle Persian gwš, (A), (A); Wakhi YIS (A), (A);

earth: Kurdish xak, erd (E), (E); Parachī khen. (K), (J); Balochi MITTI (D), (D); Ormūrī xâk (J), (H); Wazīrī KHOVRA, WATAR (G), (G); Kumzari zamiyō (def) (A), (A); Yaghnobi Guš (A), (A); Old Persian būmiš (B), (B); Pashto XAK, MDZEKA (H), (H); Sogdian γwrn(h) (M), (I); Wakhī XOK, SET (C), (C); Persian ????? (A), (A); Avestan zā (A), (A); Zazaki her (E), (E); Ossetic zaexx, SYDZYT (I), (A); Khotanese Saka uysmā,ssandā (C), (C); Ishkashmi šit (< Wakhī?) (L), (C);

eat: Kurdish xwarin (A), (A); Pashto XVAREL (A), (A); Wakhī YA- (B), (B); Zazaki werdiš (A), (A); Persian KHORDAN (A), (A); Avestan xvaraiť (A), (A); Ormūrī x̱r, x̱r-; Wazīrī KHWAREL (A), (A); Kumzari khör (imp) (A), (A); Balochi WARAGH, WARTHA (A), (A); Bashkardi xwar (A), (A); Ossetic xaeryn (A), (A); Larestani xarda (A), (A);

egg: Larestani toxm tox (A), (A); Sogdian mrγyz’tk (C), (C); Kurdish hēk (B), (B); Ossetic ajk (B), (B); Pashto HAGEJ (B), (B); Wakhī TUXM MURGH (A), (A); Ormūrī supāl (D), (D); Kumzari khaig (B), (B); Persian tokhm (A), (A); Parachī ēx (B), (B); Khwarezmian y’k (B), (B); Balochi HAIKH, ANU (B), (B); Wazīrī YOWYA, YIYA (B), (B); Shughnī tarmurx (A), (A); Khotanese Saka aḥā (B), (B); Ishkashmi akik (B), (B); Zazaki hāk (B), (B); Yaghnobi taxm, tuxm (A), (A);

eye: Pashto STERGA (B), (B); Kumzari chōm (A), (A); Khwarezmian cm (A), (A); Shughnī čaṭm, cem, ūda (A), (A); Bandari čehem (A), (A); Khotanese Saka čeemman- (A), (A); Zazaki čim (A), (A); Middle Persian čšm (A), (A); Wakhī CEZM (A), (A); Ishkashmi čtsām (A), (A); Balochi CHHAM (A), (A); Persian chasm (A), (A); Zebaki čtsām (A), (A); Yaghnobi Gurda, wenna (D), (D); Ossetic ca⁵st (A), (A); Old Persian čašna (A), (A); Bashkardi čehm (A), (A); Larestani čaš (A), (A); Ormuri čim (A), (A); Wazīrī STERGA (B), (B); Kurdish čav (A), (A); Parthian čšm (A), (A); Avestan čama (A), (A); Parachī dida (C), (C); Sogdian čšm (A), (A);

fall: Zazaki kewtiš, gunayiš (C), (C); Yaghnobi dewi-, tiraš-, ruš- (H), (H); Zebaki ēdāwî (H), (H); Parthian q- (C), (C); Khotanese Saka kas- (C), (C); Persian OFTADAN (D), (D); Avestan pataiti (A), (A); Ormūrī γuẓ (G), (G); Kumzari keft (pst) (C), (C); Balochi KHAFAGH, KHPATA (C), (C); Sogdian w’pt (A), (A); Wakhī PERWE-, PULUN-, WUZ- (B), (B); Shughnī raz (I), (I); Pashto LVEDEL (F), (F); Wazīrī PREWATEL, WALWEDEL (E), (E); Larestani kata (C), (C); Middle Persian kf- (C), (C); Ossetic xauyn (C), (C);
far: Ishkashmi dīr-šluk (A), (A); Khotanese Saka dura- (A), (A); Persian DUR (A), (A); Larestani dīr (A), (A); Shughni dar, dūr (A), (A); Parthian dwr (A), (A); Ossetic dard (B), (B); Wakhi DIR (A), (A); Balochi DIR (A), (A); Avestan dūrāē (A), (A); Sogdian dwr (A), (A); Pashto LIRI (A), (A); Old Persian dūrai (A), (A); Middle Persian dwr (A), (A); Kumzari dūr (A), (A); Yaghnobi dur, olám (A), (A); Waziri LIRE, WURIYA (A), (A); Zebari dīr (A), (A); Parachi durin (A), (A); Zazaki dūrī (A), (A);

fat: Pashto GVAR (F), (F); Avestan pīuuō, āzuūtiš (A), (A); Balochi PHIGH (B), (B); Parachi ĉarbū (E), (E); Ossetic soj (G), (G); Yaghnobi ruGen (D), (D); Persian charbi (E), (E); Larestani carbi (E), (E); Khotanese Saka pāyä, tcarba- (E), (E); Kurdish qlewe (C), (C); Shughni ĉarvi, zόγ (E), (E); Waziri WURIYA (A), (A); Sogdian crp, rwγn (E), (E); Zazaki ruGen (D), (A), (A);

father: Waziri PLOR, BABĀ (A), (A); Zebari tāt, tā (B), (B); Yaghnobi dodo (B), (B); Middle Persian pyd (A), (A); Ossetic fyd (A), (A); Parachi bāw (C), (C); Persian PEDAR (A), (A); Parachi čarbū (E), (E); Ossetic soj (G), (G); Yaghnobi ruGen (D), (D); Persian charbi (E), (E); Larestani carbi (E), (E); Khotanese Saka pāyä, tcarba- (E), (E); Kurdish qlewe (C), (C); Shughni ĉarvi, zόγ (E), (E); Waziri WURIYA (A), (A); Sogdian crp, rwγn (E), (E); Zazaki ruGen (D), (A), (A);

fear: Shughni aks ċidōw (F), (F); Parachi būm (n.) (A), (A); Yaghnobi čukāyr (D), (D); Pashto BEREDEL (A), (A); Old Persian tarsatiy (B), (B); Waziri DAREDEL, WYEREDEL (B), (B); Avestan biθāia (A), (A); Ishkashmi trās (B), (B); Persian TARSIDAN (B), (B); Sogdian pc’ykwyr (E), (E); Wakhi WESI- (C), (C); Balochi THURSAGH (B), (B); Larestani terseda, zāla ceda (B), (B); Ormuri γuš-ōk (C), (C); Ossetic taersyn (B), (B); Kumzari turšidish (pst) (B), (B);

feather: Ormuri parr (A), (A); Waziri PAKHA (B), (B); Larestani fal (A), (A); Kumzari parr (A), (A); Sogdian prn (A), (A); Avestan parām (A), (A); Zazaki pūrī (A), (A); Shughni pār (A), (A); Wakhi PUR (A), (A); Balochi PHAR, KHAMB (A), (A); Ossetic sis (D), (D); Kurdish per (A), (A); Yaghnobi bol (A), (A); Persian par (A), (A); Pashto BENĀ, BANEKA (C), (C);

ew: Pashto LEZ (B), (B); Avestan kamna- (A), (A); Yaghnobī kam,barg, andak (A), (A); Shughni andāk (E), (E); Zebaki tāy(n), kemi, senik (A), (A); Zebaki tsamend (D), (D); Sogdian kōn- (A), (A); Wakhi KUM (A), (A); Balochi KHRARDE, KHAMB (A), (A); Ossetic cysyl (C), (C); Larestani kam (A), (A); Old Persian kamā (A), (A); Persian KAM (A), (A); Waziri LEZH, LEZHKI (B), (B);

fight: Pashto DZANGIDEEL (C), (C); Middle Persian ‘yrnz (F), (F); Yaghnobi bidón nos- (E), (E); Larestani jar kerda, daavā kerda, jang kerda, jangeda (C), (C); Khotanese Saka juv- (B), (B); Kumzari jung’kin (imp sg) (C), (C); Parthian rf- (G), (G); Ossetic tox kaenyn (D), (D); Persian JANGIDAN (DA’VA KARDAN) (C), (C); Balochi MIRAGH, MIRATHA (D), (D); Old Persian -jantiy (B), (B); Shughni qastin anjīvdōw (I), (I);
Avestan porotaite, yuñieiti (A), (A); Waziri JANG, JAGGARRA BALWA (C), (C); Sogdian ’x’s (H), (H); Wakhi JUNG TSER- (C), (C);

fingernail: Shughni nôxîn (B), (B); Khotanese Saka nâhun- (B), (B); Kurdish neynûk (B), (B); Ishkashmi ingituk (D), (D); Bandari penj (C), (C); Avestan srauu¯ o (pl.) (A), (A); Khwarezmian šwk (A), (A); Kumzari nikhin (B), (B); Sogdian n’γ’n (B), (B); Larestani naxon (B), (B); Ormuri taxt ta anguš.t (D), (D); Yaghnobi naxna (B), (B); Zazaki neng¯ u (B), (B); Persian N?XUN (B), (B);

fire: Khotanese Saka dai (C), (C); Kumzari hâtish (A), (A); Sogdian ’tr (A), (A); Parachi â? (A), (A); Shughni alôw, yoc (A), (A); Ishkashmi rošni (B), (B); Avestan atarš (A), (A); Persian ATASH (A), (A); Zebaki rošni (B), (B); Bashkardi yas (A), (A); Larestani taš (A), (A); Ossetic art, CAEXAER, ZYNG (A), (A); Waziri YOR (A), (A); Balochi AS (A), (A); Old Persian * açi- (A), (A); Parthian ¯dwr (A), (A); Zazaki ādır (A), (A); Yaghnobi olow, (A), (A); Zebaki yets (A), (A); Persian rošn¯ ı (B), (B); Bashkardi yas (A), (A); Larestani taš (A), (A); Ossetic CAEUYN, kaelyn, UAJYN (C), (C); Balochi BAHAGH, BAHITHA (D), (D); Zazaki rižiyaiš (A), (A); Pashto BAHEDEL, TOJEDEL (D), (D); Khotanese Saka ttajs- (G), (G); Avestan γžâraitii, raoʃaiti (A), (A); Persian JARI SHODAN (E), (E); Larestani ‘a low kata, leta (F), (F); Old Persian danu(va)- (B), (B);

fish: Shughni möyi (loan) (E), (A); Yaghnobi mahi (loan) (F), (A); Balochi MAHI (G), (A); Larestani mâš (A), (A); Wakhi KUP (B), (B); Avestan masiø (A), (A); Ossetic kaesag (C), (C); Khotanese Saka kava (B), (B); Parthian m’sy’g (A), (A); Kurdish masî (A), (A); Pashto KAB (B), (B); Parachi masø (A), (A); Zazaki mase (A), (A); Persian mahi (A), (A); Middle Persian m’hyyg (A), (A); Kumzari mi (A), (A); Bandari müyi (A), (A); Sogdian kp’ (B), (B); Ormuri (maøi) (D), (A);

five: Avestan pançã (A), (A); Ossetic fondz (A), (A); Khotanese Saka paMjsa- (A), (A); Zebaki pûnz (A), (A); Wakhi PANZ (A), (A); Yaghnobi panç,panj (A), (A); Sogdian pnc pnj (A), (A); Waziri PINZE (A), (A); Balochi PHANCH (A), (A); Kumzari panj (A), (A); Pashto PINDZE (A), (A); Persian PANJ (A), (A); Ishkashmi pûnz (A), (A); Parthian pnj (A), (A); Shughni panj,pinţ (A), (A); Parachi pûnz (A), (A); Larestani panj (A), (A); Yazgulyam pindz (A), (A); Middle Persian pnz,pnc (A), (A);

float: Waziri BAIYEDEL (C), (C); Larestani ‘a lû’ e ’aw boda (B), (B); Ossetic naudzu kaenyn (E), (E); Pashto GERZEDEL, CALEDEL (D), (D); Balochi LURAGH, LURITHA (A), (A);

flow: Wakhi CAU, REC- (C), (C); Waziri BAIYEDEL (D), (D); Khwarezmian γ’sw, rw (E), (E); Middle Persian phryz (F), (F); Sogdian rwš’t (E), (E); Ossetic CAEUYN, kaelyn, UAJYN (C), (C); Balochi BAHAGH, BAHITHA (D), (D); Zazaki riižiøyayiš (A), (A); Pashto BAHEDEL, TOJEDEL (D), (D); Khotanese Saka ttajs- (G), (G); Avestan γžâraitii, raoʃaiti (A), (A); Persian JARI SHODAN (E), (E); Larestani ‘a low kata, leta (F), (F); Old Persian danu(va)- (B), (B);

flower: Pashto GUL (H), (A); Ishkashmi gulok (L), (A); Ossetic didinaeg (I), (C); Balochi PHUL (B), (B); Waziri GUL (G), (A); Zazaki gul,čičege (C), (A); Shughni gul (O), (A); though gul < *wRda) (D), (A); Middle Persian ‘sprhm, (A), (D); Parachi gul (K), (A);
<table>
<thead>
<tr>
<th>Word</th>
<th>Language</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>give</td>
<td>Kurdish</td>
<td>da (A), (A); Shughni dedow (A), (A); Balochi DEAGH, DATHA (A), (A); Avestan daðāitî (A), (A); Pashto VERKAVEL (B), (B); Ossetic daettyn (A), (A); Persian DADAN (A), (A); Khwarezmian hîhr (E), (E); Zazaki dayış (A), (A); Parachi baxš kan- (C), (C); Bandari dâden (A), (A); Yaghnobi tîfâr- (D), (D); Ishkashmi da- (A), (A); Middle Persian dy- (A), (A); Waziri DERKREL (A), (A); Bactrian lh- (A), (A); Zebakî da- (A), (A); Old Persian dadâtiy (A), (A);</td>
<td></td>
</tr>
<tr>
<td>good</td>
<td>Sogdian</td>
<td>šr,xwp (C), (C); Bandari xub (B), (B); Yaghnobi obadon (J), (J); Old Persian *vahu-, naiba (A), (A); Balochi šarr (C), (C); Zazaki wêš,hol,ring (D), (D); Parachi kârî,nêk,xub (H), (H); Ossetic xorz, DZAEBAEX (G), (G); Khwarezmian xwb (B), (B); Shughni bašând (D), (D); Khotanese Saka šsära (C), (C); Wakhi XUB, BUF (B), (B); Zebaki ferî (I), (I); Larestani xuš (F), (F); Middle Persian xwb (B), (B); Avestan va?huš (A), (A); Kumzari khair (&lt; Ar) (E), (E); Persian KHUB (B), (B); Ormuri şîr (C), (C); Kurdish baş, qenc, ring (D), (D); Waziri SHE (C), (C);</td>
<td></td>
</tr>
<tr>
<td>grass</td>
<td>Parachî</td>
<td>alafî,gihâi,sauza (H), (C); Waziri WOSHE (A), (A); Persian ALAF (C), (C); Larestani 'alaf (F), (E); Shughni alaf, wôx (A), (A); Khotanese Saka ggjaisai (D), (D); Pashto VASE (A), (A); Kumzari giyâ (D), (D); Wakhi WUS (A), (A); Bashkardi gidâ(h), gida (&lt; Balochi) (E), (D); Ormuri γwâši (A), (A); Middle Persian gy'w,mrw (D), (D); Zazaki vâš (A), (A); Ishkashmi uüş (A), (A); Bactrian wî (A), (A); Balochi RENV, REM (B), (B); Khwarezmian wâs (A), (A); Yaghnobi wayš,marG (A), (A); Ossetic kaerdaeg (G), (G);</td>
<td></td>
</tr>
<tr>
<td>green</td>
<td>Waziri</td>
<td>SHIN, ZARGHIN (C), (B); Ossetic c'aex (I), (D); Kurdish hêşîn, kesk (C), (B); Bandari sùz (H), (A); Waziri ŜUVZ (A), (A); Persian SABZ (E), (A); Parthian hwzrgwn (M), (F); Zazaki kîho,?ewi (D), (C); Pashto SIN (C), (B); Yaghnobi Gûra (L), (E); Parachi sabz (J), (A); Middle Persian hwzrgwn (M), (F); Larestani sowz (G), (A); Shughni savž, xum (O), (A); Khotanese Saka gîva- (N), (G); Sogdian 'xs'yn (C), (B); Kumzari sauz (F), (A); Balochi SAVZ, MAUNSHAR (B), (A); Ishkashmi (sabz) (K), (A);</td>
<td></td>
</tr>
<tr>
<td>guts</td>
<td>Balochi</td>
<td>ROTH, RODH (B), (B); Pashto KULMA (C), (C); Persian RUDE (B), (B); Waziri KULMA, LARMÎN (C), (C); Ossetic t'ang (D), (D); Wakhi SINGER (A), (A); Larestani rûda (B), (B); Shughni darûn (C), (C); Parachi rûdâ (C), (C);</td>
<td></td>
</tr>
<tr>
<td>hair</td>
<td>Zebakî</td>
<td>seGund (H), (H); Bandari müd (D), (D); Balochi PHUT (C), (C); Shughni müy, tîr, γûnj (D), (D); Parachi dôš,jâl (F), (F); Zazaki mû,por,pînê (D), (D); Yaghnobi diraw,paşm (F), (F); Larestani mû (D), (D); Persian mû (D), (D); Pashto VESTE (A), (A); Ishkashmi Genuk (G), (G); Ossetic SAERYX&quot;UYN, x'ûyn (E), (E); Kurdish mû, por (D), (D); Sogdian wîs,zw (A), (A); Wakhi SUFS, PICA (B), (B); Waziri WESHTEH (A), (A); Kumzari mû (D), (D); Khotanese Saka drau-,ggûna- (F), (F); Avestan varasó (A), (A);</td>
<td></td>
</tr>
</tbody>
</table>
hand: Waziri LOS (A), (A); Bandari dast (A), (A); Wakhi DUST (A), (A); Pashto LAS (A), (A); Parthian dst (A), (A); Ossetic k’yx (B), (B); Sogdian δst (A), (A); Ishkashmi dust (A), (A); Middle Persian dst (A), (A); Larestani dass (A), (A); Khwarezmian (y’)ost (A), (A); Shughni dst, dust, xikulta (A), (A); Avestan zastō (A), (A); Old Persian dasta (A), (A); Ormuri dest (A), (A); Persian dast (A), (A); Zazaki dest (A), (A); Balochi DAST (A), (A); Zebaki dast (A), (A); Kurdish dest (A), (A); Khotanese Saka ggoştä (C), (C); Yaghnobi dast (A), (A);

he: Zazaki o (A), (A); Ishkashmi wa (A), (A); Zebaki ao (A), (A); Waziri AGHA (A), (A); Avestan hō (A), (A); Persian u (A), (A); Larestani ‘āna (A), (A); Pashto DE (B), (B); Shughni yam, yid, yu (A), (A); Wakhi YA (A), (A); Balochi ANH, CHI, I (A), (A); Sogdian ‘γw (A), (A); Ossetic uy, uyj (A), (A); Yaghnobi ax (A), (A); Old Persian hauv(am) (A), (A); Kumzari yeh (A), (A);

head: Waziri SER (A), (A); Kumzari sor (A), (A); Parthian sr (A), (A); Zazaki sere (A), (A); Yaghnobi kallá, sarkallá, sar (B), (B); Larestani sera (A), (A); Parachi sör (A), (A); Pashto SAR (A), (A); Kurdish sar (A), (A); Waziri SAR (A), (A); Persian sar (A), (A); Middle Persian sr (A), (A); Ishkashmi sur (A), (A); Avestan sarō (A), (A); Ormuri sar (A), (A); Sogdian sr (A), (A); Zebaki sör (A), (A); Balochi SAGHAR (A), (A); Ossetic saer (A), (A); Shughni kāl, kīl (B), (B);

hear: Persian SHENIDAN (B), (B); Yaghnobi duGuš (E), (E); Old Persian xšnautiy (B), (B); Wakhi KSUI- (B), (B); Kurdish bihiştin (C), (C); Middle Persian ngwš (E), (E); Pashto ARVEDEL (C), (C); Ishkashmi šud (pst) (A), (A); Avestan surumaoiti (A), (A); Parthian ngwš, zgwl (E), (E); Sogdian prγwš- (E), (E); Yaghnobi WORWEDEL, ARWEDEL (C), (C); Balochi AKSHANAGH, ASKHUTHA (B), (B); Larestani šonofta, goş kerda (B), (B); Bandari ešnitien (B), (B); Ossetic x’yçyn (D), (D); Zazaki āšnawitiš (B), (B); Shughni āx (A), (A);

heart: Ormuri zdī (A), (A); Kumzari dil (A), (A); Khotanese Saka ysaraka (A), (A); Pashto ZRE (A), (A); Wakhi PEZUV (B), (B); Ossetic zaerd (A), (A); Kurdish dil (G), (A); Parthian zyrd (A), (A); Larestani qalb, del (A), (A); Balochi DIL (F), (A); Middle Persian dyl (A), (A); Ishkashmi avzuk (D), (D); Zazaki qelb, zeri (A), (A); Sogdian (γrd’y), õrjyy (A), (A); Bandari del (A), (A); Bashkardi der (A), (A); Yaghnobi dil (E), (A); Khwarezmian ẓz (A), (A); Avestan zorad- (A), (A); Persian QALB (C), (C); Zebaki auzak (D), (D); Shughni zōrd, dil (A), (A); Waziri ZRE (A), (A);

heavy: Khotanese Saka ggarkha- (A), (A); Wakhi GHURUNG (A), (A); Yaghnobi wazmin (D), (D); Larestani sangi (B), (B); Waziri DRIND, SAKHT (C), (C); Sogdian β’rcyk (E), (E); Shughni wazmin, zīr (D), (D); Persian SANGIN (B), (B); Kumzari san’gī (B), (B); Balochi GIRAN (A), (A); Avestan gouru- (A), (A); Pashto DRUND (C), (C); Ossetic uaezzau (D), (D); Ormuri wazmin (D), (D);

here: Ossetic am () (D); Shughni ared () (F); Waziri DELE, DOLATA () (C); Zazaki tiyā, injā, nāzā () (A); Balochi EDHA () (A); Wakhi YEM, YAEI, DREM, HADREM
left: Sogdian prγšy (E), (E); Yaghnobi čapa (C), (C); Shughni čap (C), (C); Ormuri čap (C), (C); Waziri KINR (C), (C); Wakhi CUP (B), (B); Balochi CHAP (B), (B); Ossetic galau (D), (D); Zazaki čep (B), (B); Avestan haoiš (I), (I); Persian CHAP (B), (B); Pashto KIN (C), (C); Larestani kap (B), (B);

leg: Wakhi PUED, LENG (A), (A); Ossetic k'ax, ZAENG (C), (C); Pashto PSA (A), (A); Balochi PHADH (A), (A); Zazaki ling (B), (B); Shughni pō (A), (A); Waziri PSHA, LANGRA (A), (A); Avestan paitišan (A), (A); Yaghnobi poda, ling (A), (A); Parachi leng (B), (B); Sogdian pδ (A), (A); Ishkashmi tsiw-ling (B), (B); Persian pa (A), (A); Kumzari pā (A), (A);

lie: Khwarezmian 'nbzy (H), (H); Kurdish xwe dirêj kirin (C), (C); Waziri L Mostel (E), (E); Yaghnobi nepid- (H), (H); Pashto PREVATELAJ 3 SG. (F), (F); Persian Khabidan (Deraz Kashidan) (D), (D); Shughni aγas (I), (I); Ossetic xuycyn (G), (G); Avestan saēte (A), (A); Balochi WAFSAGH (B), (B);

live: Khotanese Saka jū- (A), (A); Larestani zendegi kerda (C), (C); Balochi ZINDAGH (Living) (B), (B); Waziri PADEEL (D), (D); Parthian jyw- (A), (A); Avestan juwaiit (A), (A); Old Persian jivaity (A), (A); Pashto ZVAND KAVEL (A), (A); Ossetic caeryn (E), (E); Yaghnobi žu- (A), (A); Sogdian zw (A), (A); Middle Persian jyw- (A), (A); Shughni guzarun čidow (F), (F); Persian ZENDEGI KARDAN (A), (A);

liver: Yaghnobi šupurda,jigar (I), (D); Ossetic iqaer (A), (A); Ishkashmi göla (G), (C); Larestani jegar (A), (A); Persian JEGAR (A), (A); Waziri YENNA (A), (A); Shughni jigar,ūod (H), (E); Avestan yakara (A), (A); Kurdish ceger (D), (A); Zazaki qesibā (E), (B); Balochi JAGHAR (C), (A); Wakhi JIGAR (B), (A); Khotanese Saka gyagarrů (A), (A); Pashto INA (A), (A); Parachi jigar (F), (A),

long: Old Persian darga (A), (A); Khotanese Saka dara- (A), (A); Persian DERAZ, boland (B), (B); Pashto UZD (B), (B); Avestan darvō (A), (A); Ormuri drāγ (A), (A); Ossetic darg’ (A), (A); Wakhi DEROZ, VERZ (A), (A); Waziri WIZHD (B), (B); Parthian drG (A), (A); Shughni daroz, vōγž (B), (B); Zazaki derog (A), (A); Sogdian brzzy (B), (B); Kurdish bilind (B), (B); Balochi DRAZH (A), (A); Yaghnobi van (C), (C); Ishkashmi wuZduk (B), (B); Middle Persian dgr (A), (A);

louse: Ishkashmi spul (A), (A); Zazaki espiž (A), (A); Shughni muraš, sipaγ (A), (A); Ossetic syst (C), (A); Balochi BOT (B), (B); Larestani eš (A), (A); Parachi ispō (A), (A); Pashto SPEZA, SPEZA (A), (A); Wakhi SIS (A), (A); Yaghnobi šipūš (A), (A); Avestan spiša- (A), (A); Persian shepesh (A), (A); Bandari soš/siš (A), (A); Sogdian špšh,ršk (A), (A); Waziri SPAZHA (A), (A); Ormuri spō (A), (A); Kurdish sip (A), (A);
man: Khotanese Saka daha- (E), (E); Zebaki nar (A), (A); Wakhi MERDINA, DAEI, XULG (B), (B); Parthian myrd, nr (A), (A); Kurdish mèr (B), (B); Kumzari murtk (B), (B); Waziri SARAI, NAR, NER (C), (C); Pashto SARAJ (C), (C); Yaghnobi morti, odam (B), (B); Shughni çör, nafar (F), (F); Khwarezmian mrc (B), (B); Ossetic laeg, naelgojmag (D), (D); Sogdian nyrk (A), (A); Larestani mard (B), (B); Persian mard, ADAM (B), (B); Middle Persian myrd, nr (A), (A); Avestan vîrû, ná (A), (A); Ishkashmi nark (A), (A); Balochi MARD, MAR (B), (B); Bashkardi mòş [ < martyia] (B), (B); Parachi mè?, nar (A), (A); Old Persian *nar- (A), (A);

many: Shughni bisyôr? (M), (M);Ormuri zut (D), (D); Parachi γαλαβα (I), (I); Kumzari khaykeh (E), (E); Larestani xâli, ziyád, basi (F), (F); Balochi BA Z (C), (C); Old Persian parava (A), (A); Avestan parauû (A), (A); Yaghnobi morti (K), (K); Kurdish gelek, pirr (A), (A); Waziri DER (G), (G); Ossetic birae (H), (H); Wakhi GHUFC (B), (B); Sogdian ’tèr (L), (L); Middle Persian *wsy’r (M), (M); Pashto DER (G), (G); Persian CHANDIN, ziad (D), (D); Zebaki fai (J), (J);

meat: Ossetic fyd (B), (B); Khotanese Saka ggûśta (A), (A); Zazaki gošt (A), (A); Middle Persian pyt (B), (B); Yaghnobi yota, gušt (A), (A); Ishkashmi púdf (B), (B); Parachi γUš (A), (A); Parthian pyt (B), (B); Wakhi GOST (A), (A); Kumzari gösht (A), (A); Waziri GHOSHA, GHESHA (A), (A); Shughni guşt (A), (A); Balochi GOZHD (A), (A); Persian guşt (A), (A); Sogdian ’pt (B), (B); Pashto GVASA (A), (A); Ormuri gàka (C), (C);

moon: Kurdish meh, heyv (A), (A); Old Persian *mähâ- (A), (A); Parachi mahôk (A), (A); Ishkashmi mâ (A), (A); Yazgulyam mäst (A), (A); Middle Persian m’h (A), (A); Zazaki âšmi (B), (B); Kumzari maytâwo (A), (A); Zebaki ilmek (C), (C); Yaghnobi mahtop (A), (A); Parthian m’h (A), (A); Ormuri mätaw (A), (A); Shughni mèst (A), (A); Sogdian ’x (A), (A); Avestan mà (A), (A); Khotanese Saka mástâ (A), (A);

mother: Balochi MATH (A), (A); Waziri MOR, MER (A), (A); Kumzari mâm (A), (A); Pashto MOR (A), (A); Zazaki mâ(y), dàd (A), (A); Parachi a’í, mâcî (A), (A); Khotanese Saka nàni (B), (B); Persian madar (A), (A); Avestan màta (A), (A); Shughni nân (B), (B); Ossetic mad (A), (A); Larestani nana (B), (B); Ishkashmi nân (B), (B); Sogdian m’t (A), (A); Bandari mûm (A), (A); Wakhi NUN (B), (B); Yaghnobi oča (C), (C); Old Persian *màta (A), (A);

mountain: Old Persian kaufa (B), (B); Wakhi KU(H) (F), (B); Avestan gairiš (A), (A); Kumzari kô (B), (B); Waziri GHA (B), (A); Zazaki ko (B), (B); Shughni ku, pûxtâ, tax (E), (E); Persian KUH (B), (B); Parthian kwf (B), (B); Middle Persian kwf (B), (B); Yaghnobi Gar (A), (A); Ossetic xox (B), (B); Kurdish sax, çiya (D), (D); Ormuri giri, kô (A), (A); Sogdian γr- (A), (A); Khotanese Saka ggar- (A), (A); Larestani kû (B), (B); Pashto GAR (A), (A); Balochi PHAWAD (C), (C);

mouth: Kurdish dev (C), (C); Pashto XLA (G), (G); Balochi DAF (C), (C); Zazaki fek (D), (D); Avestan â?h- (A), (A); Khotanese Saka âhâ- (A), (A); Kumzari kâr (E), (E);
Yaghnobi rax (J), (J); Shughni γêv (L), (L); Ossetic KOM, dzyx (H), (H); Persian DAHAN (C), (C); Larestani kap, da’an (F), (F); Wakhi GHUS (B), (B); Ormuri (dân), pöṣ (I), (I); Khwarezmian k’m (H), (H); Sogdian kw’ (K), (H); Yazgulyam fôts (I), (I); Ishkashmi futs (I), (I); Waziri KHWULA (G), (G); Middle Persian dhyn (C), (C);

name: Balochi NAM (A), (A); Bandari nüm (A), (A); Pashto NUM (A), (A); Zebaki nêm (A), (A); Parthian n’m (A), (A); Ossetic nom (A), (A); Yaghnobi nom (A), (A); Persian ESM (B), (B); Khotanese Saka nâma- (A), (A); Middle Persian n’m (A), (A); Avestan n¯ama (A), (A); Saḵa nım (A), (A); Middle Persian n’m (A), (A); Old Persian nâm¯a (A), (A); Khazarian nım (A), (A);

narrow: Persian BARIK (B), (B); Yaghnobi tang, borik (A), (A); Ormuri tang (A), (A); Waziri TANG (A), (A); Shughni birik, bôrik, tâng (A), (A); Larestani tang (A), (A); Zazaki teng (A), (A); Balochi TANKH (A), (A); Ossetic uyngaeg (C), (C); Pashto TANG (A), (A); Waziri TANG, BIRIK (A), (A);

near: Ossetic XAESTAEG, cyr (C), (C); Yaghnobi qaríb (E), (E); Ishkashmi dzêa (D), (D); Waziri NEZDE, TSARMA (B), (B); Waziri NEZDE (B), (B); Kumzari nayzik (B), (B); Sogdian β’w (F), (F); Avestan asne (A), (A); Old Persian aşnay (A), (A); Balochi NAZI, NAZIKH (B), (B); Zebaki jâ (D), (D); Persian NAZDIK (B), (B); Larestani nazik (B), (B); Bandari nazik (B), (B); Zazaki nezdê (B), (B); Shughni qarib, nazdik (G), (B);

neck: Waziri GERDON, MAEYUK (B), (B); Kurdish gerden (B), (B); Yaghnobi kama (A), (A); Persian GARDAN (B), (B); Zazaki mil, vil (A), (A); Avestan mana, griuua (A), (A); Balochi GVAR (B), (B); Larestani gardû (B), (B); Kumzari gurdin (B), (B); Shughni gardan, mak (B), (B); Sogdian γ’d’k (B), (B); Waziri GHWORA, MA KANDAI, MAGHZAI (B), (B); Pashto GARA (B), (B);

new: Sogdian nwyy (A), (A); Kurdish nú (A), (A); Zazaki newe (A), (A); Waziki TOZA, SEGHD (B), (B); Ossetic naeuag (A), (A); Persian TAZE (B), (B); Parachi nú (A), (A); Kumzari nô (A), (A); Ishkashmi nawuk (A), (A); Waziri NEWAI (A), (A); Larestani now, řâza (A), (A); Shughni naw, tirâk (A), (A); Pashto NEVAJ (A), (A); Khotanese Saka nûvara (A), (A); Balochi NOKH (A), (A); Ormuri nûw (A), (A); Yaghnobi nava (A), (A); Avestan nauñô (A), (A);

night: Khotanese Saka SSava (A), (A); Waziki NUGH (B), (B); Waziri SHPA (A), (A); Middle Persian šb (A), (A); Kurdish šev (A), (A); Old Persian xšap- (A), (A); Zazaki šew, pesewe (A), (A); Balochi SHAF (A), (A); Ishkashmi šab (A), (A); Sogdian γšp (A), (A); Shughni šâb (A), (A); Avestan xšap- (A), (A); Persian SHAB (A), (A); Yaghnobi xšap, šab (A), (A); Yazgulyam Šab (A), (A); Parthian šb (A), (A); Khwarezmian ‘xîyb (A), (A); Kumzari sho (A), (A); Ossetic aexcaev (A), (A); Bandari šû (A), (A); Zebaki feršun (C), (C); Larestani šow (A), (A); Pashto SPA (A), (A);
nose: Bandari damāγ (G), (G); Avestan nā?ha (A), (A); Kumzari nökhet (A), (A); Wakhi MIS (B), (B); Shughni nēţ (A), (A); Sogdian ns (A), (A); Yaghnobi nays,dimoG (A), (A); Larestani domāg (F), (F); Ishkashmi nits (A), (A); Kurdish poz, lût (C), (C); Old Persian nāh(a)- (A), (A); Pashto PAZA (H), (H); Ossetic fyndz (H), (H); Parthian wynyg (J), (J); Balochi PHONZ (C), (C); Persian DAMAGH (E), (E); Zebaki nīts (A), (A); Wakhi PEZA, WARBIZ, WARSAK (H), (H); Parachi damāγ,nešt (I), (I); Zazaki pîrnike (D), (D);

not: Parachi na (A), (A); Persian NA (A), (A); Middle Persian ny (A), (A); Wakhi NE (A), (A); Khwarezmian n (A), (A); Old Persian naiy (A), (A); Sogdian L’ (A), (A); Zebaki na, nas (A), (A); Khotanese Saka ni (A), (A); Pashto NA (A), (A); Ossetic nae, naetae (A), (A); Kurdish na, ne (A), (A); Balochi NA, N (A), (A); Shughni na (A), (A); Waziri NA (A), (A); Parthian ny (A), (A); Kumzari nā (A), (A); Avestan nāit (A), (A); Ishkashmi na, nus (A), (A);

old: Ossetic zaerond (C), (C); Balochi PHIR (B), (B); Yaghnobi kuhna,qadim,pir (D), (D); Pashto ZOR (C), (C); Sogdian ’wny (E), (E); Khotanese Saka ysaMgara- (C), (C); Waziri ZOR (C), (C); Avestan hanō (A), (A); Shughni kina,pir (D), (D); Persian PIR (B), (B); Larestani pir (B), (B); Wakhi KONA, XAEYAR () , () ; Zazaki pîl,îxtiyâr (B), (B); Ormuri zâl (C), (C);

one: Larestani yak (A), (A); Balochi YAK, YA (A), (A); Sogdian ’yw (A), (A); Parachi žū (A), (A); Kumzari yek (A), (A); Avestan æeuū (A), (A); Wakhi YI, I, YIU (A), (A); Zebaki wok (A), (A); Ishkashmi wak (A), (A); Old Persian aïva (A), (A); Persian yek, ???? (A), (A); Waziri YO (A), (A); Kurdish yek, ēk (A), (A); Ossetic iu (A), (A); Yaghnobi yak (A), (A); Pashto JAV (A), (A); Khotanese Saka ššau- (B), (B); Yazgulyam wōG (A), (A); Shughni yak,yîw (A), (A); Bandari ya (A), (A); Ormuri sê (A), (A); Bactrian ywg (A), (A);

other: Avestan ainiiö (A), (A); Zazaki bûn (D), (D); Khwarezmian ’ny (A), (A); Yaghnobi âné,axiś (A), (A); Middle Persian ny (A), (A); Ishkashmi an (A), (A); Pashto BEL (F), (F); Wakhi DIGAR (B), (B); Ossetic aendaer, innae (A), (A); Parthian ny (A), (A); Old Persian aniya (A), (A); Sogdian ’ny’ (A), (A); Waziri BEL, NOR (F), (F); Shughni digā(r), ga, yiga (G), (G); Khotanese Saka aña- (A), (A); Balochi DOHMI, DUHMI (C), (C); Persian digar (E), (E);

person: Waziri KAS, TAN (E), (E); Zazaki însâń,merdim (A), (A); Persian SHAKHS (C), (C); Shughni čör, nafar (G), (G); Sogdian ’γryw(h) (F), (F); Old Persian martiya (A), (A); Larestani ’adam (D), (D); Balochi KHAS (B), (B); Pashto KAS, TAN (E), (E); Avestan mašiiákō, mašiio (A), (A); Kurdish meriv, mirov (A), (A);

play: Persian BAZI KARDAN (C), (C); Yaghnobi bozi (noun) (E), (E); Shughni bözi ćidîw (G), (G); Kumzari baz gu’dish (pst) (C), (C); Zazaki kây (noun) (B), (B); Balochi LEV KHANAGH (A), (A); Middle Persian w’c (C), (C); Sogdian k’t’k (F), (F); Pashto
pull: BAZI KAVEL (E), (E); Waziri MAZSHILEDEL (D), (D); Bandari gazi (noun) (C), (C); Ossetic x’azyn (B), (B);  

push: Ossetic ycxojyn (F), (F); Waziri PORI WAHEL (E), (E); Persian HOL DADAN (C), (C); Balochi TELAN (SB.) (B), (B); Larestani loodada (D), (D); Shughni barγa Dedow (G), (G); Pashto PORI VAHEL (E), (E);  

rain: Sogdian w’r (A), (A); Pashto BARAN UREZI (A), (A); Ishkashmi urnaduk (B), (B); Parthian uryn, K"AEVDA (A), (A); Zazaki variy, yaγer (A), (A); Yaghnobi borîn (A), (A); Kumzari baram (noun) (A), (A); Larestani baru baredan (A), (A); Shughni börûn (A), (A); Kurdish baran (A), (A); Waziri WAREDEL (A), (A); Parachi avγar,wâš- (A), (A); Persian BARIDAN (A), (A); Middle Persian w’r (A), (A);  

red: Zazaki sur (B), (B); Avestan raiošî tô (A), (A); Shughni lôlâ,rûšt (H), (H); Waziri SIR (B), (B); Khotanese Saka ysrû̀na- (G), (G); Larestani germez (C), (C); Kumzari sirkh (B), (B); Pashto SUR (B), (B); Balochi SUHR (B), (B); Ossetic cyrx (B), (B); Kurdish sor (B), (B); Yaghnobi kimir, surx (F), (F); Sogdian km’yr (F), (F); Wakhi SEKR (B), (B); Parachi surkhô,surku (D), (B); Ishkashmi surx (E), (B); Persian SORKH (B), (B);  

right: Zazaki doγri,râšt (A), (A); Old Persian rasta (A), (A); Balochi RAST (A), (A); Larestani râss (A), (A); Wakhi WERTS, DURUST, BUF, BEROBER (B), (B); Sogdian dšn (F), (F); Old Persian raOTA (A), (A); Ossetic rast (D), (A); Khotanese Saka rollable (A), (A); Persian DOROST (B), (B); Waziri SAHI (C), (C); Shughni dirust, mâqûl (G), (G); Pashto SAM (C), (C);  

rightside: Middle Persian dšn (F), (F); Shughni rôst, xez (H), (H); Avestan dâšnô (A), (A); Khotanese Saka Ivarandau (G), (G); Parthian dšn (F), (F); Sogdian xw’rûnt (E), (E); Balochi RAST (C), (C); Waziri SHAI (D), (D); Zazaki râšt (C), (C); Ossetic RAXIZ, rast (C), (C); Wakhi WURZGE (B), (B); Pashto SAJ (D), (D); Persian RAST (C), (C);  

criver: Parthian rwd (A), (A); Old Persian rauta (A), (A); Khotanese Saka ŋa,nâtâ (H), (H); Zazaki ro(y),la,čem (A), (A); Shughni daryô (I), (I); Persian DARYA (D), (B); Wakhi DERIO (B), (B); Sogdian rwt (A), (A); Old Persian rauta (A), (A); Ormuri wôkxânâ (G), (G); Waziri DARYOB, TOI (E), (B); Avestan rauu, ōraotô (A), (A); Yaghnobi rawt,op,dariyo,nahr (A), (A); Ossetic DON, caugaedon (F), (F); Larestani rudxuna, big (a large one) (A), (A); Pashto RUD, SIND (A), (A); Middle Persian rwd (A), (A); Bandari rûxâna (A), (A); Balochi DIRA (C), (B);
road: Parthian r’h (C), (C); Middle Persian r’h (C), (C); Shughni pun(d), ra, ro (A), (A); Khotanese Saka pada (A), (A); Zazaki kuče (D), (D); Kumzari tayra (E), (E); Bandari räh (C), (C); Waziri LYAR (C), (C); Kurdish rê (C), (C); Pashto LAR (C), (C); Wakhi VEDEK (B), (B); Larestani jâdda, râ (C), (C); Ormuri lâr,rấ (A), (A); Yaghnobi rot,rah (C), (C); Avestan “pabl- (A), (A); Sogdian r’õ (C), (C); Ossetic faendag (A), (A); Persian RAH (C), (C); Balochi DAG (B), (B); 

root: Wakhi WIUX (D), (E), Yaghnobi wîta (D), (A); Larestani rişā (A), (A); Sogdian wyx (D), (D); Balochi PAR (B), (B); Kumzari irq (C), (C); Ormuri bêx (D), (D); Pashto BEX, RISA (D), (D); Zazaki riçe (A), (A); Persian rishe (A), (A); Avestan vârâša judiciary (A), (A); Parachi bix/γ¯ıx,kördi (D), (D); Kurdish reh (A), (A); Yaghnobi rişā (E), (E); Waziri BEKH, WEKH (D), (D); 

rope: Ishkashmi vuš (C), (A); Shughni bând,sarbând (E), (E); Yaghnobi wîta (D), (D); Balochi REZ (B), (B); Zazaki resen (B), (B); Waziri PERAI (), (); Pashto PERAJ (), (); Larestani band, resmu (C), (A); Wakhi TUNOV, DEROWI, NUS, SIVEN (A), (A); Khwarezmian tšyn (B), (B); Zebaki wâš (C), (C); Kumzari bayn (C), (A); Persian bač (C), (A); Persia TANAB (A), (A); Bandari risamün (B), (B); Ossetic baendaen (C), (C); 

rotten: Sogdian pwtqy (A), (A); Ormuri šrë-bük (D), (D); Waziri WROST (D), (A); Persian PUSIDE (C), (C); Khotanese Saka haMbûta- (A), (A); Ossetic aembyd (A), (A); Yaghnobi pûta (A), (A); Avestan pûtâ (A), (A); Pashto VROST (D), (D); Balochi GALAGH (B), (B); Waziri PIFK (A), (A); 

round: Larestani gerd (B), (B); Kurdish girover (B), (B); Zazaki tekele (C), (C); Yaghnobi lunda,kursak (D), (D); Sogdian ‘skwrnkh (E), (E); Avestan skarôno (A), (A); Shughni žarn,žurn (F), (F); 

rub: Wakhi MAND, SUX (A), (A); Zazaki sâwitišt (B), (B); Larestani mâleda mošta (A), (A); Waziri MASHEL (A), (A); Parachi astar (B), (B); Sogdian ‘nsy- (B), (B); Ormuri say- (B), (B); Pashto MUSEL (A), (A); Ossetic xafyn (C), (C); Balochi MALAGH, MALITHA, MALTHA (A), (A); Khotanese Saka dar- (D), (D); Persian MALIDAN (A), (A); 

salt: Ishkashmi námulGak (A), (A); Persian NAMAK (A), (A); Khotanese Saka namva- (A), (A); Khwarezmian šwr (F), (F); Ormuri nimêk (H), (A); Balochi WHADH, WAHADH (B), (B); Pashto MALGA (A), (A); Shughni namâk (G), (A); Parachi namâ (A), (A); Kumzari khiwah (D), (D); Yaghnobi namak (A), (A); Sogdian nm’ık (A), (A); Waziri MOLGA (A), (A); Zazaki sol (C), (C); Ossetic caexx (E), (E); Larestani namak (A), (A); Wakhi NIMUK (A), (A); 

sand: Larestani šenn, mása, jong (finer one), lamr (finest one) (D), (D); Bandari γiyâm (F), (F); Shughni şôš,reg (G), (G); Khotanese Saka syatâ (G), (G); Persian SHEN (D), (D);
say: Zazaki vätiš (E), (E); Zebaki GeZ- (F), (F); Balochi GUSHAGH, GUSTA, GWASHAGH, GWASHTA (D), (D); Avestan aoxté, mraoiti (A), (A); Ossetic dzuryn (), ();

scratch: Wakhi DRUP-, CUNGOL DI- (A), (A); Zazaki wirinayıš (C), (C); Persian KCHARIDAN (B), (B); Yaghnobi ručon- (E), (E); Balochi KCHARAGH, KCHARITH (B), (B); Shughni čangowl dedow (D), (D); Pashto GERDEL (C), (C); Old Persian tātiy (A), (A);

sea: Bandari derya (A), (A); Persian DARYA, OQYANUS (A), Old Persian draya (A), (A); Khwarezmian pwrt (H), (H); Zazaki dengiz (C), (C); Waziri SAMUNDAR, DARYOB (D), (D); Ossetic dendirzyz (F), (F); Kumzari deriyö (A), (A); Larestani daryā (A), (A); Avestan zraiiöl (A), (A); Middle Persian dry'b (A), (A); Pashto BAHR (E), (E); Shughni daryō (I), (A); Balochi SAMUNDAR (B), (B);

see: Balochi DITHA, GINDAGH (B), (B); Wakhi WIN- (A), (A); Pashto LIDEL (C), (C); Bandari diden (B), (B); Sogdian wyn (A), (A); Khwarezmian wyn- (A), (A); Persian DIDAN (B), (B); Old Persian vainatiy (A), (A); Shughni win (A), (A); Zazaki ditiš (B), (B); Waziri KATEL, LIDEL (C), (C); Yaghnobi wen- (A), (A); Avestan vaenaiti (A), (A); Kurđish ditin (B), (B); Khotanese Saka tcäs (D), (D); Ossetic wynyn (A), (A);

seed: Avestan taoxma, čθram (A), (A); Shughni tuqm (A), (A); Ossetic NAEMYG, GAGA, tawinag (D), (D); Parthian kšf’n, twxm (A), (A); Parachi tuxm (A), (A); Middle Persian twhm (A), (A); Larestani toxm tox (A), (A); Kumzari baidar (C), (C); Ishkashmi teG(m) (A), (A); Wakhi TUGHUM (A), (A); Persian tokhm (A), (A); Sogdian tγmy (A), (A); Pashto DANA (A), (A); Waziri TEMNA, TEMNA (A), (A); Yaghnobi taxm, txum (A), (A); Balochi bidz (B), (B); Kurđish birz, tov (B), (B); Khotanese Saka ttima- (A), (A);

sew: Zazaki derzayiš (C), (C); Shughni pami (G), (G); Middle Persian 'bzyn- (F), (F); Waziri GANDEL (D), (D); Yaghnobi šiy- (E), (E); Sogdian šwm- (E), (E); Balochi DOSHAGH, DOKHTA (B), (B); Persian DUKHTAN (B), (B); Pashto DZORAVEL (C), (C); Wakh driv- (A), (A);

sharp: Waziri TERA (A), (A); Persian TIZ (A), (A); Wakhi TEGHD, TIZ (A), (A); Avestan tıgyro (A), (A); Ossetic cyrg (A), (A); Larestani tiz (A), (A); Balochi TEZ (A), (A); Khotanese Saka khaudala- (E), (E); Shughni tēz (C), (A); Old Persian tigra (A), (A);
short: Waziri LAND (E), (E); Zazaki kilm (D), (D); Pashto MUXTASAR, LAND (E), (E); Persian kutah (B), (B); Sogdian nwrzk (A), (A); Balochi GWAND (C), (C); Avestan mórzöt (A), (A); Parachi l nød (H), (H); Khotanese Saka mulysga- (A), (A); Ormuri länd (G), (G); Parthian qmbyg (J), (J); Shughni kat.kut (I), (I); Yaghnobî kältâ (I), (I); Wakhi KUT (B), (B); Ossetic cybyr (F), (F);

sing: Waziri SANDARA (SONG) (E), (E); Pashto VAJEL (F), (F); Sogdian p\'š- (J), (J); Larestani xanda (C), (C); Bandari šarvar zadên (D), (D); Ossetic zaryn (G), (G); Yaghnobî zoj.xofizi kun- (I), (I); Persian AVAZ KHANDAN (C), (C); Wakhi BAEXIT XAN- (A), (A); Balochi GUSHAGH, GUSHTA, GWASHTA (B), (B); Zebaki GöZâk (H), (H); Shughni luv (K), (K);

sit: Larestani šessa (C), (C); Persian NESHASTAN (C), (C); Shughni nið- (B), (B); Yaghnobî nið- (B), (B); Kurdish dañışın, rûniştin (C), (C); Bandari neşten (C), (C); Ishkashmi nêd- (B), (B); Parthian nšyy- (B), (B); Ossetic bayn (E), (E); Middle Persian nšyy- (B), (B); Zebaki nið- (B), (B); Bashkardi nen (C), (C); Pashto KSENASTEL (D), (D); Avestan âste (A), (A); Sogdian nst- (B), (B); Wakhi NEZD (B), (B); Waziri KSHENAWEL, NOSTAI, PAND, (SITTING) (D), (D); Balochi NINDAGH, NISHTA (B), (B); Khotanese Saka Nâsta (B), (B);

skin: Sogdian crm (C), (C); Larestani püss (E), (B); Waziri TSARMAN (C), (C); Pashto POST (G), (B); Kurdish çerm (D), (D); Wakhi PIST (B), (B); Ishkashmi kurust.korost (D), (D); Parachi püst (H), (B); Shughni pust (J), (B); Ossetic carm (C), (C); Persian püst (E), (B); Khwarezmian crm (C), (C); Kumzari pôst (F), (B); Khotanese Saka kängä-,chala.ttanâ (D), (D); Middle Persian pwst.crm (C), (B); Balochi PHOST (C), (B); Yaghnobî pust (I), (B); Ossetian suri (A), (A);

sky: Wakhi OSMON (A), (A); Yaghnobî osmon (A), (A); Pashto ASMAN (A), (A); Avestan asma (diìawu- Daevic) (A), (A); Bashkardi yâlmuôn (A), (A); Yazgulyam asmân (A), (A); Shughni osmun (A), (A); Waziri ASMON (A), (A); Zazaki äzmân (A), (A); Persian ASEMÁN (A), (A); Parachi âğêš < Indic? (C), (C); Ishkashmi äşmân (A), (A); Middle Persian (\')sm\'n (A), (A); Old Persian asmân (A), (A); Balochi ARSH, AZMAN (A), (A); Kumzari asmay\'no (A), (A); Ormuri âsman,falak (A), (A); Parthian (\')sm\'n (A), (A); Sogdian (\'k\'c) (D), (D); Ossetian arv (B), (B);

sleep: Ormuri xau (A), (A); Zazaki rä-kewtìš, witiš (A), (A); Shughni žofc (A), (A); Yaghnobî uñš,nepid (A), (A); Khotanese Saka üm (G), (G); Sogdian γwëzûn- (A), (A); Avestan xvafszait (A), (A); Khwarezmian 'zmxs (F), (F); Waziri KHEB (A), (A); Persian KHABIDAN (A), (A); Middle Persian hwpw- (A), (A); Kurdish xewtin, razan (A), (A); Larestani xata (A), (A); Ossetic FYNAEJ KAENYN, xuyssyn (E), (E); Pashto BIDEDEL (D), (D); Wakhi RUXP-, RUSEP- (B), (B); Balochi AKSAGH, AKASTHA
small: Ishkashmi cutökok (I), Zazaki werdı,qiž,šenik (E), Kumzari chik (D), Khotanese Saka jseıNa- (K), Larestani kəydiı (E), Waziri KAM, KAMKAI, WRIKAI WOR (D), Shugnwi -buc,žal, (G), Ormuri zärı (G), Parachi čimö (H), Wakhi ZUQ, ZUQIQ, ZUQuLAEı (B), Kurdish biçûk (D), Sogdian kən’kk (D), Balochi KIK (C), Avestan kasuš (A), Zebaki cuT (I), Persian kuchek (D), Middle Persian qwdk (E), Pashto KUCNAJ, VOR (D), Ossetic cysyl, gyccyl (F), Yaghnobi pulla,maydaak (J),

smell: Balochi BO GIRAGH (A), Shugnwi buy čidow (A), Larestani bii kerda (A), Persıan BU KARDAN (BU SHENIDAN) (A), Zazaki boy k [sic?] (A), Pashto HYS KAVEL (B),

smoke: Pashto DUD (A), (A), Wakhı DIT (A), Yaghnobi payst,dud (B), (B), Ishkashmi dıt (A), (A), Persıan DUD (A), (A), Larestani dùd (A), (A), Shugnwi ğud (A), (A), Ormuri dūd (A), (A), Sogdian pzt (B), (B), Khotanese Saka dümä (A), (A), Yagızulgam ?ąd (A), (A), Parachi dhi (A), (A), Kurdish dûkel, dûxan (A), (A), Zazaki diy (A), (A), Middle Persıan dwd (A), Ossetic faezdaeg, x’uaecae (B), (B), Balochi DUHON (A), (A), Parthıan dwd (A), (A), Kumzari dür (A), (A), Waziri DE YOR LIGAI (A), (A),

smooth: Waziri SHOE, SHWE (D), (D), Balochi LASUR (A), (A), Zazaki sermitikin (B), (B), Ossetic laeg’z (A), (A), Persıan SAF (C), (C), Yaghnobi liṅnx,hamwor (A), (A), Wakhı LUS, HUNWOR (A), (A), Pashto SAF, MUSTAVI, HAVAR (E), (E), Shugnwi anmwor, daʾxt (A), (A),

snake: Sogdian kyrım (E), (E), Khotanese Saka šsayda- (G), (G), Waziri MANGER (C), (C), Wakhı FUKS (B), (B), Ishkashmi voks (F), (F), Persıan mar (C), (C), Zazaki mär (C), (C), Larestani mär (C), (C), Balochi MAR (C), (C), Ormuri mär (C), (C), Avestan ažiz (A), (A), Pashto MAR (C), (C), Yaghnobi mor (C), (C), Parachi haždär,kirm (E), (E), Shugnwi dívıısk, mör (H), (H), Ossetic kalm (D), (D), Kumzari mär (C), (C),

snow: Larestani vafr, barf [latter < Pers?] (A), (A), Shugnwi dınıö,žınıj (D), (D), Ormuri γoš (A), (A), Persıan BARF (A), (A), Parthıan wfr,wpr (A), (A), Balochi BAWAR (A), (A), Avestan vafr- (A), (A), Ishkashmi varf (A), (A), Parachi γarp (A), (A), Ossetic mit (C), (C), Waziri WOVRA (A), (A), Yaghnobi wafr (A), (A), Middle Persıan wfr,wpr (A), (A), Sogdian wfr’ (A), (A), Wakhı ZEM (B), (B), Zazaki vewr (A), (A), Pashto VAVRA (A), (A),

some: Pashto CO, CE KADR (D), (D), Yaghnobi çof,çandin (E), (E), Persıan BA’ZI (E), (E), Shugnwi ač, báxi, çand (H), (H), Balochi KHARD-E (C), (C), Old Persıan *kaičiy (A), (A), Ossetic CYSYŁ, caldaer (G), (G), Wakhı TSUM, TSUMER, KUMD, CIZ (B), (B), Zazaki čend,tayı(n) (D), (D), Waziri DZENE, TSE (F), (F),
spit: Wakhi SEX (B), (B); Shughni tuf čidów (G), (G); Ossetic tu käynyn (C), (C); Sogdian γωρβ (E), (E); Avestan spāma- (mn.) (A), (A); Larestani tof kerda (D), (D); Parthian wf-? (E), (E); Khotanese Saka khaur-? (F), (F); Pashto TUKEl, TUKAVEL (C), (C); Waziri TIKAWEL (C), (C); Balochi THUK (SB.) (C), (C); Persian TOF KARDAN (D), (D);

split: Balochi BURAGH, BURITHA (B), (B); Ossetic FANDYN, uaryn (G), (G); Zazaki šeqnāyiš (C), (C); Pashto MATAVEL (F), (F); Avestan darōdairaiiāt (opt.) (A), (A); Sogdian kβ (H), (H); Persian NESF KARDAN (D), (D); Waziri CHAWEL (E), (E);

squeeze: Pashto KSEKSEL (D), (D); Balochi DABAGH, DABITHA (B), (B); Ossetic aelx’ivyn (E), (E);

stab: Wakhi XULA DI- (A), (A); Ossetic AERGAEVDYN, sadzyn (C), (C); Pashto HALALAVEL (B), (B);

stand: Balochi OSHTAGH, OSHTATHA (A), (A); Waziri DAREDEL (D), (D); Larestani vayseda, vā rosseda (A), (A); Ossetic laeuyun (E), (E); Yaghnobi ušt- (A), (A); Sogdian ’wst- (A), (A); Shughni wirāfč (F), (F); Kurdish rabūn, westīn (B), (B); Bandari vūstāden (A), (A); Middle Persian ’ys- (A), (A); Kumzari qawumah, sakhō (C), (C); Avestan hištaiti (A), (A); Pashto DAREDEL (D), (D); Khotanese Saka stā- (A), (A); Persian istadan (A), (A); Old Persian stā- (A), (A);

star: Waziri STORAI (A), (A); Bashkardi estāla (A), (A); Shughni āterž (A), (A); Kurdish stēr (A), (A); Balochi ISTAR (A), (A); Zebaki sitāra (A), (A); Ishkashmi struk (A), (A); Yaghnobi bildainga,sitāra (C), (C); Sogdian ’st’r’kt (A), (A); Old Persian ’st’rg (A), (A); Parachi ātēč (A), (A); Persian SETARE (A), (A); Pashto STORAJ (A), (A); Zazaki ātēr (A), (A);

stick: Balochi LATH (B), (B); Yaghnobi dork,kaltak,tayoq (I), (I); Pashto LARGAJ, LAKARA (F), (F); Ossetic laedzaeg (F), (F); Parachi kō? (G), (G); Sogdian ptr’b’k (J), (J); Waziri LARGAI (F), (F); Kumzari bākūr (E), (E); Wazhi GHUWZ, SUNG (A), (A); Khotanese Saka daula (K), (K); Zazaki čiwe,ušīr (C), (C); Ishkashmi Guča (H), (H); Persian CHUB (D), (D); Shughni nuqobā (L), (L);

stone: Avestan asma (A), (A); Ishkashmi sung (A), (A); Ormuri gap (K), (K); Pashto DABARA (I), (I); Kumzari raygh (F), (F); Khotanese Saka daḍāye (M), (M); Sogdian sng (A), (A); Shughni sang, ūr (A), (A); Bashkardi sax (G), (G); Zazaki šī (E), (E); Wazhi GHAR (B), (B); Ossetic dur (J), (J); Parachi ār (B), (B); Yaghnobi sanš (A), (A); Parthian sn(n)g (A), (A); Balochi KHOH (C), (C); Waziri KONRAI, TIZHA (H), (H); Persian SANG (A), (A); Larestani sang, kolom (a breakable one) (A), (A); Old
Persian asā, aṁā- (A), (A); Kurdish ber, kevir (D), (D); Khwarezmian snk? (A), (A); Yazgulyam Grtsōk (L), (L);

**straight**: Ossetic rast, KOMKOMMAE (A), (A); Zazaki rāšt (A), (A); Persian RAST (A), (A); Pashto SAM (D), (D); Sogdian fršty (A), (A); Waziri SAM, SIKH (D), (D); Yaghnobi rask,rost (A), (A); Wakhī ROST (A), (A); Balochi SIDHA (B), (B); Ormuri rāst (A), (A); Parthian rʿst (A), (A); Avestan ʾorzuš (A), (A); Shughni čuk, caq (E), (E); Kumzari aydil (C), (C); Middle Persian rʾst (A), (A);

**suck**: Yaghnobi zimak-,diy-,makon xwar- (E), (E); Ossetic daejyn, C"IRYN (D), (D); Wakhi SUP- (A), (A); Balochi MISHAGH, MIKHTA (B), (B); Persian MEKIDAN (B), (B); Shughni sipāf (A), (A); Pashto RAVDEL (C), (C);

**sun**: Middle Persian xwr (A), (A); Balochi RO, ROSH (B), (B); Khorasan roj, xor (B), (B); Wakhī OFTOB, YIR (A), (A); Shughni xir, ʿoftāb (A), (A); Waziri LMER, MYER (D), (D); Zebaki ʿormozd (G), (G); Khwarezmian ʿxyr (A), (A); Ishkhashmi remuz (F), (F); Pashto LMAR (D), (D); Larestani ʿaftow, xoršīd (C), (C); Yazgulyam xᵛār (A), (A); Khotanese Saka urmaysde- (G), (G); Ormuri mes (E), (E); Persian AFTAB, KHORSHID (C), (C); Kumzari īntāfū (C), (C); Sogdian γwr (A), (A); Zazaki rož (B), (B); Yaghnobi xur, ʿaftob (A), (A); Avestan huuaro (A), (A); Ossetic xur (A), (A); Parthian ʿmhr (H), (H);

**swell**: Avestan sispimnō (ptc.) (A), (A); Waziri PARSEDEL (E), (E); Persian BAD KARDAN (D), (D); Pashto PARSEDEL (E), (E); Balochi SIAGH, SITHA (C), (C); Sogdian ʿdmʾs (B), (B); Wakhī PEDEMES- (B), (B); Larestani ʿad kerda, ʿa barāʿ onda (D), (D); Ossetic RAESIJYN, dymyn (B), (B); Khotanese Saka nar- (F), (F); Shughni waram ʿcidow (G), (G);

**swim**: Persian SHENA KARDAN (A), (A); Shughni wāz, ḵinow (A), (A); Pashto LAMBO VAHEL (C), (C); Balochi TARAGH, TARATHA, THAHARAGH, THAHARTHA (B), (B); Kumzari ʿshinaw gīʾdish (pst) (A), (A); Larestani ʿšenaw kerda (A), (A); Khorasan ʿaftow, listin, mele kirin (A), (A); Wakhī USINAWERI TSER, QELOC XUS-., WEZAN DI- (A), (A); Parachā ʿubāzī- (A), (A); Zazaki ʿazne (A), (A); Sogdian fsnʾy- (A), (A); Ossetic lenk kaenyn, nakac ʿaenyn (D), (D); Waziri LAMBEYA (SWIMMING) (C), (C);

**tail**: Pashto LAKEJ (D), (D); Parachā dōγund (A), (A); Khotanese Saka dumaa- (A), (A); Wakhī KICIKA (B), (B); Balochi DUMB (A), (A); Khwarezmian ʿdmʾ (A), (A); Yaghnobi duym, dumba (A), (A); Avestan dumā- (A), (A); Khorasan ʿbç, teri (C), (C); Persian dom (A), (A); Ishkhashmi dūm (A), (A); Zazaki pōč (C), (C); Sogdian ʿdmʾ (A), (A); Larestani dom (A), (A); Ossetic dyamaeg (A), (A); Waziri LAKAI (D), (D); Bandari dom (A), (A); Shughni ʿdm (A), (A);

**that**: Kumzari ān (B), (B); Sogdian xw (A), (A); Persian AN (UN) (B), (B); Waziri AGHA, HAGHA (B), (B); Bandari ā (B), (B); Avestan hō (A), (A); Ossetic ucəy (B), (B);
Balochi AN (B), (B); Old Persian av (B), (B); Shughni yā (C), (C); Kurdish ew (B), (B); Pashto HAGA, HUGA (B), (B); Wakhi YA (C), (C);

there: Waziri WOLATA (D), (D); Zebaki tāda (A), (A); Yaghnobi wat (A), (A); Parachi ok(e)i (F), (F); Old Persian avada (A), (A); Ishkashmi wadak (A), (A); Balochi ODHA (A), (A); Shughni yamand, yamard (G), (G); Kumzari anso (C), (C); Ossetic um (E), (E); Sogdian ‘wc’ (A), (A); Persian unja (A), (A); Wakhi DRA, HUDRA, TRA, HUTRA, TRET, DRET (B), (B); Pashto HALTA (D), (D); Zazaki owrā (A), (A);

they: Waziri AGHA (C), (C); Avestan tē (A), (A); Balochi ESHAN (A), (A); Ossetic udon (E), (E); Larestani ‘anayā (B), (B); Shughni dād, wād (A), (A); Yaghnobi axtit (F), (F); Kumzari shan (A), (A); Persian ishan (A), (A); Wakhi YUST, HAEIUST (A), (A); Sogdian ‘wy(h) (G), (G); Pashto DUJ, JE (D), (D); Zazaki ē (A), (A);

thick: Persian KOLOFT (C), (C); Waziri GHWUT (D), (D); Parachi estőrő (G), (G); Balochi THULAR (B), (B); Khotanese Saka dara- (I), (I); Pashto PEND, ZAXIM (E), (E); Yaghnobi Gafs, hula, farbi (H), (H); Ossetic ystadv (F), (F); Wakhi BAJ (A), (A); Larestani koloft, got (C), (C); Shughni farbiyaki (H), (H);

thin: Larestani nazok, tanok (D), (D); Khotanese Saka ttaMga- (F), (F); Balochi LAGHAR (B), (B); Shughni birik, bōrik, nōzuk, tānuk (F), (F); Ishkashmi tānuk (F), (F); Zazaki bāriʔek (C), (C); Yaghnobi tunuk (F), (F); Persian BARIK, NAZOK (D), (D); Sogdian γγmr (G), (G); Ossetic x'uydy kaen (F), (F); Wakhi SENOR, SENUF, TENUK, BIRIK, XEROB (A), (A); Pashto NARAJ (E), (E); Waziri NARAI (E), (E);

think: Parthian nwywr- (H), (H); Old Persian maniyata (A), (A); Balochi ZANAGH, ZANTHA (B), (B); Shughni fikri ēdow, xayol ēdow (I), (I); Bandari fekri akerden (D), (D); Ossetic x'uydy kaen (F), (F); Sogdian šm'r (G), (G); Pashto FIKR KAVEL, GUMAN KAVEL (E), (E); Avestan mainiilete (A), (A); Persian FIKR KARDAN (C), (C);

this: Persian IN (A), (A); Bactrian ei(i)o (A), (A); Avestan aem (A), (A); Balochi HAM-ESH (A), (A); Zebaki am (A), (A); Waziri DAI, DA, DAGHA (B), (B); Kurdish ev (A), (A); Kumzari īyah (A), (A); Sogdian ʿyō (A), (A); Pashto DA, DAGA (B), (B); Khotanese Saka tta- (E), (E); Larestani ‘eđe, ‘e (B), (B); Ishkashmi nakav (C), (C); Wakhi YEM, HAEIEM (A), (A); Yaghnobi š (D), (D); Parachi ē (A), (A); Old Persian hauv, iyam (A), (A);

thou: Wakhi TU (A), (A); Zebaki tô (A), (A); Ossetic dy (A), (A); Sogdian tγw (A), (A); Ishkashmi tu (A), (A); Larestani to, šomā (A), (A); Yaghnobi tu (A), (A); Bandari to (A), (A); Kurdish tu (A), (A); Balochi THAU (A), (A); Kumzari tô (A), (A); Zazaki tô (A), (A); Khotanese Saka tha (A), (A); Avestan tūm, ʿθam (A), (A); Pashto TE (A), (A); Khwarezmian ‘wtk (A), (A); Old Persian tuvām, ʿθuvām (A), (A); Waziri TE (A), (A); Shughni tu, tôθ (A), (A); Persian to (A), (A);
three: Balochi SAI (A), (A); Wakhi TROI (A), (A); Middle Persian sh (A), (A); Waziri DRE (A), (A); Kumzari soh (A), (A); Old Persian *çi- (A), (A); Zebaki rāi (A), (A); Ormuri şo (A), (A); Pashto DRE (A), (A); Ossetic aertae (A), (A); Parachi ši,ši (A), (A); Persian SE (A), (A); Ishkashmi rūi (A), (A); Zazaki hire (A), (A); Yazgulyam tsoi (A), (A); Yaghnobi tiray,se (A), (A); Avestan Trāii¯ o (A), (A); Sogdian šy (A), (A); Parthian hry (A), (A); Khotanese Saka drai (A), (A); Shughni ara,aray,še (A), (A); Larestani se (A), (A); Khwarezmian šy (A), (A);

throw: Bandari kardünden (G), (G); Persian ANDAKHTAN (E), (E); Waziri ACHAWEL, WOCHAWEL, TREYEL (H), (H); Old Persian ahiyatiy (A), (A); Zazaki eštiš,vistiš (D), (D); Avestan a?hiieiti, spaiieiti (A), (A); Pashto ACAVEL, GURZAVEL (H), (H); Khotanese Saka ah (K), (A); Yaghnobi wid-, pártow- (J), (F); Balochi PHIRENAGH, PHIRENTHA (C), (C); Shughni sikawak dedow (L), (J); Ossetic aepparyn (I), (I); Larestani bessa (F), (F); Wakhi KUT- (B), (B);

tie: Ossetic battyn (A), (A); Persian BASTAN (A), (A); Zazaki bestiš (A), (A); Wakhi VUND (A), (A); Balochi BANDAGH, BASTHA (A), (A); Avestan bandaiieiti, hiš¯ aiia (pf.) (A), (A); Pashto TAREL (B), (B); Old Persian band- (A), (A); Larestani bassa (A), (A); Waziri TAREL (B), (B); Yaghnobi vant- (A), (A); Sogdian pš’y- (C), (C); Shughni vind (A), (A);

tongue: Larestani zabu (A), (A); Pashto ZEBA (A), (A); Old Persian hazāna- (A), (A); Waziri ZHEBBA (A), (A); Kurdish ziman (A), (A); Kumzari zuwān (A), (A); Avestan hizuua (A), (A); Zebaki zevuk (A), (A); Middle Persian ‘zw’n (A), (A); Ormuri zubān (B), (B); Zazaki ziwān (A), (A); Parachi (zu)bān (C), (C); Ishkashmi zivuk (A), (A); Balochi ZAWAN (A), (A); Yaghnobi zivok,zabon (A), (A); Sogdian zb’q (A), (A); Persian zaban (A), (A); Wakhi ZIK (A), (A); Parthian ‘zb’n (A), (A);

tooth: Khotanese Saka danda- (A), (A); Larestani dudu (A), (A); Zebaki dāndak (A), (A); Balochi DATHAN (A), (A); Zazaki dindān (A), (A); Sogdian dnt’ (A), (A); Khwarezmian γš (B), (B); Wakhi DENDUK (A), (A); Yaghnobi dindak (A), (A); Waziri GHOWSH, GHOSH (B), (B); Ishkashmi dānd (A), (A); Persian danān (A), (A); Avestan dantān (pl.), dātā- (), (); Ossetic daendag (A), (A); Pashto GAS (B), (B); Parachi danān (A), (A); Ormuri giši (B), (B); Shughni dindūn (A), (A);

tree: Balochi DRASHK (B), (B); Zazaki dār (B), (B); Zebaki daraxt (F), (F); Khotanese Saka kšāukā,tīrā- (G), (G); Ormuri d(a)raxt (E), (E); Ossetic balac (D), (D); Pashto DIRAXT, VENA (A), (A); Kumzari shidreh (C), (C); Shughni d(i)raxt, dorg (B), (B); Persian deraht (B), (B); Yaghnobi darāxt,dork (B), (B); Waziri WUNA (A), (A); Kurdish dar (B), (B); Larestani deraxt derax (B), (B); Parachi bhīn (A), (A); Wakhi DERUXT (B), (B); Avestan vanā (A), (A); Middle Persian *d’rw (B), (B); Khwarezmian wn’yk (A), (A); Sogdian wn’ (A), (A);
turn: Ossetic yzdaxyn (E), (E); Old Persian varataiy (A), (A); Sogdian prn’z (G), (G); Avestan uruuisieiti (A), (A); Shughni garDên (B), (B); Zazaki čerx (C), (C); Yaghnobi laksón-,tob- (F), (F); Wakhi γiir(d) (B), (B); Pashto GERZEDEL (B), (B); Persian PICHIDAN (D), (D); Waziri GERZEDEL (B), (B);

two: Parthian dw (A), (A); Yazgulyam ?au (A), (A); Wakhi BOJ (A), (A); Persian DO (A), (A); Kumzari doh (A), (A); Khotanese Saka duva (A), (A); Shughni du (A), (A); Waziri DWA (A), (A); Ishkashmi dau (A), (A); Avestan duua (A), (A); Old Persian *duva- (A), (A); Middle Persian dw (A), (A); Ossetic duuae (A), (A); Larestani dom (A), (A); Pashto DVA (A), (A); Yaghnobi du (A), (A); Zebaki dŏv (A), (A); Kurdish du (A), (A); Waziri DWA (A), (A);

vomit: Waziri KAI (F), (F); Balochi UCHALNA (C), (C); Wakhi WOQ- (B), (B); Pashto KANGI KAVEL (OQ ZADAN) (D), (D); Avestan duua (A), (A); Old Persian *duva- (A), (A); Middle Persian dw (A), (A); Ossetic duuae (A), (A); Larestani du (A), (A); Shughni du (A), (A); Pashto DVA (A), (A); Yaghnobi du (A), (A); Zebaki dŏv (A), (A); Kurdish du (A), (A); Parachi du (A), (A);

walk: Waziri SAIL (G), (G); Avestan jasaiti (A), (A); Pashto TLEL (H), (H); Persian RAH RAFTAN (D), (D); Balochi JUZAGH (A), (A); Larestani ră ceda, gešta (F), (F); Kurdish çûn (B), (B); Ossetic caeuyn (I), (I); Kumzari maysh gid’ish (pst) (E), (E); Zazaki šiyāyiš (C), (C); Yaghnobi laks- (J), (J);

warm: Yaghnobi širagarm (A), (A); Kurdish germ (A), (A); Zazaki germ (A), (A); Avestan garamo, taptō (A), (A); Pashto TOD (B), (B); Sogdian sw- (D), (D); Bandari garm (A), (A); Middle Persian grm (A), (A); Balochi GARM (A), (A); Wakhi GERM, THIN, SONDER (A), (A); Persian GARM (A), (A); Ossetic x’arm (C), (C); Waziri TOD (B), (B); Khwarezmian γmnd (A), (A); Larestani garm (A), (A); Shughni gārm (A), (A);

wash: Wakhši WUZDI- (A), (A); Shughni zini (A), (A); Avestan naenīzaiti, snaieiti (A), (A); Zazaki šitiš (B), (B); Sogdian sn’y (A), (A); Larestani sossa (B), (B); Yaghnobi sinoy (A), (A); Balochi SHODHAGH, SHUSTA (B), (B); Ossetic aexsyn (A), (A); Kumzari chór (sg) (C), (C); Middle Persian sm’y (A), (A); Pashto MINDZEL (D), (D); Bandari šušt (B), (B); Waziri WINZEL (D), (D); Persian SHOSTAN (B), (B);

water: Bashkardi SBš. yāp, NBš. yā/āu (A), (A); Khotanese Saka ūtčā (B), (B); Wakhši YUPK (A), (A); Balochi AF (A), (A); Parachi āwō (A), (A); Yaghnobi op (A), (A); Avestan āfs (A), (A); Larestani ’aw (A), (A); Pashto OBE (A), (A); Persian AB (A), (A); Ossetic don (B), (B); Zazaki āw (A), (A); Bandari hū (A), (A); Sogdian ”p’p (A), (A); Bactrian y’b (A), (A); Waziri EBO (A), (A); Kurdish av (A), (A); Ormuri wōk (A), (A); Ishkashmi wek (A), (A); Kumzari hau (A), (A); Parthian ’b (A), (A); Shughni xac (C), (C); Middle Persian ’b (A), (A); Zebaki wēk (A), (A); Old Persian āpiš (A), (A);
we: Yaghnobi mox (A), (A); Kumzari mah (A), (A); < ‘all’ (A), (A); Parachi mà (A), (A); Bashkardi yamah (A), (A); Zazaki mà (A), (A); Persian ma (A), (A); Wakhi SUK (B), (B); Kurdish em, me (A), (A); Balochi MA (A), (A); Avestan ahma, vaem (A), (A); Bandari mà (A), (A); Old Persian vayam (A), (A); Pashto MUZ (A), (A); Shughni màëš (A), (A); Sogdian m’x (A), (A); Waziri MIZH (A), (A); Ossetic max, maxtæ (A), (A); Larestani ‘amà [not cognate to Pers MA, MA;

wet: Shughni čalìn, nambdór, xist (A), (A); Bandari tar (C), (C); Ormuri šur (G), (G); Larestani tal (C), (C); Persian TAR (C), (C); Avestan dama (A), (A); Balochi THAR (C), (C); Parachi phyö (H), (H); Yaghnobi tan,obnoka,tar (C), (C); Ishkashmi shuluk (G), (G); Zazaki hi (D), (D); Ossetic xuylydz (F), (F); Pashto XIST, LUND (E), (E); Wakhi XUSC (B), (B); Waziri LIMD, TOND (E), (E); Sogdian nštct’ (A), (A);

what: Avestan kat (A), (A); Sogdian cw (A), (A); Waziri KIM, TSE (A), (A); Balochi KITHAN (A), (A); Parthian cy (A), (A); Avestan naptò (A), (A); Balochi THAR (C), (C); Parachi čhö (A), (A); Larestani ce, ci (A), (A); Yaghnobi ço (A), (A); Ossetic ca (A), (A); Persian che (A), (A); Wakhi CIZ, CICIZ (A), (A);

when: Wakhi TSOGHG, TSOGHDER, TSEWUXT (A), (A); Persian key (A), (A); Kumzari kayi (A), (A); Pashto KEŁA (A), (A); Balochi KHADHE (A), (A); Sogdian krwò (A), (A); Larestani key (A), (A); Shughni ca, ca wæxt (B), (B); Ossetic kaed, kuy (A), (A); Yaghnobi ka, kax (A), (A); Persian koja (A), (A); Wakhi CIZ, CICIZ (A), (A);

where: Sogdian ’kwrò (A), (A); Pashto CIRI, CIRTA (A), (A); Khotanese Saka ku (A), (A); Kumzari gäyä (A), (A); Shughni ar kä, kaçad (A), (A); Ossetic kaemptiony (A), (A); Yaghnobi ku (A), (A); Wakhi KUMER, KUMJAEI (A), (A); Balochi THANGO (B), (B); Larestani koja, ko (A), (A); Waziri CHERN (A), (A); Avestan kufrä, kû (A), (A); Ormuri gudä (A), (A); Bandari kojä (A), (A);

white: Ormuri spëw (A), (A); Zazaki sipë (A), (A); Waziri SPIN (A), (A); Ormuri ’sp’yt (A), (A); Bashkardi SBš. espit, NBš. espir (A), (A); Ossetic urs (C), (C); Pashto SPIN (A), (A); Avestan spaetà-, auriusö (A), (A); Persian SEFID (A), (A); Yaghnobi sipëta, safèd (A), (A); Balochi SWETH (A), (A); Zebaki surxën (E), (E); Parthian ’spyd (A), (A); Zebaki ROXUN (B), (B); Larestani safid (B), (B); Zazaki spër (A), (A); Parachi šchatö (D), (D); Khotanese Saka ššità- (A), (A); Middle Persian ’spyd (A), (A); Persian spë (A), (A); Shughni safèd (D), (D); Ishkashmi safèd (C), (C);

who: Zazaki käm (A), (A); Shughni ar tăm (B), (B); Parachi kî (A), (A); Bactrian kyd (A), (A); Persian kî (A), (A); Bandari kî (A), (A); Old Persian ka (A), (A); Zebaki kàî (A), (A); Yaghnobi ka,kax (A), (A); Parthian ky (A), (A); Kurdish kî (A), (A); Waziri TSOK (A), (A); Avestan kô (A), (A); Pashto COK (A), (A); Sogdian qy (A), (A); Balochi
KITHAN (A), (A); Wahhi KUI (A), (A); Larestani ke,ki,ka (A), (A); Ossetic ci, ciiae (A), (A); Middle Persian ky (A), (A); Ishkashmi kudum (A), (A);

wide: Zazaki herā (A), (A); Ishkashmi pām (A), (A); Balochi PRAH (A), (A); Avestan pərətš, pərənə (A), (A); Yaghnobi yaxt,pahm (A), (A); Pashto PRAX (A), (A); Waziri PLAN (A), (A); Wakhi FERUX, KESOD (A), (A); Shugnī dašt, firōx (B), (B); Sogdian əpə’y (A), (A); Persian PAHN, GOSHAD (A), (A);

wife: Balochi ZAL (C), (C); Ormuri nāk (A), (A); Kumzari zank (C), (C); Zebaki kūc (F), (F); Ossetic binojnag (C), (C); Avestan nāiri, nārika (A), (A); Waziri TABAR (D), (D); Sogdian δβ’mbn (G), (G); Yaghnobi inč,kuč (F), (F); Larestani zena (C), (C); Persian ZAN (C), (C); Wakhi JUMAUT, KEND, YUPKWOR (B), (B); Bandari zan (C), (C); Zazaki jinī (C), (C); Ishkashmi Zānj (C), (C); Pashto SEDZA (E), (E); Shugnī ɣiŋ (C), (C);

wind: Waziri BOD (A), (A); Persian BAD (A), (A); Larestani bād (A), (A); Pashto BAD (A), (A); Zazaki và (A), (A); Ossetic DYMGAE, uad (A), (A); Avestan vaṭō (A), (A); Yaghnobi wot,bōda,smol (A), (A); Sogdian w’δ (A), (A); Parachi dhamān (D), (D); Wakhi SEMOL, DUMA (B), (B); Shugnī šamol, xuţ (E), (E); Balochi GO (C), (C);

wing: Parthian b’zwr (B), (B); Sogdian w’z (B), (B); Avestan paramām (A), (A); Balochi PHAR (A), (A); Shugnī p̄r (A), (A); Zazaki per (A), (A); Ossetic BAZYR (B), (B); Larestani fal (of small birds), bāl (of big birds) (A), (A); Waziri PAR, WAZAR (A), (A); Wakhi TUP (B), (B); Middle Persian b’zwr (B), (B); Parachi bāl (A), (A); Persian PAR (A), (A); Pashto VAZAR (B), (B); Yaghnobi par,wanot,bol (A), (A);

wipe: Persian PAK KARDAN (C), (C); Yaghnobi rant- (H), (H); Pashto VUCAVEL (F), (F); Ossetic saerfn (G), (G); Waziri MASHEL (E), (E); Wakhi VISUV, TUF DI- (A), (A); Zazaki besterdiš (B), (B); Larestani xoš kerda (D), (D);

with: Balochi GO, GON, GON (C), (C); Wahhi DU (B), (B); Persian BA (D), (D); Avestan haďa, mať (A), (A); Shugnī as...ti,qati (I), (I); Pashto DE...SERA (G), (G); Waziri SARA, DE...SARA, PA...SARA (F), (F); Old Persian hadā (A), (A); Ossetic aed, -imaе (H), (H); Sogdian dn (B), (B); Kumzari wāh (E), (E); Middle Persian s’r (J), (J); Zebaki gal (C), (C); Zazaki bī,rey-di,ève (D), (D);

woman: Zazaki jīnēk (A), (A); Balochi ZAL (A), (A); Yaghnobi inčak,žan (A), (A); Ossetic binojna? (A), (A); Waziri SHEZA (C), (C); Wakhi AURUT, OJIZ, XUINUN (B), (B); Pashto SEDZA (C), (C); Ishkashmi štuk (C), (C); Shugnī awrāt, bōwā, ɣiŋik, kaxōy (A), (A); Sogdian ſtṛc (C), (C); Kurdish jīn, afret (A), (A); Persian ZAN (A), (A); Kumzari zank (A), (A); Avestan jainiš, strī, nāri (A), (A); Zebaki štāk (C), (C); Larestani zena, za’ifa (A), (A); Bandari zan (A), (A);

woods: Pashto DZANGAL (F), (F); Balochi LADH (B), (B); Persian jangal (C), (C); Ossetic x’aeda (G), (G); Khwarezmian ɻ(y)(n) (F), (F); Larestani jangal, jangal-e kaydū
worm: Ishkashmi putsuk (E), (E); Sogdian kyc’kh (F), (F); Persian KERM (A), (A); Yaghnobi kirm (A), (A); Larestani kerm (A), (A); Ormuri kirm (A), (A); Balochi KIRM (B), (B); Parachi kirmāk (A), (A); Waziri CHENJAI (D), (D); Khotanese Saka gūN¯a,pāra- (G), (G); Pashto KIRM (A), (A); Zazaki bi‘ (C), (C); Wakhi PERIC (A), (A);

year: Shughni sōl (G), (G); Yazgulyam sāuza (H), (H); Larestani sāl (B), (B); Old Persian ūard- (B), (B); Zebaki sāl (I), (I); Zazaki seri (B), (B); Waziri KOL (C), (C); Wakhi SOL (B), (B); Bashkardi sŏr/sur (B), (B); Ossetic afaedz (D), (D); Ormuri čan,sul (E), (E); Yaghnobi sol,yōso (J), (J); Balochi SAL (B), (B); Ishkashmi sāl (F), (F); Persian SAL (B), (B); Sogdian srōd (A), (A); Bandari sāl (B), (B); Pashto KAL (C), (C); Avestan yāro (A), (A); Parachi sa? (B), (B); Kumzari sāl (B), (B);

yellow: Kumzari zurd (A), (A); Ormuri zyē? (A), (A); Shughni zīrd (A), (A); Ossetic bur (B), (B); Balochi ZARD (A), (A); Avestan zairitō, zairiš (A), (A); Pashto ZER (A), (A); Waziri ZYER (A), (A); Wakhi ZERT (A), (A); Kurdish zer (A), (A); Persian ZARD (A), (A); Yaghnobi zard (A), (A); Larestani zard (A), (A); Khotanese Saka ysarūna-(A), (A); Sogdian zṛtyḥ (A), (A);

you: Shughni tama (A), (A); Waziri TUS, TOSE (A), (A); Ossetic cmax, cmxtae (A), (A); Bashkardi šomā (A), (A); Wakhi SUST (A), (A); Persian shoma (A), (A); Bandari šomā (A), (A); Yaghnobi šmox (A), (A); Larestani ‘amatū (lit. “all of you”), šomayā (A), (A); Kumzari shumā’ (A), (A); Balochi SHA (A), (A); Zazaki šimā (A), (A); Avestan yūžom, vō (encl.) (A), (A); Pashto TASI (A), (A); Sogdian šm’x (A), (A);

Typological/recurrent characters

Features: t > r; sw > sp; t, th, d > l; Retroflexion?; inherited!; traces of theta not lost (as obstruent); theta intact; Ezafe?; Dep-marking ezafe; Head-marking ezafe; Inflected ezafe; <cy> ezafe; ezafe = derivational suffix; Collective -tā; 3pl in -r-; Causative in āwaya; me-imperfective; bi-?; bi = subjunctive; bi = perfective; -ra < rādiy ?; ra marks IO?; ra marks DO?; ERG?; OBL; ENCL; VB-ENCL; CONSTITUENT ORDER sov = 1; hai; Sai
<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avestan</td>
<td>0100?110?????00000??0??1110111</td>
</tr>
<tr>
<td>Old Persian</td>
<td>0000?110?????00000??0??1110101</td>
</tr>
<tr>
<td>Wakhi</td>
<td>00011000????00100??0??11111??</td>
</tr>
<tr>
<td>Balochi</td>
<td>000100?101001000110111101101</td>
</tr>
<tr>
<td>Kurdish</td>
<td>0000?0?11000000001100110110</td>
</tr>
<tr>
<td>Zazaki</td>
<td>0100?0?110100000011011011001</td>
</tr>
<tr>
<td>New Persian</td>
<td>0000?0?11000000011101010??0101</td>
</tr>
<tr>
<td>Kumzari</td>
<td>1000?0?1100000001110???0??0110</td>
</tr>
<tr>
<td>Bashkardi</td>
<td>1000?0?1100000001110???0??0110</td>
</tr>
<tr>
<td>Larestani</td>
<td>0000?0?1100000001110???0??0110</td>
</tr>
<tr>
<td>Bandari</td>
<td>0000?0?1100000001110???0??0110</td>
</tr>
<tr>
<td>Mazandarani</td>
<td>0100?0?10100000011011110??0101</td>
</tr>
<tr>
<td>Pashto</td>
<td>011110?0?????00000??0??11101??</td>
</tr>
<tr>
<td>Ossetic</td>
<td>0100?100?????00000??0??11101??</td>
</tr>
<tr>
<td>Ormuri</td>
<td>010110?0?????10000??0??10011??</td>
</tr>
<tr>
<td>Parachi</td>
<td>010110?0?????00000??0??11101??</td>
</tr>
<tr>
<td>Ishkashmi</td>
<td>01111100?????00000??0??11101??</td>
</tr>
<tr>
<td>Shughni</td>
<td>0100?110?????00000??0??11111??</td>
</tr>
<tr>
<td>Yazgulyam</td>
<td>0100?110?????00000??0??11111??</td>
</tr>
<tr>
<td>Yaghnobi</td>
<td>0100?100?????00000??0??11111??</td>
</tr>
<tr>
<td>Sogdian</td>
<td>0100?110?????00000??0??11111??</td>
</tr>
<tr>
<td>Middle Persian</td>
<td>0000?0?0?????01000??0??11101??</td>
</tr>
<tr>
<td>Parthian</td>
<td>0100?0?0?????10000??0??1110101</td>
</tr>
<tr>
<td>Khwarezmian</td>
<td>0100?111100000000?11111110101</td>
</tr>
<tr>
<td>Bactrian</td>
<td>0110?0?11001000000??0??1110101</td>
</tr>
<tr>
<td>Khotanese Saka</td>
<td>000110?0?????01100??0??11001??</td>
</tr>
</tbody>
</table>
Appendix B: Chapter 4 Supplementary Materials

This appendix presents relevant aspects of the historical phonology of West Iranian languages observed in this Chapter 4, crucial to our understanding what irregular changes have taken place in the functional items observed in this chapter. This information was used to generate the model’s EXPECTED forms.

Persian

Persian is attested at three chronological stages. The following changes characterize the development from Proto-Iranian to Old Persian (Schmitt 1989a; Skjærvø 2009b):

1. Epenthesis of consonant + glide clusters: PIr *
\[ ^{\text{i}} \text{j}, \] *y > OP iy, uv / C_

2. Stress assignment: Primary stress is assigned to the rightmost non-final heavy syllable; if all syllables are light, stress is assigned to the antepenult (and in some cases, perhaps the pre-antepenult)

The following changes characterize the development from Old Persian to Middle Persian (Klingenschmitt 2000; Skjærvø 2009a; Cantera 2009):

1. Intervocalic lenition:
   (a) OP -b-, -d-, -g- > MP -w(-), -y(-), -y(-)
   (b) OP -p-, -t-, -k- > MP -b(-), -d(-), -g(-)

2. Syncope rule: a short, unaccented vowel in the penultimate syllable is syncopated between a non-obstruent and a plosive, or two identical plosives (Klingenschmitt 2000:210), e.g., *źárita- > zard ‘gold’

3. Elision of final rimes (in words with more than one syllable)

4. Pre-MP *wk > MP k, Pre-MP *wč > MP č, e.g., *pāva-ka- > *pāwk > MP pāk ‘pure’, OP *naiba-ka- > *naiwaka > *naiwk > MP nēk ‘good’
5. Monophthongization and vowel contraction: OP \(au > MP \tilde{a};\) OP \(ai, a(h)ya, iya > MP \tilde{e}\)

6. OP \(j > z;\) OP \(-j, -c > MP -z\)

7. OP \(viy- > MP gy-, vi- > gu-\) before other consonants

8. OP \(h > MP \emptyset\) at the beginning of syllables without primary or secondary stress, e.g.,
   OP \(*hizw\dot{a}na- > (dialectal OP hiz\dot{a}na-) > MP izw\dot{a}n/uzw\dot{a}n ‘tongue’\)

The following changes characterize the development from Middle Persian to New Persian (based on data from MacKenzie 1971a):

1. MP \(-\ddot{V}y, -\ddot{V}w > \ddot{V}\) (final glides sometimes preserved via analogical change or dialect admixture)

2. MP \(gy- > NP j-\)

3. MP \(w-, wi- > NP b-, gu-\)

4. MP \(\emptyset \Rightarrow NP h-\) in some words (sporadic “Cockney-style” prothesis)

5. MP \(-\ddot{u}g > NP -\ddot{i}, MP -\ddot{u}g > NP -\ddot{u}, MP -og > NP -o, MP -ag > NP -a(h)\)

**Historical phonology of other West Iranian languages: relevant details**

All of the West Iranian languages observed in this chapter undergo monophthongization of diphthongs and loss of final rimes (in words with more than one syllable).

**Parthian**

Parthian, the other Middle West Iranian language, differs from Middle Persian in crucial ways in its historical phonology (based on data from Durkin-Meisterernst 2004):

1. It does not undergo the syncope rule described for MP

2. PIr \(*\ddot{c}- > \ddot{z}\)

3. PIr \(*\ddot{u}- > b-\)

4. PIr \(*-b-, *-d-, *-g- > \beta, \delta, \gamma\)

\[\text{1This is separate from a similar-looking earlier development whereby OP } \emptyset \Rightarrow MP x.\]
Balochi

Balochi differs from New Persian in its historical phonology in the following ways (Korn 2003, 2005):

1. Like Parthian, it does not undergo the syncope rule, e.g., PIr *madaka- > Bal madag 'locust' (cf. NP maiq; Bal -g is irregular; see Korn 2005:164)
2. PIr *-č- is retained as č
3. PIr *-dy- > b (devoiced to p in iptī ‘second’)
4. PIr *-b-, *-d-, *-g- are retained as b, d, g
5. PIr *-p-, *-t-, *-k- are retained as p, t, k
6. PIr *ǎ(h)i> a

Kurdish

Kurdish undergoes the following relevant developments (?):

1. It undergoes the syncope rule
2. PIr *w̄i> gy- (e.g., gyān ‘soul’)
3. PIr *-č> ž
4. PIr *dy- > d
5. PIr *-d-, *-g- > y (between identical low vowels), Ø
6. PIr *-p-, *-t-, *-k- > w, y, g
7. PIr *ǎ(h)i> e

Mazandarani

Mazandarani undergoes the following relevant developments (based on data from Nawata 1984; Borjian and Borjian 2008):

1. It undergoes the syncope rule
2. PIr *-č- > j
3. PIr *dy- > b
4. PIr *-d- > Ø

125
5. PIr *-p-, *-t-, *-k- > w (between identical low vowels), g (between identical low vowels),
g  
6. PIr *˘¯ a(h)i > ā  
7. PIr *-a- is lengthened to ā in some contexts

Zazaki

Zazaki undergoes the following relevant developments (Paul 1998a):

1. It undergoes the syncope rule  
2. PIr *ui- > g- (e.g., gān ‘soul’)  
3. PIr *-c- > ẑ  
4. PIr *du- > b  
5. PIr *-d-, *-g- > y (between identical low vowels), γ (between identical low vowels)  
6. PIr *-p-, *-t-, *-k- > w, y, g  
7. PIr *˘¯ a(h)i > ā  
8. PIr *-a- > e [ɛ]

Sangesari

Sangesari appears to undergo a syncope rule, but this rule does not feed a rule simplifying *wk and *wč clusters (as in Persian), e.g., hauču ‘something’. Other relevant aspects of Sangesari historical phonology are (Azami and Windfuhr 1972:34-43):

1. PIr *-č- > ẑ  
2. PIr *du- > b  
3. PIr *-b-, *-d-, *-g- > v, y, g  
4. PIr *-p-, *-t-, *-k- > v, y, g  
5. PIr *˘¯ a(h)i > ā
**Yazdi**

Yazdi undergoes the following relevant developments (Vahman and Asatrian 2002:19-26):

1. It undergoes the syncope rule
2. PIr *-č- > j
3. PIr *du- > b
4. PIr *-b-, *-d-, *-g- > w, d, g
5. PIr *b-, *d-, *g- > b-, d-, y-
6. PIr *-p-, *-t-, *-k- > w, t, g
7. PIr *p-, *t-, *k- > b, t, x
8. PIr *˘a(h)i- > ā
9. PIr *-a- > a (often written ā [æ])

**Awromani**

Awromani undergoes the following relevant developments (Benedictsen and Christensen 1921:20-8):

1. It undergoes the syncope rule
2. PIr *wi- > gy- (e.g., gyān ‘soul’)
3. PIr *-č- > j
4. PIr *du- > b
5. PIr *-b-, *-d-, *-g- > w (between identical low vowels), y (between identical low vowels), γ
6. PIr *-p-, *-t-, *-k- > w, d, γ
7. PIr *p-, *t-, *k- > b, t, x
8. PIr *˘a(h)i- > ā
9. PIr *-a- > a (often written ā [æ])
Table 6.1: Distance measures between expected and observed functional items

<table>
<thead>
<tr>
<th>Lang</th>
<th>Fuc</th>
<th>Obs</th>
<th>Exp</th>
<th>QC</th>
<th>LD</th>
<th>RLD</th>
<th>LLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP</td>
<td>object Ay</td>
<td>Ay</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MP</td>
<td>comp</td>
<td>kE</td>
<td>kE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MP</td>
<td>something</td>
<td>Ec</td>
<td>Ec</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MP</td>
<td>here</td>
<td>En</td>
<td>gyAg</td>
<td>hEn</td>
<td>gyAg</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MP</td>
<td>from</td>
<td>az</td>
<td>haz</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>MP</td>
<td>if</td>
<td>agar</td>
<td>hagar</td>
<td>1</td>
<td>1</td>
<td>0.1428571429</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>MP</td>
<td>'with' ba</td>
<td>pad</td>
<td>pad</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MP</td>
<td>other</td>
<td>dUdIgar</td>
<td>dUdIgar</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MP</td>
<td>imperfective</td>
<td>hamEv</td>
<td>hamEv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MP</td>
<td>adversative</td>
<td>mA</td>
<td>agar</td>
<td>mA</td>
<td>hagar</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MP</td>
<td>one</td>
<td>yak</td>
<td>Ek</td>
<td>2</td>
<td>2</td>
<td>0.3333333333</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>MP</td>
<td>into</td>
<td>andar</td>
<td>andar</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MP</td>
<td>every</td>
<td>harv</td>
<td>harv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>object</td>
<td>rAD</td>
<td>rAD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>comp</td>
<td>kE</td>
<td>kE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>something</td>
<td>EviZ</td>
<td>EviZ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>from</td>
<td>aZ</td>
<td>haZ</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Pth</td>
<td>if</td>
<td>ag</td>
<td>hagar</td>
<td>2</td>
<td>3</td>
<td>0.4285714286</td>
<td>1.3862943611</td>
</tr>
<tr>
<td>Pth</td>
<td>'with' ba</td>
<td>pad</td>
<td>pad</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>other</td>
<td>bidIgar</td>
<td>bidIgar</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>imperfective</td>
<td>hamEv</td>
<td>hamEv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>one</td>
<td>Ev</td>
<td>Ev</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>into</td>
<td>andar</td>
<td>andar</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pth</td>
<td>every</td>
<td>harv</td>
<td>harv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NP</td>
<td>object</td>
<td>rA</td>
<td>rA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NP</td>
<td>comp</td>
<td>ki</td>
<td>kI</td>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>NP</td>
<td>something</td>
<td>lc</td>
<td>lc</td>
<td>1</td>
<td>1</td>
<td>0.1666666667</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>NP</td>
<td>here</td>
<td>injA</td>
<td>hInjA</td>
<td>2</td>
<td>2</td>
<td>0.1666666667</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>NP</td>
<td>from</td>
<td>az</td>
<td>haz</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>NP</td>
<td>if</td>
<td>agar</td>
<td>hagar</td>
<td>1</td>
<td>1</td>
<td>0.1428571429</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>NP</td>
<td>'with' ba</td>
<td>ba</td>
<td>pad</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>NP</td>
<td>other</td>
<td>digar</td>
<td>dUdIgar</td>
<td>1</td>
<td>3</td>
<td>0.2307692308</td>
<td>1.3862943611</td>
</tr>
<tr>
<td>NP</td>
<td>imperfective</td>
<td>mE</td>
<td>hamE</td>
<td>1</td>
<td>2</td>
<td>0.25</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>NP</td>
<td>adversative</td>
<td>magar</td>
<td>mA</td>
<td>hagar</td>
<td>2</td>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>NP</td>
<td>one</td>
<td>yak</td>
<td>Ik</td>
<td>2</td>
<td>2</td>
<td>0.3333333333</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>NP</td>
<td>into</td>
<td>dar</td>
<td>andar</td>
<td>1</td>
<td>2</td>
<td>0.4</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>NP</td>
<td>every</td>
<td>har</td>
<td>har</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bal</td>
<td>object</td>
<td>rA</td>
<td>rAd</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Bal</td>
<td>comp</td>
<td>ki</td>
<td>kA</td>
<td>2</td>
<td>1</td>
<td>0.2</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Bal</td>
<td>something</td>
<td>hc</td>
<td>Evac</td>
<td>4</td>
<td>3</td>
<td>0.5</td>
<td>1.3862943611</td>
</tr>
<tr>
<td>Bal</td>
<td>from</td>
<td>ac</td>
<td>hac</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Bal</td>
<td>if</td>
<td>agar</td>
<td>hagar</td>
<td>2</td>
<td>2</td>
<td>0.2857142857</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Bal</td>
<td>'with' ba</td>
<td>pa</td>
<td>pat</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Bal</td>
<td>other</td>
<td>digar</td>
<td>pitlkar</td>
<td>3</td>
<td>4</td>
<td>0.3076923077</td>
<td>1.6094379124</td>
</tr>
<tr>
<td>Bal</td>
<td>adversative</td>
<td>magar</td>
<td>mA</td>
<td>hagar</td>
<td>3</td>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>Bal</td>
<td>one</td>
<td>yak</td>
<td>Evak</td>
<td>4</td>
<td>2</td>
<td>0.3333333333</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Kd</td>
<td>comp</td>
<td>ka</td>
<td>ka</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kd</td>
<td>something</td>
<td>hc</td>
<td>lc</td>
<td>1</td>
<td>1</td>
<td>0.1666666667</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Keyword</td>
<td>Description</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>here</td>
<td>injA</td>
<td>hIngyA</td>
<td>2</td>
<td>4</td>
<td>0.3333333333</td>
<td>1.6094379124</td>
</tr>
<tr>
<td>Kd</td>
<td>from</td>
<td>Z</td>
<td>haZ</td>
<td>1</td>
<td>2</td>
<td>0.5</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Kd</td>
<td>if</td>
<td>agar</td>
<td>hagar</td>
<td>1</td>
<td>1</td>
<td>0.1428571429</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Kd</td>
<td>'with' ba</td>
<td>ba</td>
<td>pay</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Kd</td>
<td>adversative</td>
<td>magar</td>
<td>mA hagar</td>
<td>2</td>
<td>3</td>
<td>0.3</td>
<td>1.3862943611</td>
</tr>
<tr>
<td>Kd</td>
<td>one</td>
<td>yak</td>
<td>Ek</td>
<td>2</td>
<td>2</td>
<td>0.3333333333</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Kd</td>
<td>every</td>
<td>har</td>
<td>har</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maz</td>
<td>object</td>
<td>ra</td>
<td>ra</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maz</td>
<td>comp</td>
<td>ke</td>
<td>kE</td>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Maz</td>
<td>something</td>
<td>hacci</td>
<td>acci</td>
<td>1</td>
<td>1</td>
<td>0.1666666667</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Maz</td>
<td>here</td>
<td>InjA</td>
<td>hInjA</td>
<td>1</td>
<td>1</td>
<td>0.0833333333</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Maz</td>
<td>from</td>
<td>jA</td>
<td>haj</td>
<td>1</td>
<td>3</td>
<td>0.75</td>
<td>1.3862943611</td>
</tr>
<tr>
<td>Maz</td>
<td>if</td>
<td>Agar</td>
<td>hAgar</td>
<td>1</td>
<td>1</td>
<td>0.1428571429</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Maz</td>
<td>'with' ba</td>
<td>ba</td>
<td>pay</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Maz</td>
<td>adversative</td>
<td>mAgar</td>
<td>mA hagar</td>
<td>2</td>
<td>3</td>
<td>0.3</td>
<td>1.3862943611</td>
</tr>
<tr>
<td>Maz</td>
<td>one</td>
<td>yak</td>
<td>atta</td>
<td>3</td>
<td>4</td>
<td>0.6666666667</td>
<td>1.6094379124</td>
</tr>
<tr>
<td>Maz</td>
<td>into</td>
<td>dar</td>
<td>andar</td>
<td>1</td>
<td>2</td>
<td>0.4</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Maz</td>
<td>every</td>
<td>har</td>
<td>har</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Zaz</td>
<td>object</td>
<td>rA</td>
<td>rA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Zaz</td>
<td>comp</td>
<td>ki</td>
<td>kA</td>
<td>2</td>
<td>1</td>
<td>0.2</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Zaz</td>
<td>here</td>
<td>InjA</td>
<td>hInjA</td>
<td>3</td>
<td>2</td>
<td>0.1666666667</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Zaz</td>
<td>if</td>
<td>eger</td>
<td>heger</td>
<td>1</td>
<td>1</td>
<td>0.1428571429</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Zaz</td>
<td>'with' ba</td>
<td>bi</td>
<td>pa</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Zaz</td>
<td>one</td>
<td>Zev</td>
<td>Zev</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Zaz</td>
<td>every</td>
<td>her</td>
<td>her</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sang</td>
<td>object</td>
<td>re</td>
<td>re</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sang</td>
<td>comp</td>
<td>ku</td>
<td>kU</td>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Sang</td>
<td>something</td>
<td>havecu</td>
<td>avecu</td>
<td>1</td>
<td>1</td>
<td>0.1666666667</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Sang</td>
<td>here</td>
<td>anjU</td>
<td>hanjU</td>
<td>1</td>
<td>1</td>
<td>0.0833333333</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Sang</td>
<td>from</td>
<td>az</td>
<td>haZ</td>
<td>1</td>
<td>2</td>
<td>0.5</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Sang</td>
<td>if</td>
<td>aga</td>
<td>haga</td>
<td>1</td>
<td>1</td>
<td>0.1428571429</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Sang</td>
<td>'with' ba</td>
<td>ba</td>
<td>pe</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Sang</td>
<td>other</td>
<td>dlgar</td>
<td>bidIgar</td>
<td>1</td>
<td>2</td>
<td>0.1538461538</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Sang</td>
<td>imperfective</td>
<td>mI</td>
<td>hamI</td>
<td>1</td>
<td>2</td>
<td>0.25</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Sang</td>
<td>adversative</td>
<td>maga</td>
<td>mA hagar</td>
<td>2</td>
<td>4</td>
<td>0.4</td>
<td>1.6094379124</td>
</tr>
<tr>
<td>Sang</td>
<td>one</td>
<td>yak</td>
<td>Evk</td>
<td>3</td>
<td>2</td>
<td>0.3333333333</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Sang</td>
<td>into</td>
<td>de</td>
<td>ande</td>
<td>1</td>
<td>2</td>
<td>0.4</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Yazdi</td>
<td>object</td>
<td>ra</td>
<td>rad</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Yazdi</td>
<td>comp</td>
<td>gi</td>
<td>xl</td>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Yazdi</td>
<td>something</td>
<td>heS</td>
<td>eS</td>
<td>1</td>
<td>1</td>
<td>0.1666666667</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Yazdi</td>
<td>from</td>
<td>az</td>
<td>haj</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Yazdi</td>
<td>if</td>
<td>agar</td>
<td>hagar</td>
<td>1</td>
<td>1</td>
<td>0.1428571429</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Yazdi</td>
<td>'with' ba</td>
<td>ba</td>
<td>bat</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Yazdi</td>
<td>imperfective</td>
<td>e</td>
<td>hame</td>
<td>2</td>
<td>3</td>
<td>0.375</td>
<td>1.3862943611</td>
</tr>
<tr>
<td>Yazdi</td>
<td>adversative</td>
<td>magar</td>
<td>mA hagar</td>
<td>2</td>
<td>3</td>
<td>0.3</td>
<td>1.3862943611</td>
</tr>
<tr>
<td>Yazdi</td>
<td>one</td>
<td>yak</td>
<td>Ek</td>
<td>2</td>
<td>2</td>
<td>0.3333333333</td>
<td>1.0986122887</td>
</tr>
<tr>
<td>Yazdi</td>
<td>into</td>
<td>dar</td>
<td>edar</td>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Yazdi</td>
<td>every</td>
<td>har</td>
<td>har</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Awromani</td>
<td>object</td>
<td>ra</td>
<td>ra</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Awromani</td>
<td>comp</td>
<td>ka</td>
<td>kE</td>
<td>2</td>
<td>1</td>
<td>0.2</td>
<td>0.6931471806</td>
</tr>
<tr>
<td>Awromani</td>
<td>something</td>
<td>hlc</td>
<td>lc</td>
<td>1</td>
<td>1</td>
<td>0.1666666667</td>
<td>0.6931471806</td>
</tr>
</tbody>
</table>
Awromani here inja hIngya 3 4 0.3333333333 1.6094379124
Awromani from jA haj 2 3 0.75 1.3862943611
Awromani if agar har 2 2 0.2857142857 1.0986122887
Awromani 'with' ba ba paD 2 2 0.5 1.0986122887
Awromani imperfective ma hama 1 2 0.25 1.0986122887
Awromani one yak yoA 2 2 0.3333333333 1.0986122887

Table 6.2: Distances between Proto-Iranian etyma and reflexes, in lossless ASCII form

<table>
<thead>
<tr>
<th>Language</th>
<th>Proto-Iranian</th>
<th>Meaning</th>
<th>Reflex</th>
<th>LD</th>
<th>RLD</th>
<th>LLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>paca</td>
<td>cook</td>
<td>paz</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>kapautaka</td>
<td>blue/gray/pigeon</td>
<td>kabod</td>
<td>7</td>
<td>0.7777778</td>
<td>2.0794415</td>
</tr>
<tr>
<td>NP</td>
<td>taca</td>
<td>run</td>
<td>tAz</td>
<td>3</td>
<td>0.75</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>bUta</td>
<td>be</td>
<td>bUd</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>huSka</td>
<td>dry</td>
<td>xuSk</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>brAtar</td>
<td>brother</td>
<td>barAdar</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>xumba</td>
<td>pot</td>
<td>xum</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>dauca</td>
<td>sew</td>
<td>dOz</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
</tr>
<tr>
<td>NP</td>
<td>dantan</td>
<td>tooth</td>
<td>dandAn</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>xSaudaya</td>
<td>hunger</td>
<td>sOy</td>
<td>7</td>
<td>0.875</td>
<td>2.0794415</td>
</tr>
<tr>
<td>NP</td>
<td>pAda</td>
<td>foot</td>
<td>pAy</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>nguSa</td>
<td>listen</td>
<td>niyOS</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>kaufa</td>
<td>mountain</td>
<td>kOh</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
</tr>
<tr>
<td>NP</td>
<td>kafa</td>
<td>fall</td>
<td>kaf</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
</tr>
<tr>
<td>NP</td>
<td>paTana</td>
<td>bridge</td>
<td>pahn</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>gUTa</td>
<td>excrement</td>
<td>gUh</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>macTa</td>
<td>stay/live</td>
<td>mEha</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
</tr>
<tr>
<td>NP</td>
<td>xAda</td>
<td>bite/break</td>
<td>xAy</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>xAnya</td>
<td>well</td>
<td>xAnI</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>xara</td>
<td>donkey</td>
<td>xar</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
</tr>
<tr>
<td>NP</td>
<td>cara</td>
<td>move</td>
<td>car</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
</tr>
<tr>
<td>NP</td>
<td>caSman</td>
<td>eye</td>
<td>caSm</td>
<td>2</td>
<td>0.3333333</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>cina</td>
<td>choose</td>
<td>cIn</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>winin</td>
<td>select</td>
<td>guzIn</td>
<td>5</td>
<td>1</td>
<td>1.7917595</td>
</tr>
<tr>
<td>NP</td>
<td>rauca</td>
<td>light/day</td>
<td>roz</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
</tr>
<tr>
<td>NP</td>
<td>jana</td>
<td>offspring</td>
<td>zan</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>jani</td>
<td>woman</td>
<td>zan</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>buja</td>
<td>open</td>
<td>boz</td>
<td>3</td>
<td>0.75</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>ACuma</td>
<td>iron</td>
<td>Ahen</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>saucA</td>
<td>needle</td>
<td>sOz</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
</tr>
<tr>
<td>NP</td>
<td>paCu</td>
<td>cow</td>
<td>pah</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>wahuni</td>
<td>blood</td>
<td>xUn</td>
<td>5</td>
<td>0.8333333</td>
<td>1.7917595</td>
</tr>
<tr>
<td>NP</td>
<td>Jasta</td>
<td>hand</td>
<td>dast</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>mAh</td>
<td>moon</td>
<td>mAh</td>
<td>1</td>
<td>0.3333333</td>
<td>0.6931472</td>
</tr>
<tr>
<td>NP</td>
<td>aCanga</td>
<td>iron</td>
<td>sang</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>wAta</td>
<td>wind</td>
<td>bAd</td>
<td>3</td>
<td>0.75</td>
<td>1.3862944</td>
</tr>
<tr>
<td>NP</td>
<td>zRD</td>
<td>heart</td>
<td>dil</td>
<td>4</td>
<td>1.3333333</td>
<td>1.6094379</td>
</tr>
<tr>
<td>NP</td>
<td>waic</td>
<td>select</td>
<td>bIz</td>
<td>4</td>
<td>1</td>
<td>1.6094379</td>
</tr>
<tr>
<td>NP</td>
<td>dUta</td>
<td>smoke</td>
<td>dUd</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
</tr>
<tr>
<td>NP</td>
<td>Ayaka</td>
<td>egg</td>
<td>xAy</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
</tr>
</tbody>
</table>

130
<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>nAma</td>
<td>name</td>
<td>nAm</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>JAn</td>
<td>know</td>
<td>dAn</td>
<td>2</td>
<td>0.6666667</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>JAmAtR</td>
<td>mother-in-law</td>
<td>dAmAd</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>Traya</td>
<td>three</td>
<td>se</td>
<td>5</td>
<td>1</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>panca</td>
<td>five</td>
<td>panj</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>hafty</td>
<td>seven</td>
<td>haft</td>
<td>1</td>
<td>0.2</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>aSta</td>
<td>eight</td>
<td>haSt</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>pac</td>
<td>cook</td>
<td>pac</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>kap</td>
<td>blue/gray/pigeon</td>
<td>kapot</td>
<td>5</td>
<td>0.5555556</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>taca</td>
<td>run</td>
<td>tac</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>bUt</td>
<td>be</td>
<td>bUt</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>lnSk</td>
<td>dry</td>
<td>lnSk</td>
<td>1</td>
<td>0.2</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>brAt</td>
<td>brother</td>
<td>brAt</td>
<td>2</td>
<td>0.3333333</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>xumba</td>
<td>pot</td>
<td>kumb</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>dauca</td>
<td>sew</td>
<td>dOc</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>dant</td>
<td>tooth</td>
<td>dantAn</td>
<td>2</td>
<td>0.3333333</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>xSaudaya</td>
<td>hunger</td>
<td>sOd</td>
<td>7</td>
<td>0.875</td>
<td>2.0794415</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>pAda</td>
<td>foot</td>
<td>pAd</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>niguSa</td>
<td>listen</td>
<td>nigOS</td>
<td>2</td>
<td>0.3333333</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>kaufa</td>
<td>mountain</td>
<td>kOpag</td>
<td>5</td>
<td>1</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>kafa</td>
<td>fall</td>
<td>kap</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>paTana</td>
<td>bridge</td>
<td>patan</td>
<td>2</td>
<td>0.3333333</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>gUTa</td>
<td>excrement</td>
<td>gUt</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>macTea</td>
<td>stay/live</td>
<td>mEta</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>xAda</td>
<td>bite/break</td>
<td>kAd</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>xAnya</td>
<td>well</td>
<td>kAnG</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>xara</td>
<td>donkey</td>
<td>kar</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>cara</td>
<td>move</td>
<td>carr</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>caSman</td>
<td>eye</td>
<td>camm</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>cina</td>
<td>choose</td>
<td>cin</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>wicin</td>
<td>select</td>
<td>gicin</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>rauca</td>
<td>light/day</td>
<td>roc</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>jana</td>
<td>offspring</td>
<td>jan</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>jani</td>
<td>woman</td>
<td>jan</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>buja</td>
<td>open</td>
<td>boj</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>ACuna</td>
<td>iron</td>
<td>Asin</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>saCu</td>
<td>needle</td>
<td>sOc</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>paCu</td>
<td>cow</td>
<td>pas</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>wahuni</td>
<td>blood</td>
<td>gUn</td>
<td>5</td>
<td>0.8333333</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>Jasta</td>
<td>hand</td>
<td>dast</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>mAh</td>
<td>moon</td>
<td>mAh</td>
<td>1</td>
<td>0.3333333</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>aCanga</td>
<td>iron</td>
<td>sang</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>wAta</td>
<td>wind</td>
<td>gwAt</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>zRd</td>
<td>heart</td>
<td>zird</td>
<td>3</td>
<td>1</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>waic</td>
<td>select</td>
<td>gEc</td>
<td>4</td>
<td>1</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>dUta</td>
<td>smoke</td>
<td>dUt</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>Ayaka</td>
<td>egg</td>
<td>Ayak</td>
<td>1</td>
<td>0.2</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>nAma</td>
<td>name</td>
<td>nav</td>
<td>3</td>
<td>0.75</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>JAn</td>
<td>know</td>
<td>zAn</td>
<td>2</td>
<td>0.6666667</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>JAmAtR</td>
<td>mother-in-law</td>
<td>zAmAt</td>
<td>2</td>
<td>0.3333333</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>Traya</td>
<td>three</td>
<td>say</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>panca</td>
<td>five</td>
<td>panc</td>
<td>1</td>
<td>0.2</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>----</td>
<td>-----</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>hafta</td>
<td>seven</td>
<td>apt</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>aSta</td>
<td>eight</td>
<td>aSt</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>rauca</td>
<td>light/day</td>
<td>rOZ</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>caSman</td>
<td>eye</td>
<td>caw</td>
<td>4</td>
<td>0.6666667</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>kaufa</td>
<td>mountain</td>
<td>kOf</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>valhumi</td>
<td>blood</td>
<td>xwen</td>
<td>5</td>
<td>0.83333333</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>panca</td>
<td>five</td>
<td>penj</td>
<td>3</td>
<td>0.6666667</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>Apa</td>
<td>water</td>
<td>Av</td>
<td>2</td>
<td>0.6666667</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>nAma</td>
<td>name</td>
<td>nAv</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>dantan</td>
<td>tooth</td>
<td>dan</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>vyAna</td>
<td>soul</td>
<td>gyAn</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kd</td>
<td>vAta</td>
<td>wind</td>
<td>bA</td>
<td>3</td>
<td>0.75</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>rauca</td>
<td>light/day</td>
<td>rUj</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>caSman</td>
<td>eye</td>
<td>caS</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>kaufa</td>
<td>mountain</td>
<td>kOh</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>valhumi</td>
<td>blood</td>
<td>xUn</td>
<td>5</td>
<td>0.83333333</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>panca</td>
<td>five</td>
<td>panj</td>
<td>2</td>
<td>0.4</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>gauSa</td>
<td>ear</td>
<td>gUS</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>Apa</td>
<td>water</td>
<td>U</td>
<td>3</td>
<td>1</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>nAma</td>
<td>name</td>
<td>nAm</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>dantan</td>
<td>tooth</td>
<td>dAnUn</td>
<td>4</td>
<td>0.6666667</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>vyAna</td>
<td>soul</td>
<td>jAn</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>vAta</td>
<td>wind</td>
<td>vAd</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>visati</td>
<td>twenty</td>
<td>blst</td>
<td>4</td>
<td>0.6666667</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maz</td>
<td>pitar</td>
<td>father</td>
<td>per</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>rauca</td>
<td>light/day</td>
<td>rOZ</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>caSman</td>
<td>eye</td>
<td>Cim</td>
<td>5</td>
<td>0.83333333</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>kaufa</td>
<td>mountain</td>
<td>ko</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>valhumi</td>
<td>blood</td>
<td>gUnI</td>
<td>6</td>
<td>1</td>
<td>1.9459101</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>panca</td>
<td>five</td>
<td>panj</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>gauSa</td>
<td>ear</td>
<td>vlst</td>
<td>5</td>
<td>1</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>Apa</td>
<td>water</td>
<td>Aw</td>
<td>2</td>
<td>0.6666667</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>nAma</td>
<td>name</td>
<td>nAm</td>
<td>1</td>
<td>0.25</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>dantan</td>
<td>tooth</td>
<td>dindAn</td>
<td>4</td>
<td>0.6666667</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>vyAna</td>
<td>soul</td>
<td>gAn</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>vAta</td>
<td>wind</td>
<td>vA</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaz</td>
<td>visati</td>
<td>twenty</td>
<td>vlst</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>rauca</td>
<td>light/day</td>
<td>rUZ</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>caSman</td>
<td>eye</td>
<td>caS</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>kaufa</td>
<td>mountain</td>
<td>kUh</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>valhumi</td>
<td>blood</td>
<td>xUn</td>
<td>5</td>
<td>0.83333333</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>panca</td>
<td>five</td>
<td>panj</td>
<td>2</td>
<td>0.4</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>gauSa</td>
<td>ear</td>
<td>gUS</td>
<td>3</td>
<td>0.6</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>Apa</td>
<td>water</td>
<td>Ov</td>
<td>3</td>
<td>1</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>nAma</td>
<td>name</td>
<td>nUm</td>
<td>2</td>
<td>0.5</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>dantan</td>
<td>tooth</td>
<td>dannUn</td>
<td>3</td>
<td>0.5</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>vyAna</td>
<td>soul</td>
<td>Jun</td>
<td>4</td>
<td>0.8</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang</td>
<td>vAta</td>
<td>wind</td>
<td>ve</td>
<td>3</td>
<td>0.75</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>rauca</td>
<td>light/day</td>
<td>rUj</td>
<td>4.0</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>-----------</td>
<td>------</td>
<td>-----</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>paTana</td>
<td>bridge</td>
<td>pAn</td>
<td>4.6666667</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>caSman</td>
<td>eye</td>
<td>cam</td>
<td>3.0</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>vahuni</td>
<td>blood</td>
<td>xin</td>
<td>5.0</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>panca</td>
<td>five</td>
<td>panj</td>
<td>2.0</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>gauSa</td>
<td>ear</td>
<td>gUS</td>
<td>3.0</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>Apa</td>
<td>water</td>
<td>wuw</td>
<td>4.0</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>dantan</td>
<td>tooth</td>
<td>dEdu</td>
<td>5.0</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>vyAna</td>
<td>soul</td>
<td>Jun</td>
<td>4.0</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>vAta</td>
<td>wind</td>
<td>vUd</td>
<td>3.0</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>visati</td>
<td>twenty</td>
<td>vIst</td>
<td>3.0</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yzd</td>
<td>pitar</td>
<td>father</td>
<td>bder</td>
<td>5.0</td>
<td>1.7917595</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>rauca</td>
<td>light/day</td>
<td>rO</td>
<td>4.0</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>caSman</td>
<td>eye</td>
<td>cam</td>
<td>3.0</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>vahuni</td>
<td>blood</td>
<td>vun</td>
<td>3.0</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>panca</td>
<td>five</td>
<td>panj</td>
<td>2.0</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>gauSa</td>
<td>ear</td>
<td>gOs</td>
<td>4.0</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>Apa</td>
<td>water</td>
<td>Awr</td>
<td>3.0</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>nAma</td>
<td>name</td>
<td>nAm</td>
<td>1.0</td>
<td>0.6931472</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>dantan</td>
<td>tooth</td>
<td>dADAn</td>
<td>4.0</td>
<td>1.6094379</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>vyAna</td>
<td>soul</td>
<td>gyAn</td>
<td>2.0</td>
<td>1.0986123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awr</td>
<td>visati</td>
<td>twenty</td>
<td>bist</td>
<td>3.0</td>
<td>1.3862944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C: Chapter 5 Supplementary Materials

The following sources are unmarked for the following languages: New Persian (Steingass 1892); Larestani (Kamioka and Yamada 1979); Balochi (Barker 1969); Sangesari (Azami and Windfuhr 1972); Yazdi (Ivanow 1940); Zazaki (Paul 1998b); Sivandi (Lecoq 1979); S Tati (Yar-shater 1969); Kurdish (McCarus 1958); Mazandarani (Nawata 1984). CLI = Schmitt 1989b; FVT = Naşir Aşrafî 2002; Miller = Miller 1892; Rossi = Rossi 1998.

*zrd- 'heart': New Persian dil; Larestani del; Balochi zird,dil; Sangesari dal; Zazaki zerî; Sivandi del; Central Dialects N dil; S Tati del p. 154; Kurdish dil; Mazandarani del

*y(a)rda- 'flower': New Persian gul; Larestani gol; Balochi gul; Sangesari gol; Yazdi gülüw rosewater; Zazaki gul; Sivandi gol; Central Dialects N ; S Tati vela p. 68; Kurdish gul; Mazandarani gol FVT

*śprźan- 'spleen': New Persian supurz; Sangesari Ėspol; Yazdi sebul, svarz; Sivandi espel; S Tati supurz Miller; Kurdish sipil

*brź- 'high': New Persian buland,burz; Balochi burz; Sangesari belênd; Yazdi belend; Zazaki berz; Sivandi boland; Central Dialects N bilênd; S Tati bulund Miller; Kurdish barz,bilind; Mazandarani belan

*żar(i)l- 'yellow/gold': New Persian zard,dal; Larestani zard; Balochi zard; Sangesari zar; Yazdi zart; Sivandi zard; Central Dialects Q zārd

*marź- 'wipe': New Persian mālîdan; Larestani mâleda (rub); Sivandi mâl-; Central Dialects Kh māl-; Kurdish māl

*harź- 'release': New Persian hâlîdan; Zazaki erzen; Sivandi âl; Central Dialects Q hel-; S Tati harz; Kurdish pâl

*śata- 'hundred': New Persian sad; Balochi sad; Sangesari sey; Yazdi sad; Sivandi sad; Kurdish sad; Mazandarani sad

*laupāśa- 'fox': New Persian rūbāh; Balochi rūbā; Sangesari rēva; Yazdî ruwās; Sivandi rūbā; Central Dialects rūbā; S Tati luâs; Kurdish Rewi
*daśa- ‘ten’: New Persian dah; Larestani da; Balochi da; Sangesari das; Yazdi dah; Sivandi da; S Tati da(h) p. 144; Kurdish do; Mazandarani da

*kaśu-/kaši- ‘small/er’: New Persian kih; Sangesari kas; Yazdi kas

*gināśa- ‘sin’: New Persian gunāh; Balochi gunā; Yazdi gunū; Sivandi gona; Central Dialects Q gunāh

*aśu(a)na- ‘iron’: New Persian āhen; Larestani ā’en; Balochi ā(h)in; Sangesari @en; Yazdi wuhen; Zazaki hesin of iron; Sivandi āhän; Kurdish hesi; Mazandarani ā’en FVT

*śu- ‘louse/flea’: New Persian šepeš; Larestani eš; Sangesari ēspaz; Yazdi šeweš; S Tati espeja p. 69; Mazandarani ēspīj

*aśu- ‘horse’: New Persian savār,asb; Balochi asp; Sangesari asm; Yazdi asb; Sivandi usūr; Central Dialects Q ās; S Tati asb p. 127, suār horsemen, p. 39; Kurdish āsp; Mazandarani ās

*śuaka- ‘dog’: New Persian sag; Sangesari ēsp; Yazdi sve; Sivandi espe; S Tati esbi p. 80; Kurdish sag; Mazandarani sag

*kašiapa- tortoise/turtle: New Persian kašaf; Balochi kašib; Sangesari l@kpašt; Sivandi kalapošt; S Tati kasuya, kasawa p. 69; Kurdish küsel Behd

*śā- ‘black’: New Persian siyāh; Balochi siyā; Sangesari so; Yazdi siyū; Zazaki siyā; Sivandi siya; Central Dialects N siyāh; S Tati siā p. 262; Mazandarani siv

*māśia- ‘fish’: New Persian māhī; Larestani ma‘i; Balochi māi; Sangesari m@hi; Yazdi mūhi–mūsu; Zazaki māse; Sivandi moi; S Tati mohi Miller; Kurdish masi; Mazandarani ma‘i FVT

*u- ‘wind’: New Persian bād; Larestani bād; Balochi gwāt; Sangesari ve; Yazdi wūd; Zazaki vā; Swandi voy; Central Dialects vā; S Tati vār (CLI); Kurdish ba; Mazandarani va:

*yaśi, (?) ‘more’: New Persian beš; Balochi geš; Sangesari vešter; Yazdi veštar; Sivandi vištar; Central Dialects N vēšhtar
*غرطَکَا- ‘kidney’: New Persian gurda; Larestani gordakī; Balochi guTTig (Rossi); Sangesari vakku; Sivandi velk; Central Dialects N gurdä; Mazandarani gārda

*عیینِساتی- ‘twenty’: New Persian ɓıst; Larestani biss; Balochi ɓıst, bıst; Sangesari vist; Yazdi vist; Sivandi velk; S Tati vist; Kurdish bist; Mazandarani bist

*عَارَفَا- ‘rain’: New Persian bārān; Larestani baru; Balochi gwārān; Sangesari vārān; Yazdi wūrun; Zazaki vērun; Sivandi vārān; S Tati vāris (CLI); Kurdish baran; Mazandarani va:reš

*عَفَرا- ‘ice/snow’: New Persian barf; Larestani vafr, barf; Balochi barp; Sangesari varf, vafr; Yazdi wapr; Zazaki vērān; Sivandi vārān; Central Dialects Kh vârûn; S Tati vāra; Kurdish bāfr; Mazandarani varf

<table>
<thead>
<tr>
<th>Pair</th>
<th>greater_circle(m)</th>
<th>least_cost(m)</th>
<th>minute</th>
<th>all</th>
<th>*عَرَفَا/عَارَفَا</th>
<th>*سُوُیْ</th>
<th>*عَتْرَا</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP-Lar</td>
<td>932389.85</td>
<td>2540235.94</td>
<td>1540</td>
<td>4.623611</td>
<td>2.738613</td>
<td>2.473388</td>
<td>2.872281</td>
</tr>
<tr>
<td>NP-Bal</td>
<td>1576679.28</td>
<td>4145316.02</td>
<td>2420</td>
<td>5.110543</td>
<td>2.645751</td>
<td>3.162278</td>
<td>2.924988</td>
</tr>
<tr>
<td>NP-Kd</td>
<td>670979.10</td>
<td>1758158.98</td>
<td>11700</td>
<td>3.995095</td>
<td>2.148345</td>
<td>3.372684</td>
<td>0.000000</td>
</tr>
<tr>
<td>NP-Zaz</td>
<td>1098966.23</td>
<td>2619390.27</td>
<td>17760</td>
<td>6.836435</td>
<td>2.631174</td>
<td>4.379673</td>
<td>4.595650</td>
</tr>
<tr>
<td>NP-Tat</td>
<td>161104.44</td>
<td>44885.17</td>
<td>2340</td>
<td>5.964639</td>
<td>1.167748</td>
<td>3.824265</td>
<td>4.449719</td>
</tr>
<tr>
<td>NP-Yzd</td>
<td>501477.51</td>
<td>1654773.91</td>
<td>9840</td>
<td>4.966555</td>
<td>2.000000</td>
<td>2.549510</td>
<td>3.741657</td>
</tr>
<tr>
<td>NP-Maz</td>
<td>177039.08</td>
<td>1606831.67</td>
<td>7620</td>
<td>5.702206</td>
<td>2.645751</td>
<td>3.007238</td>
<td>4.062019</td>
</tr>
<tr>
<td>NP-Siv</td>
<td>638734.86</td>
<td>1576679.28</td>
<td>4145316.02</td>
<td>22200</td>
<td>4.966555</td>
<td>2.000000</td>
<td>2.549510</td>
</tr>
<tr>
<td>NP-Xun</td>
<td>301345.90</td>
<td>893874.32</td>
<td>5760</td>
<td>5.408864</td>
<td>1.936492</td>
<td>2.633122</td>
<td>4.260282</td>
</tr>
<tr>
<td>Lar-Bal</td>
<td>177862.35</td>
<td>332857.94</td>
<td>2820</td>
<td>5.876507</td>
<td>2.148345</td>
<td>3.509386</td>
<td>4.204589</td>
</tr>
<tr>
<td>Lar-Kd</td>
<td>1358365.75</td>
<td>44885.17</td>
<td>2340</td>
<td>5.964639</td>
<td>1.167748</td>
<td>3.824265</td>
<td>4.449719</td>
</tr>
<tr>
<td>Lar-Zaz</td>
<td>1809578.70</td>
<td>1654773.91</td>
<td>9840</td>
<td>4.966555</td>
<td>2.000000</td>
<td>2.549510</td>
<td>3.741657</td>
</tr>
<tr>
<td>Lar-Tat</td>
<td>1031479.07</td>
<td>1606831.67</td>
<td>7620</td>
<td>5.702206</td>
<td>2.645751</td>
<td>3.007238</td>
<td>4.062019</td>
</tr>
<tr>
<td>Lar-Yzd</td>
<td>469777.02</td>
<td>1576679.28</td>
<td>4145316.02</td>
<td>22200</td>
<td>4.966555</td>
<td>2.000000</td>
<td>2.549510</td>
</tr>
<tr>
<td>Lar-Maz</td>
<td>995827.23</td>
<td>1606831.67</td>
<td>7620</td>
<td>5.702206</td>
<td>2.645751</td>
<td>3.007238</td>
<td>4.062019</td>
</tr>
<tr>
<td>Lar-Siv</td>
<td>893874.32</td>
<td>1576679.28</td>
<td>4145316.02</td>
<td>22200</td>
<td>4.966555</td>
<td>2.000000</td>
<td>2.549510</td>
</tr>
<tr>
<td>Lar-Maz</td>
<td>995827.23</td>
<td>1606831.67</td>
<td>7620</td>
<td>5.702206</td>
<td>2.645751</td>
<td>3.007238</td>
<td>4.062019</td>
</tr>
<tr>
<td>Lar-Siv</td>
<td>893874.32</td>
<td>1576679.28</td>
<td>4145316.02</td>
<td>22200</td>
<td>4.966555</td>
<td>2.000000</td>
<td>2.549510</td>
</tr>
<tr>
<td>Lar-Maz</td>
<td>995827.23</td>
<td>1606831.67</td>
<td>7620</td>
<td>5.702206</td>
<td>2.645751</td>
<td>3.007238</td>
<td>4.062019</td>
</tr>
</tbody>
</table>

136
<table>
<thead>
<tr>
<th>Location</th>
<th>Distance</th>
<th>Elevation</th>
<th>Distance to Previous</th>
<th>Distance to Next</th>
<th>Width</th>
<th>Speed</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zaz-Tat</td>
<td>937952.08</td>
<td>2308467.45</td>
<td>15660</td>
<td>4.524167</td>
<td>2.860388</td>
<td>2.828427</td>
<td>2.074692</td>
</tr>
<tr>
<td>Zaz-Yzd</td>
<td>1524248.70</td>
<td>3527000.50</td>
<td>22620</td>
<td>4.966555</td>
<td>2.841993</td>
<td>3.308681</td>
<td>2.297825</td>
</tr>
<tr>
<td>Zaz-Siv</td>
<td>1527747.29</td>
<td>3281603.69</td>
<td>23400</td>
<td>5.581926</td>
<td>3.038218</td>
<td>3.699218</td>
<td>2.814249</td>
</tr>
<tr>
<td>Zaz-Maz</td>
<td>1215943.90</td>
<td>3523715.08</td>
<td>21120</td>
<td>4.412900</td>
<td>3.162278</td>
<td>2.633122</td>
<td>1.970369</td>
</tr>
<tr>
<td>Zaz-Sng</td>
<td>1264737.69</td>
<td>2816570.11</td>
<td>20580</td>
<td>4.645787</td>
<td>2.860388</td>
<td>3.271984</td>
<td>1.284523</td>
</tr>
</tbody>
</table>
Bibliography


